

SERIAL NUMBER	FILING DATE	CLASS	SUBCLASS	GROUP ART UNIT	EXAMINER
08/488,011	01/05/96	435	6	1801	Rees

APPLICANTS

MARK H. SKOLNICK, SALT LAKE CITY, UT; DAVID E. GOLDGAR, SALT LAKE CITY, UT; YOSHIO MIKI, SALT LAKE CITY, UT; JEFF SWENSON, SALT LAKE CITY, UT; ALEXANDER KAMB, SALT LAKE CITY, UT; KEITH D. HARSHMAN, SALT LAKE CITY, UT; JOSEPH M. SHATTUCK-EIDENS, SALT LAKE CITY, UT; SEAN V. TAVTIGIAN, SALT LAKE CITY, UT; ROGER W. WISEMAN, DURHAM, NC; P. ANDREW FUTREAL, DURHAM, NC.

CONTINUING DATA********* Rees 435/6

VERIFIED THIS APPLN IS A CIP OF 08/409,305 03/24/95
JDR 5/19/97 WHICH IS A CIP OF 08/348,824 11/29/94
WHICH IS A CIP OF 08/308,104 09/16/94
WHICH IS A CIP OF 08/300,266 09/02/94 ABN
WHICH IS A CIP OF 08/28[illegible]221 08/12/94

FOREIGN/PCT APPLICATIONS********

VERIFIED

NONE JDR 5/19/97

NOTE-DISCLAIMER
The term of this patent shall not extend beyond the expiration date of Pat. No. 5,710,001.

Certificate
JUN 30 2002
of Correction

FOREIGN FILING LICENSE GRANTED 04/05/96

Foreign priority claimed 35 USC 119 conditions met	Verified and Acknowledged	AS FILED	STATE OR COUNTRY	SHEETS DRWGS.	TOTAL CLAIMS	INDEP. CLAIMS	FILING FEE RECEIVED	ATTORNEY'S DOCKET NO.
☐ yes ☑ no ☐ yes ☑ no	Examiner's initials DR	→	UT	19	27	2	$1,034.00	24884-109347

ADDRESS

VENABLE, BAETJER, HOWARD AND CIVILETTI
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

TITLE

170-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

U.S. DEPT. OF COMM./ PAT. & TM—PTO-436L (Rev.12-94)

PARTS OF APPLICATION FILED SEPARATELY

Applications Examiner

NOTICE OF ALLOWANCE MAILED

5-22-97

DIANNE REES
Assistant Examiner

CLAIMS ALLOWED

Total Claims	Print Claim
37	1

ISSUE FEE

Amount Due	Date Paid
$1040 -	9-14-97

W. GARY JONES
SUPERVISORY PATENT EXAMINER
GROUP 1800

Primary Examiner

DRAWING

Sheets Drwg.	Figs. Drwg.	Print Fig.
18 19 26	19	4

PREPARED FOR ISSUE

ISSUE BATCH NUMBER N28

WARNING: The information disclosed herein may be restricted. Unauthorized disclosure may be prohibited by the United States Code Title 35, Sections 122, 181 and 368. Possession outside the U.S. Patent & Trademark Office is restricted to authorized employees and contractors only.

PTO-1267 (1-92) U.S. Department of Commerce Patent and Trademark Office

DISCLAIMER LABEL

Application No. 08/488,011

A terminal disclaimer has been entered and recorded under 37 U.S.C. 253 in this file.

DO NOT DESTROY

Form PTO-436A (Rev. 8/92)

ISSUE FEE IN FILE

PATENT APPLICATION

RECEIVED
96 JUN 13 PM 3:05
GROUP 180

1.	Application _____ papers.	
2.	Statement 1.82	6/19/96
3.		8-8-96
4.	Raw Seq. Listing (Cdc)	8-17-96
5.	Decl + Fee	12-1-95
6.	Notice of Int.	4-30-96
7.	Rej 3 mos	7-24-96
8.	Raw Seq. Lit.	1-5-96 7/11
9.	Amd. Disclosure	10-12-95
10.	Supple Disclosure	2-23-96
11.	IDS	5-1-96
12.	Petition to make special	6-6-96
13.	Pre Amdt A	6-6-96
14.	Amdt B	9-12-96
15.	Supple Amdt C	11-7-96
16.	Declaration	11-7-96
17.	Supple Amdt D	1-14-97
18.	F-Rej 3 Mos.	4-15-97
19.	Terminal Disclaimer	9/12/96 9/14
20.	Amdt E (NE) /Seq listing /disk	4/30/97
21.	Interview Summary	5/19/97
2.	Six MR's Amdt F &	5-22-97 5/21
23.	Raw Seq. List	5-6-97
24	Formal Drawings (26 sht.) sc. 1.	7/14/97
25.	PTO GRANT MAY 19 1998	7/14/97
26	PTO GRANT MAY 19 1998	
	Req. for C of C	2-15-02
	Director's Report	5-10-02

SEARCHED

Class	Sub.	Date	Exmr.
435	6	7/9/96	DR
435	91.1		
435	91.2		
435	7.1→7.9		
536	23.1		
536	24.3→24.33		
updated search		4/12/97	DR
updated search		5/18/97	DR
added	436	500	
	435	4	
	424	1, 11	
	436	548	
	~~387.2~~ 530 424	387.2 388.1 88	

SEARCH NOTES

	Date	Exmr.
APS, SN - see search notes in related cases, Seq ID's	7/9/96	DR
updated search	4/12/97	DR
updated search talked to A. Atzel		
updated search	5/18/97	DR

INTERFERENCE SEARCHED

Class	Sub.	Date	Exmr.
435	6 91.1, 91.2 7.1→7.9 4	5/18/97	DR
436	548 500		
530	387.2 388.1		
424	1, 11, 88		
536	23.1, 24.3, 24.33		
Seq ID's			

(RIGHT OUTSIDE)

Staple Issue Slip Here

POSITION		ID NO.	DATE
CLASSIFIER			
EXAMINER		401	8/[illegible]/95
TYPIST		412	4/5/96
VERIFIER			
CORPS CORR.			
SPEC. HAND		455	1-5-96
FILE MAINT.		452	8-17-95
DRAFTING			

INDEX OF CLAIMS

Claim Final	Claim Original	Date 7/5/96	Date 9/15/97	Date 9/19/99
1	(1)	✓	11	11
2	2			
3	3		✓	=
4	4		✓	=
	5			
5	6		✓	11
6	7		✓	11
7	8		✓	11
	9			
8	10		11	11
9	11		11	11
10	12		11	11
11	13		✓	11
12	14		11	11
	15			
13	16		11	11
14	17		11	11
15	18		11	11
16	19		11	11
	20			
	21			
	22			
	23			
	(24)			
	25			
	26	✓		
	27	✓		
17	28		11	11
18	29		11	
19	30		11	
20	(31)		11	
21	32		✓	
22	33		✓	
23	34		11	
24	35		11	
25	36		11	
26	37		11	
27	(38)		11	
28	39		11	
29	40		✓	
30	41		11	
31	42		11	
32	43		11	
33	44		11	
34	45		11	
35	46		11	
36	47		11	
37	48		✓	11
	49			
	50			

SYMBOLS

✓ ..... Rejected
= ..... Allowed
- (Through numberal) Canceled
+ ..... Restricted
N ..... Non-elected
I ..... Interference
A ..... Appeal
O ..... Objected

Claim Final	Claim Original	Date
	51	
	52	
	53	
	54	
	55	
	56	
	57	
	58	
	59	
	60	
	61	
	62	
	63	
	64	
	65	
	66	
	67	
	68	
	69	
	70	
	71	
	72	
	73	
	74	
	75	
	76	
	77	
	78	
	79	
	80	
	81	
	82	
	83	
	84	
	85	
	86	
	87	
	88	
	89	
	90	
	91	
	92	
	93	
	94	
	95	
	96	
	97	
	98	
	99	
	100	

(LEFT INSIDE)

Staple Issue Slip Here

PATENT NUMBER

APPLICATION SERIAL NUMBER
08488011

APPLICANT'S NAME (PLEASE PRINT)
SKOLNICK et al

REISSUE, ORIGINAL PATENT NUMBER

INTERNATIONAL CLASSIFICATION

C	12	Q	1	68
C	12	P	19	34
C	07	H	21	02
C	07	H	21	04

ORIGINAL CLASSIFICATION

CLASS	SUBCLASS
435	6

CROSS REFERENCE(S)

CLASS	SUBCLASS (ONE SUBCLASS PER BLOCK)			
435	91	91.2	7.1	7.2
435	7.9	4		
436	518	500		
530	387.2	388.1		
424	1.11	88		
536	23.1	24.3	24.33	

GROUP ART UNIT: 1807

ASSISTANT EXAMINER (PLEASE STAMP OR PRINT FULL NAME)
DIANNE REES

PRIMARY EXAMINER (PLEASE STAMP OR PRINT FULL NAME)
W. Gary Jones

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE

PTO 270
(REV. 5-91)

ISSUE CLASSIFICATION SLIP

Final	Original	3/5/96	9/15/97	9/19/99
1	(1)	✓	11	11
2	2			
3	3		✓=	
4	4		✓	=
	5			
5	6		11	11
6	7		✓	11
7	8		✓	11
	9			
8	10		11	11
9	11		11	11
10	12		11	11
11	13		✓	11
12	14		11	11
	15			
13	16		11	11
14	17		11	11
15	18		11	11
16	19		11	11
	20			
	21			
	22			
	23			
	(24)			
	25			
	26			
	27	✓		
17	28		11	11
18	29		11	
19	30		11	
20	(31)		11	
21	32		✓	
22	33		✓	
23	34		11	
24	35		11	
25	36		11	
26	37		11	
27	(38)		11	
28	39			
29	40		✓	
30	41		11	
31	42		11	
32	43		11	
33	44		11	
34	45		11	
35	46		11	
36	47		11	
37	48		✓	11
	49			
	50			

SYMBOLS

✓	Rejected
=	Allowed
- (Through numberal)	Canceled
+	Restricted
N	Non-elected
I	Interference
A	Appeal
O	Objected

Claim Final	Claim Original	Date
	51	
	52	
	53	
	54	
	55	
	56	
	57	
	58	
	59	
	60	
	61	
	62	
	63	
	64	
	65	
	66	
	67	
	68	
	69	
	70	
	71	
	72	
	73	
	74	
	75	
	76	
	77	
	78	
	79	
	80	
	81	
	82	
	83	
	84	
	85	
	86	
	87	
	88	
	89	
	90	
	91	
	92	
	93	
	94	
	95	
	96	
	97	
	98	
	99	
	100	

(LEFT INSIDE)

08/488011

TITLE OF THE INVENTION

17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

ABSTRACT OF THE DISCLOSURE

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast and ovarian cancer predisposing gene (BRCA1), some mutant alleles of which cause susceptibility to cancer, in particular breast and ovarian cancer. More specifically, the invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The present invention further relates to somatic mutations in the BRCA1 gene in human breast and ovarian cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

/

GeneDX 1009, pg. 6

MAIL ROOM
JUN 15 1995
PATENT & TRADEMARK OFF

08/488011

TITLE OF THE INVENTION

BRCA1-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of application Serial No. 08/409,305 filed on 24 March 1995, which is a continuation-in-part of application Serial No. 08/348,824 filed on 29 November 1994, which is a continuation-in-part of application Serial No. 08/308,104 filed on 16 September 1994, which is a continuation-in-part of application Serial No. 08/300,266, filed on 2 September 1994, which is a continuation-in-part of application Serial No. 08/289,221, filed on 12 August 1994, all incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast and ovarian cancer predisposing gene (BRCA1), some mutant alleles of which cause susceptibility to cancer, in particular, breast and ovarian cancer. More specifically, the invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The present invention further relates to somatic mutations in the BRCA1 gene in human breast and ovarian cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

The publications and other materials used herein to illuminate the background of the invention, and in particular, cases to provide additional details respecting the practice, are incorporated herein by reference, and for convenience, are referenced by author and date in the following text and respectively grouped in the appended List of References.

GeneDX 1009, pg. 7

BACKGROUND OF THE INVENTION

The genetics of cancer is complicated, involving multiple dominant, positive regulators of the transformed state (oncogenes) as well as multiple recessive, negative regulators (tumor suppressor genes). Over one hundred oncogenes have been characterized. Fewer than a dozen tumor suppressor genes have been identified, but the number is expected to increase beyond fifty (Knudson, 1993).

The involvement of so many genes underscores the complexity of the growth control mechanisms that operate in cells to maintain the integrity of normal tissue. This complexity is manifest in another way. So far, no single gene has been shown to participate in the development of all, or even the majority of human cancers. The most common oncogenic mutations are in the H-ras gene, found in 10-15% of all solid tumors (Anderson *et al.*, 1992). The most frequently mutated tumor suppressor genes are the TP53 gene, homozygously deleted in roughly 50% of all tumors, and CDKN2, which was homozygously deleted in 46% of tumor cell lines examined (Kamb *et al.*, 1994). Without a target that is common to all transformed cells, the dream of a "magic bullet" that can destroy or revert cancer cells while leaving normal tissue unharmed is improbable. The hope for a new generation of specifically targeted antitumor drugs may rest on the ability to identify tumor suppressor genes or oncogenes that play general roles in control of cell division.

The tumor suppressor genes which have been cloned and characterized influence susceptibility to: 1) Retinoblastoma (RB1); 2) Wilms' tumor (WT1); 3) Li-Fraumeni (TP53); 4) Familial adenomatous polyposis (APC); 5) Neurofibromatosis type 1 (NF1); 6) Neurofibromatosis type 2 (NF2); 7) von Hippel-Lindau syndrome (VHL); 8) Multiple endocrine neoplasia type 2A (MEN2A); and 9) Melanoma (CDKN2).

Tumor suppressor loci that have been mapped genetically but not yet isolated include genes for: Multiple endocrine neoplasia type 1 (MEN1); Lynch cancer family syndrome 2 (LCFS2); Neuroblastoma (NB); Basal cell nevus syndrome (BCNS); Beckwith-Wiedemann syndrome (BWS); Renal cell carcinoma (RCC); Tuberous sclerosis 1 (TSC1); and Tuberous sclerosis 2 (TSC2). The tumor suppressor genes that have been characterized to date encode products with similarities to a variety of protein types, including DNA binding proteins (WT1), ancillary transcription regulators (RB1), GTPase activating proteins or GAPs (NF1), cytoskeletal components (NF2), membrane bound receptor kinases (MEN2A), cell cycle regulators (CDKN2) and others with no obvious similarity to known proteins (APC and VHL).

GeneDX 1009, pg. 8

In many cases, the tumor suppressor gene originally identified through genetic studies has been shown to be lost or mutated in some sporadic tumors. This result suggests that regions of chromosomal aberration may signify the position of important tumor suppressor genes involved both in genetic predisposition to cancer and in sporadic cancer.

One of the hallmarks of several tumor suppressor genes characterized to date is that they are deleted at high frequency in certain tumor types. The deletions often involve loss of a single allele, a so-called loss of heterozygosity (LOH), but may also involve homozygous deletion of both alleles. For LOH, the remaining allele is presumed to be nonfunctional, either because of a preexisting inherited mutation, or because of a secondary sporadic mutation.

Breast cancer is one of the most significant diseases that affects women. At the current rate, American women have a 1 in 8 risk of developing breast cancer by age 95 (American Cancer Society, 1992). Treatment of breast cancer at later stages is often futile and disfiguring, making early detection a high priority in medical management of the disease. Ovarian cancer, although less frequent than breast cancer is often rapidly fatal and is the fourth most common cause of cancer mortality in American women. Genetic factors contribute to an ill-defined proportion of breast cancer incidence, estimated to be about 5% of all cases but approximately 25% of cases diagnosed before age 40 (Claus *et al.*, 1991). Breast cancer has been subdivided into two types, early-age onset and late-age onset, based on an inflection in the age-specific incidence curve around age 50. Mutation of one gene, BRCA1, is thought to account for approximately 45% of familial breast cancer, but at least 80% of families with both breast and ovarian cancer (Easton *et al.*, 1993).

Intense efforts to isolate the BRCA1 gene have proceeded since it was first mapped in 1990 (Hall *et al.*, 1990; Narod *et al.*, 1991). A second locus, BRCA2, has recently been mapped to chromosome 13q (Wooster *et al.*, 1994) and appears to account for a proportion of early-onset breast cancer roughly equal to BRCA1, but confers a lower risk of ovarian cancer. The remaining susceptibility to early-onset breast cancer is divided between as yet unmapped genes for familial cancer, and rarer germline mutations in genes such as TP53 (Malkin *et al.*, 1990). It has also been suggested that heterozygote carriers for defective forms of the Ataxia-Telangectasia gene are at higher risk for breast cancer (Swift *et al.*, 1976; Swift *et al.*, 1991). Late-age onset breast cancer is also often familial although the risks in relatives are not as high as those for early-onset breast cancer (Cannon-Albright *et al.*, 1994; Mettlin *et al.*, 1990). However, the percentage of such cases due to genetic susceptibility is unknown.

Breast cancer has long been recognized to be, in part, a familial disease (Anderson, 1972). Numerous investigators have examined the evidence for genetic inheritance and concluded that the data are most consistent with dominant inheritance for a major susceptibility locus or loci (Bishop and Gardner, 1980; Go *et al.*, 1983; Willams and Anderson, 1984; Bishop *et al.*, 1988; Newman *et al.*, 1988; Claus *et al.*, 1991). Recent results demonstrate that at least three loci exist which convey susceptibility to breast cancer as well as other cancers. These loci are the TP53 locus on chromosome 17p (Malkin *et al.*, 1990), a 17q-linked susceptibility locus known as BRCA1 (Hall *et al.*, 1990), and one or more loci responsible for the unmapped residual. Hall *et al.* (1990) indicated that the inherited breast cancer susceptibility in kindreds with early age onset is linked to chromosome 17q2l; although subsequent studies by this group using a more appropriate genetic model partially refuted the limitation to early onset breast cancer (Margaritte *et al.*, 1992).

Most strategies for cloning the 17q-linked breast cancer predisposing gene (BRCA1) require precise genetic localization studies. The simplest model for the functional role of BRCA1 holds that alleles of BRCA1 that predispose to cancer are recessive to wild type alleles; that is, cells that contain at least one wild type BRCA1 allele are not cancerous. However, cells that contain one wild type BRCA1 allele and one predisposing allele may occasionally suffer loss of the wild type allele either by random mutation or by chromosome loss during cell division (nondisjunction). All the progeny of such a mutant cell lack the wild type function of BRCA1 and may develop into tumors. According to this model, predisposing alleles of BRCA1 are recessive, yet susceptibility to cancer is inherited in a dominant fashion: women who possess one predisposing allele (and one wild type allele) risk developing cancer, because their mammary epithelial cells may spontaneously lose the wild type BRCA1 allele. This model applies to a group of cancer susceptibility loci known as tumor suppressors or antioncogenes, a class of genes that includes the retinoblastoma gene and neurofibromatosis gene. By inference this model may also explain the BRCA1 function, as has recently been suggested (Smith *et al.*, 1992).

A second possibility is that BRCA1 predisposing alleles are truly dominant; that is, a wild type allele of BRCA1 cannot overcome the tumor forming role of the predisposing allele. Thus, a cell that carries both wild type and mutant alleles would not necessarily lose the wild type copy of BRCA1 before giving rise to malignant cells. Instead, mammary cells in predisposed individuals would undergo some other stochastic change(s) leading to cancer.

If BRCA1 predisposing alleles are recessive, the BRCA1 gene is expected to be expressed in normal mammary tissue but not functionally expressed in mammary tumors. In contrast, if BRCA1 predisposing alleles are dominant, the wild type BRCA1 gene may or may not be expressed in normal mammary tissue. However, the predisposing allele will likely be expressed in breast tumor cells.

The 17q linkage of BRCA1 was independently confirmed in three of five kindreds with both breast cancer and ovarian cancer (Narod *et al.*, 1991). These studies claimed to localize the gene within a very large region, 15 centiMorgans (cM), or approximately 15 million base pairs, to either side of the linked marker pCMM86 (Dl7S74). However, attempts to define the region further by genetic studies, using markers surrounding pCMMS6, proved unsuccessful. Subsequent studies indicated that the gene was considerably more proximal (Easton *et al.*, 1993) and that the original analysis was flawed (Margaritte *et al.*, 1992). Hall *et al.*, (1992) recently localized the BRCA1 gene to an approximately 8 cM interval (approximately 8 million base pairs) bounded by Mfdl5 (Dl7S250) on the proximal side and the human GIP gene on the distal side. A slightly narrower interval for the BRCA1 locus, based on publicly available data, was agreed upon at the Chromosome 17 workshop in March of 1992 (Fain, 1992). The size of these regions and the uncertainty associated with them has made it exceedingly difficult to design and implement physical mapping and/or cloning strategies for isolating the BRCA1 gene.

Identification of a breast cancer susceptibility locus would permit the early detection of susceptible individuals and greatly increase our ability to understand the initial steps which lead to cancer. As susceptibility loci are often altered during tumor progression, cloning these genes could also be important in the development of better diagnostic and prognostic products, as well as better cancer therapies.

SUMMARY OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast cancer predisposing gene (BRCA1), some alleles of which cause susceptibility to cancer, in particular breast and ovarian cancer. More specifically, the present invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The invention further relates to somatic mutations in the BRCA1 gene in human breast cancer and their use

in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a diagram showing the order of loci neighboring BRCA1 as determined by the chromosome 17 workshop. Figure 1 is reproduced from Fain, 1992.

Figure 2 is a schematic map of YACs which define part of Mfd15-Mfd188 region.

Figure 3 is a schematic map of STSs, P1s and BACs in the BRCA1 region.

Figure 4 is a schematic map of human chromosome 17. The pertinent region containing BRCA1 is expanded to indicate the relative positions of two previously identified genes, CA125 and RNU2, BRCA1 spans the marker D17S855.

Figure 5 shows alignment of the BRCA1 zinc-finger domain with 3 other zinc-finger domains that scored highest in a Smith-Waterman alignment. RPT1 encodes a protein that appears to be a negative regulator of the IL-2 receptor in mouse. RIN1 encodes a DNA-binding protein that includes a RING-finger motif related to the zinc-finger. RFP1 encodes a putative transcription factor that is the N-terminal domain of the RET oncogene product. The bottom line contains the C3HC4 consensus zinc-finger sequence showing the positions of cysteines and one histidine that form the zinc ion binding pocket.

Figure 6 is a diagram of BRCA1 mRNA showing the locations of introns and the variants of BRCA1 mRNA produced by alternative splicing. Intron locations are shown by dark triangles and the exons are numbered below the line representing the cDNA. The top cDNA is the composite used to generate the peptide sequence of BRCA1. Alternative forms identified as cDNA clones or hybrid selection clones are shown below.

Figure 7 shows the tissue expression pattern of BRCA1. The blot was obtained from Clontech and contains RNA from the indicated tissues. Hybridization conditions were as recommended by the

manufacturer using a probe consisting of nucleotide positions 3631 to 3930 of BRCA1. Note that both breast and ovary are heterogeneous tissues and the percentage of relevant epithelial cells can be variable. Molecular weight standards are in kilobases.

Figure 8 is a diagram of the 5' untranslated region plus the beginning of the translated region of BRCA1 showing the locations of introns and the variants of BRCA1 mRNA produced by alternative splicing. Intron locations are shown by broken dashed lines. Six alternate splice forms are shown.

Figure 9A shows a nonsense mutation in Kindred 2082. P indicates the person originally screened, b and c are haplotype carriers, a, d, e, f, and g do not carry the BRCA1 haplotype. The C to T mutation results in a stop codon and creates a site for the restriction enzyme AvrII. PCR amplification products are cut with this enzyme. The carriers are heterozygous for the site and therefore show three bands. Non-carriers remain uncut.

Figure 9B shows a mutation and cosegregation analysis in BRCA1 kindreds. Carrier individuals are represented as filled circles and squares in the pedigree diagrams. Frameshift mutation in Kindred 1910. The first three lanes are control, noncarrier samples. Lanes labeled 1-3 contain sequences from carrier individuals. Lane 4 contains DNA from a kindred member who does not carry the BRCA1 mutation. The diamond is used to prevent identification of the kindred. The frameshift resulting from the additional C is apparent in lanes labeled 1, 2, and 3.

Figure 9C shows a mutation and cosegregation analysis in BRCA1 kindreds. Carrier individuals are represented as filled circles and squares in the pedigree diagrams. Inferred regulatory mutation in Kindred 2035. ASO analysis of carriers and noncarriers of 2 different polymorphisms (PM1 and PM7) which were examined for heterozygosity in the germline and compared to the heterozygosity of lymphocyte mRNA. The top 2 rows of each panel contain PCR products amplified from genomic DNA and the bottom 2 rows contain PCR products amplified from cDNA. "A" and "G" are the two alleles detected by the ASO. The dark spots indicate that a particular allele is present in the sample. The first three lanes of PM7 represent the three genotypes in the general population.

Figures 10A-10H show genomic sequence of BRCA1. The lower case letters denote intron sequence while the upper case letters denote exon sequence. Indefinite intervals within introns are designated with vvvvvvvvvvvvv. Known polymorphic sites are shown as underlined and boldface type.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast cancer predisposing gene (BRCA1), some alleles of which cause susceptibility to cancer, in particular breast and ovarian cancer. More specifically, the present invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The invention further relates to somatic mutations in the BRCA1 gene in human breast cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

The present invention provides an isolated polynucleotide comprising all, or a portion of the BRCA1 locus or of a mutated BRCA1 locus, preferably at least eight bases and not more than about 100 kb in length. Such polynucleotides may be antisense polynucleotides. The present invention also provides a recombinant construct comprising such an isolated polynucleotide, for example, a recombinant construct suitable for expression in a transformed host cell.

Also provided by the present invention are methods of detecting a polynucleotide comprising a portion of the BRCA1 locus or its expression product in an analyte. Such methods may further comprise the step of amplifying the portion of the BRCA1 locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the BRCA1 locus. The method is useful for either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer.

The present invention also provides isolated antibodies, preferably monoclonal antibodies, which specifically bind to an isolated polypeptide comprised of at least five amino acid residues encoded by the BRCA1 locus.

The present invention also provides kits for detecting in an analyte a polynucleotide comprising a portion of the BRCA1 locus, the kits comprising a polynucleotide complementary to the portion of the BRCA1 locus packaged in a suitable container, and instructions for its use.

The present invention further provides methods of preparing a polynucleotide comprising polymerizing nucleotides to yield a sequence comprised of at least eight consecutive nucleotides of the BRCA1 locus; and methods of preparing a polypeptide comprising polymerizing amino acids to yield a sequence comprising at least five amino acids encoded within the BRCA1 locus.

The present invention further provides methods of screening the BRCA1 gene to identify mutations. Such methods may further comprise the step of amplifying a portion of the BRCA1 locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the BRCA1 locus. The method is useful for identifying mutations for use in either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer.

The present invention further provides methods of screening suspected BRCA1 mutant alleles to identify mutations in the BRCA1 gene.

In addition, the present invention provides methods of screening drugs for cancer therapy to identify suitable drugs for restoring BRCA1 gene product function.

Finally, the present invention provides the means necessary for production of gene-based therapies directed at cancer cells. These therapeutic agents may take the form of polynucleotides comprising all or a portion of the BRCA1 locus placed in appropriate vectors or delivered to target cells in more direct ways such that the function of the BRCA1 protein is reconstituted. Therapeutic agents may also take the form of polypeptides based on either a portion of, or the entire protein sequence of BRCA1. These may functionally replace the activity of BRCA1 *in vivo*.

It is a discovery of the present invention that the BRCA1 locus which predisposes individuals to breast cancer and ovarian cancer, is a gene encoding a BRCA1 protein, which has been found to have no significant homology with known protein or DNA sequences. This gene is termed BRCA1 herein. It is a discovery of the present invention that mutations in the BRCA1 locus in the germline are indicative of a predisposition to breast cancer and ovarian cancer. Finally, it is a discovery of the present invention that somatic mutations in the BRCA1 locus are also associated with breast cancer, ovarian cancer and other cancers, which represents an indicator of these cancers or of the prognosis of these cancers. The mutational events of the BRCA1 locus can involve deletions, insertions and point mutations within the coding sequence and the non-coding sequence.

Starting from a region on the long arm of human chromosome 17 of the human genome, 17q, which has a size estimated at about 8 million base pairs, a region which contains a genetic locus, BRCA1, which causes susceptibility to cancer, including breast and ovarian cancer, has been identified.

The region containing the BRCA1 locus was identified using a variety of genetic techniques. Genetic mapping techniques initially defined the BRCA1 region in terms of recombination with genetic markers. Based upon studies of large extended families ("kindreds") with multiple cases of breast cancer (and ovarian cancer cases in some kindreds), a chromosomal region has been pinpointed that contains the BRCA1 gene as well as other putative susceptibility alleles in the BRCA1 locus. Two meiotic breakpoints have been discovered on the distal side of the BRCA1 locus which are expressed as recombinants between genetic markers and the disease, and one recombinant on the proximal side of the BRCA1 locus. Thus, a region which contains the BRCA1 locus is physically bounded by these markers.

The use of the genetic markers provided by this invention allowed the identification of clones which cover the region from a human yeast artificial chromosome (YAC) or a human bacterial artificial chromosome (BAC) library. It also allowed for the identification and preparation of more easily manipulated cosmid, P1 and BAC clones from this region and the construction of a contig from a subset of the clones. These cosmids, P1s, YACs and BACs provide the basis for cloning the BRCA1 locus and provide the basis for developing reagents effective, for example, in the diagnosis and treatment of breast and/or ovarian cancer. The BRCA1 gene and other potential susceptibility genes have been isolated from this region. The isolation was done using software trapping (a computational method for identifying sequences likely to contain coding exons, from contiguous or discontinuous genomic DNA sequences), hybrid selection techniques and direct screening, with whole or partial cDNA inserts from cosmids, P1s and BACs, in the region to screen cDNA libraries. These methods were used to obtain sequences of loci expressed in breast and other tissue. These candidate loci were analyzed to identify sequences which confer cancer susceptibility. We have discovered that there are mutations in the coding sequence of the BRCA1 locus in kindreds which are responsible for the 17q-linked cancer susceptibility known as BRCA1. This gene was not known to be in this region. The present invention not only facilitates the early detection of certain cancers, so vital to patient survival, but also permits the detection of susceptible individuals before they develop cancer.

GeneDX 1009, pg. 16

Population Resources

Large, well-documented Utah kindreds are especially important in providing good resources for human genetic studies. Each large kindred independently provides the power to detect whether a BRCA1 susceptibility allele is segregating in that family. Recombinants informative for localization and isolation of the BRCA1 locus could be obtained only from kindreds large enough to confirm the presence of a susceptibility allele. Large sibships are especially important for studying breast cancer, since penetrance of the BRCA1 susceptibility allele is reduced both by age and sex, making informative sibships difficult to find. Furthermore, large sibships are essential for constructing haplotypes of deceased individuals by inference from the haplotypes of their close relatives.

While other populations may also provide beneficial information, such studies generally require much greater effort, and the families are usually much smaller and thus less informative. Utah's age-adjusted breast cancer incidence is 20% lower than the average U.S. rate. The lower incidence in Utah is probably due largely to an early age at first pregnancy, increasing the probability that cases found in Utah kindreds carry a genetic predisposition.

Genetic Mapping

Given a set of informative families, genetic markers are essential for linking a disease to a region of a chromosome. Such markers include restriction fragment length polymorphisms (RFLPs) (Botstein *et al.*, 1980), markers with a variable number of tandem repeats (VNTRs) (Jeffreys *et al.*, 1985; Nakamura *et al.*, 1987), and an abundant class of DNA polymorphisms based on short tandem repeats (STRs), especially repeats of CpA (Weber and May, 1989; Litt *et al.*, 1989). To generate a genetic map, one selects potential genetic markers and tests them using DNA extracted from members of the kindreds being studied.

Genetic markers useful in searching for a genetic locus associated with a disease can be selected on an *ad hoc* basis, by densely covering a specific chromosome, or by detailed analysis of a specific region of a chromosome. A preferred method for selecting genetic markers linked with a disease involves evaluating the degree of informativeness of kindreds to determine the ideal distance between genetic markers of a given degree of polymorphism, then selecting markers from known genetic maps which are ideally spaced for maximal efficiency. Informativeness of kindreds is measured by the probability that the markers will be heterozygous in unrelated individuals. It is also most efficient to use STR markers which are detected by amplification of the target nucleic acid sequence using PCR;

GeneDX 1009, pg. 17

such markers are highly informative, easy to assay (Weber and May, 1989), and can be assayed simultaneously using multiplexing strategies (Skolnick and Wallace, 1988), greatly reducing the number of experiments required.

Once linkage has been established, one needs to find markers that flank the disease locus, i.e., one or more markers proximal to the disease locus, and one or more markers distal to the disease locus. Where possible, candidate markers can be selected from a known genetic map. Where none is known, new markers can be identified by the STR technique, as shown in the Examples.

Genetic mapping is usually an iterative process. In the present invention, it began by defining flanking genetic markers around the BRCA1 locus, then replacing these flanking markers with other markers that were successively closer to the BRCA1 locus. As an initial step, recombination events, defined by large extended kindreds, helped specifically to localize the BRCA1 locus as either distal or proximal to a specific genetic marker (Goldgar *et al.*, 1994).

The region surrounding BRCA1, until the disclosure of the present invention, was not well mapped and there were few markers. Therefore, short repetitive sequences on cosmids subcloned from YACs, which had been physically mapped, were analyzed in order to develop new genetic markers. Using this approach, one marker of the present invention, 42D6, was discovered which replaced pCMM86 as the distal flanking marker for the BRCA1 region. Since 42D6 is approximately 14 cM from pCMM86, the BRCA1 region was thus reduced by approximately 14 centiMorgans (Easton *et al.*, 1993). The present invention thus began by finding a much more closely linked distal flanking marker of the BRCA1 region. BRCA1 was then discovered to be distal to the genetic marker Mfdl5. Therefore, BRCA1 was shown to be in a region of 6 to 10 million bases bounded by Mfdl5 and 42D6. Marker Mfdl9l was subsequently discovered to be distal to Mfdl5 and proximal to BRCA1. Thus, Mfdl5 was replaced with Mfd191 as the closest proximal genetic marker. Similarly, it was discovered that genetic marker Mfdl88 could replace genetic marker 42D6, narrowing the region containing the BRCA1 locus to approximately 1.5 million bases. Then the marker Mfdl9l was replaced with tdj1474 as the proximal marker and Mfdl88 was replaced with U5R as the distal marker, further narrowing the BRCA1 region to a small enough region to allow isolation and characterization of the BRCA1 locus (see Figure 3), using techniques known in the art and described herein.

GeneDX 1009, pg. 18

Physical Mapping

Three distinct methods were employed to physically map the region. The first was the use of yeast artificial chromosomes (YACs) to clone the region which is flanked by tdj1474 and U5R. The second was the creation of a set of P1, BAC and cosmid clones which cover the region containing the BRCA1 locus.

Yeast Artificial Chromosomes (YACs). Once a sufficiently small region containing the BRCA1 locus was identified, physical isolation of the DNA in the region proceeded by identifying a set of overlapping YACs which covers the region. Useful YACs can be isolated from known libraries, such as the St. Louis and CEPH YAC libraries, which are widely distributed and contain approximately 50,000 YACs each. The YACs isolated were from these publicly accessible libraries and can be obtained from a number of sources including the Michigan Genome Center. Clearly, others who had access to these YACs, without the disclosure of the present invention, would not have known the value of the specific YACs we selected since they would not have known which YACs were within, and which YACs outside of, the smallest region containing the BRCA1 locus.

Cosmid, P1 and BAC Clones. In the present invention, it is advantageous to proceed by obtaining cosmid, P1, and BAC clones to cover this region. The smaller size of these inserts, compared to YAC inserts, makes them more useful as specific hybridization probes. Furthermore, having the cloned DNA in bacterial cells, rather than in yeast cells, greatly increases the ease with which the DNA of interest can be manipulated, and improves the signal-to-noise ratio of hybridization assays. For cosmid subclones of YACs, the DNA is partially digested with the restriction enzyme Sau3A and cloned into the BamHI site of the pWE15 cosmid vector (Stratagene, cat. #1251201). The cosmids containing human sequences are screened by hybridization with human repetitive DNA (e.g., Gibco/BRL, Human C_0t-1 DNA, cat. 5279SA), and then fingerprinted by a variety of techniques, as detailed in the Examples.

P1 and BAC clones are obtained by screening libraries constructed from the total human genome with specific sequence tagged sites (STSs) derived from the YACs, cosmids or P1s and BACs, isolated as described herein.

These P1, BAC and cosmid clones can be compared by interspersed repetitive sequence (IRS) PCR and/or restriction enzyme digests followed by gel electrophoresis and comparison of the resulting DNA fragments ("fingerprints") (Maniatis *et al.*, 1982). The clones can also be characterized by the presence of STSs. The fingerprints are used to define an overlapping contiguous set of clones which

GeneDX 1009, pg. 19

covers the region but is not excessively redundant, referred to herein as a "minimum tiling path". Such a minimum tiling path forms the basis for subsequent experiments to identify cDNAs which may originate from the BRCA1 locus.

Coverage of the Gap with Pl and BAC Clones. To cover any gaps in the BRCA1 contig between the identified cosmids with genomic clones, clones in P1 and BAC vectors which contain inserts of genomic DNA roughly twice as large as cosmids for P1s and still greater for BACs (Sternberg, 1990; Sternberg *et al.*, 1990; Pierce *et al.*, 1992; Shizuya *et al.*, 1992) were used. P1 clones were isolated by Genome Sciences using PCR primers provided by us for screening. BACs were provided by hybridization techniques in Dr. Mel Simon's laboratory. The strategy of using P1 clones also permitted the covering of the genomic region with an independent set of clones not derived from YACs. This guards against the possibility of other deletions in YACs that have not been detected. These new sequences derived from the P1 clones provide the material for further screening for candidate genes, as described below.

Gene Isolation.

There are many techniques for testing genomic clones for the presence of sequences likely to be candidates for the coding sequence of a locus one is attempting to isolate, including but not limited to:

a. zoo blots
b. identifying HTF islands
c. exon trapping
d. hybridizing cDNA to cosmids or YACs.
e. screening cDNA libraries.

(a) Zoo blots. The first technique is to hybridize cosmids to Southern blots to identify DNA sequences which are evolutionarily conserved, and which therefore give positive hybridization signals with DNA from species of varying degrees of relationship to humans (such as monkey, cow, chicken, pig, mouse and rat). Southern blots containing such DNA from a variety of species are commercially available (Clonetech, Cat. 7753-1).

(b) Identifying HTF islands. The second technique involves finding regions rich in the nucleotides C and G, which often occur near or within coding sequences. Such sequences are called

HTF (HpaI tiny fragment) or CpG islands, as restriction enzymes specific for sites which contain CpG dimers cut frequently in these regions (Lindsay *et al.*, 1987).

(c) Exon trapping. The third technique is exon trapping, a method that identifies sequences in genomic DNA which contain splice junctions and therefore are likely to comprise coding sequences of genes. Exon amplification (Buckler *et al.*, 1991) is used to select and amplify exons from DNA clones described above. Exon amplification is based on the selection of RNA sequences which are flanked by functional 5' and/or 3' splice sites. The products of the exon amplification are used to screen the breast cDNA libraries to identify a manageable number of candidate genes for further study. Exon trapping can also be performed on small segments of sequenced DNA using computer programs or by software trapping.

(d) Hybridizing cDNA to Cosmids, P1s, BACs or YACs. The fourth technique is a modification of the selective enrichment technique which utilizes hybridization of cDNA to cosmids, P1s, BACs or YACs and permits transcribed sequences to be identified in, and recovered from cloned genomic DNA (Kandpal *et al.*, 1990). The selective enrichment technique, as modified for the present purpose, involves binding DNA from the region of BRCA1 present in a YAC to a column matrix and selecting cDNAs from the relevant libraries which hybridize with the bound DNA, followed by amplification and purification of the bound DNA, resulting in a great enrichment for cDNAs in the region represented by the cloned genomic DNA.

(e) Identification of cDNAs. The fifth technique is to identify cDNAs that correspond to the BRCA1 locus. Hybridization probes containing putative coding sequences, selected using any of the above techniques, are used to screen various libraries, including breast tissue cDNA libraries, ovarian cDNA libraries, and any other necessary libraries.

Another variation on the theme of direct selection of cDNA was also used to find candidate genes for BRCA1 (Lovett *et al.*, 1991; Futreal, 1993). This method uses cosmid, P1 or BAC DNA as the probe. The probe DNA is digested with a blunt cutting restriction enzyme such as HaeIII. Double stranded adapters are then ligated onto the DNA and serve as binding sites for primers in subsequent PCR amplification reactions using biotinylated primers. Target cDNA is generated from mRNA derived from tissue samples, e.g., breast tissue, by synthesis of either random primed or oligo(dT) primed first strand followed by second strand synthesis. The cDNA ends are rendered blunt and ligated onto double-stranded adapters. These adapters serve as amplification sites for PCR. The target and probe sequences are denatured and mixed with human C_0t-1 DNA to block repetitive sequences.

Solution hybridization is carried out to high C_0t-1/2 values to ensure hybridization of rare target cDNA molecules. The annealed material is then captured on avidin beads, washed at high stringency and the retained cDNAs are eluted and amplified by PCR. The selected cDNA is subjected to further rounds of enrichment before cloning into a plasmid vector for analysis.

Testing the cDNA for Candidacy

Proof that the cDNA is the BRCA1 locus is obtained by finding sequences in DNA extracted from affected kindred members which create abnormal BRCA1 gene products or abnormal levels of BRCA1 gene product. Such BRCA1 susceptibility alleles will co-segregate with the disease in large kindreds. They will also be present at a much higher frequency in non-kindred individuals with breast and ovarian cancer then in individuals in the general population. Finally, since tumors often mutate somatically at loci which are in other instances mutated in the germline, we expect to see normal germline BRCA1 alleles mutated into sequences which are identical or similar to BRCA1 susceptibility alleles in DNA extracted from tumor tissue. Whether one is comparing BRCA1 sequences from tumor tissue to BRCA1 alleles from the germline of the same individuals, or one is comparing germline BRCA1 alleles from cancer cases to those from unaffected individuals, the key is to find mutations which are serious enough to cause obvious disruption to the normal function of the gene product. These mutations can take a number of forms. The most severe forms would be frame shift mutations or large deletions which would cause the gene to code for an abnormal protein or one which would significantly alter protein expression. Less severe disruptive mutations would include small in-frame deletions and nonconservative base pair substitutions which would have a significant effect on the protein produced, such as changes to or from a cysteine residue, from a basic to an acidic amino acid or vice versa, from a hydrophobic to hydrophilic amino acid or vice versa, or other mutations which would affect secondary, tertiary or quaternary protein structure. Silent mutations or those resulting in conservative amino acid substitutions would not generally be expected to disrupt protein function.

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type BRCA1 locus is detected. In addition, the method can be performed by detecting the wild-type BRCA1 locus and confirming the lack of a predisposition to cancer at the BRCA1 locus. "Alteration of a wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or of only a portion of the gene. Point mutations may result in stop codons, frameshift mutations or amino acid

substitutions. Somatic mutations are those which occur only in certain tissues, e.g., in the tumor tissue, and are not inherited in the germline. Germline mutations can be found in any of a body's tissues and are inherited. If only a single allele is somatically mutated, an early neoplastic state is indicated. However, if both alleles are somatically mutated, then a late neoplastic state is indicated. The finding of BRCA1 mutations thus provides both diagnostic and prognostic information. A BRCA1 allele which is not deleted (e.g., found on the sister chromosome to a chromosome carrying a BRCA1 deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. It is believed that many mutations found in tumor tissues will be those leading to decreased expression of the BRCA1 gene product. However, mutations leading to non-functional gene products would also lead to a cancerous state. Point mutational events may occur in regulatory regions, such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the BRCA1 gene product, or to a decrease in mRNA stability or translation efficiency.

Useful diagnostic techniques include, but are not limited to fluorescent *in situ* hybridization (FISH), direct DNA sequencing, PFGE analysis, Southern blot analysis, single stranded conformation analysis (SSCA), RNase protection assay, allele-specific oligonucleotide (ASO), dot blot analysis and PCR-SSCP, as discussed in detail further below.

Predisposition to cancers, such as breast and ovarian cancer, and the other cancers identified herein, can be ascertained by testing any tissue of a human for mutations of the BRCA1 gene. For example, a person who has inherited a germline BRCA1 mutation would be prone to develop cancers. This can be determined by testing DNA from any tissue of the person's body. Most simply, blood can be drawn and DNA extracted from the cells of the blood. In addition, prenatal diagnosis can be accomplished by testing fetal cells, placental cells or amniotic cells for mutations of the BRCA1 gene. Alteration of a wild-type BRCA1 allele, whether, for example, by point mutation or deletion, can be detected by any of the means discussed herein.

There are several methods that can be used to detect DNA sequence variation. Direct DNA sequencing, either manual sequencing or automated fluorescent sequencing can detect sequence variation. For a gene as large as BRCA1, manual sequencing is very labor-intensive, but under optimal conditions, mutations in the coding sequence of a gene are rarely missed. Another approach is the single-stranded conformation polymorphism assay (SSCA) (Orita *et al.*, 1989). This method does not detect all sequence changes, especially if the DNA fragment size is greater than 200 bp, but can be

optimized to detect most DNA sequence variation. The reduced detection sensitivity is a disadvantage, but the increased throughput possible with SSCA makes it an attractive, viable alternative to direct sequencing for mutation detection on a research basis. The fragments which have shifted mobility on SSCA gels are then sequenced to determine the exact nature of the DNA sequence variation. Other approaches based on the detection of mismatches between the two complementary DNA strands include clamped denaturing gel electrophoresis (CDGE) (Sheffield *et al.*, 1991), heteroduplex analysis (HA) (White *et al.*, 1992) and chemical mismatch cleavage (CMC) (Grompe *et al.*, 1989). None of the methods described above will detect large deletions, duplications or insertions, nor will they detect a regulatory mutation which affects transcription or translation of the protein. Other methods which might detect these classes of mutations such as a protein truncation assay or the asymmetric assay, detect only specific types of mutations and would not detect missense mutations. A review of currently available methods of detecting DNA sequence variation can be found in a recent review by Grompe (1993). Once a mutation is known, an allele specific detection approach such as allele specific oligonucleotide (ASO) hybridization can be utilized to rapidly screen large numbers of other samples for that same mutation.

In order to detect the alteration of the wild-type BRCA1 gene in a tissue, it is helpful to isolate the tissue free from surrounding normal tissues. Means for enriching tissue preparation for tumor cells are known in the art. For example, the tissue may be isolated from paraffin or cryostat sections. Cancer cells may also be separated from normal cells by flow cytometry. These techniques, as well as other techniques for separating tumor cells from normal cells, are well known in the art. If the tumor tissue is highly contaminated with normal cells, detection of mutations is more difficult.

A rapid preliminary analysis to detect polymorphisms in DNA sequences can be performed by looking at a series of Southern blots of DNA cut with one or more restriction enzymes, preferably with a large number of restriction enzymes. Each blot contains a series of normal individuals and a series of cancer cases, tumors, or both. Southern blots displaying hybridizing fragments (differing in length from control DNA when probed with sequences near or including the BRCA1 locus) indicate a possible mutation. If restriction enzymes which produce very large restriction fragments are used, then pulsed field gel electrophoresis (PFGE) is employed.

Detection of point mutations may be accomplished by molecular cloning of the BRCA1 allele(s) and sequencing the allele(s) using techniques well known in the art. Alternatively, the gene sequences

can be amplified directly from a genomic DNA preparation from the tumor tissue, using known techniques. The DNA sequence of the amplified sequences can then be determined.

There are six well known methods for a more complete, yet still indirect, test for confirming the presence of a susceptibility allele: 1) single stranded conformation analysis (SSCA) (Orita *et al.*, 1989); 2) denaturing gradient gel electrophoresis (DGGE) (Wartell *et al.*, 1990; Sheffield *et al.*, 1989); 3) RNase protection assays (Finkelstein *et al.*, 1990; Kinszler *et al.*, 1991); 4) allele-specific oligonucleotides (ASOs) (Conner *et al.*, 1983); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich, 1991); and 6) allele-specific PCR (Rano & Kidd, 1989). For allele-specific PCR, primers are used which hybridize at their 3' ends to a particular BRCA1 mutation. If the particular BRCA1 mutation is not present, an amplification product is not observed. Amplification Refractory Mutation System (ARMS) can also be used, as disclosed in European Patent Application Publication No. 0332435 and in Newton *et al.*, 1989. Insertions and deletions of genes can also be detected by cloning, sequencing and amplification. In addition, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to score alteration of an allele or an insertion in a polymorphic fragment. Such a method is particularly useful for screening relatives of an affected individual for the presence of the BRCA1 mutation found in that individual. Other techniques for detecting insertions and deletions as known in the art can be used.

In the first three methods (SSCA, DGGE and RNase protection assay), a new electrophoretic band appears. SSCA detects a band which migrates differentially because the sequence change causes a difference in single-strand, intramolecular base pairing. RNase protection involves cleavage of the mutant polynucleotide into two or more smaller fragments. DGGE detects differences in migration rates of mutant sequences compared to wild-type sequences, using a denaturing gradient gel. In an allele-specific oligonucleotide assay, an oligonucleotide is designed which detects a specific sequence, and the assay is performed by detecting the presence or absence of a hybridization signal. In the mutS assay, the protein binds only to sequences that contain a nucleotide mismatch in a heteroduplex between mutant and wild-type sequences.

Mismatches, according to the present invention, are hybridized nucleic acid duplexes in which the two strands are not 100% complementary. Lack of total homology may be due to deletions, insertions, inversions or substitutions. Mismatch detection can be used to detect point mutations in the gene or in its mRNA product. While these techniques are less sensitive than sequencing, they are simpler to perform on a large number of tumor samples. An example of a mismatch cleavage technique

GeneDX 1009, pg. 25

is the RNase protection method. In the practice of the present invention, the method involves the use of a labeled riboprobe which is complementary to the human wild-type BRCA1 gene coding sequence. The riboprobe and either mRNA or DNA isolated from the tumor tissue are annealed (hybridized) together and subsequently digested with the enzyme RNase A which is able to detect some mismatches in a duplex RNA structure. If a mismatch is detected by RNase A, it cleaves at the site of the mismatch. Thus, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch has been detected and cleaved by RNase A, an RNA product will be seen which is smaller than the full length duplex RNA for the riboprobe and the mRNA or DNA. The riboprobe need not be the full length of the BRCA1 mRNA or gene but can be a segment of either. If the riboprobe comprises only a segment of the BRCA1 mRNA or gene, it will be desirable to use a number of these probes to screen the whole mRNA sequence for mismatches.

In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. See, e.g., Cotton *et al.*, 1988; Shenk *et al.*, 1975; Novack *et al.*, 1986. Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. See, e.g., Cariello, 1988. With either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR (see below) before hybridization. Changes in DNA of the BRCA1 gene can also be detected using Southern hybridization, especially if the changes are gross rearrangements, such as deletions and insertions.

DNA sequences of the BRCA1 gene which have been amplified by use of PCR may also be screened using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the BRCA1 gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length, corresponding to a portion of the BRCA1 gene sequence. By use of a battery of such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously identified mutation in the BRCA1 gene. Hybridization of allele-specific probes with amplified BRCA1 sequences can be performed, for example, on a nylon filter. Hybridization to a particular probe under stringent hybridization conditions indicates the presence of the same mutation in the tumor tissue as in the allele-specific probe.

The most definitive test for mutations in a candidate locus is to directly compare genomic BRCA1 sequences from cancer patients with those from a control population. Alternatively, one could sequence messenger RNA after amplification, e.g., by PCR, thereby eliminating the necessity of determining the exon structure of the candidate gene.

Mutations from cancer patients falling outside the coding region of BRCA1 can be detected by examining the non-coding regions, such as introns and regulatory sequences near or within the BRCA1 gene. An early indication that mutations in noncoding regions are important may come from Northern blot experiments that reveal messenger RNA molecules of abnormal size or abundance in cancer patients as compared to control individuals.

Alteration of BRCA1 mRNA expression can be detected by any techniques known in the art. These include Northern blot analysis, PCR amplification and RNase protection. Diminished mRNA expression indicates an alteration of the wild-type BRCA1 gene. Alteration of wild-type BRCA1 genes can also be detected by screening for alteration of wild-type BRCA1 protein. For example, monoclonal antibodies immunoreactive with BRCA1 can be used to screen a tissue. Lack of cognate antigen would indicate a BRCA1 mutation. Antibodies specific for products of mutant alleles could also be used to detect mutant BRCA1 gene product. Such immunological assays can be done in any convenient formats known in the art. These include Western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered BRCA1 protein can be used to detect alteration of wild-type BRCA1 genes. Functional assays, such as protein binding determinations, can be used. In addition, assays can be used which detect BRCA1 biochemical function. Finding a mutant BRCA1 gene product indicates alteration of a wild-type BRCA1 gene.

Mutant BRCA1 genes or gene products can also be detected in other human body samples, such as serum, stool, urine and sputum. The same techniques discussed above for detection of mutant BRCA1 genes or gene products in tissues can be applied to other body samples. Cancer cells are sloughed off from tumors and appear in such body samples. In addition, the BRCA1 gene product itself may be secreted into the extracellular space and found in these body samples even in the absence of cancer cells. By screening such body samples, a simple early diagnosis can be achieved for many types of cancers. In addition, the progress of chemotherapy or radiotherapy can be monitored more easily by testing such body samples for mutant BRCA1 genes or gene products.

The methods of diagnosis of the present invention are applicable to any tumor in which BRCA1 has a role in tumorigenesis. The diagnostic method of the present invention is useful for clinicians, so they can decide upon an appropriate course of treatment.

The primer pairs of the present invention are useful for determination of the nucleotide sequence of a particular BRCA1 allele using PCR. The pairs of single-stranded DNA primers can be annealed to sequences within or surrounding the BRCA1 gene on chromosome 17q21 in order to prime amplifying

DNA synthesis of the BRCA1 gene itself. A complete set of these primers allows synthesis of all of the nucleotides of the BRCA1 gene coding sequences, i.e., the exons. The set of primers preferably allows synthesis of both intron and exon sequences. Allele-specific primers can also be used. Such primers anneal only to particular BRCA1 mutant alleles, and thus will only amplify a product in the presence of the mutant allele as a template.

In order to facilitate subsequent cloning of amplified sequences, primers may have restriction enzyme site sequences appended to their 5' ends. Thus, all nucleotides of the primers are derived from BRCA1 sequences or sequences adjacent to BRCA1, except for the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art. The primers themselves can be synthesized using techniques which are well known in the art. Generally, the primers can be made using oligonucleotide synthesizing machines which are commercially available. Given the sequence of the BRCA1 open reading frame shown in SEQ ID NO:1, design of particular primers is well within the skill of the art.

The nucleic acid probes provided by the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA and in the RNase protection method for detecting point mutations already discussed above. The probes can be used to detect PCR amplification products. They may also be used to detect mismatches with the BRCA1 gene or mRNA using other techniques.

It has been discovered that individuals with the wild-type BRCA1 gene do not have cancer which results from the BRCA1 allele. However, mutations which interfere with the function of the BRCA1 protein are involved in the pathogenesis of cancer. Thus, the presence of an altered (or a mutant) BRCA1 gene which produces a protein having a loss of function, or altered function, directly correlates to an increased risk of cancer. In order to detect a BRCA1 gene mutation, a biological sample is prepared and analyzed for a difference between the sequence of the BRCA1 allele being analyzed and the sequence of the wild-type BRCA1 allele. Mutant BRCA1 alleles can be initially identified by any of the techniques described above. The mutant alleles are then sequenced to identify the specific mutation of the particular mutant allele. Alternatively, mutant BRCA1 alleles can be initially identified by identifying mutant (altered) BRCA1 proteins, using conventional techniques. The mutant alleles are then sequenced to identify the specific mutation for each allele. The mutations, especially those which lead to an altered function of the BRCA1 protein, are then used for the diagnostic and prognostic methods of the present invention.

Definitions

The present invention employs the following definitions:

"**Amplification of Polynucleotides**" utilizes methods such as the polymerase chain reaction (PCR), ligation amplification (or ligase chain reaction, LCR) and amplification methods based on the use of Q-beta replicase. These methods are well known and widely practiced in the art. See, e.g., U.S. Patents 4,683,195 and 4,683,202 and Innis *et al.*, 1990 (for PCR); and Wu *et al.*, 1989a (for LCR). Reagents and hardware for conducting PCR are commercially available. Primers useful to amplify sequences from the BRCA1 region are preferably complementary to, and hybridize specifically to sequences in the BRCA1 region or in regions that flank a target region therein. BRCA1 sequences generated by amplification may be sequenced directly. Alternatively, but less desirably, the amplified sequence(s) may be cloned prior to sequence analysis. A method for the direct cloning and sequence analysis of enzymatically amplified genomic segments has been described by Scharf, 1986.

"**Analyte polynucleotide**" and "**analyte strand**" refer to a single- or double-stranded polynucleotide which is suspected of containing a target sequence, and which may be present in a variety of types of samples, including biological samples.

"**Antibodies.**" The present invention also provides polyclonal and/or monoclonal antibodies and fragments thereof, and immunologic binding equivalents thereof, which are capable of specifically binding to the BRCA1 polypeptides and fragments thereof or to polynucleotide sequences from the BRCA1 region, particularly from the BRCA1 locus or a portion thereof. The term "**antibody**" is used both to refer to a homogeneous molecular entity, or a mixture such as a serum product made up of a plurality of different molecular entities. Polypeptides may be prepared synthetically in a peptide synthesizer and coupled to a carrier molecule (e.g., keyhole limpet hemocyanin) and injected over several months into rabbits. Rabbit sera is tested for immunoreactivity to the BRCA1 polypeptide or fragment. Monoclonal antibodies may be made by injecting mice with the protein polypeptides, fusion proteins or fragments thereof. Monoclonal antibodies will be screened by ELISA and tested for specific immunoreactivity with BRCA1 polypeptide or fragments thereof. See, Harlow & Lane, 1988. These antibodies will be useful in assays as well as pharmaceuticals.

Once a sufficient quantity of desired polypeptide has been obtained, it may be used for various purposes. A typical use is the production of antibodies specific for binding. These antibodies may be either polyclonal or monoclonal, and may be produced by *in vitro* or *in vivo* techniques well known in the art. For production of polyclonal antibodies, an appropriate target immune system, typically mouse

or rabbit, is selected. Substantially purified antigen is presented to the immune system in a fashion determined by methods appropriate for the animal and by other parameters well known to immunologists. Typical sites for injection are in footpads, intramuscularly, intraperitoneally, or intradermally. Of course, other species may be substituted for mouse or rabbit. Polyclonal antibodies are then purified using techniques known in the art, adjusted for the desired specificity.

An immunological response is usually assayed with an immunoassay. Normally, such immunoassays involve some purification of a source of antigen, for example, that produced by the same cells and in the same fashion as the antigen. A variety of immunoassay methods are well known in the art. See, e.g., Harlow & Lane, 1988, or Goding, 1986.

Monoclonal antibodies with affinities of 10^{-8} M^{-1} or preferably 10^{-9} to 10^{-10} M^{-1} or stronger will typically be made by standard procedures as described, e.g., in Harlow & Lane, 1988 or Goding, 1986. Briefly, appropriate animals will be selected and the desired immunization protocol followed. After the appropriate period of time, the spleens of such animals are excised and individual spleen cells fused, typically, to immortalized myeloma cells under appropriate selection conditions. Thereafter, the cells are clonally separated and the supernatants of each clone tested for their production of an appropriate antibody specific for the desired region of the antigen.

Other suitable techniques involve *in vitro* exposure of lymphocytes to the antigenic polypeptides, or alternatively, to selection of libraries of antibodies in phage or similar vectors. See Huse *et al.*, 1989. The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, magnetic particles and the like. Patents teaching the use of such labels include U.S. Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced (see U.S. Patent 4,816,567).

"Binding partner" refers to a molecule capable of binding a ligand molecule with high specificity, as for example, an antigen and an antigen-specific antibody or an enzyme and its inhibitor. In general, the specific binding partners must bind with sufficient affinity to immobilize the analyte copy/complementary strand duplex (in the case of polynucleotide hybridization) under the isolation conditions. Specific binding partners are known in the art and include, for example, biotin and avidin

or streptavidin, IgG and protein A, the numerous, known receptor-ligand couples, and complementary polynucleotide strands. In the case of complementary polynucleotide binding partners, the partners are normally at least about 15 bases in length, and may be at least 40 bases in length. The polynucleotides may be composed of DNA, RNA, or synthetic nucleotide analogs.

A **"biological sample"** refers to a sample of tissue or fluid suspected of containing an analyte polynucleotide or polypeptide from an individual including, but not limited to, e.g., plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, blood cells, tumors, organs, tissue and samples of *in vitro* cell culture constituents.

As used herein, the terms "**diagnosing**" or "**prognosing**," as used in the context of neoplasia, are used to indicate 1) the classification of lesions as neoplasia, 2) the determination of the severity of the neoplasia, or 3) the monitoring of the disease progression, prior to, during and after treatment.

"Encode". A polynucleotide is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the mRNA for and/or the polypeptide or a fragment thereof. The anti-sense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

"Isolated" or **"substantially pure"**. An "isolated" or "substantially pure" nucleic acid (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components which naturally accompany a native human sequence or protein, e.g., ribosomes, polymerases, many other human genome sequences and proteins. The term embraces a nucleic acid sequence or protein which has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogs or analogs biologically synthesized by heterologous systems.

"BRCA1 Allele" refers to normal alleles of the BRCA1 locus as well as alleles carrying variations that predispose individuals to develop cancer of many sites including, for example, breast, ovarian, colorectal and prostate cancer. Such predisposing alleles are also called **"BRCA1 susceptibility alleles"**.

"BRCA1 Locus," "BRCA1 Gene," "BRCA1 Nucleic Acids" or **"BRCA1 Polynucleotide"** each refer to polynucleotides, all of which are in the BRCA1 region, that are likely to be expressed in normal tissue, certain alleles of which predispose an individual to develop breast, ovarian, colorectal and prostate cancers. Mutations at the BRCA1 locus may be involved in the initiation and/or progression of other types of tumors. The locus is indicated in part by mutations that predispose

GeneDX 1009, pg. 31

individuals to develop cancer. These mutations fall within the BRCA1 region described *infra*. The BRCA1 locus is intended to include coding sequences, intervening sequences and regulatory elements controlling transcription and/or translation. The BRCA1 locus is intended to include all allelic variations of the DNA sequence.

These terms, when applied to a nucleic acid, refer to a nucleic acid which encodes a BRCA1 polypeptide, fragment, homolog or variant, including, e.g., protein fusions or deletions. The nucleic acids of the present invention will possess a sequence which is either derived from, or substantially similar to a natural BRCA1-encoding gene or one having substantial homology with a natural BRCA1-encoding gene or a portion thereof. The coding sequence for a BRCA1 polypeptide is shown in SEQ ID NO:1, with the amino acid sequence shown in SEQ ID NO:2.

The polynucleotide compositions of this invention include RNA, cDNA, genomic DNA, synthetic forms, and mixed polymers, both sense and antisense strands, and may be chemically or biochemically modified or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those skilled in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.), charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids, etc.). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The present invention provides recombinant nucleic acids comprising all or part of the BRCA1 region. The recombinant construct may be capable of replicating autonomously in a host cell. Alternatively, the recombinant construct may become integrated into the chromosomal DNA of the host cell. Such a recombinant polynucleotide comprises a polynucleotide of genomic, cDNA, semi-synthetic, or synthetic origin which, by virtue of its origin or manipulation, 1) is not associated with all or a portion of a polynucleotide with which it is associated in nature; 2) is linked to a polynucleotide other than that to which it is linked in nature; or 3) does not occur in nature.

GeneDX 1009, pg. 32

Therefore, recombinant nucleic acids comprising sequences otherwise not naturally occurring are provided by this invention. Although the wild-type sequence may be employed, it will often be altered, e.g., by deletion, substitution or insertion.

cDNA or genomic libraries of various types may be screened as natural sources of the nucleic acids of the present invention, or such nucleic acids may be provided by amplification of sequences resident in genomic DNA or other natural sources, e.g., by PCR. The choice of cDNA libraries normally corresponds to a tissue source which is abundant in mRNA for the desired proteins. Phage libraries are normally preferred, but other types of libraries may be used. Clones of a library are spread onto plates, transferred to a substrate for screening, denatured and probed for the presence of desired sequences.

The DNA sequences used in this invention will usually comprise at least about five codons (15 nucleotides), more usually at least about 7-15 codons, and most preferably, at least about 35 codons. One or more introns may also be present. This number of nucleotides is usually about the minimal length required for a successful probe that would hybridize specifically with a BRCA1-encoding sequence.

Techniques for nucleic acid manipulation are described generally, for example, in Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992. Reagents useful in applying such techniques, such as restriction enzymes and the like, are widely known in the art and commercially available from such vendors as New England BioLabs, Boehringer Mannheim, Amersham, Promega Biotec, U. S. Biochemicals, New England Nuclear, and a number of other sources. The recombinant nucleic acid sequences used to produce fusion proteins of the present invention may be derived from natural or synthetic sequences. Many natural gene sequences are obtainable from various cDNA or from genomic libraries using appropriate probes. See, GenBank, National Institutes of Health.

"BRCA1 Region" refers to a portion of human chromosome 17q21 bounded by the markers tdj1474 and U5R. This region contains the BRCA1 locus, including the BRCA1 gene.

As used herein, the terms **"BRCA1 locus," "BRCA1 allele"** and **"BRCA1 region"** all refer to the double-stranded DNA comprising the locus, allele, or region, as well as either of the single-stranded DNAs comprising the locus, allele or region.

As used herein, a **"portion"** of the BRCA1 locus or region or allele is defined as having a minimal size of at least about eight nucleotides, or preferably about 15 nucleotides, or more preferably at least about 25 nucleotides, and may have a minimal size of at least about 40 nucleotides.

GeneDX 1009, pg. 33

"BRCA1 protein" or **"BRCA1 polypeptide"** refer to a protein or polypeptide encoded by the BRCA1 locus, variants or fragments thereof. The term "polypeptide" refers to a polymer of amino acids and its equivalent and does not refer to a specific length of the product; thus, peptides, oligopeptides and proteins are included within the definition of a polypeptide. This term also does not refer to, or exclude modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations, and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages as well as other modifications known in the art, both naturally and non-naturally occurring. Ordinarily, such polypeptides will be at least about 50% homologous to the native BRCA1 sequence, preferably in excess of about 90%, and more preferably at least about 95% homologous. Also included are proteins encoded by DNA which hybridize under high or low stringency conditions, to BRCA1-encoding nucleic acids and closely related polypeptides or proteins retrieved by antisera to the BRCA1 protein(s).

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acids, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression.

"Probes". Polynucleotide polymorphisms associated with BRCA1 alleles which predispose to certain cancers or are associated with most cancers are detected by hybridization with a polynucleotide probe which forms a stable hybrid with that of the target sequence, under stringent to moderately stringent hybridization and wash conditions. If it is expected that the probes will be perfectly complementary to the target sequence, stringent conditions will be used. Hybridization stringency may be lessened if some mismatching is expected, for example, if variants are expected with the result that the probe will not be completely complementary. Conditions are chosen which rule out nonspecific/adventitious bindings, that is, which minimize noise. Since such indications identify neutral DNA polymorphisms as well as mutations, these indications need further analysis to demonstrate detection of a BRCA1 susceptibility allele.

Probes for BRCA1 alleles may be derived from the sequences of the BRCA1 region or its cDNAs. The probes may be of any suitable length, which span all or a portion of the BRCA1 region,

and which allow specific hybridization to the BRCA1 region. If the target sequence contains a sequence identical to that of the probe, the probes may be short, e.g., in the range of about 8-30 base pairs, since the hybrid will be relatively stable under even stringent conditions. If some degree of mismatch is expected with the probe, i.e., if it is suspected that the probe will hybridize to a variant region, a longer probe may be employed which hybridizes to the target sequence with the requisite specificity.

The probes will include an isolated polynucleotide attached to a label or reporter molecule and may be used to isolate other polynucleotide sequences, having sequence similarity by standard methods. For techniques for preparing and labeling probes see, e.g., Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992. Other similar polynucleotides may be selected by using homologous polynucleotides. Alternatively, polynucleotides encoding these or similar polypeptides may be synthesized or selected by use of the redundancy in the genetic code. Various codon substitutions may be introduced, e.g., by silent changes (thereby producing various restriction sites) or to optimize expression for a particular system. Mutations may be introduced to modify the properties of the polypeptide, perhaps to change ligand-binding affinities, interchain affinities, or the polypeptide degradation or turnover rate.

Probes comprising synthetic oligonucleotides or other polynucleotides of the present invention may be derived from naturally occurring or recombinant single- or double-stranded polynucleotides, or be chemically synthesized. Probes may also be labeled by nick translation, Klenow fill-in reaction, or other methods known in the art.

Portions of the polynucleotide sequence having at least about eight nucleotides, usually at least about 15 nucleotides, and fewer than about 6 kb, usually fewer than about 1.0 kb, from a polynucleotide sequence encoding BRCA1 are preferred as probes. The probes may also be used to determine whether mRNA encoding BRCA1 is present in a cell or tissue.

"Protein modifications or fragments" are provided by the present invention for BRCA1 polypeptides or fragments thereof which are substantially homologous to primary structural sequence but which include, e.g., *in vivo* or *in vitro* chemical and biochemical modifications or which incorporate unusual amino acids. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, e.g., with radionuclides, and various enzymatic modifications, as will be readily appreciated by those well skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{32}P, ligands which bind to labeled antiligands (e.g.,

antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods of labeling polypeptides are well known in the art. See, e.g., Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992.

Besides substantially full-length polypeptides, the present invention provides for biologically active fragments of the polypeptides. Significant biological activities include ligand-binding, immunological activity and other biological activities characteristic of BRCA1 polypeptides. Immunological activities include both immunogenic function in a target immune system, as well as sharing of immunological epitopes for binding, serving as either a competitor or substitute antigen for an epitope of the BRCA1 protein. As used herein, "epitope" refers to an antigenic determinant of a polypeptide. An epitope could comprise three amino acids in a spatial conformation which is unique to the epitope. Generally, an epitope consists of at least five such amino acids, and more usually consists of at least 8-10 such amino acids. Methods of determining the spatial conformation of such amino acids are known in the art.

For immunological purposes, tandem-repeat polypeptide segments may be used as immunogens, thereby producing highly antigenic proteins. Alternatively, such polypeptides will serve as highly efficient competitors for specific binding. Production of antibodies specific for BRCA1 polypeptides or fragments thereof is described below.

The present invention also provides for fusion polypeptides, comprising BRCA1 polypeptides and fragments. Homologous polypeptides may be fusions between two or more BRCA1 polypeptide sequences or between the sequences of BRCA1 and a related protein. Likewise, heterologous fusions may be constructed which would exhibit a combination of properties or activities of the derivative proteins. For example, ligand-binding or other domains may be "swapped" between different new fusion polypeptides or fragments. Such homologous or heterologous fusion polypeptides may display, for example, altered strength or specificity of binding. Fusion partners include immunoglobulins, bacterial β-galactosidase, trpE, protein A, β-lactamase, alpha amylase, alcohol dehydrogenase and yeast alpha mating factor. See, e.g., Godowski *et al.*, 1988.

Fusion proteins will typically be made by either recombinant nucleic acid methods, as described below, or may be chemically synthesized. Techniques for the synthesis of polypeptides are described, for example, in Merrifield, 1963.

"Protein purification" refers to various methods for the isolation of the BRCA1 polypeptides from other biological material, such as from cells transformed with recombinant nucleic acids encoding BRCA1, and are well known in the art. For example, such polypeptides may be purified by immuno-affinity chromatography employing, e.g., the antibodies provided by the present invention. Various methods of protein purification are well known in the art, and include those described in Deutscher, 1990 and Scopes, 1982.

The terms **"isolated"**, **"substantially pure"**, and **"substantially homogeneous"** are used interchangeably to describe a protein or polypeptide which has been separated from components which accompany it in its natural state. A monomeric protein is substantially pure when at least about 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure protein will typically comprise about 60 to 90% W/W of a protein sample, more usually about 95%, and preferably will be over about 99% pure. Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art which are utilized for purification.

A BRCA1 protein is substantially free of naturally associated components when it is separated from the native contaminants which accompany it in its natural state. Thus, a polypeptide which is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be substantially free from its naturally associated components. A protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art.

A polypeptide produced as an expression product of an isolated and manipulated genetic sequence is an "isolated polypeptide," as used herein, even if expressed in a homologous cell type. Synthetically made forms or molecules expressed by heterologous cells are inherently isolated molecules.

"Recombinant nucleic acid" is a nucleic acid which is not naturally occurring, or which is made by the artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. Such is usually done to replace a codon with a redundant codon encoding the same or a conservative amino acid, while

typically introducing or removing a sequence recognition site. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions.

"Regulatory sequences" refers to those sequences normally within 100 kb of the coding region of a locus, but they may also be more distant from the coding region, which affect the expression of the gene (including transcription of the gene, and translation, splicing, stability or the like of the messenger RNA).

"Substantial homology or similarity". A nucleic acid or fragment thereof is "substantially homologous" ("or substantially similar") to another if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases.

Alternatively, substantial homology or (similarity) exists when a nucleic acid or fragment thereof will hybridize to another nucleic acid (or a complementary strand thereof) under selective hybridization conditions, to a strand, or to its complement. Selectivity of hybridization exists when hybridization which is substantially more selective than total lack of specificity occurs. Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 14 nucleotides, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90%. See, Kanehisa, 1984. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will often be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions will generally include temperatures in excess of 30°C, typically in excess of 37°C, and preferably in excess of 45°C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Wetmur & Davidson, 1968.

Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

The terms **"substantial homology"** or **"substantial identity"**, when referring to polypeptides, indicate that the polypeptide or protein in question exhibits at least about 30% identity with an entire naturally-occurring protein or a portion thereof, usually at least about 70% identity, and preferably at least about 95% identity.

"**Substantially similar function**" refers to the function of a modified nucleic acid or a modified protein, with reference to the wild-type BRCA1 nucleic acid or wild-type BRCA1 polypeptide. The modified polypeptide will be substantially homologous to the wild-type BRCA1 polypeptide and will have substantially the same function. The modified polypeptide may have an altered amino acid sequence and/or may contain modified amino acids. In addition to the similarity of function, the modified polypeptide may have other useful properties, such as a longer half-life. The similarity of function (activity) of the modified polypeptide may be substantially the same as the activity of the wild-type BRCA1 polypeptide. Alternatively, the similarity of function (activity) of the modified polypeptide may be higher than the activity of the wild-type BRCA1 polypeptide. The modified polypeptide is synthesized using conventional techniques, or is encoded by a modified nucleic acid and produced using conventional techniques. The modified nucleic acid is prepared by conventional techniques. A nucleic acid with a function substantially similar to the wild-type BRCA1 gene function produces the modified protein described above.

Homology, for polypeptides, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisconsin 53705. Protein analysis software matches similar sequences using measure of homology assigned to various substitutions, deletions and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

A polypeptide "**fragment**," "**portion**" or "**segment**" is a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids.

GeneDX 1009, pg. 39

The polypeptides of the present invention, if soluble, may be coupled to a solid-phase support, e.g., nitrocellulose, nylon, column packing materials (e.g., Sepharose beads), magnetic beads, glass wool, plastic, metal, polymer gels, cells, or other substrates. Such supports may take the form, for example, of beads, wells, dipsticks, or membranes.

"Target region" refers to a region of the nucleic acid which is amplified and/or detected. The term "**target sequence**" refers to a sequence with which a probe or primer will form a stable hybrid under desired conditions.

The practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, genetics, and immunology. See, e.g., Maniatis *et al.*, 1982; Sambrook *et al.*, 1989; Ausubel *et al.*, 1992; Glover, 1985; Anand, 1992; Guthrie & Fink, 1991. A general discussion of techniques and materials for human gene mapping, including mapping of human chromosome 17q, is provided, e.g., in White and Lalouel, 1988.

Preparation of recombinant or chemically synthesized nucleic acids; vectors, transformation, host cells

Large amounts of the polynucleotides of the present invention may be produced by replication in a suitable host cell. Natural or synthetic polynucleotide fragments coding for a desired fragment will be incorporated into recombinant polynucleotide constructs, usually DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell. Usually the polynucleotide constructs will be suitable for replication in a unicellular host, such as yeast or bacteria, but may also be intended for introduction to (with and without integration within the genome) cultured mammalian or plant or other eukaryotic cell lines. The purification of nucleic acids produced by the methods of the present invention is described, e.g., in Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992.

The polynucleotides of the present invention may also be produced by chemical synthesis, e.g., by the phosphoramidite method described by Beaucage & Carruthers, 1981 or the triester method according to Matteucci and Caruthers, 1981, and may be performed on commercial, automated oligonucleotide synthesizers. A double-stranded fragment may be obtained from the single-stranded product of chemical synthesis either by synthesizing the complementary strand and annealing the strands together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

GeneDX 1009, pg. 40

Polynucleotide constructs prepared for introduction into a prokaryotic or eukaryotic host may comprise a replication system recognized by the host, including the intended polynucleotide fragment encoding the desired polypeptide, and will preferably also include transcription and translational initiation regulatory sequences operably linked to the polypeptide encoding segment. Expression vectors may include, for example, an origin of replication or autonomously replicating sequence (ARS) and expression control sequences, a promoter, an enhancer and necessary processing information sites, such as ribosome-binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, and mRNA stabilizing sequences. Secretion signals may also be included where appropriate, whether from a native BRCA1 protein or from other receptors or from secreted polypeptides of the same or related species, which allow the protein to cross and/or lodge in cell membranes, and thus attain its functional topology, or be secreted from the cell. Such vectors may be prepared by means of standard recombinant techniques well known in the art and discussed, for example, in Sambrook *et al.*, 1989 or Ausubel *et al.* 1992.

An appropriate promoter and other necessary vector sequences will be selected so as to be functional in the host, and may include, when appropriate, those naturally associated with BRCA1 genes. Examples of workable combinations of cell lines and expression vectors are described in Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992; see also, e.g., Metzger *et al.*, 1988. Many useful vectors are known in the art and may be obtained from such vendors as Stratagene, New England Biolabs, Promega Biotech, and others. Promoters such as the trp, lac and phage promoters, tRNA promoters and glycolytic enzyme promoters may be used in prokaryotic hosts. Useful yeast promoters include promoter regions for metallothionein, 3-phosphoglycerate kinase or other glycolytic enzymes such as enolase or glyceraldehyde-3-phosphate dehydrogenase, enzymes responsible for maltose and galactose utilization, and others. Vectors and promoters suitable for use in yeast expression are further described in Hitzeman *et al.*, EP 73,675A. Appropriate non-native mammalian promoters might include the early and late promoters from SV40 (Fiers *et al.*, 1978) or promoters derived from murine Moloney leukemia virus, mouse tumor virus, avian sarcoma viruses, adenovirus II, bovine papilloma virus or polyoma. In addition, the construct may be joined to an amplifiable gene (e.g., DHFR) so that multiple copies of the gene may be made. For appropriate enhancer and other expression control sequences, see also *Enhancers and Eukaryotic Gene Expression*, Cold Spring Harbor Press, Cold Spring Harbor, New York (1983).

While such expression vectors may replicate autonomously, they may also replicate by being inserted into the genome of the host cell, by methods well known in the art.

Expression and cloning vectors will likely contain a selectable marker, a gene encoding a protein necessary for survival or growth of a host cell transformed with the vector. The presence of this gene ensures growth of only those host cells which express the inserts. Typical selection genes encode proteins that a) confer resistance to antibiotics or other toxic substances, e.g. ampicillin, neomycin, methotrexate, etc.; b) complement auxotrophic deficiencies, or c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for *Bacilli*. The choice of the proper selectable marker will depend on the host cell, and appropriate markers for different hosts are well known in the art.

The vectors containing the nucleic acids of interest can be transcribed *in vitro*, and the resulting RNA introduced into the host cell by well-known methods, e.g., by injection (see, Kubo *et al.*, 1988), or the vectors can be introduced directly into host cells by methods well known in the art, which vary depending on the type of cellular host, including electroporation; transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; infection (where the vector is an infectious agent, such as a retroviral genome); and other methods. See generally, Sambrook *et al.*, 1989 and Ausubel *et al.*, 1992. The introduction of the polynucleotides into the host cell by any method known in the art, including, *inter alia*, those described above, will be referred to herein as "transformation." The cells into which have been introduced nucleic acids described above are meant to also include the progeny of such cells.

Large quantities of the nucleic acids and polypeptides of the present invention may be prepared by expressing the BRCA1 nucleic acids or portions thereof in vectors or other expression vehicles in compatible prokaryotic or eukaryotic host cells. The most commonly used prokaryotic hosts are strains of *Escherichia coli*, although other prokaryotes, such as *Bacillus subtilis* or *Pseudomonas* may also be used.

Mammalian or other eukaryotic host cells, such as those of yeast, filamentous fungi, plant, insect, or amphibian or avian species, may also be useful for production of the proteins of the present invention. Propagation of mammalian cells in culture is *per se* well known. See, Jakoby and Pastan, 1979. Examples of commonly used mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cells, and WI38, BHK, and COS cell lines, although it will be appreciated by the

skilled practitioner that other cell lines may be appropriate, e.g., to provide higher expression, desirable glycosylation patterns, or other features.

Clones are selected by using markers depending on the mode of the vector construction. The marker may be on the same or a different DNA molecule, preferably the same DNA molecule. In prokaryotic hosts, the transformant may be selected, e.g., by resistance to ampicillin, tetracycline or other antibiotics. Production of a particular product based on temperature sensitivity may also serve as an appropriate marker.

Prokaryotic or eukaryotic cells transformed with the polynucleotides of the present invention will be useful not only for the production of the nucleic acids and polypeptides of the present invention, but also, for example, in studying the characteristics of BRCA1 polypeptides.

Antisense polynucleotide sequences are useful in preventing or diminishing the expression of the BRCA1 locus, as will be appreciated by those skilled in the art. For example, polynucleotide vectors containing all or a portion of the BRCA1 locus or other sequences from the BRCA1 region (particularly those flanking the BRCA1 locus) may be placed under the control of a promoter in an antisense orientation and introduced into a cell. Expression of such an antisense construct within a cell will interfere with BRCA1 transcription and/or translation and/or replication.

The probes and primers based on the BRCA1 gene sequences disclosed herein are used to identify homologous BRCA1 gene sequences and proteins in other species. These BRCA1 gene sequences and proteins are used in the diagnostic/prognostic, therapeutic and drug screening methods described herein for the species from which they have been isolated.

Methods of Use: Nucleic Acid Diagnosis and Diagnostic Kits

In order to detect the presence of a BRCA1 allele predisposing an individual to cancer, a biological sample such as blood is prepared and analyzed for the presence or absence of susceptibility alleles of BRCA1. In order to detect the presence of neoplasia, the progression toward malignancy of a precursor lesion, or as a prognostic indicator, a biological sample of the lesion is prepared and analyzed for the presence or absence of mutant alleles of BRCA1. Results of these tests and interpretive information are returned to the health care provider for communication to the tested individual. Such diagnoses may be performed by diagnostic laboratories, or, alternatively, diagnostic kits are manufactured and sold to health care providers or to private individuals for self-diagnosis.

Initially, the screening method involves amplification of the relevant BRCA1 sequences. In another preferred embodiment of the invention, the screening method involves a non-PCR based strategy. Such screening methods include two-step label amplification methodologies that are well known in the art. Both PCR and non-PCR based screening strategies can detect target sequences with a high level of sensitivity.

The most popular method used today is target amplification. Here, the target nucleic acid sequence is amplified with polymerases. One particularly preferred method using polymerase-driven amplification is the polymerase chain reaction (PCR). The polymerase chain reaction and other polymerase-driven amplification assays can achieve over a million-fold increase in copy number through the use of polymerase-driven amplification cycles. Once amplified, the resulting nucleic acid can be sequenced or used as a substrate for DNA probes.

When the probes are used to detect the presence of the target sequences (for example, in screening for cancer susceptibility), the biological sample to be analyzed, such as blood or serum, may be treated, if desired, to extract the nucleic acids. The sample nucleic acid may be prepared in various ways to facilitate detection of the target sequence; e.g. denaturation, restriction digestion, electrophoresis or dot blotting. The targeted region of the analyte nucleic acid usually must be at least partially single-stranded to form hybrids with the targeting sequence of the probe. If the sequence is naturally single-stranded, denaturation will not be required. However, if the sequence is double-stranded, the sequence will probably need to be denatured. Denaturation can be carried out by various techniques known in the art.

Analyte nucleic acid and probe are incubated under conditions which promote stable hybrid formation of the target sequence in the probe with the putative targeted sequence in the analyte. The region of the probes which is used to bind to the analyte can be made completely complementary to the targeted region of human chromosome 17q. Therefore, high stringency conditions are desirable in order to prevent false positives. However, conditions of high stringency are used only if the probes are complementary to regions of the chromosome which are unique in the genome. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, base composition, probe length, and concentration of formamide. These factors are outlined in, for example, Maniatis *et al.*, 1982 and Sambrook *et al.*, 1989. Under certain circumstances, the formation of higher order hybrids, such as triplexes, quadraplexes, etc., may be desired to provide the means of detecting target sequences.

Detection, if any, of the resulting hybrid is usually accomplished by the use of labeled probes. Alternatively, the probe may be unlabeled, but may be detectable by specific binding with a ligand which is labeled, either directly or indirectly. Suitable labels, and methods for labeling probes and ligands are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (e.g., nick translation, random priming or kinasing), biotin, fluorescent groups, chemiluminescent groups (e.g., dioxetanes, particularly triggered dioxetanes), enzymes, antibodies and the like. Variations of this basic scheme are known in the art, and include those variations that facilitate separation of the hybrids to be detected from extraneous materials and/or that amplify the signal from the labeled moiety. A number of these variations are reviewed in, e.g., Matthews & Kricka, 1988; Landegren *et al.*, 1988; Mittlin, 1989; U.S. Patent 4,868,105, and in EPO Publication No. 225,807.

As noted above, non-PCR based screening assays are also contemplated in this invention. An exemplary non-PCR based procedure is provided in Example 11. This procedure hybridizes a nucleic acid probe (or an analog such as a methyl phosphonate backbone replacing the normal phosphodiester), to the low level DNA target. This probe may have an enzyme covalently linked to the probe, such that the covalent linkage does not interfere with the specificity of the hybridization. This enzyme-probe-conjugate-target nucleic acid complex can then be isolated away from the free probe enzyme conjugate and a substrate is added for enzyme detection. Enzymatic activity is observed as a change in color development or luminescent output resulting in a 10^3-10^6 increase in sensitivity. For an example relating to the preparation of oligodeoxynucleotide-alkaline phosphatase conjugates and their use as hybridization probes see Jablonski *et al.*, 1986.

Two-step label amplification methodologies are known in the art. These assays work on the principle that a small ligand (such as digoxigenin, biotin, or the like) is attached to a nucleic acid probe capable of specifically binding BRCA1. Exemplary probes are provided in Table 9 of this patent application and additionally include the nucleic acid probe corresponding to nucleotide positions 3631 to 3930 of SEQ ID NO:1. Allele specific probes are also contemplated within the scope of this example and exemplary allele specific probes include probes encompassing the predisposing mutations summarized in Tables 11 and 12 of this patent application.

In one example, the small ligand attached to the nucleic acid probe is specifically recognized by an antibody-enzyme conjugate. In one embodiment of this example, digoxigenin is attached to the nucleic acid probe. Hybridization is detected by an antibody-alkaline phosphatase conjugate which

GeneDX 1009, pg. 45

turns over a chemiluminescent substrate. For methods for labeling nucleic acid probes according to this embodiment see Martin *et al.*, 1990. In a second example, the small ligand is recognized by a second ligand-enzyme conjugate that is capable of specifically complexing to the first ligand. A well known embodiment of this example is the biotin-avidin type of interactions. For methods for labeling nucleic acid probes and their use in biotin-avidin based assays see Rigby *et al.*, 1977 and Nguyen *et al.*, 1992.

It is also contemplated within the scope of this invention that the nucleic acid probe assays of this invention will employ a cocktail of nucleic acid probes capable of detecting BRCA1. Thus, in one example to detect the presence of BRCA1 in a cell sample, more than one probe complementary to BRCA1 is employed and in particular the number of different probes is alternatively 2, 3, or 5 different nucleic acid probe sequences. In another example, to detect the presence of mutations in the BRCA1 gene sequence in a patient, more than one probe complementary to BRCA1 is employed where the cocktail includes probes capable of binding to the allele-specific mutations identified in populations of patients with alterations in BRCA1. In this embodiment, any number of probes can be used, and will preferably include probes corresponding to the major gene mutations identified as predisposing an individual to breast cancer. Some candidate probes contemplated within the scope of the invention include probes that include the allele-specific mutations identified in Tables 11 and 12 and those that have the BRCA1 regions corresponding to SEQ ID NO:1 both 5' and 3' to the mutation site.

Methods of Use: Peptide Diagnosis and Diagnostic Kits

The neoplastic condition of lesions can also be detected on the basis of the alteration of wild-type BRCA1 polypeptide. Such alterations can be determined by sequence analysis in accordance with conventional techniques. More preferably, antibodies (polyclonal or monoclonal) are used to detect differences in, or the absence of BRCA1 peptides. The antibodies may be prepared as discussed above under the heading "Antibodies" and as further shown in Examples 12 and 13. Other techniques for raising and purifying antibodies are well known in the art and any such techniques may be chosen to achieve the preparations claimed in this invention. In a preferred embodiment of the invention, antibodies will immunoprecipitate BRCA1 proteins from solution as well as react with BRCA1 protein on Western or immunoblots of polyacrylamide gels. In another preferred embodiment, antibodies will detect BRCA1 proteins in paraffin or frozen tissue sections, using immunocytochemical techniques.

GeneDX 1009, pg. 46

Preferred embodiments relating to methods for detecting BRCA1 or its mutations include enzyme linked immunosorbent assays (ELISA), radioimmunoassays (RIA), immunoradiometric assays (IRMA) and immunoenzymatic assays (IEMA), including sandwich assays using monoclonal and/or polyclonal antibodies. Exemplary sandwich assays are described by David *et al.* in U.S. Patent Nos. 4,376,110 and 4,486,530, hereby incorporated by reference, and exemplified in Example 14.

Methods of Use: Drug Screening

This invention is particularly useful for screening compounds by using the BRCA1 polypeptide or binding fragment thereof in any of a variety of drug screening techniques.

The BRCA1 polypeptide or fragment employed in such a test may either be free in solution, affixed to a solid support, or borne on a cell surface. One method of drug screening utilizes eucaryotic or procaryotic host cells which are stably transformed with recombinant polynucleotides expressing the polypeptide or fragment, preferably in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, for the formation of complexes between a BRCA1 polypeptide or fragment and the agent being tested, or examine the degree to which the formation of a complex between a BRCA1 polypeptide or fragment and a known ligand is interfered with by the agent being tested.

Thus, the present invention provides methods of screening for drugs comprising contacting such an agent with a BRCA1 polypeptide or fragment thereof and assaying (i) for the presence of a complex between the agent and the BRCA1 polypeptide or fragment, or (ii) for the presence of a complex between the BRCA1 polypeptide or fragment and a ligand, by methods well known in the art. In such competitive binding assays the BRCA1 polypeptide or fragment is typically labeled. Free BRCA1 polypeptide or fragment is separated from that present in a protein:protein complex, and the amount of free (i.e., uncomplexed) label is a measure of the binding of the agent being tested to BRCA1 or its interference with BRCA1:ligand binding, respectively.

Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to the BRCA1 polypeptides and is described in detail in Geysen, PCT published application WO 84/03564, published on September 13, 1984. Briefly stated, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with BRCA1 polypeptide and washed. Bound BRCA1 polypeptide is then detected by methods well known in the art.

GeneDX 1009, pg. 47

Purified BRCA1 can be coated directly onto plates for use in the aforementioned drug screening techniques. However, non-neutralizing antibodies to the polypeptide can be used to capture antibodies to immobilize the BRCA1 polypeptide on the solid phase.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of specifically binding the BRCA1 polypeptide compete with a test compound for binding to the BRCA1 polypeptide or fragments thereof. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants of the BRCA1 polypeptide.

A further technique for drug screening involves the use of host eukaryotic cell lines or cells (such as described above) which have a nonfunctional BRCA1 gene. These host cell lines or cells are defective at the BRCA1 polypeptide level. The host cell lines or cells are grown in the presence of drug compound. The rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of BRCA1 defective cells.

Methods of Use: Rational Drug Design

The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which, e.g., enhance or interfere with the function of a polypeptide *in vivo*. See, e.g., Hodgson, 1991. In one approach, one first determines the three-dimensional structure of a protein of interest (e.g., BRCA1 polypeptide) or, for example, of the BRCA1-receptor or ligand complex, by x-ray crystallography, by computer modeling or most typically, by a combination of approaches. Less often, useful information regarding the structure of a polypeptide may be gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson *et al.*, 1990). In addition, peptides (e.g., BRCA1 polypeptide) are analyzed by an alanine scan (Wells, 1991). In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

It is also possible to isolate a target-specific antibody, selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idio-

GeneDX 1009, pg. 48

typic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original receptor. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

Thus, one may design drugs which have, e.g., improved BRCA1 polypeptide activity or stability or which act as inhibitors, agonists, antagonists, etc. of BRCA1 polypeptide activity. By virtue of the availability of cloned BRCA1 sequences, sufficient amounts of the BRCA1 polypeptide may be made available to perform such analytical studies as x-ray crystallography. In addition, the knowledge of the BRCA1 protein sequence provided herein will guide those employing computer modeling techniques in place of, or in addition to x-ray crystallography.

Methods of Use: Gene Therapy

According to the present invention, a method is also provided of supplying wild-type BRCA1 function to a cell which carries mutant BRCA1 alleles. Supplying such a function should suppress neoplastic growth of the recipient cells. The wild-type BRCA1 gene or a part of the gene may be introduced into the cell in a vector such that the gene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. If a gene fragment is introduced and expressed in a cell carrying a mutant BRCA1 allele, the gene fragment should encode a part of the BRCA1 protein which is required for non-neoplastic growth of the cell. More preferred is the situation where the wild-type BRCA1 gene or a part thereof is introduced into the mutant cell in such a way that it recombines with the endogenous mutant BRCA1 gene present in the cell. Such recombination requires a double recombination event which results in the correction of the BRCA1 gene mutation. Vectors for introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector may be used. Methods for introducing DNA into cells such as electroporation, calcium phosphate co-precipitation and viral transduction are known in the art, and the choice of method is within the competence of the routineer. Cells transformed with the wild-type BRCA1 gene can be used as model systems to study cancer remission and drug treatments which promote such remission.

As generally discussed above, the BRCA1 gene or fragment, where applicable, may be employed in gene therapy methods in order to increase the amount of the expression products of such genes in cancer cells. Such gene therapy is particularly appropriate for use in both cancerous and pre-cancerous

GeneDX 1009, pg. 49

cells, in which the level of BRCA1 polypeptide is absent or diminished compared to normal cells. It may also be useful to increase the level of expression of a given BRCA1 gene even in those tumor cells in which the mutant gene is expressed at a "normal" level, but the gene product is not fully functional.

Gene therapy would be carried out according to generally accepted methods, for example, as described by Friedman, 1991. Cells from a patient's tumor would be first analyzed by the diagnostic methods described above, to ascertain the production of BRCA1 polypeptide in the tumor cells. A virus or plasmid vector (see further details below), containing a copy of the BRCA1 gene linked to expression control elements and capable of replicating inside the tumor cells, is prepared. Suitable vectors are known, such as disclosed in U.S. Patent 5,252,479 and PCT published application WO 93/07282. The vector is then injected into the patient, either locally at the site of the tumor or systemically (in order to reach any tumor cells that may have metastasized to other sites). If the transfected gene is not permanently incorporated into the genome of each of the targeted tumor cells, the treatment may have to be repeated periodically.

Gene transfer systems known in the art may be useful in the practice of the gene therapy methods of the present invention. These include viral and nonviral transfer methods. A number of viruses have been used as gene transfer vectors, including papovaviruses, e.g., SV40 (Madzak *et al.*, 1992), adenovirus (Berkner, 1992; Berkner *et al.*, 1988; Gorziglia and Kapikian, 1992; Quantin *et al.*, 1992; Rosenfeld *et al.*, 1992; Wilkinson *et al.*, 1992; Stratford-Perricaudet *et al.*, 1990), vaccinia virus (Moss, 1992), adeno-associated virus (Muzyczka, 1992; Ohi *et al.*, 1990), herpesviruses including HSV and EBV (Margolskee, 1992; Johnson *et al.*, 1992; Fink *et al.*, 1992; Breakfield and Geller, 1987; Freese *et al.*, 1990), and retroviruses of avian (Brandyopadhyay and Temin, 1984; Petropoulos *et al.*, 1992), murine (Miller, 1992; Miller *et al.*, 1985; Sorge *et al.*, 1984; Mann and Baltimore, 1985; Miller *et al.*, 1988), and human origin (Shimada *et al.*, 1991; Helseth *et al.*, 1990; Page *et al.*, 1990; Buchschacher and Panganiban, 1992). Most human gene therapy protocols have been based on disabled murine retroviruses.

Nonviral gene transfer methods known in the art include chemical techniques such as calcium phosphate coprecipitation (Graham and van der Eb, 1973; Pellicer *et al.*, 1980); mechanical techniques, for example microinjection (Anderson *et al.*, 1980; Gordon *et al.*, 1980; Brinster *et al.*, 1981; Constantini and Lacy, 1981); membrane fusion-mediated transfer via liposomes (Felgner *et al.*, 1987; Wang and Huang, 1989; Kaneda et al, 1989; Stewart *et al.*, 1992; Nabel *et al.*, 1990; Lim *et al.*, 1992); and direct DNA uptake and receptor-mediated DNA transfer (Wolff *et al.*, 1990; Wu *et al.*, 1991;

GeneDX 1009, pg. 50

Zenke *et al.*, 1990; Wu *et al.*, 1989b; Wolff *et al.*, 1991; Wagner *et al.*, 1990; Wagner *et al.*, 1991; Cotten *et al.*, 1990; Curiel *et al.*, 1991a; Curiel *et al.*, 1991b). Viral-mediated gene transfer can be combined with direct *in vivo* gene transfer using liposome delivery, allowing one to direct the viral vectors to the tumor cells and not into the surrounding nondividing cells. Alternatively, the retroviral vector producer cell line can be injected into tumors (Culver *et al.*, 1992). Injection of producer cells would then provide a continuous source of vector particles. This technique has been approved for use in humans with inoperable brain tumors.

In an approach which combines biological and physical gene transfer methods, plasmid DNA of any size is combined with a polylysine-conjugated antibody specific to the adenovirus hexon protein, and the resulting complex is bound to an adenovirus vector. The trimolecular complex is then used to infect cells. The adenovirus vector permits efficient binding, internalization, and degradation of the endosome before the coupled DNA is damaged.

Liposome/DNA complexes have been shown to be capable of mediating direct *in vivo* gene transfer. While in standard liposome preparations the gene transfer process is nonspecific, localized *in vivo* uptake and expression have been reported in tumor deposits, for example, following direct *in situ* administration (Nabel, 1992).

Gene transfer techniques which target DNA directly to breast and ovarian tissues, e.g., epithelial cells of the breast or ovaries, is preferred. Receptor-mediated gene transfer, for example, is accomplished by the conjugation of DNA (usually in the form of covalently closed supercoiled plasmid) to a protein ligand via polylysine. Ligands are chosen on the basis of the presence of the corresponding ligand receptors on the cell surface of the target cell/tissue type. One appropriate receptor/ligand pair may include the estrogen receptor and its ligand, estrogen (and estrogen analogues). These ligand-DNA conjugates can be injected directly into the blood if desired and are directed to the target tissue where receptor binding and internalization of the DNA-protein complex occurs. To overcome the problem of intracellular destruction of DNA, coinfection with adenovirus can be included to disrupt endosome function.

The therapy involves two steps which can be performed singly or jointly. In the first step, prepubescent females who carry a BRCA1 susceptibility allele are treated with a gene delivery vehicle such that some or all of their mammary ductal epithelial precursor cells receive at least one additional copy of a functional normal BRCA1 allele. In this step, the treated individuals have reduced risk of breast cancer to the extent that the effect of the susceptible allele has been countered by the presence of

the normal allele. In the second step of a preventive therapy, predisposed young females, in particular women who have received the proposed gene therapeutic treatment, undergo hormonal therapy to mimic the effects on the breast of a full term pregnancy.

Methods of Use: Peptide Therapy

Peptides which have BRCA1 activity can be supplied to cells which carry mutant or missing BRCA1 alleles. The sequence of the BRCA1 protein is disclosed (SEQ ID NO:2). Protein can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors. Alternatively, BRCA1 polypeptide can be extracted from BRCA1-producing mammalian cells. In addition, the techniques of synthetic chemistry can be employed to synthesize BRCA1 protein. Any of such techniques can provide the preparation of the present invention which comprises the BRCA1 protein. The preparation is substantially free of other human proteins. This is most readily accomplished by synthesis in a microorganism or *in vitro*.

Active BRCA1 molecules can be introduced into cells by microinjection or by use of liposomes, for example. Alternatively, some active molecules may be taken up by cells, actively or by diffusion. Extracellular application of the BRCA1 gene product may be sufficient to affect tumor growth. Supply of molecules with BRCA1 activity should lead to partial reversal of the neoplastic state. Other molecules with BRCA1 activity (for example, peptides, drugs or organic compounds) may also be used to effect such a reversal. Modified polypeptides having substantially similar function are also used for peptide therapy.

Methods of Use: Transformed Hosts

Similarly, cells and animals which carry a mutant BRCA1 allele can be used as model systems to study and test for substances which have potential as therapeutic agents. The cells are typically cultured epithelial cells. These may be isolated from individuals with BRCA1 mutations, either somatic or germline. Alternatively, the cell line can be engineered to carry the mutation in the BRCA1 allele, as described above. After a test substance is applied to the cells, the neoplastically transformed phenotype of the cell is determined. Any trait of neoplastically transformed cells can be assessed, including anchorage-independent growth, tumorigenicity in nude mice, invasiveness of cells, and growth factor dependence. Assays for each of these traits are known in the art.

GeneDX 1009, pg. 52

Animals for testing therapeutic agents can be selected after mutagenesis of whole animals or after treatment of germline cells or zygotes. Such treatments include insertion of mutant BRCA1 alleles, usually from a second animal species, as well as insertion of disrupted homologous genes. Alternatively, the endogenous BRCA1 gene(s) of the animals may be disrupted by insertion or deletion mutation or other genetic alterations using conventional techniques (Capecchi, 1989; Valancius and Smithies, 1991; Hasty *et al.*, 1991; Shinkai *et al.*, 1992; Mombaerts *et al.*, 1992; Philpott *et al.*, 1992; Snouwaert *et al.*, 1992; Donehower *et al.*, 1992). After test substances have been administered to the animals, the growth of tumors must be assessed. If the test substance prevents or suppresses the growth of tumors, then the test substance is a candidate therapeutic agent for the treatment of the cancers identified herein. These animal models provide an extremely important testing vehicle for potential therapeutic products.

The present invention is described by reference to the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described below were utilized.

EXAMPLE 1

Ascertain and Study Kindreds Likely to Have a 17q-Linked Breast Cancer Susceptibility Locus

by our University of Utah collaborators

Extensive cancer prone kindreds were ascertained from a defined population providing a large set of extended kindreds with multiple cases of breast cancer and many relatives available to study. The large number of meioses present in these large kindreds provided the power to detect whether the BRCA1 locus was segregating, and increased the opportunity for informative recombinants to occur within the small region being investigated. This vastly improved the chances of establishing linkage to the BRCA1 region, and greatly facilitated the reduction of the BRCA1 region to a manageable size, which permits identification of the BRCA1 locus itself.

by our collaborators

Each kindred was extended through all available connecting relatives, and to all informative first degree relatives of each proband or cancer case. For these kindreds, additional breast cancer cases and individuals with cancer at other sites of interest (e.g. ovarian) who also appeared in the kindreds were identified through the tumor registry linked files. All breast cancers reported in the kindred which were

not confirmed in the Utah Cancer Registry were researched. Medical records or death certificates were obtained for confirmation of all cancers. Each key connecting individual and all informative individuals were invited to participate by providing a blood sample from which DNA was extracted. We also sampled spouses and relatives of deceased cases so that the genotype of the deceased cases could be inferred from the genotypes of their relatives.

Ten kindreds which had three or more cancer cases with inferable genotypes were selected for linkage studies to 17q markers from a set of 29 kindreds originally ascertained ~~from the linked databases~~ for a study of proliferative breast disease and breast cancer (Skolnick *et al.*, 1990). The criterion for selection of these kindreds was the presence of two sisters or a mother and her daughter with breast cancer. Additionally, two kindreds which have been studied since 1980 as part of ~~our~~ breast cancer linkage studies (K1001, K9018), six kindreds ascertained from ~~the linked databases~~ for the presence of clusters of breast and/or ovarian cancer (K2019, K2073, K2079, K2080, K2039, K2082) and a self-referred kindred with early onset breast cancer (K2035) were included. These kindreds were investigated and expanded in our clinic in the manner described above. Table 1 displays the characteristics of these 19 kindreds which are the subject of subsequent examples. In Table 1, for each kindred the total number of individuals in our database, the number of typed individuals, and the minimum, median, and maximum age at diagnosis of breast/ovarian cancer are reported. Kindreds are sorted in ascending order of median age at diagnosis of breast cancer. Four women diagnosed with both ovarian and breast cancer are counted in both categories.

TABLE 1

Description of the 19 Kindreds

KINDRED	No. of Individuals		Breast				Ovarian			
				Age at Dx				Age at Dx		
	Total	Sample	# Aff.	Min.	Med.	Max.	# Aff.	Min.	Med.	Max.
1910	15	10	4	27	34	49	-	-	-	-
1001	133	98	13	28	37	64	-	-	-	-
2035	42	25	8	28	37	45	1	-	60	-
2027	21	11	4	34	38	41	-	-	-	-
9018	54	17	9	30	40	72	2	46	48	50
1925	50	27	4	39	42	53	-	-	-	-
1927	49	29	5	32	42	51	-	-	-	-
1911	28	21	7	28	42	76	-	-	-	-
1929	16	11	4	34	43	73	-	-	-	-
1901	35	19	10	31	44	76	-	-	-	-
2082	180	105	20	27	47	67	10	45	52	66
2019	42	19	10	42	53	79	-	-	-	-
1900	70	23	8	45	55	70	1	-	78	-
2080	264	74	22+	27	55	92	4	45	53	71
2073	57	29	9	35	57	80	-	-	-	-
1917	16	6	4	43	58	61	-	-	-	-
1920	22	14	3	62	63	68	-	-	-	-
2079	136	18	14	38	66	84	4	52	59	65
2039	87	40	14	44	68	88	4	41	51	75

+ Includes one case of male breast cancer.

EXAMPLE 2

Selection of Kindreds Which are Linked to Chromosome 17q and Localization of BRCA1 to the Interval Mfd15 - Mfd188

For each sample collected in these 19 kindreds, DNA was extracted from blood (or in two cases from paraffin-embedded tissue blocks) using standard laboratory protocols. Genotyping in this study was restricted to short tandem repeat (STR) markers since, in general, they have high heterozygosity and PCR methods offer rapid turnaround while using very small amounts of DNA. To aid in this effort, four such STR markers on chromosome 17 were developed by screening a chromosome specific cosmid library for CA positive clones. Three of these markers localized to the long arm: (46E6, Easton *et al.*, 1993); (42D6, Easton *et al.*, 1993); 26C2 (D17S514, Oliphant *et al.*, 1991), while the other, 12G6 (D17S513, Oliphant *et al.*, 1991), localized to the short arm near the p53 tumor suppressor locus. Two of these, 42D6 and 46E6, were submitted to the Breast Cancer Linkage Consortium for typing of breast cancer families by investigators worldwide. Oligonucleotide sequences for markers not developed in our laboratory were obtained from published reports, or as part of the Breast Cancer Linkage Consortium, or from other investigators. All genotyping films were scored blindly with a standard lane marker used to maintain consistent coding of alleles. Key samples in the four kindreds presented here underwent duplicate typing for all relevant markers. All 19 kindreds have been typed for two polymorphic CA repeat markers: 42D6 (D17S588), a CA repeat isolated in our laboratory, and Mfd15 (D17S250), a CA repeat provided by J. Weber (Weber *et al.*, 1990). Several sources of probes were used to create genetic markers on chromosome 17, specifically chromosome 17 cosmid and lambda phage libraries created from sorted chromosomes by the Los Alamos National Laboratories (van Dilla *et al.*, 1986).

LOD scores for each kindred with these two markers (42D6, Mfd15) and a third marker, Mfd188 (D17S579, Hall *et al.*, 1992), located roughly midway between these two markers, were calculated for two values of the recombination fraction, 0.001 and 0.1. (For calculation of LOD scores, *see* Oh, 1985). Likelihoods were computed under the model derived by Claus *et al.*, 1991, which assumes an estimated gene frequency of 0.003, a lifetime risk in gene carriers of about 0.80, and population based age-specific risks for breast cancer in non-gene carriers. Allele frequencies for the three markers used for the LOD score calculations were calculated from our own laboratory typings of unrelated

individuals in the CEPH panel (White and Lalouel, 1988). Table 2 shows the results of the pairwise linkage analysis of each kindred with the three markers 42D6, Mfdl88, and Mfdl5.

TABLE 2

Pairwise Linkage Analysis of Kindreds

KINDRED	Mfd15 (D17S250) Recombination		Mfd188 (D17S579) Recombination		42D6 (D17S588) Recombination	
	0.001	0.1	0.001	0.1	0.001	0.1
1910	0.06	0.30	0.06	0.30	0.06	0.30
1001	-0.30	-0.09	NT	NT	-0.52	-0.19
2035	2.34	1.85	0.94	0.90	2.34	1.82
2027	-1.22	-0.33	-1.20	-0.42	-1.16	-0.33
9018	-0.54	-0.22	-0.17	-0.10	0.11	0.07
1925	1.08	0.79	0.55	0.38	-0.11	-0.07
1927	-0.41	0.01	-0.35	0.07	-0.44	-0.02
1911	-0.27	-0.13	-0.43	-0.23	0.49	0.38
1929	-0.49	-0.25	NT	NT	-0.49	-0.25
1901	1.50	1.17	0.78	0.57	0.65	0.37
2082	4.25	3.36	6.07	5.11	2.00	3.56
2019	-0.10	-0.01	-0.11	-0.05	-0.18	-0.10
1900	-0.14	-0.11	NT	NT	-0.12	-0.05
2080	-0.16	-0.04	0.76	0.74	-1.25	-0.58
2073	-0.41	-0.29	0.63	0.49	-0.23	-0.13
1917	-0.02	-0.02	NT	NT	-0.01	0.00
1920	-0.03	-0.02	NT	NT	0.00	0.00
2079	0.02	0.01	-0.01	-0.01	0.01	0.01
2039	-1.67	-0.83	0.12	0.59	-1.15	0.02

NT - Kindred not typed for Mfd188.

Using a criterion for linkage to 17q of a LOD score > 1.0 for at least one locus under the CASH model (Claus *et al.*, 1991), four of the 19 kindreds appeared to be linked to 17q (K1901, K1925, K2035, K2082). A number of additional kindreds showed some evidence of linkage but at this time could not be definitively assigned to the linked category. These included kindreds K1911, K2073, K2039, and K2080. Three of the 17q-linked kindreds had informative recombinants in this region and these are detailed below.

Kindred 2082 is the largest 17q-linked breast cancer family reported to date by any group. The kindred contains 20 cases of breast cancer, and ten cases of ovarian cancer. Two cases have both ovarian and breast cancer. The evidence of linkage to 17q for this family is overwhelming; the LOD score with the linked haplotype is over 6.0, despite the existence of three cases of breast cancer which appear to be sporadic, i.e., these cases share no part of the linked haplotype between Mfd15 and 42D6. These three sporadic cases were diagnosed with breast cancer at ages 46, 47, and 54. In smaller kindreds, sporadic cancers of this type greatly confound the analysis of linkage and the correct identification of key recombinants. The key recombinant in the 2082 kindred is a woman who developed ovarian cancer at age 45 whose mother and aunt had ovarian cancer at ages 58 and 66, respectively. She inherited the linked portion of the haplotype for both Mfd188 and 42D6 while inheriting unlinked alleles at Mfd15; this recombinant event placed BRCA1 distal to Mfd15.

K1901 is typical of early-onset breast cancer kindreds. The kindred contains 10 cases of breast cancer with a median age at diagnosis of 43.5 years of age; four cases were diagnosed under age 40. The LOD score for this kindred with the marker 42D6 is 1.5, resulting in a posterior probability of 17q-linkage of 0.96. Examination of haplotypes in this kindred identified a recombinant haplotype in an obligate male carrier and his affected daughter who was diagnosed with breast cancer at age 45. Their linked allele for marker Mfd15 differs from that found in all other cases in the kindred (except one case which could not be completely inferred from her children). The two haplotypes are identical for Mfd188 and 42D6. Accordingly, data from Kindred 1901 would also place the BRCA1 locus distal to Mfd15.

Kindred 2035 is similar to K1901 in disease phenotype. The median age of diagnosis for the eight cases of breast cancer in this kindred is 37. One case also had ovarian cancer at age 60. The breast cancer cases in this family descend from two sisters who were both unaffected with breast cancer until their death in the eighth decade. Each branch contains four cases of breast cancer with at least one case in each branch having markedly early onset. This kindred has a LOD score of 2.34 with Mfd15.

GeneDX 1009, pg. 59

The haplotypes segregating with breast cancer in the two branches share an identical allele at Mfd15 but differ for the distal loci Mfd188 and NM23 (a marker typed as part of the consortium which is located just distal to 42D6 (Hall *et al.*, 1992)). Although the two haplotypes are concordant for marker 42D6, it is likely that the alleles are shared identical by state (the same allele but derived from different ancestors), rather than identical by descent (derived from a common ancestor) since the shared allele is the second most common allele observed at this locus. By contrast the linked allele shared at Mfd15 has a frequency of 0.04. This is a key recombinant in our dataset as it is the sole recombinant in which BRCA1 segregated with the proximal portion of the haplotype, thus setting the distal boundary to the BRCA1 region. For this event not to be a key recombinant requires that a second mutant BRCA1 gene be present in a spouse marrying into the kindred who also shares the rare Mfd15 allele segregating with breast cancer in both branches of the kindred. This event has a probability of less than one in a thousand. The evidence from this kindred therefore placed the BRCA1 locus proximal to Mfd188.

EXAMPLE 3

Creation of a Fine Structure Map and Refinement of the BRCA1 Region to Mfd191-Mfd188 using Additional STR Polymorphisms

In order to improve the characterization of our recombinants and define closer flanking markers, a dense map of this relatively small region on chromosome 17q was required. The chromosome 17 workshop has produced a consensus map of this region (Figure 1) based on a combination of genetic and physical mapping studies (Fain, 1992). This map contains both highly polymorphic STR polymorphisms, and a number of nonpolymorphic expressed genes. Because this map did not give details on the evidence for this order nor give any measure of local support for inversions in the order of adjacent loci, we viewed it as a rough guide for obtaining resources to be used for the development of new markers and construction of our own detailed genetic and physical map of a small region containing BRCA1. Our approach was to analyze existing STR markers provided by other investigators and any newly developed markers from our laboratory with respect to both a panel of meiotic (genetic) breakpoints identified using DNA from the CEPH reference families and a panel of somatic cell hybrids (physical breakpoints) constructed for this region. These markers included 26C2 developed in our laboratory which maps proximal to Mfd15, Mfd19l (provided by James Weber), THRA1 (Futreal

et al., 1992a), and three polymorphisms kindly provided to us by Dr. Donald Black, NM23 (Hall *et al.* 1992), SCG40 (D17S181), and 6C1 (D17S293).

Genetic localization of markers. In order to localize new markers genetically within the region of interest, we have identified a number of key meiotic breakpoints within the region, both in the CEPH reference panel and in our large breast cancer kindred (K2082). Given the small genetic distance in this region, they are likely to be only a relatively small set of recombinants which can be used for this purpose, and they are likely to group markers into sets. The orders of the markers within each set can only be determined by physical mapping. However the number of genotypings necessary to position a new marker is minimized. These breakpoints are illustrated in Tables 3 and 4. Using this approach we were able to genetically order the markers THRA1, 6C1, SCG40, and Mfd19l. As can be seen from Tables 3 and 4, THRA1 and MFD191 both map inside the Mfd15-Mfd188 region we had previously identified as containing the BRCA1 locus. In Tables 3 and 4, M/P indicates a maternal or paternal recombinant. A "1" indicates inherited allele is of grandpaternal origin, while a "0" indicates grandmaternal origin, and "-" indicates that the locus was untyped or uninformative.

TABLE 3

CEPH Recombinants

Family	ID	M/P	Mfd15	THRA1	Mfd191	Mfd188	SCG40	6C1	42D6
13292	4	M	1	1	1	0	0	0	0
13294	4	M	1	1	1	0	0	0	0
13294	6	M	0	0	1	1	-	-	-
1334	3	M	1	1	1	1	1	0	0
1333	4	M	1	1	1	0	-	-	0
1333	6	M	0	0	1	1	-	-	1
1333	8	P	1	0	0	0	-	-	0
1377	8	M	0	-	0	0	0	0	1

GeneDX 1009, pg. 62

TABLE 4

Kindred 2082 Recombinants

Family	ID	M/P	Mfd15	Mfd191	Mfd188	SCG40	6C1	42D6
75		M	0	1	1	1	-	-
63		M	0	0	1	1	-	1
125		M	1	1	1	0	-	0
40		M	1	1	0	0	-	0

GeneDX 1009, pg. 63

Analysis of markers Mfd15, Mfd188, Mfd191, and THRA1 in our recombinant families. Mfd15, Mfd188, Mfd191 and THRA1 were typed in our recombinant families and examined for additional information to localize the BRCA1 locus. In kindred 1901, the Mfdl5 recombinant was recombinant for THRA1 but uninformative for Mfd19l, thus placing BRCA1 distal to THRA1. In K2082, the recombinant with Mfd15 also was recombinant with Mfd191, thus placing the BRCA1 locus distal to Mfd191 (Goldgar *et al.*, 1994). Examination of THRA1 and Mfd191 in kindred K2035 yielded no further localization information as the two branches were concordant for both markers. However, SCG40 and 6C1 both displayed the same pattern as Mfd188, thus increasing our confidence in the localization information provided by the Mfd188 recombinant in this family. The BRCA1 locus, or at least a portion of it, therefore lies within an interval bounded by Mfd191 on the proximal side and Mfd188 on the distal side.

EXAMPLE 4

Development of Genetic and Physical Resources in the Region of Interest

To increase the number of highly polymorphic loci in the Mfd19l-Mfd188 region, we developed a number of STR markers in our laboratory from cosmids and YACs which physically map to the region. These markers allowed us to further refine the region.

STSs were identified from genes known to be in the desired region to identify YACs which contained these loci, which were then used to identify subclones in cosmids, P1s or BACs. These subclones were then screened for the presence of a CA tandem repeat using a $(CA)_n$ oligonucleotide (Pharmacia). Clones with a strong signal were selected preferentially, since they were more likely to represent CA-repeats which have a large number of repeats and/or are of near-perfect fidelity to the $(CA)_n$ pattern. Both of these characteristics are known to increase the probability of polymorphism (Weber, 1990). These clones were sequenced directly from the vector to locate the repeat. We obtained a unique sequence on one side of the CA-repeat by using one of a set of possible primers complementary to the end of a CA-repeat, such as $(GT)_{10}T$. Based on this unique sequence, a primer was made to sequence back across the repeat in the other direction, yielding a unique sequence for design of a second primer flanking the CA-repeat. STRs were then screened for polymorphism on a small group of unrelated individuals and tested against the hybrid panel to confirm their physical

localization. New markers which satisfied these criteria were then typed in a set of 40 unrelated individuals from the Utah and CEPH families to obtain allele frequencies appropriate for the study population. Many of the other markers reported in this study were tested in a smaller group of CEPH unrelated individuals to obtain similarly appropriate allele frequencies.

Using the procedure described above, a total of eight polymorphic STRs was found from these YACS. Of the loci identified in this manner, four were both polymorphic and localized to the BRCA1 region. Four markers did not localize to chromosome 17, reflecting the chimeric nature of the YACs used. The four markers which were in the region were denoted AA1, ED2, 4-7, and YM29. AAl and ED2 were developed from YACs positive for the RNU2 gene, 4-7 from an EPB3 YAC and YM29 from a cosmid which localized to the region by the hybrid panel. A description of the number of alleles, heterozygosity and source of these four and all other STR polymorphisms analyzed in the breast cancer kindreds is given below in Table 5.

TABLE 5

Polymorphic Short Tandem Repeat Markers Used
for Fine Structure Mapping of the BRCA1 Locus

Clone	Gene	Na**	Hetero-zygosity	Allele* Frequency (%) 1	2	3	4	5	6
Mfd15	D17S250	10	0.82	26	22	15	7	7	23
THRA1	THRA1	5							
Mfd191	D17S776	7	0.55	48	20	11	7	7	7
ED2	D17S1327	12	0.55	62	9	8	5	5	11
AA1	D17S1326	7	0.83	28	28	25	8	6	5
CA375	D17S184	10	0.75	26	15	11	9	9	20
4-7	D17S1183	9	0.50	63	15	8	6	4	4
YM29	--	9	0.62	42	24	12	7	7	8
Mfd188	D17S579	12	0.92	33	18	8	8	8	25
SCG40	D17S181	14	0.90	20	18	18	10	8	35
42D6	D17S588	11	0.86	21	17	11	10	9	32
6C1	D17S293	7	0.75	30	30	11	11	9	9
Z109	D17S750	9	0.70	33	27	7	7	7	19
tdj1475	D17S1321	13	0.84	21	16	11	11	8	33
CF4	D17S1320	6	0.63	50	27	9	7	4	3
tdj1239	D17S1328	10	0.80	86	10	9	7	4	14
U5	D17S1325	13	0.83	19	16	12	10	9	34

* Allele codes 1-5 are listed in decreasing frequency; allele numbers do not correspond to fragment sizes. Allele 6 frequency is the joint frequency of all other alleles for each locus.

** Number of alleles seen in the genetically independent DNA samples used for calculating allele frequencies.

GeneDX 1009, pg. 66

The four STR polymorphisms which mapped physically to the region (4-7, ED2, AA1, YM29) were analyzed in the meiotic, breakpoint panel shown initially in Tables 3 and 4. Tables 6 and 7 contain the relevant CEPH data and Kindred 2082 data for localization of these four markers. In the tables, M/P indicates a maternal or paternal recombinant. A "1" indicates inherited allele is of grandpaternal origin, while a "0" indicates grandmaternal origin, and "-" indicates that the locus was untyped or uninformative.

TABLE 6

Key Recombinants Used for Genetic
Ordering of New STR Loci Developed in
Our Laboratory Within the BRCA1 Region of 17q

CEPH Family	ID	M/P	Mfd15	THRA1	Mfd191	ED2	AA1	Z109	4-7	YM29	Mfd188	SCG40	42D6
13292	4	M	1	1	1	1	1	0	0	0	0	0	0
13294	4	M	1	0	0	-	0	-	-	-	0	-	-
13294	6	M	0	0	1	-	1	-	-	-	1	-	-
1333	4	M	1	1	1	-	0	-	-	0	0	-	0
1333	6	M	0	0	1	-	1	-	-	1	1	-	1
1333	3	M	0	0	1	-	-	-	1	1	1	-	1

GeneDX 1009, pg. 68

TABLE 7

Kindred 2082 Recombinants

ID	M/P	Mfd15	Mfd191	ED2	AA1	4-7	YM29	Mfd188	SCG40	42D6
63	M	0	0	1	-	1	1	1	1	1
125	M	1	1	1	-	1	1	1	0	0
40	M	1	1	0	-	0	-	0	0	0
22	P	0	0	1	1	1	1	1	1	1

GeneDX 1009, pg. 69

From CEPH 1333-04, we see that AAl and YM29 must lie distal to Mfd191. From 13292, it can be inferred that both AA1 and ED2 are proximal to 4-7, YM29, and Mfd188. The recombinants found in K2082 provide some additional ordering information. Three independent observations (individual numbers 22, 40, & 63) place AA1, ED2, 4-7, and YM29, and Mfdl88 distal to Mfd191, while ID 125 places 4-7, YM29, and Mfd188 proximal to SCG40. No genetic information on the relative ordering within the two clusters of markers AA1/ED2 and 4-7/YM29/Mfd188 was obtained from the genetic recombinant analysis. Although ordering loci with respect to hybrids which are known to contain "holes" in which small pieces of interstitial human DNA may be missing is problematic, the hybrid patterns indicate that 4-7 lies above both YM29 and Mfd188.

EXAMPLE 5

Genetic Analyses of Breast Cancer
Kindreds with Markers AA1, 4-7, ED2, and YM29

In addition to the three kindreds containing key recombinants which have been discussed previously, kindred K2039 was shown through analysis of the newly developed STR markers to be linked to the region and to contain a useful recombinant.

Table 8 defines the haplotypes (shown in coded form) of the kindreds in terms of specific marker alleles at each locus and their respective frequencies. In Table 8, alleles are listed in descending order of frequency; frequencies of alleles 1-5 for each locus are given in Table 5. Haplotypes coded H are BRCA1 associated haplotypes, P designates a partial H haplotype, and an R indicates an observable recombinant haplotype. As evident in Table 8, not all kindreds were typed for all markers; moreover, not all individuals within a kindred were typed for an identical set of markers, especially in K2082. With one exception, only haplotypes inherited from affected or at-risk kindred members are shown; haplotypes from spouses marrying into the kindred are not described. Thus in a given sibship, the appearance of haplotypes X and Y indicates that both haplotypes from the affected/at-risk individual were seen and neither was a breast cancer associated haplotype.

TABLE 8

Breast Cancer Linked Haplotypes
Found in the Three Kindreds

Kin.	HAP	Mfd 15	THRA1	Mfd 191	tdj 1475	ED2	AA1	Z109	CA375	4-7	YM29	Mfd 188	SCG40	6C1	42D6
1901	H1	1	5	5	3	1	4	NI	NI	1	1	3	NI	NI	1
	R2	9	2	5	6	1	4	NI	NI	1	1	3	NI	NI	1
2082	H1	3	NI	4	6	6	1	NI	NI	2	1	4	2	NI	1
	P1	3	NI	4	NI	NI	NI	NI	NI	NI	NI	4	2	NI	1
	P2	3	NI	NI	NI	NI	NI	NI	NI	NI	NI	4	NI	NI	NI
	R1	6	NI	1	5	6	1	NI	NI	2	1	4	2	NI	1
	R2	6	NI	4	6	6	1	NI	NI	2	1	4	2	NI	1
	R3	3	NI	4	NI	6	1	NI	NI	2	1	4	1	NI	7
	R4	7	NI	1	NI	1	5	NI	NI	4	6	1	2	NI	1
	R5	3	NI	4	NI	NI	NI	NI	NI	NI	2	1	NI	NI	NI
	R6	3	NI	4	3	1	2	NI	NI	1	2	2	6	NI	6
	R7	3	NI	4	3	7	1	NI	NI	1	1	3	7	NI	4
2035	HI	8	2	1	NI	5	1	1	4	3	1	6	8	2	4
	H2	8	2	1	NI	5	1	1	2	1	1	2	3	1	4
	R2	8	2	1	NI	5	1	1	2	1	1	2	3	6	1

GeneDX 1009, pg. 71

In kindred K1901, the new markers showed no observable recombination with breast cancer susceptibility, indicating that the recombination event in this kindred most likely took place between THRA1 and ED2. Thus, no new BRCA1 localization information was obtained based upon studying the four new markers in this kindred. In kindred 2082 the key recombinant individual has inherited the linked alleles for ED2, 4-7, AA1, and YM29, and was recombinant for tdj1474 indicating that the recombination event occurred in this individual between tdj1474 and ED2/AA1.

There are three haplotypes of interest in kindred K2035, Hl, H2, and R2 shown in Table 8. H1 is present in the four cases and one obligate male carrier descendant from individual 17 while H2 is present or inferred in two cases and two obligate male carriers in descendants of individual 10. R2 is identical to H2 for loci between and including Mfd15 and SCG40, but has recombined between SCG40 and 42D6. Since we have established that BRCA1 is proximal to 42D6, this H2/R2 difference adds no further localization information. H1 and R2 share an identical allele at Mfd15, THRA1, AA1, and ED2 but differ for loci presumed distal to ED2, i.e., 4-7, Mfd188, SCG40, and 6C1. Although the two haplotypes are concordant for the 5th allele for marker YM29, a marker which maps physically between 4-7 and Mfd188, it is likely that the alleles are shared identical by state rather than identical by descent since this allele is the most common allele at this locus with a frequency estimated in CEPH parents of 0.42. By contrast, the linked alleles shared at the Mfd15 and ED2 loci have frequencies of 0.04 and 0.09, respectively. They also share more common alleles at Mfdl9l (frequency = 0.52), THRA1, and AA1 (frequency = 0.28). This is the key recombinant in the set as it is the sole recombinant in which breast cancer segregated with the proximal portion of the haplotype, thus setting the distal boundary. The evidence from this kindred therefore places the BRCA1 locus proximal to 4-7.

The recombination event in kindred 2082 which places BRCA1 distal to tdj1474 is the only one of the four events described which can be directly inferred; that is, the affected mother's genotype can be inferred from her spouse and offspring, and the recombinant haplotype can be seen in her affected daughter. In this family the odds in favor of affected individuals carrying BRCA1 susceptibility alleles are extremely high; the only possible interpretations of the data are that BRCA1 is distal to Mfd191 or alternatively that the purported recombinant is a sporadic case of ovarian cancer at age 44. Rather than a directly observable or inferred recombinant, interpretation of kindred 2035 depends on the observation of distinct 17q-haplotypes segregating in different and sometimes distantly related branches of the kindred. The observation that portions of these haplotypes have alleles in common for some markers while they differ at other markers places the BRCA1 locus in the shared region. The

GeneDX 1009, pg. 72

confidence in this placement depends on several factors: the relationship between the individuals carrying the respective haplotypes, the frequency of the shared allele, the certainty with which the haplotypes can be shown to segregate with the BRCA1 locus, and the density of the markers in the region which define the haplotype. In the case of kindred 2035, the two branches are closely related, and each branch has a number of early onset cases which carry the respective haplotype. While two of the shared alleles are common, (Mfd191, THRA1), the estimated frequencies of the shared alleles at Mfd15, AA1, and ED2 are 0.04, 0.28, and 0.09, respectively. It is therefore highly likely that these alleles are identical by descent (derived from a common ancestor) rather than identical by state (the same allele but derived from the general population).

EXAMPLE 6

Refined Physical Mapping Studies Place the BRCA1 Gene in a Region Flanked by tdj1474 and U5R

Since its initial localization to chromosome 17q in 1990 (Hall *et al.*, 1990) a great deal of effort has gone into localizing the BRCA1 gene to a region small enough to allow implementation of effective positional cloning strategies to isolate the gene. The BRCA1 locus was first localized to the interval Mfd15 (D17S250) - 42D6 (D17S588) by multipoint linkage analysis (Easton *et al.*, 1993) in the collaborative Breast Cancer Linkage Consortium dataset consisting of 214 families collected worldwide. Subsequent refinements of the localization have been based upon individual recombinant events in specific families. The region THRA1 - D17S183 was defined by Bowcock *et al.*, 1993; and the region THRA1 - D17S78 was defined by Simard *et al.*, 1993.

We further showed that the BRCA1 locus must lie distal to the marker Mfd191 (D17S776) (Goldgar *et al.*, 1994). This marker is known to lie distal to THRA1 and RARA. The smallest published region for the BRCA1 locus is thus between D17S776 and D17S78. This region still contains approximately 1.5 million bases of DNA, making the isolation and testing of all genes in the region a very difficult task. We have therefore undertaken the tasks of constructing a physical map of the region, isolating a set of polymorphic STR markers located in the region, and analyzing these new markers in a set of informative families to refine the location of the BRCA1 gene to a manageable interval.

GeneDX 1009, pg. 73

Four families provide important genetic evidence for localization of BRCA1 to a sufficiently small region for the application of positional cloning strategies. Two families (K2082, K1901) provide data relating to the proximal boundary for BRCA1 and the other two (K2035, K1813) fix the distal boundary. These families are discussed in detail below. A total of 15 Short Tandem Repeat markers assayable by PCR were used to refine this localization in the families studied. These markers include DS17S7654, DS17S975, tdj1474, and tdj1239. Primer sequences for these markers are provided in SEQ ID NO:3 and SEQ ID NO:4 for DS17S754; in SEQ ID NO:5 and SEQ ID NO:6 for DS17S975; in SEQ ID NO:7 and SEQ ID NO:8 for tdj1474; and, in SEQ ID NO:9 and SEQ ID NO:10 for tdj1239.

Kindred 2082

Kindred 2082 is the largest BRCA1-linked breast/ovarian cancer family studied to date. It has a LOD score of 8.6, providing unequivocal evidence for 17q linkage. This family has been previously described and shown to contain a critical recombinant placing BRCA1 distal to MFD191 (D17S776). This recombinant occurred in a woman diagnosed with ovarian cancer at age 45 whose mother had ovarian cancer at age 63. The affected mother was deceased; however, from her children, she could be inferred to have the linked haplotype present in the 30 other linked cases in the family in the region between Mfd15 and Mfd188. Her affected daughter received the linked allele at the loci ED2, 4-7, and Mfd188, but received the allele on the non-BRCA1 chromosome at Mfd15 and Mfd191. In order to further localize this recombination breakpoint, we tested DNA from the key members of this family for the following markers derived from physical mapping resources: tdj1474, tdj1239, CF4, D17S855. For the markers tdj1474 and CF4, the affected daughter did not receive the linked allele. For the STR locus tdj1239, however, the mother could be inferred to be informative and her daughter did receive the BRCA1-associated allele. D17S855 was not informative in this family. Based on this analysis, the order is 17q centromere - Mfd191 - 17HSD - CF4 - tdj1474 - tdj1239 - D17S855 - ED2 - 4-7 - Mfd188 - 17q telomere. The recombinant described above therefore places BRCA1 distal to tdj1474, and the breakpoint is localized to the interval between tdj1474 and tdj1239. The only alternative explanation for the data in this family other than that of BRCA1 being located distal to tdj1474, is that the ovarian cancer present in the recombinant individual is caused by reasons independent of the BRCA1 gene. Given that ovarian cancer diagnosed before age 50 is rare, this alternate explanation is exceedingly unlikely.

GeneDX 1009, pg. 74

Kindred 1901

Kindred 1901 is an early-onset breast cancer family with 7 cases of breast cancer diagnosed before 50, 4 of which were diagnosed before age 40. In addition, there were three cases of breast cancer diagnosed between the ages of 50 and 70. One case of breast cancer also had ovarian cancer at age 61. This family currently has a LOD score of 1.5 with D17S855. Given this linkage evidence and the presence of at lease one ovarian cancer case, this family has a posterior probability of being due to BRCA1 of over 0.99. In this family, the recombination comes from the fact that an individual who is the brother of the ovarian cancer case from which the majority of the other cases descend, only shares a portion of the haplotype which is cosegregating with the other cases in the family. However, he passed this partial haplotype to his daughter who developed breast cancer at age 44. If this case is due to the BRCA1 gene, then only the part of the haplotype shared between this brother and his sister can contain the BRCA1 gene. The difficulty in interpretation of this kind of information is that while one can be sure of the markers which are not shared and therefore recombinant, markers which are concordant can either be shared because they are non-recombinant, or because their parent was homozygous. Without the parental genotypic data it is impossible to discriminate between these alternatives. Inspection of the haplotype in K1901, shows that he does not share the linked allele at Mfd15 (D17S250), THRA1, CF4 (D17S1320), and tdj1474 (17DS1321). He does share the linked allele at Mfd191 (D17S776), ED2 (D17S1327), tdj1239 (D17S1328), and Mfd188 (D17S579). Although the allele shared at Mfd191 is relatively rare (0.07), we would presume that the parent was homozygous since they are recombinant with markers located nearby on either side, and a double recombination event in this region would be extremely unlikely. Thus the evidence in this family would also place the BRCA1 locus distal to tdj1474. However, the lower limit of this breakpoint is impossible to determine without parental genotype information. It is intriguing that the key recombinant breakpoint in this family confirms the result in Kindred 2082. As before, the localization information in this family is only meaningful if the breast cancer was due to the BRCA1 gene. However, her relatively early age at diagnosis (44) makes this seem very likely since the risk of breast cancer before age 45 in the general population is low (approximately 1%).

Kindred 2035

This family is similar to K1901 in that the information on the critical recombinant events is not directly observed but is inferred from the observation that the two haplotypes which are cosegregating

GeneDX 1009, pg. 75

with the early onset breast cancer in the two branches of the family appear identical for markers located in the proximal portion of the 17q BRCA1 region but differ at more distal loci. Each of these two haplotypes occurs in at least four cases of early-onset or bilateral breast cancer. The overall LOD score with ED2 in this family is 2.2, and considering that there is a case of ovarian cancer in the family (indicating a prior probability of BRCA1 linkage of 80%), the resulting posterior probability that this family is linked to BRCA1 is 0.998. The haplotypes are identical for the markers Mfd15, THRA1, Mfd191, ED2, AA1, D17S858 and D17S902. The common allele at Mfd15 and ED2 are both quite rare, indicating that this haplotype is shared identical by descent. The haplotypes are discordant, however, for CA375, 4-7, and Mfd188, and several more distal markers. This indicates that the BRCA1 locus must lie above the marker CA-375. This marker is located approximately 50 kb below D17S78, so it serves primarily as additional confirmation of this previous lower boundary as reported in Simard *et al.* (1993).

Kindred 1813

Kindred 1813 is a small family with four cases of breast cancer diagnosed ~~under the age of 40~~ at very early ages whose mother also had breast cancer diagnosed ~~at age 45~~ at an early age some years later and ovarian cancer at age 61. This situation is somewhat complicated by the fact the four cases appear to have three different fathers, only one of whom has been genotyped. However, by typing a number of different markers in the BRCA1 region as well as highly polymorphic markers elsewhere in the genome, the paternity of all children in the family has been determined with a high degree of certainty. This family yields a maximum multipoint LOD score of 0.60 with 17q markers and, given that there is at least one case of ovarian cancer, results in a posterior probability of being a BRCA1 linked family of 0.93. This family contains a directly observable recombination event in individual 18 (see Figure 5 in Simard *et al.*, *Human Mol. Genet.* 2:1193-1199 (1993)), who developed breast cancer at age 34. The genotype of her affected mother at the relevant 17q loci can be inferred from her genotypes, her affected sister's genotypes, and the genotypes of three other unaffected siblings. Individual 18 inherits the BRCA1-linked alleles for the following loci: Mfd15, THRA1, D17S800, D17S855, AA1, and D17S931. However, for markers below D17S931, i.e., U5R, vrs31, D17S858, and D17S579, she has inherited the alleles located on the non-disease bearing chromosome. The evidence from this family therefore would place the BRCA1 locus proximal to the marker U5R. Because of her early age at diagnosis (34) it is extremely unlikely that the recombinant individual's cancer is not due to the gene responsible for the other cases of

GeneDX 1009, pg. 76

breast/ovarian cancer in this family; the uncertainty in this family comes from our somewhat smaller amount of evidence that breast cancer in this family is due to BRCA1 rather than a second, as yet unmapped, breast cancer susceptibility locus.

Size of the region containing BRCA1

Based on the genetic data described in detail above, the BRCA1 locus must lie in the interval between the markers tdj1474 and U5R, both of which were isolated in our laboratory. Based upon the physical maps shown in Figures 2 and 3, we can try to estimate the physical distance between these two loci. It takes approximately 14 P1 clones with an average insert size of approximately 80 kb to span the region. However, because all of these P1s overlap to some unknown degree, the physical region is most likely much smaller than 14 times 80 kb. Based on restriction maps of the clones covering the region, we estimate the size of the region containing BRCA1 to be approximately 650 kb.

EXAMPLE 7

Identification of Candidate cDNA Clones for the BRCA1 Locus by Genomic Analysis of the Contig Region

Complete screen of the plausible region. The first method to identify candidate cDNAs, although labor intensive, used known techniques. The method comprised the screening of cosmids and P1 and BAC clones in the contig to identify putative coding sequences. The clones containing putative coding sequences were then used as probes on filters of cDNA libraries to identify candidate cDNA clones for future analysis. The clones were screened for putative coding sequences by either of two methods.

Zoo blots. The first method for identifying putative coding sequences was by screening the cosmid and Pl clones for sequences conserved through evolution across several species. This technique is referred to as "zoo blot analysis" and is described by Monaco, 1986. Specifically, DNAs from cow, chicken, pig, mouse and rat were digested with the restriction enzymes EcoRI and HindIII (8 μg of DNA per enzyme). The digested DNAs were separated overnight on an 0.7% gel at 20 volts for 16 hours (14 cm gel), and the DNA transferred to Nylon membranes using standard Southern blot techniques. For example, the zoo blot filter was treated at 65°C in 0.1 x SSC, 0.5% SDS, and 0.2M Tris, pH 8.0, for 30 minutes and then blocked overnight at 42°C in 5x SSC, 10% PEG 8000, 20 mM

GeneDX 1009, pg. 77

$NaPO_4$ pH 6.8, 100 μg/ml Salmon Sperm DNA, lx Denhardt's, 50% formamide, 0.1% SDS, and 2 μg/ml C_0t-1 DNA.

The cosmid and Pl clones to be analyzed were digested with a restriction enzyme to release the human DNA from the vector DNA. The DNA was separated on a 14 cm, 0.5% agarose gel run overnight at 20 volts for 16 hours. The human DNA bands were cut out of the gel and electroeluted from the gel wedge at 100 volts for at least two hours in 0.5x Tris Acetate buffer (Maniatis *et al.*, 1982). The eluted Not I digested DNA (~15 kb to 25 kb) was then digested with EcoRI restriction enzyme to give smaller fragments (~0.5 kb to 5.0 kb) which melt apart more easily for the next step of labeling the DNA with radionucleotides. The DNA fragments were labeled by means of the hexamer random prime labeling method (Boehringer-Mannheim, Cat. #1004760). The labeled DNA was spermine precipitated (add 100 μl TE, 5 μl 0.1 M spermine, and 5 μl of 10 mg/ml salmon sperm DNA) to remove unincorporated radionucleotides. The labeled DNA was then resuspended in 100 μl TE, 0.5 M NaCl at 65°C for 5 minutes and then blocked with Human C_0t-1 DNA for 2-4 hrs. as per the manufacturer's instructions (Gibco/BRL, Cat. #5279SA). The C_0t-1 blocked probe was incubated on the zoo blot filters in the blocking solution overnight at 42°C. The filters were washed for 30 minutes at room temperature in 2 x SSC, 0.1% SDS, and then in the same buffer for 30 minutes at 55°C. The filters were then exposed 1 to 3 days at -70°C to Kodak XAR-5 film with an intensifying screen. Thus, the zoo blots were hybridized with either the pool of Eco-Rl fragments from the insert, or each of the fragments individually.

HTF island analysis. The second method for identifying cosmids to use as probes on the cDNA libraries was HTF island analysis. Since the pulsed-field map can reveal HTF islands, cosmids that map to these HTF island regions were analyzed with priority. HTF islands are segments of DNA which contain a very high frequency of unmethylated CpG dinucleotides (Tonolio *et al.*, 1990) and are revealed by the clustering of restriction sites of enzymes whose recognition sequences include CpG dinucleotides. Enzymes known to be useful in HTF-island analysis are AscI, NotI, BssHII, EagI, SacII, NaeI, NarI, SmaI, and MluI (Anand, 1992). A pulsed-field map was created using the enzymes NotI, NruI, EagI, SacII, and SalI, and two HTF islands were found. These islands are located in the distal end of the region, one being distal to the GP2B locus, and the other being proximal to the same locus, both outside the BRCA1 region. The cosmids derived from the YACs that cover these two locations were analyzed to identify those that contain these restriction sites, and thus the HTF islands.

GeneDX 1009, pg. 78

cDNA screening. Those clones that contain HTF islands or show hybridization to other species DNA besides human are likely to contain coding sequences. The human DNA from these clones was isolated as whole insert or as EcoRl fragments and labeled as described above. The labeled DNA was used to screen filters of various cDNA libraries under the same conditions as the zoo blots except that the cDNA filters undergo a more stringent wash of 0.1 x SSC, 0.1% SDS at 65°C for 30 minutes twice.

Most of the cDNA libraries used to date in our studies (libraries from normal breast tissue, breast tissue from a woman in her eighth month of pregnancy and a breast malignancy) were prepared at Clonetech, Inc. The cDNA library generated from breast tissue of an 8 month pregnant woman is available from Clonetech (Cat. #HL1037a) in the Lambda gt-10 vector, and is grown in C600Hfl bacterial host cells. Normal breast tissue and malignant breast tissue samples were isolated from a 37 year old Caucasian female and one-gram of each tissue was sent to Clonetech for mRNA processing and cDNA library construction. The latter two libraries were generated using both random and oligo-dT priming, with size selection of the final products which were then cloned into the Lambda Zap II vector, and grown in XL1-blue strain of bacteria as described by the manufacturer. Additional tissue-specific cDNA libraries include human fetal brain (Stratagene, Cat. 936206), human testis (Clonetech Cat. HL3024), human thymus (Clonetech Cat. HL1127n), human brain (Clonetech Cat. HL11810), human placenta (Clonetech Cat 1075b), and human skeletal muscle (Clonetech Cat. HL1124b).

The cDNA libraries were plated with their host cells on NZCYM plates, and filter lifts are made in duplicate from each plate as per Maniatis *et al.* (1982). Insert (human) DNA from the candidate genomic clones was purified and radioactively labeled to high specific activity. The radioactive DNA was then hybridized to the cDNA filters to identify those cDNAs which correspond to genes located within the candidate cosmid clone. cDNAs identified by this method were picked, replated, and screened again with the labeled clone insert or its derived EcoRl fragment DNA to verify their positive status. Clones that were positive after this second round of screening were then grown up and their DNA purified for Southern blot analysis and sequencing. Clones were either purified as plasmid through *in vivo* excision of the plasmid from the Lambda vector as described in the protocols from the manufacturers, or isolated from the Lambda vector as a restriction fragment and subcloned into plasmid vector.

The Southern blot analysis was performed in duplicate, one using the original genomic insert DNA as a probe to verify that cDNA insert contains hybridizing sequences. The second blot was hybridized with cDNA insert DNA from the largest cDNA clone to identify which clones represent the

GeneDX 1009, pg. 79

same gene. All cDNAs which hybridize with the genomic clone and are unique were sequenced and the DNA analyzed to determine if the sequences represent known or unique genes. All cDNA clones which appear to be unique were further analyzed as candidate BRCA1 loci. Specifically, the clones are hybridized to Northern blots to look for breast specific expression and differential expression in normal versus breast tumor RNAs. They are also analyzed by PCR on clones in the BRCA1 region to verify their location. To map the extent of the locus, full length cDNAs are isolated and their sequences used as PCR probes on the YACs and the clones surrounding and including the original identifying clones. Intron-exon boundaries are then further defined through sequence analysis.

We have screened the normal breast, 8 month pregnant breast and fetal brain cDNA libraries with zoo blot-positive Eco Rl fragments from cosmid BAC and P1 clones in the region. Potential BRCA1 cDNA clones were identified among the three libraries. Clones were picked, replated, and screened again with the original probe to verify that they were positive.

Analysis of hybrid-selected cDNA. cDNA fragments obtained from direct selection were checked by Southern blot hybridization against the probe DNA to verify that they originated from the contig. Those that passed this test were sequenced in their entirety. The set of DNA sequences obtained in this way were then checked against each other to find independent clones that overlapped. For example, the clones 694-65, 1240-1 and 1240-33 were obtained independently and subsequently shown to derive from the same contiguous cDNA sequence which has been named EST:489:1.

Analysis of candidate clones. One or more of the candidate genes generated from above were sequenced and the information used for identification and classification of each expressed gene. The DNA sequences were compared to known genes by nucleotide sequence comparisons and by translation in all frames followed by a comparison with known amino acid sequences. This was accomplished using Genetic Data Environment (GDE) version 2.2 software and the Basic Local Alignment Search Tool (Blast) series of client/server software packages (e.g., BLASTN 1.3.13MP), for sequence comparison against both local and remote sequence databases (e.g., GenBank), running on Sun SPARC workstations. Sequences reconstructed from collections of cDNA clones identified with the cosmids and P1s have been generated. All candidate genes that represented new sequences were analyzed further to test their candidacy for the putative BRCA1 locus.

Mutation screening. To screen for mutations in the affected pedigrees, two different approaches were followed. First, genomic DNA isolated from family members known to carry the susceptibility allele of BRCA1 was used as a template for amplification of candidate gene sequences by PCR. If the

GeneDX 1009, pg. 80

PCR primers flank or overlap an intron/exon boundary, the amplified fragment will be larger than predicted from the cDNA sequence or will not be present in the amplified mixture. By a combination of such amplification experiments and sequencing of Pl, BAC or cosmid clones using the set of designed primers it is possible to establish the intron/exon structure and ultimately obtain the DNA sequences of genomic DNA from the pedigrees.

A second approach that is much more rapid if the intron/exon structure of the candidate gene is complex involves sequencing fragments amplified from pedigree lymphocyte cDNA. cDNA synthesized from lymphocyte mRNA extracted from pedigree blood was used as a substrate for PCR amplification using the set of designed primers. If the candidate gene is expressed to a significant extent in lymphocytes, such experiments usually produce amplified fragments that can be sequenced directly without knowledge of intron/exon junctions.

The products of such sequencing reactions were analyzed by gel electrophoresis to determine positions in the sequence that contain either mutations such as deletions or insertions, or base pair substitutions that cause amino acid changes or other detrimental effects.

Any sequence within the BRCA1 region that is expressed in breast is considered to be a candidate gene for BRCA1. Compelling evidence that a given candidate gene corresponds to BRCA1 comes from a demonstration that pedigree families contain defective alleles of the candidate.

EXAMPLE 8

Identification of BRCA1

Identification of BRCA1. Using several strategies, a detailed map of transcripts was developed for the 600 kb region of 17q21 between D17S1321 and D17S1324. Candidate expressed sequences were defined as DNA sequences obtained from: 1) direct screening of breast, fetal brain, or lymphocyte cDNA libraries, 2) hybrid selection of breast, lymphocyte or ovary cDNAs, or 3) random sequencing of genomic DNA and prediction of coding exons by XPOUND (Thomas and Skolnick, 1994). These expressed sequences in many cases were assembled into contigs composed of several independently identified sequences. Candidate genes may comprise more than one of these candidate expressed sequences. Sixty-five candidate expressed sequences within this region were identified by hybrid selection, by direct screening of cDNA libraries, and by random sequencing of P1 subclones. Expressed sequences were characterized by transcript size, DNA sequence, database comparison, expression

GeneDX 1009, pg. 81

pattern, genomic structure, and, most importantly, DNA sequence analysis in individuals from kindreds segregating 17q-linked breast and ovarian cancer susceptibility.

Three independent contigs of expressed sequence, 1141:1 (649 bp), 694:5 (213 bp) and 754:2 (1079 bp) were isolated and eventually shown to represent portions of BRCA1. When ESTs for these contigs were used as hybridization probes for Northern analysis, a single transcript of approximately 7.8 kb was observed in normal breast mRNA, suggesting that they encode different portions of a single gene. Screens of breast, fetal brain, thymus, testes, lymphocyte and placental cDNA libraries and PCR experiments with breast mRNA linked the 1141:1, 694:5 and 754:2 contigs. 5' RACE experiments with thymus, testes, and breast mRNA extended the contig to the putative 5' end, yielding a composite full length sequence. PCR and direct sequencing of P1s and BACs in the region were used to identify the location of introns and allowed the determination of splice donor and acceptor sites. These three expressed sequences were merged into a single transcription unit that proved in the final analysis to be BRCA1. This transcription unit is located adjacent to D17S855 in the center of the 600 kb region (Fig. 4).

Combination of sequences obtained from cDNA clones, hybrid selection sequences, and amplified PCR products allowed construction of a composite full length BRCA1 cDNA (SEQ ID NO:1). The sequence of the BRCA1 cDNA (up through the stop codon) has also been deposited with GenBank and assigned accession number U-14680. This deposited sequence is incorporated herein by reference. The cDNA clone extending farthest in the 3' direction contains a poly(A) tract preceded by a polyadenylation signal. Conceptual translation of the cDNA revealed a single long open reading frame of 208 kilodaltons (amino acid sequence: SEQ ID NO:2) with a potential initiation codon flanked by sequences resembling the Kozak consensus sequence (Kozak, 1987). Smith-Waterman (Smith and Waterman, 1981) and BLAST (Altschul *et al.*, 1990) searches identified a sequence near the amino terminus with considerable homology to zinc-finger domains (Fig. 5). This sequence contains cysteine and histidine residues present in the consensus C3HC4 zinc-finger motif and shares multiple other residues with zinc-finger proteins in the databases. The BRCA1 gene is composed of 23 coding exons arrayed over more than 100 kb of genomic DNA (Fig. 6). Northern blots using fragments of the BRCA1 cDNA as probes identified a single transcript of about 7.8 kb, present most abundantly in breast, thymus and testis, and also present in ovary (Fig. 7). Four alternatively spliced products were observed as independent cDNA clones; 3 of these were detected in breast and 2 in ovary mRNA (Fig. 6). A PCR survey from tissue cDNAs further supports the idea that there is considerable heterogeneity

GeneDX 1009, pg. 82

near the 5' end of transcripts from this gene; the molecular basis for the heterogeneity involves differential choice of the first splice donor site, and the changes detected all alter the transcript in the region 5' of the identified start codon. We have detected six potential alternate splice donors in this 5' untranslated region, with the longest deletion being 1,155 bp. The predominant form of the BRCA1 protein in breast and ovary lacks exon 4. The nucleotide sequence for BRCA1 exon 4 is shown in SEQ ID NO:11, with the predicted amino acid sequence shown in SEQ ID NO:12.

Additional 5' sequence of BRCA1 genomic DNA is set forth in SEQ ID NO:13. The G at position 1 represents the potential start site in testis. The A in position 140 represents the potential start site in somatic tissue. There are six alternative splice forms of this 5' sequence as shown in Figure 8. The G at position 356 represents the canonical first splice donor site. The G at position 444 represents the first splice donor site in two clones (testis 1 and testis 2). The G at position 889 represents the first splice donor site in thymus 3. A fourth splice donor site is the G at position 1230. The T at position 1513 represents the splice acceptor site for all of the above splice donors. A fifth alternate splice form has a first splice donor site at position 349 with a first acceptor site at position 591 and a second splice donor site at position 889 and a second acceptor site at position 1513. A sixth alternate form is unspliced in this 5' region. The A at position 1532 is the canonical start site, which appears at position 120 of SEQ ID NO:1. Partial genomic DNA sequences determined for BRCA1 are set forth in Figures 10A-10H and SEQ ID Numbers:14-34. The lower case letters (in figures 10A-10H) denote intron sequence while the upper case letters denote exon sequence. Indefinite intervals within introns are designated with vvvvvvvvvvvvv in Figures 10A-10H. The intron/exon junctions are shown in Table 9. The CAG found at the 5' end of exons 8 and 14 is found in some cDNAs but not in others. Known polymorphic sites are shown in Figures 10A-10H in boldface type and are underlined.

GeneDX 1009, pg. 83

TABLE 9

Exon No.	Base Position* 5'	3'	Length	Intron Borders 5'	3'
e1	1	100	100	GATAAATTAAAACTGCGACTGCGCGGCGTG[35*]	GTAGTAGAGTCCCGGGAAAGGGACAGGGGG[36]
e2	101	199	99	ATATATATATGTTTTTCTAATGTGTTAAAG[37]	GTAAGTCAGCACAAGAGTGTATTAATTTGG[38]
e3	200	253	54	TTTCTTTTTCTCCCCCCCCTACCCTGCTAG[39]	GTAAGTTTGAATGTGTTATGTGGCTCCATT[40]
e4	***	***	111	AGCTACTTTTTTTTTTTTTTTTTTGAGACAG[41]	GTAAGTGCACACCACCATATCCAGCTAAAT[42]
e5	254	331	78	AATTGTTCTTTCTTTCTTTATAATTTATAG[43]	GTATATAATTTGGTAATGATGCTAGGTTGG[44]
e6	332	420	89	GAGTGTGTTTCTCAAACAATTTAATTTCAG[45]	GTAAGTGTTGAATATCCCAAGAATGACACT[46]
e7	421	560	140	AAACATAATGTTTTCCCTTGTATTTTACAG[47]	GTAAAACCATTTGTTTTCTTCTTCTTCTTC[48]
e8	561	666	106	TGCTTGACTGTTCTTTACCATACTGTTTAG[49]	GTAAGGGTCTCAGGTTTTTTAAGTATTTAA[50]
e9	667	712	46	TGATTTATTTTTTGGGGGGAAATTTTTTAG[51]	GTGAGTCAAAGAGAACCTTTGTCTATGAAG[52]
e10	713	789	77	TCTTATTAGGACTCTGTCTTTTCCCTATAG[53]	GTAATGGCAAAGTTTGCCAACTTAACAGGC[54]
e11	790	4215	3426	GAGTACCTTGTTATTTTTGTATATTTTCAG[55]	GTATTGGAACCAGGTTTTTTGTGTTTGCCCC[56]
e12	4216	4302	87	ACATCTGAACCTCTGTTTTTGTTATTTAAG[57]	AGGTAAAAAGCGTGTGTGTGTGTGCACATG[58]
e13	4303	4476	174	CATTTTCTTGGTACCATTTATCGTTTTTGA[59]	GTGTGTATTGTTGGCCAAACACTGATATCT[60]
e14	4477	4603	127	AGTAGATTTGTTTTCTCATTCCATTTAAAG[61]	GTAAGAAACATCAATGTAAAGATGCTGTGG[62]

* Base numbers in SEQ ID NO:1.
** Numbers in superscript refer to SEQ ID NOS.
*** e4 from SEQ ID NO:11.

GeneDX 1009, pg. 84

TABLE 9 (cont'd)

Exon No.	Base Position* 5'	3'	Length	Intron Borders 5'	3'
e15	4604	4794	191	ATGGTTTTCTCCTTCCATTTATCTTTCTAG[63**]	GTAATATTTCATCTGCTGTATTGGAACAAA[64]
e16	4795	5105	311	TGTAAATTAAACTTCTCCCATTCCTTTCAG[65]	GTGAGTGTATCCATATGTATCTCCCTAATG[66]
e17	5106	5193	88	ATGATAATGGAATATTTGATTTAATTTCAG[67]	GTATACCAAGAACCTTTACAGAATACCTTG[68]
e18	5194	5271	78	CTAATCCTTTGAGTGTTTTTCATTCTGCAG[69]	GTAAGTATAATACTATTTCTCCCCTCCTCC[70]
e19	5272	5312	41	TGTAACCTGTCTTTTCTATGATCTCTTTAG[71]	GTAAGTACTTGATGTTACAAACTAACCAGA[72]
e20	5313	5396	84	TCCTGATGGGTTGTGTTTGGTTTCTTTCAG[73]	GTAAAGCTCCCTCCCTCAAGTTGACAAAAA[74]
e21	5397	5451	55	CTGTCCCTCTCTCTTCCTCTCTTCTTCCAG[75]	GTAAGAGCCTGGGAGAACCCCAGAGTTCCA[76]
e22	5452	5525	74	AGTGATTTTACATGTAAATGTCCATTTTAG[77]	GTAAGTATTGGGTGCCCTGTCAGTGTGGGA[78]
e23	5526	5586	61	TTGAATGCTCTTTCCTTCCTGGGGATCCAG[79]	GTAAGGTGCCTCGCATGTACCTGTGCTATT[80]
e24	5587	5914	328	CTAATCTCTGCTTGTGTTCTCTGTCTCCAG[81]	

* Base numbers in SEQ ID NO:1.
** Numbers in superscript refer to SEQ ID NOS.

Low stringency blots in which genomic DNA from organisms of diverse phylogenetic background were probed with BRCA1 sequences that lack the zinc-finger region revealed strongly hybridizing fragments in human, monkey, sheep and pig, and very weak hybridization signals in rodents. This result indicates that, apart from the zinc-finger domain, BRCA1 is conserved only at a moderate level through evolution.

Germline BRCA1 mutations in 17q-linked kindreds. The most rigorous test for BRCA1 candidate genes is to search for potentially disruptive mutations in carrier individuals from kindreds that segregate 17q-linked susceptibility to breast and ovarian cancer. Such individuals must contain BRCA1 alleles that differ from the wildtype sequence. The set of DNA samples used in this analysis consisted of DNA from individuals representing 8 different BRCA1 kindreds (Table 10).

T8104

TABLE 10

KINDRED DESCRIPTIONS AND ASSOCIATED LOD SCORES

Kindred	Cases (n)			Sporadic Cases[1] (n)	LOD Score	Marker(s)
	Br	Br<50	Ov			
2082	31	20	22	7	9.49	D17S1327
2099	22	14	2*	0	2.36	D17S800/D17S855[2]
2035	10	8	1*	0	2.25	D17S1327
1901	10	7	1*	0	1.50	D17S855
1925	4	3	0	0	0.55	D17S579
1910	5	4	0	0	0.36	D17S579/D17S250[2]
1927	5	4	0	1	-0.44	D17S250
1911	8	5	0	2	-0.20	D17S250

[1] Number of women with breast cancer (diagnosed under age 50) or ovarian cancer (diagnosed at any age) who do not share the BRCA1-linked haplotype segregating in the remainder of the cases in the kindred.

[2] Multipoint LOD score calculated using both markers

* kindred contains one individual who had both breast and ovarian cancer; this individual is counted as a breast cancer case and as an ovarian cancer case.

81

The logarithm of the odds (LOD) scores in these kindreds range from 9.49 to -0.44 for a set of markers in 17q21. Four of the families have convincing LOD scores for linkage, and 4 have low positive or negative LOD scores. The latter kindreds were included because they demonstrate haplotype sharing at chromosome 17q21 for at least 3 affected members. Furthermore, all kindreds in the set display early age of breast cancer onset and 4 of the kindreds include at least one case of ovarian cancer, both hallmarks of BRCA1 kindreds. One kindred, 2082, has nearly equal incidence of breast and ovarian cancer, an unusual occurrence given the relative rarity of ovarian cancer in the population. All of the kindreds except two were ascertained in Utah. K2035 is from the midwest. K2099 is an African-American kindred from the southern USA.

In the initial screen for predisposing mutations in BRCA1, DNA from one individual who carries the predisposing haplotype in each kindred was tested. The 23 coding exons and associated splice junctions were amplified either from genomic DNA samples or from cDNA prepared from lymphocyte mRNA. When the amplified DNA sequences were compared to the wildtype sequence, 4 of the 8 kindred samples were found to contain sequence variants (Table 11).

TABLE 11

PREDISPOSING MUTATIONS

Kindred Number	Mutation	Coding Effect	Location*
2082	C→T	Gln→Stop	4056
1910	extra C	frameshift	5385
2099	T→G	Met→Arg	5443
2035	?	loss of transcript	
1901	11 bp deletion	frameshift	189

* In Sequence ID NO:1

All four sequence variants are heterozygous and each appears in only one of the kindreds. Kindred 2082 contains a nonsense mutation in exon 11 (Fig. 9A), Kindred 1910 contains a single nucleotide insertion in exon 20 (Fig. 9B), and Kindred 2099 contains a missense mutation in exon 21, resulting in a Met→Arg substitution. The frameshift and nonsense mutations are likely disruptive to the function of the BRCA1 product. The peptide encoded by the frameshift allele in Kindred 1910

would contain an altered amino acid sequence beginning 108 residues from the wildtype C-terminus. The peptide encoded by the frameshift allele in Kindred 1901 would contain an altered amino acid sequence beginning with the 24th residue from the wildtype N-terminus. The mutant allele in Kindred 2082 would encode a protein missing 551 residues from the C-terminus. The missense substitution observed in Kindred 2099 is potentially disruptive as it causes the replacement of a small hydrophobic amino acid (Met), by a large charged residue (Arg). Eleven common polymorphisms were also identified, 8 in coding sequence and 3 in introns.

The individual studied in Kindred 2035 evidently contains a regulatory mutation in BRCA1. In her cDNA, a polymorphic site (A→G at base 3667) appeared homozygous, whereas her genomic DNA revealed heterozygosity at this position (Fig. 9C). A possible explanation for this observation is that mRNA from her mutated BRCA1 allele is absent due to a mutation that affects its production or stability. This possibility was explored further by examining 5 polymorphic sites in the BRCA1 coding region, which are separated by as much as 3.5 kb in the BRCA1 transcript. In all cases where her genomic DNA appeared heterozygous for a polymorphism, cDNA appeared homozygous. In individuals from other kindreds and in non-haplotype carriers in Kindred 2035, these polymorphic sites could be observed as heterozygous in cDNA, implying that amplification from cDNA was not biased in favor of one allele. This analysis indicates that a BRCA1 mutation in Kindred 2035 either prevents transcription or causes instability or aberrant splicing of the BRCA1 transcript.

Cosegregation of BRCA1 mutations with BRCA1 haplotypes and population frequency analysis. In addition to potentially disrupting protein function, two criteria must be met for a sequence variant to qualify as a candidate predisposing mutation. The variant must: 1) be present in individuals from the kindred who carry the predisposing BRCA1 haplotype and absent in other members of the kindred, and 2) be rare in the general population.

Each mutation was tested for cosegregation with BRCA1. For the frameshift mutation in Kindred 1910, two other haplotype carriers and one non-carrier were sequenced (Fig. 9B). Only the carriers exhibited the frameshift mutation. The C to T change in Kindred 2082 created a new AvrII restriction site. Other carriers and non-carriers in the kindred were tested for the presence of the restriction site (Fig. 9A). An allele-specific oligonucleotide (ASO) was designed to detect the presence of the sequence variant in Kindred 2099. Several individuals from the kindred, some known to carry the haplotype associated with the predisposing allele, and others known not to carry the associated

haplotype, were screened by ASO for the mutation previously detected in the kindred. In each kindred, the corresponding mutant allele was detected in individuals carrying the BRCA1-associated haplotype, and was not detected in noncarriers. In the case of the potential regulatory mutation observed in the individual from Kindred 2035, cDNA and genomic DNA from carriers in the kindred were compared for heterozygosity at polymorphic sites. In every instance, the extinguished allele in the cDNA sample was shown to lie on the chromosome that carries the BRCA1 predisposing allele (Fig. 9C).

To exclude the possibility that the mutations were simply common polymorphisms in the population, ASOs for each mutation were used to screen a set of normal DNA samples. Gene frequency estimates in Caucasians were based on random samples from the Utah population. Gene frequency estimates in African-Americans were based on 39 samples provided by M. Peracek-Vance which originate from African-Americans used in her linkage studies and 20 newborn Utah African-Americans. None of the 4 potential predisposing mutations was found in the appropriate control population, indicating that they are rare in the general population. Thus, two important requirements for BRCA1 susceptibility alleles were fulfilled by the candidate predisposing mutations: 1) cosegregation of the mutant allele with disease, and 2) absence of the mutant allele in controls, indicating a low gene frequency in the general population.

Phenotypic Expression of BRCA1 Mutations. The effect of the mutations on the BRCA1 protein correlated with differences in the observed phenotypic expression in the BRCA1 kindreds. Most BRCA1 kindreds have a moderately increased ovarian cancer risk, and a smaller subset have high risks of ovarian cancer, comparable to those for breast cancer (Easton *et al.*, 1993). Three of the four kindreds in which BRCA1 mutations were detected fall into the former category, while the fourth (K2082) falls into the high ovarian cancer risk group. Since the BRCA1 nonsense mutation found in K2082 lies closer to the amino terminus than the other mutations detected, it might be expected to have a different phenotype. In fact, Kindred K2082 mutation has a high incidence of ovarian cancer, and a later mean age at diagnosis of breast cancer cases than the other kindreds (Goldgar *et al.*, 1994). This difference in age of onset could be due to an ascertainment bias in the smaller, more highly penetrant families, or it could reflect tissue-specific differences in the behavior of BRCA1 mutations. The other 3 kindreds that segregate known BRCA1 mutations have, on average, one ovarian cancer for every 10 cases of breast cancer, but have a high proportion of breast cancer cases diagnosed in their late 20's or early 30's. Kindred 1910, which has a frameshift mutation, is noteworthy because three of the four

GeneDX 1009, pg. 89

affected individuals had bilateral breast cancer, and in each case the second tumor was diagnosed within a year of the first occurrence. Kindred 2035, which segregates a potential regulatory BRCA1 mutation, might also be expected to have a dramatic phenotype. Eighty percent of breast cancer cases in this kindred occur under age 50. This figure is as high as any in the set, suggesting a BRCA1 mutant allele of high penetrance (Table 10).

Although the mutations described above clearly are deleterious, causing breast cancer in women at very young ages, each of the four kindreds with mutations includes at least one woman who carries the mutation who lived until age 80 without developing a malignancy. It will be of utmost importance in the studies that follow to identify other genetic or environmental factors that may ameliorate the effects of BRCA1 mutations.

In four of the eight putative BRCA1-linked kindreds, potential predisposing mutations were not found. Three of these four have LOD scores for BRCA1-linked markers of less than 0.55. Thus, these kindreds may not in reality segregate BRCA1 predisposing alleles. Alternatively, the mutations in these four kindreds may lie in regions of BRCA1 that, for example, affect the level of transcript and therefore have thus far escaped detection.

Role of BRCA1 in Cancer. Most tumor suppressor genes identified to date give rise to protein products that are absent, nonfunctional, or reduced in function. The majority of TP53 mutations are missense; some of these have been shown to produce abnormal p53 molecules that interfere with the function of the wildtype product (Shaulian *et al.*, 1992; Srivastava *et al.*, 1993). A similar dominant negative mechanism of action has been proposed for some adenomatous polyposis coli (APC) alleles that produce truncated molecules (Su *et al.*, 1993), and for point mutations in the Wilms' tumor gene (WT1) that alter DNA binding of the protein (Little *et al.*, 1993). The nature of the mutations observed in the BRCA1 coding sequence is consistent with production of either dominant negative proteins or nonfunctional proteins. The regulatory mutation inferred in Kindred 2035 cannot be a dominant negative; rather, this mutation likely causes reduction or complete loss of BRCA1 expression from the affected allele.

The BRCA1 protein contains a C_3HC_4 zinc-finger domain, similar to those found in numerous DNA binding proteins and implicated in zinc-dependent binding to nucleic acids. The first 180 amino acids of BRCA1 contain five more basic residues than acidic residues. In contrast, the remainder of the molecule is very acidic, with a net excess of 70 acidic residues. The excess negative charge is

particularly concentrated near the C-terminus. Thus, one possibility is that BRCA1 encodes a transcription factor with an N-terminal DNA binding domain and a C-terminal transactivational "acidic blob" domain. Interestingly, another familial tumor suppressor gene, WT1, also contains a zinc-finger motif (Haber *et al.*, 1990). Many cancer predisposing mutations in WT1 alter zinc-finger domains (Little *et al.*, 1993; Haber *et al.*, 1990; Little *et al.*, 1992). WT1 encodes a transcription factor, and alternative splicing of exons that encode parts of the zinc-finger domain alter the DNA binding properties of WT1 (Bickmore *et al.*, 1992). Some alternatively spliced forms of WT1 mRNA generate molecules that act as transcriptional repressors (Drummond *et al.*, 1994). Some BRCA1 splicing variants may alter the zinc-finger motif, raising the possibility that a regulatory mechanism similar to that which occurs in WT1 may apply to BRCA1.

EXAMPLE 9

Analysis of Tumors for BRCA1 Mutations

To focus the analysis on tumors most likely to contain BRCA1 mutations, primary breast and ovarian carcinomas were typed for LOH in the BRCA1 region. Three highly polymorphic, simple tandem repeat markers were used to assess LOH: D17S1323 and D17S855, which are intragenic to BRCA1, and D17S1327, which lies approximately 100 kb distal to BRCA1. The combined LOH frequency in informative cases (i.e., where the germline was heterozygous) was 32/72 (44%) for the breast carcinomas and 12/21 (57%) for the ovarian carcinomas, consistent with previous measurements of LOH in the region (Futreal *et al.*, 1992b; Jacobs *et al.*, 1993; Sato *et al.*, 1990; Eccles *et al.*, 1990; Cropp *et al.*, 1994). The analysis thus defined a panel of 32 breast tumors and 12 ovarian tumors of mixed race and age of onset to be examined for BRCA mutations. The complete 5,589 bp coding region and intron/exon boundary sequences of the gene were screened in this tumor set by direct sequencing alone or by a combination of single-strand conformation analysis (SSCA) and direct sequencing.

A total of six mutations (of which two are identical) was found, one in an ovarian tumor, four in breast tumors and one in a male unaffected haplotype carrier (Table 12). One mutation, Glu1541Ter, introduced a stop codon that would create a truncated protein missing 323 amino acids at the carboxy terminus. In addition, two missense mutations were identified. These are Ala1708Glu and Met1775Arg and involve substitutions of small, hydrophobic residues by charged residues. Patients

17764 and 19964 are from the same family. In patient OV24 nucleotide 2575 is deleted and in patients 17764 and 19964 nucleotides 2993-2996 are deleted.

T87CY

TABLE 12

Predisposing Mutations

Patient	Codon	Nucleotide Change	Amino Acid Change	Age of Onset	Family History
BT098	1541	GAG → TAG	Glu → Stop	39	-
OV24	819	1 bp deletion	frameshift	44	-
BT106	1708	GCG → GAG	Ala → Glu	24	+
MC44	1775	ATG → AGG	Met → Arg	42	+
17764	958	4 bp deletion	frameshift	31	+
19964	958	4 bp deletion	frameshift		+*

* Unaffected haplotype carrier, male

Several lines of evidence suggest that all five mutations represent BRCA1 susceptibility alleles:

(i) all mutations are present in the germline;

(ii) all are absent in appropriate control populations, suggesting they are not common polymorphisms;

(iii) each mutant allele is retained in the tumor, as is the case in tumors from patients belonging to kindreds that segregate BRCA1 susceptibility alleles (Smith *et al.,* 1992; Kelsell *et al.,* 1993) (if the mutations represented neutral polymorphisms, they should be retained in only 50% of the cases);

(iv) the age of onset in the four breast cancer cases with mutations varied between 24 and 42 years of age, consistent with the early age of onset of breast cancer in individuals with BRCA1 susceptibility; similarly, the ovarian cancer case was diagnosed at 44, an age that falls in the youngest 13% of all ovarian cancer cases; and finally,

(v) three of the five cases have positive family histories of breast or ovarian cancer found retrospectively in their medical records, although the tumor set was not selected with regard to this criterion.

BT106 was diagnosed at ~~age 24~~ a very early age with breast cancer. Her mother had ovarian cancer, her father had melanoma, and her paternal grandmother also had breast cancer. Patient MC44, an African-American, had bilateral breast cancer at ~~age 42~~ an early age. This patient had a sister who died of breast cancer at ~~age 34~~ a very early age, another sister who died of lymphoma, and a brother who died of lung cancer. Her mutation (Met1775Arg) had been detected previously in Kindred 2099, an African-American family that segregates a BRCA1 susceptibility allele, and was absent in African-American and Caucasian controls. Patient MC44, to our knowledge, is unrelated to Kindred 2099. The detection of a rare mutant allele, once in a BRCA1 kindred and once in the germline of an apparently unrelated early-onset breast cancer case, suggests that the Met1775Arg change may be a common predisposing mutation in African-Americans. Collectively, these observations indicate that all four BRCA1 mutations in tumors represent susceptibility alleles; no somatic mutations were detected in the samples analyzed.

The paucity of somatic BRCA1 mutations is unexpected, given the frequency of LOH on 17q, and the usual role of susceptibility genes as tumor suppressors in cancer progression. There are three possible explanations for this result: (i) some BRCA1 mutations in coding sequences were missed by our screening procedure; (ii) BRCA1 somatic mutations fall primarily outside the coding exons; and (iii) LOH events in 17q do not reflect BRCA1 somatic mutations.

If somatic BRCA1 mutations truly are rare in breast and ovary carcinomas, this would have strong implications for the biology of BRCA1. The apparent lack of somatic BRCA1 mutations implies that there may be some fundamental difference in the genesis of tumors in genetically predisposed BRCA1 carriers, compared with tumors in the general population. For example, mutations in BRCA1 may have an effect only on tumor formation at a specific stage early in breast and ovarian development. This possibility is consistent with a primary function for BRCA1 in premenopausal breast cancer. Such a model for the role of BRCA1 in breast and ovarian cancer predicts an interaction between reproductive hormones and BRCA1 function. However, no clinical or pathological differences in familial versus sporadic breast and ovary tumors, other than age of onset, have been described (Lynch *et al.*, 1990). On the other hand, the recent finding of increased TP53 mutation and microsatellite instability in breast tumors from patients with a family history of breast cancer (Glebov *et al.*, 1994) may reflect some difference in tumors that arise in genetically predisposed persons. The involvement of BRCA1 in this phenomenon can now be addressed directly. Alternatively, the lack of somatic BRCA1 mutations may result from the existence of multiple genes that function in the same pathway of tumor suppression as BRCA1, but which collectively represent a more favored target for

mutation in sporadic tumors. Since mutation of a single element in a genetic pathway is generally sufficient to disrupt the pathway, BRCA1 might mutate at a rate that is far lower than the sum of the mutational rates of the other elements.

EXAMPLE 10

Analysis of the BRCA1 Gene

The structure and function of BRCA 1 gene are determined according to the following methods.

Biological Studies. Mammalian expression vectors containing BRCA1 cDNA are constructed and transfected into appropriate breast carcinoma cells with lesions in the gene. Wild-type BRCA1 cDNA as well as altered BRCA1 cDNA are utilized. The altered BRCA1 cDNA can be obtained from altered BRCA1 alleles or produced as described below. Phenotypic reversion in cultures (e.g., cell morphology, doubling time, anchorage-independent growth) and in animals (e.g., tumorigenicity) is examined. The studies will employ both wild-type and mutant forms (Section B) of the gene.

Molecular Genetics Studies. *In vitro* mutagenesis is performed to construct deletion mutants and missense mutants (by single base-pair substitutions in individual codons and cluster charged → alanine scanning mutagenesis). The mutants are used in biological, biochemical and biophysical studies.

Mechanism Studies. The ability of BRCA1 protein to bind to known and unknown DNA sequences is examined. Its ability to transactivate promoters is analyzed by transient reporter expression systems in mammalian cells. Conventional procedures such as particle-capture and yeast two-hybrid system are used to discover and identify any functional partners. The nature and functions of the partners are characterized. These partners in turn are targets for drug discovery.

Structural Studies. Recombinant proteins are produced in *E. coli*, yeast, insect and/or mammalian cells and are used in crystallographical and NMR studies. Molecular modeling of the proteins is also employed. These studies facilitate structure-driven drug design.

EXAMPLE 11

Two Step Assay to Detect the Presence of BRCA1 in a Sample

Patient sample is processed according to the method disclosed by Antonarakis *et al.* (1985), separated through a 1% agarose gel and transferred to nylon membrane for Southern blot analysis.

Membranes are UV cross linked at 150 mJ using a GS Gene Linker (Bio-Rad). BRCA1 probe corresponding to nucleotide positions 3631-3930 of SEQ ID NO:1 is subcloned into pTZ18U. The phagemids are transformed into *E. coli* MV1190 infected with M13KO7 helper phage (Bio-Rad, Richmond, CA). Single stranded DNA is isolated according to standard procedures (see Sambrook *et al.*, 1989).

Blots are prehybridized for 15-30 min at 65°C in 7% sodium dodecyl sulfate (SDS) in 0.5 M $NaPO_4$. The methods follow those described by Nguyen *et al.*, 1992. The blots are hybridized overnight at 65°C in 7% SDS, 0.5 M $NaPO_4$ with 25-50 ng/ml single stranded probe DNA. Post-hybridization washes consist of two 30 min washes in 5% SDS, 40 mM $NaPO_4$ at 65°C, followed by two 30 min washes in 1% SDS, 40 mM $NaPO_4$ at 65°C.

Next the blots are rinsed with phosphate buffered saline (pH 6.8) for 5 min at room temperature and incubated with 0.2% casein in PBS for 30-60 min at room temperature and rinsed in PBS for 5 min. The blots are then preincubated for 5-10 minutes in a shaking water bath at 45°C with hybridization buffer consisting of 6 M urea, 0.3 M NaCl, and 5X Denhardt's solution (see Sambrook, *et al.*, 1989). The buffer is removed and replaced with 50-75 $\mu l/cm^2$ fresh hybridization buffer plus 2.5 nM of the covalently cross-linked oligonucleotide-alkaline phosphatase conjugate with the nucleotide sequence complementary to the universal primer site (UP-AP, Bio-Rad). The blots are hybridized for 20-30 min at 45°C and post hybridization washes are incubated at 45°C as two 10 min washes in 6 M urea, 1x standard saline citrate (SSC), 0.1% SDS and one 10 min wash in 1x SSC, 0.1% ~~Triton~~ TRITON® X-100, detergent. The blots are rinsed for 10 min at room temperature with 1x SSC.

Blots are incubated for 10 min at room temperature with shaking in the substrate buffer consisting of 0.1 M diethanolamine, 1 mM $MgCl_2$, 0.02% sodium azide, pH 10.0. Individual blots are placed in heat sealable bags with substrate buffer and 0.2 mM AMPPD (3-(2'-spiroadamantane)-4-methoxy-4-(3'-phosphoryloxy)phenyl-1,2-dioxetane, disodium salt, Bio-Rad). After a 20 min incubation at room temperature with shaking, the excess AMPPD solution is removed. The blot is exposed to X-ray film overnight. Positive bands indicate the presence of BRCA1.

EXAMPLE 12

Generation of Polyclonal Antibody against BRCA1

Segments of BRCA1 coding sequence were expressed as fusion protein in *E. coli*. The overexpressed protein was purified by gel elution and used to immunize rabbits and mice using a procedure similar to the one described by Harlow and Lane, 1988. This procedure has been shown to generate Abs against various other proteins (for example, see Kraemer *et al.*, 1993).

Briefly, a stretch of BRCA1 coding sequence was cloned as a fusion protein in plasmid PET5A (Novagen, Inc., Madison, WI). The BRCA1 incorporated sequence includes the amino acids corresponding to #1361-1554 of SEQ ID NO:2. After induction with IPTG, the overexpression of a fusion protein with the expected molecular weight was verified by SDS/PAGE. Fusion protein was purified from the gel by electroelution. The identification of the protein as the BRCA1 fusion product was verified by protein sequencing at the N-terminus. Next, the purified protein was used as immunogen in rabbits. Rabbits were immunized with 100 μg of the protein in complete Freund's adjuvant and boosted twice in 3 week intervals, first with 100 μg of immunogen in incomplete Freund's adjuvant followed by 100 μg of immunogen in PBS. Antibody containing serum is collected two weeks thereafter.

This procedure is repeated to generate antibodies against the mutant forms of the BRCA1 gene. These antibodies, in conjunction with antibodies to wild type BRCA1, are used to detect the presence and the relative level of the mutant forms in various tissues and biological fluids.

EXAMPLE 13

Generation of Monoclonal Antibodies Specific for BRCA1

Monoclonal antibodies are generated according to the following protocol. Mice are immunized with immunogen comprising intact BRCA1 or BRCA1 peptides (wild type or mutant) conjugated to keyhole limpet hemocyanin using glutaraldehyde or EDC as is well known.

The immunogen is mixed with an adjuvant. Each mouse receives four injections of 10 to 100 μg of immunogen and after the fourth injection blood samples are taken from the mice to determine if the serum contains antibody to the immunogen. Serum titer is determined by ELISA or RIA. Mice with sera indicating the presence of antibody to the immunogen are selected for hybridoma production.

GeneDX 1009, pg. 96

Spleens are removed from immune mice and a single cell suspension is prepared (see Harlow and Lane, 1988). Cell fusions are performed essentially as described by Kohler and Milstein, 1975. Briefly, P3.65.3 myeloma cells (American Type Culture Collection, Rockville, MD) are fused with immune spleen cells using polyethylene glycol as described by Harlow and Lane, 1988. Cells are plated at a density of $2x10^5$ cells/well in 96 well tissue culture plates. Individual wells are examined for growth and the supernatants of wells with growth are tested for the presence of BRCA1 specific antibodies by ELISA or RIA using wild type or mutant BRCA1 target protein. Cells in positive wells are expanded and subcloned to establish and confirm monoclonality.

Clones with the desired specificities are expanded and grown as ascites in mice or in a hollow fiber system to produce sufficient quantities of antibody for characterization and assay development.

EXAMPLE 14

Sandwich Assay for BRCA1

Monoclonal antibody is attached to a solid surface such as a plate, tube, bead, or particle. Preferably, the antibody is attached to the well surface of a 96-well ELISA plate. 100 µl sample (e.g., serum, urine, tissue cytosol) containing the BRCA1 peptide/protein (wild-type or mutant) is added to the solid phase antibody. The sample is incubated for 2 hrs at room temperature. Next the sample fluid is decanted, and the solid phase is washed with buffer to remove unbound material. 100 µl of a second monoclonal antibody (to a different determinant on the BRCA1 peptide/protein) is added to the solid phase. This antibody is labeled with a detector molecule (e.g., ^{125}I, enzyme, fluorophore, or a chromophore) and the solid phase with the second antibody is incubated for two hrs at room temperature. The second antibody is decanted and the solid phase is washed with buffer to remove unbound material.

The amount of bound label, which is proportional to the amount of BRCA1 peptide/protein present in the sample, is quantitated. Separate assays are performed using monoclonal antibodies which are specific for the wild-type BRCA1 as well as monoclonal antibodies specific for each of the mutations identified in BRCA1.

GeneDX 1009, pg. 97

Industrial Utility

As previously described above, the present invention provides materials and methods for use in testing BRCA1 alleles of an individual and an interpretation of the normal or predisposing nature of the alleles. Individuals at higher than normal risk might modify their lifestyles appropriately. In the case of BRCA1, the most significant non-genetic risk factor is the protective effect of an early, full term pregnancy. Therefore, women at risk could consider early childbearing or a therapy designed to simulate the hormonal effects of an early full-term pregnancy. Women at high risk would also strive for early detection and would be more highly motivated to learn and practice breast self examination. Such women would also be highly motivated to have regular mammograms, perhaps starting at an earlier age than the general population. Ovarian screening could also be undertaken at greater frequency. Diagnostic methods based on sequence analysis of the BRCA1 locus could also be applied to tumor detection and classification. Sequence analysis could be used to diagnose precursor lesions. With the evolution of the method and the accumulation of information about BRCA1 and other causative loci, it could become possible to separate cancers into benign and malignant.

Women with breast cancers may follow different surgical procedures if they are predisposed, and therefore likely to have additional cancers, than if they are not predisposed. Other therapies may be developed, using either peptides or small molecules (rational drug design). Peptides could be the missing gene product itself or a portion of the missing gene product. Alternatively, the therapeutic agent could be another molecule that mimics the deleterious gene's function, either a peptide or a nonpeptidic molecule that seeks to counteract the deleterious effect of the inherited locus. The therapy could also be gene based, through introduction of a normal BRCA1 allele into individuals to make a protein which will counteract the effect of the deleterious allele. These gene therapies may take many forms and may be directed either toward preventing the tumor from forming, curing a cancer once it has occurred, or stopping a cancer from metastasizing.

It will be appreciated that the methods and compositions of the instant invention can be incorporated in the form of a variety of embodiments, only a few of which are disclosed herein. It will be apparent to the artisan that other embodiments exist and do not depart from the spirit of the invention. Thus, the described embodiments are illustrative and should not be construed as restrictive.

GeneDX 1009, pg. 98

LIST OF REFERENCES

Altschul, S.F. *et al.* (1990). *J. Mol. Biol.* **215**: 195-197.

American Cancer Society, Cancer Facts & Figures - 1992. (American Cancer Society, Atlanta, GA).

Anand, R. (1992). *Techniques for the Analysis of Complex Genomes*, (Academic Press).

Anderson, *et al.* (1980). *Proc. Natl. Acad. Sci. USA* 77:5399-5403.

Anderson, D.E. (1972). *J. Natl. Cancer Inst.* **48**:1029-1034.

Anderson, J.A., *et al.* (1992). *J. Otolaryngology* **21**:321.

Antonarakis, S.E., *et al.* (1985). *New Eng. J. Med.* **313**:842-848.

Ausubel, F.M., *et al.* (1992). *Current Protocols in Molecular Biology*, (J. Wiley and Sons, N.Y.)

Beaucage & Carruthers (1981). *Tetra. Letts.* **22**:1859-1862.

Berkner (1992). *Curr. Top. Microbiol. Immunol.* **158**:39-61.

Berkner, *et al.* (1988). *BioTechniques* **6**:616-629.

Bickmore, W.A., *et al.* (1992). *Science* **257**:235-7.

Bishop, D.T., *et al.* (1988). *Genet. Epidemiol.* **5**:151-169.

Bishop, D.T. and Gardner, E.J. (1980). In: Banbury Report 4: Cancer Incidence in Defined Populations (J. Cairns, J.L. Lyon, M. Skolnick, eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 309-408.

Botstein, *et al.* (1980). *Am. J. Hum. Genet.* **32**:314-331.

Bowcock, A.M., *et al.* (1993). *Am. J. Hum. Genet.* **52**:718.

Brandyopadhyay and Temin (1984). *Mol. Cell. Biol.* **4**:749-754.

Breakfield and Geller (1987). *Mol. Neurobiol.* **1**:337-371.

Brinster, *et al.* (1981). *Cell* **27**:223-231.

Buchschacher and Panganiban (1992). *J. Virol.* **66**:2731-2739.

Buckler, *et al.* (1991). *Proc. Natl. Acad. Sci. USA* **88**:4005-4009.

GeneDX 1009, pg. 99

Cannon-Albright, L., *et al.* (1994). *Cancer Research* **54**:2378-2385.

Capecchi, M.R. (1989). *Science* **244**:1288.

Cariello (1988). *Human Genetics* **42**:726.

Claus, E., *et al.* (1991). *Am. J. Hum. Genet.* **48**:232-242.

Conner, B.J., *et al.* (1983). *Proc. Natl. Acad. Sci. USA* **80**:278-282.

Constantini and Lacy (1981). *Nature* **294**:92-94.

Cotten, *et al.* (1990). *Proc. Natl. Acad. Sci. USA* **87**:4033-4037.

Cotton, *et al.* (1988). *Proc. Natl. Acad. Sci. USA* **85**:4397.

Cropp, C.S., *et al.* (1994). *Cancer Res.* **54**:2548-2551.

Culver, *et al.* (1992). *Science* **256**:1550-1552.

Curiel, *et al.* (1991a). *Proc. Natl. Acad. Sci. USA* **88**:8850-8854.

Curiel, *et al.* (1991b). *Hum. Gene Ther.* **3**:147-154.

Deutscher, M. (1990). *Meth. Enzymology* **182** (Academic Press, San Diego, Cal.).

Donehower, L.A., *et al.* (1992). *Nature* **356**:215.

Drummond, I. A., *et al.* (1994). *Mol. Cell Biol.* **14**:3800-9.

Easton, D., *et al.* (1993). *Am. J. Hum. Genet.* **52**:678-701.

Eccles, D.M., *et al.* (1990). *Oncogene* **5**:1599-1601.

Enhancers and Eurkaryotic Gene Expression, Cold Spring Harbor Press, Cold Spring Harbor, New York (1983).

Erickson, J. *et al.*, (1990). *Science* **249**:527-533.

Fain, P.R. (1992). *Cytogen. Cell Genet.* **60**:178.

Felgner, *et al.* (1987). *Proc. Natl. Acad. Sci. USA* **84**:7413-7417.

Fiers, *et al.* (1978). *Nature* **273**:113.

Fink, *et al.* (1992). *Hum. Gene Ther.* **3**:11-19.

Finkelstein, J., *et al.* (1990). *Genomics* 7:167-172.

Freese, *et al.* (1990). *Biochem. Pharmacol.* **40**:2189-2199.

Friedman, T. (1991). In *Therapy for Genetic Diseases*, T. Friedman, ed., Oxford University Press, pp. 105-121.

Futreal (1993). Ph.D. Thesis, University of North Carolina, Chapel Hill.

Futreal, A., *et al.* (1992a). *Hum. Molec. Genet.* **1**:66.

Futreal, P.A., *et al.* (1992b). *Cancer Res.* **52**:2624-2627.

Glebov, O.K., *et al.* (1994). *Cancer Res.* **54**:3703-3709.

Glover, D. (1985). *DNA Cloning*, **I** and **II** (Oxford Press).

Go, R.C.P., *et al.* (1983). *J. Natl. Cancer Inst.* **71**:455-461.

Goding (1986). *Monoclonal Antibodies: Principles and Practice*, 2d ed. (Academic Press, N.Y.).

Godowski, *et al.* (1988). *Science* **241**:812-816.

Goldgar, D.E., *et al.* (1994). *J. Natl. Can. Inst.* **86**:3:200-209.

Gordon, *et al.* (1980). *Proc. Natl. Acad. Sci. USA* 77:7380-7384.

Gorziglia and Kapikian (1992). *J. Virol.* **66**:4407-4412.

Graham and van der Eb (1973). *Virology* **52**:456-467.

Grompe, M., (1993). *Nature Genetics* **5**:111-117.

Grompe, M., *et al.*, (1989). *Proc. Natl. Acad. Sci. USA* **86**:5855-5892.

Guthrie, G. & Fink, G.R. (1991). *Guide to Yeast Genetics and Molecular Biology* (Academic Press).

Haber, D. A., *et al.* (1990). *Cell* **61**:1257-69.

Hall, J.M., *et al.* (1990). *Science* **250**:1684-1689.

Hall, J.M., *et al.* (1992). *Am. J. Hum. Genet.* **50**:1235-1241.

Harlow & Lane (1988). Antibodies: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Hasty, P., K., *et al.* (1991). *Nature* **350**:243.

Helseth, *et al.* (1990). *J. Virol.* **64**:2416-2420.

Hodgson, J. (1991). *Bio/Technology* **9**:19-21.

Huse, *et al.* (1989). *Science* **246**:1275-1281.

Innis *et al.* (1990). *PCR Protocols: A Guide to Methods and Applications* (Academic Press, San Diego, Cal.).

Jablonski, E., *et al.* (1986). *Nuc. Acids Res.* **14**:6115-6128.

Jacobs, I.J., *et al.* (1993). *Cancer Res.* **53**:1218-1221.

Jakoby, W.B. and Pastan, I.H. (eds.) (1979). *Cell Culture. Methods in Enzymology*, volume 58 (Academic Press, Inc., Harcourt Brace Jovanovich (New York)).

Jeffreys, *et al.* (1985). *Nature* **314**:67-73.

Johnson, *et al.* (1992). *J. Virol.* **66**:2952-2965.

Kamb, A. *et al.* (1994). *Science* **264**:436-440.

Kandpal, *et al.* (1990). *Nucl. Acids Res.* **18**:1789-1795.

Kaneda, *et al.* (1989). *J. Biol. Chem.* **264**:12126-12129.

Kanehisa (1984). *Nucl. Acids Res.* **12**:203-213.

Kelsell, D.P., *et al.* (1993). *Human Mol. Genet.* **2**:1823-1828.

Kinszler, K.W., *et al.* (1991). *Science* **251**:1366-1370.

Knudson, A.G. (1993). *Nature Genet.* **5**:103.

Kohler, G. and Milstein, C. (1975). *Nature* **256**:495-497.

Kozak, M. (1987). *Nucleic Acids Res.* **15**:8125-8148.

Kraemer, F.B. *et al.* (1993). *J. Lipid Res.* **34**:663-672.

Kubo, T., *et al.* (1988). *FEBS Letts.* **241**:119.

GeneDX 1009, pg. 102

Landegren, *et al.* (1988). *Science* **242**:229.

Lim, *et al.* (1992). *Circulation* **83**:2007-2011.

Lindsay, S., *et al.* (1987). *Nature* **327**:336-368.

Litt, *et al.* (1989). *Am. J. Hum. Genet.* **44**:397-401.

Little, M.H., *et al.* (1992). *Proc. Natl. Acad. Sci. USA* **89**:4791.

Little, M.H., *et al.* (1993). *Hum. Mol. Genet.* **2**:259.

Lovett, *et al.* (1991). *Proc. Natl. Acad. Sci. USA* **88**:9628-9632.

Lynch, H.T., *et al.* (1990). *Gynecol. Oncol.* **36**:48-55.

Madzak, *et al.* (1992). *J. Gen. Virol.* **73**:1533-1536.

Malkin, D., *et al.* (1990). *Science* **250**:1233-1238.

Maniatis. T., *et al.* (1982). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Mann and Baltimore (1985). *J. Virol.* **54**:401-407.

Margaritte, *et al.* (1992). *Am. J. Hum. Genet.* **50**:1231-1234.

Margolskee (1992). *Curr. Top. Microbiol. Immunol.* **158**:67-90.

Martin, R., *et al.* (1990). *BioTechniques* **9**:762-768.

Matteucci, M.D. and Caruthers, M.H. (1981). *J. Am. Chem. Soc.* **103**:3185.

Matthews & Kricka (1988). *Anal. Biochem.* **169**:1.

Merrifield (1963). *J. Am. Chem. Soc.* **85**:2149-2156.

Mettlin, C., *et al.* (1990). *American Journal of Epidemiology* **131**:973-983.

Metzger, *et al.* (1988). *Nature* **334**:31-36.

Miller (1992). *Curr. Top. Microbiol. Immunol.* **158**:1-24.

Miller, *et al.* (1985). *Mol. Cell. Biol.* **5**:431-437.

98

Miller, *et al.* (1988). *J. Virol.* **62**:4337-4345.

Mittlin (1989). *Clinical Chem.* **35**:1819.

Modrich, P. (1991). *Ann. Rev. Genet.* **25**:229-253.

Mombaerts, P., *et al.* (1992). *Cell* **68**:869.

Monaco, *et al.* (1986). *Nature* **323**:646.

Moss (1992). *Curr. Top. Microbiol. Immunol.* **158**:25-38.

Muzyczka (1992). *Curr. Top. Microbiol. Immunol.* **158**:97-123.

Nabel (1992). *Hum. Gene Ther.* **3**:399-410.

Nabel, *et al.* (1990). *Science* **249**:1285-1288.

Nakamura, *et al.* (1987). *Science* **235**:1616-1622.

Narod, S.A., *et al.* (1991). *The Lancet* **338**:82-83.

Newman, B., *et al.* (1988). *Proc. Natl. Acad. Sci. USA* **85**:3044-3048.

Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C., and Markham, A.F. (1989). *Nucl. Acids Res.* **17**:2503-2516.

Nguyen, Q., *et al.* (1992). *BioTechniques* **13**:116-123.

Novack, *et al.* (1986). *Proc. Natl. Acad. Sci. USA* **83**:586.

Oh, J. (1985). *Analysis of Human Genetic Linkage*, Johns Hopkins University Press, Baltimore, Md, pp. 1-216.

Ohi, *et al.* (1990). *Gene* **89**:279-282.

Oliphant, A., *et al.* (1991). *Nucleic Acid Res.* **19**:4794.

Oliphant, A., *et al.* (1991). *Nucleic Acid Res.* **19**:4795.

Orita, *et al.* (1989). *Proc. Natl. Acad. Sci. USA* **86**:2776-2770.

Page, *et al.* (1990). *J. Virol.* **64**:5370-5276.

Pellicer, *et al.* (1980). *Science* **209**:1414-1422.

GeneDX 1009, pg. 104

Petropoulos, *et al.* (1992). *J. Virol.* **66**:3391-3397.

Philpott, K.L., *et al.* (1992). *Science* **256**:1448.

Pierce, *et al.* (1992). *Proc. Natl. Acad. Sci. USA* **89**:2056-2060.

Quantin, *et al.* (1992). *Proc. Natl. Acad. Sci. USA* **89**:2581-2584.

Rano & Kidd (1989). *Nucl. Acids Res.* **17**:8392.

Rigby, P.W.J., *et al.* (1977). *J. Mol. Biol.* **113**:237-251.

Rosenfeld, *et al.* (1992). *Cell* **68**:143-155.

Sambrook, J., *et al.* (1989). *Molecular Cloning: A Laboratory Manual*, 2nd Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Sato, T., *et al.* (1990). *Cancer Res.* **50**:7184-7189.

Scharf (1986). *Science* **233**:1076.

Scopes, R. (1982). *Protein Purification: Principles and Practice*, (Springer-Verlag, N.Y.).

Shaulian, E., *et al.* (1992). *Mol. Cell Biol.* **12**:5581-92.

Sheffield, V.C., *et al.* (1989). *Proc. Natl. Acad. Sci. USA* **86**:232-236.

Sheffield, V.C., *et al.* (1991). *Am. J. Hum. Genet.* **49**:699-706.

Shenk, *et al.* (1975). *Proc. Natl. Acad. Sci. USA* **72**:989.

Shimada, *et al.* (1991). *J. Clin. Invest.* **88**:1043-1047.

Shinkai, Y., *et al.* (1992). *Cell* **68**:855.

Shizuya, H., *et al.* (1992). *Proc. Natl. Acad. Sci. USA* **89**:8794-8797.

Simard, J., *et al.* (1993). *Human Mol. Genet.* **2**:1193-1199.

Skolnick, M.H. and Wallace, B.R. (1988). *Genomics* **2**:273-279.

Skolnick, M.H., *et al.* (1990). *Science* **250**:1715-1720.

Smith, S.A., *et al.* (1992). *Nature Genetics* **2**:128-131.

Smith, T.F. and Waterman, M.S. (1981). *J. Mol. Biol.* **147**:195-197.

102

Snouwaert, J.N., *et al.* (1992). *Science* **257**:1083.

Sorge, *et al.* (1984). *Mol. Cell. Biol.* **4**:1730-1737.

Srivastava, S., *et al.* (1993). *Cancer Res.* **53**:4452-5.

Sternberg (1990). *Proc. Natl. Acad. Sci. USA* **87**:103-107.

Sternberg, *et al.* (1990). *The New Biologist* **2**:151-162.

Stewart, *et al.* (1992). *Hum. Gene Ther.* **3**:267-275.

Stratford-Perricaudet, *et al.* (1990). *Hum. Gene Ther.* **1**:241-256.

Swift, M., *et al.* (1991). *N. Engl. J. Med.* **325**:1831-1836.

Swift, M., *et al.* (1976). *Cancer Res.* **36**:209-215.

Su, L. K., *et al.* (1993). *Cancer Res.* **53**:2728-31.

Thomas, A. and Skolnick, M.H. (1994). *IMA Journal of Mathematics Applied in Medicine and Biology* (in press).

Tonolio, D., *et al.* (1990). Cold Spring Harbor Conference.

Valancius, V. & Smithies, O. (1991). *Mol. Cell Biol.* **11**:1402.

van Dilla, *et al.* (1986). *Biotechnology* **4**:537-552.

Wagner, *et al.* (1990). *Proc. Natl. Acad. Sci. USA* **87**:3410-3414.

Wagner, *et al.* (1991). *Proc. Natl. Acad. Sci. USA* **88**:4255-4259.

Wang and Huang (1989). *Biochemistry* **28**:9508-9514.

Wartell, R.M., *et al.* (1990). *Nucl. Acids Res.* **18**:2699-2705.

Weber, J.L. (1990). *Genomics* 7:524-530.

Weber and May (1989). *Am. J. Hum. Genet.* **44**:388-396.

Weber, J.L., *et al.* (1990). *Nucleic Acid Res.* **18**:4640.

Wells, J.A. (1991). *Methods in Enzymol.* **202**:390-411.

Wetmur & Davidson (1968). *J. Mol. Biol.* **31**:349-370.

White, M.B., *et al.*, (1992). *Genomics* **12**:301-306.

White and Lalouel (1988). *Ann. Rev. Genet.* **22**:259-279.

Wilkinson, *et al.* (1992). *Nucleic Acids Res.* **20**:2233-2239.

Willams and Anderson (1984). *Genet. Epidemiol.* **1**:7-20.

Wolff, *et al.* (1990). *Science* **247**:1465-1468.

Wolff, *et al.* (1991). *BioTechniques* **11**:474-485.

Wooster, R., *et al.* (1994). *Science* **265**:2088.

Wu, *et al.* (1989a). *Genomics* **4**:560-569.

Wu, *et al.* (1989b). *J. Biol. Chem.* **264**:16985-16987.

Wu, *et al.* (1991). *J. Biol. Chem.* **266**:14338-14342.

Zenke, *et al.* (1990). *Proc. Natl. Acad. Sci. USA* **87**:3655-3659.

List of Patents and Patent Applications:

U.S. Patent No. 3,817,837

U.S. Patent No. 3,850,752

U.S. Patent No. 3,939,350

U.S. Patent No. 3,996,345

U.S. Patent No. 4,275,149

U.S. Patent No. 4,277,437

U.S. Patent No. 4,366,241

U.S. Patent No. 4,376,110

U.S. Patent No. 4,486,530

U.S. Patent No. 4,683,195

GeneDX 1009, pg. 107

U.S. Patent No. 4,683,202

U.S. Patent No. 4,816,567

U.S. Patent No. 4,868,105

U.S. Patent No. 5,252,479

EPO Publication No. 225,807

European Patent Application Publication No. 0332435

Geysen, H., PCT published application WO 84/03564, published 13 September 1984

Hitzeman *et al.*, EP 73,675A

PCT published application WO 93/07282

GeneDX 1009, pg. 108

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Skolnick, Mark H.
Goldgar, David E.
Miki, Yoshio
Swenson, Jeff
Kamb, Alexander
Harshman, Keith D.
Shattuck-Eidens, Donna M.
Tavtigian, Sean V.
Wiseman, Roger W.
Futreal, P. Andrew

(ii) TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer Susceptibility Gene

(iii) NUMBER OF SEQUENCES: 85

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
(B) STREET: 1201 New York Avenue, N.W., Suite 1000
(C) CITY: Washington
(D) STATE: DC
(E) COUNTRY: USA
(F) ZIP: 20005

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/488,011
(B) FILING DATE: 07-JUN-1995
(C) CLASSIFICATION: 435

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/409,305
(B) FILING DATE: 24-MAR-1995

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/348,824
(B) FILING DATE: 29-NOV-1994

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/308,104
(B) FILING DATE: 16-SEP-1994

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/300,266
(B) FILING DATE: 02-SEP-1994

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/289,221
(B) FILING DATE: 12-AUG-1994

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Ihnen, Jeffrey L.
(B) REGISTRATION NUMBER: 28,957
(C) REFERENCE/DOCKET NUMBER: 24884-109347-09

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 202-962-4810
(B) TELEFAX: 202-962-8300

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5914 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

E6 cont'd

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 120..5708

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
AGCTCGCTGA GACTTCCTGG ACCCCGCACC AGGCTGTGGG GTTTCTCAGA TAACTGGGCC      60

CCTGCGCTCA GGAGGCCTTC ACCCTCTGCT CTGGGTAAAG TTCATTGGAA CAGAAAGAA      119

ATG GAT TTA TCT GCT CTT CGC GTT GAA GAA GTA CAA AAT GTC ATT AAT        167
Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
  1               5                  10                  15

GCT ATG CAG AAA ATC TTA GAG TGT CCC ATC TGT CTG GAG TTG ATC AAG       215
Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
             20                  25                  30
```


GeneDX 1009, pg. 110

```
GAA CCT GTC TCC ACA AAG TGT GAC CAC ATA TTT TGC AAA TTT TGC ATG        263
Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
          35                  40                  45

CTG AAA CTT CTC AAC CAG AAG AAA GGG CCT TCA CAG TGT CCT TTA TGT        311
Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
     50                  55                  60

AAG AAT GAT ATA ACC AAA AGG AGC CTA CAA GAA AGT ACG AGA TTT AGT        359
Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
 65                  70                  75                  80

CAA CTT GTT GAA GAG CTA TTG AAA ATC ATT TGT GCT TTT CAG CTT GAC        407
Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
                 85                  90                  95

ACA GGT TTG GAG TAT GCA AAC AGC TAT AAT TTT GCA AAA AAG GAA AAT        455
Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
            100                 105                 110

AAC TCT CCT GAA CAT CTA AAA GAT GAA GTT TCT ATC ATC CAA AGT ATG        503
Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
        115                 120                 125

GGC TAC AGA AAC CGT GCC AAA AGA CTT CTA CAG AGT GAA CCC GAA AAT        551
Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
    130                 135                 140

CCT TCC TTG CAG GAA ACC AGT CTC AGT GTC CAA CTC TCT AAC CTT GGA        599
Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
145                 150                 155                 160

ACT GTG AGA ACT CTG AGG ACA AAG CAG CGG ATA CAA CCT CAA AAG ACG        647
Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
                165                 170                 175

TCT GTC TAC ATT GAA TTG GGA TCT GAT TCT TCT GAA GAT ACC GTT AAT        695
Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
            180                 185                 190

AAG GCA ACT TAT TGC AGT GTG GGA GAT CAA GAA TTG TTA CAA ATC ACC        743
Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
        195                 200                 205

CCT CAA GGA ACC AGG GAT GAA ATC AGT TTG GAT TCT GCA AAA AAG GCT        791
Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
    210                 215                 220

GCT TGT GAA TTT TCT GAG ACG GAT GTA ACA AAT ACT GAA CAT CAT CAA        839
Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
225                 230                 235                 240

CCC AGT AAT AAT GAT TTG AAC ACC ACT GAG AAG CGT GCA GCT GAG AGG        887
Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg
                245                 250                 255
```

E6 cont'd

106

```
CAT CCA GAA AAG TAT CAG GGT AGT TCT GTT TCA AAC TTG CAT GTG GAG       935
His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu
            260                 265                 270

CCA TGT GGC ACA AAT ACT CAT GCC AGC TCA TTA CAG CAT GAG AAC AGC       983
Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser
        275                 280                 285

AGT TTA TTA CTC ACT AAA GAC AGA ATG AAT GTA GAA AAG GCT GAA TTC      1031
Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe
    290                 295                 300

TGT AAT AAA AGC AAA CAG CCT GGC TTA GCA AGG AGC CAA CAT AAC AGA      1079
Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg
305                 310                 315                 320

TGG GCT GGA AGT AAG GAA ACA TGT AAT GAT AGG CGG ACT CCC AGC ACA      1127
Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr
                325                 330                 335

GAA AAA AAG GTA GAT CTG AAT GCT GAT CCC CTG TGT GAG AGA AAA GAA      1175
Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu
            340                 345                 350

TGG AAT AAG CAG AAA CTG CCA TGC TCA GAG AAT CCT AGA GAT ACT GAA      1223
Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu
        355                 360                 365

GAT GTT CCT TGG ATA ACA CTA AAT AGC AGC ATT CAG AAA GTT AAT GAG      1271
Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu
    370                 375                 380

TGG TTT TCC AGA AGT GAT GAA CTG TTA GGT TCT GAT GAC TCA CAT GAT      1319
Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp
385                 390                 395                 400

GGG GAG TCT GAA TCA AAT GCC AAA GTA GCT GAT GTA TTG GAC GTT CTA      1367
Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu
                405                 410                 415

AAT GAG GTA GAT GAA TAT TCT GGT TCT TCA GAG AAA ATA GAC TTA CTG      1415
Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu
            420                 425                 430

GCC AGT GAT CCT CAT GAG GCT TTA ATA TGT AAA AGT GAA AGA GTT CAC      1463
Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Glu Arg Val His
        435                 440                 445

TCC AAA TCA GTA GAG AGT AAT ATT GAA GAC AAA ATA TTT GGG AAA ACC      1511
Ser Lys Ser Val Glu Ser Asn Ile Glu Asp Lys Ile Phe Gly Lys Thr
    450                 455                 460

TAT CGG AAG AAG GCA AGC CTC CCC AAC TTA AGC CAT GTA ACT GAA AAT      1559
Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn
465                 470                 475                 480
```

E6 cont'd

```
CTA ATT ATA GGA GCA TTT GTT ACT GAG CCA CAG ATA ATA CAA GAG CGT      1607
Leu Ile Ile Gly Ala Phe Val Thr Glu Pro Gln Ile Ile Gln Glu Arg
                485                 490                 495

CCC CTC ACA AAT AAA TTA AAG CGT AAA AGG AGA CCT ACA TCA GGC CTT      1655
Pro Leu Thr Asn Lys Leu Lys Arg Lys Arg Arg Pro Thr Ser Gly Leu
            500                 505                 510

CAT CCT GAG GAT TTT ATC AAG AAA GCA GAT TTG GCA GTT CAA AAG ACT      1703
His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr
        515                 520                 525

CCT GAA ATG ATA AAT CAG GGA ACT AAC CAA ACG GAG CAG AAT GGT CAA      1751
Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln
    530                 535                 540

GTG ATG AAT ATT ACT AAT AGT GGT CAT GAG AAT AAA ACA AAA GGT GAT      1799
Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp
545                 550                 555                 560

TCT ATT CAG AAT GAG AAA AAT CCT AAC CCA ATA GAA TCA CTC GAA AAA      1847
Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys
                565                 570                 575

GAA TCT GCT TTC AAA ACG AAA GCT GAA CCT ATA AGC AGC AGT ATA AGC      1895
Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser
            580                 585                 590

AAT ATG GAA CTC GAA TTA AAT ATC CAC AAT TCA AAA GCA CCT AAA AAG      1943
Asn Met Glu Leu Glu Leu Asn Ile His Asn Ser Lys Ala Pro Lys Lys
        595                 600                 605

AAT AGG CTG AGG AGG AAG TCT TCT ACC AGG CAT ATT CAT GCG CTT GAA      1991
Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Ala Leu Glu
    610                 615                 620

CTA GTA GTC AGT AGA AAT CTA AGC CCA CCT AAT TGT ACT GAA TTG CAA      2039
Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln
625                 630                 635                 640

ATT GAT AGT TGT TCT AGC AGT GAA GAG ATA AAG AAA AAA AAG TAC AAC      2087
Ile Asp Ser Cys Ser Ser Ser Glu Glu Ile Lys Lys Lys Lys Tyr Asn
                645                 650                 655

CAA ATG CCA GTC AGG CAC AGC AGA AAC CTA CAA CTC ATG GAA GGT AAA      2135
Gln Met Pro Val Arg His Ser Arg Asn Leu Gln Leu Met Glu Gly Lys
            660                 665                 670

GAA CCT GCA ACT GGA GCC AAG AAG AGT AAC AAG CCA AAT GAA CAG ACA      2183
Glu Pro Ala Thr Gly Ala Lys Lys Ser Asn Lys Pro Asn Glu Gln Thr
        675                 680                 685

AGT AAA AGA CAT GAC AGC GAT ACT TTC CCA GAG CTG AAG TTA ACA AAT      2231
Ser Lys Arg His Asp Ser Asp Thr Phe Pro Glu Leu Lys Leu Thr Asn
    690                 695                 700
```

E6 contd

```
GCA CCT GGT TCT TTT ACT AAG TGT TCA AAT ACC AGT GAA CTT AAA GAA      2279
Ala Pro Gly Ser Phe Thr Lys Cys Ser Asn Thr Ser Glu Leu Lys Glu
705                 710                 715                 720

TTT GTC AAT CCT AGC CTT CCA AGA GAA GAA AAA GAA GAG AAA CTA GAA      2327
Phe Val Asn Pro Ser Leu Pro Arg Glu Glu Lys Glu Glu Lys Leu Glu
                725                 730                 735

ACA GTT AAA GTG TCT AAT AAT GCT GAA GAC CCC AAA GAT CTC ATG TTA      2375
Thr Val Lys Val Ser Asn Asn Ala Glu Asp Pro Lys Asp Leu Met Leu
            740                 745                 750

AGT GGA GAA AGG GTT TTG CAA ACT GAA AGA TCT GTA GAG AGT AGC AGT      2423
Ser Gly Glu Arg Val Leu Gln Thr Glu Arg Ser Val Glu Ser Ser Ser
        755                 760                 765

ATT TCA TTG GTA CCT GGT ACT GAT TAT GGC ACT CAG GAA AGT ATC TCG      2471
Ile Ser Leu Val Pro Gly Thr Asp Tyr Gly Thr Gln Glu Ser Ile Ser
    770                 775                 780

TTA CTG GAA GTT AGC ACT CTA GGG AAG GCA AAA ACA GAA CCA AAT AAA      2519
Leu Leu Glu Val Ser Thr Leu Gly Lys Ala Lys Thr Glu Pro Asn Lys
785                 790                 795                 800

TGT GTG AGT CAG TGT GCA GCA TTT GAA AAC CCC AAG GGA CTA ATT CAT      2567
Cys Val Ser Gln Cys Ala Ala Phe Glu Asn Pro Lys Gly Leu Ile His
                805                 810                 815

GGT TGT TCC AAA GAT AAT AGA AAT GAC ACA GAA GGC TTT AAG TAT CCA      2615
Gly Cys Ser Lys Asp Asn Arg Asn Asp Thr Glu Gly Phe Lys Tyr Pro
            820                 825                 830

TTG GGA CAT GAA GTT AAC CAC AGT CGG GAA ACA AGC ATA GAA ATG GAA      2663
Leu Gly His Glu Val Asn His Ser Arg Glu Thr Ser Ile Glu Met Glu
        835                 840                 845

GAA AGT GAA CTT GAT GCT CAG TAT TTG CAG AAT ACA TTC AAG GTT TCA      2711
Glu Ser Glu Leu Asp Ala Gln Tyr Leu Gln Asn Thr Phe Lys Val Ser
    850                 855                 860

AAG CGC CAG TCA TTT GCT CCG TTT TCA AAT CCA GGA AAT GCA GAA GAG      2759
Lys Arg Gln Ser Phe Ala Pro Phe Ser Asn Pro Gly Asn Ala Glu Glu
865                 870                 875                 880

GAA TGT GCA ACA TTC TCT GCC CAC TCT GGG TCC TTA AAG AAA CAA AGT      2807
Glu Cys Ala Thr Phe Ser Ala His Ser Gly Ser Leu Lys Lys Gln Ser
                885                 890                 895

CCA AAA GTC ACT TTT GAA TGT GAA CAA AAG GAA GAA AAT CAA GGA AAG      2855
Pro Lys Val Thr Phe Glu Cys Glu Gln Lys Glu Glu Asn Gln Gly Lys
            900                 905                 910

AAT GAG TCT AAT ATC AAG CCT GTA CAG ACA GTT AAT ATC ACT GCA GGC      2903
Asn Glu Ser Asn Ile Lys Pro Val Gln Thr Val Asn Ile Thr Ala Gly
        915                 920                 925
```

E6 cont'd

```
TTT CCT GTG GTT GGT CAG AAA GAT AAG CCA GTT GAT AAT GCC AAA TGT      2951
Phe Pro Val Val Gly Gln Lys Asp Lys Pro Val Asp Asn Ala Lys Cys
    930                 935                 940

AGT ATC AAA GGA GGC TCT AGG TTT TGT CTA TCA TCT CAG TTC AGA GGC      2999
Ser Ile Lys Gly Gly Ser Arg Phe Cys Leu Ser Ser Gln Phe Arg Gly
945                 950                 955                 960

AAC GAA ACT GGA CTC ATT ACT CCA AAT AAA CAT GGA CTT TTA CAA AAC      3047
Asn Glu Thr Gly Leu Ile Thr Pro Asn Lys His Gly Leu Leu Gln Asn
                965                 970                 975

CCA TAT CGT ATA CCA CCA CTT TTT CCC ATC AAG TCA TTT GTT AAA ACT      3095
Pro Tyr Arg Ile Pro Pro Leu Phe Pro Ile Lys Ser Phe Val Lys Thr
            980                 985                 990

AAA TGT AAG AAA AAT CTG CTA GAG GAA AAC TTT GAG GAA CAT TCA ATG      3143
Lys Cys Lys Lys Asn Leu Leu Glu Glu Asn Phe Glu Glu His Ser Met
        995                 1000                1005

TCA CCT GAA AGA GAA ATG GGA AAT GAG AAC ATT CCA AGT ACA GTG AGC      3191
Ser Pro Glu Arg Glu Met Gly Asn Glu Asn Ile Pro Ser Thr Val Ser
    1010                1015                1020

ACA ATT AGC CGT AAT AAC ATT AGA GAA AAT GTT TTT AAA GAA GCC AGC      3239
Thr Ile Ser Arg Asn Asn Ile Arg Glu Asn Val Phe Lys Glu Ala Ser
1025                1030                1035                1040

TCA AGC AAT ATT AAT GAA GTA GGT TCC AGT ACT AAT GAA GTG GGC TCC      3287
Ser Ser Asn Ile Asn Glu Val Gly Ser Ser Thr Asn Glu Val Gly Ser
                1045                1050                1055

AGT ATT AAT GAA ATA GGT TCC AGT GAT GAA AAC ATT CAA GCA GAA CTA      3335
Ser Ile Asn Glu Ile Gly Ser Ser Asp Glu Asn Ile Gln Ala Glu Leu
            1060                1065                1070

GGT AGA AAC AGA GGG CCA AAA TTG AAT GCT ATG CTT AGA TTA GGG GTT      3383
Gly Arg Asn Arg Gly Pro Lys Leu Asn Ala Met Leu Arg Leu Gly Val
        1075                1080                1085

TTG CAA CCT GAG GTC TAT AAA CAA AGT CTT CCT GGA AGT AAT TGT AAG      3431
Leu Gln Pro Glu Val Tyr Lys Gln Ser Leu Pro Gly Ser Asn Cys Lys
    1090                1095                1100

CAT CCT GAA ATA AAA AAG CAA GAA TAT GAA GAA GTA GTT CAG ACT GTT      3479
His Pro Glu Ile Lys Lys Gln Glu Tyr Glu Glu Val Val Gln Thr Val
1105                1110                1115                1120

AAT ACA GAT TTC TCT CCA TAT CTG ATT TCA GAT AAC TTA GAA CAG CCT      3527
Asn Thr Asp Phe Ser Pro Tyr Leu Ile Ser Asp Asn Leu Glu Gln Pro
                1125                1130                1135

ATG GGA AGT AGT CAT GCA TCT CAG GTT TGT TCT GAG ACA CCT GAT GAC      3575
Met Gly Ser Ser His Ala Ser Gln Val Cys Ser Glu Thr Pro Asp Asp
            1140                1145                1150
```

E6 control

```
CTG TTA GAT GAT GGT GAA ATA AAG GAA GAT ACT AGT TTT GCT GAA AAT      3623
Leu Leu Asp Asp Gly Glu Ile Lys Glu Asp Thr Ser Phe Ala Glu Asn
        1155                1160                1165

GAC ATT AAG GAA AGT TCT GCT GTT TTT AGC AAA AGC GTC CAG AAA GGA      3671
Asp Ile Lys Glu Ser Ser Ala Val Phe Ser Lys Ser Val Gln Lys Gly
    1170                1175                1180

GAG CTT AGC AGG AGT CCT AGC CCT TTC ACC CAT ACA CAT TTG GCT CAG      3719
Glu Leu Ser Arg Ser Pro Ser Pro Phe Thr His Thr His Leu Ala Gln
1185                1190                1195                1200

GGT TAC CGA AGA GGG GCC AAG AAA TTA GAG TCC TCA GAA GAG AAC TTA      3767
Gly Tyr Arg Arg Gly Ala Lys Lys Leu Glu Ser Ser Glu Glu Asn Leu
                1205                1210                1215

TCT AGT GAG GAT GAA GAG CTT CCC TGC TTC CAA CAC TTG TTA TTT GGT      3815
Ser Ser Glu Asp Glu Glu Leu Pro Cys Phe Gln His Leu Leu Phe Gly
            1220                1225                1230

AAA GTA AAC AAT ATA CCT TCT CAG TCT ACT AGG CAT AGC ACC GTT GCT      3863
Lys Val Asn Asn Ile Pro Ser Gln Ser Thr Arg His Ser Thr Val Ala
        1235                1240                1245

ACC GAG TGT CTG TCT AAG AAC ACA GAG GAG AAT TTA TTA TCA TTG AAG      3911
Thr Glu Cys Leu Ser Lys Asn Thr Glu Glu Asn Leu Leu Ser Leu Lys
    1250                1255                1260

AAT AGC TTA AAT GAC TGC AGT AAC CAG GTA ATA TTG GCA AAG GCA TCT      3959
Asn Ser Leu Asn Asp Cys Ser Asn Gln Val Ile Leu Ala Lys Ala Ser
1265                1270                1275                1280

CAG GAA CAT CAC CTT AGT GAG GAA ACA AAA TGT TCT GCT AGC TTG TTT      4007
Gln Glu His His Leu Ser Glu Glu Thr Lys Cys Ser Ala Ser Leu Phe
                1285                1290                1295

TCT TCA CAG TGC AGT GAA TTG GAA GAC TTG ACT GCA AAT ACA AAC ACC      4055
Ser Ser Gln Cys Ser Glu Leu Glu Asp Leu Thr Ala Asn Thr Asn Thr
            1300                1305                1310

CAG GAT CCT TTC TTG ATT GGT TCT TCC AAA CAA ATG AGG CAT CAG TCT      4103
Gln Asp Pro Phe Leu Ile Gly Ser Ser Lys Gln Met Arg His Gln Ser
        1315                1320                1325

GAA AGC CAG GGA GTT GGT CTG AGT GAC AAG GAA TTG GTT TCA GAT GAT      4151
Glu Ser Gln Gly Val Gly Leu Ser Asp Lys Glu Leu Val Ser Asp Asp
    1330                1335                1340

GAA GAA AGA GGA ACG GGC TTG GAA GAA AAT AAT CAA GAA GAG CAA AGC      4199
Glu Glu Arg Gly Thr Gly Leu Glu Glu Asn Asn Gln Glu Glu Gln Ser
1345                1350                1355                1360

ATG GAT TCA AAC TTA GGT GAA GCA GCA TCT GGG TGT GAG AGT GAA ACA      4247
Met Asp Ser Asn Leu Gly Glu Ala Ala Ser Gly Cys Glu Ser Glu Thr
                1365                1370                1375
```

E6 cont'd

```
AGC GTC TCT GAA GAC TGC TCA GGG CTA TCC TCT CAG AGT GAC ATT TTA      4295
Ser Val Ser Glu Asp Cys Ser Gly Leu Ser Ser Gln Ser Asp Ile Leu
            1380                1385                1390

ACC ACT CAG CAG AGG GAT ACC ATG CAA CAT AAC CTG ATA AAG CTC CAG      4343
Thr Thr Gln Gln Arg Asp Thr Met Gln His Asn Leu Ile Lys Leu Gln
        1395                1400                1405

CAG GAA ATG GCT GAA CTA GAA GCT GTG TTA GAA CAG CAT GGG AGC CAG      4391
Gln Glu Met Ala Glu Leu Glu Ala Val Leu Glu Gln His Gly Ser Gln
    1410                1415                1420

CCT TCT AAC AGC TAC CCT TCC ATC ATA AGT GAC TCT TCT GCC CTT GAG      4439
Pro Ser Asn Ser Tyr Pro Ser Ile Ile Ser Asp Ser Ser Ala Leu Glu
1425                1430                1435                1440

GAC CTG CGA AAT CCA GAA CAA AGC ACA TCA GAA AAA GCA GTA TTA ACT      4487
Asp Leu Arg Asn Pro Glu Gln Ser Thr Ser Glu Lys Ala Val Leu Thr
                1445                1450                1455

TCA CAG AAA AGT AGT GAA TAC CCT ATA AGC CAG AAT CCA GAA GGC CTT      4535
Ser Gln Lys Ser Ser Glu Tyr Pro Ile Ser Gln Asn Pro Glu Gly Leu
            1460                1465                1470

TCT GCT GAC AAG TTT GAG GTG TCT GCA GAT AGT TCT ACC AGT AAA AAT      4583
Ser Ala Asp Lys Phe Glu Val Ser Ala Asp Ser Ser Thr Ser Lys Asn
        1475                1480                1485

AAA GAA CCA GGA GTG GAA AGG TCA TCC CCT TCT AAA TGC CCA TCA TTA      4631
Lys Glu Pro Gly Val Glu Arg Ser Ser Pro Ser Lys Cys Pro Ser Leu
    1490                1495                1500

GAT GAT AGG TGG TAC ATG CAC AGT TGC TCT GGG AGT CTT CAG AAT AGA      4679
Asp Asp Arg Trp Tyr Met His Ser Cys Ser Gly Ser Leu Gln Asn Arg
1505                1510                1515                1520

AAC TAC CCA TCT CAA GAG GAG CTC ATT AAG GTT GTT GAT GTG GAG GAG      4727
Asn Tyr Pro Ser Gln Glu Glu Leu Ile Lys Val Val Asp Val Glu Glu
                1525                1530                1535

CAA CAG CTG GAA GAG TCT GGG CCA CAC GAT TTG ACG GAA ACA TCT TAC      4775
Gln Gln Leu Glu Glu Ser Gly Pro His Asp Leu Thr Glu Thr Ser Tyr
            1540                1545                1550

TTG CCA AGG CAA GAT CTA GAG GGA ACC CCT TAC CTG GAA TCT GGA ATC      4823
Leu Pro Arg Gln Asp Leu Glu Gly Thr Pro Tyr Leu Glu Ser Gly Ile
        1555                1560                1565

AGC CTC TTC TCT GAT GAC CCT GAA TCT GAT CCT TCT GAA GAC AGA GCC      4871
Ser Leu Phe Ser Asp Asp Pro Glu Ser Asp Pro Ser Glu Asp Arg Ala
    1570                1575                1580

CCA GAG TCA GCT CGT GTT GGC AAC ATA CCA TCT TCA ACC TCT GCA TTG      4919
Pro Glu Ser Ala Arg Val Gly Asn Ile Pro Ser Ser Thr Ser Ala Leu
1585                1590                1595                1600
```

E6 cont'd

GeneDX 1009, pg. 117

```
AAA GTT CCC CAA TTG AAA GTT GCA GAA TCT GCC CAG AGT CCA GCT GCT      4967
Lys Val Pro Gln Leu Lys Val Ala Glu Ser Ala Gln Ser Pro Ala Ala
                1605                1610                1615

GCT CAT ACT ACT GAT ACT GCT GGG TAT AAT GCA ATG GAA GAA AGT GTG      5015
Ala His Thr Thr Asp Thr Ala Gly Tyr Asn Ala Met Glu Glu Ser Val
            1620                1625                1630

AGC AGG GAG AAG CCA GAA TTG ACA GCT TCA ACA GAA AGG GTC AAC AAA      5063
Ser Arg Glu Lys Pro Glu Leu Thr Ala Ser Thr Glu Arg Val Asn Lys
        1635                1640                1645

AGA ATG TCC ATG GTG GTG TCT GGC CTG ACC CCA GAA GAA TTT ATG CTC      5111
Arg Met Ser Met Val Val Ser Gly Leu Thr Pro Glu Glu Phe Met Leu
    1650                1655                1660

GTG TAC AAG TTT GCC AGA AAA CAC CAC ATC ACT TTA ACT AAT CTA ATT      5159
Val Tyr Lys Phe Ala Arg Lys His His Ile Thr Leu Thr Asn Leu Ile
1665                1670                1675                1680

ACT GAA GAG ACT ACT CAT GTT GTT ATG AAA ACA GAT GCT GAG TTT GTG      5207
Thr Glu Glu Thr Thr His Val Val Met Lys Thr Asp Ala Glu Phe Val
                1685                1690                1695

TGT GAA CGG ACA CTG AAA TAT TTT CTA GGA ATT GCG GGA GGA AAA TGG      5255
Cys Glu Arg Thr Leu Lys Tyr Phe Leu Gly Ile Ala Gly Gly Lys Trp
            1700                1705                1710

GTA GTT AGC TAT TTC TGG GTG ACC CAG TCT ATT AAA GAA AGA AAA ATG      5303
Val Val Ser Tyr Phe Trp Val Thr Gln Ser Ile Lys Glu Arg Lys Met
        1715                1720                1725

CTG AAT GAG CAT GAT TTT GAA GTC AGA GGA GAT GTG GTC AAT GGA AGA      5351
Leu Asn Glu His Asp Phe Glu Val Arg Gly Asp Val Val Asn Gly Arg
    1730                1735                1740

AAC CAC CAA GGT CCA AAG CGA GCA AGA GAA TCC CAG GAC AGA AAG ATC      5399
Asn His Gln Gly Pro Lys Arg Ala Arg Glu Ser Gln Asp Arg Lys Ile
1745                1750                1755                1760

TTC AGG GGG CTA GAA ATC TGT TGC TAT GGG CCC TTC ACC AAC ATG CCC      5447
Phe Arg Gly Leu Glu Ile Cys Cys Tyr Gly Pro Phe Thr Asn Met Pro
                1765                1770                1775

ACA GAT CAA CTG GAA TGG ATG GTA CAG CTG TGT GGT GCT TCT GTG GTG      5495
Thr Asp Gln Leu Glu Trp Met Val Gln Leu Cys Gly Ala Ser Val Val
            1780                1785                1790

AAG GAG CTT TCA TCA TTC ACC CTT GGC ACA GGT GTC CAC CCA ATT GTG      5543
Lys Glu Leu Ser Ser Phe Thr Leu Gly Thr Gly Val His Pro Ile Val
        1795                1800                1805

GTT GTG CAG CCA GAT GCC TGG ACA GAG GAC AAT GGC TTC CAT GCA ATT      5591
Val Val Gln Pro Asp Ala Trp Thr Glu Asp Asn Gly Phe His Ala Ile
    1810                1815                1820
```

E6 cont'd

GeneDX 1009, pg. 118

```
GGG CAG ATG TGT GAG GCA CCT GTG GTG ACC CGA GAG TGG GTG TTG GAC      5639
Gly Gln Met Cys Glu Ala Pro Val Val Thr Arg Glu Trp Val Leu Asp
1825                 1830                 1835                 1840

AGT GTA GCA CTC TAC CAG TGC CAG GAG CTG GAC ACC TAC CTG ATA CCC      5687
Ser Val Ala Leu Tyr Gln Cys Gln Glu Leu Asp Thr Tyr Leu Ile Pro
                 1845                 1850                 1855

CAG ATC CCC CAC AGC CAC TAC TGA CTGCAGCCAG CCACAGGTAC AGAGCCACAG     5741
Gln Ile Pro His Ser His Tyr
            1860

GACCCCAAGA ATGAGCTTAC AAAGTGGCCT TTCCAGGCCC TGGGAGCTCC TCTCACTCTT    5801

CAGTCCTTCT ACTGTCCTGG CTACTAAATA TTTTATGTAC ATCAGCCTGA AAAGGACTTC    5861

TGGCTATGCA AGGGTCCCTT AAAGATTTTC TGCTTGAAGT CTCCCTTGGA AAT           5914
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1863 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

E6 cont'd

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
  1               5                  10                  15

Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
             20                  25                  30

Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
         35                  40                  45

Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
     50                  55                  60

Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
 65                  70                  75                  80

Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
                 85                  90                  95

Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
            100                 105                 110

Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
        115                 120                 125

Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
    130                 135                 140
```

114

GeneDX 1009, pg. 119

```
Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
145                 150                 155                 160

Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
                165                 170                 175

Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
            180                 185                 190

Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
        195                 200                 205

Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
    210                 215                 220

Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
225                 230                 235                 240

Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg
                245                 250                 255

His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu
            260                 265                 270

Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser
        275                 280                 285

Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe
    290                 295                 300

Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg
305                 310                 315                 320

Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr
                325                 330                 335

Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu
            340                 345                 350

Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu
        355                 360                 365

Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu
    370                 375                 380

Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp
385                 390                 395                 400

Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu
                405                 410                 415

Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu
            420                 425                 430
```

E6 cont'd

```
Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Glu Arg Val His
        435                 440                 445

Ser Lys Ser Val Glu Ser Asn Ile Glu Asp Lys Ile Phe Gly Lys Thr
    450                 455                 460

Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn
465                 470                 475                 480

Leu Ile Ile Gly Ala Phe Val Thr Glu Pro Gln Ile Ile Gln Glu Arg
                485                 490                 495

Pro Leu Thr Asn Lys Leu Lys Arg Lys Arg Arg Pro Thr Ser Gly Leu
            500                 505                 510

His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr
        515                 520                 525

Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln
    530                 535                 540

Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp
545                 550                 555                 560

Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys
                565                 570                 575

Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser
            580                 585                 590

Asn Met Glu Leu Glu Leu Asn Ile His Asn Ser Lys Ala Pro Lys Lys
        595                 600                 605

Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Ala Leu Glu
    610                 615                 620

Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln
625                 630                 635                 640

Ile Asp Ser Cys Ser Ser Ser Glu Glu Ile Lys Lys Lys Lys Tyr Asn
                645                 650                 655

Gln Met Pro Val Arg His Ser Arg Asn Leu Gln Leu Met Glu Gly Lys
            660                 665                 670

Glu Pro Ala Thr Gly Ala Lys Lys Ser Asn Lys Pro Asn Glu Gln Thr
        675                 680                 685

Ser Lys Arg His Asp Ser Asp Thr Phe Pro Glu Leu Lys Leu Thr Asn
    690                 695                 700

Ala Pro Gly Ser Phe Thr Lys Cys Ser Asn Thr Ser Glu Leu Lys Glu
705                 710                 715                 720

Phe Val Asn Pro Ser Leu Pro Arg Glu Glu Lys Glu Glu Lys Leu Glu
                725                 730                 735
```

E6 contd

116

```
Thr Val Lys Val Ser Asn Asn Ala Glu Asp Pro Lys Asp Leu Met Leu
            740                 745                 750

Ser Gly Glu Arg Val Leu Gln Thr Glu Arg Ser Val Glu Ser Ser Ser
        755                 760                 765
Ile Ser Leu Val Pro Gly Thr Asp Tyr Gly Thr Gln Glu Ser Ile Ser
    770                 775                 780

Leu Leu Glu Val Ser Thr Leu Gly Lys Ala Lys Thr Glu Pro Asn Lys
785                 790                 795                 800

Cys Val Ser Gln Cys Ala Ala Phe Glu Asn Pro Lys Gly Leu Ile His
                805                 810                 815

Gly Cys Ser Lys Asp Asn Arg Asn Asp Thr Glu Gly Phe Lys Tyr Pro
            820                 825                 830

Leu Gly His Glu Val Asn His Ser Arg Glu Thr Ser Ile Glu Met Glu
        835                 840                 845

Glu Ser Glu Leu Asp Ala Gln Tyr Leu Gln Asn Thr Phe Lys Val Ser
    850                 855                 860

Lys Arg Gln Ser Phe Ala Pro Phe Ser Asn Pro Gly Asn Ala Glu Glu
865                 870                 875                 880

Glu Cys Ala Thr Phe Ser Ala His Ser Gly Ser Leu Lys Lys Gln Ser
                885                 890                 895

Pro Lys Val Thr Phe Glu Cys Glu Gln Lys Glu Glu Asn Gln Gly Lys
            900                 905                 910

Asn Glu Ser Asn Ile Lys Pro Val Gln Thr Val Asn Ile Thr Ala Gly
        915                 920                 925

Phe Pro Val Val Gly Gln Lys Asp Lys Pro Val Asp Asn Ala Lys Cys
    930                 935                 940

Ser Ile Lys Gly Gly Ser Arg Phe Cys Leu Ser Ser Gln Phe Arg Gly
945                 950                 955                 960

Asn Glu Thr Gly Leu Ile Thr Pro Asn Lys His Gly Leu Leu Gln Asn
                965                 970                 975

Pro Tyr Arg Ile Pro Pro Leu Phe Pro Ile Lys Ser Phe Val Lys Thr
            980                 985                 990

Lys Cys Lys Lys Asn Leu Leu Glu Glu Asn Phe Glu Glu His Ser Met
        995                 1000                1005

Ser Pro Glu Arg Glu Met Gly Asn Glu Asn Ile Pro Ser Thr Val Ser
    1010                1015                1020

Thr Ile Ser Arg Asn Asn Ile Arg Glu Asn Val Phe Lys Glu Ala Ser
1025                1030                1035                1040
```

E6 cont'd

117

Ser Ser Asn Ile Asn Glu Val Gly Ser Ser Thr Asn Glu Val Gly Ser
1045 1050 1055

Ser Ile Asn Glu Ile Gly Ser Ser Asp Glu Asn Ile Gln Ala Glu Leu
1060 1065 1070
Gly Arg Asn Arg Gly Pro Lys Leu Asn Ala Met Leu Arg Leu Gly Val
1075 1080 1085

Leu Gln Pro Glu Val Tyr Lys Gln Ser Leu Pro Gly Ser Asn Cys Lys
1090 1095 1100

His Pro Glu Ile Lys Lys Gln Glu Tyr Glu Glu Val Val Gln Thr Val
1105 1110 1115 1120

Asn Thr Asp Phe Ser Pro Tyr Leu Ile Ser Asp Asn Leu Glu Gln Pro
1125 1130 1135

Met Gly Ser Ser His Ala Ser Gln Val Cys Ser Glu Thr Pro Asp Asp
1140 1145 1150

Leu Leu Asp Asp Gly Glu Ile Lys Glu Asp Thr Ser Phe Ala Glu Asn
1155 1160 1165

Asp Ile Lys Glu Ser Ser Ala Val Phe Ser Lys Ser Val Gln Lys Gly
1170 1175 1180

Glu Leu Ser Arg Ser Pro Ser Pro Phe Thr His Thr His Leu Ala Gln
1185 1190 1195 1200

E6 cont'd

Gly Tyr Arg Arg Gly Ala Lys Lys Leu Glu Ser Ser Glu Glu Asn Leu
1205 1210 1215

Ser Ser Glu Asp Glu Glu Leu Pro Cys Phe Gln His Leu Leu Phe Gly
1220 1225 1230

Lys Val Asn Asn Ile Pro Ser Gln Ser Thr Arg His Ser Thr Val Ala
1235 1240 1245

Thr Glu Cys Leu Ser Lys Asn Thr Glu Glu Asn Leu Leu Ser Leu Lys
1250 1255 1260

Asn Ser Leu Asn Asp Cys Ser Asn Gln Val Ile Leu Ala Lys Ala Ser
1265 1270 1275 1280

Gln Glu His His Leu Ser Glu Glu Thr Lys Cys Ser Ala Ser Leu Phe
1285 1290 1295

Ser Ser Gln Cys Ser Glu Leu Glu Asp Leu Thr Ala Asn Thr Asn Thr
1300 1305 1310

Gln Asp Pro Phe Leu Ile Gly Ser Ser Lys Gln Met Arg His Gln Ser
1315 1320 1325

Glu Ser Gln Gly Val Gly Leu Ser Asp Lys Glu Leu Val Ser Asp Asp
1330 1335 1340

115

```
Glu Glu Arg Gly Thr Gly Leu Glu Glu Asn Asn Gln Glu Glu Gln Ser
1345                1350                1355                1360

Met Asp Ser Asn Leu Gly Glu Ala Ala Ser Gly Cys Glu Ser Glu Thr
                1365                1370                1375

Ser Val Ser Glu Asp Cys Ser Gly Leu Ser Ser Gln Ser Asp Ile Leu
            1380                1385                1390

Thr Thr Gln Gln Arg Asp Thr Met Gln His Asn Leu Ile Lys Leu Gln
        1395                1400                1405

Gln Glu Met Ala Glu Leu Glu Ala Val Leu Glu Gln His Gly Ser Gln
    1410                1415                1420

Pro Ser Asn Ser Tyr Pro Ser Ile Ile Ser Asp Ser Ser Ala Leu Glu
1425                1430                1435                1440

Asp Leu Arg Asn Pro Glu Gln Ser Thr Ser Glu Lys Ala Val Leu Thr
                1445                1450                1455

Ser Gln Lys Ser Ser Glu Tyr Pro Ile Ser Gln Asn Pro Glu Gly Leu
            1460                1465                1470

Ser Ala Asp Lys Phe Glu Val Ser Ala Asp Ser Ser Thr Ser Lys Asn
        1475                1480                1485

Lys Glu Pro Gly Val Glu Arg Ser Ser Pro Ser Lys Cys Pro Ser Leu
    1490                1495                1500

Asp Asp Arg Trp Tyr Met His Ser Cys Ser Gly Ser Leu Gln Asn Arg
1505                1510                1515                1520

Asn Tyr Pro Ser Gln Glu Glu Leu Ile Lys Val Val Asp Val Glu Glu
                1525                1530                1535

Gln Gln Leu Glu Glu Ser Gly Pro His Asp Leu Thr Glu Thr Ser Tyr
            1540                1545                1550

Leu Pro Arg Gln Asp Leu Glu Gly Thr Pro Tyr Leu Glu Ser Gly Ile
        1555                1560                1565

Ser Leu Phe Ser Asp Asp Pro Glu Ser Asp Pro Ser Glu Asp Arg Ala
    1570                1575                1580

Pro Glu Ser Ala Arg Val Gly Asn Ile Pro Ser Ser Thr Ser Ala Leu
1585                1590                1595                1600

Lys Val Pro Gln Leu Lys Val Ala Glu Ser Ala Gln Ser Pro Ala Ala
                1605                1610                1615

Ala His Thr Thr Asp Thr Ala Gly Tyr Asn Ala Met Glu Glu Ser Val
            1620                1625                1630

Ser Arg Glu Lys Pro Glu Leu Thr Ala Ser Thr Glu Arg Val Asn Lys
        1635                1640                1645
```

E6 / cont'd

```
Arg Met Ser Met Val Val Ser Gly Leu Thr Pro Glu Glu Phe Met Leu
    1650                1655                1660

Val Tyr Lys Phe Ala Arg Lys His His Ile Thr Leu Thr Asn Leu Ile
1665                1670                1675                1680

Thr Glu Glu Thr Thr His Val Val Met Lys Thr Asp Ala Glu Phe Val
                1685                1690                1695

Cys Glu Arg Thr Leu Lys Tyr Phe Leu Gly Ile Ala Gly Gly Lys Trp
            1700                1705                1710

Val Val Ser Tyr Phe Trp Val Thr Gln Ser Ile Lys Glu Arg Lys Met
        1715                1720                1725

Leu Asn Glu His Asp Phe Glu Val Arg Gly Asp Val Val Asn Gly Arg
    1730                1735                1740

Asn His Gln Gly Pro Lys Arg Ala Arg Glu Ser Gln Asp Arg Lys Ile
1745                1750                1755                1760

Phe Arg Gly Leu Glu Ile Cys Cys Tyr Gly Pro Phe Thr Asn Met Pro
                1765                1770                1775

Thr Asp Gln Leu Glu Trp Met Val Gln Leu Cys Gly Ala Ser Val Val
            1780                1785                1790

Lys Glu Leu Ser Ser Phe Thr Leu Gly Thr Gly Val His Pro Ile Val
        1795                1800                1805

Val Val Gln Pro Asp Ala Trp Thr Glu Asp Asn Gly Phe His Ala Ile
    1810                1815                1820

Gly Gln Met Cys Glu Ala Pro Val Val Thr Arg Glu Trp Val Leu Asp
1825                1830                1835                1840

Ser Val Ala Leu Tyr Gln Cys Gln Glu Leu Asp Thr Tyr Leu Ile Pro
                1845                1850                1855

Gln Ile Pro His Ser His Tyr
            1860
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

GeneDX 1009, pg. 125

(vii) IMMEDIATE SOURCE:
(B) CLONE: s754 A

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CTAGCCTGGG CAACAAACGA 20

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
(B) CLONE: s754 B

E6) cont'd

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GCAGGAAGCA GGAATGGAAC 20

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
(B) CLONE: s975 A

121

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TAGGAGATGG ATTATTGGTG 20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
(B) CLONE: s975 B

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGGCAACTTT GCAATGAGTG 20

E6 contd

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
(B) CLONE: tdj1474 A

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CAGAGTGAGA CCTTGTCTCA AA 22

122

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
(B) CLONE: tdj1474 B

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TTCTGCAAAC ACCTTAAACT CAG 23

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
(B) CLONE: tdj1239 A

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

AACCTGGAAG GCAGAGGTTG 20

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

c6 cont'd

123

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
(B) CLONE: tdj1239 B

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TCTGTACCTG CTAAGCAGTG G 21

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 111 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

E6 contd

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..111

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
G GKC TTA CTC TGT TGT CCC AGC TGG AGT ACA GWG TGC GAT CAT GAG          46
  Xaa Leu Leu Cys Cys Pro Ser Trp Ser Thr Xaa Cys Asp His Glu
  1865                1870                1875

GCT TAC TGT TGC TTG ACT CCT AGG CTC AAG CGA TCC TAT CAC CTC AGT        94
Ala Tyr Cys Cys Leu Thr Pro Arg Leu Lys Arg Ser Tyr His Leu Ser
1880                1885                1890                1895

CTC CAA GTA GCT GGA  CT                                           111
Leu Gln Val Ala Gly
                1900
```


GeneDX 1009, pg. 129

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Xaa Leu Leu Cys Cys Pro Ser Trp Ser Thr Xaa Cys Asp His Glu Ala
1 5 10 15

Tyr Cys Cys Leu Thr Pro Arg Leu Lys Arg Ser Tyr His Leu Ser Leu
20 25 30

Gln Val Ala Gly
35

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1534 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

E6 cont'd

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAGGCTAGAG GGCAGGCACT TTATGGCAAA CTCAGGTAGA ATTCTTCCTC TTCCGTCTCT 60

TTCCTTTTAC GTCATCGGGG AGACTGGGTG GCAATCGCAG CCCGAGAGAC GCATGGCTCT 120

TTCTGCCCTC CATCCTCTGA TGTACCTTGA TTTCGTATTC TGAGAGGCTG CTGCTTAGCG 180

GTAGCCCCTT GGTTTCCGTG GCAACGGAAA AGCGCGGGAA TTACAGATAA ATTAAAACTG 240

CGACTGCGCG GCGTGAGCTC GCTGAGACTT CCTGGACCCC GCACCAGGCT GTGGGGTTTC 300

TCAGATAACT GGGCCCCTGC GCTCAGGAGG CCTTCACCCT CTGCTCTGGG TAAAGGTAGT 360

AGAGTCCCGG GAAAGGGACA GGGGGCCCAA GTGATGCTCT GGGGTACTGG CGTGGGAGAG 420

TGGATTTCCG AAGCTGACAG ATGGGTATTC TTTGACGGGG GGTAGGGGCG GAACCTGAGA 480

125

GGCGTAAGGC GTTGTGAACC CTGGGGAGGG GGGCAGTTTG TAGGTCGCGA GGGAAGCGCT 540

GAGGATCAGG AAGGGGGCAC TGAGTGTCCG TGGGGGAATC CTCGTGATAG GAACTGGAAT 600

ATGCCTTGAG GGGGACACTA TGTCTTTAAA AACGTCGGCT GGTCATGAGG TCAGGAGTTC 660

CAGACCAGCC TGACCAACGT GGTGAAACTC CGTCTCTACT AAAAATACNA AAATTAGCCG 720

GGCGTGGTGC CGCTCCAGCT ACTCAGGAGG CTGAGGCAGG AGAATCGCTA GAACCCGGGA 780

GGCGGAGGTT GCAGTGAGCC GAGATCGCGC CATTGCACTC CAGCCTGGGC GACAGAGCGA 840

GACTGTCTCA AAACAAAACA AAACAAAACA AAACAAAAAA CACCGGCTGG TATGTATGAG 900

AGGATGGGAC CTTGTGGAAG AAGAGGTGCC AGGAATATGT CTGGGAAGGG GAGGAGACAG 960

GATTTTGTGG GAGGGAGAAC TTAAGAACTG GATCCATTTG CGCCATTGAG AAAGCGCAAG 1020

AGGGAAGTAG AGGAGCGTCA GTAGTAACAG ATGCTGCCGG CAGGGATGTG CTTGAGGAGG 1080

ATCCAGAGAT GAGAGCAGGT CACTGGGAAA GGTTAGGGGC GGGGAGGCCT TGATTGGTGT 1140

TGGTTTGGTC GTTGTTGATT TTGGTTTTAT GCAAGAAAAA GAAAACAACC AGAAACATTG 1200

GAGAAAGCTA AGGCTACCAC CACCTACCCG GTCAGTCACT CCTCTGTAGC TTTCTCTTTC 1260

TTGGAGAAAG GAAAAGACCC AAGGGGTTGG CAGCGATATG TGAAAAAATT CAGAATTTAT 1320

GTTGTCTAAT TACAAAAAGC AACTTCTAGA ATCTTTAAAA ATAAAGGACG TTGTCATTAG 1380

TTCTTCTGGT TTGTATTATT CTAAAACCTT CCAAATCTTC AAATTTACTT TATTTTAAAA 1440

TGATAAAATG AAGTTGTCAT TTTATAAACC TTTTAAAAAG ATATATATAT ATGTTTTTCT 1500

AATGTGTTAA AGTTCATTGG AACAGAAAGA AATG 1534

EL cont'd

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1924 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

126

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAGGCTAGAG GGCAGGCACT TTATGGCAAA CTCAGGTAGA ATTCTTCCTC TTCCGTCTCT 60

TTCCTTTTAC GTCATCGGGG AGACTGGGTG GCAATCGCAG CCCGAGAGAC GCATGGCTCT 120

TTCTGCCCTC CATCCTCTGA TGTACCTTGA TTTCGTATTC TGAGAGGCTG CTGCTTAGCG 180

GTAGCCCCTT GGTTTCCGTG GCAACGGAAA AGCGCGGGAA TTACAGATAA ATTAAAACTG 240

CGACTGCGCG GCGTGAGCTC GCTGAGACTT CCTGGACCCC GCACCAGGCT GTGGGGTTTC 300

TCAGATAACT GGGCCCCTGC GCTCAGGAGG CCTTCACCCT CTGCTCTGGG TAAAGGTAGT 360

AGAGTCCCGG GAAAGGGACA GGGGGCCCAA GTGATGCTCT GGGGTACTGG CGTGGGAGAG 420

TGGATTTCCG AAGCTGACAG ATGGGTATTC TTTGACGGGG GGTAGGGGCG GAACCTGAGA 480

GGCGTAAGGC GTTGTGAACC CTGGGGAGGG GGGCAGTTTG TAGGTCGCGA GGGAAGCGCT 540

GAGGATCAGG AAGGGGGCAC TGAGTGTCCG TGGGGGAATC CTCGTGATAG GAACTGGAAT 600

ATGCCTTGAG GGGGACACTA TGTCTTTAAA AACGTCGGCT GGTCATGAGG TCAGGAGTTC 660

CAGACCAGCC TGACCAACGT GGTGAAACTC CGTCTCTACT AAAAATACNA AAATTAGCCG 720

GGCGTGGTGC CGCTCCAGCT ACTCAGGAGG CTGAGGCAGG AGAATCGCTA GAACCCGGGA 780

GGCGGAGGTT GCAGTGAGCC GAGATCGCGC CATTGCACTC CAGCCTGGGC GACAGAGCGA 840

E6 contd

GACTGTCTCA AAACAAAACA AAACAAAACA AAACAAAAAA CACCGGCTGG TATGTATGAG 900

AGGATGGGAC CTTGTGGAAG AAGAGGTGCC AGGAATATGT CTGGGAAGGG GAGGAGACAG 960

GATTTTGTGG GAGGGAGAAC TTAAGAACTG GATCCATTTG CGCCATTGAG AAAGCGCAAG 1020

AGGGAAGTAG AGGAGCGTCA GTAGTAACAG ATGCTGCCGG CAGGGATGTG CTTGAGGAGG 1080

ATCCAGAGAT GAGAGCAGGT CACTGGGAAA GGTTAGGGGC GGGGAGGCCT TGATTGGTGT 1140

TGGTTTGGTC GTTGTTGATT TTGGTTTTAT GCAAGAAAAA GAAAACAACC AGAAACATTG 1200

GAGAAAGCTA AGGCTACCAC CACCTACCCG GTCAGTCACT CCTCTGTAGC TTTCTCTTTC 1260

TTGGAGAAAG GAAAAGACCC AAGGGGTTGG CAGCGATATG TGAAAAAATT CAGAATTTAT 1320

GTTGTCTAAT TACAAAAAGC AACTTCTAGA ATCTTTAAAA ATAAAGGACG TTGTCATTAG 1380

TTCTTCTGGT TTGTATTATT CTAAAACCTT CCAAATCTTC AAATTTACTT TATTTTAAAA 1440

TGATAAAATG AAGTTGTCAT TTTATAAACC TTTTAAAAAG ATATATATAT ATGTTTTTCT 1500

AATGTGTTAA AGTTCATTGG AACAGAAAGA AATGGATTTA TCTGCTCTTC GCGTTGAAGA 1560

AGTACAAAAT GTCATTAATG CTATGCAGAA AATCTTAGAG TGTCCCATCT GGTAAGTCAG 1620

127

CACAAGAGTG TATTAATTTG GGATTCCTAT GATTATCTCC TATGCAAATG AACAGAATTG 1680

ACCTTACATA CTAGGGAAGA AAAGACATGT CTAGTAAGAT TAGGCTATTG TAATTGCTGA 1740

TTTTCTTAAC TGAAGAACTT TAAAAATATA GAAAATGATT CCTTGTTCTC CATCCACTCT 1800

GCCTCTCCCA CTCCTCTCCT TTTCAACACA ATCCTGTGGT CCGGGAAAGA CAGGGCTCTG 1860

TCTTGATTGG TTCTGCACTG GGCAGGATCT GTTAGATACT GCATTTGCTT TCTCCAGCTC 1920

TAAA 1924

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 631 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

E6 contd

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AAATGCTGAT GATAGTATAG AGTATTGAAG GGATCAATAT AATTCTGTTT TGATATCTGA 60

AAGCTCACTG AAGGTAAGGA TCGTATTCTC TGCTGTATTC TCAGTTCCTG ACACAGCAGA 120

CATTTAATAA ATATTGAACG AACTTGAGGC CTTATGTTGA CTCAGTCATA ACAGCTCAAA 180

GTTGAACTTA TTCACTAAGA ATAGCTTTAT TTTTAAATAA ATTATTGAGC CTCATTTATT 240

TTCTTTTTCT CCCCCCCCTA CCCTGCTAGT CTGGAGTTGA TCAAGGAACC TGTCTCCACA 300

AAGTGTGACC ACATATTTTG CAAGTAAGTT TGAATGTGTT ATGTGGCTCC ATTATTAGCT 360

TTTGTTTTTG TCCTTCATAA CCCAGGAAAC ACCTAACTTT ATAGAAGCTT TACTTTCTTC 420

AATTAAGTGA GAACGAAAAT CCAACTCCAT TTCATTCTTT CTCAGAGAGT ATATAGTTAT 480

CAAAAGTTGG TTGTAATCAT AGTTCCTGGT AAAGTTTTGA CATATATTAT CTTTTTTTTT 540

TTTTGAGACA AGTCTCGCTC TGTCGCCCAG GCTGGAGTGC AGTGGCATGA GGCTTGCTCA 600

CTGCACCTCC GCCCCCGAGT TCAGCGACTC T 631

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 481 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TGAGATCTAG ACCACATGGT CAAAGAGATA GAATGTGAGC AATAAATGAA CCTTAAATTT 60

TTCAACAGCT ACTTTTTTTT TTTTTTTTTG AGACAGGGKC TTACTCTGTT GTCCCAGCTG 120

GAGTACAGWG TGCGATCATG AGGCTTACTG TTGCTTGACT CCTAGGCTCA AGCGATCCTA 180

TCACCTCAGT CTCCAAGTAG CTGGACTGTA AGTGCACACC ACCATATCCA GCTAAATTTT 240

GTGTTTTCTG TAGAGACGGG GTTTCGCCAT GTTTCCCAGG CTGGTCTTGA ACTTTGGGCT 300

TAACCCGTCT GCCCACCTAG GCATCCCAAA GTGCTAGGAT TACAGGTGTG AGTCATCATG 360

CCTGGCCAGT ATTTTAGTTA GCTCTGTCTT TTCAAGTCAT ATACAAGTTC ATTTTCTTTT 420

AAGTTTAGTT AACAACCTTA TATCATGTAT TCTTTTCTAG CATAAAGAAA GATTCGAGGC 480

C 481

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 522 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TGTGATCATA ACAGTAAGCC ATATGCATGT AAGTTCAGTT TTCATAGATC ATTGCTTATG 60

TAGTTTAGGT TTTTGCTTAT GCAGCATCCA AAAACAATTA GGAAACTATT GCTTGTAATT 120

CACCTGCCAT TACTTTTTAA ATGGCTCTTA AGGGCAGTTG TGAGATTATC TTTTCATGGC 180

TATTTGCCTT TTGAGTATTC TTTCTACAAA AGGAAGTAAA TTAAATTGTT CTTTCTTTCT 240

TTATAATTTA TAGATTTTGC ATGCTGAAAC TTCTCAACCA GAAGAAAGGG CCTTCACAGT 300

GTCCTTTATG TAAGAATGAT ATAACCAAAA GGTATATAAT TTGGTAATGA TGCTAGGTTG 360

GAAGCAACCA CAGTAGGAAA AAGTAGAAAT TATTTAATAA CATAGCGTTC CTATAAAACC 420

ATTCATCAGA AAAATTTATA AAAGAGTTTT TAGCACACAG TAAATTATTT CCAAAGTTAT 480

TTTCCTGAAA GTTTTATGGG CATCTGCCTT ATACAGGTAT TG 522

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 465 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

E6 cont'd

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GGTAGGCTTA AATGAATGAC AAAAAGTTAC TAAATCACTG CCATCACACG GTTTATACAG 60

ATGTCAATGA TGTATTGATT ATAGAGGTTT TCTACTGTTG CTGCATCTTA TTTTTATTTG 120

TTTACATGTC TTTTCTTATT TTAGTGTCCT TAAAAGGTTG ATAATCACTT GCTGAGTGTG 180

TTTCTCAAAC AATTTAATTT CAGGAGCCTA CAAGAAAGTA CGAGATTTAG TCAACTTGTT 240

GAAGAGCTAT TGAAAATCAT TTGTGCTTTT CAGCTTGACA CAGGTTTGGA GTGTAAGTGT 300

TGAATATCCC AAGAATGACA CTCAAGTGCT GTCCATGAAA ACTCAGGAAG TTTGCACAAT 360

TACTTTCTAT GACGTGGTGA TAAGACCTTT TAGTCTAGGT TAATTTTAGT TCTGTATCTG 420

TAATCTATTT TAAAAAATTA CTCCCACTGG TCTCACACCT TATTT 465

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 513 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

AAAAAATCAC AGGTAACCTT AATGCATTGT CTTAACACAA CAAAGAGCAT ACATAGGGTT 60

TCTCTTGGTT TCTTTGATTA TAATTCATAC ATTTTTCTCT AACTGCAAAC ATAATGTTTT 120

CCCTTGTATT TTACAGATGC AAACAGCTAT AATTTTGCAA AAAAGGAAAA TAACTCTCCT 180

GAACATCTAA AAGATGAAGT TTCTATCATC CAAAGTATGG GCTACAGAAA CCGTGCCAAA 240

AGACTTCTAC AGAGTGAACC CGAAAATCCT TCCTTGGTAA AACCATTTGT TTTCTTCTTC 300

TTCTTCTTCT TCTTTTCTTT TTTTTTTCTT TTTTTTTTTG AGATGGAGTC TTGCTCTGTG 360

GCCCAGGCTA GAAGCAGTCC TCCTGCCTTA GCCNCCTTAG TAGCTGGGAT TACAGGCACG 420

CGCACCATGC CAGGCTAATT TTTGTATTTT TAGTAGAGAC GGGGTTTCAT CATGTTGGCC 480

AGGCTGGTCT CGAACTCCTA ACCTCAGGTG ATC 513

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6769 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

EL cont'd

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATGATGGAGA TCTTAAAAAG TAATCATTCT GGGGCTGGGC GTAGTAGCTT GCACCTGTAA 60

TCCCAGCACT TCGGGAGGCT GAGGCAGGCA GATAATTTGA GGTCAGGAGT TTGAGACCAG 120

CCTGGCCAAC ATGGTGAAAC CCATCTCTAC TAAAAATACA AAAATTAGCT GGGTGTGGTG 180

GCACGTACCT GTAATCCCAG CTACTCGGGA GGCGGAGGCA CAAGAATTGC TTGAACCTAG 240

GACGCGGAGG TTGCAGCGAG CCAAGATCGC GCCACTGCAC TCCAGCCTGG GCCGTAGAGT 300

GAGACTCTGT CTCAAAAAAG AAAAAAAAGT AATTGTTCTA GCTGGGCGCA GTGGCTCTTG 360

CCTGTAATCC CAGCACTTTG GGAGGCCAAG GCGGGTGGAT CTCGAGTCCT AGAGTTCAAG 420

ACCAGCCTAG GCAATGTGGT GAAACCCCAT CGCTACAAAA AATACAAAAA TTAGCCAGGC 480

ATGGTGGCGT GCGCATGTAG TCCCAGCTCC TTGGGAGGCT GAGGTGGGAG GATCACTTGA 540

ACCCAGGAGA CAGAGGTTGC AGTGAACCGA GATCACGCCA CCACGCTCCA GCCTGGGCAA 600

CAGAACAAGA CTCTGTCTAA AAAAATACAA ATAAAATAAA AGTAGTTCTC ACAGTACCAG 660

CATTCATTTT TCAAAAGATA TAGAGCTAAA AAGGAAGGAA AAAAAAAGTA ATGTTGGGCT 720

E.C. cont'd

TTTAAATACT CGTTCCTATA CTAAATGTTC TTAGGAGTGC TGGGGTTTTA TTGTCATCAT 780

TTATCCTTTT TAAAAATGTT ATTGGCCAGG CACGGTGGCT CATGGCTGTA ATCCCAGCAC 840

TTTGGGAGGC CGAGGCAGGC AGATCACCTG AGGTCAGGAG TGTGAGACCA GCCTGGCCAA 900

CATGGCGAAA CCTGTCTCTA CTAAAAATAC AAAAATTAAC TAGGCGTGGT GGTGTACGCC 960

TGTAGTCCCA GCTACTCGGG AGGCTGAGGC AGGAGAATCA ACTGAACCAG GGAGGTGGAG 1020

GTTGCAGTGT GCCGAGATCA CGCCACTGCA CTCTAGCCTG GCAACAGAGC AAGATTCTGT 1080

CTCAAAAAAA AAAAACATAT ATACACATAT ATCCCAAAGT GCTGGGATTA CATATATATA 1140

TATATATATA TATTATATAT ATATATATAT ATATATGTGA TATATATGTG ATATATATAT 1200

AACATATATA TATGTAATAT ATATGTGATA TATATATAAT ATATATATGT AATATATATG 1260

TGATATATAT ATATACACAC ACACACACAT ATATATGTAT GTGTGTGTAC ACACACACAC 1320

ACAAATTAGC CAGGCATAGT TGCACACGCT TGGTAGACCC AGCTACTCAG GAGGCTGAGG 1380

GAGGAGAATC TCTTGAACTT AGGAGGCGGA GGTTGCAGTG AGCTGAGATT GCGCCACTGC 1440

ACTCCAGCCT GGGTGACAGA GCAGGACTCT GTACACCCCC CAAAACAAAA AAAAAAGTTA 1500

132

TCAGATGTGA TTGGAATGTA TATCAAGTAT CAGCTTCAAA ATATGCTATA TTAATACTTC 1560

AAAAATTACA CAAATAATAC ATAATCAGGT TTGAAAAATT TAAGACAACM SAARAAAAAA 1620

WYCMAATCAC AMATATCCCA CACATTTTAT TATTMCTMCT MCWATTATTT TGWAGAGMCT 1680

GGGTCTCACY CYKTTGCTWA TGCTGGTCTT TGAACYCCYK GCCYCAARCA RTCCTSCTCC 1740

ABCCTCCCAA RGTGCTGGGG ATWATAGGCA TGARCTAACC GCACCCAGCC CCAGACATTT 1800

TAGTGTGTAA ATTCCTGGGC ATTTTTTCAA GGCATCATAC ATGTTAGCTG ACTGATGATG 1860

GTCAATTTAT TTTGTCCATG GTGTCAAGTT TCTCTTCAGG AGGAAAAGCA CAGAACTGGC 1920

CAACAATTGC TTGACTGTTC TTTACCATAC TGTTTAGCAG GAAACCAGTC TCAGTGTCCA 1980

ACTCTCTAAC CTTGGAACTG TGAGAACTCT GAGGACAAAG CAGCGGATAC AACCTCAAAA 2040

GACGTCTGTC TACATTGAAT TGGGTAAGGG TCTCAGGTTT TTTAAGTATT TAATAATAAT 2100

TGCTGGATTC CTTATCTTAT AGTTTTGCCA AAAATCTTGG TCATAATTTG TATTTGTGGT 2160

AGGCAGCTTT GGGAAGTGAA TTTTATGAGC CCTATGGTGA GTTATAAAAA ATGTAAAAGA 2220

CGCAGTTCCC ACCTTGAAGA ATCTTACTTT AAAAAGGGAG CAAAAGAGGC CAGGCATGGT 2280

GGCTCACACC TGTAATCCCA GCACTTTGGG AGGCCAAAGT GGGTGGATCA CCTGAGGTCG 2340

GGAGTTCGAG ACCAGCCTAG CCAACATGGA GAAACTCTGT CTGTACCAAA AAATAAAAAA 2400

TTAGCCAGGT GTGGTGGCAC ATAACTGTAA TCCCAGCTAC TCGGGAGGCT GAGGCAGGAG 2460

E6 contd

AATCACTTGA ACCCGGGAGG TGGAGGTTGC GGTGAACCGA GATCGCACCA TTGCACTCCA 2520

GCCTGGGCAA AAATAGCGAA ACTCCATCTA AAAAAAAAAA AGAGAGCAAA AGAAAGAMTM 2580

TCTGGTTTTA AMTMTGTGTA AATATGTTTT TGGAAAGATG GAGAGTAGCA ATAAGAAAAA 2640

ACATGATGGA TTGCTACAGT ATTTAGTTCC AAGATAAATT GTACTAGATG AGGAAGCCTT 2700

TTAAGAAGAG CTGAATTGCC AGGCGCAGTG GCTCACGCCT GTAATCCCAG CACTTTGGGA 2760

GGCCGAGGTG GGCGGATCAC CTGAGGTCGG GAGTTCAAGA CCAGCCTGAC CAACATGGAG 2820

AAACCCCATC TCTACTAAAA AAAAAAAAAA AAAAATTAGC CGGGGTGGTG GCTTATGCCT 2880

GTAATCCCAG CTACTCAGGA GGCTGAGGCA GGAGAATCGC TTGAACCCAG GAAGCAGAGG 2940

TTGCAGTGAG CCAAGATCGC ACCATTGCAC TCCAGCCTAG GCAACAAGAG TGAAACTCCA 3000

TCTCAAAAAA AAAAAAAAAG AGCTGAATCT TGGCTGGGCA GGATGGCTCG TGCCTGTAAT 3060

CCTAACGCTT TGGAAGACCG AGGCAGAAGG ATTGGTTGAG TCCACGAGTT TAAGACCAGC 3120

CTGGCCAACA TAGGGGAACC CTGTCTCTAT TTTTAAAATA ATAATACATT TTTGGCCGGT 3180

GCGGTGGCTC ATGCCTGTAA TCCCAATACT TTGGGAGGCT GAGGCAGGTA GATCACCTGA 3240

GGTCAGAGTT CGAGACCAGC CTGGATAACC TGGTGAAACC CCTCTTTACT AAAAATACAA 3300

AAAAAAAAAA AAATTAGCTG GGTGTGGTAG CACATGCTTG TAATCCCAGC TACTTGGGAG 3360

GCTGAGGCAG GAGAATCGCT TGAACCAGGG AGGCGGAGGT TACAATGAGC CAACACTACA 3420

CCACTGCACT CCAGCCTGGG CAATAGAGTG AGACTGCATC TCAAAAAAAT AATAATTTTT 3480

AAAAATAATA AATTTTTTTA AGCTTATAAA AAGAAAAGTT GAGGCCAGCA TAGTAGCTCA 3540

CATCTGTAAT CTCAGCAGTG GCAGAGGATT GCTTGAAGCC AGGAGTTTGA GACCAGCCTG 3600

GGCAACATAG CAAGACCTCA TCTCTACAAA AAAATTTCTT TTTTAAATTA GCTGGGTGTG 3660

GTGGTGTGCA TCTGTAGTCC CAGCTACTCA GGAGGCAGAG GTGAGTGGAT ACATTGAACC 3720

CAGGAGTTTG AGGCTGTAGT GAGCTATGAT CATGCCACTG CACTCCAACC TGGGTGACAG 3780

AGCAAGACCT CCAAAAAAAA AAAAAAAAGA GCTGCTGAGC TCAGAATTCA AACTGGGCTC 3840

TCAAATTGGA TTTTCTTTTA GAATATATTT ATAATTAAAA AGGATAGCCA TCTTTTGAGC 3900

TCCCAGGCAC CACCATCTAT TTATCATAAC ACTTACTGTT TTCCCCCCTT ATGATCATAA 3960

ATTCCTAGAC AACAGGCATT GTAAAAATAG TTATAGTAGT TGATATTTAG GAGCACTTAA 4020

CTATATTCCA GGCACTATTG TGCTTTTCTT GTATAACTCA TTAGATGCTT GTCAGACCTC 4080

TGAGATTGTT CCTATTATAC TTATTTTACA GATGAGAAAA TTAAGGCACA GAGAAGTTAT 4140

GAAATTTTTC CAAGGTATTA AACCTAGTAA GTGGCTGAGC CATGATTCAA ACCTAGGAAG 4200

TTAGATGTCA GAGCCTGTGC TTTTTTTTTG TTTTTGTTTT TGTTTTCAGT AGAAACGGGG 4260

GTCTCACTTT GTTGGCCAGG CTGGTCTTGA ACTCCTAACC TCAAATAATC CACCCATCTC 4320

GGCCTCCTCA AGTGCTGGGA TTACAGGTGA GAGCCACTGT GCCTGGCGAA GCCCATGCCT 4380

TTAACCACTT CTCTGTATTA CATACTAGCT TAACTAGCAT TGTACCTGCC ACAGTAGATG 4440

CTCAGTAAAT ATTTCTAGTT GAATATCTGT TTTTCAACAA GTACATTTTT TTAACCCTTT 4500

TAATTAAGAA AACTTTTATT GATTTATTTT TTGGGGGGAA ATTTTTTAGG ATCTGATTCT 4560

TCTGAAGATA CCGTTAATAA GGCAACTTAT TGCAGGTGAG TCAAAGAGAA CCTTTGTCTA 4620

TGAAGCTGGT ATTTTCCTAT TTAGTTAATA TTAAGGATTG ATGTTTCTCT CTTTTTAAAA 4680

ATATTTTAAC TTTTATTTTA GGTTCAGGGA TGTATGTGCA GTTTGTTATA TAGGTAAACA 4740

CACGACTTGG GATTTGGTGT ATAGATTTTT TTCATCATCC GGGTACTAAG CATACCCCAC 4800

AGTTTTTTGT TTGCTTTCTT TCTGAATTTC TCCCTCTTCC CACCTTCCTC CCTCAAGTAG 4860

EL cont'd

124

GCTGGTGTTT CTCCAGACTA GAATCATGGT ATTGGAAGAA ACCTTAGAGA TCATCTAGTT 4920

TAGTTCTCTC ATTTTATAGT GGAGGAAATA CCCTTTTTGT TTGTTGGATT TAGTTATTAG 4980

CACTGTCCAA AGGAATTTAG GATAACAGTA GAACTCTGCA CATGCTTGCT TCTAGCAGAT 5040

TGTTCTCTAA GTTCCTCATA TACAGTAATA TTGACACAGC AGTAATTGTG ACTGATGAAA 5100

ATGTTCAAGG ACTTCATTTT CAACTCTTTC TTTCCTCTGT TCCTTATTTC CACATATCTC 5160

TCAAGCTTTG TCTGTATGTT ATATAATAAA CTACAAGCAA CCCCAACTAT GTTACCTACC 5220

TTCCTTAGGA ATTATTGCTT GACCCAGGTT TTTTTTTTTT TTTTTTTGGA GACGGGGTCT 5280

TGCCCTGTTG CCAGGATGGA GTGTAGTGGC GCCATCTCGG CTCACTGCAA TCTCCAACTC 5340

CCTGGTTCAA GCGATTCTCC TGTCTCAATC TCACGAGTAG CTGGGACTAC AGGTATACAC 5400

CACCACGCCC GGTTAATTGA CCATTCCATT TCTTTCTTTC TCTCTTTTTT TTTTTTTTTT 5460

TTGAGACAGA GTCTTGCTCT GTTGCCCAGG CTGGAGTACA GAGGTGTGAT CTCACCTCTC 5520

CGCAACGTCT GCCTCCCAGG TTGAAGCCAT ACTCCTGCCT CAGCCTCTCT AGTAGCTGGG 5580

ACTACAGGCG CGCGCCACCA CACCCGGCTA ATTTTTGTAT TTTTAGTAGA GATGGGGTTT 5640

CACCATGTTG GCCAGGCTGG TCTTGAACTC ATGACCTCAA GTGGTCCACC CGCCTCAGCC 5700

TCCCAAAGTG CTGGAATTAC AGGCTTGAGC CACCGTGCCC AGCAACCATT TCATTTCAAC 5760

E6 cont'd

TAGAAGTTTC TAAAGGAGAG AGCAGCTTTC ACTAACTAAA TAAGATTGGT CAGCTTTCTG 5820

TAATCGAAAG AGCTAAAATG TTTGATCTTG GTCATTTGAC AGTTCTGCAT ACATGTAACT 5880

AGTGTTTCTT ATTAGGACTC TGTCTTTTCC CTATAGTGTG GGAGATCAAG AATTGTTACA 5940

AATCACCCCT CAAGGAACCA GGGATGAAAT CAGTTTGGAT TCTGCAAAAA AGGGTAATGG 6000

CAAAGTTTGC CAACTTAACA GGCACTGAAA AGAGAGTGGG TAGATACAGT ACTGTAATTA 6060

GATTATTCTG AAGACCATTT GGGACCTTTA CAACCCACAA AATCTCTTGG CAGAGTTAGA 6120

GTATCATTCT CTGTCAAATG TCGTGGTATG GTCTGATAGA TTTAAATGGT ACTAGACTAA 6180

TGTACCTATA ATAAGACCTT CTTGTAACTG ATTGTTGCCC TTTCGCTTTT TTTTTTGTTT 6240

GTTTGTTTGT TTTTTTTTGA GATGGGGTCT CACTCTGTTG CCCAGGCTGG AGTGCAGTGA 6300

TGCAATCTTG GCTCACTGCA ACCTCCACCT CCAAAGGCTC AAGCTATCCT CCCACTTCAG 6360

CCTCCTGAGT AGCTGGGACT ACAGGCGCAT GCCACCACAC CCGGTTAATT TTTTGTGGTT 6420

TTATAGAGAT GGGGTTTCAC CATGTTACCG AGGCTGGTCT CAAACTCCTG GACTCAAGCA 6480

GTCTGCCCAC TTCAGCCTCC CAAAGTGCTG CAGTTACAGG CTTGAGCCAC TGTGCCTGGC 6540

CTGCCCTTTA CTTTTAATTG GTGTATTTGT GTTTCATCTT TTACCTACTG GTTTTTAAAT 6600

ATAGGGAGTG GTAAGTCTGT AGATAGAACA GAGTATTAAG TAGACTTAAT GGCCAGTAAT 6660

CTTTAGAGTA CATCAGAACC AGTTTTCTGA TGGCCAATCT GCTTTTAATT CACTCTTAGA 6720

CGTTAGAGAA ATAGGTGTGG TTTCTGCATA GGGAAAATTC TGAAATTAA 6769

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4249 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GATCCTAAGT GGAAATAATC TAGGTAAATA GGAATTAAAT GAAAGAGTAT GAGCTACATC 60

TTCAGTATAC TTGGTAGTTT ATGAGGTTAG TTTCTCTAAT ATAGCCAGTT GGTTGATTTC 120

CACCTCCAAG GTGTATGAAG TATGTATTTT TTTAATGACA ATTCAGTTTT TGAGTACCTT 180

GTTATTTTTG TATATTTTCA GCTGCTTGTG AATTTTCTGA GACGGATGTA ACAAATACTG 240

AACATCATCA ACCCAGTAAT AATGATTTGA ACACCACTGA GAAGCGTGCA GCTGAGAGGC 300

ATCCAGAAAA GTATCAGGGT AGTTCTGTTT CAAACTTGCA TGTGGAGCCA TGTGGCACAA 360

ATACTCATGC CAGCTCATTA CAGCATGAGA ACAGCAGTTT ATTACTCACT AAAGACAGAA 420

TGAATGTAGA AAAGGCTGAA TTCTGTAATA AAAGCAAACA GCCTGGCTTA GCAAGGAGCC 480

AACATAACAG ATGGGCTGGA AGTAAGGAAA CATGTAATGA TAGGCGGACT CCCAGCACAG 540

AAAAAAAGGT AGATCTGAAT GCTGATCCCC TGTGTGAGAG AAAAGAATGG AATAAGCAGA 600

AACTGCCATG CTCAGAGAAT CCTAGAGATA CTGAAGATGT TCCTTGGATA ACACTAAATA 660

GCAGCATTCA GAAAGTTAAT GAGTGGTTTT CCAGAAGTGA TGAACTGTTA GGTTCTGATG 720

ACTCACATGA TGGGGAGTCT GAATCAAATG CCAAAGTAGC TGATGTATTG GACGTTCTAA 780

E6 cont'd

ATGAGGTAGA TGAATATTCT GGTTCTTCAG AGAAAATAGA CTTACTGGCC AGTGATCCTC 840

ATGAGGCTTT AATATGTAAA AGTGAAAGAG TTCACTCCAA ATCAGTAGAG AGTAATATTG 900

AAGGCCAAAT ATTTGGGAAA ACCTATCGGA AGAAGGCAAG CCTCCCCAAC TTAAGCCATG 960

TAACTGAAAA TCTAATTATA GGAGCATTTG TTACTGAGCC ACAGATAATA CAAGAGCGTC 1020

CCCTCACAAA TAAATTAAAG CGTAAAAGGA GACCTACATC AGGCCTTCAT CCTGAGGATT 1080

TTATCAAGAA AGCAGATTTG GCAGTTCAAA AGACTCCTGA AATGATAAAT CAGGGAACTA 1140

ACCAAACGGA GCAGAATGGT CAAGTGATGA ATATTACTAA TAGTGGTCAT GAGAATAAAA 1200

CAAAAGGTGA TTCTATTCAG AATGAGAAAA ATCCTAACCC AATAGAATCA CTCGAAAAAG 1260

AATCTGCTTT CAAAACGAAA GCTGAACCTA TAAGCAGCAG TATAAGCAAT ATGGAACTCG 1320

AATTAAATAT CCACAATTCA AAAGCACCTA AAAAGAATAG GCTGAGGAGG AAGTCTTCTA 1380

CCAGGCATAT TCATGCGCTT GAACTAGTAG TCAGTAGAAA TCTAAGCCCA CCTAATTGTA 1440

CTGAATTGCA AATTGATAGT TGTTCTAGCA GTGAAGAGAT AAAGAAAAAA AAGTACAACC 1500

AAATGCCAGT CAGGCACAGC AGAAACCTAC AACTCATGGA AGGTAAAGAA CCTGCAACTG 1560

GAGCCAAGAA GAGTAACAAG CCAAATGAAC AGACAAGTAA AAGACATGAC AGCGATACTT 1620

TCCCAGAGCT GAAGTTAACA AATGCACCTG GTTCTTTTAC TAAGTGTTCA AATACCAGTG 1680

AACTTAAAGA ATTTGTCAAT CCTAGCCTTC CAAGAGAAGA AAAAGAAGAG AACTAGAAAC 1740

AGTTAAAGTG TCTAATAATG CTGAAGACCC CAAAGATCTC ATGTTAAGTG GAGAAAGGGT 1800

TTTGCAAACT GAAAGATCTG TAGAGAGTAG CAGTATTTCA TTGGTACCTG GTACTGATTA 1860

TGGCACTCAG GAAAGTATCT CGTTACTGGA AGTTAGCACT CTAGGGAAGG CAAAAACAGA 1920

ACCAAATAAA TGTGTGAGTC AGTGTGCAGC ATTTGAAAAC CCCAAGGGAC TAATTCATGG 1980

TTGTTCCAAA GATAATAGAA ATGACACAGA AGGCTTTAAG TATCCATTGG GACATGAAGT 2040

TAACCACAGT CGGGAAACAA GCATAGAAAT GGAAGAAAGT GAACTTGATG CTCAGTATTT 2100

GCAGAATACA TTCAAGGTTT CAAAGCGCCA GTCATTTGCT CCGTTTTCAA ATCCAGGAAA 2160

TGCAGAAGAG GAATGTGCAA CATTCTCTGC CCACTCTGGG TCCTTAAAGA AACAAAGTCC 2220

AAAAGTCACT TTTGAATGTG AACAAAAGGA AGAAAATCAA GGAAAGAATG AGTCTAATAT 2280

CAAGCCTGTA CAGACAGTTA ATATCACTGC AGGCTTTCCT GTGGTTGGTC AGAAAGATAA 2340

GCCAGTTGAT AATGCCAAAT GTAGTATCAA AGGAGGCTCT AGGTTTTGTC TATCATCTCA 2400

GTTCAGAGGC AACGAAACTG GACTCATTAC TCCAAATAAA CATGGACTTT TACAAAACCC 2460

E6 cont'd

GeneDX 1009, pg. 142

ATATCGTATA CCACCACTTT TTCCCATCAA GTCATTTGTT AAAACTAAAT GTAAGAAAAA 2520

TCTGCTAGAG GAAAACTTTG AGGAACATTC AATGTCACCT GAAAGAGAAA TGGGAAATGA 2580

GAACATTCCA AGTACAGTGA GCACAATTAG CCGTAATAAC ATTAGAGAAA ATGTTTTTAA 2640

AGAAGCCAGC TCAAGCAATA TTAATGAAGT AGGTTCCAGT ACTAATGAAG TGGGCTCCAG 2700

TATTAATGAA ATAGGTTCCA GTGATGAAAA CATTCAAGCA GAACTAGGTA GAAACAGAGG 2760

GCCAAAATTG AATGCTATGC TTAGATTAGG GGTTTTGCAA CCTGAGGTCT ATAAACAAAG 2820

TCTTCCTGGA AGTAATTGTA AGCATCCTGA AATAAAAAAG CAAGAATATG AAGAAGTAGT 2880

TCAGACTGTT AATACAGATT TCTCTCCATA TCTGATTTCA GATAACTTAG AACAGCCTAT 2940

GGGAAGTAGT CATGCATCTC AGGTTTGTTC TGAGACACCT GATGACCTGT TAGATGATGG 3000

TGAAATAAAG GAAGATACTA GTTTTGCTGA AAATGACATT AAGGAAAGTT CTGCTGTTTT 3060

TAGCAAAAGC GTCCAGAAAG GAGAGCTTAG CAGGAGTCCT AGCCCTTTCA CCCATACACA 3120

TTTGGCTCAG GGTTACCGAA GAGGGGCCAA GAAATTAGAG TCCTCAGAAG AGAACTTATC 3180

TAGTGAGGAT GAAGAGCTTC CCTGCTTCCA ACACTTGTTA TTTGGTAAAG TAAACAATAT 3240

ACCTTCTCAG TCTACTAGGC ATAGCACCGT TGCTACCGAG TGTCTGTCTA AGAACACAGA 3300

GGAGAATTTA TTATCATTGA AGAATAGCTT AAATGACTGC AGTAACCAGG TAATATTGGC 3360

AAAGGCATCT CAGGAACATC ACCTTAGTGA GGAAACAAAA TGTTCTGCTA GCTTGTTTTC 3420

TTCACAGTGC AGTGAATTGG AAGACTTGAC TGCAAATACA AACACCCAGG ATCCTTTCTT 3480

GATTGGTTCT TCCAAACAAA TGAGGCATCA GTCTGAAAGC CAGGGAGTTG GTCTGAGTGA 3540

CAAGGAATTG GTTTCAGATG ATGAAGAAAG AGGAACGGGC TTGGAAGAAA ATAATCAAGA 3600

AGAGCAAAGC ATGGATTCAA ACTTAGGTAT TGGAACCAGG TTTTTGTGTT TGCCCCAGTC 3660

TATTTATAGA AGTGAGCTAA ATGTTTATGC TTTTGGGGAG CACATTTTAC AAATTTCCAA 3720

GTATAGTTAA AGGAACTGCT TCTTAAACTT GAAACATGTT CCTCCTAAGG TGCTTTTCAT 3780

AGAAAAAAGT CCTTCACACA GCTAGGACGT CATCTTTGAC TGAATGAGCT TTAACATCCT 3840

AATTACTGGT GGACTTACTT CTGGTTTCAT TTTATAAAGC AAATCCCGGT GTCCCAAAGC 3900

AAGGAATTTA ATCATTTTGT GTGACATGAA AGTAAATCCA GTCCTGCCAA TGAGAAGAAA 3960

AAGACACAGC AAGTTGCAGC GTTTATAGTC TGCTTTTACA TCTGAACCTC TGTTTTTGTT 4020

ATTTAAGGTG AAGCAGCATC TGGGTGTGAG AGTGAAACAA GCGTCTCTGA AGACTGCTCA 4080

GGGCTATCCT CTCAGAGTGA CATTTTAACC ACTCAGGTAA AAAGCGTGTG TGTGTGTGCA 4140

fig 2 cont'd

CATGCGTGTG TGTGGTGTCC TTTGCATTCA GTAGTATGTA TCCCACATTC TTAGGTTTGC 4200

TGACATCATC TCTTTGAATT AATGGCACAA TTGTTTGTGG TTCATTGTC 4249

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 710 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

NGNGAATGTA ATCCTAATAT TTCNCNCCNA CTTAAAAGAA TACCACTCCA ANGGCATCNC 60

AATACATCAA TCAATTGGGG AATTGGGATT TTCCCTCNCT AACATCANTG GAATAATTTC 120

ATGGCATTAA TTGCATGAAT GTGGTTAGAT TAAAAGGTGT TCATGCTAGA ACTTGTAGTT 180

E6 cont'd

CCATACTAGG TGATTTCAAT TCCTGTGCTA AAATTAATTT GTATGATATA TTNTCATTTA 240

ATGGAAAGCT TCTCAAAGTA TTTCATTTTC TTGGTACCAT TTATCGTTTT TGAAGCAGAG 300

GGATACCATG CAACATAACC TGATAAAGCT CCAGCAGGAA ATGGCTGAAC TAGAAGCTGT 360

GTTAGAACAG CATGGGAGCC AGCCTTCTAA CAGCTACCCT TCCATCATAA GTGACTCTTC 420

TGCCCTTGAG GACCTGCGAA ATCCAGAACA AAGCACATCA GAAAAAGGTG TGTATTGTTG 480

GCCAAACACT GATATCTTAA GCAAAATTCT TTCCTTCCCC TTTATCTCCT TCTGAAGAGT 540

AAGGACCTAG CTCCAACATT TTATGATCCT TGCTCAGCAC ATGGGTAATT ATGGAGCCTT 600

GGTTCTTGTC CCTGCTCACA ACTAATATAC CAGTCAGAGG GACCCAAGGC AGTCATTCAT 660

GTTGTCATCT GAGATACCTA CAACAAGTAG ATGCTATGGG GAGCCCATGG 710

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 473 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCATTGGTGC TAGCATCTGT CTGTTGCATT GCTTGTGTTT ATAAAATTCT GCCTGATATA 60

CTTGTTAAAA ACCAATTTGT GTATCATAGA TTGATGCTTT TGAAAAAAAT CAGTATTCTA 120

ACCTGAATTA TCACTATCAG AACAAAGCAG TAAAGTAGAT TTGTTTTCTC ATTCCATTTA 180

AAGCAGTATT AACTTCACAG AAAAGTAGTG AATACCCTAT AAGCCAGAAT CCAGAAGGCC 240

TTTCTGCTGA CAAGTTTGAG GTGTCTGCAG ATAGTTCTAC CAGTAAAAAT AAAGAACCAG 300

GAGTGGAAAG GTAAGAAACA TCAATGTAAA GATGCTGTGG TATCTGACAT CTTTATTTAT 360

ATTGAACTCT GATTGTTAAT TTTTTTCACC ATACTTTCTC CAGTTTTTTT GCATACAGGC 420

ATTTATACAC TTTTATTGCT CTAGGATACT TCTTTTGTTT AATCCTATAT AGG 473

E6 cont'd

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 421 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GGATAAGNTC AAGAGATATT TTGATAGGTG ATGCAGTGAT NAATTGNGAA AATTTNCTGC 60

CTGCTTTTAA TCTTCCCCCG TTCTTTCTTC CTNCCTCCCT CCCTTCCTNC CTCCCGTCCT 120

TNCCTTTCCT TTCCCTCCCT TCCNCCTTCT TTCCNTCTNT CTTTCCTTTC TTTCCTGTCT 180

GeneDX 1009, pg. 145

ACCTTTCTTT CCTTCCTCCC TTCCTTTTCT TTTCTTTCTT TCCTTTCCTT TTCTTTCCTT 240

TCTTTCCTTT CCTTTCTTTC TTGACAGAGT CTTGCTCTGT CACTCAGGCT GGAGTGCAGT 300

GGCGTGATCT CGNCTCACTG CAACCTCTGT CTCCCAGGTT CAAGCAATTT TCCTGCCTCA 360

GCCTCCCGAG TAGCTGAGAT TACAGGCGCC AGCCACCACA CCCAGCTACT GACCTGCTTT 420

T 421

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 997 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

E6 cont'd

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

AAACAGCTGG GAGATATGGT GCCTCAGACC AACCCCATGT TATATGTCAA CCCTGACATA 60

TTGGCAGGCA ACATGAATCC AGACTTCTAG GCTGTCATGC GGGCTCTTTT TTGCCAGTCA 120

TTTCTGATCT CTCTGACATG AGCTGTTTCA TTTATGCTTT GGCTGCCCAG CAAGTATGAT 180

TTGTCCTTTC ACAATTGGTG GCGATGGTTT TCTCCTTCCA TTTATCTTTC TAGGTCATCC 240

CCTTCTAAAT GCCCATCATT AGATGATAGG TGGTACATGC ACAGTTGCTC TGGGAGTCTT 300

CAGAATAGAA ACTACCCATC TCAAGAGGAG CTCATTAAGG TTGTTGATGT GGAGGAGCAA 360

CAGCTGGAAG AGTCTGGGCC ACACGATTTG ACGGAAACAT CTTACTTGCC AAGGCAAGAT 420

CTAGGTAATA TTTCATCTGC TGTATTGGAA CAAACACTYT GATTTTACTC TGAATCCTAC 480

ATAAAGATAT TCTGGTTAAC CAACTTTTAG ATGTACTAGT CTATCATGGA CACTTTTGTT 540

ATACTTAATT AAGCCCACTT TAGAAAAATA GCTCAAGTGT TAATCAAGGT TTACTTGAAA 600

ATTATTGAAA CTGTTAATCC ATCTATATTT TAATTAATGG TTTAACTAAT GATTTTGAGG 660

ATGWGGGAGT CKTGGTGTAC TCTAMATGTA TTATTTCAGG CCAGGCATAG TGGCTCACGC 720

CTGGTAATCC CAGTAYYCMR GAGCCCGAGG CAGGTGGAGC CAGCTGAGGT CAGGAGTTCA 780

AGACCTGTCT TGGCCAACAT GGGNGAAACC CTGTCTTCTT CTTAAAAAAN ACAAAAAAAA 840

TTAACTGGGT TGTGCTTAGG TGNATGCCCC GNATCCTAGT TNTTCTTGNG GGTTGAGGGA 900

GGAGATCACN TTGGACCCCG GAGGGGNGGG TGGGGGNGAG CAGGNCAAAA CACNGACCCA 960

GCTGGGGTGG AAGGGAAGCC CACTCNAAAA AANNTTN 997

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 639 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

EG contid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TTTTTAGGAA ACAAGCTACT TTGGATTTCC ACCAACACCT GTATTCATGT ACCCATTTTT 60

CTCTTAACCT AACTTTATTG GTCTTTTTAA TTCTTAACAG AGACCAGAAC TTTGTAATTC 120

AACATTCATC GTTGTGTAAA TTAAACTTCT CCCATTCCTT TCAGAGGGAA CCCCTTACCT 180

GGAATCTGGA ATCAGCCTCT TCTCTGATGA CCCTGAATCT GATCCTTCTG AAGACAGAGC 240

CCCAGAGTCA GCTCGTGTTG GCAACATACC ATCTTCAACC TCTGCATTGA AAGTTCCCCA 300

ATTGAAAGTT GCAGAATCTG CCCAGAGTCC AGCTGCTGCT CATACTACTG ATACTGCTGG 360

GTATAATGCA ATGGAAGAAA GTGTGAGCAG GGAGAAGCCA GAATTGACAG CTTCAACAGA 420

AAGGGTCAAC AAAAGAATGT CCATGGTGGT GTCTGGCCTG ACCCCAGAAG AATTTGTGAG 480

TGTATCCATA TGTATCTCCC TAATGACTAA GACTTAACAA CATTCTGGAA AGAGTTTTAT 540

GTAGGTATTG TCAATTAATA ACCTAGAGGA AGAAATCTAG AAAACAATCA CAGTTCTGTG 600

TAATTTAATT TCGATTACTA ATTTCTGAAA ATTTAGAAY 639

172

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 922 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

NCCCNNCCCC CNAATCTGAA ATGGGGGTAA CCCCCCCCCA ACCGANACNT GGGTNGCNTA 60

GAGANTTTAA TGGCCCNTTC TGAGGNACAN AAGCTTAAGC CAGGNGACGT GGANCNATGN 120

GTTGTTTNTT GTTTGGTTAC CTCCAGCCTG GGTGACAGAG CAAGACTCTG TCTAAAAAAA 180

AAAAAAAAAA AAATCGACTT TAAATAGTTC CAGGACACGT GTAGAACGTG CAGGATTGCT 240

ACGTAGGTAA ACATATGCCA TGGTGGGATA ACTAGTATTC TGAGCTGTGT GCTAGAGGTA 300

ACTCATGATA ATGGAATATT TGATTTAATT TCAGATGCTC GTGTACAAGT TTGCCAGAAA 360

ACACCACATC ACTTTAACTA ATCTAATTAC TGAAGAGACT ACTCATGTTG TTATGAAAAC 420

AGGTATACCA AGAACCTTTA CAGAATACCT TGCATCTGCT GCATAAAACC ACATGAGGCG 480

AGGCACGGTG GCGCATGCCT GTAATCGCAG CACTTTGGGA GGCCGAGGCG GGCAGATCAC 540

GAGATTAGGA GATCGAGACC ATCCTGGCCA GCATGGTGAA ACCCCGTCTC TACTANNAAA 600

TGGNAAAATT ANCTGGGTGT GGTCGCGTGC NCCTGTAGTC CCAGCTACTC GTGAGGCTGA 660

GGCAGGAGAA TCACTTGAAC CGGGGAAATG GAGGTTTCAG TGAGCAGAGA TCATNCCCCT 720

NCATTCCAGC CTGGCGACAG AGCAAGGCTC CGTCNCCNAA AAAATAAAAA AAAACGTGAA 780

CAAATAAGAA TATTTGTTGA GCATAGCATG GATGATAGTC TTCTAATAGT CAATCAATTA 840

CTTTATGAAA GACAAATAAT AGTTTTGCTG CTTCCTTACC TCCTTTTGTT TTGGGTTAAG 900

ATTTGGAGTG TGGGCCAGGC AC 922

E6 cont'd

143

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 867 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GATCTATAGC TAGCCTTGGC GTCTAGAAGA TGGGTGTTGA GAAGAGGGAG TGGAAAGATA 60

TTTCCTCTGG TCTTAACTTC ATATCAGCCT CCCCTAGACT TCCAAATATC CATACCTGCT 120

GGTTATAATT AGTGGTGTTT TCAGCCTCTG ATTCTGTCAC CAGGGGTTTT AGAATCATAA 180

ATCCAGATTG ATCTTGGGAG TGTAAAAAAC TGAGGCTCTT TAGCTTCTTA GGACAGCACT 240

TCCTGATTTT GTTTTCAACT TCTAATCCTT TGAGTGTTTT TCATTCTGCA GATGCTGAGT 300

TTGTGTGTGA ACGGACACTG AAATATTTTC TAGGAATTGC GGGAGGAAAA TGGGTAGTTA 360

GCTATTTCTG TAAGTATAAT ACTATTTCTC CCCTCCTCCC TTTAACACCT CAGAATTGCA 420

TTTTTACACC TAACATTTAA CACCTAAGGT TTTTGCTGAT GCTGAGTCTG AGTTACCAAA 480

AGGTCTTTAA ATTGTAATAC TAAACTACTT TTATCTTTAA TATCACTTTG TTCAAGATAA 540

GCTGGTGATG CTGGGAAAAT GGGTCTCTTT TATAACTAAT AGGACCTAAT CTGCTCCTAG 600

CAATGTTAGC ATATGAGCTA GGGATTTATT TAATAGTCGG CAGGAATCCA TGTGCARCAG 660

NCAAACTTAT AATGTTTAAA TTAAACATCA ACTCTGTCTC CAGAAGGAAA CTGCTGCTAC 720

AAGCCTTATT AAAGGGCTGT GGCTTTAGAG GGAAGGACCT CTCCTCTGTC ATTCTTCCTG 780

TGCTCTTTTG TGAATCGCTG ACCTCTCTAT CTCCGTGAAA AGAGCACGTT CTTCTGCTGT 840

ATGTAACCTG TCTTTTCTAT GATCTCT 867

E6 cont'd

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 561 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

NAAAAACGGG GNNGGGANTG GGCCTTAAAN CCAAAGGGCN AACTCCCCAA CCATTNAAAA 60

ANTGACNGGG GATTATTAAA ANCGGCGGGA AACATTTCAC NGCCCAACTA ATATTGTTAA 120

ATTAAAACCA CCACCNCTGC NCCAAGGAGG GAAACTGCTG CTACAAGCCT TATTAAAGGG 180

CTGTGGCTTT AGAGGGAAGG ACCTCTCCTC TGTCATTCTT CCTGTGCTCT TTTGTGAATC 240

GCTGACCTCT CTATGTCCGT GAAAAGAGCA CGTTCTTCGT CTGTATGTAA CCTGTCTTTT 300

CTATGATCTC TTTAGGGGTG ACCCAGTCTA TTAAAGAAAG AAAAATGCTG AATGAGGTAA 360

GTACTTGATG TTACAAACTA ACCAGAGATA TTCATTCAGT CATATAGTTA AAAATGTATT 420

TGCTTCCTTC CATCAATGCA CCACTTTCCT TAACAATGCA CAAATTTTCC ATGATAATGA 480

GGATCATCAA GAATTATGCA GGCCTGCACT GTGGCTCATA CCTATAATCC CAGCGCTTTG 540

GGAGGCTGAG GCGCTTGGAT C 561

E6 cont'd

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 567 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

145

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

AATTTTTTGT ATTTTTAGTA GAGATGAGGT TCACCATGTT GGTCTAGATC TGGTGTCGAA 60

CGTCCTGACC TCAAGTGATC TGCCAGCCTC AGTCTCCCAA AGTGCTAGGA TTACAGGGGT 120

GAGCCACTGC GCCTGGCCTG AATGCCTAAA ATATGACGTG TCTGCTCCAC TTCCATTGAA 180

GGAAGCTTCT CTTTCTCTTA TCCTGATGGG TTGTGTTTGG TTTCTTTCAG CATGATTTTG 240

AAGTCAGAGG AGATGTGGTC AATGGAAGAA ACCACCAAGG TCCAAAGCGA GCAAGAGAAT 300

CCCAGGACAG AAAGGTAAAG CTCCCTCCCT CAAGTTGACA AAAATCTCAC CCCACCACTC 360

TGTATTCCAC TCCCCTTTGC AGAGATGGGC CGCTTCATTT TGTAAGACTT ATTACATACA 420

TACACAGTGC TAGATACTTT CACACAGGTT CTTTTTTCAC TCTTCCATCC CAACCACATA 480

AATAAGTATT GTCTCTACTT TATGAATGAT AAAACTAAGA GATTTAGAGA GGCTGTGTAA 540

TTTGGATTCC CGTCTCGGGT TCAGATC 567

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 633 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

E6 cont'd

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TTGGCCTGAT TGGTGACAAA AGTGAGATGC TCAGTCCTTG AATGACAAAG AATGCCTGTA 60

GAGTTGCAGG TCCAACTACA TATGCACTTC AAGAAGATCT TCTGAAATCT AGTAGTGTTC 120

TGGACATTGG ACTGCTTGTC CCTGGGAAGT AGCAGCAGAA ATGATCGGTG GTGAACAGAA 180

GAAAAAGAAA AGCTCTTCCT TTTTGAAAGT CTGTTTTTTG AATAAAAGCC AATATTCTTT 240

146

TATAACTAGA TTTTCCTTCT CTCCATTCCC CTGTCCCTCT CTCTTCCTCT CTTCTTCCAG 300

ATCTTCAGGG GGCTAGAAAT CTGTTGCTAT GGGCCCTTCA CCAACATGCC CACAGGTAAG 360

AGCCTGGGAG AACCCCAGAG TTCCAGCACC AGCCTTTGTC TTACATAGTG GAGTATTATA 420

AGCAAGGTCC CACGATGGGG GTTCCTCAGA TTGCTGAAAT GTTCTAGAGG CTATTCTATT 480

TCTCTACCAC TCTCCAAACA AAACAGCACC TAAATGTTAT CCTATGGCAA AAAAAAACTA 540

TACCTTGTCC CCCTTCTCAA GAGCATGAAG GTGGTTAATA GTTAGGATTC AGTATGTTAT 600

GTGTTCAGAT GGCGTTGAGC TGCTGTTAGT GCC 633

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 470 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

E6 contd

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TTTGAGAGAC TATCAAACCT TATACCAAGT GGCCTTATGG AGACTGATAA CCAGAGTACA 60

TGGCATATCA GTGGCAAATT GACTTAAAAT CCATACCCCT ACTATTTTAA GACCATTGTC 120

CTTTGGAGCA GAGAGACAGA CTCTCCCATT GAGAGGTCTT GCTATAAGCC TTCATCCGGA 180

GAGTGTAGGG TAGAGGGCCT GGGTTAAGTA TGCAGATTAC TGCAGTGATT TTACATGTAA 240

ATGTCCATTT TAGATCAACT GGAATGGATG GTACAGCTGT GTGGTGCTTC TGTGGTGAAG 300

GAGCTTTCAT CATTCACCCT TGGCACAGTA AGTATTGGGT GCCCTGTCAG TGTGGGAGGA 360

CACAATATTC TCTCCTGTGA GCAAGACTGG CACCTGTCAG TCCCTATGGA TGCCCCTACT 420

GTAGCCTCAG AAGTCTTCTC TGCCCACATA CCTGTGCCAA AAGACTCCAT 470

147

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 517 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGTGGTACGT GTCTGTAGTT CCAGCTACTT GGGAGGCTGA GATGGAAGGA TTGCTTGAGC 60

CCAGGAGGCA GAGGTGGNAN NTTACGCTGA GATCACACCA CTGCACTCCA GCCTGGGTGA 120

CAGAGCAAGA CCCTGTCTCA AAAACAAACA AAAAAAATGA TGAAGTGACA GTTCCAGTAG 180

TCCTACTTTG ACACTTTGAA TGCTCTTTCC TTCCTGGGGA TCCAGGGTGT CCACCCAATT 240

GTGGTTGTGC AGCCAGATGC CTGGACAGAG GACAATGGCT TCCATGGTAA GGTGCCTCGC 300

ATGTACCTGT GCTATTAGTG GGGTCCTTGT GCATGGGTTT GGTTTATCAC TCATTACCTG 360

GTGCTTGAGT AGCACAGTTC TTGGCACATT TTTAAATATT TGTTGAATGA ATGGCTAAAA 420

TGTCTTTTTG ATGTTTTTAT TGTTATTTGT TTTATATTGT AAAAGTAATA CATGAACTGT 480

TTCCATGGGG TGGGAGTAAG ATATGAATGT TCATCAC 517

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 434 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

EC control

GeneDX 1009, pg. 153

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CAGTAATCCT NAGAACTCAT ACGACCGGGC CCCTGGAGTC GNTGNTTNGA GCCTAGTCCN 60

GGAGAATGAA TTGACACTAA TCTCTGCTTG TGTTCTCTGT CTCCAGCAAT TGGGCAGATG 120

TGTGAGGCAC CTGTGGTGAC CCGAGAGTGG GTGTTGGACA GTGTAGCACT CTACCAGTGC 180

CAGGAGCTGG ACACCTACCT GATACCCCAG ATCCCCCACA GCCACTACTG ACTGCAGCCA 240

GCCACAGGTA CAGAGCCACA GGACCCCAAG AATGAGCTTA CAAAGTGGCC TTTCCAGGCC 300

CTGGGAGCTC CTCTCACTCT TCAGTCCTTC TACTGTCCTG GCTACTAAAT ATTTTATGTA 360

CATCAGCCTG AAAAGGACTT CTGGCTATGC AAGGGTCCCT TAAAGATTTT CTGCTTGAAG 420

TCTCCCTTGG AAAT 434

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

E6 cont'd

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GATAAATTAA AACTGCGACT GCGCGGCGTG 30

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

149

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GTAGTAGAGT CCCGGGAAAG GGACAGGGGG 30

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

ATATATATAT GTTTTTCTAA TGTGTTAAAG 30

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GTAAGTCAGC ACAAGAGTGT ATTAATTTGG 30

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

GeneDX 1009, pg. 155

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TTTCTTTTTC TCCCCCCCCT ACCCTGCTAG 30

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

E6 cont'd

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GTAAGTTTGA ATGTGTTATG TGGCTCCATT 30

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AGCTACTTTT TTTTTTTTTT TTTGAGACAG 30

GeneDX 1009, pg. 156

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GTAAGTGCAC ACCACCATAT CCAGCTAAAT 30

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

E6 cont'd

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

AATTGTTCTT TCTTTCTTTA TAATTTATAG 30

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

152

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GTATATAATT TGGTAATGAT GCTAGGTTGG 30

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GAGTGTGTTT CTCAAACAAT TTAATTTCAG 30

(2) INFORMATION FOR SEQ ID NO:46:

E6 cont'd

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GTAAGTGTTG AATATCCCAA GAATGACACT 30

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

153

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

AAACATAATG TTTTCCCTTG TATTTTACAG 30

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

E6 contid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GTAAAACCAT TTGTTTTCTT CTTCTTCTTC 30

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

TGCTTGACTG TTCTTTACCA TACTGTTTAG 30

154

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GTAAGGGTCT CAGGTTTTTT AAGTATTTAA 30

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

E6 cont'd

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

TGATTTATTT TTTGGGGGGA AATTTTTTAG 30

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GTGAGTCAAA GAGAACCTTT GTCTATGAAG 30

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

TCTTATTAGG ACTCTGTCTT TTCCCTATAG 30

(2) INFORMATION FOR SEQ ID NO:54:

F6 cont'd

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GTAATGGCAA AGTTTGCCAA CTTAACAGGC 30

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

156

GeneDX 1009, pg. 161

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GAGTACCTTG TTATTTTTGT ATATTTTCAG 30

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

E6 cont'd

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GTATTGGAAC CAGGTTTTTG TGTTTGCCCC 30

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

ACATCTGAAC CTCTGTTTTT GTTATTTAAG 30

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

AGGTAAAAAG CGTGTGTGTG TGTGCACATG 30

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

E6 Cont'd

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CATTTTCTTG GTACCATTTA TCGTTTTTGA 30

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

157

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

GTGTGTATTG TTGGCCAAAC ACTGATATCT 30

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

AGTAGATTTG TTTTCTCATT CCATTTAAAG 30

(2) INFORMATION FOR SEQ ID NO:62:

E6 contd

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GTAAGAAACA TCAATGTAAA GATGCTGTGG 30

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

159

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

ATGGTTTTCT CCTTCCATTT ATCTTTCTAG 30

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

E6 cont'd

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GTAATATTTC ATCTGCTGTA TTGGAACAAA 30

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

TGTAAATTAA ACTTCTCCCA TTCCTTTCAG 30

160

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GTGAGTGTAT CCATATGTAT CTCCCTAATG 30

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

F6 cont'd

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

ATGATAATGG AATATTTGAT TTAATTTCAG 30

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GTATACCAAG AACCTTTACA GAATACCTTG 30

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CTAATCCTTT GAGTGTTTTT CATTCTGCAG 30

(2) INFORMATION FOR SEQ ID NO:70:

E6 cont'd

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

GTAAGTATAA TACTATTTCT CCCCTCCTCC 30

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

TGTAACCTGT CTTTTCTATG ATCTCTTTAG 30

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

E6 cont'd

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

GTAAGTACTT GATGTTACAA ACTAACCAGA 30

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

TCCTGATGGG TTGTGTTTGG TTTCTTTCAG 30

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

GTAAAGCTCC CTCCCTCAAG TTGACAAAAA 30

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

#16 Cont'd

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

CTGTCCCTCT CTCTTCCTCT CTTCTTCCAG 30

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

GeneDX 1009, pg. 169

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

GTAAGAGCCT GGGAGAACCC CAGAGTTCCA 30

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

AGTGATTTTA CATGTAAATG TCCATTTTAG 30

E6 Contid

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

GTAAGTATTG GGTGCCCTGT CAGTGTGGGA 30

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

TTGAATGCTC TTTCCTTCCT GGGGATCCAG 30

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

E6 contd

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

GTAAGGTGCC TCGCATGTAC CTGTGCTATT 30

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

CTAATCTCTG CTTGTGTTCT CTGTCTCCAG 30

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

```
Cys Pro Ile Cys Leu Glu Leu Ile Lys Glu Pro Val Ser Thr Lys Cys
1               5                  10                  15

Asp His Ile Phe Cys Lys Phe Cys Met Leu Lys Leu Leu Asn Gln Lys
            20                  25                  30

Lys Gly Pro Ser Gln Cys Pro Leu Cys Lys
        35                  40
```

E6 contd

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

```
Cys Pro Ile Cys Leu Glu Leu Leu Lys Glu Pro Val Ser Ala Asp Cys
1               5                  10                  15

Asn His Ser Phe Cys Arg Ala Cys Ile Thr Leu Asn Tyr Glu Ser Asn
            20                  25                  30

Arg Asn Thr Asp Gly Lys Gly Asn Cys Pro Val Cys Arg
        35                  40                  45
```

167

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Cys Pro Ile Cys Leu Asp Met Leu Lys Asn Thr Met Thr Thr Lys Glu
1 5 10 15

Cys Leu His Arg Phe Cys Ser Asp Cys Ile Val Thr Ala Leu Arg Ser
20 25 30

Gly Asn Lys Glu Cys Pro Thr Cys Arg
35 40

(2) INFORMATION FOR SEQ ID NO:85:

E6 contd

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

Cys Pro Val Cys Leu Gln Tyr Phe Ala Glu Pro Met Met Leu Asp Cys
1 5 10 15

Gly His Asn Ile Cys Cys Ala Cys Leu Ala Arg Cys Trp Gly Thr Ala
20 25 30

Cys Thr Asn Val Ser Cys Pro Gln Cys Arg
35 40

WHAT IS CLAIMED IS:

1. A method for diagnosing a predisposition for breast and ovarian cancer in a human subject which comprises detecting a germline alteration in the sequence of the BRCA1 gene or its expression products in said subject by comparing the germline sequence of the BRCA1 gene or its expression products in a tissue sample from said subject with the germline sequence of the wild-type BRCA1 gene or its expression products, wherein an alteration in the germline sequence of the subject indicates a predisposition to said cancer.

2. The method of claim 1 wherein the wild-type BRCA1 gene has the sequence set forth in SEQ ID NO:1.

3. The method of claim 1 wherein said expression product is mRNA of the BRCA1 gene.

4. The method of claim 3 wherein the alteration of the mRNA is detected by hybridization of the mRNA of said sample with a BRCA1 gene probe.

5. The method of claim 1 wherein said expression product is a BRCA1 polypeptide encoded by the BRCA1 gene.

6. The method of claim 1 wherein an alteration is detected in the regulatory regions of the BRCA1 gene.

7. The method of claim 1 wherein the alteration in the germline sequence is detected by observing shifts in electrophoretic mobility of single-stranded DNA on non-denaturing polyacrylamide gels.

8. The method of claim 1 wherein the alteration in the germline sequence is detected by hybridization of a BRCA1 gene probe to genomic DNA isolated from said tissue.

GeneDX 1009, pg. 174

9. The method of claim 1 wherein the alteration in the germline sequence is detected by hybridization with an allele-specific probe.

10. The method of claim 1 wherein the alteration in the germline sequence is detected by amplifying all or part of the BRCA1 gene from said sample to produce an amplified sequence and sequencing the amplified sequence.

11. The method of claim 1 wherein the alteration in the germline sequence is detected by amplifying all or part of the BRCA1 gene for a specific BRCA1 mutant allele.

12. The method of claim 1 wherein the alteration in the germline sequence is detected by molecularly cloning all or part of the BRCA1 gene from said sample to produce a cloned sequence and sequencing the cloned sequence.

13. The method of claim 1 wherein the alteration in the germline sequence is detected by identifying a mismatch between molecules (1) BRCA1 gene genomic DNA or BRCA1 mRNA isolated from said sample, and (2) a nucleic acid probe complementary to the human wild-type BRCA1 gene DNA, when molecules (1) and (2) are hybridized to each other to form a duplex.

14. The method of claim 1 wherein the alteration in the germline sequence is detected by amplification of BRCA1 gene sequences in said sample and hybridization of the amplified sequences to nucleic acid probes which comprise wild-type BRCA1 gene sequences.

15. The method of claim 1 wherein the alteration in the germline sequence is detected by amplification of BRCA1 gene sequences in said tissue and hybridization of the amplified sequences to nucleic acid probes which comprise mutant BRCA1 gene sequences.

16. The method of claim 1 wherein the alteration in the germline sequence is detected by screening for a deletion mutation.

17. The method of claim 1 wherein the alteration in the germline sequence is detected by screening for a point mutation.

18. The method of claim 1 wherein the alteration in the germline sequence is detected by screening for an insertion mutation.

19. The method of claim 1 wherein the alteration in the germline sequence is detected by *in situ* hybridization of the BRCA1 gene with nucleic acid probes which comprise the BRCA1 gene.

20. The method of claim 5 wherein the alteration in the germline sequence is detected by immunoblotting.

21. The method of claim 5 wherein the alteration in the germline sequence is detected by immunocytochemistry.

22. The method of claim 5 wherein the alteration in the germline sequence is detected by assaying for binding interactions between BRCA1 gene protein isolated from said tissue and its ligand.

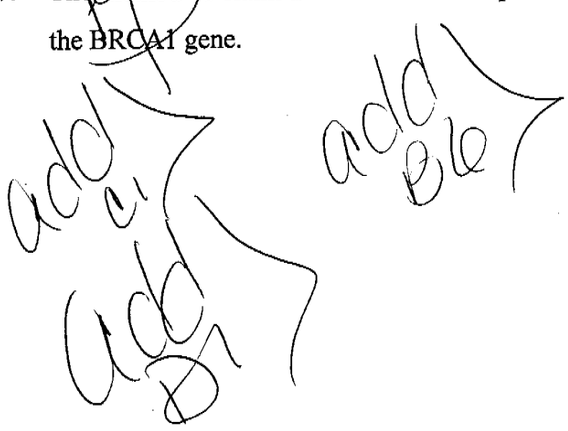
23. The method of claim 5 wherein the alteration in the germline sequence is detected by assaying for the inhibition of biochemical activity of its binding partner.

24. A method for confirming the lack of a predisposition for cancer at the BRCA1 locus in a human subject which comprises detecting a germline wild-type sequence of the BRCA1 gene or its expression products in said subject by comparing the germline sequence of the BRCA1 gene or its expression products in a tissue sample from said subject with the germline sequence of the wild-type BRCA1 gene or its expression products, wherein the presence of the wild-type germline sequence indicates no predisposition to cancer at the BRCA1 locus.

25. The method of claim 24 wherein the wild-type BRCA1 gene has the sequence set forth in SEQ ID NO:1.

26. The method of claim 24 wherein said expression product is mRNA of the BRCA1 gene.

27. The method of claim 24 wherein said expression product is a BRCA1 polypeptide encoded by the BRCA1 gene.



MAIL ROOM
DEC
1
1995

DECLARATION AND POWER OF ATTORNEY

As below-named inventor, I hereby declare that my residence, post office address and citizenship are as stated below next to my name; I believe I am an original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE, the specification of which was filed on 7 June 1995 under Serial No. 08/488,011.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) on which priority is claimed:

NONE

I hereby claim the benefit under Title 35, United States Code, §120 of any United States applications that are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which occurred between the filing date of the prior applications and the national or PCT international filing date of this application:

U.S. Serial No. 08/409,305, filed on 24 March 1995

U.S. Serial No. 08/348,824, filed on 24 November 1994

U.S. Serial No. 08/308,104, filed on 16 September 1994

U.S. Serial No. 08/300,266, filed on 2 September 1994

U.S. Serial No. 08/289,221, filed on 12 August 1994

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

I hereby appoint Jeffrey L. Ihnen, Registration No. 28,957; E. Brendan Magrab, Registration No. 36,205; and Stephen A. Saxe, Registration No. 38,609; all of Venable, Baetjer, Howard & Civiletti, whose address is 1201 New York Avenue, N.W., Suite 1000, Washington, D.C. 20005, telephone number (202) 962-4800, my attorneys and/or agents with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

3

Please direct all correspondence and telephone calls to Venable, Baetjer, Howard & Civiletti at the above address and telephone number.

Full Name of Sole or First Inventor	Mark H. SKOLNICK 1-00
Inventor's Signature	Mark Skolnick
Date of Signature	Oct 9 1995
Residence Address	460 Northmount Way Salt Lake City, Utah 84103
Post Office Address	same UT
Country of Citizenship	United States

#11 see paper

Full Name of Second Inventor	David E. GOLDGAR 2-00
Inventor's Signature	David E Goldgar
Date of Signature	October 9 1995
Residence Address	2707 East Kelly Lane Salt Lake City, Utah 84117 UT
Post Office Address	same
Country of Citizenship	United States

Full Name of Third Inventor	Yoshio MIKI 3-00
Inventor's Signature	_____
Date of Signature	_____
Residence Address	40 South 900 East Salt Lake City, Utah 84102 UT
Post Office Address	same
Country of Citizenship	Japan

see paper #11

Full Name of Fourthh Inventor	Jeff SWENSON 4-00
Inventor's Signature	Jeff Swenson
Date of Signature	10-10-95
Residence Address	745 West Creek Bend, #2807 Salt Lake City, Utah 84117 UT
Post Office Address	same
Country of Citizenship	United States

Full Name of Fifth Inventor	Alexander KAMB 5-00
Inventor's Signature	
Date of Signature	10/9/95
Residence Address	1103 East 600 South Salt Lake City, Utah 84102 UT
Post Office Address	same
Country of Citizenship	United States

Full Name of Sixth Inventor	Keith D. HARSHMAN 6-00
Inventor's Signature	
Date of Signature	10.9.95
Residence Address	159 North F Street Salt Lake City, Utah 841032 UT
Post Office Address	same
Country of Citizenship	United States

Full Name of Seventh Inventor	Donna M. SHATTUCK-EIDENS 7-00
Inventor's Signature	
Date of Signature	10.9.95
Residence Address	1874 Texas Street Salt Lake City, Utah 84108 UT
Post Office Address	same
Country of Citizenship	United States

Full Name of Eighth Inventor	Sean V. TAVTIGIAN 8-00
Inventor's Signature	
Date of Signature	10/9/95
Residence Address	557 East 1st Ave., #3 Salt Lake City, Utah 841023 UT
Post Office Address	same
Country of Citizenship	United States

Full Name of Ninth Inventor	Roger W. WISEMAN 9-00
Inventor's Signature	
Date of Signature	10/20/95
Residence Address	~~515 Woodwinds Drive~~ 16 Stream View Court Rev Durham, North Carolina 27713
Post Office Address	same NC
Country of Citizenship	United States

Full Name of Tenth Inventor	P. Andrew FUTREAL
Inventor's Signature	_____
Date of Signature	_____
Residence Address	220 East Markham Ave. Durham, North Carolina 27701
Post Office Address	same
Country of Citizenship	United States

Full Name of Ninth Inventor	Roger W. WISEMAN
Inventor's Signature	_____
Date of Signature	_____
Residence Address	515 Woodwinds Drive Durham, North Carolina 27713
Post Office Address	same
Country of Citizenship	United States

Full Name of Tenth Inventor	P. Andrew FUTREAL 10-00
Inventor's Signature	P Andrew Futreal
Date of Signature	11/14/95
Residence Address	~~220 East Markham Ave.~~ ~~Durham, North Carolina 27701~~
Post Office Address	same 311 S. LaSalle 28B Durham NC 27705
Country of Citizenship	United States NC PAF

PRINT OF DR[illegible]
AS ORIGINA[illegible]LED

08/488011

Map of the early onset breast and ovarian cancer region (BRCA1)

FIGURE 1

PRINT OF DRAWINGS
AS ORIGINALLY FILED

08/488011

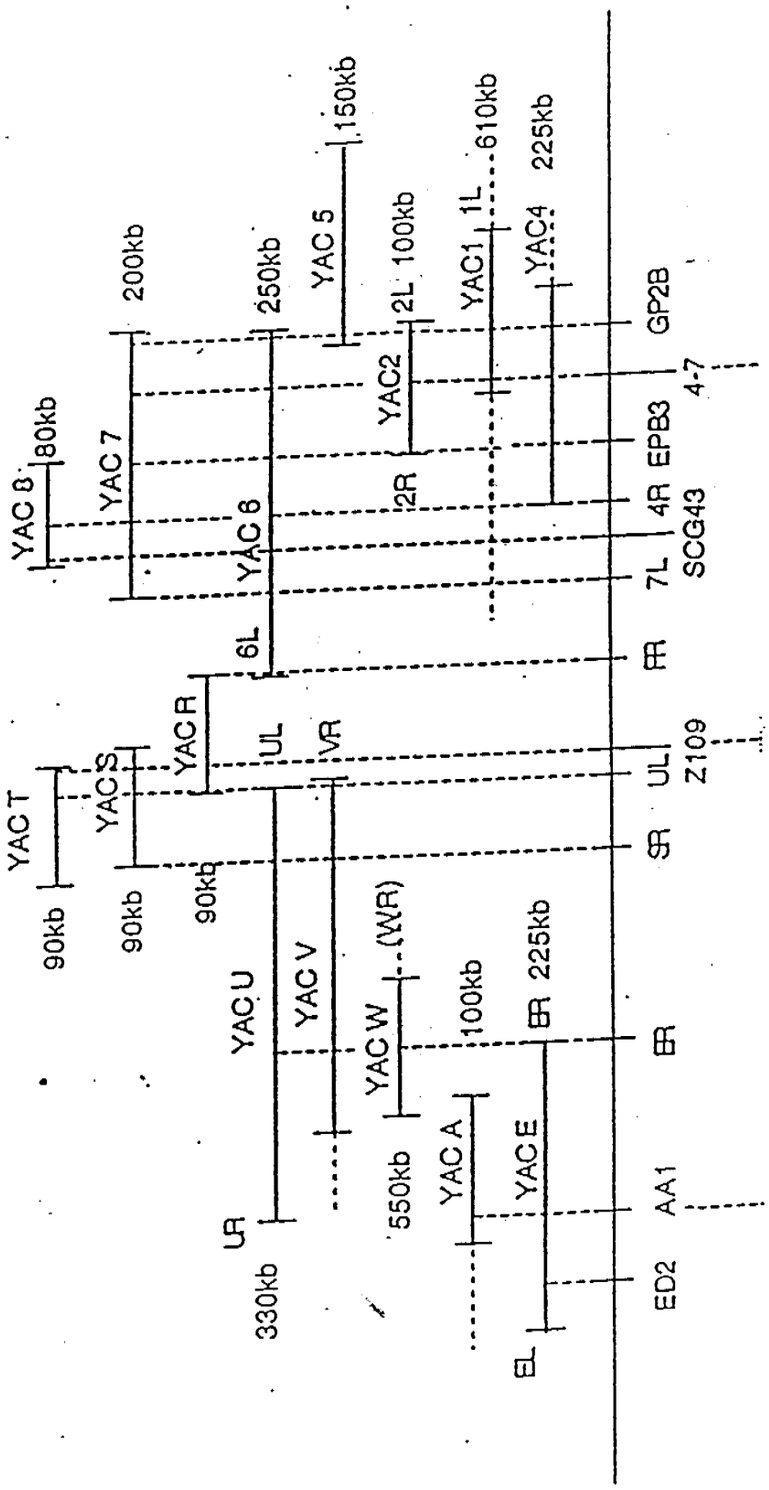


FIGURE 2

Schematic Map of STS', P1s, and Bac's in the Region

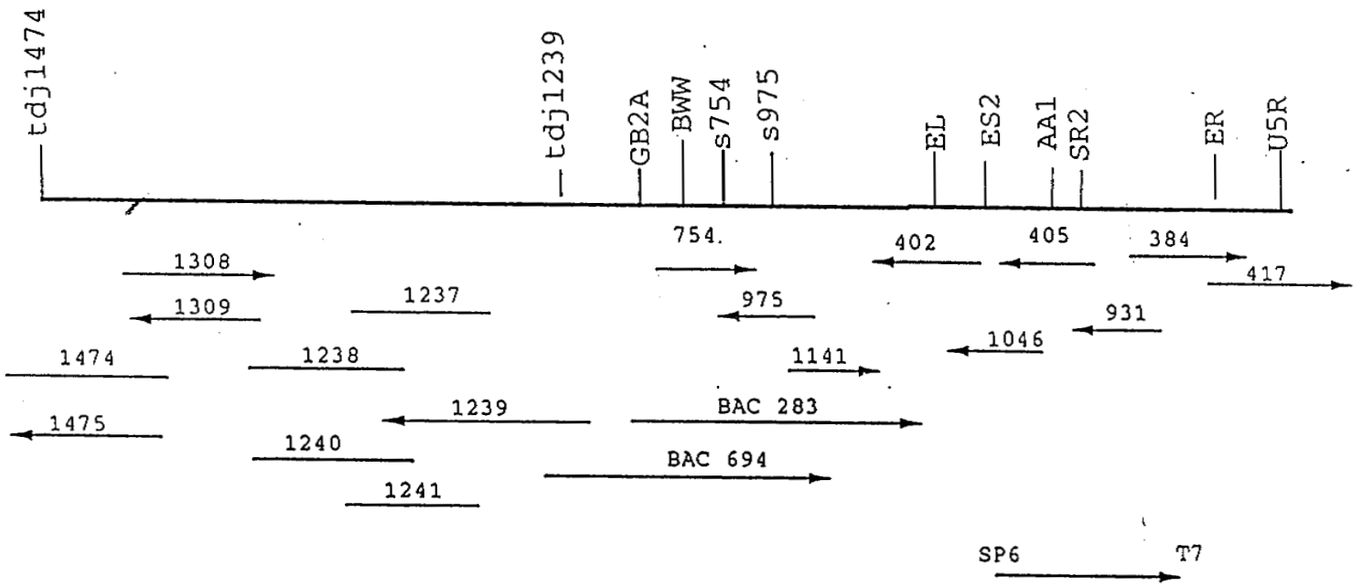


FIGURE 3

PRINT OF DR
AS ORIGINA
FILED
INGS

PRINT OF DRAWINGS
AS ORIGINAL FILED

08/488011

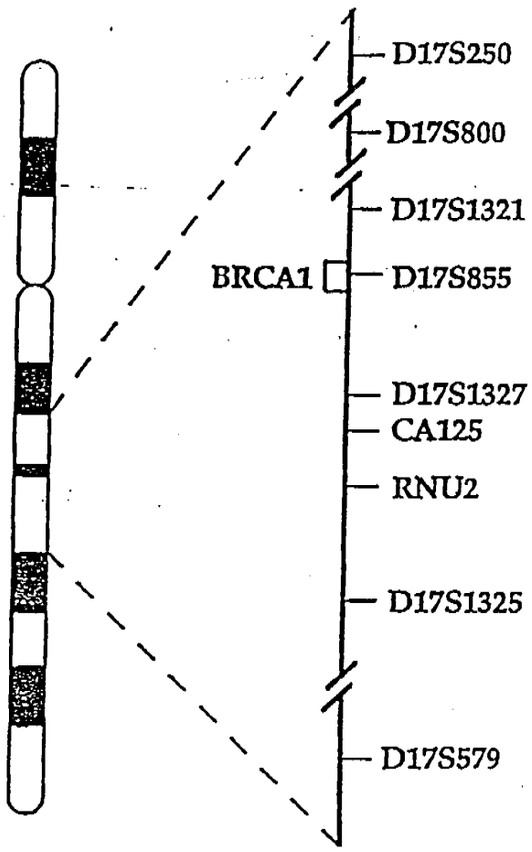


FIGURE 4

PRINT OF DRAWINGS
AS ORIGINALLY FILED

SEQ ID NO:		
82	BRCA1	CPICLELIKEPVSTK-CDHIFCKFCMLKLLNQKK---GPSQCPLCK
83	RPT1	CPICLELLKEPVSAD-CNHSFCRACITLNYESNRNTDGKGNCPVCR
84	RIN1	CPICLDMLKNTMTTKECLHRFCSDCIVTALRS-----GNKECPTCR
85	RFP1	CPVCLQYFAEPMMLD-CGHNICCACLARCWGTAC---TNVSCPQCR
	C3HC4 motif	C--C----------C-H--C--C----------------C--C

FIGURE 5

08/48801

GeneDX 1009, pg. 188

PRINT OF DRAWINGS
AS ORIGINAL FILED

08/488011

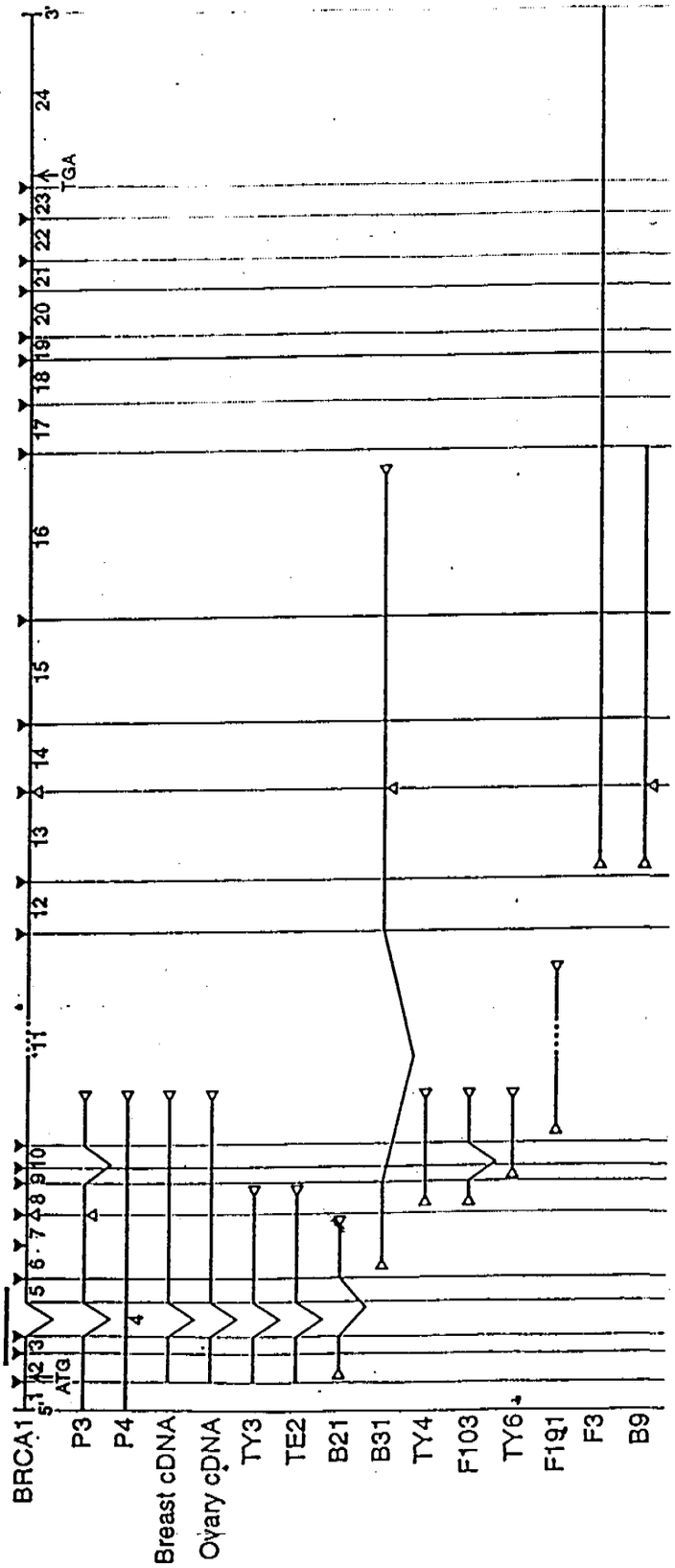


FIGURE 6

PRINT OF DRAWINGS
AS ORIGINAL FILED

OR 4880

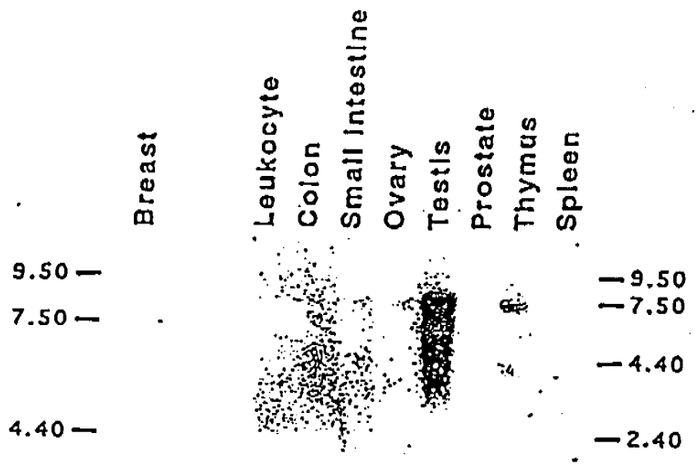


FIGURE 7

PRINT OF DRAWINGS
AS ORIGINAL FILED

FIGURE 8

PRINT OF DR WINGS
AS ORIGIN / FILED

08/488011

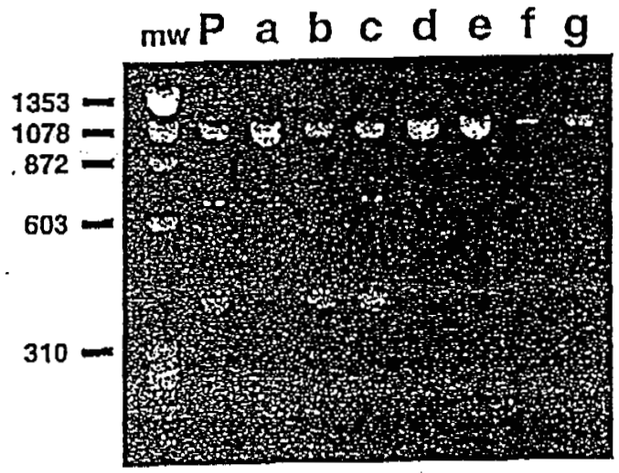


FIGURE 9A

GeneDX 1009, pg. 192

PRINT OF DRAWINGS
AS ORIGINAL FILED

08/488011

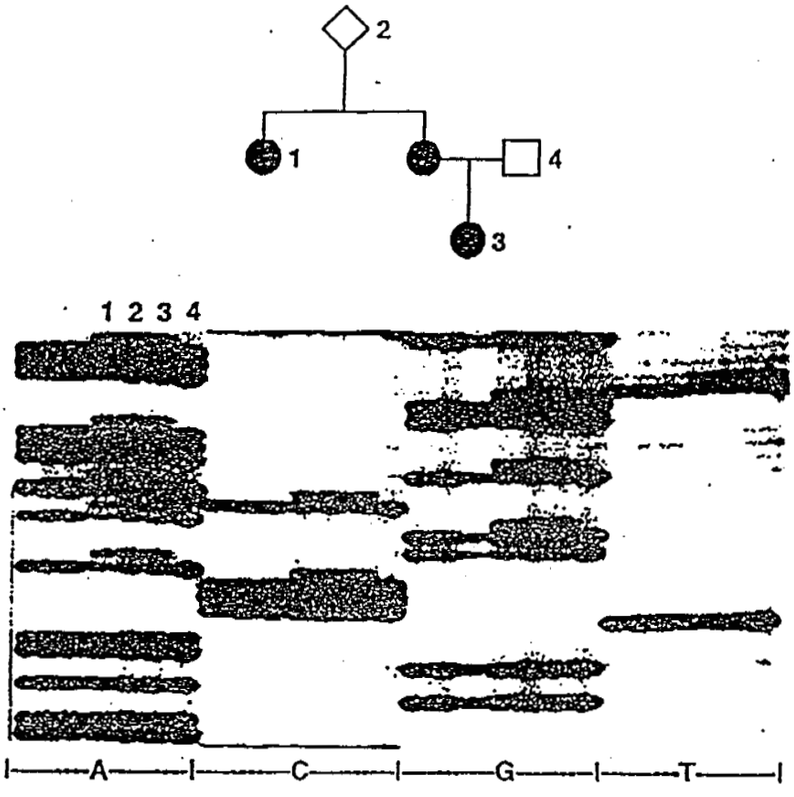


FIGURE 9B

PRINT OF DRAWINGS
AS ORIGINAL FILED

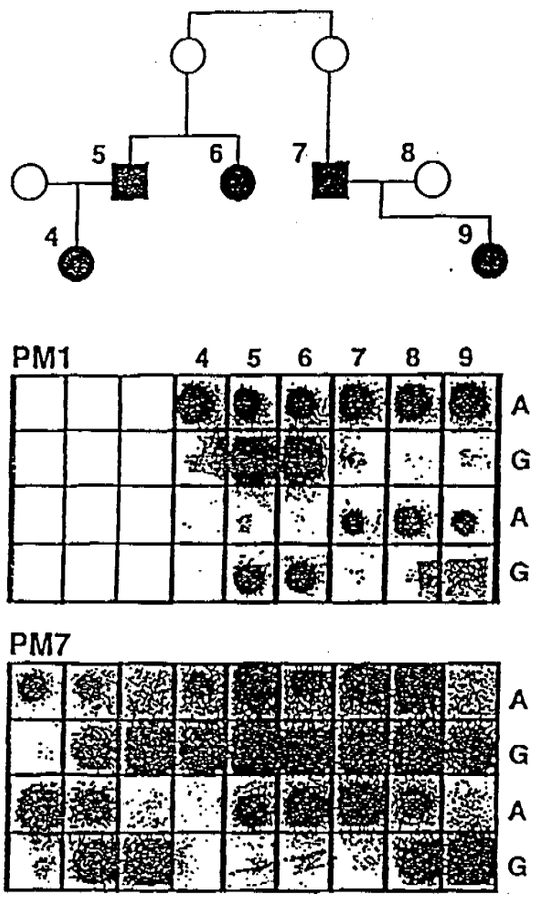


FIGURE 9C

```
   1  gaggctagagggcaggcactttatggcaaactcaggtagaattcttcctcttccgtctct
  61  ttccttttacgtcatcggggagactgggtggcaatcgcagcccgagagacgcatggctct
 121  ttctgccctccatcctctgatgtaccttgatttcgtattctgagaggctgctgcttagcg
 181  gtagccccttggtttccgtggcaacggaaaagcgcgggaattacagataaattaaaactg
 241  cgactgcgcggcgtgAGCTCGCTGAGACTTCCTGGACCCCGCACCAGGCTGTGGGGTTTC
 301  TCAGATAACTGGGCCCCTGCGCTCAGGAGGCCTTCACCCTCTGCTCTGGGTAAAGgtagt
 361  agagtcccgggaaagggacaggggcccaagtgatgctctggggtactggcgtgggagag
 421  tggatttccgaagctgacagatgggtattctttgacgggggggtaggggcggaacctgaga
 481  ggcgtaaggcgttgtgaaccctggggagggggcagtttgtaggtcgcgagggaagcgct
 541  gaggatcaggaaggggcactgagtgtccgtgggggaatcctcgtgataggaactggaat
 601  atgccttgaggggacactatgtctttaaaaacgtcggctggtcatgaggtcaggagttc
 661  cagaccagcctgaccaacgtggtgaaactccgtctctactaaaaatacnaaaattagccg
 721  ggcgtggtgccgctccagctactcaggaggctgaggcaggagaatcgctagaacccggga
 781  ggcggaggttgcagtgagccgagatcgcgccattgcactccagcctgggcgacagagcga
 841  gactgtctcaaaacaaaacaaaacaaaacaaaacaaaacaaaaaacaccggctggtatgtatgag
 901  aggatgggaccttgtggaagaagaggtgccaggaatatgtctgggaaggggaggagacag
 961  gattttgtgggagggagaacttaagaactggatccatttgcgccattgagaaagcgcaag
1021  agggaagtagaggagcgtcagtagtaacagatgctgccggcagggatgtgcttgaggagg
1081  atccagagatgagagcaggtcactgggaaaggttaggggcggggaggccttgattggtgt
1141  tggtttggtcgttgttgattttggttttatgcaagaaaaagaaaacaaccagaaacattg
1201  gagaaagctaaggctaccaccacctacccggtcagtcactcctctgtagctttctctttc
1261  ttggagaaaggaaaagacccaaggggttggcagcgatatgtgaaaaaattcagaatttat
1321  gttgtctaattacaaaaagcaacttctagaatctttaaaaataaaggacgttgtcattag
1381  ttcttctggtttgtattattctaaaaccttccaaatcttcaaatttacttttatttttaaaa
1441  tgataaaatgaagttgtcattttataaaccttttaaaaagatatatatatatgtttttct
1501  aatgtgttaaagTTCATTGGAACAGAAAGAAATGGATTTATCTGCTCTTCGCGTTGAAGA
1561  AGTACAAAATGTCATTAATGCTATGCAGAAAATCTTAGAGTGTCCCATCTGgtaagtcag
1621  cacaagagtgtattaatttgggattcctatgattatctcctatgcaaatgaacagaattg
1681  accttacatactagggaagaaaagacatgtctagtaagattaggctattgtaattgctga
1741  ttttcttaactgaagaactttaaaaatatagaaaatgattccttgttctccatccactct
1801  gcctctcccactcctctccttttcaacacaatcctgtggtccgggaaagacagggctctg
1861  tcttgattggttctgcactgggcaggatctgttagatactgcatttgctttctccagctc
1921  taaavvvvvvvvvvvvvaaatgctgatgatagtatagagtattgaagggatcaatataat
1981  tctgttttgatatctgaaagctcactgaaggtaaggatcgtattctctgctgtattctca
2041  gttcctgacacagcagacatttaataaatattgaacgaacttgaggccttatgttgactc
2101  agtcataacagctcaaagttgaacttattcactaagaatagctttatttttaaataaatt
2161  attgagcctcatttatttttcttttttctccccccccctaccctgctagTCTGGAGTTGATCA
2221  AGGAACCTGTCTCCACAAAGTGTGACCACATATTTTGCAAgtaagtttgaatgtgttatg
2281  tggctccattattagcttttgtttttgtccttcataacccaggaaacacctaactttata
2341  gaagctttacttttcttcaattaagtgagaacgaaaatccaactccatttcattctttctc
2401  agagagtatatagttatcaaaagttggttgtaatcatagttcctggtaaagttttgacat
2461  atattatctttttttttttttgagacaagtctcgctctgtcgcccaggctggagtgcagt
2521  ggcatgaggcttgctcactgcacctccgcccccgagttcagcgactctvvvvvvvvvvvv
2581  vtgagatctagaccacatggtcaaagagatagaatgtgagcaataaatgaaccttaaatt
2641  tttcaacagctacttttttttttttttttttgagacagGGKCTTACTCTGTTGTCCCAGCT
2701  GGAGTACAGWGTGCGATCATGAGGCTTACTGTTGCTTGACTCCTAGGCTCAAGCGATCCT
2761  ATCACCTCAGTCTCCAAGTAGCTGGACTgtaagtgcacaccaccatatccagctaaattt
2821  tgtgttttctgtagagacggggtttcgccatgtttcccaggctggtcttgaactttgggc
2881  ttaacccgtctgcccacctaggcatcccaaagtgctaggattacaggtgtgagtcatcat
2941  gcctggccagtattttagttagctctgtcttttcaagtcatatacaagttcattttcttt
3001  taagtttagttaacaaccttatatcatgtattcttttctagcataaagaaagattcgagg
```

Figure 10A

PRINT OF DRAWINGS
AS ORIGINAL FILED


```
3061  ccvvvvvvvvvvvvvvtgtgatcataacagtaagccatatgcatgtaagttcagttttcat
3121  agatcattgcttatgtagtttaggtttttgcttatgcagcatccaaaaacaattaggaaa
3181  ctattgcttgtaattcacctgccattacttttaaatggctcttaagggcagttgtgaga
3241  ttatcttttcatggctatttgccttttgagtattctttctacaaaaggaagtaaattaaa
3301  ttgttctttctttctttataatttatagATTTTGCATGCTGAAACTTCTCAACCAGAAGA
3361  AAGGGCCTTCACAGTGTCCTTTATGTAAGAATGATATAACCAAAAGgtatataatttggt
3421  aatgatgctaggttggaagcaaccacagtaggaaaaagtagaaattatttaataacatag
3481  cgttcctataaaaccattcatcagaaaaatttataaaagagtttttagcacacagtaaat
3541  tatttccaaagttatttttcctgaaagttttatgggcatctgccttatacaggtattgvvv
3601  vvvvvvvvvvggtaggcttaaatgaatgacaaaaagttactaaatcactgccatcacacg
3661  gtttatacagatgtcaatgatgtattgattatagaggttttctactgttgctgcatctta
3721  tttttatttgtttacatgtcttttcttattttagtgtccttaaaaggttgataatcactt
3781  gctgagtgtgtttctcaaacaatttaatttcagGAGCCTACAAGAAAGTACGAGATTTAG
3841  TCAACTTGTTGAAGAGCTATTGAAAATCATTTGTGCTTTTCAGCTTGACACAGGTTTGGA
3901  GTgtaagtgttgaatatcccaagaatgacactcaagtgctgtccatgaaaactcaggaag
3961  tttgcacaattactttctatgacgtggtgataagacctttagtctaggttaattttagt
4021  tctgtatctgtaatctattttaaaaaattactcccactggtctcacaccttatttvvvvv
4081  vvvvvvvvaaaaaatcacaggtaaccttaatgcattgtcttaacacaacaaagagcatac
4141  atagggtttctcttggtttctttgattataattcatacatttttctctaactgcaaacat
4201  aatgttttcccttgtattttacagATGCAAACAGCTATAATTTTGCAAAAAAGGAAAATA
4261  ACTCTCCTGAACATCTAAAAGATGAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACC
4321  GTGCCAAAAGACTTCTACAGAGTGAACCCGAAAATCCTTCCTTGgtaaaaccatttgttt
4381  tcttcttcttgttcttcttcttttctttttttttttctttttttttttgagatggagtctt
4441  gctctgtggcccaggctagaagcagtcctcctgccttagccncccttagtagctgggatta
4501  caggcacgcgcaccatgccaggctaatttttgtatttttagtagagacggggtttcatca
4561  tgttggccaggctggtctcgaactcctaacctcaggtgatcvvvvvvvvvvvvvvatgatg
4621  gagatcttaaaaagtaatcatttctggggctgggcgtagtagcttgcacctgtaatcccag
4681  cacttcgggaggctgaggcaggcagataatttgaggtcaggagtttgagaccagcctggc
4741  caacatggtgaaacccatctctactaaaaatacaaaaattagctgggtgtggtggcacgt
4801  acctgtaatcccagctactcgggaggcggaggcacaagaattgcttgaacctaggacgcg
4861  gaggttgcagcgagccaagatcgcgccactgcactccagcctgggccgtagagtgagact
4921  ctgtctcaaaaaagaaaaaaaagtaattgttctagctgggcgcagtggctcttgcctgta
4981  atcccagcactttgggaggccaaggcgggtggatctcgagtcctagagttcaagaccagc
5041  ctaggcaatgtggtgaaaccccatcgctacaaaaaatacaaaaattagccaggcatggtg
5101  gcgtgcgcatgtagtcccagctccttgggaggctgaggtgggaggatcacttgaacccag
5161  gagacagaggttgcagtgaaccgagatcacgccaccacgctccagcctgggcaacagaac
5221  aagactctgtctaaaaaaatacaaataaaataaaagtagttctcacagtaccagcattca
5281  tttttcaaaagatatagagctaaaaaggaaggaaaaaaaaagtaatgttgggctttaaa
5341  tactcgttcctatactaaatgttcttaggagtgctggggttttattgtcatcatttatcc
5401  tttttaaaaatgttattggccaggcacggtggctcatggctgtaatcccagcactttggg
5461  aggccgaggcaggcagatcacctgaggtcaggagtgtgagaccagcctggccaacatggc
5521  gaaacctgtctctactaaaaatacaaaaattaactaggcgtggtggtgtacgcctgtagt
5581  cccagctactcgggaggctgaggcaggagaatcaactgaaccagggaggtggaggttgca
5641  gtgtgccgagatcacgccactgcactctagcctggcaacagagcaagattctgtctcaaa
5701  aaaaaaaaacatatatacacatatatcccaaagtgctgggattacatatatatatatata
5761  tatatattatatatatatatatatatatatgtgatatatatgtgatatatatataacata
5821  tatatatgtaatatatatgtgatatatatataatatatatatgtaatatatatgtgatat
5881  atatatatacacacacacacacatatatatgtatgtgtgtgtacacacacacacacaaat
5941  tagccaggcatagttgcacacgcctggtagacccagctactcaggaggctgagggaggag
6001  aatctcttgaacttaggaggcggaggttgcagtgagctgagattgcgccactgcactcca
6061  gcctgggtgacagagcaggactctgtacaccccccaaaacaaaaaaaaaaagttatcagat
```

Figure 10B

```
6121  gtgattggaatgtatatcaagtatcagcttcaaaatatgctatattaatacttcaaaaat
6181  tacacaaataatacataatcaggtttgaaaaatttaagacaacmsaaraaaaaawycmaa
6241  tcacamatatcccacacattttattattmctmctmcwattattttgwagagmctgggtct
6300  cacycykttgctwatgctggtctttgaacyccykgccycaarcartcctsctccabcctc
6361  ccaargtgctggggatwataggcatgarctaaccgcacccagccccagacattttagtgt
6421  gtaaattcctgggcatttttttcaaggcatcatacatgttagctgactgatgatggtcaat
6481  ttattttgtccatggtgtcaagtttctcttcaggaggaaaagcacagaactggccaacaa
6541  ttgcttgactgttctttaccatactgtttagCAGGAAACCAGTCTCAGTGTCCAACTCTC
6601  TAACCTTGGAACTGTGAGAACTCTGAGGACAAAGCAGCGGATACAACCTCAAAAGACGTC
6661  TGTCTACATTGAATTGGgtaagggtctcaggtttttttaagtatttaataataattgctgg
6721  attccttatcttatagttttgccaaaaatcttggtcataatttgtatttgtggtaggcag
6781  ctttgggaagtgaattttatgagccctatggtgagttataaaaaaatgtaaaagacgcagt
6841  tcccaccttgaagaatcttactttaaaaagggagcaaaagaggccaggcatggtggctca
6901  cacctgtaatcccagcactttgggaggccaaagtgggtggatcacctgaggtcgggagtt
6961  cgagaccagcctagccaacatggagaaactctgtctgtaccaaaaaatataaaaaattagcc
7021  aggtgtggtggcacataactgtaatcccagctactcgggaggctgaggcaggagaatcac
7081  ttgaacccgggaggtggaggttgcggtgaaccgagatcgcaccattgcactccagcctgg
7141  gcaaaaatagcgaaactccatctaaaaaaaaaaaagagagcaaaagaaagamtmtctggt
7201  tttaamtmtgtgtaaatatgtttttggaaagatggagagtagcaataagaaaaaacatga
7261  tggattgctacagtatttagttccaagataaattgtactagatgaggaagccttttaaga
7321  agagctgaattgccaggcgcagtggctcacgcctgtaatcccagcactttgggaggccga
7381  ggtgggcggatcacctgaggtcgggagttcaagaccagcctgaccaacatggagaaaccc
7441  catctctactaaaaaaaaaaaaaaaaaaattagccggggtggtggcttatgcctgtaatc
7501  ccagctactcaggaggctgaggcaggagaatcgcttgaacccaggaagcagaggttgcag
7561  tgagccaagatcgcaccattgcactccagcctaggcaacaagagtgaaactccatctcaa
7621  aaaaaaaaaaaaagagctgaatcttggctgggcaggatggctcgtgcctgtaatcctaac
7681  gctttggaagaccgaggcagaaggattggttgagtccacgagtttaagaccagcctggcc
7741  aacatagggaaccctgtctctatttttaaaataataatacatttttggccggtgcggtg
7801  gctcatgcctgtaatcccaatactttgggaggctgaggcaggtagatcacctgaggtcag
7861  agttcgagaccagcctggataacctggtgaaaccctctttactaaaaatacaaaaaaaa
7921  aaaaaaattagctgggtgtggtagcacatgcttgtaatcccagctacttgggaggctgag
7981  gcaggagaatcgcttgaaccagggaggcggaggttacaatgagccaacactacaccactg
8041  cactccagcctgggcaatagagtgagactgcatctcaaaaaaataataatttttaaaaat
8101  aataaatttttttaagcttataaaaagaaaagttgaggccagcatagtagctcacatctg
8161  taatctcagcagtggcagaggattgcttgaagccaggagtttgagaccagcctgggcaac
8221  atagcaagacctcatctctacaaaaaaatttcttttttaaattagctgggtgtggtggtg
8281  tgcatctgtagtcccagctactcaggaggcagaggtgagtggatacattgaacccaggag
8341  tttgaggctgtagtgagctatgatcatgccactgcactccaacctgggtgacagagcaag
8401  acctccaaaaaaaaaaaaaaaaagagctgctgagctcagaattcaaactgggctctcaaat
8461  tggattttcttttagaatatatttataattaaaaaggatagccatcttttgagctcccag
8521  gcaccaccatctatttatcataacacttactgttttccccccttatgatcataaattcct
8581  agacaacaggcattgtaaaaatagttatagtagttgatatttaggagcacttaactatat
8641  tccaggcactattgtgcttttcttgtataactcattagatgcttgtcagacctctgagat
8701  tgttcctattatacttattttacagatgagaaaattaaggcacagagaagttatgaaatt
8761  tttccaaggtattaaacctagtaagtggctgagccatgattcaaacctaggaagttagat
8821  gtcagagcctgtgctttttttttgttttgttttgttttcagtagaaacgggggtctca
8881  ctttgttggccaggctggtcttgaactcctaacctcaaataatccacccatctcggcctc
8941  ctcaagtgctgggattacaggtgagagccactgtgcctggcgaagcccatgcctttaacc
9001  acttctctgtattacatactagcttaactagcattgtacctgccacagtagatgctcagt
9061  aaatatttctagttgaatatctgtttttcaacaagtacatttttttaaccctttaatta
9121  agaaaactttattgatttatttttgggggaaatttttagGATCTGATTCTTCTGAA
```

Figure 10C

```
 9181  GATACCGTTAATAAGGCAACTTATTGCAGgtgagtcaaagagaacctttgtctatgaagc
 9241  tggtattttcctatttagttaatattaaggattgatgtttctctcttttaaaaatatttt
 9301  taacttttatttaggttcagggatgtatgtgcagtttgttatataggtaaacacacgac
 9361  ttgggatttggtgtatagatttttttcatcatccgggtactaagcataccccacagtttt
 9421  ttgtttgctttctttctgaatttctccctcttcccaccttcctccctcaagtaggctggt
 9481  gtttctccagactagaatcatggtattggaagaaaccttagagatcatctagtttagttc
 9541  tctcattttatagtggaggaaatacccttttgtttgttggatttagttattagcactgt
 9601  ccaaaggaatttaggataacagtagaactctgcacatgcttgcttctagcagattgttct
 9661  ctaagttcctcatatacagtaatattgacacagcagtaattgtgactgatgaaaatgttc
 9721  aaggacttcattttcaactctttctttcctctgttccttatttccacatatctctcaagc
 9781  tttgtctgtatgttatataataaactacaagcaaccccaactatgttacctaccttcctt
 9841  aggaattattgcttgacccaggtttttttttttttttttttggagacggggtcttgccct
 9901  gttgccaggatggagtgtagtggcgccatctcggctcactgcaatctccaactccctggt
 9961  tcaagcgattctcctgtctcaatctcacgagtagctgggactacaggtatacaccaccac
10021  gcccggttaattgaccattccatttctttctttctctctttttttttttttttttgaga
10081  cagagtcttgctctgttgcccaggctggagtacagaggtgtgatctcacctctccgcaac
10141  gtctgcctcccaggttgaagccatactcctgcctcagcctctctagtagctgggactaca
10201  ggcgcgcgccaccacacccggctaatttttgtatttttagtagagatggggtttcaccat
10261  gttggccaggctggtcttgaactcatgacctcaagtggtccacccgcctcagcctcccaa
10321  agtgctggaattacaggcttgagccaccgtgcccagcaaccatttcatttcaactagaag
10381  tttctaaaggagagagcagctttcactaactaaataagattggtcagctttctgtaatcg
10441  aaagagctaaaatgtttgatcttggtcatttgacagttctgcatacatgtaactagtgtt
10501  tcttattaggactctgtcttttccctatagTGTGGGAGATCAAGAATTGTTACAAATCAC
10561  CCCTCAAGGAACCAGGGATGAAATCAGTTTGGATTCTGCAAAAAAGGgtaatggcaaagt
10621  ttgccaacttaacaggcactgaaaagagagtgggtagatacagtactgtaattagattat
10681  tctgaagaccatttgggacctttacaacccacaaaatctcttggcagagttagagtatca
10741  ttctctgtcaaatgtcgtggtatggtctgatagatttaaatggtactagactaatgtacc
10801  tataataagaccttcttgtaactgattgttgccctttcgcttttttttttgtttgtttgt
10861  ttgtttttttttgagatggggtctcactctgttgcccaggctggagtgcagtgatgcaat
10921  cttggctcactgcaacctccacctccaaaggctcaagctatcctcccacttcagcctcct
10981  gagtagctgggactacaggcgcatgccaccacacccggttaatttttgtggttttatag
11041  agatggggtttcaccatgttaccgaggctggtctcaaactcctggactcaagcagtctgc
11101  ccacttcagcctcccaaagtgctgcagttacaggcttgagccactgtgcctggcctgccc
11161  tttacttttaattggtgtatttgtgtttcatcttttacctactggtttttaaatatagggg
11221  agtggtaagtctgtagatagaacagagtattaagtagacttaatggccagtaatctttag
11281  agtacatcagaaccagttttctgatggccaatctgcttttaattcactcttagacgttag
11341  agaaataggtgtggtttctgcatagggaaaattctgaaattaavvvvvvvvvvvvvvgatc
11401  ctaagtggaaataatctaggtaaataggaattaaatgaaagagtatgagctacatcttca
11461  gtatacttggtagtttatgaggttagtttctctaatatagccagttggttgatttccacc
11521  tccaaggtgtatgaagtatgtatttttttaatgacaattcagtttttgagtaccttgtta
11581  tttttgtatattttcagCTGCTTGTGAATTTTCTGAGACGGATGTAACAAATACTGAACA
11641  TCATCAACCCAGTAATAATGATTTGAACACCACTGAGAAGCGTGCAGCTGAGAGGCATCC
11701  AGAAAAGTATCAGGGTAGTTCTGTTTCAAACTTGCATGTGGAGCCATGTGGCACAAATAC
11761  TCATGCCAGCTCATTACAGCATGAGAACAGCAGTTTATTACTCACTAAAGACAGAATGAA
11821  TGTAGAAAAGGCTGAATTCTGTAATAAAAGCAAACAGCCTGGCTTAGCAAGGAGCCAACA
11881  TAACAGATGGGCTGGAAGTAAGGAAACATGTAATGATAGGCGGACTCCCAGCACAGAAAA
11941  AAAGGTAGATCTGAATGCTGATCCCCTGTGTGAGAGAAAAGAATGGAATAAGCAGAAACT
12001  GCCATGCTCAGAGAATCCTAGAGATACTGAAGATGTTCCTTGGATAACACTAAATAGCAG
12061  CATTCAGAAAGTTAATGAGTGGTTTTCCAGAAGTGATGAACTGTTAGGTTCTGATGACTC
12121  ACATGATGGGGAGTCTGAATCAAATGCCAAAGTAGCTGATGTATTGGACGTTCTAAATGA
12181  GGTAGATGAATATTCTGGTTCTTCAGAGAAAATAGACTTACTGGCCAGTGATCCTCATGA
```

Figure 10D

```
12241  GGCTTTAATATGTAAAAGTGAAAGAGTTCACTCCAAATCAGTAGAGAGTAATATTGAAGG
12301  CCAAATATTTGGGAAAACCTATCGGAAGAAGGCAAGCCTCCCCAACTTAAGCCATGTAAC
12361  TGAAAATCTAATTATAGGAGCATTTGTTACTGAGCCACAGATAATACAAGAGCGTCCCCT
12421  CACAAATAAATTAAAGCGTAAAAGGAGACCTACATCAGGCCTTCATCCTGAGGATTTTAT
12481  CAAGAAAGCAGATTTGGCAGTTCAAAAGACTCCTGAAATGATAAATCAGGGAACTAACCA
12541  AACGGAGCAGAATGGTCAAGTGATGAATATTACTAATAGTGGTCATGAGAATAAAACAAA
12601  AGGTGATTCTATTCAGAATGAGAAAAATCCTAACCCAATAGAATCACTCGAAAAAGAATC
12661  TGCTTTCAAAACGAAAGCTGAACCTATAAGCAGCAGTATAAGCAATATGGAACTCGAATT
12721  AAATATCCACAATTCAAAAGCACCTAAAAAGAATAGGCTGAGGAGGAAGTCTTCTACCAG
12781  GCATATTCATGCGCTTGAACTAGTAGTCAGTAGAAATCTAAGCCCACCTAATTGTACTGA
12841  ATTGCAAATTGATAGTTGTTCTAGCAGTGAAGAGATAAAGAAAAAAAAGTACAACCAAAT
12901  GCCAGTCAGGCACAGCAGAAACCTACAACTCATGGAAGGTAAAGAACCTGCAACTGGAGC
12961  CAAGAAGAGTAACAAGCCAAATGAACAGACAAGTAAAAGACATGACAGCGATACTTTCCC
13021  AGAGCTGAAGTTAACAAATGCACCTGGTTCTTTTACTAAGTGTTCAAATACCAGTGAACT
13081  TAAAGAATTTGTCAATCCTAGCCTTCCAAGAGAAGAAAAAGAAGAGAAACTAGAAACAGT
13141  TAAAGTGTCTAATAATGCTGAAGACCCCAAAGATCTCATGTTAAGTGGAGAAAGGGTTTT
13201  GCAAACTGAAAGATCTGTAGAGAGTAGCAGTATTTCATTGGTACCTGGTACTGATTATGG
13261  CACTCAGGAAAGTATCTCGTTACTGGAAGTTAGCACTCTAGGGAAGGCAAAAACAGAACC
13321  AAATAAATGTGTGAGTCAGTGTGCAGCATTTGAAAACCCCAAGGGACTAATTCATGGTTG
13381  TTCCAAAGATAATAGAAATGACACAGAAGGCTTTAAGTATCCATTGGGACATGAAGTTAA
13441  CCACAGTCGGGAAACAAGCATAGAAATGGAAGAAAGTGAACTTGATGCTCAGTATTTGCA
13501  GAATACATTCAAGGTTTCAAAGCGCCAGTCATTTGCTCCGTTTTCAAATCCAGGAAATGC
13561  AGAAGAGGAATGTGCAACATTCTCTGCCCACTCTGGGTCCTTAAAGAAACAAAGTCCAAA
13621  AGTCACTTTTGAATGTGAACAAAAGGAAGAAAATCAAGGAAAGAATGAGTCTAATATCAA
13681  GCCTGTACAGACAGTTAATATCACTGCAGGCTTTCCTGTGGTTGGTCAGAAAGATAAGCC
13741  AGTTGATAATGCCAAATGTAGTATCAAAGGAGGCTCTAGGTTTTGTCTATCATCTCAGTT
13801  CAGAGGCAACGAAACTGGACTCATTACTCCAAATAAACATGGACTTTTACAAAACCCATA
13861  TCGTATACCACCACTTTTTCCCATCAAGTCATTTGTTAAAACTAAATGTAAGAAAAATCT
13921  GCTAGAGGAAAACTTTGAGGAACATTCAATGTCACCTGAAAGAGAAATGGGAAATGAGAA
13981  CATTCCAAGTACAGTGAGCACAATTAGCCGTAATAACATTAGAGAAAATGTTTTTAAAGA
14041  AGCCAGCTCAAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGGCTCCAGTAT
14101  TAATGAAATAGGTTCCAGTGATGAAAACATTCAAGCAGAACTAGGTAGAAACAGAGGGCC
14161  AAAATTGAATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGTCTATAAACAAAGTCT
14221  TCCTGGAAGTAATTGTAAGCATCCTGAAATAAAAAAGCAAGAATATGAAGAAGTAGTTCA
14281  GACTGTTAATACAGATTTCTCTCCATATCTGATTTCAGATAACTTAGAACAGCCTATGGG
14341  AAGTAGTCATGCATCTCAGGTTTGTTCTGAGACACCTGATGACCTGTTAGATGATGGTGA
14401  AATAAAGGAAGATACTAGTTTTGCTGAAAATGACATTAAGGAAAGTTCTGCTGTTTTTAG
14461  CAAAAGCGTCCAGAAAGGAGAGCTTAGCAGGAGTCCTAGCCCTTTCACCCATACACATTT
14521  GGCTCAGGGTTACCGAAGAGGGGCCAAGAAATTAGAGTCCTCAGAAGAGAACTTATCTAG
14581  TGAGGATGAAGAGCTTCCCTGCTTCCAACACTTGTTATTTGGTAAAGTAAACAATATACC
14641  TTCTCAGTCTACTAGGCATAGCACCGTTGCTACCGAGTGTCTGTCTAAGAACACAGAGGA
14701  GAATTTATTATCATTGAAGAATAGCTTAAATGACTGCAGTAACCAGGTAATATTGGCAAA
14761  GGCATCTCAGGAACATCACCTTAGTGAGGAAACAAAATGTTCTGCTAGCTTGTTTTCTTC
14821  ACAGTGCAGTGAATTGGAAGACTTGACTGCAAATACAAACACCCAGGATCCTTTCTTGAT
14881  TGGTTCTTCCAAACAAATGAGGCATCAGTCTGAAAGCCAGGGAGTTGGTCTGAGTGACAA
14941  GGAATTGGTTTCAGATGATGAAGAAAGAGGAACGGGCTTGGAAGAAAATAATCAAGAAGA
15001  GCAAAGCATGGATTCAAACTTAGgtattggaaccaggtttttgtgtttgccccagtctat
15061  ttatagaagtgagctaaatgtttatgcttttggggagcacattttacaaatttccaagta
15121  tagttaaaggaactgcttcttaaacttgaaacatgttcctcctaaggtgcttttcataga
15181  aaaaagtccttcacacagctaggacgtcatctttgactgaatgagctttaacatcctaat
15241  tactggtggacttacttctggtttcattttataaagcaaatccCggtgtcccaaagcaag
```

Figure 10E

PRINT OF DRAWINGS
AS ORIGINAL FILED


```
15301  gaatttaatcattttgtgtgacatgaaagtaaatccagtcctgccaatgagaagaaaaag
15361  acacagcaagttgcagcgtttatagtctgcttttacatctgaacctctgtttttgttatt
15421  taagGTGAAGCAGCATCTGGGTGTGAGAGTGAAACAAGCGTCTCTGAAGACTGCTCAGGG
15481  CTATCCTCTCAGAGTGACATTTTAACCACTCaggtaaaaagcgtgtgtgtgtgtgcacat
15541  gcgtgtgtgtggtgtcctttgcattcagtagtatgtatcccacattcttaggtttgctga
15601  catcatctctttgaattaatggcacaattgtttgtggttcattgtcvvvvvvvvvvvvvn
15661  gngaatgtaatcctaatatttcncnccnacttaaaagaataccactccaanggcatcnca
15721  atacatcaatcaattggggaattgggattttccctcnctaacatcantggaataatttca
15781  tggcattaattgcatgaatgtggttagattaaaaggtgttcatgctagaacttgtagttc
15841  catactaggtgatttcaattcctgtgctaaaattaatttgtatgatatattntcatttaa
15901  tggaaagcttctcaaagtatttcattttcttggtaccatttatcgtttttgaAGCAGAGG
15961  GATACCATGCAACATAACCTGATAAAGCTCCAGCAGGAAATGGCTGAACTAGAAGCTGTG
16021  TTAGAACAGCATGGGAGCCAGCCTTCTAACAGCTACCCTTCCATCATAAGTGACTCTTCT
16081  GCCCTTGAGGACCTGCGAAATCCAGAACAAAGCACATCAGAAAAAGgtgtgtattgttgg
16141  ccaaacactgatatcttaagcaaaattctttccttcccctttatctccttctgaagagta
16201  aggacctagctccaacattttatgatccttgctcagcacatgggtaattatggagccttg
16261  gttcttgtccctgctcacaactaatataccagtcagagggacccaaggcagtcattcatg
16321  ttgtcatctgagatacctacaacaagtagatgctatggggagcccatggvvvvvvvvvvv
16381  vvccattggtgctagcatctgtctgttgcattgcttgtgtttataaaattctgcctgata
16441  tacttgttaaaaaccaatttgtgtatcatagattgatgcttttgaaaaaaatcagtattc
16501  taacctgaattatcactatcagaacaaagcagtaaagtagatttgttttctcattccatt
16561  taaagCAGTATTAACTTCACAGAAAAGTAGTGAATACCCTATAAGCCAGAATCCAGAAGG
16621  CCTTTCTGCTGACAAGTTTGAGGTGTCTGCAGATAGTTCTACCAGTAAAAATAAAGAACC
16681  AGGAGTGGAAAGgtaagaaacatcaatgtaaagatgctgtggtatctgacatctttattt
16741  atattgaactctgattgttaatttttttcaccatactttctccagtttttttgcatacag
16801  gcatttatacacttttattgctctaggatacttcttttgtttaatcctatataggvvvvv
16861  vvvvvvvvggataagntcaagagatattttgataggtgatgcagtgatnaattgngaaaa
16921  tttnctgcctgcttttaatcttccccgttctttcttcctnccteccteccttcctncct
16981  cccgtccttnccttttcctttccctccctccnccttctttccntctntctttcctttctt
17041  tcctgtctacctttctttccttcctccctccttttctttttctttctttcctttcctttt
17101  ctttcctttctttccttttcctttctttcttgacagagtcttgctctgtcactcaggctgg
17161  agtgcagtggcgtgatctcgnctcactgcaacctctgtctcccaggttcaagcaattttc
17221  ctgcctcagcctcccgagtagctgagattacaggcgccagccaccacacccagctactga
17281  cctgcttttvvvvvvvvvvvvvvaaacagctgggagatatggtgcctcagaccaaccccat
17341  gttatatgtcaaccctgacatattggcaggcaacatgaatccagacttctaggctgtcat
17401  gcgggctctttttgccagtcatttctgatctctctgacatgagctgtttcatttatgct
17461  ttggctgcccagcaagtatgatttgtcctttcacaattggtggcgatggttttctccttc
17521  catttatctttctagGTCATCCCCTTCTAAATGCCCATCATTAGATGATAGGTGGTACAT
17581  GCACAGTTGCTCTGGGAGTCTTCAGAATAGAAACTACCCATCTCAAGAGGAGCTCATTAA
17641  GGTTGTTGATGTGGAGGAGCAACAGCTGGAAGAGTCTGGGCCACACGATTTGACGGAAAC
17701  ATCTTACTTGCCAAGGCAAGATCTAGgtaatatttcatctgctgtattggaacaaacact
17761  ytgattttactctgaatcctacataaagatattctggttaaccaacttttagatgtacta
17821  gtctatcatggacacttttgttatacttaattaagcccactttagaaaaatagctcaagt
17881  gttaatcaaggtttacttgaaaattattgaaactgttaatccatctatattttaattaat
17941  ggtttaactaatgattttgaggatgwgggagtcktggtgtactctamatgtattatttca
18001  ggccaggcatagtggctcacgcctggtaatcccagtayycmrgagcccgaggcaggtgga
18061  gccagctgaggtcaggagttcaagacctgtcttggccaacatgggngaaaccctgtcttc
18121  ttcttaaaaaanacaaaaaaaattaactgggttgtgcttaggtgnatgccccgnatccta
18181  gttnttcttgnggggttgagggaggagatcacnttggaccccggaggggngggtgggggng
18241  agcaggncaaaacacngacccagctggggtggaagggaagcccactcnaaaaaannttnv
18301  vvvvvvvvvvvvttttaggaaacaagctactttggatttccaccaacacctgtattcat
```

Figure 10F

PRINT OF DRAWINGS
AS ORIGINALLY FILED


```
18361  gtacccatttttctcttaacctaactttattggtctttttaattcttaacagagaccaga
18421  actttgtaattcaacattcatcgttgtgtaaattaaacttctcccattcctttcagAGGG
18481  AACCCCTTACCTGGAATCTGGAATCAGCCTCTTCTCTGATGACCCTGAATCTGATCCTTC
18541  TGAAGACAGAGCCCCAGAGTCAGCTCGTGTTGGCAACATACCATCTTCAACCTCTGCATT
18601  GAAAGTTCCCCAATTGAAAGTTGCAGAATCTGCCCAGAGTCCAGCTGCTGCTCATACTAC
18661  TGATACTGCTGGGTATAATGCAATGGAAGAAAGTGTGAGCAGGGAGAAGCCAGAATTGAC
18721  AGCTTCAACAGAAAGGGTCAACAAAAGAATGTCCATGGTGGTGTCTGGCCTGACCCCAGA
18781  AGAATTTgtgagtgtatccatatgtatctccctaatgactaagacttaacaacattctgg
18841  aaagagttttatgtaggtattgtcaattaataacctagaggaagaaatctagaaaacaat
18901  cacagttctgtgtaatttaatttcgattactaatttctgaaaatttagaaayvvvvvvvvv
18961  vvvvncccnncccccnaatctgaaatgggggtaacccccccccaaccganacntgggtng
19021  cntagagantttaatggcccnttctgaggnacanaagcttaagccaggngacgtggancn
19081  atgngttgtttnttgtttggttacctccagcctgggtgacagagcaagactctgtctaaa
19141  aaaaaaaaaaaaaaaaatcgactttaaatagttccaggacacgtgtagaacgtgcaggat
19201  tgctacgtaggtaaacatatgccatggtgggataactagtattctgagctgtgtgctaga
19261  ggtaactcatgataatggaatatttgatttaatttcagATGCTCGTGTACAAGTTTGCCA
19321  GAAAACACCACATCACTTTAACTAATCTAATTACTGAAGAGACTACTCATGTTGTTATGA
19381  AAACAGgtataccaagaaccttacagaataccttgcatctgctgcataaaaccacatga
19441  ggcgaggcacggtggcgcatgcctgtaatcgcagcactttgggaggccgaggcgggcaga
19501  tcacgagattaggagatcgagaccatcctggccagcatggtgaaaccccgtctctactan
19561  naaatggnaaaattanctgggtgtggtcgcgtgcnnctgtagtcccagctactcgtgagg
19621  ctgaggcaggagaatcacttgaaccggggaaatggaggtttcagtgagcagagatcatnc
19681  ccctncattccagcctggcgacagagcaaggctccgtcnccnaaaaaataaaaaaaaacg
19741  tgaacaaataagaatatttgttgagcatagcatggatgatagtcttctaatagtcaatca
19801  attactttatgaaagacaaataatagttttgctgcttccttacctccttttgttttgggt
19861  taagatttggagtgtgggccaggcacvvvvvvvvvvvvvvgatctatagctagccttggcg
19921  tctagaagatgggtgttgagaagagggagtggaaagatatttcctctggtcttaacttca
19981  tatcagcctcccctagacttccaaatatccatacctgctggttataattagtggtgtttt
20041  cagcctctgattctgtcaccaggggttttagaatcataaatccagattgatcttgggagt
20101  gtaaaaaactgaggctctttagcttcttaggacagcagttcctgattttgttttcaactt
20161  ctaatcctttgagtgtttttcattctgcagATGCTGAGTTTGTGTGTGAACGGACACTGA
20221  AATATTTTCTAGGAATTGCGGGAGGAAAATGGGTAGTTAGCTATTTCTgtaagtataata
20281  ctatttctcccctcctcccttaacacctcagaattgcatttttacacctaacatttaac
20341  acctaaggtttttgctgatgctgagtctgagttaccaaaaggtctttaaattgtaatact
20401  aaactacttttatctttaatatcactttgttcaagataagctggtgatgctgggaaaatg
20461  ggtctcttttataactaataggacctaatctgctcctagcaatgttagcatatgagctag
20521  ggatttatttaatagtcggcaggaatccatgtgcarcagncaaacttataatgtttaaat
20581  taaacatcaactctgtctccagaaggaaactgctgctacaagccttattaaagggctgtg
20641  gctttagagggaaggacctctcctctgtcattcttcctgtgctcttttgtgaatcgctga
20701  cctctctatctccgtgaaaagagcacgttcttctgctgtatgtaacctgtctttttctatg
20761  atctctvvvvvvvvvvvvvvnaaaaacggggnngggantgggccttaaanccaaagggcna
20821  actccccaaccattnaaaaantgacnggggattattaaaancggcgggaaacatttcacn
20881  gcccaactaatattgttaaattaaaaccaccaccnctgcnccaaggagggaaactgctgc
20941  tacaagccttattaaagggctgtggctttagagggaaggacctctcctctgtcattcttc
21001  ctgtgctctttttgtgaatcgctgacctctctatgtccgtgaaaagagcacgttcttcgtc
21061  tgtatgtaacctgtctttttctatgatctctttagGGGTGACCCAGTCTATTAAAGAAAGA
21121  AAAATGCTGAATGAGgtaagtacttgatgttacaaactaaccagagatattcattcagtc
21181  atatagttaaaaatgtatttgcttccttccatcaatgcaccactttccttaacaatgcac
21241  aaattttccatgataatgaggattcatcaagaattatgcaggcctgcactgtggctcatac
21301  ctataatcccagcgctttgggaggctgaggcgcttggatcvvvvvvvvvvvvvvaattttt
21361  tgtatttttagtagagatgaggttcaccatgttggtctagatctggtgtcgaacgtcctg
```

Figure 10G

PRINT OF DRAWINGS
AS ORIGINA FILED


```
21421  acctcaagtgatctgccagcctcagtctcccaaagtgctaggattacaggggtgagccac
21481  tgcgcctggcctgaatgcctaaaatatgacgtgtctgctccacttccattgaaggaagct
21541  tctctttctcttatcctgatgggttgtgtttggtttctttcagCATGATTTTGAAGTCAG
21601  AGGAGATGTGGTCAATGGAAGAAACCACCAAGGTCCAAAGCGAGCAAGAGAATCCCAGGA
21661  CAGAAAGgtaaagctccctccctcaagttgacaaaaatctcaccccaccactctgtattc
21721  cactccccttgcagagatgggccgcttcattttgtaagacttattacatacatacacag
21781  tgctagatactttcacacaggttctttttcactcttccatcccaaccacataaataagt
21841  attgtctctactttatgaatgataaaactaagagatttagagaggctgtgtaatttggat
21901  tcccgtctcgggttcagatcvvvvvvvvvvvvvttggcctgattggtgacaaaagtgaga
21961  tgctcagtccttgaatgacaaagaatgcctgtagagttgcaggtccaactacatatgcac
22021  ttcaagaagatcttctgaaatctagtagtgttctggacattggactgcttgtccctggga
22081  agtagcagcagaaatgatcggtggtgaacagaagaaaaagaaaagctcttccttttgaa
22141  agtctgtttttgaataaaagccaatattcttttataactagattttccttctctccatt
22201  cccctgtccctctctcttcctctcttcttccagATCTTCAGGGGGCTAGAAATCTGTTGC
22261  TATGGGCCCTTCACCAACATGCCCACAGgtaagagcctgggagaaccccagagttccagc
22321  accagcctttgtcttacatagtggagtattataagcaaggtcccacgatgggggttcctc
22381  agattgctgaaatgttctagaggctattctatttctctaccactctccaaacaaaacagc
22441  acctaaatgttatcctatggcaaaaaaaaactataccttgtccccttctcaagagcatg
22501  aaggtggttaatagttaggattcagtatgttatgtgttcagatggcgttgagctgctgtt
22561  agtgccvvvvvvvvvvvvvvtttgagagactatcaaaccttataccaagtggccttatgga
22621  gactgataaccagagtacatggcatatcagtggcaaattgacttaaaatccataccccta
22681  ctattttaagaccattgtcctttggagcagagagacagactctcccattgagaggtcttg
22741  ctataagccttcatccggagagtgtagggtagagggcctgggttaagtatgcagattact
22801  gcagtgattttacatgtaaatgtccattttagATCAACTGGAATGGATGGTACAGCTGTG
22861  TGGTGCTTCTGTGGTGAAGGAGCTTTCATCATTCACCCTTGGCACAgtaagtattgggtg
22921  ccctgtcagtgtgggaggacacaatattctctcctgtgagcaagactggcacctgtcagt
22981  ccctatggatgcccctactgtagcctcagaagtcttctctgcccacatacctgtgccaaa
23041  agactccatvvvvvvvvvvvvvvggtggtacgtgtctgtagttccagctacttgggaggct
23101  gagatggaaggattgcttgagcccaggaggcagaggtggnannttacgctgagatcacac
23161  cactgcactccagcctgggtgacagagcaagaccctgtctcaaaaacaaacaaaaaaaat
23221  gatgaagtgacagttccagtagtcctactttgacactttgaatgctctttccttcctggg
23281  gatccagGGTGTCCACCCAATTGTGGTTGTGCAGCCAGATGCCTGGACAGAGGACAATGG
23341  CTTCCATGgtaaggtgcctcgcatgtacctgtgctattagtggggtccttgtgcatgggt
23401  ttggtttatcactcattacctggtgcttgagtagcacagttcttggcacatttttaaata
23461  tttgttgaatgaatggctaaaatgtctttttgatgttttattgttatttgttttatatt
23521  gtaaaagtaatacatgaactgtttccatggggtgggagtaagatatgaatgttcatcacv
23581  vvvvvvvvvvvvcagtaatcctnagaactcatacgaccgggcccctggagtcgntgnttn
23641  gagcctagtccnggagaatgaattgacactaatctctgcttgtgttctctgtctccagCA
23701  ATTGGGCAGATGTGTGAGGCACCTGTGGTGACCCGAGAGTGGGTGTTGGACAGTGTAGCA
23761  CTCTACCAGTGCCAGGAGCTGGACACCTACCTGATACCCCAGATCCCCCACAGCCACTAC
23821  TGACTGCAGCCAGCCACAGGTACAGAGCCACAGGACCCCAAGAATGAGCTTACAAAGTGG
23881  CCTTTCCAGGCCCTGGGAGCTCCTCTCACTCTTCAGTCCTTCTACTGTCCTGGCTACTAA
23941  ATATTTTATGTACATCAGCCTGAAAAGGACTTCTGGCTATGCAAGGGTCCCTTAAAGATT
24001  TTCTGCTTGAAGTCTCCCTTGGAAAT
```

Figure 10H

08/488011

MAIL ROOM
SEP 15
1995
TRADEMARK OFFICE

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Re. application of:

Mark H. SKOLNICK et al.

Serial No. (to be assigned)

Filed: herewith

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

STATEMENT PURSUANT TO 37 CFR §1.821(f)

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

In the matter of the above-identified application, presented herewith is a computer-readable copy of the Sequence Listing pursuant to 37 CFR §1.821(f). It is hereby certified that these copies are identical.

Respectfully submitted,

Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI
1201 New York Avenue, N.W., #1000
Washington, D.C. 20005
(202) 962-4810

Atty. Docket No. 24884-109347-B
Dated: 07 June 1995

UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

APPLICATION NUMBER	FILING DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
08/488,011	06/07/95	SKOLNICK M	24684-109347

0242/0818
VENABLE BAETJER HOWARD AND CIVILETTI
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

#3

0000

DATE MAILED: 08/18/95

NOTICE TO FILE MISSING PARTS OF APPLICATION
FILING DATE GRANTED

An Application Number and Filing Date have been assigned to this application. However, the items indicated below are missing. The required items and fees identified below must be timely submitted **ALONG WITH THE PAYMENT OF A SURCHARGE** for items 1 and 3-6 only of $ 130 for large entities or $ 65 for small entities who have filed a verified statement claiming such status. The surcharge is set forth in 37 CFR 1.16(e).

If all required items on this form are filed within the period set below, the total amount owed by applicant as a ☒ large entity, ☐ small entity (verified statement filed), is $ 1014.

Applicant is given **ONE MONTH FROM THE DATE OF THIS LETTER, OR TWO MONTHS FROM THE FILING DATE** of this application, **WHICHEVER IS LATER,** within which to file all required items and pay any fees required above to avoid abandonment. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

1. ☒ The statutory basic filing fee is: ☒ missing ☐ insufficient. Applicant as a ☒ large entity ☐ small entity, must submit $ 730.00 to complete the basic filing fee.

2. ☒ Additional claim fees of $ 154.00 as a ☒ large entity, ☐ small entity, including any required multiple dependent claim fee, are required. Applicant must submit the additional claim fees or cancel the additional claims for which fees are due.

3. ☒ The oath or declaration:
   ☒ is missing.
   ☐ does not cover the newly submitted items.

   An oath or declaration in compliance with 37 CFR 1.63, identifying the application by the above Application Number and Filing Date is required.

4. ☐ The oath or declaration does not identify the application to which it applies. An oath or declaration in compliance with 37 CFR 1.63, identifying the application by the above Application Number and Filing Date, is required.

5. ☐ The signature(s) to the oath or declaration is/are: ☐ missing; ☐ by a person other than the inventor or a person qualified under 37 CFR 1.42, 1.43, or 1.47. A properly signed oath or declaration in compliance with 37 CFR 1.63, identifying the application by the above Application Number and Filing Date, is required.

6. ☐ The signature of the following joint inventor(s) is missing from the oath or declaration:

   _____ An oath or declaration listing the names of all inventors and signed by the omitted inventor(s), identifying this application by the above Application Number and Filing Date, is required.

7. ☐ The application was filed in a language other than English. Applicant must file a verified English translation of the application and a fee of $_____ under 37 CFR 1.17(k), unless this fee has already been paid.

8. ☐ A $_____ processing fee is required since your check was returned without payment. (37 CFR 1.21(m)).

9. ☐ Your filing receipt was mailed in error because your check was returned without payment.

10. ☒ The application does not comply with the Sequence Rules. See attached Notice to Comply with Sequence Rules 37 CFR 1.821-1.825.

11. ☐ Other.

Direct the response to Box Missing Part and refer any questions to the Customer Service Center at (703) 308-1202.

A copy of this notice MUST be returned with the response.

PTO-1533 (REV. 11-94) OFFICE COPY

This Raw Listing contains the General Information Section and up to the first 5 pages.

#24

ENTERED

SEQUENCE LISTING

(1) General Information:

(i) APPLICANT: Skolnick, Mark H.
Goldgar, David E.
Miki, Yoshio
Swenson, Jeff
Kamb, Alexander
Harshman, Keith D.
Shattuck-Eidens, Donna M.
Tavtigian, Sean V.
Wiseman, Roger W.
Futreal, P. Andrew

(ii) TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer Susceptibility Gene

(iii) NUMBER OF SEQUENCES: 85

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
(B) STREET: 1201 New York Avenue, N.W., Suite 1000
(C) CITY: Washington
(D) STATE: DC
(E) COUNTRY: USA
(F) ZIP: 20005

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
--> **(A) APPLICATION NUMBER: US**
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/409,305
(B) FILING DATE: 24-MAR-1995

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/348,824
(B) FILING DATE: 29-NOV-1994

```
47
48      (vii) PRIOR APPLICATION DATA:
49            (A) APPLICATION NUMBER: US 08/308,104
50            (B) FILING DATE: 16-SEP-1994
51
52      (vii) PRIOR APPLICATION DATA:
53            (A) APPLICATION NUMBER: US 08/300,266
54            (B) FILING DATE: 02-SEP-1994
55
56      (vii) PRIOR APPLICATION DATA:
57            (A) APPLICATION NUMBER: US 08/289,221
58            (B) FILING DATE: 12-AUG-1994
59
60       (viii) ATTORNEY/AGENT INFORMATION:
61            (A) NAME: Ihnen, Jeffrey L.
62            (B) REGISTRATION NUMBER: 28,957
63            (C) REFERENCE/DOCKET NUMBER: 24884-109347
64
65       (ix) TELECOMMUNICATION INFORMATION:
66            (A) TELEPHONE: 202-962-4810
67            (B) TELEFAX: 202-962-8300
68
69
70   (2) INFORMATION FOR SEQ ID NO:1:
71
72        (i) SEQUENCE CHARACTERISTICS:
73            (A) LENGTH: 5914 base pairs
74            (B) TYPE: nucleic acid
75            (C) STRANDEDNESS: double
76            (D) TOPOLOGY: linear
77
78       (ii) MOLECULE TYPE: cDNA
79
80      (iii) HYPOTHETICAL: NO
81
82       (iv) ANTI-SENSE: NO
83
84       (vi) ORIGINAL SOURCE:
85            (A) ORGANISM: Homo sapiens
86
87       (ix) FEATURE:
88            (A) NAME/KEY: CDS
89            (B) LOCATION: 120..5711
90
91
92       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
93
94   AGCTCGCTGA GACTTCCTGG ACCCCGCACC AGGCTGTGGG GTTTCTCAGA TAACTGGGCC      60
95
96   CCTGCGCTCA GGAGGCCTTC ACCCTCTGCT CTGGGTAAAG TTCATTGGAA CAGAAAGAA      119
97
98   ATG GAT TTA TCT GCT CTT CGC GTT GAA GAA GTA CAA AAT GTC ATT AAT         167
99   Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
```

```
100     1               5                   10                  15
101
102   GCT ATG CAG AAA ATC TTA GAG TGT CCC ATC TGT CTG GAG TTG ATC AAG        215
103   Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
104               20                  25                  30
105
106   GAA CCT GTC TCC ACA AAG TGT GAC CAC ATA TTT TGC AAA TTT TGC ATG        263
107   Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
108            35                  40                  45
109
110   CTG AAA CTT CTC AAC CAG AAG AAA GGG CCT TCA CAG TGT CCT TTA TGT        311
111   Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
112        50                  55                  60
113
114   AAG AAT GAT ATA ACC AAA AGG AGC CTA CAA GAA AGT ACG AGA TTT AGT        359
115   Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
116    65                  70                  75                  80
117
118   CAA CTT GTT GAA GAG CTA TTG AAA ATC ATT TGT GCT TTT CAG CTT GAC        407
119   Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
120                   85                  90                  95
121
122   ACA GGT TTG GAG TAT GCA AAC AGC TAT AAT TTT GCA AAA AAG GAA AAT        455
123   Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
124               100                 105                 110
125
126   AAC TCT CCT GAA CAT CTA AAA GAT GAA GTT TCT ATC ATC CAA AGT ATG        503
127   Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
128            115                 120                 125
129
130   GGC TAC AGA AAC CGT GCC AAA AGA CTT CTA CAG AGT GAA CCC GAA AAT        551
131   Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
132        130                 135                 140
133
134   CCT TCC TTG CAG GAA ACC AGT CTC AGT GTC CAA CTC TCT AAC CTT GGA        599
135   Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
136   145                 150                 155                 160
137
138   ACT GTG AGA ACT CTG AGG ACA AAG CAG CGG ATA CAA CCT CAA AAG ACG        647
139   Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
140                   165                 170                 175
141
142   TCT GTC TAC ATT GAA TTG GGA TCT GAT TCT TCT GAA GAT ACC GTT AAT        695
143   Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
144               180                 185                 190
145
146   AAG GCA ACT TAT TGC AGT GTG GGA GAT CAA GAA TTG TTA CAA ATC ACC        743
147   Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
148            195                 200                 205
149
150   CCT CAA GGA ACC AGG GAT GAA ATC AGT TTG GAT TCT GCA AAA AAG GCT        791
151   Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
152        210                 215                 220
```

```
153
154  GCT TGT GAA TTT TCT GAG ACG GAT GTA ACA AAT ACT GAA CAT CAT CAA       839
155  Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
156  225                 230                 235                 240
157
158  CCC AGT AAT AAT GAT TTG AAC ACC ACT GAG AAG CGT GCA GCT GAG AGG       887
159  Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg
160                  245                 250                 255
161
162  CAT CCA GAA AAG TAT CAG GGT AGT TCT GTT TCA AAC TTG CAT GTG GAG       935
163  His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu
164              260                 265                 270
165
166  CCA TGT GGC ACA AAT ACT CAT GCC AGC TCA TTA CAG CAT GAG AAC AGC       983
167  Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser
168          275                 280                 285
169
170  AGT TTA TTA CTC ACT AAA GAC AGA ATG AAT GTA GAA AAG GCT GAA TTC      1031
171  Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe
172      290                 295                 300
173
174  TGT AAT AAA AGC AAA CAG CCT GGC TTA GCA AGG AGC CAA CAT AAC AGA      1079
175  Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg
176  305                 310                 315                 320
177
178  TGG GCT GGA AGT AAG GAA ACA TGT AAT GAT AGG CGG ACT CCC AGC ACA      1127
179  Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr
180                  325                 330                 335
181
182  GAA AAA AAG GTA GAT CTG AAT GCT GAT CCC CTG TGT GAG AGA AAA GAA      1175
183  Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu
184              340                 345                 350
185
186  TGG AAT AAG CAG AAA CTG CCA TGC TCA GAG AAT CCT AGA GAT ACT GAA      1223
187  Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu
188          355                 360                 365
189
190  GAT GTT CCT TGG ATA ACA CTA AAT AGC AGC ATT CAG AAA GTT AAT GAG      1271
191  Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu
192      370                 375                 380
193
194  TGG TTT TCC AGA AGT GAT GAA CTG TTA GGT TCT GAT GAC TCA CAT GAT      1319
195  Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp
196  385                 390                 395                 400
197
198  GGG GAG TCT GAA TCA AAT GCC AAA GTA GCT GAT GTA TTG GAC GTT CTA      1367
199  Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu
200                  405                 410                 415
201
202  AAT GAG GTA GAT GAA TAT TCT GGT TCT TCA GAG AAA ATA GAC TTA CTG      1415
203  Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu
204              420                 425                 430
205
```

```
206  GCC AGT GAT CCT CAT GAG GCT TTA ATA TGT AAA AGT GAA AGA GTT CAC      1463
207  Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Glu Arg Val His
208          435                 440                 445
209
210  TCC AAA TCA GTA GAG AGT AAT ATT GAA GAC AAA ATA TTT GGG AAA ACC      1511
211  Ser Lys Ser Val Glu Ser Asn Ile Glu Asp Lys Ile Phe Gly Lys Thr
212      450                 455                 460
213
214  TAT CGG AAG AAG GCA AGC CTC CCC AAC TTA AGC CAT GTA ACT GAA AAT      1559
215  Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn
216  465                 470                 475                 480
217
218  CTA ATT ATA GGA GCA TTT GTT ACT GAG CCA CAG ATA ATA CAA GAG CGT      1607
219  Leu Ile Ile Gly Ala Phe Val Thr Glu Pro Gln Ile Ile Gln Glu Arg
220                  485                 490                 495
221
222  CCC CTC ACA AAT AAA TTA AAG CGT AAA AGG AGA CCT ACA TCA GGC CTT      1655
223  Pro Leu Thr Asn Lys Leu Lys Arg Lys Arg Arg Pro Thr Ser Gly Leu
224              500                 505                 510
225
226  CAT CCT GAG GAT TTT ATC AAG AAA GCA GAT TTG GCA GTT CAA AAG ACT      1703
227  His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr
228          515                 520                 525
229
230  CCT GAA ATG ATA AAT CAG GGA ACT AAC CAA ACG GAG CAG AAT GGT CAA      1751
231  Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln
232      530                 535                 540
233
234  GTG ATG AAT ATT ACT AAT AGT GGT CAT GAG AAT AAA ACA AAA GGT GAT      1799
235  Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp
236  545                 550                 555                 560
237
238  TCT ATT CAG AAT GAG AAA AAT CCT AAC CCA ATA GAA TCA CTC GAA AAA      1847
239  Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys
240                  565                 570                 575
241
242  GAA TCT GCT TTC AAA ACG AAA GCT GAA CCT ATA AGC AGC AGT ATA AGC      1895
243  Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser
244              580                 585                 590
245
246  AAT ATG GAA CTC GAA TTA AAT ATC CAC AAT TCA AAA GCA CCT AAA AAG      1943
247  Asn Met Glu Leu Glu Leu Asn Ile His Asn Ser Lys Ala Pro Lys Lys
248          595                 600                 605
249
250  AAT AGG CTG AGG AGG AAG TCT TCT ACC AGG CAT ATT CAT GCG CTT GAA      1991
251  Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Ala Leu Glu
252      610                 615                 620
253
254  CTA GTA GTC AGT AGA AAT CTA AGC CCA CCT AAT TGT ACT GAA TTG CAA      2039
255  Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln
256  625                 630                 635                 640
257
258  ATT GAT AGT TGT TCT AGC AGT GAA GAG ATA AAG AAA AAA AAG TAC AAC      2087
```

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION ***US/08/488,011***

DATE: 08/17/95
TIME: 10:53:37

INPUT SET: S5614.raw

Line	Error	Original Text
36	Wrong application Serial Number	(A) APPLICATION NUMBER: US
932	Stop Codon at end of sequence removed - no error	

MAIL ROOM
DEC
1
1995
PTO

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. patent application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011 ***Attn:*** *BOX MISSING PART*

Filed: 07 June 1995

#5

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

TRANSMITTAL OF FORMAL DOCUMENTS

Hon. Commissioner of Patents and Trademarks
Washington, D.C. 20231

Sir:

In response to the Notice to File Missing Parts mailed 18 August 1995, submitted herewith are the following:

(1) Four executed Declarations and Powers of Attorney;
(2) Three executed Assignment documents, with cover sheets; and
(3) A Statement Pursuant to 37 CFR §1.821(f), including a computer readable copy of the Sequence Listing;
(4) A copy of Form PTO-1533; and
(5) A copy of the Notice to Comply (Applicants note that the application as filed contained an initial paper copy of the Sequence Listing).

Please charge the combined filing/surcharge fee to our Deposit Account No. 22-0261. This transmittal letter is submitted in triplicate.

Respectfully submitted,

Jeffrey L. Ihnen, Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, *LLP*
1201 New York Avenue, N.W., #1000
Washington, D.C. 20005
(202) 962-4810
Atty. Docket No. 24884-109347-B
Dated: 1 December 1995
647.01

E.O.T.

MAIL ROOM
DEC 1
1995

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re patent application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011

Filed: 07 June 1995

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

Assignment To D

PETITION FOR EXTENSION OF TIME

Hon. Commissioner of Patents and Trademarks
Washington, D.C. 20231

Sir:

In response to the Notice to File Missing Parts mailed 18 August 1995, petition is hereby made for an extension of three (3) months in which to reply. Please charge the requisite extension fee of $900.00 to our Deposit Account No. 22-0261. This petition is submitted in triplicate.

Respectfully submitted,

Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, *LLP*
1201 New York Avenue, N.W., #1000
Washington, D.C. 20005
(202) 962-4810
Atty. Docket No. 24884-109347-B
Dated: 1 December 1995
1125.01

290 KH 22-0261 12/28/95 08488011
29014 117 900.00CH

MAIL ROOM
DEC
1
1995

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. patent application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011

Filed: 07 June 1995

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

PETITION FOR EXTENSION OF TIME

Hon. Commissioner of Patents and Trademarks
Washington, D.C. 20231

Sir:

In response to the Notice to File Missing Parts mailed 18 August 1995, petition is hereby made for an extension of three (3) months in which to reply. Please charge the requisite extension fee of $900.00 to our Deposit Account No. 22-0261. This petition is submitted in triplicate.

Respectfully submitted,

Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, *LLP*
1201 New York Avenue, N.W., #1000
Washington, D.C. 20005
(202) 962-4810
Atty. Docket No. 24884-109347-B
Dated: 1 December 1995
1125.01

MAIL ROOM
DEC
1
1995

DECLARATION AND POWER OF ATTORNEY

#5

As a named inventor, I hereby declare that my residence, post office address and citizenship are as stated below next to my name; I believe I am an original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE, the specification of which was filed on 7 June 1995 under Serial No. 08/488,011.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) on which priority is claimed:

NONE

I hereby claim the benefit under Title 35, United States Code, §120 of any United States applications that are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which occurred between the filing date of the prior applications and the national or PCT international filing date of this application:

U.S. Serial No. 08/409,305, filed on 24 March 1995

U.S. Serial No. 08/348,824, filed on 24 November 1994

U.S. Serial No. 08/308,104, filed on 16 September 1994

U.S. Serial No. 08/300,266, filed on 2 September 1994

U.S. Serial No. 08/289,221, filed on 12 August 1994

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

I hereby appoint Jeffrey L. Ihnen, Registration No. 28,957; E. Brendan Magrab, Registration No. 36,205; and Stephen A. Saxe, Registration No. 38,609; all of Venable, Baetjer, Howard & Civiletti, whose address is 1201 New York Avenue, N.W., Suite 1000, Washington, D.C. 20005, telephone number (202) 962-4800, my attorneys and/or agents with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

Please direct all correspondence and telephone calls to Venable, Baetjer, Howard & Civiletti at the above address and telephone number.

Full Name of Sole or First Inventor	Mark H. SKOLNICK
Inventor's Signature	Mark Skolnick
Date of Signature	Oct 9, 1995
Residence Address	460 Northmount Way Salt Lake City, Utah 84103
Post Office Address	same
Country of Citizenship	United States

Full Name of Second Inventor	David E. GOLDGAR
Inventor's Signature	David E Goldgar
Date of Signature	October 9 1995
Residence Address	2707 East Kelly Lane Salt Lake City, Utah 84117
Post Office Address	same
Country of Citizenship	United States

Full Name of Third Inventor: Yoshio MIKI

Inventor's Signature: Yoshio Miki

Date of Signature: 10/13/95

Residence Address: 40 South 900 East
Salt Lake City, Utah 84102

Post Office Address: same

Country of Citizenship: Japan

Full Name of Fourthh Inventor: Jeff SWENSON

Inventor's Signature: _____

Date of Signature: _____

Residence Address: 745 West Creek Bend, #2807
Salt Lake City, Utah 84117

Post Office Address: same

Country of Citizenship: United States

Full Name of Fifth Inventor: Alexander KAMB

Inventor's Signature: [signature]

Date of Signature: 10/9/95

Residence Address: 1103 East 600 South
Salt Lake City, Utah 84102

Post Office Address: same

Country of Citizenship: United States

Full Name of Sixth Inventor	Keith D. HARSHMAN
Inventor's Signature	
Date of Signature	10-9-95
Residence Address	159 North F Street Salt Lake City, Utah 841032
Post Office Address	same
Country of Citizenship	United States

Full Name of Seventh Inventor	Donna M. SHATTUCK-EIDENS
Inventor's Signature	
Date of Signature	10.9.95
Residence Address	1874 Texas Street Salt Lake City, Utah 84108
Post Office Address	same
Country of Citizenship	United States

Full Name of Eighth Inventor	Sean V. TAVTIGIAN
Inventor's Signature	
Date of Signature	10/9/95
Residence Address	557 East 1st Ave., #3 Salt Lake City, Utah 841023
Post Office Address	same
Country of Citizenship	United States

Full Name of Ninth Inventor	Roger W. WISEMAN
Inventor's Signature	_____
Date of Signature	_____
Residence Address	515 Woodwinds Drive Durham, North Carolina 27713
Post Office Address	same
Country of Citizenship	United States

Full Name of Tenth Inventor	P. Andrew FUTREAL
Inventor's Signature	_____
Date of Signature	_____
Residence Address	220 East Markham Ave. Durham, North Carolina 27701
Post Office Address	same
Country of Citizenship	United States

I hereby appoint Jeffrey L. Ihnen, Registration No. 28,957; E. Brendan Magrab, Registration No. 36,205; and Stephen A. Saxe, Registration No. 38,609; all of Venable, Baetjer, Howard & Civiletti, whose address is 1201 New York Avenue, N.W., Suite 1000, Washington, D.C. 20005, telephone number (202) 962-4800, my attorneys and/or agents with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

Please direct all correspondence and telephone calls to Venable, Baetjer, Howard & Civiletti at the above address and telephone number.

Full Name of Sole or First Inventor	Mark H. SKOLNICK
Inventor's Signature	_____
Date of Signature	_____
Residence Address	460 Northmount Way Salt Lake City, Utah 84103
Post Office Address	same
Country of Citizenship	United States

Full Name of Second Inventor	David E. GOLDGAR
Inventor's Signature	_____
Date of Signature	_____
Residence Address	2707 East Kelly Lane Salt Lake City, Utah 84117
Post Office Address	same
Country of Citizenship	United States

Full Name of Third Inventor: Yoshio MIKI

Inventor's Signature: _____

Date of Signature: _____

Residence Address: 40 South 900 East
Salt Lake City, Utah 84102

Post Office Address: same

Country of Citizenship: Japan

Full Name of Fourthh Inventor: Jeff SWENSON

Inventor's Signature: _____

Date of Signature: _____

Residence Address: 745 West Creek Bend, #2807
Salt Lake City, Utah 84117

Post Office Address: same

Country of Citizenship: United States

Full Name of Fifth Inventor: Alexander KAMB

Inventor's Signature: _____

Date of Signature: _____

Residence Address: 1103 East 600 South
Salt Lake City, Utah 84102

Post Office Address: same

Country of Citizenship: United States

Full Name of Sixth Inventor	Keith D. HARSHMAN
Inventor's Signature	_____
Date of Signature	_____
Residence Address	159 North F Street Salt Lake City, Utah 841032
Post Office Address	same
Country of Citizenship	United States

Full Name of Seventh Inventor	Donna M. SHATTUCK-EIDENS
Inventor's Signature	_____
Date of Signature	_____
Residence Address	1874 Texas Street Salt Lake City, Utah 84108
Post Office Address	same
Country of Citizenship	United States

Full Name of Eighth Inventor	Sean V. TAVTIGIAN
Inventor's Signature	_____
Date of Signature	_____
Residence Address	557 East 1st Ave., #3 Salt Lake City, Utah 841023
Post Office Address	same
Country of Citizenship	United States

MAIL ROOM
DEC
1
1995
PAT. & TRADEMARK OFF.

DECLARATION AND POWER OF ATTORNEY

As below-named inventor, I hereby declare that my residence, post office address and citizenship are as stated below next to my name; I believe I am an original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE, the specification of which was filed on 7 June 1995 under Serial No. 08/488,011.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) on which priority is claimed:

NONE

I hereby claim the benefit under Title 35, United States Code, §120 of any United States applications that are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which occurred between the filing date of the prior applications and the national or PCT international filing date of this application:

U.S. Serial No. 08/409,305, filed on 24 March 1995

U.S. Serial No. 08/348,824, filed on 24 November 1994

U.S. Serial No. 08/308,104, filed on 16 September 1994

U.S. Serial No. 08/300,266, filed on 2 September 1994

U.S. Serial No. 08/289,221, filed on 12 August 1994

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

I hereby appoint Jeffrey L. Ihnen, Registration No. 28,957; E. Brendan Magrab, Registration No. 36,205; and Stephen A. Saxe, Registration No. 38,609; all of Venable, Baetjer, Howard & Civiletti, whose address is 1201 New York Avenue, N.W., Suite 1000, Washington, D.C. 20005, telephone number (202) 962-4800, my attorneys and/or agents with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

Please direct all correspondence and telephone calls to Venable, Baetjer, Howard & Civiletti at the above address and telephone number.

Full Name of Sole or First Inventor	Mark H. SKOLNICK
Inventor's Signature	_____
Date of Signature	_____
Residence Address	460 Northmount Way Salt Lake City, Utah 84103
Post Office Address	same
Country of Citizenship	United States

Full Name of Second Inventor	David E. GOLDGAR
Inventor's Signature	_____
Date of Signature	_____
Residence Address	2707 East Kelly Lane Salt Lake City, Utah 84117
Post Office Address	same
Country of Citizenship	United States

Full Name of Third Inventor	Yoshio MIKI
Inventor's Signature	_____
Date of Signature	_____
Residence Address	40 South 900 East Salt Lake City, Utah 84102
Post Office Address	same
Country of Citizenship	Japan

Full Name of Fourthh Inventor	Jeff SWENSON
Inventor's Signature	_____
Date of Signature	_____
Residence Address	745 West Creek Bend, #2807 Salt Lake City, Utah 84117
Post Office Address	same
Country of Citizenship	United States

Full Name of Fifth Inventor	Alexander KAMB
Inventor's Signature	_____
Date of Signature	_____
Residence Address	1103 East 600 South Salt Lake City, Utah 84102
Post Office Address	same
Country of Citizenship	United States

Full Name of Sixth Inventor	Keith D. HARSHMAN
Inventor's Signature	_____
Date of Signature	_____
Residence Address	159 North F Street Salt Lake City, Utah 841032
Post Office Address	same
Country of Citizenship	United States

Full Name of Seventh Inventor	Donna M. SHATTUCK-EIDENS
Inventor's Signature	_____
Date of Signature	_____
Residence Address	1874 Texas Street Salt Lake City, Utah 84108
Post Office Address	same
Country of Citizenship	United States

Full Name of Eighth Inventor	Sean V. TAVTIGIAN
Inventor's Signature	_____
Date of Signature	_____
Residence Address	557 East 1st Ave., #3 Salt Lake City, Utah 841023
Post Office Address	same
Country of Citizenship	United States

Full Name of Ninth Inventor	Roger W. WISEMAN
Inventor's Signature	_____
Date of Signature	_____
Residence Address	515 Woodwinds Drive Durham, North Carolina 27713
Post Office Address	same
Country of Citizenship	United States

Full Name of Tenth Inventor	P. Andrew FUTREAL
Inventor's Signature	_____
Date of Signature	_____
Residence Address	220 East Markham Ave. Durham, North Carolina 27701
Post Office Address	same
Country of Citizenship	United States

MAIL ROOM
DEC
1
1995

DECLARATION AND POWER OF ATTORNEY

As below named inventor, I hereby declare that my residence, post office address and citizenship are as stated below next to my name; I believe I am an original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE, the specification of which was filed on 7 June 1995 under Serial No. 08/488,011.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) on which priority is claimed:

NONE

I hereby claim the benefit under Title 35, United States Code, §120 of any United States applications that are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which occurred between the filing date of the prior applications and the national or PCT international filing date of this application:

U.S. Serial No. 08/409,305, filed on 24 March 1995

U.S. Serial No. 08/348,824, filed on 24 November 1994

U.S. Serial No. 08/308,104, filed on 16 September 1994

U.S. Serial No. 08/300,266, filed on 2 September 1994

U.S. Serial No. 08/289,221, filed on 12 August 1994

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office

08/488,011 01/06/96 SKOLNICK [illegible] 20584-10

COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

APPLICATION NUMBER	FILING DATE	FIRST NAMED APPLICANT	ATTORNEY DOCKET NO./TITLE
		[illegible]430	

VENABLE BAETJER HOWARD AND CIVILETTI
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

EXAMINER	
1803	

ART UNIT	PAPER NUMBER
	04/30/96

#6

DATE MAILED:

NOTICE OF FILING DATE GRANTED INCOMPLETE RESPONSE

Dear Sir,

Applicant's response of 08/488011 has been entered into the record. However, the application still remains incomplete for the following reason(s)

___ 1) The filing fee has not been received by the Office. The amount of _____ is due.

___ 2) The surcharge of _____ has not been received by the Office.

___ 3) The declaration was not received by the Office.

___ 4) The inventor's signature is missing on the Declaration

___ 5) The declaration must be signed by the inventor or a qualified person designated in the application.

X 6) Inventor number 3-00 signature is missing on the declaration.

___ 7) Signature of legal representative to process the application of a deceased applicant.

The response period remains as set forth in the office letter mailed on _____. However, you can obtain EXTENSIONS OF TIME under the provisions of 37 CFR 1.136 by filing a petition accompanied by the appropriate fee (37 CFR 1.17).

Please submit your response to the Special Processing Branch.

A copy of this notice MUST be returned with your response

SPECIAL PROCESSING AND CORRESPONDENCE BRANCH
ONAR, APPLICATION PROCESSING DIVISION
(703) 308-1202

RECEIVED
MAY 06 1996
By

PTOL-470 (REV. 12-90) PART I - OFFICE COPY

GeneDX 1009, pg. 228

UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR		ATTORNEY DOCKET NO.
08/488,011	01/05/96	SKOLNICK	M	24884-109347

18M2/0724

VENABLE BAETJER HOWARD AND CIVILETTI
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

EXAMINER	
REES,	

ART UNIT	PAPER NUMBER
1807	7

DATE MAILED: 07/24/96

Please find below and/or attached an Office communication concerning this application or proceeding.

Commissioner of Patents and Trademarks

PTO-90C (REV. 2/95)

1 - File Copy

Office Action Summary	Application No. 08/488,011	Applicant(s) SKOLNICK ET AL.
	Examiner Dianne Rees	Group Art Unit 1807

☒ Responsive to communication(s) filed on *Dec 1, 1995*.

☐ This action is **FINAL.**

☐ Since this application is in condition for allowance except for formal matters, **prosecution as to the merits is closed** in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11; 453 O.G. 213.

A shortened statutory period for response to this action is set to expire *3* month(s), or thirty days, whichever is longer, from the mailing date of this communication. Failure to respond within the period for response will cause the application to become abandoned. (35 U.S.C. § 133). Extensions of time may be obtained under the provisions of 37 CFR 1.136(a).

Disposition of Claims

☒ Claim(s) *1-27* is/are pending in the application.

Of the above, claim(s) _____ is/are withdrawn from consideration.

☐ Claim(s) _____ is/are allowed.

☒ Claim(s) *1-27* is/are rejected.

☐ Claim(s) _____ is/are objected to.

☐ Claims _____ are subject to restriction or election requirement.

Application Papers

☒ See the attached Notice of Draftsperson's Patent Drawing Review, PTO-948.

☐ The drawing(s) filed on _____ is/are objected to by the Examiner.

☐ The proposed drawing correction, filed on _____ is ☐ approved ☐ disapproved.

☐ The specification is objected to by the Examiner.

☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. § 119

☐ Acknowledgement is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d).

☐ All ☐ Some* ☐ None of the CERTIFIED copies of the priority documents have been

☐ received.

☐ received in Application No. (Series Code/Serial Number) _____.

☐ received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

*Certified copies not received: _____.

☐ Acknowledgement is made of a claim for domestic priority under 35 U.S.C. § 119(e).

Attachment(s)

☐ Notice of References Cited, PTO-892

☐ Information Disclosure Statement(s), PTO-1449, Paper No(s). _____

☐ Interview Summary, PTO-413

☒ Notice of Draftsperson's Patent Drawing Review, PTO-948

☐ Notice of Informal Patent Application, PTO-152

--- SEE OFFICE ACTION ON THE FOLLOWING PAGES ---

Part III DETAILED ACTION

Specification

1. The disclosure is objected to because of the following informalities: The specification refers to "our Utah collaborators" without providing names (a reference or a footnote for example) to indicate who these collaborators are. Appropriate correction is required.

2. The following is a quotation of the first paragraph of 35 U.S.C. § 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

The specification is objected to under 35 U.S.C. § 112, first paragraph, as failing to enable a diagnostic test for breast cancer.

Factors to be considered in determining whether a disclosure meets the enablement requirement of 35 USC 112, first paragraph, have been described in *In Re Colianni*, 195 USPQ 150 (CCPA 1977) and have been adopted by the Board of Patent Appeals and Interferences in *Ex Parte Forman*, 230 USPQ 546 (BPAI 1986). Among these factors are: the nature of the invention, the state of the prior art, the predictability or lack thereof in the art, the breadth of the claims, the amount of direction or guidance present, and the presence or absence of working examples.

The claims are drawn to methods of "diagnosing a predisposition for breast and ovarian cancer at the BRCA1 gene" in a human subject by detecting alterations in the sequence of the BRCA1 gene or its expression products (RNA or protein) by a variety of assays which include detecting alterations in the regulatory regions of BRCA1, determining shifts in electrophoretic mobility of single stranded DNA, by general hybridization assays which may further include the use of allele specific oligos as probes, amplification assays which may be further combined with hybridization assays with probes directed against wild type or mutant sequences, cloning of BRCA1 sequences, detection of mismatches, in situ hybridization, immunoblotting, immunohistochemistry, "assaying binding interactions between BRCA1 protein and its ligand", assaying the "inhibition of biochemical activity of a binding partner of BRCA1". Additional embodiments of the invention are recited in claims which are drawn to confirming "the lack of a predisposition for cancer at the BRCA1 locus". This assay may include the detection of mRNA products of BRCA1 or BRCA1 polypeptides.

The claims are inclusive of the detection of any breast and ovarian cancer and the methods of the claims have a very specific outcome- providing for a diagnosis of either the presence or absence of a predisposition to cancer associated with

an alteration in the BRCA1 gene. A subset of the claims are inclusive of the detection of somatic as well as germline mutations in BRCA1. A subset of the claims are also inclusive of detecting a lack of predisposition of any cancer "at the BRCA locus" wherein the presence of a wild type germline sequence indicates no predisposition to cancer at the BRCA1 locus.

The specification defines diagnosis or prognosis as "the classification of lesions as neoplasia, the determination of the severity of neoplasia or the monitoring of disease progression prior to, during and after treatment" . This definition does not distinguish between "diagnosis" and "prognosis" (page 25) so this examiner's interpretation of diagnosis as recited is based on "the classification of lesions as neoplasia" given that this is closest to the definition of diagnosis in the art (see Dorland's medical Dictionary). The specification also discloses that "the methods of diagnosis of the present invention are applicable to any tumor in which BRCA1 has a role in tumorigenesis." (page 21). "The diagnostic method of the present invention is useful for clinicians , so that they can decide upon the appropriate course of treatment". The specification discloses there may be a role for BRCA1 in breast, ovarian, colorectal, and prostate cancer.

The specification does not enable a *diagnostic* method for the detection of breast or ovarian cancers.

The specification and the art teach that multiple genes are likely to contribute to the etiology of breast and ovarian cancer. In addition to BRCA1, other genes such as BRCA2, (Wooster et al. Science 265:2088-2090, 1994), Prohibitin, (Sato et al. Cancer Research 52: 1643-1646, 1992), p53 (Malkin et al, Science 250: 1233-1238), and genes identified by as yet unmapped mutations, play a role either singly or in combination. Further the genetic influences on late onset breast cancer are yet uncharacterized (page 3).

The art teaches that BRCA1 mutations have not been found in sporadic breast cancer (see Boyd, Nature Genetics 9: 335-336, 1995) and in less than half the families displaying dominant transmission of breast cancer have mutations in BRCA1 been found. The art does teach that BRCA1 may be implicated in ovarian cancers, in that there is a higher overall rate of ovarian cancer in individuals with BRCA1 mutations, however the incidence of ovarian cancer varies unpredictably with the nature of the mutant allele (see Merajver et al,. Nature Genetics 9: 439-443). The art teaches that "the molecular basis for heterogeneity of penetrance, organ of origin of the cancer, age of onset and laterality of the lesions has not yet been elucidated (see page 439). Given the unpredictable utility of mutant alleles of the BRCA1 gene as diagnostic indicators, the specification is looked to provide guidance and/or examples to enable the breadth of the claims.

The specification teaches mutations in the BRCA1 gene in one individual who carries a predisposing haplotype in each of 8 kindreds. 23 coding exons and associated splice junctions were amplified from either genomic DNA samples or from cDNA prepared from lymphocyte mRNA. Amplified DNA sequences were then compared with the wild type sequences and 4 of 8 were found to contain an identifiable sequence variant. All of the four sequence variants are heterozygous and each appeared in only one of the kindreds, (page 81). However each of the four kindreds includes at least one woman who carries the mutation who lived until age 80 without developing a malignancy (page 84). The specification further teaches the additional screening of 32 breast tumors and 12 ovarian tumors obtained from individuals of mixed race and age of onset for BRCA1 mutations. The complete 5589 bp coding region and intron-exon boundary sequences of the gene were screened in this tumor set by direct sequencing alone or by an combination of SSCP and direct sequencing. Among the tumors analyzed , six mutations were found: one in ovarian tumors, 5 in breast tumors (two of which were identical)(page 85, Table 12, page 86).
The specification teaches therefore that somatic mutations in BRCA1 are rare. Among the possibilities raised for the paucity of BRCA1 mutations in somatic cells are the insensitivity of the assays used, the presence of mutations in noncoding regions, and LOH events in 17q that do not reflect BRCA1 mutations (page 87). Interactions with other loci are also a distinct probability.

The specification raises relevant criteria for the assessment of a predisposing mutation:

1) The variant must be present in individuals from the kindred who carry the predisposing BRCA1 haplotype and absent in other members of the kindred, and 2) be rare in the general population. For somatic mutations the same general rules apply "cosegregation of the mutant allele with the disease and absence of the mutant alleles in controls indicating a low gene frequency in the population". These criteria are met for the specific mutations that are taught, however it is noted that, as discussed above, the germline mutations are unique to specific kindreds taught and the somatic mutations are represented at low frequency in patient samples with breast tumors in general. The specification therefore does not provide for a diagnostic for susceptibility to breast or ovarian cancer, as clearly there would be a high false negative rate of mutations using even these specific alleles and a measurable false positive rate given the teaching of women who have specific mutations in BRCA1, yet at age 80 have not yet developed cancer (i.e to say that a test provides a diagnostic indicia of risk implies that some value can be put on this risk-some statistical probability that the individual will actually develop breast and/or ovarian cancer). This does not detract from the utility of the detection of these specific mutations in a screening assay where one is assessing risk among the contribution of many factors that might be

involved in the transition from a normal tissue to a neoplastic one; the claims therefore might be amended to overcome this rejection by reciting a screening assay for the specific mutant alleles disclosed.

The claims, however, are inclusive of diagnostic assays of a diverse population of alterations and mutations in BRCA1 sequences that may be manifested additionally at the mRNA or protein level. All the assays described and claimed are well known in the art and the specification provides a detailed review of methods of detection of altered sequences and proteins available to one of skill in the art. It is also taken that the skill in the art is high. Yet while there is no doubt that one may detect alterations in the BRCA1 gene, there is considerable doubt, raised by both the art and the teachings of the specification, that a mutation thus isolated will be predictably diagnostic of a susceptibility to cancer, even cancer limited to breast or ovarian cancer. The criteria, as discussed above that a mutation will cosegregate with the phenotype of breast cancer and be rare in the general population would have to be met for each of the mutations identified. In the assessment of the experimentation required to determine the diagnostic utility of a given allele, it is not simply the amount of labor that one is required to perform , as it is a given that one of skill in the art would be able to screen affected kindreds and patient samples available from tumor registries, rather it is the

unpredictability that any mutation obtained from this screening would in fact be diagnostic, or even useful as a screening tool, that is considered. For any mutation identified, the enabling requirement of the claim is that the individual who is being tested will be provided with some indicia of what their risk is, what their susceptibility to breast or ovarian cancer is, or as other claims recite, what their lack of risk is -the specification does not provide this enablement for any generic mutation in BRCA1 and it is not predictable as discussed above which mutations will be predisposing and which mutations will not. It is this unpredictability that makes this experimentation undue.

Enablement issues related to the specific assays claimed are now considered:

A variety of hybridization assays are disclosed to detect alterations in BRCA1 sequences using BRCA1 gene probes. The specification defines that BRCA1 genes, nucleic acids and polynucleotides , are those "which are likely to be expressed in normal tissue, certain alleles of which predispose an individual to develop breast, ovarian, colorectal, and prostate cancers. Mutations in the BRCA1 locus may be involved in the initiation and/or progression of other types of tumors. The locus is indicated in part by mutations that predispose individuals to developing cancer. These mutations fall within the BRCA1 region described infra. The BRCA1 locus is intended to include coding

sequences, intervening sequences, and regulatory sequences controlling transcription or translation. The BRCA1 locus is intended to include all allelic variations of the DNA sequence" (see page 25, 26).

Coding sequences are taught (figure 1,2) by the specification, as well as intron-exon boundaries, however the specification discloses only a limited number of variants of BRCA1 while intending clearly to encompass all variants. The design of allele specific probes therefore is enabled only for the specific mutations disclosed as one of skill in the art would be unable to envisage what other sequences or alterations could be predictably used in screening or diagnostic assays. The specification does not provide teaching as to the extent or structure of regulatory regions of the BRCA1 gene although claim 6 is directed to the detection of alterations in the regulatory regions of BRCA1 . While there is limited information provided as to the transcriptional start sites of the gene in somatic tissues and testes, no other characterization of the promoter region is provided nor is any information provided as to potential enhancer sequences upstream or even potentially within intronic sequences. Thus one of skill in the art has no guidance as to what sequences to analyze to determine whether alterations are present in regulatory regions.

Additional claims are directed to the detection of alterations in BRCA1 transcripts (by hybridization assays which

may include in situ hybridization assays, or by the detection of mismatches) and it is noted that these assays may provide an indirect means to assess a mutation in a regulatory region. The specification teaches that "Northern blots using BRCA1 cDNA as probes identified a single transcript of about 7.8 kb, present most abundantly in breast, thymus and testes (figure 7). Four alternatively spliced products were observed as independent cDNA clones; 3 of these were detected in breast and 2 in ovary mRNA (Fig 6). A PCR survey from different tissues further supports the idea that there is considerable heterogeneity near the 5′ end of the gene. The specification teaches that there are six *potential* alternate splice donors in the 5′ untranslated region, with the longest deletion being 1,155 bp. The predominant form of BRCA1 protein in breast and ovary lacks exon 4 (Seq. ID, 11 and 12).

The data provided in Figure 7 is difficult to assess given the reproduction provided in the instant specification. In the figure provided, no band is visible in breast tissue, and a smear is observed in testes starting at about 7.8 kb, while two discrete bands are observed in thymus and a faint smear is detected in ovaries. It is difficult therefore to assess what wild type levels of BRCA transcripts are and further no teaching is provided as to what levels or alterations of BRCA1 transcripts are found in individuals with even BRCA1 linked breast cancer. Thus the specification has not enabled a diagnostic assay based

on the detection of alterations in BRCA1 transcripts other than to say generally that one might perform such an assay. The specification would enable a screening assay for a subset of the specific mutations identified based on an altered transcription pattern given that these mutant genes might be reliably predicted to have alterations in transcripts (either an absence or an altered size) and would enable the detection of mismatches for the remainder.

The BRCA polypeptide is similarly defined in broad terms in the specification as referring to "a protein or polypeptide encoded by the BRCA1 locus, variants, or fragments thereof. The term polypeptide refers to a polymer of amino acids and its equivalent and does not refer to a specific length of product; thus peptides, oligopeptides, and proteins are included within the definition of a polypeptide."" ...Ordinarily such polypeptides will be at least 50% homologous to the native BRCA1 sequence, preferably in excess of 90% and more preferably at least about 95% homologous". The specification also states that "Also included are proteins encoded by DNA which hybridize under high or low stringency conditions, to BRCA nucleic acids and closely related polypeptides or proteins retrieved by antisera to the BRCA1 protein (s). (page 28) (however it is noted that zooblot analysis of BRCA1 sequences only teaches that there is cross reactivity at low stringencies between BRCA1 and sequences in other species. No further characterization of these sequences

are provided and no teaching as to related proteins are provided).

The specification provides no teaching as to the characterization of either the wild type or the mutant protein other than to provide teachings as to the predicted amino acid sequence of the wild type BRCA1 protein and the predicted amino acid alterations in the mutants. On page 23, the specification teaches prophetical examples of the use of antibodies and immunoassays -however no teaching is provided as to the actual molecular weight or number of species of proteins identified by such antibodies. In other words the specification does not clearly establish what a baseline level for normal expression of BRCA1 is and therefore it is not clear what alterations of BRCA1 protein expression would be indicative of pathology. The generation of epitope specific antibodies to identify specific mutations is especially problematical, in that while it is true that the specification has identified regions the predicted BRCA1 protein sequence that would be altered in a subset of BRCA1 mutations and hence might be associated with an increased risk of neoplasia, it is not predictable whether peptides containing these sequences would in fact be immunogenic. It would require undue experimentation, again in view of the unpredictability of the research, to isolate such antibodies.

The specification does not enable assays to detect altered binding interactions between BRCA1 and its ligands. The scope of

this claim encompasses the detection of altered BRCA1 interactions with nucleic acid as well as protein ligands. The specification, while identifying a role for BRAC1 as a transcription factor based on its homology to zinc finger proteins, does not identify any sequences to which BRCA1 protein binds or genes containing these sequences. No teachings are provided as to the nature of potential protein ligands of BRAC1. Therefore one of skill in the art does not have guidance as to what ligands are to even be targeted much less what alterations in these ligands are to be indicative of altered BRCA1 function. The claims might are also inclusive of a detection of altered interactions between BRCA1 and an antibody. As discussed above no teachings of antibody interactions with mutant proteins are provided other than to state that one may determine such interactions.

Claims 24-28 are drawn to a method of confirming the lack of cancer "wherein the presence of a wild type sequence indicates no cancer at the BRCA1 gene". It is somewhat unclear how cancer "at the BRCA1 gene" is determined; the claim is interpreted as being drawn to cancer linked to mutations at the BRCA1 gene. Given the genetic complexity of cancer, the presence or absence of a wild type sequence is not necessarily diagnostic of the presence or absence of cancer. Further the diagnostic utility of an assay to determine the absence of neoplasia will be very much dependent on how the sequence of the wild type gene is determined. Thus there

might be many situations where one could envision an apparently wild type transcript or protein and still have a mutation in the BRCA1 gene. Thus claims whereby detection itself is the only step of the assay are not enabled by the teachings of the specification or the prior art. The only absolute way to make such a determination would be by obtaining the sequence of the entire BRCA1 gene in an individual, a daunting task given its large size.

It is therefore the position of the examiner that the teachings of the specification do not overcome doubts of record in the art concerning the enablement of BRCA1 probes (whether nucleic acid probes or proteins) as diagnostic indicators of neoplasia or an absence or neoplasia and that it would require undue experimentation to perform the methods of the claims as broadly written.

3. Claims 1-27 are rejected under 35 U.S.C. § 112, first paragraph, for the reasons set forth in the objection to the specification.

Claim Rejections - 35 USC § 112

4. Claims 1-27 are rejected under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

The following phrases render the claims vague and indefinite:

a) In claim 1, the recitation of "the recitation of comparing the germline sequence of the BRCA1 gene or its expression products in a tissue sample from said subject with the germline sequence of the wild type BRCA1 gene or its expression products" in that the recitation of "in a tissue sample" because the sequence is not compared "in the tissue sample per se" -i.e the phrase appears to be a nonsequitor since no further steps are recited where nucleic acids are actually isolated from the tissues.

b) In claim 7, the recitation of "wherein the alterations in the germline sequence is detected by observing shifts in electrophoretic mobility of single stranded DNA on non-denaturing polyacrylamide gels, renders the claim indefinite because it is not clear where the single stranded DNA is derived from in this claim and further what the point of comparison is to allow one to determine a "shift in electrophoretic mobility".

c) Claim 22 is indefinite in the recitation of "BRCA1 gene protein" as the claim appears to intend --BRCA gene product--. Claim 22 is further indefinite in the recitation of "assaying for binding interactions" in that it is not clear what "interactions" are actually being measured.

d) Claim 24 is indefinite in the recitation of "no cancer at the BRCA1 gene" as neoplasia occurs within the cells/tissues of an organism not at the gene per se.

5. Claims 7-23, 27 and 28 are rejected under 35 U.S.C. § 112, second paragraph, as being incomplete for omitting essential steps, such omission amounting to a gap between the steps. See MPEP § 706.03(f).

The claims all recite the detection of an alteration by a test (such as hybridization, immunoblotting, assaying binding interactions, inhibition of biochemical activity etc.) but fail to recite a step which one extrapolates from the results of the test to the detection of an alteration.

Double Patenting

A rejection based on double patenting of the "same invention" type finds its support in the language of 35 U.S.C. 101 which states that "whoever invents or discovers any new and useful process ... may obtain a patent therefor ..." (Emphasis added). Thus, the term "same invention," in this context, means an invention drawn to identical subject matter. *Miller v. Eagle Mfg. Co.*, 151 U.S. 186 (1894); *In re Ockert*, 245 F.2d 467, 114 USPQ 330 (CCPA 1957); and *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970).

The non-statutory double patenting rejection, whether of the obvious-type or non-obvious-type, is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent. *In re Thorington,* 418 F.2d 528, 163 USPQ 644 (CCPA 1969); *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); *In re Van Ornam*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); and *In re Goodman*, 29 USPQ2d 2010 (Fed. Cir. 1993).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321 (b) and (c) may be used to overcome an actual or provisional rejection based on a non-statutory double patenting ground provided the conflicting application or patent is shown to be commonly owned with this application. See 37 CFR 1.78 (d).

Effective January 1, 1994, a registered attorney or agent of record may sign a Terminal Disclaimer. A Terminal Disclaimer signed by the assignee must fully comply with 37 CFR 3.73(b).

Claims 1-23 are provisionally rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1-28 of copending application Serial No.

08487002. Although the conflicting claims are not identical, they are not patentably distinct from each other because the methods recited in 08487002 meet all of the limitations of claims 1-28 in the present application (i.e a method of diagnosing susceptibility to any cancer by detecting mutations in BRCA1 encompasses a method of detecting breast and ovarian cancer).

Claims 1-19 are provisionally rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 3-18 of copending application Serial No. 08409305. Although the conflicting claims are not identical, they are not patentably distinct from each other because the methods recited in 08409305 meet all of the limitations of claims 1-19 in the present application.

This is a *provisional* obviousness-type double patenting rejection because the conflicting claims have not in fact been patented.

The examiner notes that the above rejections are not an exhaustive list of examples of double patenting related to the above application and reminds the applicant that as the prosecution advances, conflicting claims in all related applications will have to be either cancelled, or changed in scope (with the filing of a terminal disclaimer).

The following references are cited as relevant intervening art:

Castilla et al. Nature Genetics 8: 387-391, Dec, 1994, teaches mutations in BRCA1 in families with early onset breast and ovarian cancer.

Friedman et al. Nature Genetics 8: 1-6, Dec 1994 teaches the confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families.

Simard et al. Nature Genetics 8: 392-397 teaches BRCA mutations in Canadian breast and ovarian cancer families.

Goldgar et al. J. of the National Cancer Institute 86: 200-209, Feb 2, 1994, teaches a large kindred with 17q Linked Breast and Ovarian Cancer.

Futreal et al. Science 266 120-122, 1994, teaches BRCA mutations in Primary Breast and Ovarian Carcinomas.

Hosking et al. Nature Genetics 9: 343-344, April 1995 teaches a somatic BRCA1 mutation in an ovarian tumor.

Holt et al.WO 95/19369 teaches a method detecting and diagnosing preinvasive breast cancer by identifying differentially expressed genes in early, pre invasive breast cancer tissue and methods for screening for compounds which induce the expression of BRCA1.

6. No claims are allowed.

7. Papers related to this application may be submitted to Group 1800 by facsimile transmission via the P.T.O. Fax Center located in Crystal Mall 1. The CM1 Fax Center number is (703) 305-7401. Please note that the faxing of such papers must conform with the notice to Comply published in the Official Gazette, 1096 OG 30 (Nov 15, 1989).

An inquiry regarding this communication should be directed to examiner Dianne Rees, Ph.D., whose telephone number is (703) 308-6565. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, W. Gary Jones, can be reached on (703) 308-1156.

Calls of a general nature may be directed to the Group receptionist who may be reached at (703) 308-0196.

Dianne Rees

Dianne Rees

July 8, 1996

B. L. Sisson

BRADLEY L. SISSON
PATENT EXAMINER
GROUP 1800

7-11-96

1800 5/24 1807 **TEAM 2**

RAW SEQUENCE LISTING
PATENT APPLICATION *US/08/488,011A*

DATE: 01/05/96
TIME: 14:21:20

INPUT SET: S8111.raw

This Raw Listing contains the General Information Section and those Sequences containing ERRORS.

#8

see p. 2 →

SEQUENCE LISTING

(1) General Information:

(i) APPLICANT: Skolnick, Mark H.
Goldgar, David E.
Miki, Yoshio
Swenson, Jeff
Kamb, Alexander
Harshman, Keith D.
Shattuck-Eidens, Donna M.
Tavtigian, Sean V.
Wiseman, Roger W.
Futreal, P. Andrew

(ii) TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer Susceptibility Gene

(iii) NUMBER OF SEQUENCES: 85

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
(B) STREET: 1201 New York Avenue, N.W., Suite 1000
(C) CITY: Washington
(D) STATE: DC
(E) COUNTRY: USA
(F) ZIP: 20005

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
--> **(A) APPLICATION NUMBER: US**
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/409,305
(B) FILING DATE: 24-MAR-1995

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/348,824

```
        (B) FILING DATE: 29-NOV-1994

    (vii) PRIOR APPLICATION DATA:
          (A) APPLICATION NUMBER: US 08/308,104
          (B) FILING DATE: 16-SEP-1994

    (vii) PRIOR APPLICATION DATA:
          (A) APPLICATION NUMBER: US 08/300,266
          (B) FILING DATE: 02-SEP-1994

    (vii) PRIOR APPLICATION DATA:
          (A) APPLICATION NUMBER: US 08/289,221
          (B) FILING DATE: 12-AUG-1994

   (viii) ATTORNEY/AGENT INFORMATION:
          (A) NAME: Ihnen, Jeffrey L.
          (B) REGISTRATION NUMBER: 28,957
          (C) REFERENCE/DOCKET NUMBER: 24884-109347

     (ix) TELECOMMUNICATION INFORMATION:
          (A) TELEPHONE: 202-962-4810
          (B) TELEFAX: 202-962-8300
```

ERRORED SEQUENCES FOLLOW:

```
(2) INFORMATION FOR SEQ ID NO:2:

      (i) SEQUENCE CHARACTERISTICS:
-->     (A) LENGTH: 1864 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: protein

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
 1               5                  10                  15

Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
            20                  25                  30

Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
        35                  40                  45

Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
    50                  55                  60
```

Only 1863 showed up. Is one missing?

```
598  Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
599   65                  70                  75                  80
600
601  Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
602                  85                  90                  95
603
604  Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
605              100                 105                 110
606
607  Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
608          115                 120                 125
609
610  Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
611      130                 135                 140
612
613  Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
614  145                 150                 155                 160
615
616  Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
617                  165                 170                 175
618
619  Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
620              180                 185                 190
621
622  Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
623          195                 200                 205
624
625  Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
626      210                 215                 220
627
628  Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
629  225                 230                 235                 240
630
631  Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg
632                  245                 250                 255
633
634  His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu
635              260                 265                 270
636
637  Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser
638          275                 280                 285
639
640  Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe
641      290                 295                 300
642
643  Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg
644  305                 310                 315                 320
645
646  Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr
647                  325                 330                 335
648
649  Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu
650              340                 345                 350
```

```
651
652  Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu
653          355                 360                 365
654
655  Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu
656      370                 375                 380
657
658  Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp
659  385                 390                 395                 400
660
661  Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu
662                  405                 410                 415
663
664  Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu
665              420                 425                 430
666
667
668  Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Glu Arg Val His
669          435                 440                 445
670
671  Ser Lys Ser Val Glu Ser Asn Ile Glu Asp Lys Ile Phe Gly Lys Thr
672      450                 455                 460
673
674  Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn
675  465                 470                 475                 480
676
677  Leu Ile Ile Gly Ala Phe Val Thr Glu Pro Gln Ile Ile Gln Glu Arg
678                  485                 490                 495
679
680  Pro Leu Thr Asn Lys Leu Lys Arg Lys Arg Arg Pro Thr Ser Gly Leu
681              500                 505                 510
682
683  His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr
684          515                 520                 525
685
686  Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln
687      530                 535                 540
688
689  Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp
690  545                 550                 555                 560
691
692  Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys
693                  565                 570                 575
694
695  Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser
696              580                 585                 590
697
698  Asn Met Glu Leu Glu Leu Asn Ile His Asn Ser Lys Ala Pro Lys Lys
699          595                 600                 605
700
701  Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Ala Leu Glu
702      610                 615                 620
703
```

```
704  Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln
705  625                 630                 635                 640
706
707  Ile Asp Ser Cys Ser Ser Ser Glu Glu Ile Lys Lys Lys Lys Tyr Asn
708                  645                 650                 655
709
710  Gln Met Pro Val Arg His Ser Arg Asn Leu Gln Leu Met Glu Gly Lys
711              660                 665                 670
712
713  Glu Pro Ala Thr Gly Ala Lys Lys Ser Asn Lys Pro Asn Glu Gln Thr
714          675                 680                 685
715
716  Ser Lys Arg His Asp Ser Asp Thr Phe Pro Glu Leu Lys Leu Thr Asn
717      690                 695                 700
718
719  Ala Pro Gly Ser Phe Thr Lys Cys Ser Asn Thr Ser Glu Leu Lys Glu
720  705                 710                 715                 720
721
722  Phe Val Asn Pro Ser Leu Pro Arg Glu Glu Lys Glu Glu Lys Leu Glu
723                  725                 730                 735
724
725  Thr Val Lys Val Ser Asn Asn Ala Glu Asp Pro Lys Asp Leu Met Leu
726              740                 745                 750
727
728  Ser Gly Glu Arg Val Leu Gln Thr Glu Arg Ser Val Glu Ser Ser Ser
729          755                 760                 765
730  Ile Ser Leu Val Pro Gly Thr Asp Tyr Gly Thr Gln Glu Ser Ile Ser
731      770                 775                 780
732
733  Leu Leu Glu Val Ser Thr Leu Gly Lys Ala Lys Thr Glu Pro Asn Lys
734  785                 790                 795                 800
735
736  Cys Val Ser Gln Cys Ala Ala Phe Glu Asn Pro Lys Gly Leu Ile His
737                  805                 810                 815
738
739  Gly Cys Ser Lys Asp Asn Arg Asn Asp Thr Glu Gly Phe Lys Tyr Pro
740              820                 825                 830
741
742  Leu Gly His Glu Val Asn His Ser Arg Glu Thr Ser Ile Glu Met Glu
743          835                 840                 845
744
745  Glu Ser Glu Leu Asp Ala Gln Tyr Leu Gln Asn Thr Phe Lys Val Ser
746      850                 855                 860
747
748  Lys Arg Gln Ser Phe Ala Pro Phe Ser Asn Pro Gly Asn Ala Glu Glu
749  865                 870                 875                 880
750
751  Glu Cys Ala Thr Phe Ser Ala His Ser Gly Ser Leu Lys Lys Gln Ser
752                  885                 890                 895
753
754  Pro Lys Val Thr Phe Glu Cys Glu Gln Lys Glu Glu Asn Gln Gly Lys
755              900                 905                 910
756
```

```
Asn Glu Ser Asn Ile Lys Pro Val Gln Thr Val Asn Ile Thr Ala Gly
        915                 920                 925

Phe Pro Val Val Gly Gln Lys Asp Lys Pro Val Asp Asn Ala Lys Cys
    930                 935                 940

Ser Ile Lys Gly Gly Ser Arg Phe Cys Leu Ser Ser Gln Phe Arg Gly
945                 950                 955                 960

Asn Glu Thr Gly Leu Ile Thr Pro Asn Lys His Gly Leu Leu Gln Asn
                965                 970                 975

Pro Tyr Arg Ile Pro Pro Leu Phe Pro Ile Lys Ser Phe Val Lys Thr
            980                 985                 990

Lys Cys Lys Lys Asn Leu Leu Glu Glu Asn Phe Glu Glu His Ser Met
        995                 1000                1005

Ser Pro Glu Arg Glu Met Gly Asn Glu Asn Ile Pro Ser Thr Val Ser
    1010                1015                1020

Thr Ile Ser Arg Asn Asn Ile Arg Glu Asn Val Phe Lys Glu Ala Ser
1025                1030                1035                1040

Ser Ser Asn Ile Asn Glu Val Gly Ser Ser Thr Asn Glu Val Gly Ser
                1045                1050                1055

Ser Ile Asn Glu Ile Gly Ser Ser Asp Glu Asn Ile Gln Ala Glu Leu
            1060                1065                1070
Gly Arg Asn Arg Gly Pro Lys Leu Asn Ala Met Leu Arg Leu Gly Val
        1075                1080                1085

Leu Gln Pro Glu Val Tyr Lys Gln Ser Leu Pro Gly Ser Asn Cys Lys
    1090                1095                1100

His Pro Glu Ile Lys Lys Gln Glu Tyr Glu Glu Val Val Gln Thr Val
1105                1110                1115                1120

Asn Thr Asp Phe Ser Pro Tyr Leu Ile Ser Asp Asn Leu Glu Gln Pro
                1125                1130                1135

Met Gly Ser Ser His Ala Ser Gln Val Cys Ser Glu Thr Pro Asp Asp
            1140                1145                1150

Leu Leu Asp Asp Gly Glu Ile Lys Glu Asp Thr Ser Phe Ala Glu Asn
        1155                1160                1165

Asp Ile Lys Glu Ser Ser Ala Val Phe Ser Lys Ser Val Gln Lys Gly
    1170                1175                1180

Glu Leu Ser Arg Ser Pro Ser Pro Phe Thr His Thr His Leu Ala Gln
1185                1190                1195                1200
```

```
Gly Tyr Arg Arg Gly Ala Lys Lys Leu Glu Ser Ser Glu Glu Asn Leu
                1205                1210                1215

Ser Ser Glu Asp Glu Glu Leu Pro Cys Phe Gln His Leu Leu Phe Gly
            1220                1225                1230

Lys Val Asn Asn Ile Pro Ser Gln Ser Thr Arg His Ser Thr Val Ala
        1235                1240                1245

Thr Glu Cys Leu Ser Lys Asn Thr Glu Glu Asn Leu Leu Ser Leu Lys
    1250                1255                1260

Asn Ser Leu Asn Asp Cys Ser Asn Gln Val Ile Leu Ala Lys Ala Ser
1265                1270                1275                1280

Gln Glu His His Leu Ser Glu Glu Thr Lys Cys Ser Ala Ser Leu Phe
                1285                1290                1295

Ser Ser Gln Cys Ser Glu Leu Glu Asp Leu Thr Ala Asn Thr Asn Thr
            1300                1305                1310

Gln Asp Pro Phe Leu Ile Gly Ser Ser Lys Gln Met Arg His Gln Ser
        1315                1320                1325

Glu Ser Gln Gly Val Gly Leu Ser Asp Lys Glu Leu Val Ser Asp Asp
    1330                1335                1340

Glu Glu Arg Gly Thr Gly Leu Glu Glu Asn Asn Gln Glu Glu Gln Ser
1345                1350                1355                1360

Met Asp Ser Asn Leu Gly Glu Ala Ala Ser Gly Cys Glu Ser Glu Thr
                1365                1370                1375

Ser Val Ser Glu Asp Cys Ser Gly Leu Ser Ser Gln Ser Asp Ile Leu
            1380                1385                1390

Thr Thr Gln Gln Arg Asp Thr Met Gln His Asn Leu Ile Lys Leu Gln
        1395                1400                1405

Gln Glu Met Ala Glu Leu Glu Ala Val Leu Glu Gln His Gly Ser Gln
    1410                1415                1420

Pro Ser Asn Ser Tyr Pro Ser Ile Ile Ser Asp Ser Ser Ala Leu Glu
1425                1430                1435                1440

Asp Leu Arg Asn Pro Glu Gln Ser Thr Ser Glu Lys Ala Val Leu Thr
                1445                1450                1455

Ser Gln Lys Ser Ser Glu Tyr Pro Ile Ser Gln Asn Pro Glu Gly Leu
            1460                1465                1470

Ser Ala Asp Lys Phe Glu Val Ser Ala Asp Ser Ser Thr Ser Lys Asn
        1475                1480                1485
```

```
863
864  Lys Glu Pro Gly Val Glu Arg Ser Ser Pro Ser Lys Cys Pro Ser Leu
865      1490                1495                1500
866
867  Asp Asp Arg Trp Tyr Met His Ser Cys Ser Gly Ser Leu Gln Asn Arg
868  1505                1510                1515                1520
869
870  Asn Tyr Pro Ser Gln Glu Glu Leu Ile Lys Val Val Asp Val Glu Glu
871                  1525                1530                1535
872
873  Gln Gln Leu Glu Glu Ser Gly Pro His Asp Leu Thr Glu Thr Ser Tyr
874              1540                1545                1550
875
876  Leu Pro Arg Gln Asp Leu Glu Gly Thr Pro Tyr Leu Glu Ser Gly Ile
877          1555                1560                1565
878
879  Ser Leu Phe Ser Asp Asp Pro Glu Ser Asp Pro Ser Glu Asp Arg Ala
880      1570                1575                1580
881
882  Pro Glu Ser Ala Arg Val Gly Asn Ile Pro Ser Ser Thr Ser Ala Leu
883  1585                1590                1595                1600
884
885  Lys Val Pro Gln Leu Lys Val Ala Glu Ser Ala Gln Ser Pro Ala Ala
886                  1605                1610                1615
887
888  Ala His Thr Thr Asp Thr Ala Gly Tyr Asn Ala Met Glu Glu Ser Val
889              1620                1625                1630
890
891  Ser Arg Glu Lys Pro Glu Leu Thr Ala Ser Thr Glu Arg Val Asn Lys
892          1635                1640                1645
893
894  Arg Met Ser Met Val Val Ser Gly Leu Thr Pro Glu Glu Phe Met Leu
895      1650                1655                1660
896
897  Val Tyr Lys Phe Ala Arg Lys His His Ile Thr Leu Thr Asn Leu Ile
898  1665                1670                1675                1680
899
900  Thr Glu Glu Thr Thr His Val Val Met Lys Thr Asp Ala Glu Phe Val
901                  1685                1690                1695
902
903  Cys Glu Arg Thr Leu Lys Tyr Phe Leu Gly Ile Ala Gly Gly Lys Trp
904              1700                1705                1710
905
906  Val Val Ser Tyr Phe Trp Val Thr Gln Ser Ile Lys Glu Arg Lys Met
907          1715                1720                1725
908
909  Leu Asn Glu His Asp Phe Glu Val Arg Gly Asp Val Val Asn Gly Arg
910      1730                1735                1740
911
912  Asn His Gln Gly Pro Lys Arg Ala Arg Glu Ser Gln Asp Arg Lys Ile
913  1745                1750                1755                1760
914
915  Phe Arg Gly Leu Glu Ile Cys Cys Tyr Gly Pro Phe Thr Asn Met Pro
```

```
916                 1765                1770                1775
917
918  Thr Asp Gln Leu Glu Trp Met Val Gln Leu Cys Gly Ala Ser Val Val
919              1780                1785                1790
920
921  Lys Glu Leu Ser Ser Phe Thr Leu Gly Thr Gly Val His Pro Ile Val
922          1795                1800                1805
923
924  Val Val Gln Pro Asp Ala Trp Thr Glu Asp Asn Gly Phe His Ala Ile
925      1810                1815                1820
926
927  Gly Gln Met Cys Glu Ala Pro Val Val Thr Arg Glu Trp Val Leu Asp
928  1825                1830                1835                1840
929
930  Ser Val Ala Leu Tyr Gln Cys Gln Glu Leu Asp Thr Tyr Leu Ile Pro
931                 1845                1850                1855
932
933  Gln Ile Pro His Ser His Tyr
934              1860
935
```

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION *US/08/488,011A*

DATE: 01/05/96
TIME: 14:22:03

INPUT SET: S8111.raw

Line	Error	Original Text
36	Wrong application Serial Number	(A) APPLICATION NUMBER: US
578	Entered (1864) and Calc. Seq. Length (1863) differ	(A) LENGTH: 1864 amino acids

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

MAIL ROOM
OCT
52 12
1995
PAT. & TRADEMARK OFF.

In re patent application of:

Mark H. SKOLNICK et al.

Serial No.: 08/488,011 Group Art Unit:

Filed: 7 June 1995

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

INFORMATION DISCLOSURE STATEMENT

RECEIVED
MAY 2 1 1996
GROUP 1800

Hon. Commissioner of Patents
and Trademarks
Washington, DC 20231

Sir:

Transmitted herewith for the Examiner's consideration are copies of the references listed below, which are believed relevant to the above invention, along with a completed Form PTO-1449 listing the references. The Examiner is requested to return an initialed copy of this form indicating his review of the same.

Domestic Patents

1. 5,236,844 Basset, P., et al. Issued 08-17-93

Literature References

2. Albertsen, H., et al. (1994). "Genetic Mapping of the BRCA1 Region on Chromosome 17q21," *Am. J. Hum. Genet. 54:*516-525.

3. Anderson, D.E. (1972). "A Genetic Study of Human Breast Cancer," *J. Natl. Cancer Inst. 48:*1029-1034.

4. Arason, A., et al. (1993). "Linkage Analysis of Chromosome 17q Markers and Breast-Ovarian Cancer in Icelandic Families, and Possible Relationship to Prostatic Cancer," *Am. J. Hum. Genet. 52:*711-717.

5. Bishop, T.D. and Gardner, E.J. (1980). "Analysis of the Genetic Predisposition to Cancer in Individual Pedigrees," Banbury Report 4: Cancer Incidence in Defined Populations, Cairns et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 389-408.

6. Bishop, D.T., et al. (1988). "Segregation and Linkage Analysis of Nine Utah Breast Cancer Pedigrees," *Genet. Epidemiol. 5:*151-169.

7. Black, D.M., et al. (1993). "A Somatic Cell Hybrid Map of the Long Arm of Human Chromosome 17, Containing the Familial Breast Cancer Locus (BRCA1)," *Am. J. Hum. Genet. 52:*702-710.

8. Bowcock, A.M., et al. (1993). "THRA1 and D17S183 Flank an Interval of <4cM for the Breast-Ovarian Cancer Gene (BRCA1) on Chromosome 17q21," *Am. J. Hum. Genet. 52:* 718-722.

9. Boyd, J. (1995). "*BRCA1:* More than a hereditary breast cancer gene?" *Nature Genetics 9:*335-336.

10. Chamberlain, J.S., et al. (1993). "BRCA1 Maps Proximal to D17S579 on Chromosome 17q21 by Genetic Analysis," *Am. J. Hum. Genet. 52:*792-798.

11. Claus, E.B., et al. (1991). "Genetic Analysis of Breast Cancer in the Cancer and Steroid Hormone Study," *Am. J. Hum. Genet. 48:*232-242.

12. Cohen, B.B., et al. (1993). "Linkage of a Major Breast Cancer Gene to Chromosome 17q12-21: Results from 15 Edinburgh Families," *Am. J. Hum. Genet. 52:*723-729.

13. Devilee, P., et al. (1993). "Linkage to Markers for the Chromosome Region 17q12-q21 in 13 Dutch Breast Cancer Kindreds," *Am. J. Hum. Genet. 52:*730-735.

14. Dunphy, W.G. and Newport, J.W. (1989). "Fission Yeast p13 Blocks Mitotic Activation and Tyrosine Dephosphorylation of the Xenopus cdc2 Protein Kinase," *Cell 58:*181-191.

15. Easton, D.F., et al. (1993). "Genetic Linkage Analysis in Familial Breast and Ovarian Cancer: Results from 214 Families," *Am. J. Hum. Genet. 52:*678-701.

16. Fain, P.R. (1992). "Third international workshop on human chromosome 17 mapping," *Cytogen. Cell Genet. 60:*178-186.

17. Feunteun, J., et al. (1993). "A Breast-Ovarian Cancer Susceptibility Gene Maps to Chromosome 17q21," *Am. J. Hum. Genet. 52:*736-742.

18. Ford, D., et al. (1993). "The risks of cancer in BRCA1 mutation carriers," *Am. J. Hum. Genet. 53(supplement):*abstract no. 298.

19. Futreal, P.A., et al. (1994). "*BRCA1* Mutations in Primary Breast and Ovarian Carcinomas," *Science 266:*120-122.

20. Go, R.C.P., et al. (1983). "Genetic Epidemiology of Breast Cancer and Associated Cancers in High-Risk Families. I. Segregation Analysis," *J. Natl. Cancer Inst. 71:*455-461.

21. Goldgar, D.E., et al. (1993). "Chromosome 17q Linkage Studies of 18 Utah Breast Cancer Kindreds," *Am. J. Hum. Genet. 52:*743-748.

22. Gould, K.L. and Nurse, P. (1989). "Tyrosine phosphorylation of the fission yeast $cdc2^{+}$ protein kinase regulates entry into mitosis," *Nature 342:*39-45.

23. Hall, J.M., et al. (1990). "Linkage of Early-Onset Familial Breast Cancer to Chromosome 17q21," *Science 250:*1684-1689.

24. Hall, J.M., et al. (1992). "Closing in on a Breast Cancer Gene on Chromosome 17q," *Am. J. Hum. Genet. 50:*1235-1241.

25. Hosking, L., et al. (1995). "A somatic *BRCA1* mutation in an ovarian tumour," *Nature Genetics 9:*343-344.

26. Ishibashi, T., et al. (1992). "Expression cloning of a human dual-specificity phosphatase," *Proc. Natl. Acad. Sci. USA 89:*12170-12174.

27. Kamb, A., et al. (1994). "Localization of the VHR Phosphatase Gene and Its Analysis as a Candidate for BRCA1," *Genomics 23:*163-167.

28. Kumagai, A. and Dunphy, W.G. (1991). "The cdc25 Protein Controls Tyrosine Dephosphorylation of the cdc2 Protein in a Cell-Free System," *Cell 64:*903-914.

29. Lindblom, A., et al. (1993). "Linkage Analysis with Markers on 17q in 29 Swedish Breast Cancer Families," *Am. J. Hum. Genet. 52:*749-753.

30. Malkin, D., et al. (1990). "Germ Line p53 Mutations in a Familial Syndrome of Breast Cancer, Sarcomas, and Other Neoplasms," *Science 250:*1233-1238.

31. Margaritte, P., et al. (1992). "Linkage of Familial Breast Cancer to Chromosome 17q21 May Not be Restricted to Early-Onset Disease," *Am. J. Hum. Genet. 50:*1231-1234.

32. Mazoyer, S., et al. (1993). "Linkage Analysis of 19 French Breast Cancer Families, with Five Chromosome 17q Markers," *Am. J. Hum. Genet. 52:*754-760.

33. Merajver, S.D., et al. (1995). "Somatic mutations in the *BRCA1* gene in sporadic ovarian tumours," *Nature Genetices 9:*439-443.

34. Miki, Y., et al. (1994). "A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene *BRCA1*," *Science 266:*66-71.

35. Milner, B.J., et al. (1993). "Linkage "Studies with 17q and 18q Markers in a Breast/Ovarian Cancer Family," *Am. J. Hum. Genet. 52:*761-766.

36. Narod, S.A., et al. (1991). "Familial breast-ovarian cancer locus on chromosome 17q12-q23," *Lancet 338:*82-83.

37. Newman, B. et al. (1988). "Inheritance of human breast cancer: Evidence for autosomal dominant transmission in high-risk families," *Proc. Natl. Acad. Sci. USA 85:*3044-3048.

38. O'Connell, P., et al. (1994). "A Radiation Hybrid Map of the BRCA1 Region," *Am. J. Hum. Genet. 54:*526-534.

39. Roberts, L. (1993). "Zeroing In on a Breast Cancer Susceptibility Gene," *Science 259:*622-625.

40. Shattuck-Eidens, D., et al. (1995). "A Collaborative Survey of 80 Mutations in the BRCA1 Breast and Ovarian Cancer Susceptibility Gene," JAMA *273:*535-541.

41. Simard, J., et al. (1994). "Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families," *Nature Genetics 8:*392-398.

42. Smith, S.A., et al. (1992). "Allele losses in the region 17q12-21 in familial breast and ovarian cancer involve the wild-type chromosome," *Nature Genetics 2:*128-131.

43. Smith, S.A., et al. (1993). "Genetic Heterogeneity and Localization of a Familial Breast-Ovarian Cancer Gene on Chromosome 17q12-q21," *Am. J. Hum. Genet. 52:*767-776.

44. Spurr, N.K., et al. (1993). "Linkage Analysis of Early-Onset Breast and Ovarian Cancer Families, with Markers on the Long Arm of Chromosome 17," *Am. J. Hum. Genet. 52:*777-785.

45. Teare, M.D., et al. (1993). "A Linkage Study in Seven Breast Cancer Families," *Am. J. Hum. Genet. 52:* 786-788.

46. Thompson, M.E., et al. (1995). "Decreased expression of *BRCA1* accelerates growth and is often present during sporadic breast cancer progression," *Nature Genetics 9:*444-450.

47. Ullrich, A. and Schlessinger, J. (1990). "Signal Transductin by Receptors with Tyrosine Kinase Activity," *Cell 61:*203-212.

48. Williams, W.R. and Anderson, D.E. (1984). "Genetic Epidemiology of Breast Cancer: Segregation Analysis of 200 Danish Pedigrees," *Genet. Epidemiol. 1:*7-20.

49. Zimmermann, W., et al. (1993). "Linkage Analysis in German Breast Cancer Families with Early Onset of the Disease, Using Highly Polymorphic Markers from the Chromosome 17q11-q24 Region," *Am. J. Hum. Genet. 52:*789-791.

Respectfully submitted,



Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI
1201 New York Avenue, N.W., #1000
Washington, DC 20005
(202) 962-4810
Atty. Docket No. 24884-109347-B
Dated: 12 October 1995
11743.01

COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

Docket No. 24884-109347-B

FORM PTO-1449 (Modified)
INFORMATION DISCLOSURE STATEMENT
Sheet 1 of 9

In re Application of:

Mark H. SKOLNICK et al.

Serial No.: 08/488,011 Group Art Unit: 1807

Filed: 7 June 1995 Examiner: DRees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

U.S. PATENT DOCUMENTS

Ref. Desig.	Examiner's Initials	Document Number	Name	Date	Class-Subclass
1.	DR ✓	5,236,844	Basset et al.	08-17-93	X

DOCUMENTS (including Author, Title, Date, Pages, etc.)

Ref. Desig.	Examiner's Initials	
2.	DR ✓	Albertsen, H., et al. (1994). "Genetic Mapping of the BRCA1 Region on Chromosome 17q21," *Am. J. Hum. Genet. 54:*516-525.
3.	DR ✓	Anderson, D.E. (1972). "A Genetic Study of Human Breast Cancer," *J. Natl. Cancer Inst. 48:*1029-1034.
4.	DR ✓	Arason, A., et al. (1993). "Linkage Analysis of Chromosome 17q Markers and Breast-Ovarian Cancer in Icelandic Families, and Possible Relationship to Prostatic Cancer," *Am. J. Hum. Genet. 52:*711-717.
5.	DR ✓	Bishop, T.D. and Gardner, E.J. (1980). "Analysis of the Genetic Predisposition to Cancer in Individual Pedigrees," Banbury Report 4: Cancer Incidence in Defined Populations, Cairns et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 389-408.

Examiner: Dianne Rees Date Considered: 5/19/97

EXAMINER: Initial if citation considered whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

Docket No. 24484-109347-B

FORM PTO-1449 (Modified)
INFORMATION DISCLOSURE STATEMENT
Sheet 2 of 9

In re Application of:

Mark H. SKOLNICK et al.

Serial No.: 08/488,011 Group Art Unit: 1807

Filed: 7 June 1995 Examiner: DRees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

OTHER DOCUMENTS (Continued)

Ref. Desig.	Examiner's Initials	
6.	DR ✓	Bishop, D.T., et al. (1988). "Segregation and Linkage Analysis of Nine Utah Breast Cancer Pedigrees," *Genet. Epidemiol. 5:*151-169.
7.	DR ✓	Black, D.M., et al. (1993). "A Somatic Cell Hybrid Map of the Long Arm of Human Chromosome 17, Containing the Familial Breast Cancer Locus (BRCA1)," *Am. J. Hum. Genet. 52:*702-710.
8.	DR ✓	Bowcock, A.M., et al. (1993). "THRA1 and D17S183 Flank an Interval of <4cM for the Breast-Ovarian Cancer Gene (BRCA1) on Chromosome 17q21," *Am. J. Hum. Genet. 52:* 718-722.
9.	DR ✓	Boyd, J. (1995). "*BRCA1:* More than a hereditary breast cancer gene?" *Nature Genetics 9:*335-336.
10.	DR ✓	Chamberlain, J.S., et al. (1993). "BRCA1 Maps Proximal to D17S579 on Chromosome 17q21 by Genetic Analysis," *Am. J. Hum. Genet. 52:*792-798.

Examiner: Dianne Rees Date Considered: 5/15/97

EXAMINER: Initial if citation considered whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

MAIL ROOM
OCT
12
1995
52
PAT. & TRADEMARK OFF.

Docket No. 24484-109347-B

FORM PTO-1449 (Modified)
INFORMATION DISCLOSURE STATEMENT
Sheet 3 of 9

In re Application of:

Mark H. SKOLNICK et al.

Serial No.: 08/488,011 Group Art Unit: 1807

Filed: 7 June 1995 Examiner: Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

OTHER DOCUMENTS (Continued)

Ref. Desig.	Examiner's Initials	
11.	DR ✓	Claus, E.B., et al. (1991). "Genetic Analysis of Breast Cancer in the Cancer and Steroid Hormone Study," *Am. J. Hum. Genet. 48:*232-242.
12.	DR ✓	Cohen, B.B., et al. (1993). "Linkage of a Major Breast Cancer Gene to Chromosome 17q12-21: Results from 15 Edinburgh Families," *Am. J. Hum. Genet. 52:*723-729.
13.	DR ✓	Devilee, P., et al. (1993). "Linkage to Markers for the Chromosome Region 17q12-q21 in 13 Dutch Breast Cancer Kindreds," *Am. J. Hum. Genet. 52:*730-735.
14.	DR ✓	Dunphy, W.G. and Newport, J.W. (1989). "Fission Yeast p13 Blocks Mitotic Activation and Tyrosine Dephosphorylation of the Xenopus cdc2 Protein Kinase," *Cell 58:*181-191.
15.	DR ✓	Easton, D.F., et al. (1993). "Genetic Linkage Analysis in Familial Breast and Ovarian Cancer: Results from 214 Families," *Am. J. Hum. Genet. 52:*678-701.

Examiner: Diane Rees Date Considered: 5/19/97

EXAMINER: Initial if citation considered whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

MAIL ROOM
OCT
12
1995
PAT. & TRADEMARK OFF.

Docket No. 24484-109347-B

FORM PTO-1449 (Modified)
INFORMATION DISCLOSURE STATEMENT
Sheet 4 of 9

In re Application of:

Mark H. SKOLNICK et al.

Serial No.: 08/488,011 Group Art Unit: 1807

Filed: 7 June 1995 Examiner: DRees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

OTHER DOCUMENTS (Continued)

Ref. Desig.	Examiner's Initials	
16.	DR ✓	Fain, P.R. (1992). "Third international workshop on human chromosome 17 mapping," *Cytogen. Cell Genet. 60:*178-186.
17.	DR ✓	Feunteun, J., et al. (1993). "A Breast-Ovarian Cancer Susceptibility Gene Maps to Chromosome 17q21," *Am. J. Hum. Genet. 52:*736-742.
18.	DR ✓	Ford, D., et al. (1993). "The risks of cancer in BRCA1 mutation carriers," *Am. J. Hum. Genet. 53(supplement):*abstract no. 298.
19.	DR ✓	Futreal, P.A., et al. (1994). "*BRCA1* Mutations in Primary Breast and Ovarian Carcinomas," *Science 266:*120-122.
20.	DR ✓	Go, R.C.P., et al. (1983). "Genetic Epidemiology of Breast Cancer and Associated Cancers in High-Risk Families. I. Segregation Analysis," *J. Natl. Cancer Inst. 71:*455-461.
21.	DR	Goldgar, D.E., et al. (1993). "Chromosome 17q Linkage Studies of 18 Utah Breast Cancer Kindreds," *Am. J. Hum. Genet. 52:*743-748.

Examiner: Dianne Rees Date Considered: 5/9/97

EXAMINER: Initial if citation considered whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

MAIL ROO
OCT
52
1995
PAT. & TRADEMARK OFF.

COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

Docket No. 24484-109347-B

FORM PTO-1449 (Modified)
INFORMATION DISCLOSURE STATEMENT
Sheet 5 of 9

In re Application of:

Mark H. SKOLNICK et al.

Serial No.: 08/488,011 Group Art Unit: 1807

Filed: 7 June 1995 Examiner: DRees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

OTHER DOCUMENTS (Continued)

Ref. Desig.	Examiner's Initials	
22.	DR ✓	Gould, K.L. and Nurse, P. (1989). "Tyrosine phosphorylation of the fission yeast cdc2$^+$ protein kinase regulates entry into mitosis," *Nature 342:*39-45.
23.	DR ✓	Hall, J.M., et al. (1990). "Linkage of Early-Onset Familial Breast Cancer to Chromosome 17q21," *Science 250:*1684-1689.
24.	DR ✓	Hall, J.M., et al. (1992). "Closing in on a Breast Cancer Gene on Chromosome 17q," *Am. J. Hum. Genet. 50:*1235-1241.
25.	DR ✓	Hosking, L., et al. (1995). "A somatic *BRCA1* mutation in an ovarian tumour," *Nature Genetics 9:*343-344.
26.	DR ✓	Ishibashi, T., et al. (1992). "Expression cloning of a human dual-specificity phosphatase," *Proc. Natl. Acad. Sci. USA 89:*12170-12174.
27.	DR ✓	Kamb, A., et al. (1994). "Localization of the VHR Phosphatase Gene and Its Analysis as a Candidate for BRCA1," *Genomics 23:*163-167.
28.	DR ✓	Kumagai, A. and Dunphy, W.G. (1991). "The cdc25 Protein Controls Tyrosine Dephosphorylation of the cdc2 Protein in a Cell-Free System," *Cell 64:*903-914.

Examiner: Dianne Rees Date Considered: 5/19/97

EXAMINER: Initial if citation considered whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

MAIL ROOM
OCT
52 12
1995
PAT. & TRADEMARK OFF.

COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

Docket No. 24484-109347-B

FORM PTO-1449 (Modified)
INFORMATION DISCLOSURE STATEMENT
Sheet 6 of 9

In re Application of:

Mark H. SKOLNICK et al.

Serial No.: 08/488,011 Group Art Unit: 1807

Filed: 7 June 1995 Examiner: DRees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

OTHER DOCUMENTS (Continued)

Ref. Desig.	Examiner's Initials	
29.	DR ✓	Lindblom, A., et al. (1993). "Linkage Analysis with Markers on 17q in 29 Swedish Breast Cancer Families," *Am. J. Hum. Genet. 52:*749-753.
30.	DR ✓	Malkin, D., et al. (1990). "Germ Line p53 Mutations in a Familial Syndrome of Breast Cancer, Sarcomas, and Other Neoplasms," *Science 250:*1233-1238.
31.	DR ✓	Margaritte, P., et al. (1992). "Linkage of Familial Breast Cancer to Chromosome 17q21 May Not be Restricted to Early-Onset Disease," *Am. J. Hum. Genet. 50:*1231-1234.
32.	DR ✓	Mazoyer, S., et al. (1993). "Linkage Analysis of 19 French Breast Cancer Families, with Five Chromosome 17q Markers," *Am. J. Hum. Genet. 52:*754-760.
33.	DR ✓	Merajver, S.D., et al. (1995). "Somatic mutations in the *BRCA1* gene in sporadic ovarian tumours," *Nature Genetices 9:*439-443.
34.	DR ✓	Miki, Y., et al. (1994). "A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene *BRCA1*," *Science 266:*66-71.

Examiner: Dianne Rees Date Considered: 5/19/97

EXAMINER: Initial if citation considered whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231



Docket No. 24484-109347-B

FORM PTO-1449 (Modified)
INFORMATION DISCLOSURE STATEMENT
Sheet 7 of 9

In re Application of:

Mark H. SKOLNICK et al.

Serial No.: 08/488,011 Group Art Unit: 1807

Filed: 7 June 1995 Examiner: DReed

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

OTHER DOCUMENTS (Continued)

Ref. Desig.	Examiner's Initials	
35.	DR ✓	Milner, B.J., et al. (1993). "Linkage "Studies with 17q and 18q Markers in a Breast/Ovarian Cancer Family," *Am. J. Hum. Genet. 52:*761-766.
36.	DR ✓	Narod, S.A., et al. (1991). "Familial breast-ovarian cancer locus on chromosome 17q12-q23," *Lancet 338:*82-83.
37.	DR ✓	Newman, B. et al. (1988). "Inheritance of human breast cancer: Evidence for autosomal dominant transmission in high-risk families," *Proc. Natl. Acad. Sci. USA 85:*3044-3048.
38.	DR ✓	O'Connell, P., et al. (1994). "A Radiation Hybrid Map of the BRCA1 Region," *Am. J. Hum. Genet. 54:*526-534.
39.	DR ✓	Roberts, L. (1993). "Zeroing In on a Breast Cancer Susceptibility Gene," *Science 259:*622-625.
40.	DR ✓	Shattuck-Eidens, D., et al. (1995). "A Collaborative Survey of 80 Mutations in the BRCA1 Breast and Ovarian Cancer Susceptibility Gene," JAMA *273:*535-541.
41.	DR ✓	Simard, J., et al. (1994). "Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families," *Nature Genetics 8:*392-398.

Examiner: Dianne Reed Date Considered: 5/19/97

EXAMINER: Initial if citation considered whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

MAIL ROOM
52 OCT 12 1995
PAT. & TRADEMARK OFF.

Docket No. 24484-109347-B

FORM PTO-1449 (Modified)
INFORMATION DISCLOSURE STATEMENT
Sheet 8 of 9

In re Application of:

Mark H. SKOLNICK et al.

Serial No.: 08/488,011 Group Art Unit: 1807

Filed: 7 June 1995 Examiner: D. Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

OTHER DOCUMENTS (Continued)

Ref. Desig.	Examiner's Initials	
42.	DR ✓	Smith, S.A., et al. (1992). "Allele losses in the region 17q12-21 in familial breast and ovarian cancer involve the wild-type chromosome," *Nature Genetics 2:*128-131.
43.	DR ✓	Smith, S.A., et al. (1993). "Genetic Heterogeneity and Localization of a Familial Breast-Ovarian Cancer Gene on Chromosome 17q12-q21," *Am. J. Hum. Genet. 52:*767-776.
44.	DR ✓	Spurr, N.K., et al. (1993). "Linkage Analysis of Early-Onset Breast and Ovarian Cancer Families, with Markers on the Long Arm of Chromosome 17," *Am. J. Hum. Genet. 52:*777-785.
45.	DR ✓	Teare, M.D., et al. (1993). "A Linkage Study in Seven Breast Cancer Families," *Am. J. Hum. Genet. 52:* 786-788.
46.	DR ✓	Thompson, M.E., et al. (1995). "Decreased expression of *BRCA1* accelerates growth and is often present during sporadic breast cancer progression," *Nature Genetics 9:*444-450.
47.	DR ✓	Ullrich, A. and Schlessinger, J. (1990). "Signal Transduction by Receptors with Tyrosine Kinase Activity," *Cell 61:*203-212.

Examiner: Dianne Rees Date Considered: 5/19/97

EXAMINER: Initial if citation considered whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

MAIL ROOM
OCT
12
1995
PAT. & TRADEMARK OFF.
52

COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

Docket No. 24484-109347-B

FORM PTO-1449 (Modified)
INFORMATION DISCLOSURE STATEMENT
Sheet 9 of 9

In re Application of:

Mark H. SKOLNICK et al.

Serial No.: 08/488,011 Group Art Unit: 1807

Filed: 7 June 1995 Examiner: D. Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

OTHER DOCUMENTS (Continued)

Ref. Desig.	Examiner's Initials	
48.	DR	Williams, W.R. and Anderson, D.E. (1984). "Genetic Epidemiology of Breast Cancer: Segregation Analysis of 200 Danish Pedigrees," *Genet. Epidemiol. 1*:7-20.
49.	DR	Zimmermann, W., et al. (1993). "Linkage Analysis in German Breast Cancer Families with Early Onset of the Disease, Using Highly Polymorphic Markers from the Chromosome 17q11-q24 Region," *Am. J. Hum. Genet. 52*:789-791.

Examiner: Dianne Rees Date Considered: 5/19/97

EXAMINER: Initial if citation considered whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

Inv. 022 a
2/24/96

01/05/96

T-2

MAIL ROOM
FEB
39 23
1996
PAT. & TRADEMARK OFF.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

1807
Rees

In re patent application of:

Mark H. SKOLNICK et al.

Serial No.: 08/488,011 Group Art Unit: 1862 1807 1800

Filed: 7 June 1995 Examiner: Kim 1806

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

SUPPLEMENTAL INFORMATION DISCLOSURE STATEMENT

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

Transmitted herewith for the Examiner's consideration are copies of the references which are believed relevant to the above invention, and which are listed on the attached Form PTO-1449. The Examiner is requested to return an initialled copy of this form indicating his review of the same.

Respectfully submitted,

Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI
1201 New York Avenue, N.W., #1000
Washington, D.C. 20005
(202) 962-4810

Atty. Docket No. 24884-109347-09
Dated: 23 February 1996
DC2\21237.01

COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

Docket No. 24884-109347-09

FORM PTO-1449 (Modified)
INFORMATION DISCLOSURE STATEMENT
(Use several sheets if necessary)
Sheet 1 of 2

In re Application of:

Mark H. SKOLNICK et al.

Serial No.: 08/488,011

Group Art Unit: 1807

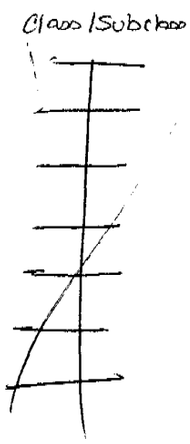
Filed: 7 June 1995

Examiner: D. Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

FOREIGN PATENT DOCUMENTS

Ref. Desig.	Examiner's Initials	Country	Document Number	Publication Date	Class/Subclass
1.	DR ✓	EPO	0 518 650	12-16-92	
2.	DR ✓	WO	91/09964	07-11-91	
3.	DR ✓	WO	92/00311	01-09-92	
4.	DR ✓	WO	92/11874	07-23-92	
5.	DR ✓	WO	94/00764	01-06-94	
6.	DR ✓	WO	95/19369	07-20-95	



OTHER DOCUMENTS (including Author, Title, Date, Pages, etc.)

Ref. Desig.	Examiner's Initials	
7.	DR ✓	Castilla, L.H. et al. (1994). "Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer," *Nature Genetics* 8:87-391.
8.	DR ✓	Friedman, L.S. et al. (1994). "Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families," *Nature Genetics* 8:399-404.

Examiner: Dianne Rees Date Considered: 5/19/97

EXAMINER: Initial if citation considered whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

Docket No. 24884-109347-09

FORM PTO-1449 (Modified)
INFORMATION DISCLOSURE STATEMENT
(Use several sheets if necessary)
Sheet 2 of 2

In re Application of:

Mark H. SKOLNICK et al.

Serial No.: 08/488,011

Group Art Unit:

Filed: 7 June 1995

Examiner:

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

OTHER DOCUMENTS (including Author, Title, Date, Pages, etc.)

9. DR ✓ Goldgar, D.E. et al. (1994). "A Large Kindred With 17q-Linked Breast and Ovarian Cancer: Genetic, Phenotypic, and Genealogical Analysis," *J. Natl. Cancer Institute 86*:200-209.

10. DR ✓ Liang, P. et al. (1992). "Differential Display and Cloning of Messenger RNAs from Human Breast Cancer *versus* Mammary Epithelial Cells," *Cancer Research 52*:6966-6968.

11. DR ✓ Neuhausen, S.L. et al. (1994). "A P1-based physical map of the region from D17S776 to D17378 containing the breast cancer susceptibility gene BRCA1," *Hum. Mol. Gen.* 3:1919-1926.

12. DR ✓ Sato, T. et al. (1992). "The Human Prohibitin Gene Located on Chromosome 17q21 is Mutated in Sporadic Breast Cancer," *Cancer Research 52*:1543-1646.

13. DR ✓ Stampfer, M.R. et al. (1993). "Culture Systems for Study of Human Mammary Epithelial Cell Proliferation, Differentiation and Transformation," *Cancer Surveys 18*:7-34.

14. DR ✓ Wooster, R. et al. (1994). "Localization of a Breast Cancer Susceptibility Gene, BRCA2, to Chromosome 13q12-13," *Science 265*:2088-2090.

Examiner: Dianne Rees Date Considered: 5/19/97

EXAMINER: Initial if citation considered whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

Notice of References Cited

Application No.	Applicant(s)		
08/~~487002~~ 489011	SKOLNICK ET AL.		
Examiner		Group Art Unit	Page 1 of 3
Dianne Rees		1807	

U.S. PATENT DOCUMENTS

*		DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS
	A					
	B					
	C					
	D					
	E					
	F					
	G					
	H					
	I					
	J					
	K					
	L					
	M					

FOREIGN PATENT DOCUMENTS

*		DOCUMENT NO.	DATE	COUNTRY	NAME	CLASS	SUBCLASS
~~X~~	N	~~WO 95/19369~~		PCT	~~HOLT ET AL.~~	~~---~~	
	O				(cited in 487002)		
	P						
	Q						
	R						
	S						
	T						

NON-PATENT DOCUMENTS

*		DOCUMENT (Including Author, Title, Source, and Pertinent Pages)	DATE
~~X~~	U	~~DORLAND'S MEDICAL DICTIONARY~~ DR-5/19/97 cited in '487002	
~~X~~	V	GOLDBERG ET AL. CLINICAL CHEMISTRY 39: 2360-2374 cited in '487002	1993
X	W	WOOSTER ET AL., SCIENCE 265:2088-2090 " "	1994
X	X	SATO ET AL., CANCER RES. 52: 1643-1646 " "	1992

*A copy of this reference is not being furnished with this Office action.
(See Manual of Patent Examining Procedure, Section 707.05(a).)

GeneDX 1009, pg. 277

Notice of References Cited

Application No.	Applicant(s)		
08/~~487,002~~ 488011	SKOLNICK ET AL.		
Examiner		Group Art Unit	Page 2 of 3
Dianne Rees		1807	

U.S. PATENT DOCUMENTS

*		DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS
	A					
	B					
	C					
	D					
	E					
	F					
	G					
	H					
	I					
	J					
	K					
	L					
	M					

FOREIGN PATENT DOCUMENTS

*		DOCUMENT NO.	DATE	COUNTRY	NAME	CLASS	SUBCLASS
	N						
	O						
	P						
	Q						
	R						
	S						
	T						

NON-PATENT DOCUMENTS

*		DOCUMENT (Including Author, Title, Source, and Pertinent Pages)	DATE
x	U	MALKIN ET AL.,SCIENCE 250: 1233-1238	1990
x	V	BOYD, NATURE GENETICS 9: 335-336	1995
x	W	MERAJVER ET AL, NATURE GENETICS 9;439-443	1995
	X	CASTILLA ET AL NATURE GENETICS 8: 387-391	1994

* A copy of this reference is not being furnished with this Office action.
(See Manual of Patent Examining Procedure, Section 707.05(a).)

Notice of References Cited

Application No.	Applicant(s)		
08/~~487002~~ 488011	SKOLNICK ET AL.		
Examiner		Group Art Unit	Page 3 of 3
Dianne Rees		1807	

U.S. PATENT DOCUMENTS

		DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS
	A					
	B					
	C					
	D					
	E					
	F					
	G					
	H					
	I					
	J					
	K					
	L					
	M					

FOREIGN PATENT DOCUMENTS

		DOCUMENT NO.	DATE	COUNTRY	NAME	CLASS	SUBCLASS
	N						
	O						
	P						
	Q						
	R						
	S						
	T						

NON-PATENT DOCUMENTS

		DOCUMENT (Including Author, Title, Source, and Pertinent Pages)	DATE
~~X~~	U	FRIEDMAN ET AL., NATURE GENETICS 8: 1-6 cited in '487002	1994
~~X~~	V	GOLDGAR ET AL. J. OF NAT. CANCER INST, 86: 200-209 cited in '487002	1994
	W		
	X		

GeneDX 1009, pg. 279

MAIL ROOM
MAY
7
1996
39
PAT. & TRADEMARK OFF.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Re: Application of:

Mark H. SKOLNICK, et al.

Serial No. 08/488,011

Filed: 07 June 1995

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

COMMUNICATION

In response to the Notice of Filing Date Granted Incomplete Response mailed 30 April 1996, Applicant states that the signature for the third inventor is not missing from the Declaration and Power of Attorney. A copy of the papers submitted on 1 December 1995 in response to the Notice to File Missing Parts mailed 18 August 1995 is attached. As you will note, there were four (4) signed Declarations submitted and the signature for the third inventor appears on one of these declarations. A copy of the date stamped filing receipt is also attached evidencing the Patent Office's receipt of these documents.

Please correct the records in the patent office to reflect timely submission of the missing signature, and issue a filing receipt for this case.

Respectfully submitted

Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, LLP
1201 New York Avenue, NW, Suite 1000
Washington, DC 20005
(202) 962-4810
Attorney Docket No. 24884-109347-09
Dated: 7 May 1996

DECLARATION AND POWER OF ATTORNEY

As below-named inventor, I hereby declare that my residence, post office address and citizenship are as stated below next to my name; I believe I am an original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE, the specification of which was filed on 7 June 1995 under Serial No. 08/488,011.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) on which priority is claimed:

NONE

I hereby claim the benefit under Title 35, United States Code, §120 of any United States applications that are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which occurred between the filing date of the prior applications and the national or PCT international filing date of this application:

U.S. Serial No. 08/409,305, filed on 24 March 1995

U.S. Serial No. 08/348,824, filed on 24 November 1994

U.S. Serial No. 08/308,104, filed on 16 September 1994

U.S. Serial No. 08/300,266, filed on 2 September 1994

U.S. Serial No. 08/289,221, filed on 12 August 1994

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

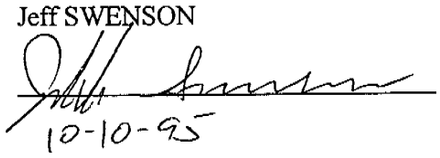
I hereby appoint Jeffrey L. Ihnen, Registration No. 28,957; E. Brendan Magrab, Registration No. 36,205; and Stephen A. Saxe, Registration No. 38,609; all of Venable, Baetjer, Howard & Civiletti, whose address is 1201 New York Avenue, N.W., Suite 1000, Washington, D.C. 20005, telephone number (202) 962-4800, my attorneys and/or agents with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

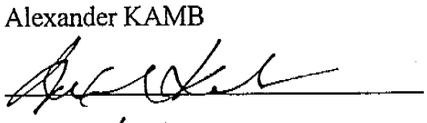
Please direct all correspondence and telephone calls to Venable, Baetjer, Howard & Civiletti at the above address and telephone number.

Full Name of Sole or First Inventor	Mark H. SKOLNICK
Inventor's Signature	Mark Skolnick
Date of Signature	Oct 9 1995
Residence Address	460 Northmount Way Salt Lake City, Utah 84103
Post Office Address	same
Country of Citizenship	United States

Full Name of Second Inventor	David E. GOLDGAR
Inventor's Signature	David E Goldgar
Date of Signature	October 9 1995
Residence Address	2707 East Kelly Lane Salt Lake City, Utah 84117
Post Office Address	same
Country of Citizenship	United States

Full Name of Third Inventor	Yoshio MIKI
Inventor's Signature	_____
Date of Signature	_____
Residence Address	40 South 900 East Salt Lake City, Utah 84102
Post Office Address	same
Country of Citizenship	Japan

Full Name of Fourthh Inventor	Jeff SWENSON
Inventor's Signature	
Date of Signature	10-10-95
Residence Address	745 West Creek Bend, #2807 Salt Lake City, Utah 84117
Post Office Address	same
Country of Citizenship	United States

Full Name of Fifth Inventor	Alexander KAMB
Inventor's Signature	
Date of Signature	10/9/95
Residence Address	1103 East 600 South Salt Lake City, Utah 84102
Post Office Address	same
Country of Citizenship	United States

Full Name of Sixth Inventor: Keith D. HARSHMAN

Inventor's Signature: _____

Date of Signature: 10.9.95

Residence Address: 159 North F Street
Salt Lake City, Utah 841032

Post Office Address: same

Country of Citizenship: United States

Full Name of Seventh Inventor: Donna M. SHATTUCK-EIDENS

Inventor's Signature: _____

Date of Signature: 10.9.95

Residence Address: 1874 Texas Street
Salt Lake City, Utah 84108

Post Office Address: same

Country of Citizenship: United States

Full Name of Eighth Inventor: Sean V. TAVTIGIAN

Inventor's Signature: _____

Date of Signature: 10/9/95

Residence Address: 557 East 1st Ave., #3
Salt Lake City, Utah 841023

Post Office Address: same

Country of Citizenship: United States

Full Name of Ninth Inventor	Roger W. WISEMAN
Inventor's Signature	_____
Date of Signature	_____
Residence Address	515 Woodwinds Drive Durham, North Carolina 27713
Post Office Address	same
Country of Citizenship	United States

Full Name of Tenth Inventor	P. Andrew FUTREAL
Inventor's Signature	_____
Date of Signature	_____
Residence Address	220 East Markham Ave. Durham, North Carolina 27701
Post Office Address	same
Country of Citizenship	United States

DECLARATION AND POWER OF ATTORNEY

As below-named inventor, I hereby declare that my residence, post office address and citizenship are as stated below next to my name; I believe I am an original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE, the specification of which was filed on 7 June 1995 under Serial No. 08/488,011.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) on which priority is claimed:

NONE

I hereby claim the benefit under Title 35, United States Code, §120 of any United States applications that are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which occurred between the filing date of the prior applications and the national or PCT international filing date of this application:

U.S. Serial No. 08/409,305, filed on 24 March 1995

U.S. Serial No. 08/348,824, filed on 24 November 1994

U.S. Serial No. 08/308,104, filed on 16 September 1994

U.S. Serial No. 08/300,266, filed on 2 September 1994

U.S. Serial No. 08/289,221, filed on 12 August 1994

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

I hereby appoint Jeffrey L. Ihnen, Registration No. 28,957; E. Brendan Magrab, Registration No. 36,205; and Stephen A. Saxe, Registration No. 38,609; all of Venable, Baetjer, Howard & Civiletti, whose address is 1201 New York Avenue, N.W., Suite 1000, Washington, D.C. 20005, telephone number (202) 962-4800, my attorneys and/or agents with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

Please direct all correspondence and telephone calls to Venable, Baetjer, Howard & Civiletti at the above address and telephone number.

Full Name of Sole or First Inventor	Mark H. SKOLNICK
Inventor's Signature	Mark Skolnick
Date of Signature	Oct 9 1995
Residence Address	460 Northmount Way Salt Lake City, Utah 84103
Post Office Address	same
Country of Citizenship	United States

Full Name of Second Inventor	David E. GOLDGAR
Inventor's Signature	David E Goldgar
Date of Signature	October 9 1995
Residence Address	2707 East Kelly Lane Salt Lake City, Utah 84117
Post Office Address	same
Country of Citizenship	United States

Full Name of Third Inventor	Yoshio MIKI
Inventor's Signature	Yoshio Miki
Date of Signature	10/13/95
Residence Address	40 South 900 East Salt Lake City, Utah 84102
Post Office Address	same
Country of Citizenship	Japan

Full Name of Fourthh Inventor	Jeff SWENSON
Inventor's Signature	_____
Date of Signature	_____
Residence Address	745 West Creek Bend, #2807 Salt Lake City, Utah 84117
Post Office Address	same
Country of Citizenship	United States

Full Name of Fifth Inventor	Alexander KAMB
Inventor's Signature	
Date of Signature	10/9/95
Residence Address	1103 East 600 South Salt Lake City, Utah 84102
Post Office Address	same
Country of Citizenship	United States

Full Name of Sixth Inventor: Keith D. HARSHMAN

Inventor's Signature: Keith D. Harshman

Date of Signature: 10.9.95

Residence Address: 159 North F Street
Salt Lake City, Utah 841032

Post Office Address: same

Country of Citizenship: United States

Full Name of Seventh Inventor: Donna M. SHATTUCK-EIDENS

Inventor's Signature: Donna M. Shattuck-Eidens

Date of Signature: 10.9.95

Residence Address: 1874 Texas Street
Salt Lake City, Utah 84108

Post Office Address: same

Country of Citizenship: United States

Full Name of Eighth Inventor: Sean V. TAVTIGIAN

Inventor's Signature: SVT

Date of Signature: 10/9/95

Residence Address: 557 East 1st Ave., #3
Salt Lake City, Utah 841023

Post Office Address: same

Country of Citizenship: United States

Full Name of Ninth Inventor	Roger W. WISEMAN
Inventor's Signature	_____
Date of Signature	_____
Residence Address	515 Woodwinds Drive Durham, North Carolina 27713
Post Office Address	same
Country of Citizenship	United States

Full Name of Tenth Inventor	P. Andrew FUTREAL
Inventor's Signature	_____
Date of Signature	_____
Residence Address	220 East Markham Ave. Durham, North Carolina 27701
Post Office Address	same
Country of Citizenship	United States

DECLARATION AND POWER OF ATTORNEY

As below-named inventor, I hereby declare that my residence, post office address and citizenship are as stated below next to my name; I believe I am an original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE, the specification of which was filed on 7 June 1995 under Serial No. 08/488,011.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) on which priority is claimed:

NONE

I hereby claim the benefit under Title 35, United States Code, §120 of any United States applications that are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which occurred between the filing date of the prior applications and the national or PCT international filing date of this application:

U.S. Serial No. 08/409,305, filed on 24 March 1995

U.S. Serial No. 08/348,824, filed on 24 November 1994

U.S. Serial No. 08/308,104, filed on 16 September 1994

U.S. Serial No. 08/300,266, filed on 2 September 1994

U.S. Serial No. 08/289,221, filed on 12 August 1994

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

I hereby appoint Jeffrey L. Ihnen, Registration No. 28,957; E. Brendan Magrab, Registration No. 36,205; and Stephen A. Saxe, Registration No. 38,609; all of Venable, Baetjer, Howard & Civiletti, whose address is 1201 New York Avenue, N.W., Suite 1000, Washington, D.C. 20005, telephone number (202) 962-4800, my attorneys and/or agents with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

Please direct all correspondence and telephone calls to Venable, Baetjer, Howard & Civiletti at the above address and telephone number.

Full Name of Sole or First Inventor	Mark H. SKOLNICK
Inventor's Signature	_____
Date of Signature	_____
Residence Address	460 Northmount Way Salt Lake City, Utah 84103
Post Office Address	same
Country of Citizenship	United States

Full Name of Second Inventor	David E. GOLDGAR
Inventor's Signature	_____
Date of Signature	_____
Residence Address	2707 East Kelly Lane Salt Lake City, Utah 84117
Post Office Address	same
Country of Citizenship	United States

Full Name of Third Inventor	Yoshio MIKI
Inventor's Signature	_____
Date of Signature	_____
Residence Address	40 South 900 East Salt Lake City, Utah 84102
Post Office Address	same
Country of Citizenship	Japan

Full Name of Fourthh Inventor	Jeff SWENSON
Inventor's Signature	_____
Date of Signature	_____
Residence Address	745 West Creek Bend, #2807 Salt Lake City, Utah 84117
Post Office Address	same
Country of Citizenship	United States

Full Name of Fifth Inventor	Alexander KAMB
Inventor's Signature	_____
Date of Signature	_____
Residence Address	1103 East 600 South Salt Lake City, Utah 84102
Post Office Address	same
Country of Citizenship	United States

Full Name of Sixth Inventor	Keith D. HARSHMAN
Inventor's Signature	_____
Date of Signature	_____
Residence Address	159 North F Street Salt Lake City, Utah 841032
Post Office Address	same
Country of Citizenship	United States

Full Name of Seventh Inventor	Donna M. SHATTUCK-EIDENS
Inventor's Signature	_____
Date of Signature	_____
Residence Address	1874 Texas Street Salt Lake City, Utah 84108
Post Office Address	same
Country of Citizenship	United States

Full Name of Eighth Inventor	Sean V. TAVTIGIAN
Inventor's Signature	_____
Date of Signature	_____
Residence Address	557 East 1st Ave., #3 Salt Lake City, Utah 841023
Post Office Address	same
Country of Citizenship	United States

Full Name of Ninth Inventor: Roger W. WISEMAN

Inventor's Signature: _____

Date of Signature: _____

Residence Address: 515 Woodwinds Drive
Durham, North Carolina 27713

Post Office Address: same

Country of Citizenship: United States

Full Name of Tenth Inventor: P. Andrew FUTREAL

Inventor's Signature: P Andrew Futreal

Date of Signature: 11/14/95

Residence Address: ~~220 East Markham Ave.~~
~~Durham, North Carolina 27701~~

Post Office Address: same 311 S. LaSalle 28B
Durham NC 27705

Country of Citizenship: United States

PAF

DECLARATION AND POWER OF ATTORNEY

As below-named inventor, I hereby declare that my residence, post office address and citizenship are as stated below next to my name; I believe I am an original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE, the specification of which was filed on 7 June 1995 under Serial No. 08/488,011.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) on which priority is claimed:

NONE

I hereby claim the benefit under Title 35, United States Code, §120 of any United States applications that are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose all information known to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which occurred between the filing date of the prior applications and the national or PCT international filing date of this application:

U.S. Serial No. 08/409,305, filed on 24 March 1995

U.S. Serial No. 08/348,824, filed on 24 November 1994

U.S. Serial No. 08/308,104, filed on 16 September 1994

U.S. Serial No. 08/300,266, filed on 2 September 1994

U.S. Serial No. 08/289,221, filed on 12 August 1994

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

I hereby appoint Jeffrey L. Ihnen, Registration No. 28,957; E. Brendan Magrab, Registration No. 36,205; and Stephen A. Saxe, Registration No. 38,609; all of Venable, Baetjer, Howard & Civiletti, whose address is 1201 New York Avenue, N.W., Suite 1000, Washington, D.C. 20005, telephone number (202) 962-4800, my attorneys and/or agents with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

Please direct all correspondence and telephone calls to Venable, Baetjer, Howard & Civiletti at the above address and telephone number.

Full Name of Sole or First Inventor	Mark H. SKOLNICK
Inventor's Signature	_____
Date of Signature	_____
Residence Address	460 Northmount Way Salt Lake City, Utah 84103
Post Office Address	same
Country of Citizenship	United States

Full Name of Second Inventor	David E. GOLDGAR
Inventor's Signature	_____
Date of Signature	_____
Residence Address	2707 East Kelly Lane Salt Lake City, Utah 84117
Post Office Address	same
Country of Citizenship	United States

Full Name of Third Inventor	Yoshio MIKI
Inventor's Signature	_____
Date of Signature	_____
Residence Address	40 South 900 East Salt Lake City, Utah 84102
Post Office Address	same
Country of Citizenship	Japan

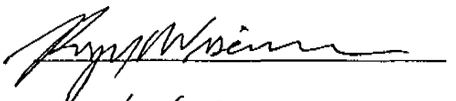
Full Name of Fourthh Inventor	Jeff SWENSON
Inventor's Signature	_____
Date of Signature	_____
Residence Address	745 West Creek Bend, #2807 Salt Lake City, Utah 84117
Post Office Address	same
Country of Citizenship	United States

Full Name of Fifth Inventor	Alexander KAMB
Inventor's Signature	_____
Date of Signature	_____
Residence Address	1103 East 600 South Salt Lake City, Utah 84102
Post Office Address	same
Country of Citizenship	United States

Full Name of Sixth Inventor	Keith D. HARSHMAN
Inventor's Signature	_____
Date of Signature	_____
Residence Address	159 North F Street Salt Lake City, Utah 841032
Post Office Address	same
Country of Citizenship	United States

Full Name of Seventh Inventor	Donna M. SHATTUCK-EIDENS
Inventor's Signature	_____
Date of Signature	_____
Residence Address	1874 Texas Street Salt Lake City, Utah 84108
Post Office Address	same
Country of Citizenship	United States

Full Name of Eighth Inventor	Sean V. TAVTIGIAN
Inventor's Signature	_____
Date of Signature	_____
Residence Address	557 East 1st Ave., #3 Salt Lake City, Utah 841023
Post Office Address	same
Country of Citizenship	United States

Full Name of Ninth Inventor	Roger W. WISEMAN
Inventor's Signature	
Date of Signature	10/20/95
Residence Address	~~515 Woodwinds Drive~~ 16 Stream View Court RW Durham, North Carolina 27713
Post Office Address	same
Country of Citizenship	United States

Full Name of Tenth Inventor	P. Andrew FUTREAL
Inventor's Signature	_____
Date of Signature	_____
Residence Address	220 East Markham Ave. Durham, North Carolina 27701
Post Office Address	same
Country of Citizenship	United States

DOCKETED
Reminder
Action Due
Due Date
Client

(PTO Receipt Stamp Here)

Receipt for Documents Filed

MAIL ROOM DEC 1 1995 EXP 13 PAT. & TRADEMARK OFF.

IN RE. [xx] PATENT [ ] TM
APPLICATION OF INVENTOR(S): Mark H. SKOLNICK et al.

APPLICANT (FOR PCT OR TM):

SERIAL NO.: 08/488,011 REG. NO.
PAT. NO.

FILING DATE: 07 June 1995

FOR: [ ] MARK [xx] TITLE 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

DOCUMENTS FILED:

==========

TRANSMITTAL OF FORMAL DOCUMENTS, including cover letter (in triplicate), four signed Declarations and Powers of Attorney, three Assignments (with cover sheets), Statement Pursuant to 37 CFR 1.821(f) with computer diskette of sequences, copy of form PTO-1533 and copy of Notice to Comply; and with authorization to charge $750 filing fee/$130 surcharge/$120 assignments recordation to Deposit Account; AND

PETITION FOR EXTENSION OF TIME (3 Mos), in triplicate, with authorization to charge Deposit Account for $900 fee

==========

DATE FILED: 01 December 1995

DUE DATE: 18 December 1995

ATTY/SEC. JLI:ch

CLIENT/
MATTER NO. 24884-109347-B

CLIENT NAME: Myriad

MAIL ROOM
JUN 6 1996
PATENT & TRADEMARK OFF

Gp 1803

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re: Application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011 Group Art Unit:

Filed: 07 June 1995

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

#12
8-28-96

PETITION TO MAKE SPECIAL

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

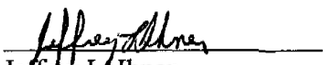
Sir:

In accordance with the O.G. Notice of 21 March 1995, Applicants hereby petition the Commissioner to afford the above-identified application "special" status and to advance the examination of the application. The instant invention is directed to the identification of the BRCA1 gene which is associated with breast and ovarian cancer. The identification of this gene and its mutations enables the diagnosis of individuals who carry a mutated BRCA1 gene. Individuals carrying a mutated BRCA1 gene are at risk of developing breast and/or ovarian cancer. Diagnosis of at risk individuals will have a significant impact in the monitoring individuals for the onset of such cancer. Diagnosis of a mutated BRCA1 gene in individuals who have had breast cancer will have a significant impact in the monitoring for reoccurrence of breast cancer or the development of ovarian cancer.

The Commissioner is hereby authorized to charge the requisite petition fee of $130.00 to Deposit Account No. 22-0261. The Commissioner is further authorized to charge any additional

fees which may be due or to credit any overpayments to said deposit account. This Petition is submitted in triplicate.

Respectfully submitted,



Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, LLP
1201 New York Avenue, N.W., Suite 1000
Washington, D.C. 20005
(202) 962-4810

Attorney Docket No.: 24884-109347-09
Dated: 6 June 1996
DC2\19979.01

MAIL ROOM JUN 6 1996 PAT. & TRADEMARK OFF

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011 Group Art Unit:

Filed: 07 June 1995

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

PRELIMINARY AMENDMENT

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

Prior to the examination of the above-identified application, please amend the application as follows:

IN THE SPECIFICATION

On page 47, line 21, please insert -- by our University of Utah collaborators -- after "ascertained".

On page 47, line 28, please insert -- by our collaborators -- after "relatives".

On page 48, line 3, please insert -- by our collaborators -- after "invited".

On page 48, line 4, please substitute -- They -- for "We".

On page 48, lines 7-8, please delete "the linked databases".

On page 48, line 10 please insert -- by our collaborators -- after "studied".

On page 48, line 10, please substitute -- their -- for "our".

On page 48, line 11, please delete "the linked databases".

On page 48, line 14 please insert -- collaborators' -- after "our".

On page 68, line 19, please insert -- DNA from -- after "tested".

On page 70, line 15, please delete "under the age of 40" and substitute therefor -- at very early ages --.

On page 70, line 16, after "mother" please insert -- also --.

On page 70, line 16, please delete "age of 45" and substitute therefor -- at an early age --.

On page 70, line 16, please delete "at age 61" and substitute therefor -- some years later --.

On page 70, lines 16-20, please delete "This ... certainty.".

On page 87, line 1, please delete "age 24" and substitute therefor -- a very early age --.

On page 87, line 3, please delete "age 42" and substitute therefor -- an early age --.

On page 87, line 4, please delete "age 34" and substitute therefor -- a very early age --.

On page 87, line 4, please delete ", another ... cancer".

REMARKS

The specification has been amended in an effort to maintain the confidentiality of the individuals and their families. The specification has also been amended to indicate that the inventors did not have access or names of the individuals or families studied. These changes have been made to comply with federal, state and University guidelines for maintaining the confidentiality of individuals and families involved in studies such as disclosed in the instant application. It is believed that none of these amendments constitute new matter and the entry of these amendments is requested.

Respectfully submitted,

Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, LLP
1201 New York Avenue, N.W., Suite 1000
Washington, D.C. 20005
(202) 962-4810

Attorney Docket No.: 24884-109347-09
Dated: 6 June 1996
DC2\19977.01

SEP 12 1996

1807

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. Application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011 Group Art Unit: 1807

Filed: 07 June 1995 Examiner: D. Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

14 P
8/488,011
1825-96

AMENDMENT

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

In response to the Office Action mailed 24 July 1996, please amend the above-identified application as follows:

IN THE SPECIFICATION

On page 1, line 7, after "1995," please insert -- now abandoned, --.

On page 1, line 7, after "1994," please insert -- now abandoned, --.

On page 1, line 8, after "1994," please insert -- now abandoned, --.

On page 1, line 9, after "1994," please insert -- now abandoned, --.

On page 1, line 10, after "1994," please insert -- now abandoned, --.

IN THE CLAIMS

Please cancel claims 5, 9, 15 and 20-27.

Please amend the claims as follows:

360 FL 22-0261 10/07/96 08488011
36025 148 110.00CH

-- Claim 1 (amended). A method for screening germline of [diagnosing for a predisposition for breast and ovarian cancer in] a human subject for an alteration of a BRCA1 gene which comprises [detecting an alteration in the sequence of the BRCA1 gene or its expression products in said subject by] comparing [the] germline sequence of a [the] BRCA1 gene or BRCA1 RNA [its expression products in] from a tissue sample from said subject or a sequence of BRCA1 cDNA made from mRNA from said sample with [the] germline sequences [sequence] of [the] wild-type BRCA1 gene, wild-type BRCA1 RNA or wild-type BRCA1 cDNA [its expression products], wherein [an alteration] a difference in the sequence of the BRCA1 gene, BRCA1 RNA or BRCA1 cDNA of the subject from [the] wild-type indicates an alteration in the BRCA1 gene in said subject [a predisposition to said cancer]. --

-- Claim 3 (amended). The method of claim 1 wherein [said expression products is mRNA of the BRCA1 gene] the nucleic acid sequence of BRCA1 RNA from the subject is compared to nucleic acid sequences of wild-type BRCA1 gene, BRCA1 RNA or BRCA1 cDNA. --

-- Claim 4 (amended). The method of claim 3 wherein the [alteration of the mRNA is detected] nucleic acid sequence is compared by hybridizing [hybridization of the mRNA of said sample with] a BRCA1 gene probe ~~specific for~~ which specifically hybridizes to a BRCA1 allele to RNA isolated from said subject and detecting of the presence of a hybridization product, wherein a presence of said product indicates the presence of said allele in the subject. --

-- Claim 6 (amended). The method of claim 1 wherein [an alteration is detected in the] a regulatory [regions] region of the BRCA1 gene from said subject is compared with a regulatory region of wild-type BRCA1 gene sequences, said regulatory region corresponding to nucleotides 1-1531 of SEQ ID NO:13. --

-- Claim 7 (amended). The method of claim 1 wherein a [the alteration in the] germline nucleic acid sequence is [detected] compared by forming single-stranded DNA from a BRCA1 gene fragment from said human sample and from a corresponding fragment of a wild-type BRCA1 gene, electrophoresing said single-stranded DNAs on a non-denaturing polyacrylamide gel, comparing the [observing shifts in electrophoretic] mobility of said single-stranded [DNA]

DNAs on [non-denaturing polyacrylamide gels] said gel to determine if said single-stranded DNA is shifted relative to said probe and sequencing said single-stranded DNA having a shift in electrophoretic mobility. --

-- Claim ~~8~~ 7 (amended). The method of claim ~~1~~ 1 wherein a [the alteration in the] germline nucleic acid sequence is [detected] compared by [hybridization of] hybridizing a BRCA1 gene probe ~~specific for~~ which specifically hybridizes to a BRCA1 allele to genomic DNA isolated from said [tissue] sample and detecting the presence of a hybridization product, wherein a presence of said product indicates the presence of said allele in the subject. --

-- Claim ~~10~~ 8 (amended). The method of claim ~~1~~ 1 wherein a [the alteration in the] germline nucleic acid sequence is [detected] compared by amplifying all or part of a [the] BRCA1 gene from said sample using a set of primers to produce [an] amplified nucleic acids [sequence] and sequencing the amplified nucleic acids [sequence]. --

-- Claim ~~11~~ 9 (amended). The method of claim ~~1~~ 1 wherein a [the alteration in the] germline nucleic acid sequence is [detected] compared by amplifying all or part of a [the] BRCA1 gene using a primer specific for a specific BRCA1 mutant allele and detecting the presence of an amplified product, wherein the presence of said product indicates the presence of said specific allele. --

-- Claim ~~12~~ 10 (amended). The method of claim ~~1~~ 1 wherein a [the alteration in the] germline nucleic acid sequence is [detected] compared by molecularly cloning all or part of a [the] BRCA1 gene from said sample to produce a cloned nucleic acid [sequence] and sequencing the cloned nucleic acid [sequence]. --

-- Claim 13 (amended). The method of claim 1 wherein a [the alteration in the] germline nucleic acid sequence is [detected] compared by forming a heteroduplex consisting of a first strand of nucleic acid selected from the group consisting of [identifying a mismatch between molecules (1)] BRCA1 gene genomic DNA fragment isolated from said sample, BRCA1 RNA fragment isolated from said sample [or] and BRCA1 cDNA fragment made from mRNA

GeneDX 1009, pg. 308

[isolated] from said sample and a second strand of [(2)] a nucleic acid consisting of [probe complementary to the] a corresponding human wild-type BRCA1 gene fragment, [DNA when molecules (1) and (2 are hybridized to each other to form a duplex] analyzing for the presence of a mismatch in said heteroduplex and sequencing said first strand of nucleic acid having a mismatch. --

-- Claim ~~14~~ 12 (amended). The method of claim ~~1~~ 1 wherein a [the alteration in the] germline nucleic acid sequence is [detected] compared by amplifying [amplification of] BRCA1 nucleic acids [gene sequences in] from said sample to produce amplified nucleic acids, [and hybridization of the] hybridizing the amplified nucleic acids [sequences] to [nucleic acid probes which comprise wild-type BRCA1 gene sequences] a BRCA1 DNA probe specific for a BRCA1 allele and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said allele in the subject. --

-- Claim ~~16~~ 13 (amended). The method of claim ~~1~~ 1 wherein a [the alteration in the] germline nucleic acid sequence is [detected] compared by [screening] analyzing BRCA1 nucleic acids in said sample for a deletion mutation. --

-- Claim ~~17~~ 14 (amended). The method of claim ~~1~~ 1 wherein a [the alteration in the] germline nucleic acid sequence is [detected] compared by [screening] analyzing BRCA1 nucleic acids in said sample for a point mutation. --

-- Claim ~~18~~ 15 (amended). The method of claim ~~1~~ 1 wherein a [the alteration in the] germline nucleic acid sequence is [detected] compared by [screening] analyzing BRCA1 nucleic acids in said sample for an insertion mutation. --

-- Claim ~~19~~ 16 (amended). The method of claim ~~1~~ 1 wherein a [the alteration in the] germline nucleic acid sequence is [detected] compared by hybridizing the tissue sample *in situ* [hybridization of the BRCA1 gene] with a nucleic acid [probes which comprise the BRCA1 gene] probe specific for a BRCA1 allele and detecting the presence of a hybridization product, wherein a presence of said product indicates the presence of said allele in the subject. --

Please add the following new claims:

-- 28. The method of claim 1 wherein a nucleic acid of a germline ~~BRCA1 cDNA~~ made from mRNA from said sample is compared to nucleic acid sequences of wild-type BRCA1 gene, BRCA1 RNA or BRCA1 cDNA. --

-- 29. The method of claim 1 wherein a nucleic acid sequence of a germline BRCA1 gene from the subject is compared to nucleic acid sequences of wild-type BRCA1 gene, BRCA1 RNA or BRCA1 cDNA. --

-- 30. The method of claim 1 wherein said difference is selected from the group consisting of missense mutations within the zinc finger motif, deletions, insertions, frameshift mutations, nonsense mutations and splice site mutations. --

-- 31. A method for detecting a germline alteration in a BRCA1 gene, said alteration selected from the group consisting of the alterations set forth in Tables 11 and 12 which comprises analyzing a sequence of the BRCA1 gene or BRCA1 RNA from a human sample or analyzing the sequence of BRCA1 cDNA made from mRNA from said sample. --

-- 32. The method of claim 31 wherein a germline alteration is detected by hybridizing a BRCA1 gene probe ~~specific for~~ which specifically hybridizes to an allele of one of said alterations to RNA isolated from said human sample and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said allele in the sample. --

-- 33. The method of claim 31 wherein a germline alteration is detected by hybridizing a BRCA1 gene probe ~~specific for~~ which specifically hybridizes to one of said alterations to genomic DNA isolated from said sample and detecting of the presence of a hybridization product, wherein the presence of said product indicates the presence of said alteration in the sample. --

-- ~~34~~ 23. The method of claim ~~31~~ 20 wherein a germline alteration is detected by amplifying all or part of a BRCA1 gene in said sample using a set of primers to produce amplified nucleic acids and sequencing the amplified nucleic acids. --

-- ~~35~~ 24. The method of claim ~~31~~ 20 wherein a germline alteration is detected by amplifying part of a BRCA1 gene in said sample using a primer specific for an allele having one of said alterations and detecting the presence of an amplified product, wherein the presence of said product indicates the presence of said allele in the sample. --

-- ~~36~~ 25. The method of claim ~~31~~ 20 wherein a germline alteration is detected by molecularly cloning all or part of a BRCA1 gene in said sample to produce a cloned nucleic acid and sequencing the cloned nucleic acid. --

-- ~~37~~ 26. The method of claim ~~31~~ 20 wherein a germline alteration is detected by amplifying BRCA1 gene nucleic acids in said sample, hybridizing the amplified nucleic acids to a BRCA1 DNA probe specific for one of said alterations and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said alteration. –

-- 38. A method for screening for a germline alteration in a BRCA1 gene in a human subject which comprises analyzing a BRCA1 polypeptide from a tissue sample from said subject for an altered BRCA1 polypeptide by (i) detecting a truncated BRCA1 polypeptide or (ii) binding an antibody specific to an epitope of an altered BRCA1 polypeptide to the BRCA1 polypeptide from said sample and detecting bound antibody, wherein the presence of a truncated protein or bound antibody indicates the presence of a germline alteration in the BRCA1 gene. --

-- ~~39~~ 27. The method of claim ~~38~~ 28 wherein a BRCA1 polypeptide is analyzed by detecting a truncated BRCA1 polypeptide. --

-- 40. The method of claim 38 wherein a BRCA1 polypeptide is analyzed by binding an antibody specific to an epitope of an altered BRCA1 polypeptide to the BRCA1 polypeptide from said sample. --

REMARKS

The undersigned wishes to thank the Examiner for the courtesies extended to Dr. Mark Skolnick and himself during the interview on 10 September 1996 for Serial No. 08/487,002 during which the above application was also discussed.

The specification has been amended to indicate the current status of the parent applications. In accordance with the Examiner's suggestion, the claims have been amended to be directed to a method for screening for a germline alteration in a BRCA1 gene. The claims have further been amended to set forth each of the process steps as set forth in the specification and known in the art.

Claim 1 has further been amended so that it is directed to nucleic acid analysis. Claim 6 has been amended to specify a regulatory region of a BRCA1 gene to correspond to a sequence shown in SEQ ID NO:13. Support for this amendment can be found at page 77, lines 7-22. New claims 28 and 29 have been added to claim the material tested, similar to claim 3 and new claim 30 has been added to claim the type of alteration. New claims 31-37 have been added to claim a method for detecting the specific alterations set forth in Table 11 and 12. New claims 38-40 have been added to claim the analysis of a BRCA1 polypeptide to screen for the alterations, and find their support in claims 5, 20 and 21 as originally filed. The analysis is specifically directed to detecting a truncated protein or binding an antibody specific to an epitope of an altered BRCA1 polypeptide. It is believed that none of these amendments constitute new matter.

It is believed that the above amendments obviate the 35 USC §112, second paragraph rejection of the claims. Withdrawal of this rejection is requested.

Similarly, it is believed that the above amendments obviate the 35 USC §112, first paragraph rejection of the claims. The general nature of this rejection and claim amendments were discussed during the interview as noted on the interview summary record. Dr. Skolnick and the Examiner each noted the recognized utility for alterations in the BRCA1 gene which lead to deleterious effects in the protein. Dr. Skolnick also discussed the utility for other alterations including polymorphisms. He noted that the three types of alterations are (a) those that are neutral, (b) those that are causal and (c) those that are not clear. Each of these types of alterations are useful for making a judgment concerning the risk of developing cancer from the involvement of the BRCA1 gene. Dr. Skolnick indicated that any risk assessment has utility and

any alteration provides screening data which is used in assessing or weighing the risk of developing cancer. He noted, for example, that a person may have an alteration for which the effect is not clear. However, this person would know that she did not have a wild-type gene and therefore, on average, would have a higher risk than someone with no alteration. A sister of this person who did not have the alteration would then know that she has a normal risk from this factor. In addition, Dr. Skolnick noted that each type of alteration is useful in the analysis of ethnic background and penetrance of the BRCA1 gene in different populations. He also noted that although a BRCA1 gene heterozygosity may exist in the germline, a homozygosity in expression has been seen. Thus, it is possible to screen for a lack of a transcript. He indicated that heterozygosity in germline to homozygosity in expression may occur about 10-15% of the time.

Dr. Skolnick noted that a detection of an alteration in the germline is an indication of predisposition to cancer and a detection of a somatic alteration in a tumor is an indication of progression towards neoplasia. He indicated that a somatic mutation, including loss of all or part of the BRCA1 gene, is generally seen in the chromosome opposite the chromosome with a germline alteration in the BRCA1 gene, ans is also rarely seen in ovarian tumors.

Dr Skolnick also briefly discussed how the BRCA1 gene is currently analyzed for alterations. He indicated that 35 primers are currently used to amplify the DNA and then this DNA is sequenced. Generally one pair of primers is used per exon and surrounding intron. Since exon 11 is large, it is broken up into several overlapping primer pairs. The BRCA1 gene is amplified with each primer pair. The resulting amplified DNA is the sequenced in both the forward and reverse directions. Dr. Skolnick indicated that they are currently processing 25 BRCA1 patient samples per day. The technique of using a set of primers allows for the amplification of all or part of the BRCA1 gene as now set forth in claims 10 and 34.

The enablement of the screening for alterations in the BRCA1 gene using a polypeptide was also briefly discussed. Dr. Skolnick noted that the molecular weight of the BRCA1 polypeptide is known and is around 220 kilodaltons. He also noted that antibodies to BRCA1 polypeptide are commercially available from a number of labs. The Examiner noted that the detection of a truncated protein was enabled. She also noted that the skill in the art for the development of antibodies to proteins was really high. The nature of antibodies to detect an alteration was discussed. New claims 38 and 40 which are directed to binding an antibody to a

BRCA1 polypeptide contains the limitation that the antibody is specific for an epitope of an altered BRCA1 polypeptide. Thus, the antibody recognizes and binds to a BRCA1 polypeptide with an altered sequence. It is well known in the art that the epitope must be accessible to the antibody in order for it to bind. In view of the high skill in the art for making antibodies against proteins and determining epitopes to which they bind, it is submitted that it is not undue experimentation to prepare an antibody which would be specific for an altered BRCA1 polypeptide. To further demonstrate these facts, Applicants are in the process of preparing an appropriate declaration and will submit it as soon as it is completed.

In view of the amendments to the claims and the above remarks, it is believed that the claims are enabled in accordance with 35 USC §112, first paragraph. Withdrawal of this rejection is requested.

The claims were rejected for obviousness-type double patenting over Serial Nos. 08/409,305. Serial No. 08/409,305 was expressly abandoned on 10 September 1996. Withdrawal of this rejection is requested.

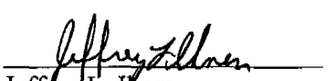
Applicants note that the claims of related application Serial No. 08/487,002 were similarly rejected over Serial No. 08/409,305 and Serial No. 08/483,553. In view of this rejection in the related application, Applicants note that the present claims are patentably distinct from the claims of Serial No. 08/483,553. The present claims are directed to methods for detecting germline alterations in the BRCA1 gene or specific germline alterations set forth in Tables 11 or 12. Serial No. 08/483,553 is directed to methods for detecting specific germline mutations set forth in Tables 12A, 14, 18 or 19. The Tables of 08/483,553 do not contain any of the alterations set forth in Tables 11 or 12 of the present application. There is no disclosure in the present application which would make the specific alterations of Serial No. 08/483,553 obvious or would suggest such specific alterations to a skilled artisan. Consequently, it is submitted that the claims of the present application are patentably distinct over the claims of Serial No. 08/483,553.

The claims of this application were also rejected for obviousness-type double patenting over the claims of Serial No. 08/487,002. In view of this rejection, Applicants are submitting a Terminal Disclaimer with this Amendment. In addition, Applicants direct the attention of the Examiner to a further related application Serial No. 08/483,554 which contains claims to wild-type BRCA1, mutant BRCA1 and a method of identifying mutant BRCA1.

Finally, Applicants note that an Information Disclosure Statement was filed on 12 October 1996 and a Supplemental Information Disclosure Statement was filed on 23 February 1996. If either of these papers are not in the file, the Examiner is requested to telephone the undersigned so that he may provide the Examiner with copies of the papers filed.

In view of the above amendments and remarks, it is believed that the claims satisfy the provisions of 35 USC §112 and are in condition for allowance. Since it is believed that all of the issues raised in the Office Action have been overcome, the Examiner is invited to telephone the undersigned to resolve any further issues and to expedite the prosecution of this application. Reconsideration of this application and early notice of allowance is requested.

Respectfully submitted,



Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, LLP
1201 New York Avenue, N.W., Suite 1000
Washington, D.C. 20005
(202) 962-4810

Attorney Docket No.: 24884-109347-09
Dated: 12 September 1996
DC2/29139.01

MAIL ROOM
SEP
5 17
1996
PAT. & TRADEMARK OFF.

AF 1807

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. Application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011 Group Art Unit: 1807

Filed: 07 June 1995 Examiner: D. Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

RECEIVED
SEP 2 6 1996
GROUP 1800

COMMUNICATION

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

Further to the Amendment filed 12 September 1996 for the above application, the fee for additional claims has been calculated as shown below:

	Claims Remaining After Amendment	Highest Number Previous	Present No. Extra Claims	Rate	Large Entity Fee Due
TOTAL	29	27	2	x $22	$ 44.00
INDEP.	3	2	0	x $37	-0-
			TOTAL DUE		$ 44.00

Please charge the above fee, and any additional filing fees required under 37 CFR 1.116, to Deposit Account No. 22-0261. Two (2) additional copies of this letter are attached.

Respectfully submitted,

Jeffrey L. Ihnen
Reg. No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI
1201 New York Avenue N.W., Suite 1000
Washington, D.C. 20005
202-962-4810

360 BA 22-0261 09/25/96 08488011
36010 102 44.00CH

Filed: 17 September 1996
Atty. Docket No. 24884-109347-09

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. Application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011 Group Art Unit: 1807

Filed: 07 June 1995 Examiner: D. Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

15/C
3800
11-19-96

SUPPLEMENTAL AMENDMENT

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

Further to the Amendment filed 12 September 1996, please amend the above-identified application as follows:

IN THE CLAIMS

Please add the following new claims:

-- 41. The method of claim 31 wherein said alteration comprises the alteration comprising a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1. --

-- 42. The method of claim 32 wherein said alteration comprises the alteration comprising a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1. --

175

-- ~~43~~ 32. The method of claim ~~33~~ 22 wherein said alteration ~~comprises the alteration~~ consists of ~~comprising~~ a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1. --

-- ~~44~~ 33. The method of claim ~~34~~ 23 wherein said alteration ~~comprises the alteration~~ consists of ~~comprising~~ a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1. --

-- ~~45~~ 34. The method of claim ~~35~~ 24 wherein said alteration ~~comprises the alteration~~ consists of ~~comprising~~ a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1. --

-- ~~46~~ 35. The method of claim ~~36~~ 25 wherein said alteration ~~comprises the alteration~~ consists of ~~comprising~~ a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1. --

-- ~~47~~ 36. The method of claim ~~37~~ 26 wherein said alteration ~~comprises the alteration~~ consists of ~~comprising~~ a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1. --

REMARKS

The claims have been amended to add dependent claims directed to detecting one of the specific alterations set forth in Table 11. It is believed that this amendment does not constitute new matter.

Also submitted herewith is a Wolfert Declaration to support the statements made on pages 8 and 9 of the Amendment filed on 12 September 1996.

GeneDX 1009, pg. 318

Favorable consideration of this application is requested.

Respectfully submitted,

Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, LLP
1201 New York Avenue, N.W., Suite 1000
Washington, D.C. 20005
(202) 962-4810

Attorney Docket No.: 24884-109347-08
Dated: 7 November 1996
DC2/29139.01

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. Application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011 Group Art Unit: 1807

Filed: 07 June 1995 Examiner: D. Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

#16 Declaration 11-01-96

DECLARATION OF ROBERT L. WOLFERT

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

I, Robert L. Wolfert, hereby declare as follows:

1. I am employed as Manager, Immunodiagnostics Program at Hybritech/Beckman. I have been an employee of Hybritech for over 12 years, and have headed the monoclonal antibody development group for over 10 years. I hold a Ph.D. in biochemistry, received from Tufts University School of Medicine in 1981 , and completed a post-doctoral fellowship in Immunology at Scripps Clinic and Research Foundation. I have published several articles in the field of monoclonal antibody development. My brief C.V. is attached as Exhibit 1.

2. Hybritech is a licensee of the above-identified patent application for detection of the BRCA1 gene product.

3. In my opinion, it would not require undue experimentation to identify and develop monoclonal antibodies against the epitopes of an altered (i.e., mutated) BRCA1 polypeptide, in light of the high level of skill in the art. This process involves well-documented techniques in immunization, hybridoma cell fusion technology, and antibody screening strategies, to generate and select monoclonal antibodies that can serve as immunological probes with highly defined specificities. Indeed, in the literature over more than a decade, monoclonal antibodies have been described that can differentiate between minute changes in molecular structure, including single amino acid differences in proteins, oxidation states of lipids, stereoisomers and post-translational

modification of identical gene products, even down to a single glucose molecule. A partial list of such literature is attached as Exhibit 2.

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Dated: 10/8/96

Robert L. Wolfert

Exhibit 1

ROBERT L. WOLFERT
10495 Canyon Lake Drive
San Diego, CA 92131
619/578-0151 (Home)
619/535-8652 (Office)

Education

1981 Tufts University School of Medicine, Boston, Massachusetts
Ph.D., Department of Biochemistry
Dissertation title: Antigen binding and complement fixation studies on the mouse myeloma IgM A22

1975 Cornell University, Ithaca, New York
Bachelor of Arts, majors in chemistry and biology

Professional Experience

1984-Present Hybritech Incorporated, San Diego, California
Director of Research, 1990-Present
Research Manager, 1986-1990
Research Scientist, 1984-1986

1981-1984 Scripps Clinic and Research Foundation, La Jolla, California
Post-doctoral fellow, Department of Immunology
Research concentration: Biochemical mechanisms of signal transduction in the activation of B cells

Awards/Honors

NIH, National Research Service Award, 1982-1984
Leukemia Society of America Fellowship, 1981-1982

Professional Associations

American Association for the Advancement of Science
American Society of Protein Chemists
American Association of Clinical Chemists
American Association of Cancer Research

Exhibit 2

REFERENCES

1. **Huse, W., L. Sastry, S. Iverson, A. Kang, M. Alting-Mees, D. Burton, S. Benkovic, and R. Lerner.** 1989. Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. *Science* **246:**1275-1281.

2. **Lerner, R., S. Benkovic, and P. Schultz.** 1991. At the crossroads of chemistry and immunology: catalytic antibodies. *Science* **252:**659-67.

3. **McCormack, R. T., J. R. Ludwig, and R. L. Wolfert.** 1992. Advances in design, generation, and manipulation of monoclonal antibodies. In R. M. Nakamura, Y. Kasahara, and G. A. Rechnitz (Eds.), *Immunochemical assays and biosensor technology for the 1990s* (pp.57-82). Washington, D.C: American Society for Microbiology.

4. **McKenzie, S. J., P. Marks, T. Lam, J. Morgan, D. Panicali, K. Trimpe, and W. Carney.** 1989. Generation and characterization of monoclonal antibodies specific for the human neu oncogene product, p185. *Oncogene* **4:** 543-548.

5. **Mudgett-Hunter, M., W. Anderson, E. Haber, and M. N. Margolies.** 1985. Binding and structural diversity among high-affinity monoclonal anti-digoxin antibodies. *Molecular Immunology***22,** No. 4, 477-488.

6. **Posner, B., J. Smiley, I. Lee, and S. Benkovic.** 1994. Catalytic antibodies: perusing combinatorial libraries. *TIBS* **19**:145-50.

7. **Schmitz, H., H. Atassi, and M. Atassi.** 1983. Production of monoclonal antibodies with preselected submolecular binding specificities to protein antigenic sites: antibodies to sperm whale myoglobin sites. *Mol. Immunol.*B **20**:719-726.

8. **Tanaka, T., D. Slamon, and M. Cline.** 1985. Efficient generation of antibodies to oncoproteins by using synthetic peptide antigens. *Proc. Natl. Acad. Sci. USA* **82:**3400-3411.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. application of

Mark H. SKOLNICK et al.

Serial No. 08/488,011 Group Art Unit: 1807

Filed: 07 June 1995 Examiner: D. Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

RECEIVED
NOV 7 1996
MATRIX CUSTOMER SERVICE CENTER

TRANSMITTAL OF AMENDMENT

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

Transmitted herewith is an amendment in the above-identified application. The fee has been calculated as shown below:

	Claims Remaining After Amendment	Highest Number Previous	Present No. Extra Claims	Large Entity Rate	Large Entity Fee Due
INDEP.	3	3	0	x $80.00	$ 0.00
TOTAL	36	29	7	x $22.00	154.00
			TOTAL DUE		$154.00

Please charge the above fee, and any additional filing fees required under 37 CFR 1.16 or 1.17 to Deposit Account No. 22-0261. Two (2) additional copies of this letter are attached.

Respectfully submitted,

Jeffrey L. Ihnen, Reg. No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, *LLP*
1201 New York Avenue N.W., Suite 1000
Washington, D.C. 20005
202-962-4848

Atty. Docket No. 24884-109347-09
Filed: 7 November 1996

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. Application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011 Group Art Unit: 1807

Filed: 07 June 1995 Examiner: D. Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

SECOND SUPPLEMENTAL AMENDMENT

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

Further to the Amendment filed 12 September 1996 and the subsequently filed Supplemental Amendment, please amend the above-identified application as follows:

IN THE CLAIMS

Please add the following new claim:

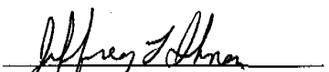
-- 48. A kit for screening for a germline alterationin a BRCA1 gene in a human subject which comprises at least one antibody (i) specific for a truncated BRCA1 polypeptide or (ii) specific to an epitope of an altered BRCA1 polypeptide. --

REMARKS

The claims have been amended to add a claim to a kit for detecting a germline alteration which contains an antibody. It is believed that this amendment does not constitute new matter.

Favorable consideration of this application is requested.

Respectfully submitted,



Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, LLP
1201 New York Avenue, N.W., Suite 1000
Washington, D.C. 20005
(202) 962-4810

Attorney Docket No.: 24884-109347-09
Dated: 14 January 1997
DC2/29139.01

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. application of

Mark H. SKOLNICK et al.

Serial No. 08/488,011 Group Art Unit: 1807

Filed: 07 June 1995 Examiner: D. Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

TRANSMITTAL OF AMENDMENT

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

Transmitted herewith is an amendment in the above-identified application. The fee has been calculated as shown below:

	Claims Remaining After Amendment	Highest Number Previous	Present No. Extra Claims	Large Entity Rate	Large Entity Fee Due
INDEP.	4	3	1	x $80.00	$ 80.00
TOTAL	37	36	1	x $22.00	22.00
			TOTAL DUE		$102.00

Please charge the above fee, and any additional filing fees required under 37 CFR 1.16 or 1.17 to Deposit Account No. 22-0261. Two (2) additional copies of this letter are attached.

Respectfully submitted,

Jeffrey L. Ihnen, Reg. No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, *LLP*
1201 New York Avenue N.W., Suite 1000
Washington, D.C. 20005
202-962-4810

Atty. Docket No. 24884-109347-09
Filed: 14 January 1997

UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, DC 20231

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.
08/[illegible]11	01/[illegible]/96	SKOLNICK M	24884-109347

18M2/0415

VENABLE BAETJER HOWARD AND CIVILETTI
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

EXAMINER	
REES,D	
ART UNIT	**PAPER NUMBER**
1807	

DATE MAILED: 04/15/97

Please find below and/or attached an Office communication concerning this application or proceeding.

Commissioner of Patents and Trademarks

PTO-90C (REV. 2/95) 1-File Copy

☆U.S.GOVERNMENT PRINTING OFFICE 1995-319-826

Office Action Summary

Application No.	Applicant(s)
08/488,011	Skolnick et al.
Examiner	**Group Art Unit**
Dianne Rees	1807

☒ Responsive to communication(s) filed on *6/6/96,912/96,11/7/96* .

☒ This action is **FINAL.**

☐ Since this application is in condition for allowance except for formal matters, **prosecution as to the merits is closed** in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11; 453 O.G. 213.

A shortened statutory period for response to this action is set to expire *3* month(s), or thirty days, whichever is longer, from the mailing date of this communication. Failure to respond within the period for response will cause the application to become abandoned. (35 U.S.C. § 133). Extensions of time may be obtained under the provisions of 37 CFR 1.136(a).

Disposition of Claims JR 4/12/97

☒ Claim(s) *1-4, 6-8, 10-14, 16-19, and 28-48* is/are pending in the application.

Of the above, claim(s) _____ is/are withdrawn from consideration.

☒ Claim(s) *1-3, 6, 10-12, 14, 16-19, 28-31, 34-39, and 41-47* is/are allowed.

☒ Claim(s) *4, 7, 8, 13, 32, 33, 40, and 48* is/are rejected.

☐ Claim(s) _____ is/are objected to.

☐ Claims _____ are subject to restriction or election requirement.

Application Papers

☐ See the attached Notice of Draftsperson's Patent Drawing Review, PTO-948.

☐ The drawing(s) filed on _____ is/are objected to by the Examiner.

☐ The proposed drawing correction, filed on _____ is ☐ approved ☐ disapproved.

☐ The specification is objected to by the Examiner.

☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. § 119

☐ Acknowledgement is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d).

☐ All ☐ Some* ☐ None of the CERTIFIED copies of the priority documents have been

☐ received.

☐ received in Application No. (Series Code/Serial Number) _____ .

☐ received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

*Certified copies not received: _____ .

☐ Acknowledgement is made of a claim for domestic priority under 35 U.S.C. § 119(e).

Attachment(s)

☐ Notice of References Cited, PTO-892

☐ Information Disclosure Statement(s), PTO-1449, Paper No(s). _____

☐ Interview Summary, PTO-413

☐ Notice of Draftsperson's Patent Drawing Review, PTO-948

☐ Notice of Informal Patent Application, PTO-152

--- SEE OFFICE ACTION ON THE FOLLOWING PAGES ---

GeneDX 1009, pg. 329

DETAILED ACTION

The Applicant's amendments and arguments filed 9/12/96, and 11/7/96 have been thoroughly reviewed. Rejections and/or objections not reiterated from the previous office action are hereby withdrawn. The following rejections and/or objections are either newly applied or are reiterated. They constitute the complete set being presently applied to the present application. Response to applicant's arguments follow.

Claim Rejections - 35 USC § 112

1. Claims 4,7,8,13,32,33,40 and 48 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

The following phrases render the claims vague and indefinite:

In claim 4,the recitation of 'specific for" is indefinite as it is unclear what the basis of the specificity is. The claim might be amended to recite --which specifically hybridizes--.

See claim 8, 32,33

Claim 7 is indefinite in the recitation of "forming a single stranded DNA from a BRCA1 gene" and "from a corresponding fragment" in that a fragment if the wild type gene in the sample was not previously recited so it is not clear what a corresponding fragments , Further the same problem arises as above in that the fragments of the genes are not normally part of a biological sample.

See claim 13 as well.

Claim 40 is indefinite in the recitation of "binding an antibody specific to an epitope of an altered BRCA1 is indefinite" as it is unclear if the antibody is specific binds the epitope that defines the alteration or if the antibody may recognize nonaltered BRCA1 polypeptides as well.

Claim 48 is indefinite in the recitation of "alterationin" as the claim appears to intend -- alteration in--. IT is further unclear what the specificity of the antibodies being claimed are . Do they recognize wild type as well as mutant proteins or only mutant proteins? (This is more relevant to (ii) than (i))

Claims 1-3,6,10-12,14,16-19,28-31,34-39,41-47 are allowed.

Response to Applicant's arguments:

The rejections of claims under 35 USC 112 first paragraph and second paragraph are withdrawn in view of Applicant's amendments and arguments.

The Opinion Declaration of Robert L. Wolfert attesting to the skill in the art in generating epitope specific antibodies is considered and found persuasive. It is further noted by the Examiner that workers in the field have currently had success in generating antibodies to the BRCA1 protein and various domains of the protein using Applicant's methodology and therefore it is the position of the Examiner that given the high skill in the art and the amount of guidance in the specification, one would be able to make and use the antibodies claimed.

Applicant's amendments have necessitated a new grounds of rejection under 35 USC 112 second paragraph.

2. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for response to this final action is set to expire THREE MONTHS from the date of this action. In the event a first response is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event will

the statutory period for response expire later than SIX MONTHS from the date of this final action.

Papers related to this application may be submitted to Group 1800 by facsimile transmission via the P.T.O. Fax Center located in Crystal Mall 1. The CM1 Fax Center number is (703) 305-7939. Please note that the faxing of such papers must conform with the notice to Comply published in the Official Gazette, 1096 OG 30 (Nov 15, 1989).

An inquiry regarding this communication should be directed to examiner Dianne Rees, Ph.D., whose telephone number is (703) 308-6565. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, W. Gary Jones, can be reached on (703) 308-1152.

Calls of a general nature may be directed to the Group receptionist who may be reached at (703) 308-0196.

Dianne Rees
4/12/97

W. GARY JONES
SUPERVISORY PATENT EXAMINER
GROUP 1800
4/14/97

Form PTO 948 (Rev. 10-94) U.S. DEPARTMENT OF COMMERCE - Patent and Trademark Office Application No. 488,011

NOTICE OF DRAFTSPERSON'S PATENT DRAWING REVIEW

PTO Draftpersons review all originally filed drawings regardless of whether they are designated as formal or informal. Additionally, patent Examiners will review the drawings for compliance with the regulations. Direct telephone inquiries concerning this review to the Drawing Review Branch, 703-305-8404.

The drawings filed (insert date) 1/5/96, are
A. ___ not objected to by the Draftsperson under 37 CFR 1.84 or 1.152.
B. X objected to by the Draftsperson under 37 CFR 1.84 or 1.152 as indicated below. The Examiner will require submission of new, corrected drawings when necessary. Corrected drawings must be submitted according to the instructions on the back of this Notice.

1. DRAWINGS. 37 CFR 1.84(a): Acceptable categories of drawings: Black ink. Color.
   - ___ Not black solid lines. Fig(s)___
   - ___ Color drawings are not acceptable until petition is granted. Fig(s) ___
2. PHOTOGRAPHS. 37 CFR 1.84(b)
   - ___ Photographs are not acceptable until petition is granted. Fig(s) ___
   - ___ Photographs not properly mounted (must use brystol board or photographic double-weight paper). Fig(s)___
   - ___ Poor quality (half-tone). Fig(s)___
3. GRAPHIC FORMS. 37 CFR 1.84 (d)
   - ___ Chemical or mathematical formula not labeled as separate figure. Fig(s)___
   - ___ Group of waveforms not presented as a single figure, using common vertical axis with time extending along horizontal axis. Fig(s)___
   - ___ Individuals waveform not identified with a separate letter designation adjacent to the vertical axis. Fig(s)___
4. TYPE OF PAPER. 37 CFR 1.84(c)
   - ___ Paper not flexible, strong, white, smooth, nonshiny, and durable. Sheet(s)___
   - X Erasures, alterations, overwritings, interlineations, cracks, creases, and folds copy machine marks not accepted. Fig(s) ___
   - ___ Mylar, velum paper is not acceptable (too thin). Fig(s) ___
5. SIZE OF PAPER. 37 CFR 1.84(f): Acceptable sizes:
   - 21.6 cm. by 35.6 cm. (8 1/2 by 14 inches)
   - 21.6 cm. by 33.1 cm. (8 1/2 by 13 inches)
   - 21.6 cm. by 27.9 cm. (8 1/2 by 11 inches)
   - 21.0 cm. by 29.7 cm. (DIN size A4)
   - ___ All drawing sheets not the same size. Sheet(s)___
   - ___ Drawing sheet not an acceptable size. Sheet(s) ___
6. MARGINS. 37 CFR 1.84(g): Acceptable margins:

Paper size	21.6 cm. X 35.6 cm. (8 1/2 X 14 inches)	21.6 cm X 33.1 cm. (8 1/2 X 13 inches)	21.6 cm. X 27.9 cm. (8 1/2 X 11 inches)	21.0 cm. X 29.7 cm. (DIN Size A4)
T	5.1 cm. (2")	2.5 cm. (1")	2.5 cm. (1")	2.5cm.
L	.64 cm. (1/4")	.64 cm. (1/4")	.64 cm. (1/4")	2.5 cm.
R	.64 cm. (1/4")	.64 cm. (1/4")	.64 cm. (1/4")	1.5 cm.
B	.64 cm. (1/4")	.64 cm. (1/4")	.64 cm. (1/4")	1.0 cm.

Margins do not conform to chart above.
Sheet(s)___
___Top (T) ___Left (L) ___Right (R) ___Bottom (B)

7. VIEWS. 37 CFR 1.84(h)
   REMINDER: Specification may require revision to correspond to drawing changes.
   - ___ All views not grouped together. Fig(s)___
   - ___ Views connected by projection lines or lead lines. Fig(s)___
   - Partial views. 37 CFR 1.84(h) 2
   - ___ View and enlarged view not labled separatly or properly. Fig(s)___
   - Sectional views. 37 CFR 1.84 (h) 3
   - ___ Hatching not indicated for sectional portions of an object. Fig(s)___
   - ___ Cross section not drawn same as view with parts in cross section with regularly spaced parallel oblique strokes. Fig(s)___
8. ARRANGEMENT OF VIEWS. 37 CFR 1.84(i)
   - ___ Words do not appear on a horizontal, left-to-right fashion when page is either upright or turned so that the top becomes the right side, except for graphs. Fig(s)___
9. SCALE. 37 CFR 1.84(k)
   - ___ Scale not large enough to show mechanism with crowding when drawing is reduced in size to two-thirds in reproduction. Fig(s)___
   - ___ Indication such as "actual size" or scale 1/2" not permitted. Fig(s)___
10. CHARACTER OF LINES, NUMBERS, & LETTERS. 37 CFR 1.84(l)
    - X Lines, numbers & letters not uniformly thick and well defined, clean, durable, and black (except for color drawings). Fig(s) 1-10A
11. SHADING. 37 CFR 1.84(m)
    - ___ Solid black shading areas not permitted. Fig(s)___
    - ___ Shade lines, pale, rough and blurred. Fig(s)___
12. NUMBERS, LETTERS, & REFERENCE CHARACTERS. 37 CFR 1.84(p)
    - ___ Numbers and reference characters not plain and legible. 37 CFR 1.84(p)(l) Fig(s)___
    - ___ Numbers and reference characters not oriented in same direction as the view. 37 CFR 1.84(p)(l) Fig(s)___
    - ___ English alphabet not used. 37 CFR 1.84(p)(2) Fig(s)___
    - ___ Numbers, letters, and reference characters do not measure at least .32 cm. (1/8 inch) in height. 37 CFR(p)(3) Fig(s)___
13. LEAD LINES. 37 CFR 1.84(q)
    - ___ Lead lines cross each other. Fig(s)___
    - ___ Lead lines missing. Fig(s)___
14. NUMBERING OF SHEETS OF DRAWINGS. 37 CFR 1.84(t)
    - ___ Sheets not numbered consecutively, and in Arabic numerals, beginning with number 1. Sheet(s)___
15. NUMBER OF VIEWS. 37 CFR 1.84(u)
    - ___ Views not numbered consecutively, and in Arabic numerals, beginning with number 1. Fig(s)___
    - ___ View numbers not preceded by the abbreviation Fig. Fig(s)___
16. CORRECTIONS. 37 CFR 1.84(w)
    - ___ Corrections not made from prior PTO-948. Fig(s)___
17. DESIGN DRAWING. 37 CFR 1.152
    - ___ Surface shading shown not appropriate. Fig(s)___
    - ___ Solid black shading not used for color contrast. Fig(s)___

COMMENTS:

ATTACHMENT TO PAPER NO.___ REVIEWER___ DATE 5/

PTO Copy

MAIL ROOM
SEP
39 12
1996
PAT. & TRADEMARK OFF.

1807

19
LF
4/29/97

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011 Group Art Unit: 1807

Filed: 07 June 1995 Examiner: D. Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

TERMINAL DISCLAIMER

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

The owners, UNIVERSITY OF UTAH RESEARCH FOUNDATION, THE UNITED STATES OF AMERICA AS REPRESENTED BY THE SECRETARY OF THE DEPARTMENT OF HEALTH AND HUMAN SERVICES, and MYRIAD GENETICS, INC., of one hundred percent (100%) interest in the instant application, hereby disclaim, except as provided below, the terminal part of the statutory term of any patent granted on the instant application, which would extend beyond the expiration date of the full statutory term defined in 35 U.S.C. §§154 to 156 and 173, as shortened by any terminal disclaimer filed prior to the grant of any patent granted on pending second Application Serial No. 08/487,002, filed on 07 June 1995. The owners hereby agree that any patent so granted on the instant application shall be enforceable only for and during such period that it and any patent granted on the second application are commonly owned. This agreement runs with any patent granted on the instant application, and is binding upon grantee, its successors or assigns.

In making the above disclaimer, the owners do not disclaim the terminal part of any patent granted on the instant application that would extend to the expiration date of the full statutory term as defined in 35 U.S.C. §154 to 156 and 173 of any patent granted on the second application, as shortened by any terminal disclaimer filed prior to the patent grant, in the event that any such granted patent: (1) expires for failure to pay a maintenance fee, (2) is held unenforceable, (3) is

230 EK 22 04/18/97 08488011
03080 145 110.00CH

found invalid by a court of competent jurisdiction, (4) is statutorily disclaimed in whole or terminally disclaimed under 37 C.F.R. §1.321, (5) has all claims cancelled by a reexamination certificate, (6) is reissued, or (7) is in any manner terminated prior to the expiration of its full statutory term as shortened by any terminal disclaimer filed prior to its grant.

The undersigned is an attorney of record in the instant application. Please charge the requisite fee to Deposit Account No. 22-0261. Two additional copies of this document are enclosed.

Respectfully submitted,



Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, *LLP*
1201 New York Avenue, N.W., #1000
Washington, D.C. 20005
(202) 962-4810

Atty. Docket No. 24884-109347-09
Dated: 12 September 1996
22887.01

MAIL ROOM
APR 30 1997
PAT. & TRADEMARK OFF.

Initial Review
BOX AF

BOX AF
AMENDMENT AFTER FINAL
EXPEDITED PROCESSING

20/E
A.G.J
5/3/97
X(NE)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. Application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011

Group Art Unit: 1807

Filed: 07 June 1995

Examiner: D. Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

RECEIVED
MAY 02 1997
GROUP 1800

AMENDMENT AFTER FINAL

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

OK TO ENTER DR 5/18/97

In response to the Office Action mailed 15 April 1997, please amend the above-identified application as follows:

IN THE SPECIFICATION:

Please replace pages 103-167 with new pages 103-167, which are submitted herewith.

IN THE CLAIMS

In claim 4, line 3, change "specific for" to read -- which specifically hybridizes to --.

In claim 8, line 3, change "specific for" to read -- which specifically hybridizes to --.

In claim 32, line 2, change "specific for" to read -- which specifically hybridizes to --.

In claim 33, line 2, change "specific for" to read -- which specifically hybridizes to --.

Please amend claims 7, 13, 40 and 48 as follows:

E

E1

-- Claim 7 (twice amended). The method of claim 1 wherein a germline nucleic acid sequence is compared by obtaining a first BRCA1 gene fragment from a BRCA1 gene from a human sample and a second BRCA1 gene fragment from a wild-type BRCA1 gene, said second fragment corresponding to said first fragment, forming single-stranded DNA from said first [a] BRCA1 gene fragment [from said human sample] and from said second BRCA1 gene [a corresponding] fragment [of a wild-type BRCA1 gene], electrophoresing said single-stranded DNAs on a non-denaturing polyacrylamide gel, comparing the mobility of said single-stranded DNAs on said gel to determine if said single-stranded DNA from said first BRCA1 gene fragment is shifted relative to said [probe] second BRCA1 gene fragment and sequencing said single-stranded DNA from said first BRCA1 gene fragment having a shift in electrophoretic mobility. --

E2

-- Claim 13 (twice amended). The method of claim 1 wherein a germline nucleic acid sequence is compared by [forming a heteroduplex consisting of] obtaining a first BRCA1 gene fragment [strand of nucleic acid selected] from [the group consisting of] (a) BRCA1 gene genomic DNA [fragment] isolated from said sample, (b) BRCA1 RNA [fragment] isolated from said sample [and] or (c) BRCA1 cDNA [fragment] made from mRNA isolated from said sample, obtaining [and] a second BRCA1 gene fragment from (a) [strand of a nucleic acid consisting of a corresponding human] wild-type BRCA1 genomic DNA, (b) wild-type BRCA1 RNA or (c) wild-type cDNA made from wild-type mRNA, said second BRCA1 gene fragment corresponding to said first BRCA1 gene fragment, forming single-stranded DNA from said first BRCA1 gene fragment and from said second BRCA1 gene fragment, forming a heteroduplex consisting of single-stranded DNA from said first BRCA1 gene fragment and single-stranded DNA from said second BRCA1 gene fragment, [gene fragment,] analyzing the heteroduplex to determine if said single-stranded DNA from said first BRCA1 gene fragment has [for the presence of] a mismatch [in said heteroduplex] relative to said single-stranded DNA from said second BRCA1 gene fragment and sequencing said first single-stranded DNA from said first BRCA1 gene fragment [strand of nucleic acid] having a mismatch. --

E3

-- Claim 38 (amended). A method for screening for a germline alteration in a BRCA1 gene in a human subject which comprises analyzing a BRCA1 polypeptide from a tissue sample

GeneDX 1009, pg. 338

from said subject for an altered BRCA1 polypeptide by (i) detecting either a full length BRCA1 polypeptide or a truncated BRCA1 polypeptide or (ii) [binding] contacting an antibody [specific] which specifically binds to an epitope of an altered BRCA1 polypeptide to the BRCA1 polypeptide from said sample and detecting bound antibody, wherein the presence of a truncated protein or bound antibody indicates the presence of a germline alteration in the BRCA1 gene. --

E3 cont'd

-- Claim ~~40~~ 29 (amended). The method of claim ~~38~~ 27 wherein a BRCA1 polypeptide is analyzed by [binding] contacting an antibody [specific] which specifically binds to an epitope of an altered BRCA1 polypeptide to the BRCA1 polypeptide from said sample. --

E4

-- Claim ~~48~~ 37 (amended). A kit for screening for [a germline alterationin] an alteration in a BRCA1 gene in a human subject which comprises at least one antibody (i) [specific for] which specifically binds to wild-type BRCA1 polypeptide but not a truncated BRCA1 polypeptide or (ii) [specific] which specifically binds to an epitope of an altered BRCA1 polypeptide. --

E5

REMARKS

A substitute computer readable form and a substitute paper copy of the Sequence Listing are submitted with this Final Amendment in response to a discussion with the Examiner. The original Sequence Listing had an error in listing the length of SEQ ID NO:2 as 1864 amino acids, when in fact it is 1863 amino acids. This was because of an error in the PatentIn program which asked for the stop codon to be included as part of the coding region, and then counted this as an amino acid residue. To correct this error, the Feature for SEQ ID NO:1 regarding CDS (coding sequence) has been changed to indicate that the coding sequence is from nucleotide 120 to nucleotide 5708 rather than between 120 and 5711. Nucleotides 5709-5711 correspond to the stop codon. By making this change, SEQ ID NO:2 (which is automatically prepared by the PatentIn program) correctly shows a length of 1863 amino acids. Other changes made in the Sequence Listing are that information (specifically, the application number, filing date and classification information) concerning the Current Application Data were added. This information was unknown at the time the original Sequence Listing was prepared.

It is urged that the changes and additions add no new matter to the application. No changes other than those noted above have been made in the Sequence Listing which is

178

GeneDX 1009, pg. 339

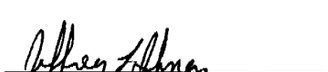
submitted herewith, as compared to the originally filed Sequence Listing. It is requested that the changes made by this Amendment be entered into the application.

The claims have been amended in view of the Examiner's suggestions and comments concerning the claim language. Specifically, claims 4, 8, 32 and 33 have been amended to change the language for the gene probe from "specific for" to "specifically hybridizes to." Claims 7 and 13 have been amended to set forth a step of obtaining fragments. Claim 13 has further been amended to include a step of forming single-stranded DNA and to provide proper antecedent basis for a "fragment having a mismatch." Claim 38 has been amended to indicate that the antibody comparison is by "contacting an antibody which specifically binds to an epitope of an altered BRCA1 polypeptide with a BRCA1 polypeptide from said sample and detecting bound antibody." Claim 40 has been similarly amended as to the specificity of antibody binding. Claim 48 has been amended to indicate that the antibody "specifically binds" to the items specified in the claim. It is believed that none of these amendments constitute new matter.

It is believed that the above amendments obviate the 35 USC §112, second paragraph rejection of the claims. Withdrawal of this rejection is requested.

In view of the above amendments and remarks, it is believed that the claims satisfy the provisions of 35 USC §112 and are in condition for allowance. Since it is believed that all of the issues raised in the Office Action have been overcome, the Examiner is invited to telephone the undersigned to resolve any further issues and to expedite the prosecution of this application. Reconsideration of this application and early notice of allowance are requested.

Respectfully submitted,



Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, LLP
1201 New York Avenue, N.W., Suite 1000
Washington, D.C. 20005
(202) 962-4810

Attorney Docket No.: 24884-109347-09
Dated: 30 April 1997
DC2/34222.01

Interview Summary	Application No. 08/488,011	Applicant(s) Skolnick et al.
	Examiner Dianne Rees	Group Art Unit 1807

All participants (applicant, applicant's representative, PTO personnel):

(1) *Dianne Rees* (3) _____

(2) *Jeffrey Inhen* (4) _____

Date of Interview *May 19, 1997*

Type: ☒ Telephonic ☐ Personal (copy is given to ☐ applicant ☐ applicant's representative).

Exhibit shown or demonstration conducted: ☐ Yes ☒ No. If yes, brief description:

Agreement ☒ was reached. ☐ was not reached.

Claim(s) discussed: *41-47*

Identification of prior art discussed:

Description of the general nature of what was agreed to if an agreement was reached, or any other comments:
Obtained authorization for Examiner's amendment to claims 41-47 to put the case in condition for allowance.

(A fuller description, if necessary, and a copy of the amendments, if available, which the examiner agreed would render the claims allowable must be attached. Also, where no copy of the amendents which would render the claims allowable is available, a summary thereof must be attached.)

1. ☒ It is not necessary for applicant to provide a separate record of the substance of the interview.

Unless the paragraph above has been checked to indicate to the contrary, A FORMAL WRITTEN RESPONSE TO THE LAST OFFICE ACTION IS NOT WAIVED AND MUST INCLUDE THE SUBSTANCE OF THE INTERVIEW. (See MPEP Section 713.04). If a response to the last Office action has already been filed, APPLICANT IS GIVEN ONE MONTH FROM THIS INTERVIEW DATE TO FILE A STATEMENT OF THE SUBSTANCE OF THE INTERVIEW.

2. ☒ Since the Examiner's interview summary above (including any attachments) reflects a complete response to each of the objections, rejections and requirements that may be present in the last Office action, and since the claims are now allowable, this completed form is considered to fulfill the response requirements of the last Office action. Applicant is not relieved from providing a separate record of the interview unless box 1 above is also checked.

Examiner Note: You must sign and stamp this form unless it is an attachment to a signed Office action.

GeneDX 1009, pg. 341

Notice of Allowability	Application No. 08/488,011	Applicant(s) Skolnick et al.
	Examiner Dianne Rees	Group Art Unit 1807

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance and Issue Fee Due or other appropriate communication will be mailed in due course.

☒ This communication is responsive to *9/12/96,4/30/97*.

☒ The allowed claim(s) is/are *1-4, 6-8, 10-13, 13, 14, 14-19, and 28-48*.

☐ The drawings filed on _____ are acceptable.

☐ Acknowledgement is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d).

☐ All ☐ Some* ☐ None of the CERTIFIED copies of the priority documents have been

☐ received.

☐ received in Application No. (Series Code/Serial Number) _____.

☐ received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

*Certified copies not received: _____.

☐ Acknowledgement is made of a claim for domestic priority under 35 U.S.C. § 119(e).

A SHORTENED STATUTORY PERIOD FOR RESPONSE to comply with the requirements noted below is set to EXPIRE **THREE MONTHS** FROM THE "DATE MAILED" of this Office action. Failure to timely comply will result in ABANDONMENT of this application. Extensions of time may be obtained under the provisions of 37 CFR 1.136(a).

☐ Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL APPLICATION, PTO-152, which discloses that the oath or declaration is deficient. A SUBSTITUTE OATH OR DECLARATION IS REQUIRED.

☒ Applicant MUST submit NEW FORMAL DRAWINGS

☐ because the originally filed drawings were declared by applicant to be informal.

☒ including changes required by the Notice of Draftsperson's Patent Drawing Review, PTO-948, attached hereto or to Paper No. *7*.

☐ including changes required by the proposed drawing correction filed on _____, which has been approved by the examiner.

☐ including changes required by the attached Examiner's Amendment/Comment.

Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the reverse side of the drawings. The drawings should be filed as a separate paper with a transmittal lettter addressed to the Official Draftsperson.

☐ Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Any response to this letter should include, in the upper right hand corner, the APPLICATION NUMBER (SERIES CODE/SERIAL NUMBER). If applicant has received a Notice of Allowance and Issue Fee Due, the ISSUE BATCH NUMBER and DATE of the NOTICE OF ALLOWANCE should also be included.

Attachment(s)

☐ Notice of References Cited, PTO-892

☒ Information Disclosure Statement(s), PTO-1449, Paper No(s). *9,10*

☐ Notice of Draftsperson's Patent Drawing Review, PTO-948

☐ Notice of Informal Patent Application, PTO-152

☒ Interview Summary, PTO-413

☒ Examiner's Amendment/Comment

☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material

☒ Examiner's Statement of Reasons for Allowance

GeneDX 1009, pg. 342

DETAILED ACTION

1. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

2. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

Authorization for this examiner's amendment was given in a telephone interview with Jeffrey Inhen on May 19, 1997

The application has been amended as follows:

In claims 41-47, the words "comprises the alteration comprising" has been deleted and --consists of-- has been inserted.

3. The following is an examiner's statement of reasons for allowance:

The claims are drawn to methods of detecting germline alterations in the BRCA1 gene by detecting alterations in BRCA1 nucleic acids or in the products expressed by these nucleic acids, such as by the use of eptiope specific antibodies. In a further embodiment of the invention , the method is used to detect specific polymorphisms in the BRCA1 gene. Applicants are first to discover the sequence of the wild type BRCA1 gene and disclose and enable the methods of detection recited broadly herein (Applicant's declaration by Dr. Wolfert was considered persuasive with regards to method claims employing altered -epitope specific antibodies). The skill in the art in this technology is high and such antibodies have been routinely made for a number of years. It is further noted that workers in the field have generated antibodies to various domains of the BRCA1 protein subsequent to the filing of this application using the same methodology disclosed in the instant application. Detection of germline mutations in the BRCA1 gene provide an indication of increased susceptibility to breast and ovarian cancer.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dianne Rees whose telephone number is (703) 308-6565.

Dianne Rees
5/19/97

W. GARY JONES
SUPERVISORY PATENT EXAMINER
GROUP 1800
5/20/97

May 19, 1997

DETAILED ACTION

1. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

2. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

Authorization for this examiner's amendment was given in a telephone interview with Jeffrey Inhen on May 19, 1997

The application has been amended as follows:

In claims 41-47, the words "comprises the alteration comprising" has been deleted and --consists of-- has been inserted.

3. The following is an examiner's statement of reasons for allowance:

The claims are drawn to methods of detecting germline alterations in the BRCA1 gene by detecting alterations in BRCA1 nucleic acids or in the products expressed by these nucleic acids, such as by the use of eptiope specific antibodies. In a further embodiment of the invention , the method is used to detect specific polymorphisms in the BRCA1 gene. Applicants are first to discover the sequence of the wild type BRCA1 gene and disclose and enable the methods of detection recited broadly herein (Applicant's declaration by Dr. Wolfert was considered persuasive with regards to method claims employing altered -epitope specific antibodies). The skill in the art in this technology is high and such antibodies have been routinely made for a number of years. It is further noted that workers in the field have generated antibodies to various domains of the BRCA1 protein subsequent to the filing of this application using the same methodology disclosed in the instant application. Detection of germline mutations in the BRCA1 gene provide an indication of increased susceptibility to breast and ovarian cancer.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dianne Rees whose telephone number is (703) 308-6565.

W. GARY JONES
SUPERVISORY PATENT EXAMINER
GROUP 1800
5/20/97

Dianne Rees
5/19/97

May 19, 1997

UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office

Address: Box ISSUE FEE
ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, D.C. 20231

NOTICE OF ALLOWANCE AND ISSUE FEE DUE

18M2/0522
VENABLE BAETJER HOWARD AND CIVILETTI
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

APPLICATION NO.	FILING DATE	TOTAL CLAIMS	EXAMINER AND GROUP ART UNIT	DATE MAILED
08/488,011	01/05/96	037	REES, D 1807	05/22/97

First Named Applicant: SKOLNICK, MARK H.

TITLE OF INVENTION: 17Q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

ATTY'S DOCKET NO.	CLASS-SUBCLASS	BATCH NO.	APPLN. TYPE	SMALL ENTITY	FEE DUE	DATE DUE
1 24884-109347	435-006.000	N28	UTILITY	NO	$1290.00	08/22/97

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED.

THE ISSUE FEE MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED.

HOW TO RESPOND TO THIS NOTICE:

I. Review the SMALL ENTITY status shown above.
If the SMALL ENTITY is shown as yes, verify your current SMALL ENTITY status:

A. If the status is changed, pay twice the amount of the FEE DUE shown and notify the Patent and Trademark Office of the change in status, or
B. If the status is the same, pay the FEE DUE shown above.

If the SMALL ENTITY is shown as NO:

A. Pay FEE DUE shown above, or

B. File verified statementof Small Entity Status before, or with, payment of 1/2 the FEE DUE shown above.

II. Part B of this notice should be completed and returned to the Patent and Trademark Office (PTO) with your ISSUE FEE. Even if the ISSUE FEE has already been paid by charge to deposit account, Part B should be completed and returned. If you are charging the ISSUE FEE to your deposit account, section "6b" of Part B should be completed.

III. All communications regarding this application must give application number and batch number. Please direct all communication prior to issuance to Box ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: ***Patents issuing on applications filed on or after Dec. 12, 1980 may require payment of maintenance fees. It is patentee's responsibility to ensure timely payment of maintenance fees when due.***

RAW SEQUENCE LISTING
PATENT APPLICATION *US/08/488,011B*

INPUT SET: S17442.raw

This Raw Listing contains the General Information Section and up to the first 5 pages.

#23

ENTERED

SEQUENCE LISTING

(1) General Information:

(i) APPLICANT: Skolnick, Mark H.
Goldgar, David E.
Miki, Yoshio
Swenson, Jeff
Kamb, Alexander
Harshman, Keith D.
Shattuck-Eidens, Donna M.
Tavtigian, Sean V.
Wiseman, Roger W.
Futreal, P. Andrew

(ii) TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer Susceptibility Gene

(iii) NUMBER OF SEQUENCES: 85

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
(B) STREET: 1201 New York Avenue, N.W., Suite 1000
(C) CITY: Washington
(D) STATE: DC
(E) COUNTRY: USA
(F) ZIP: 20005

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/488,011
(B) FILING DATE: 07-JUN-1995
(C) CLASSIFICATION: 435

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/409,305
(B) FILING DATE: 24-MAR-1995

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/348,824
(B) FILING DATE: 29-NOV-1994

```
47
48      (vii) PRIOR APPLICATION DATA:
49            (A) APPLICATION NUMBER: US 08/308,104
50            (B) FILING DATE: 16-SEP-1994
51
52      (vii) PRIOR APPLICATION DATA:
53            (A) APPLICATION NUMBER: US 08/300,266
54            (B) FILING DATE: 02-SEP-1994
55
56      (vii) PRIOR APPLICATION DATA:
57            (A) APPLICATION NUMBER: US 08/289,221
58            (B) FILING DATE: 12-AUG-1994
59
60       (viii) ATTORNEY/AGENT INFORMATION:
61            (A) NAME: Ihnen, Jeffrey L.
62            (B) REGISTRATION NUMBER: 28,957
63            (C) REFERENCE/DOCKET NUMBER: 24884-109347-09
64
65       (ix) TELECOMMUNICATION INFORMATION:
66            (A) TELEPHONE: 202-962-4810
67            (B) TELEFAX: 202-962-8300
68
69
70   (2) INFORMATION FOR SEQ ID NO:1:
71
72        (i) SEQUENCE CHARACTERISTICS:
73            (A) LENGTH: 5914 base pairs
74            (B) TYPE: nucleic acid
75            (C) STRANDEDNESS: double
76            (D) TOPOLOGY: linear
77
78      (ii) MOLECULE TYPE: cDNA
79
80     (iii) HYPOTHETICAL: NO
81
82      (iv) ANTI-SENSE: NO
83
84      (vi) ORIGINAL SOURCE:
85            (A) ORGANISM: Homo sapiens
86
87      (ix) FEATURE:
88            (A) NAME/KEY: CDS
89            (B) LOCATION: 120..5708
90
91
92      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
93
94   AGCTCGCTGA GACTTCCTGG ACCCCGCACC AGGCTGTGGG GTTTCTCAGA TAACTGGGCC      60
95
96   CCTGCGCTCA GGAGGCCTTC ACCCTCTGCT CTGGGTAAAG TTCATTGGAA CAGAAAGAA      119
97
98   ATG GAT TTA TCT GCT CTT CGC GTT GAA GAA GTA CAA AAT GTC ATT AAT         167
99   Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
```

```
100     1               5                  10                  15
101
102   GCT ATG CAG AAA ATC TTA GAG TGT CCC ATC TGT CTG GAG TTG ATC AAG        215
103   Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
104                20                  25                  30
105
106
107   GAA CCT GTC TCC ACA AAG TGT GAC CAC ATA TTT TGC AAA TTT TGC ATG        263
108   Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
109             35                  40                  45
110
111   CTG AAA CTT CTC AAC CAG AAG AAA GGG CCT TCA CAG TGT CCT TTA TGT        311
112   Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
113        50                  55                  60
114
115   AAG AAT GAT ATA ACC AAA AGG AGC CTA CAA GAA AGT ACG AGA TTT AGT        359
116   Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
117    65                  70                  75                  80
118
119   CAA CTT GTT GAA GAG CTA TTG AAA ATC ATT TGT GCT TTT CAG CTT GAC        407
120   Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
121                  85                  90                  95
122
123   ACA GGT TTG GAG TAT GCA AAC AGC TAT AAT TTT GCA AAA AAG GAA AAT        455
124   Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
125              100                 105                 110
126
127   AAC TCT CCT GAA CAT CTA AAA GAT GAA GTT TCT ATC ATC CAA AGT ATG        503
128   Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
129          115                 120                 125
130
131   GGC TAC AGA AAC CGT GCC AAA AGA CTT CTA CAG AGT GAA CCC GAA AAT        551
132   Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
133      130                 135                 140
134
135   CCT TCC TTG CAG GAA ACC AGT CTC AGT GTC CAA CTC TCT AAC CTT GGA        599
136   Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
137   145                 150                 155                 160
138
139   ACT GTG AGA ACT CTG AGG ACA AAG CAG CGG ATA CAA CCT CAA AAG ACG        647
140   Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
141                  165                 170                 175
142
143   TCT GTC TAC ATT GAA TTG GGA TCT GAT TCT TCT GAA GAT ACC GTT AAT        695
144   Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
145              180                 185                 190
146
147   AAG GCA ACT TAT TGC AGT GTG GGA GAT CAA GAA TTG TTA CAA ATC ACC        743
148   Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
149          195                 200                 205
150
151   CCT CAA GGA ACC AGG GAT GAA ATC AGT TTG GAT TCT GCA AAA AAG GCT        791
152   Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
```

```
153      210                 215                 220
154
155  GCT TGT GAA TTT TCT GAG ACG GAT GTA ACA AAT ACT GAA CAT CAT CAA       839
156  Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
157  225                 230                 235                 240
158
159  CCC AGT AAT AAT GAT TTG AAC ACC ACT GAG AAG CGT GCA GCT GAG AGG       887
160  Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg
161                  245                 250                 255
162
163  CAT CCA GAA AAG TAT CAG GGT AGT TCT GTT TCA AAC TTG CAT GTG GAG       935
164  His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu
165              260                 265                 270
166
167  CCA TGT GGC ACA AAT ACT CAT GCC AGC TCA TTA CAG CAT GAG AAC AGC       983
168  Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser
169          275                 280                 285
170
171  AGT TTA TTA CTC ACT AAA GAC AGA ATG AAT GTA GAA AAG GCT GAA TTC      1031
172  Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe
173      290                 295                 300
174
175  TGT AAT AAA AGC AAA CAG CCT GGC TTA GCA AGG AGC CAA CAT AAC AGA      1079
176  Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg
177  305                 310                 315                 320
178
179  TGG GCT GGA AGT AAG GAA ACA TGT AAT GAT AGG CGG ACT CCC AGC ACA      1127
180  Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr
181                  325                 330                 335
182
183  GAA AAA AAG GTA GAT CTG AAT GCT GAT CCC CTG TGT GAG AGA AAA GAA      1175
184  Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu
185              340                 345                 350
186
187  TGG AAT AAG CAG AAA CTG CCA TGC TCA GAG AAT CCT AGA GAT ACT GAA      1223
188  Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu
189          355                 360                 365
190
191  GAT GTT CCT TGG ATA ACA CTA AAT AGC AGC ATT CAG AAA GTT AAT GAG      1271
192  Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu
193      370                 375                 380
194
195  TGG TTT TCC AGA AGT GAT GAA CTG TTA GGT TCT GAT GAC TCA CAT GAT      1319
196  Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp
197  385                 390                 395                 400
198
199  GGG GAG TCT GAA TCA AAT GCC AAA GTA GCT GAT GTA TTG GAC GTT CTA      1367
200  Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu
201                  405                 410                 415
202
203  AAT GAG GTA GAT GAA TAT TCT GGT TCT TCA GAG AAA ATA GAC TTA CTG      1415
204  Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu
205              420                 425                 430
```

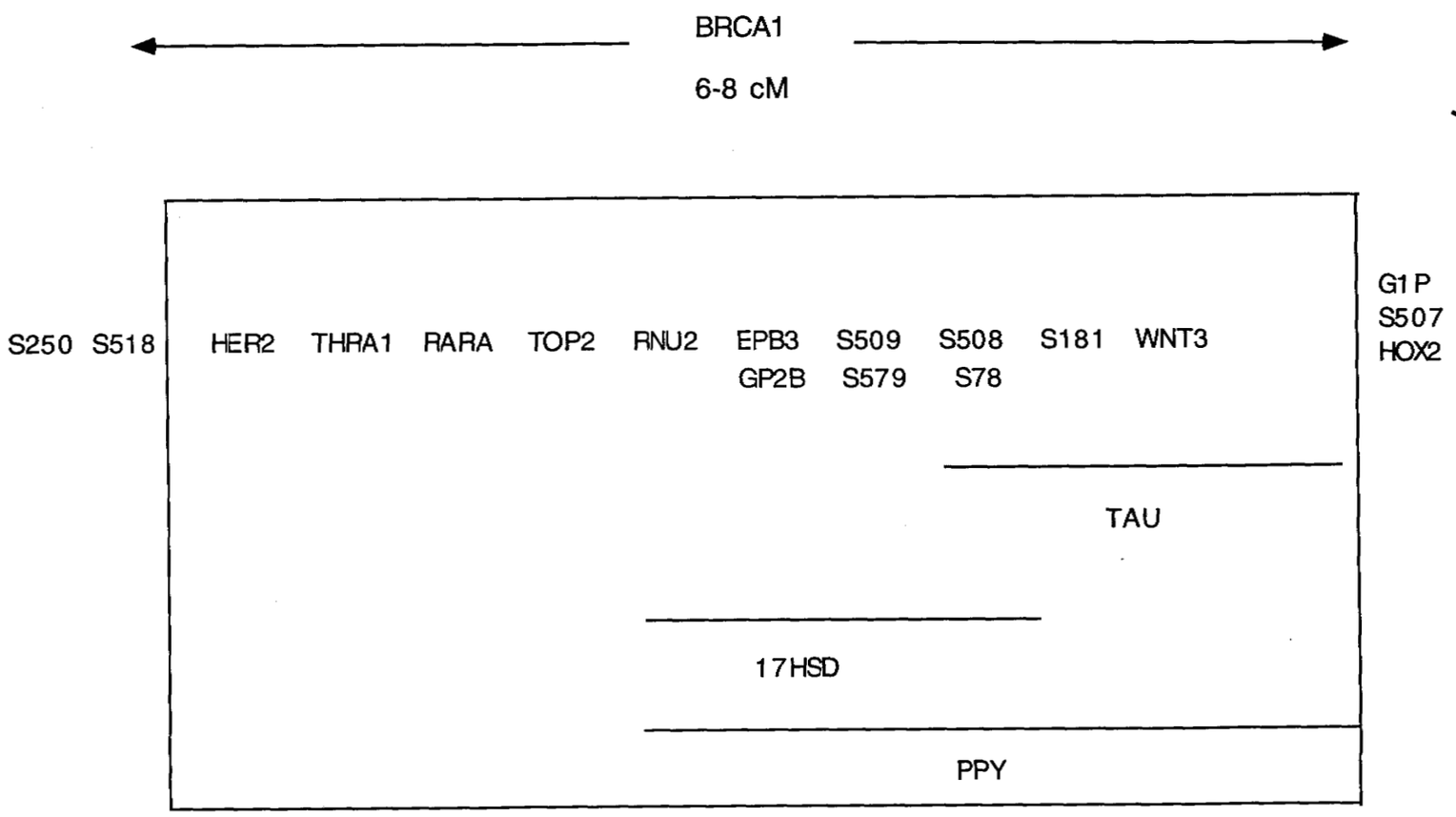
```
206
207  GCC AGT GAT CCT CAT GAG GCT TTA ATA TGT AAA AGT GAA AGA GTT CAC      1463
208  Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Glu Arg Val His
209          435                 440                 445
210
211  TCC AAA TCA GTA GAG AGT AAT ATT GAA GAC AAA ATA TTT GGG AAA ACC      1511
212  Ser Lys Ser Val Glu Ser Asn Ile Glu Asp Lys Ile Phe Gly Lys Thr
213      450                 455                 460
214
215  TAT CGG AAG AAG GCA AGC CTC CCC AAC TTA AGC CAT GTA ACT GAA AAT      1559
216  Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn
217  465                 470                 475                 480
218
219
220  CTA ATT ATA GGA GCA TTT GTT ACT GAG CCA CAG ATA ATA CAA GAG CGT      1607
221  Leu Ile Ile Gly Ala Phe Val Thr Glu Pro Gln Ile Ile Gln Glu Arg
222                  485                 490                 495
223
224  CCC CTC ACA AAT AAA TTA AAG CGT AAA AGG AGA CCT ACA TCA GGC CTT      1655
225  Pro Leu Thr Asn Lys Leu Lys Arg Lys Arg Arg Pro Thr Ser Gly Leu
226              500                 505                 510
227
228  CAT CCT GAG GAT TTT ATC AAG AAA GCA GAT TTG GCA GTT CAA AAG ACT      1703
229  His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr
230          515                 520                 525
231
232  CCT GAA ATG ATA AAT CAG GGA ACT AAC CAA ACG GAG CAG AAT GGT CAA      1751
233  Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln
234      530                 535                 540
235
236  GTG ATG AAT ATT ACT AAT AGT GGT CAT GAG AAT AAA ACA AAA GGT GAT      1799
237  Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp
238  545                 550                 555                 560
239
240  TCT ATT CAG AAT GAG AAA AAT CCT AAC CCA ATA GAA TCA CTC GAA AAA      1847
241  Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys
242                  565                 570                 575
243
244  GAA TCT GCT TTC AAA ACG AAA GCT GAA CCT ATA AGC AGC AGT ATA AGC      1895
245  Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser
246              580                 585                 590
247
248  AAT ATG GAA CTC GAA TTA AAT ATC CAC AAT TCA AAA GCA CCT AAA AAG      1943
249  Asn Met Glu Leu Glu Leu Asn Ile His Asn Ser Lys Ala Pro Lys Lys
250          595                 600                 605
251
252  AAT AGG CTG AGG AGG AAG TCT TCT ACC AGG CAT ATT CAT GCG GTT GAA      1991
253  Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Ala Leu Glu
254      610                 615                 620
255
256  CTA GTA GTC AGT AGA AAT CTA AGC CCA CCT AAT TGT ACT GAA TTG CAA      2039
257  Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln
258  625                 630                 635                 640
```

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION *US/08/488,011B*

DATE: 05/06/97
TIME: 11:43:10

INPUT SET: S17442.raw

Line	Error	Original Text



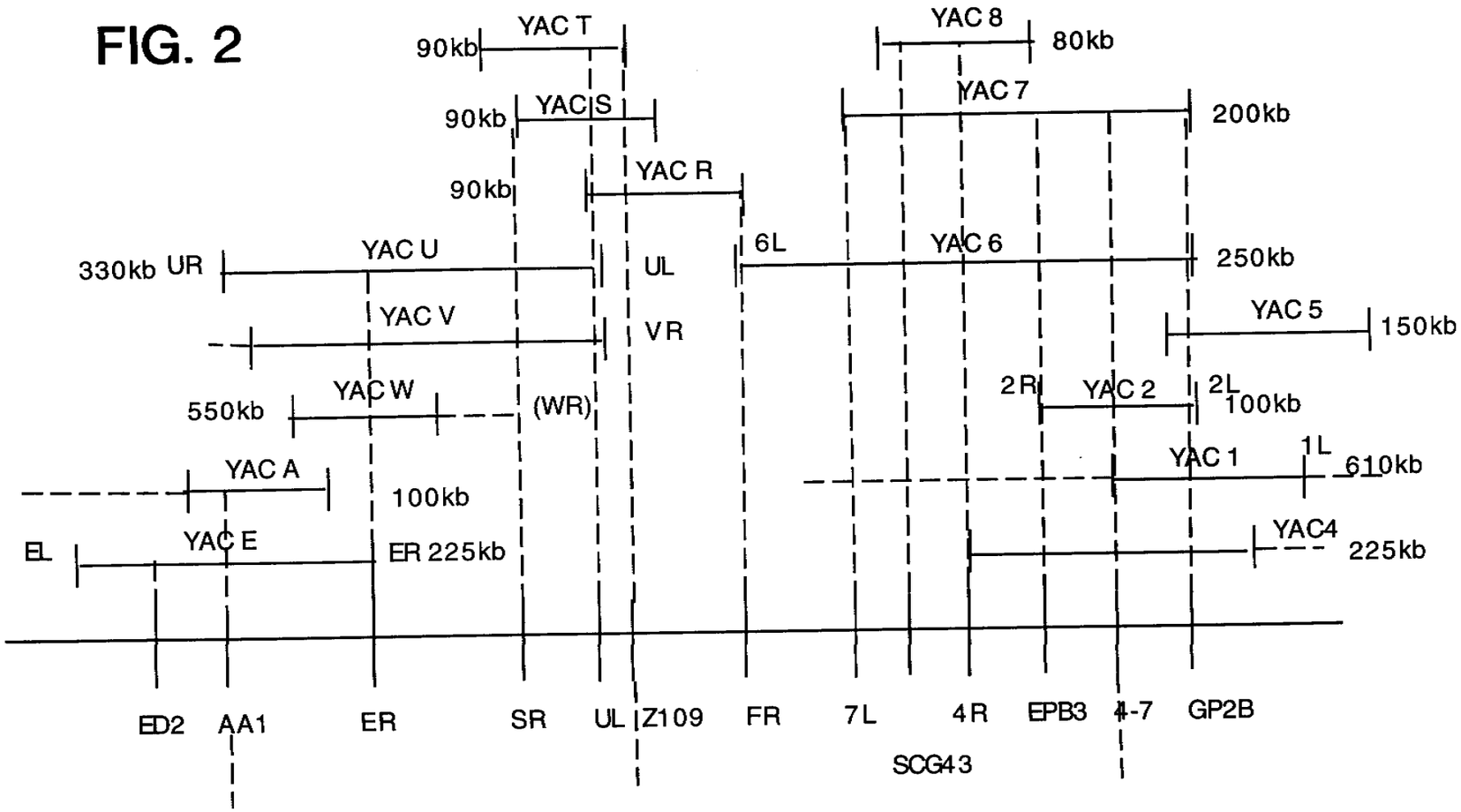
Map of the early onset breast and ovarian cancer region (BRCA1)

FIG. 1

GeneDX 1009, pg. 356

APPROVED BY DRAFTSPERSON
PATENTS
CLASS

FIG. 2



GeneDX 1009, pg. 357

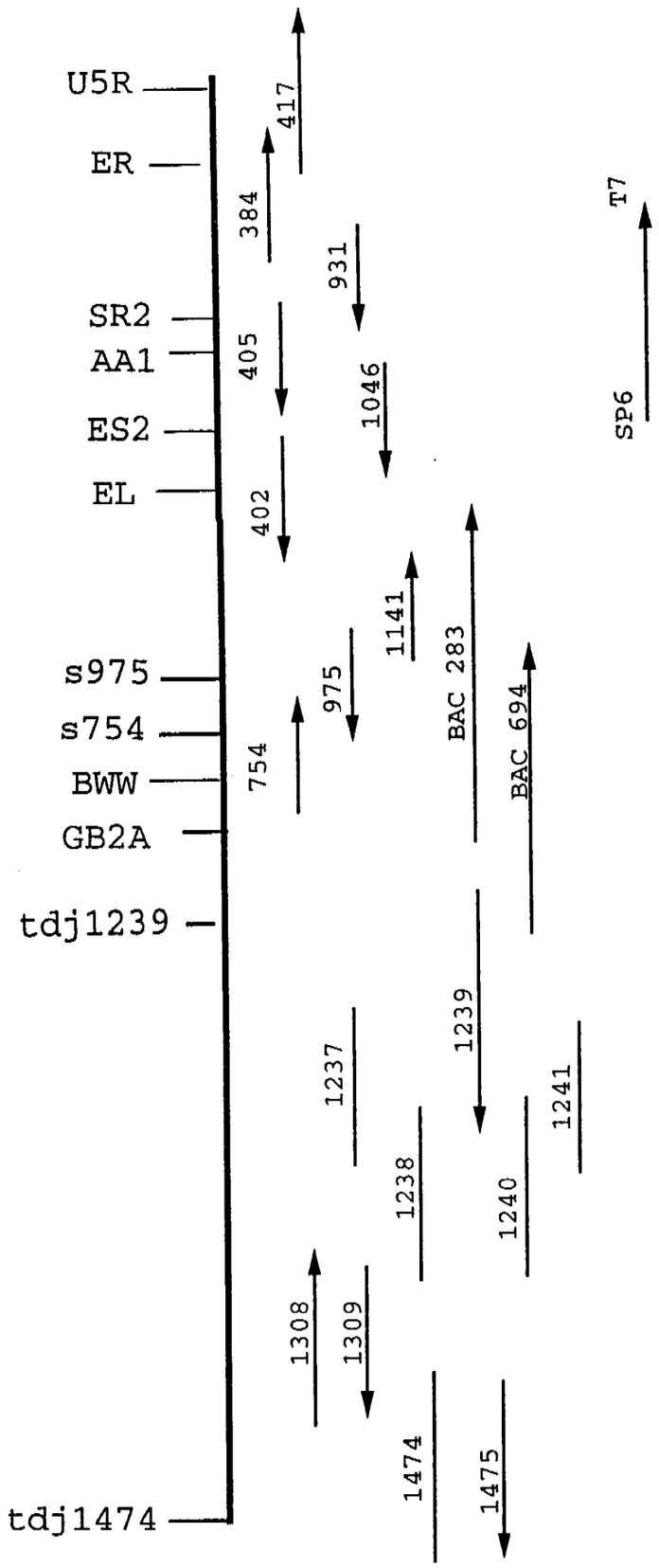

U5R
ER
SR2
AA1
ES2
EL
s975
s754
BWW
GB2A
tdj1239
tdj1474
417
384
931
405
1046
402
1141
975
754
BAC 283
BAC 694
1239
1237
1241
1238
1240
1308
1309
1474
1475
T7
SP6

FIG. 3

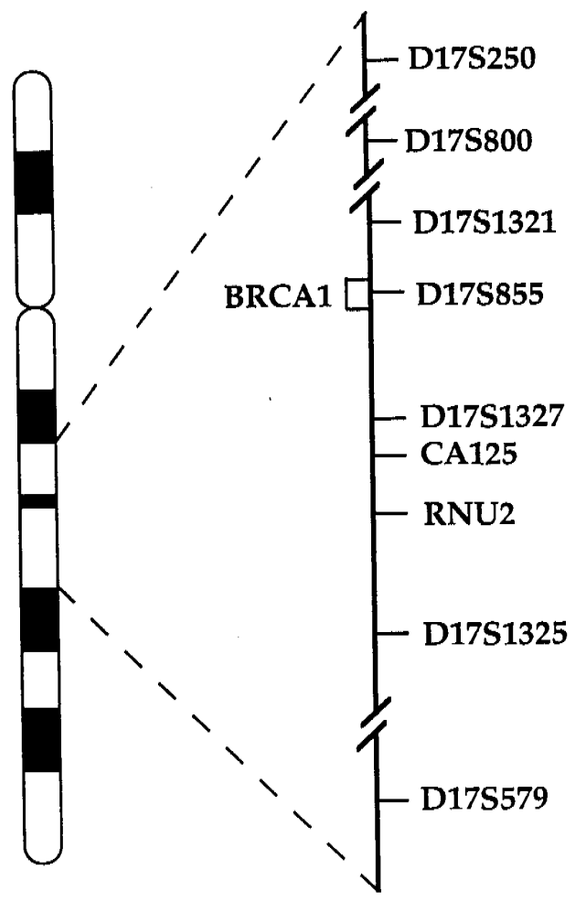


FIG. 4

SEQ. ID NO:		
82	BRCA1	CPICLELIKEPVSTK-CDHIFCKFCMLKLLNQKK---GPSQCPLCK
83	RPT1	CPICLELLKEPVSAD-CNHSFCRACITLNYESNRNTDGKGNCPVCR
84	RIN1	CPICLDMLKNTMTTKECLHRFCSDCIVTALRS-----GNKECPTCR
85	RFP1	CPVCLQYFAEPMMLD-CGHNICCACLARCWGTAC---TNVSCPQCR
	C3HC4 motif	C--C----------C-H--C--C----------C--C

FIG. 5

GeneDX 1009, pg. 359

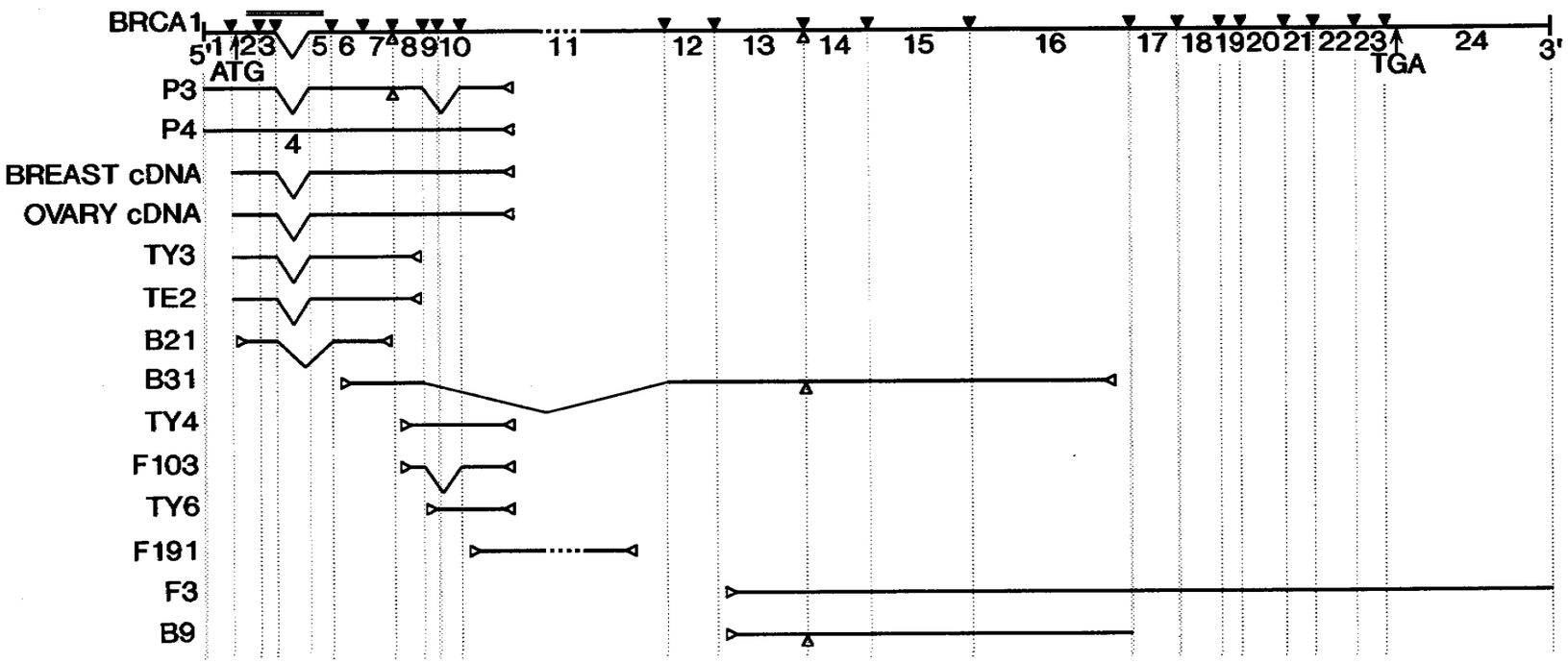

BRCA1
5'
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
3'
ATG
TGA
P3
P4
BREAST cDNA
OVARY cDNA
TY3
TE2
B21
B31
TY4
F103
TY6
F191
F3
B9

FIG. 6

GeneDX 1009, pg. 361

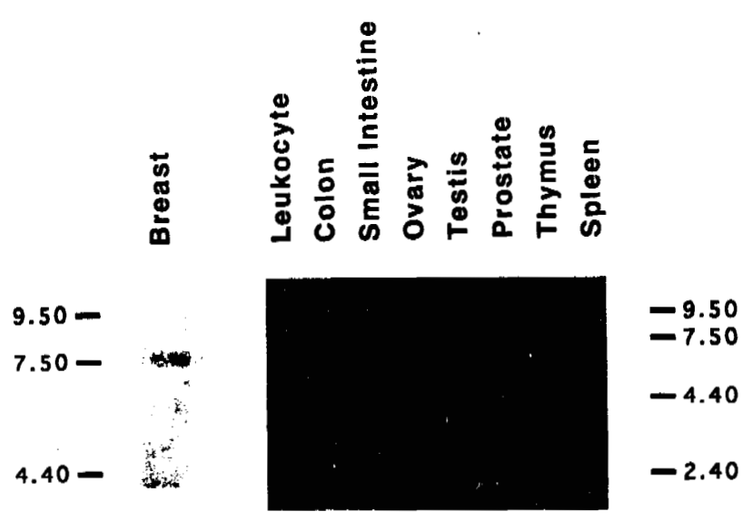


FIG. 7

ART UNIT | FIG.
BY | CLASS | SUBCLASS
DRAFTSMAN

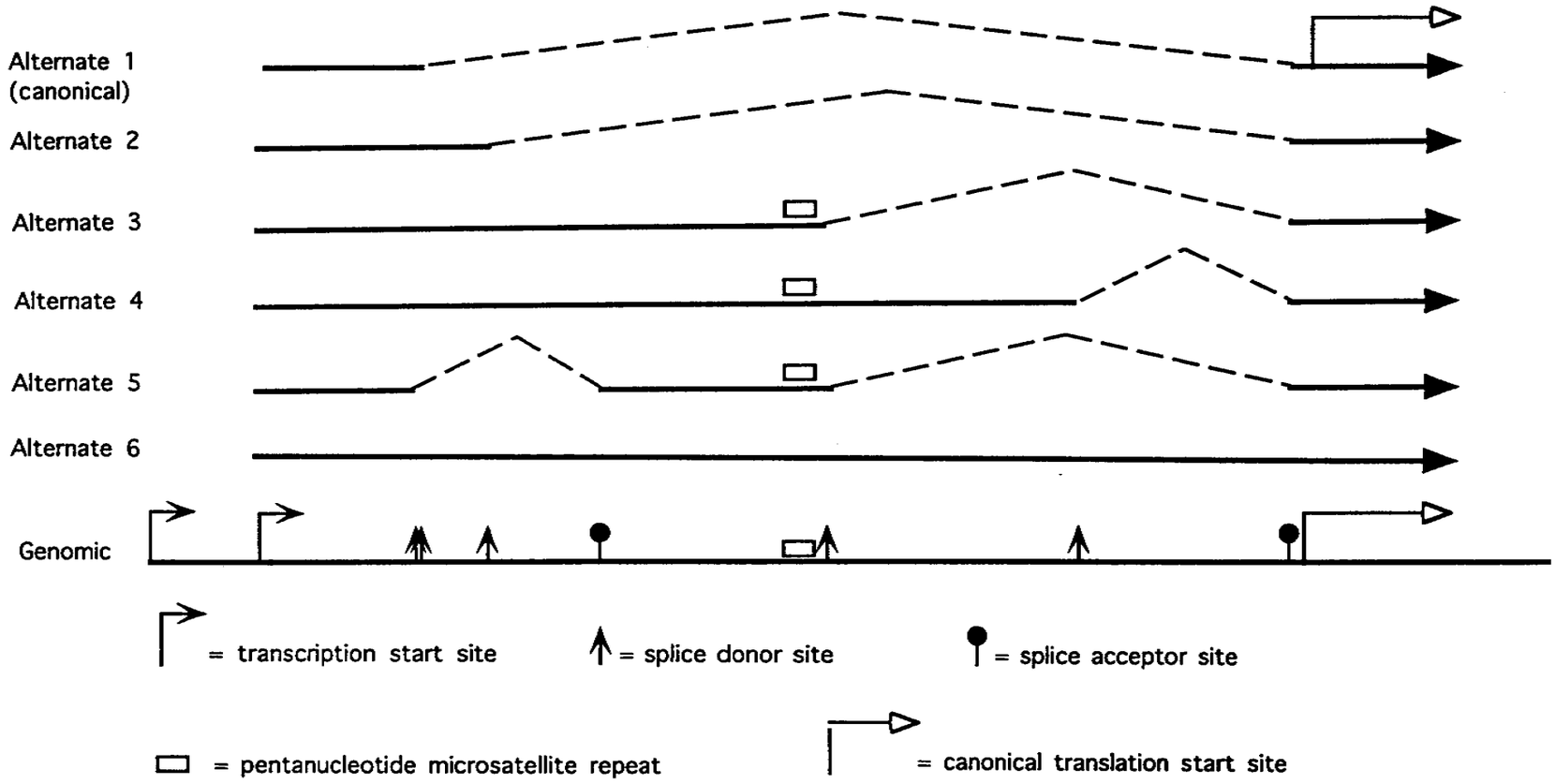


FIG. 8

GeneDX 1009, pg. 362

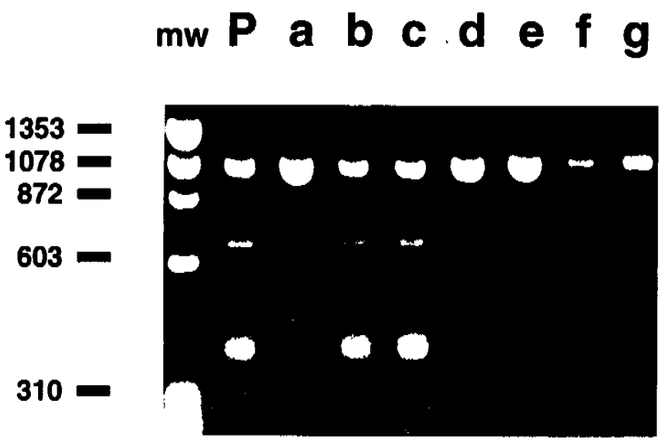
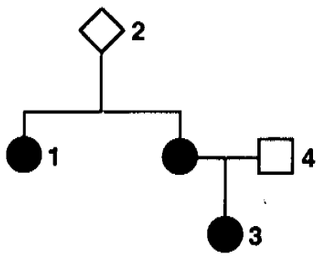


FIG. 9A



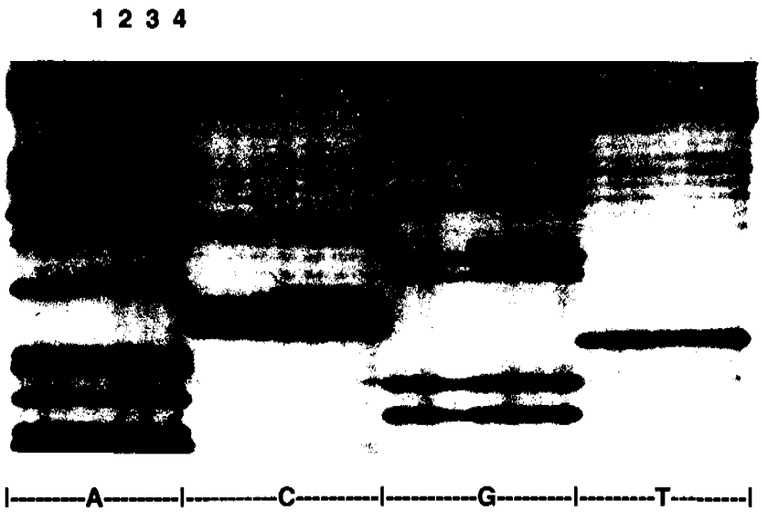
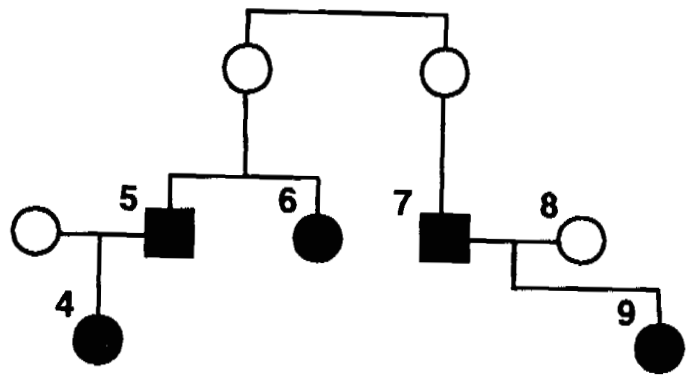


FIG. 9B



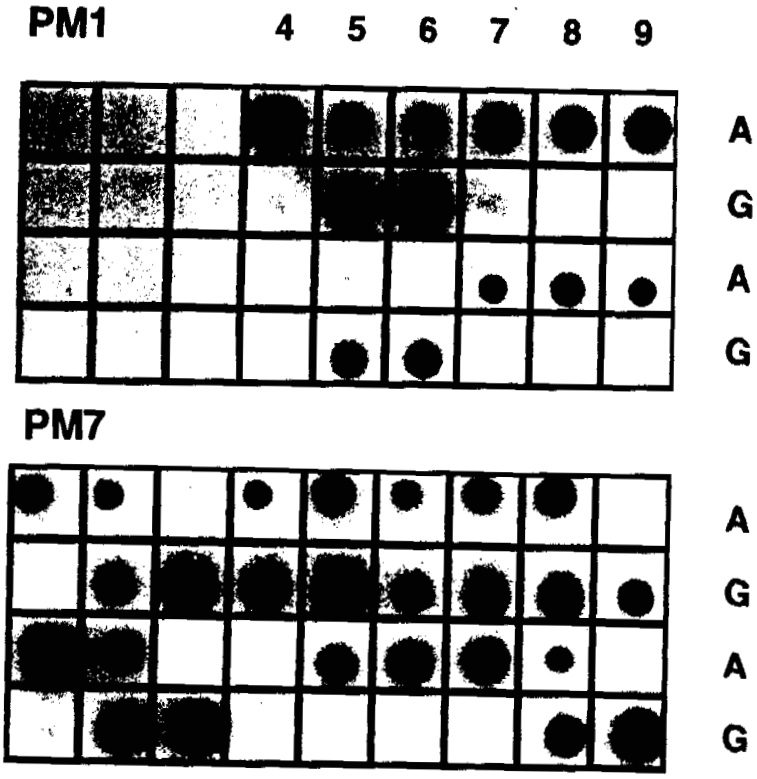


FIG. 9C

```
   1  gaggctagagggcaggcactttatggcaaactcaggtagaattcttcctcttccgtctct
  61  ttccttttacgtcatcggggagactgggtggcaatcgcagcccgagagacgcatggctct
 121  ttctgccctccatcctctgatgtaccttgatttcgtattctgagaggctgctgcttagcg
 181  gtagccccttggtttccgtggcaacggaaaagcgcgggaattacagataaattaaaactg
 241  cgactgcgcggcgtgAGCTCGCTGAGACTTCCTGGACCCCGCACCAGGCTGTGGGGTTTC
 301  TCAGATAACTGGGCCCCTGCGCTCAGGAGGCCTTCACCCTCTGCTCTGGGTAAAGgtagt
 361  agagtcccgggaaagggacaggggcccaagtgatgctctggggtactggcgtgggagag
 421  tggatttccgaagctgacagatgggtattctttgacgggggggtaggggcggaacctgaga
 481  ggcgtaaggcgttgtgaaccctggggaggggggcagtttgtaggtcgcgagggaagcgct
 541  gaggatcaggaaggggcactgagtgtccgtgggggaatcctcgtgataggaactggaat
 601  atgccttgaggggacactatgtctttaaaaacgtcggctggtcatgaggtcaggagttc
 661  cagaccagcctgaccaacgtggtgaaactccgtctctactaaaaatacnaaaattagccg
 721  ggcgtggtgccgctccagctactcaggaggctgaggcaggagaatcgctagaacccggga
 781  ggcggaggttgcagtgagccgagatcgcgccattgcactccagcctgggcgacagagcga
 841  gactgtctcaaaacaaaacaaaacaaaacaaaacaaaaaacaccggctggtatgtatgag
 901  aggatgggaccttgtggaagaagaggtgccaggaatatgtctgggaaggggaggagacag
 961  gattttgtgggagggagaacttaagaactggatccatttgcgccattgagaaagcgcaag
1021  agggaagtagaggagcgtcagtagtaacagatgctgccggcagggatgtgcttgaggagg
1081  atccagagatgagagcaggtcactgggaaaggttaggggcggggaggccttgattggtgt
1141  tggtttggtcgttgttgattttggttttatgcaagaaaaagaaaacaaccagaaacattg
1201  gagaaagctaaggctaccaccacctacccggtcagtcactcctctgtagctttctctttc
1261  ttggagaaaggaaaagacccaaggggttggcagcgatatgtgaaaaaattcagaatttat
1321  gttgtctaattacaaaaagcaacttctagaatctttaaaaataaaggacgttgtcattag
1381  ttcttctggtttgtattattctaaaaccttccaaatcttcaaatttacttttattttaaaa
1441  tgataaaatgaagttgtcattttataaacctttaaaaagatatatatatatgtttttct
1501  aatgtgttaaagTTCATTGGAACAGAAAGAAATGGATTTATCTGCTCTTCGCGTTGAAGA
1561  AGTACAAAATGTCATTAATGCTATGCAGAAAATCTTAGAGTGTCCCATCTGgtaagtcag
1621  cacaagagtgtattaatttgggattcctatgattatctcctatgcaaatgaacagaattg
1681  accttacatactagggaagaaaagacatgtctagtaagattaggctattgtaattgctga
1741  ttttcttaactgaagaactttaaaaatatagaaaatgattccttgttctccatccactct
1801  gcctctcccactcctctccttttcaacacaatcctgtggtccgggaaagacagggctctg
1861  tcttgattggttctgcactgggcaggatctgttagatactgcatttgctttctccagctc
1921  taaavvvvvvvvvvvvvaaatgctgatgatagtatagagtattgaagggatcaatataat
1981  tctgttttgatatctgaaagctcactgaaggtaaggatcgtattctctgctgtattctca
2041  gttcctgacacagcagacatttaataaatattgaacgaacttgaggccttatgttgactc
2101  agtcataacagctcaaagttgaacttattcactaagaatagctttatttttaaataaatt
2161  attgagcctcatttatttttcttttttctcccccccctaccctgctagTCTGGAGTTGATCA
2221  AGGAACCTGTCTCCACAAAGTGTGACCACATATTTTGCAAgtaagtttgaatgtgttatg
2281  tggctccattattagcttttgtttttgtccttcataacccaggaaacacctaactttata
2341  gaagctttacttttcttcaattaagtgagaacgaaaatccaactccatttcattctttctc
2401  agagagtatatagttatcaaaagttggttgtaatcatagttcctggtaaagttttgacat
2461  atattatctttttttttttttgagacaagtctcgctctgtcgcccaggctggagtgcagt
2521  ggcatgaggcttgctcactgcacctccgcccccgagttcagcgactctvvvvvvvvvvvv
2581  vtgagatctagaccacatggtcaaagagatagaatgtgagcaataaatgaaccttaaatt
2641  tttcaacagctactttttttttttttttttgagacagGGKCTTACTCTGTTGTCCCAGCT
2701  GGAGTACAGWGTGCGATCATGAGGCTTACTGTTGCTTGACTCCTAGGCTCAAGCGATCCT
2761  ATCACCTCAGTCTCCAAGTAGCTGGACTgtaagtgcacaccaccatatccagctaaattt
2821  tgtgttttctgtagagacggggtttcgccatgtttcccaggctggtcttgaactttgggc
2881  ttaacccgtctgcccacctaggcatcccaaagtgctaggattacaggtgtgagtcatcat
2941  gcctggccagtattttagttagctctgtcttttcaagtcatatacaagttcattttcttt
3001  taagtttagttaacaaccttatatcatgtattcttttctagcataaagaaagattcgagg
```

FIG. 10A

3061 ccvvvvvvvvvvvvvvtgtgatcataacagtaagccatatgcatgtaagttcagttttcat
3121 agatcattgcttatgtagtttaggtttttgcttatgcagcatccaaaaacaattaggaaa
3181 ctattgcttgtaattcacctgccattacttttttaaatggctcttaagggcagttgtgaga
3241 ttatcttttcatggctatttgccttttgagtattctttctacaaaaggaagtaaattaaa
3301 ttgttctttctttctttataatttatagATTTTGCATGCTGAAACTTCTCAACCAGAAGA
3361 AAGGGCCTTCACAGTGTCCTTTATGTAAGAATGATATAACCAAAAGgtatataatttggt
3421 aatgatgctaggttggaagcaaccacagtaggaaaaagtagaaattatttaataacatag
3481 cgttcctataaaaccattcatcagaaaaatttataaaagagtttttagcacacagtaaat
3541 tatttccaaagttatttttcctgaaagttttatgggcatctgccttatacaggtattgvvv
3601 vvvvvvvvvvggtaggcttaaatgaatgacaaaaagttactaaatcactgccatcacacg
3661 gtttatacagatgtcaatgatgtattgattatagaggttttctactgttgctgcatctta
3721 tttttatttgtttacatgtcttttcttattttagtgtccttaaaaggttgataatcactt
3781 gctgagtgtgtttctcaaacaatttaatttcagGAGCCTACAAGAAAGTACGAGATTTAG
3841 TCAACTTGTTGAAGAGCTATTGAAAATCATTTGTGCTTTTCAGCTTGACACAGGTTTGGA
3901 GTgtaagtgttgaatatcccaagaatgacactcaagtgctgtccatgaaaactcaggaag
3961 tttgcacaattactttctatgacgtggtgataagacctttttagtctaggttaattttagt
4021 tctgtatctgtaatctatttttaaaaaattactcccactggtctcacaccttatttvvvvv
4081 vvvvvvvvaaaaaatcacaggtaaccttaatgcattgtcttaacacaacaaagagcatac
4141 atagggtttctcttggtttctttgattataattcatacatttttctctaactgcaaacat
4201 aatgttttcccttgtattttacagATGCAAACAGCTATAATTTTGCAAAAAAGGAAAATA
4261 ACTCTCCTGAACATCTAAAAGATGAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACC
4321 GTGCCAAAAGACTTCTACAGAGTGAACCCGAAAATCCTTCCTTGgtaaaaccatttgttt
4381 tcttcttcttctttcttcttcttttctttttttttttctttttttttttgagatggagtctt
4441 gctctgtggcccaggctagaagcagtcctcctgccttagccnccttagtagctgggatta
4501 caggcacgcgcaccatgccaggctaatttttgtatttttagtagagacggggtttcatca
4561 tgttggccaggctggtctcgaactcctaacctcaggtgatcvvvvvvvvvvvvvatgatg
4621 gagatcttaaaaagtaatcattctggggctgggcgtagtagcttgcacctgtaatcccag
4681 cacttcgggaggctgaggcaggcagataatttgaggtcaggagtttgagaccagcctggc
4741 caacatggtgaaacccatctctactaaaaatacaaaaattagctgggtgtggtggcacgt
4801 acctgtaatcccagctactcgggaggcggaggcacaagaattgcttgaacctaggacgcg
4861 gaggttgcagcgagccaagatcgcgccactgcactccagcctgggccgtagagtgagact
4921 ctgtctcaaaaaagaaaaaaaagtaattgttctagctgggcgcagtggctcttgcctgta
4981 atcccagcactttgggaggccaaggcgggtggatctcgagtcctagagttcaagaccagc
5041 ctaggcaatgtggtgaaaccccatcgctacaaaaaatacaaaaattagccaggcatggtg
5101 gcgtgcgcatgtagtcccagctccttgggaggctgaggtgggaggatcacttgaacccag
5161 gagacagaggttgcagtgaaccgagatcacgccaccacgctccagcctgggcaacagaac
5221 aagactctgtctaaaaaaatacaaataaaataaaagtagttctcacagtaccagcattca
5281 tttttcaaaagatatagagctaaaaaggaaggaaaaaaaaagtaatgttgggctttaaa
5341 tactcgttcctatactaaatgttcttaggagtgctggggttttattgtcatcatttatcc
5401 tttttaaaaatgttattggccaggcacggtggctcatggctgtaatcccagcactttggg
5461 aggccgaggcaggcagatcacctgaggtcaggagtgtgagaccagcctggccaacatggc
5521 gaaacctgtctctactaaaaatacaaaaattaactaggcgtggtggtgtacgcctgtagt
5581 cccagctactcgggaggctgaggcaggagaatcaactgaaccagggaggtggaggttgca
5641 gtgtgccgagatcacgccactgcactctagcctggcaacagagcaagattctgtctcaaa
5701 aaaaaaaaacatatatacacatatatcccaaagtgctgggattacatatatatatatata
5761 tatatattatatatatatatatatatatatgtgatatatatgtgatatatatataacata
5821 tatatatgtaatatatatgtgatatatatataatatatatatgtaatatatatgtgatat
5881 atatatatacacacacacacatatatatgtatgtgtgtgtacacacacacacacaaat
5941 tagccaggcatagttgcacacgcttggtagacccagctactcaggaggctgagggaggag
6001 aatctcttgaacttaggaggcggaggttgcagtgagctgagattgcgccactgcactcca
6061 gcctgggtgacagagcaggactctgtacacccccaaaacaaaaaaaaaagttatcagat

FIG. 10B

GeneDX 1009, pg. 367



```
6121  gtgattggaatgtatatcaagtatcagcttcaaaatatgctatattaatacttcaaaaat
6181  tacacaaataatacataatcaggtttgaaaaatttaagacaacmsaaraaaaaawycmaa
6241  tcacamatatcccacacattttattattmctmctmcwattattttgwagagmctgggtct
6300  cacycykttgctwatgctggtctttgaacyccykgccycaarcartcctsctccabcctc
6361  ccaargtgctggggatwataggcatgarctaaccgcacccagccccagacattttagtgt
6421  gtaaattcctgggcatttttcaaggcatcatacatgttagctgactgatgatggtcaat
6481  ttattttgtccatggtgtcaagtttctcttcaggaggaaaagcacagaactggccaacaa
6541  ttgcttgactgttctttaccatactgtttagCAGGAAACCAGTCTCAGTGTCCAACTCTC
6601  TAACCTTGGAACTGTGAGAACTCTGAGGACAAAGCAGCGGATACAACCTCAAAAGACGTC
6661  TGTCTACATTGAATTGGGtaagggtctcaggtttttaagtatttaataataattgctgg
6721  attccttatcttatagttttgccaaaaatcttggtcataatttgtatttgtggtaggcag
6781  ctttgggaagtgaattttatgagccctatggtgagttataaaaaaatgtaaaagacgcagt
6841  tcccaccttgaagaatcttactttaaaaagggagcaaaagaggccaggcatggtggctca
6901  cacctgtaatcccagcactttgggaggccaaagtgggtggatcacctgaggtcgggagtt
6961  cgagaccagcctagccaacatggagaaactctgtctgtaccaaaaaataaaaaattagcc
7021  aggtgtggtggcacataactgtaatcccagctactcgggaggctgaggcaggagaatcac
7081  ttgaacccgggaggtggaggttgcggtgaaccgagatcgcaccattgcactccagcctgg
7141  gcaaaaatagcgaaactccatctaaaaaaaaaaaaagagagcaaaagaaagamtmtctggt
7201  tttaamtmtgtgtaaatatgtttttggaaagatggagagtagcaataagaaaaaacatga
7261  tggattgctacagtatttagttccaagataaattgtactagatgaggaagccttttaaga
7321  agagctgaattgccaggcgcagtggctcacgcctgtaatcccagcactttgggaggccga
7381  ggtgggcggatcacctgaggtcgggagttcaagaccagcctgaccaacatggagaaaccc
7441  catctctactaaaaaaaaaaaaaaaaaaattagccggggtggtggcttatgcctgtaatc
7501  ccagctactcaggaggctgaggcaggagaatcgcttgaacccaggaagcagaggttgcag
7561  tgagccaagatcgcaccattgcactccagcctaggcaacaagagtgaaactccatctcaa
7621  aaaaaaaaaaaaagagctgaatcttggctgggcaggatggctcgtgcctgtaatcctaac
7681  gctttggaagaccgaggcagaaggattggttgagtccacgagtttaagaccagcctggcc
7741  aacataggggaaccctgtctctatttttaaaataataatacatttttggccggtgcggtg
7801  gctcatgcctgtaatcccaatactttgggaggctgaggcaggtagatcacctgaggtcag
7861  agttcgagaccagcctggataacctggtgaaacccctctttactaaaaatacaaaaaaaa
7921  aaaaaaattagctgggtgtggtagcacatgcttgtaatcccagctacttgggaggctgag
7981  gcaggagaatcgcttgaaccagggaggcggaggttacaatgagccaacactacaccactg
8041  cactccagcctgggcaatagagtgagactgcatctcaaaaaaataataatttttaaaaat
8101  aataaatttttttaagcttataaaaagaaaagttgaggccagcatagtagctcacatctg
8161  taatctcagcagtggcagaggattgcttgaagccaggagtttgagaccagcctgggcaac
8221  atagcaagacctcatctctacaaaaaaatttctttttttaaattagctgggtgtggtggtg
8281  tgcatctgtagtcccagctactcaggaggcagaggtgagtggatacattgaacccaggag
8341  tttgaggctgtagtgagctatgatcatgccactgcactccaacctgggtgacagagcaag
8401  acctccaaaaaaaaaaaaaaaaaagagctgctgagctcagaattcaaactgggctctcaaat
8461  tggatttttcttttagaatatatttataattaaaaaggatagccatcttttgagctcccag
8521  gcaccaccatctatttatcataacacttactgttttccccccttatgatcataaattcct
8581  agacaacaggcattgtaaaaatagttatagtagttgatatttaggagcacttaactatat
8641  tccaggcactattgtgcttttcttgtataactcattagatgcttgtcagacctctgagat
8701  tgttcctattatacttatttacagatgagaaaattaaggcacagagaagttatgaaatt
8761  tttccaaggtattaaacctagtaagtggctgagccatgattcaaacctaggaagttagat
8821  gtcagagcctgtgctttttttttgttttgttttgttttcagtagaaacgggggtctca
8881  ctttgttggccaggctggtcttgaactcctaacctcaaataatccacccatctcggcctc
8941  ctcaagtgctgggattacaggtgagagccactgtgcctggcgaagcccatgcctttaacc
9001  acttctctgtattacatactagcttaactagcattgtacctgccacagtagatgctcagt
9061  aaatatttctagttgaatatctgtttttcaacaagtacatttttttaaccctttaatta
9121  agaaaacttttattgatttatttttgggggaaatttttagGATCTGATTCTTCTGAA
```

FIG. 10C

GeneDX 1009, pg. 368

```
 9181  GATACCGTTAATAAGGCAACTTATTGCAGgtgagtcaaagagaaccctttgtctatgaagc
 9241  tggtattttcctatttagttaatattaaggattgatgtttctctcttttaaaaatatttt
 9301  taacttttatttttaggttcagggatgtatgtgcagtttgttatataggtaaacacacgac
 9361  ttgggatttggtgtatagatttttttcatcatccgggtactaagcataccccacagtttt
 9421  ttgtttgctttctttctgaatttctccctcttcccaccttcctccctcaagtaggctggt
 9481  gtttctccagactagaatcatggtattggaagaaaccttagagatcatctagtttagttc
 9541  tctcattttatagtggaggaaataccctttttgtttgttggatttagttattagcactgt
 9601  ccaaaggaatttaggataacagtagaactctgcacatgcttgcttctagcagattgttct
 9661  ctaagttcctcatatacagtaatattgacacagcagtaattgtgactgatgaaaatgttc
 9721  aaggacttcattttcaactctttctttcctctgttccttatttccacatatctctcaagc
 9781  tttgtctgtatgttatataataaactacaagcaaccccaactatgttacctaccttcctt
 9841  aggaattattgcttgacccaggtttttttttttttttttggagacggggtcttgccct
 9901  gttgccaggatggagtgtagtggcgccatctcggctcactgcaatctccaactccctggt
 9961  tcaagcgattctcctgtctcaatctcacgagtagctgggactacaggtatacaccaccac
10021  gcccggttaattgaccattccatttctttctttctctctttttttttttttttttgaga
10081  cagagtcttgctctgttgcccaggctggagtacagaggtgtgatctcacctctccgcaac
10141  gtctgcctcccaggttgaagccatactcctgcctcagcctctctagtagctgggactaca
10201  ggcgcgcgccaccacacccggctaatttttgtatttttagtagagatggggtttcaccat
10261  gttggccaggctggtcttgaactcatgacctcaagtggtccacccgcctcagcctcccaa
10321  agtgctggaattacaggcttgagccaccgtgcccagcaaccatttcatttcaactagaag
10381  tttctaaaggagagagcagctttcactaactaaataagattggtcagctttctgtaatcg
10441  aaagagctaaaatgtttgatcttggtcatttgacagttctgcatacatgtaactagtgtt
10501  tcttattaggactctgtcttttccctatagTGTGGGAGATCAAGAATTGTTACAAATCAC
10561  CCCTCAAGGAACCAGGGATGAAATCAGTTTGGATTCTGCAAAAAAGGgtaatggcaaagt
10621  ttgccaacttaacaggcactgaaaagagagtgggtagatacagtactgtaattagattat
10681  tctgaagaccatttgggacctttacaacccacaaaatctcttggcagagttagagtatca
10741  ttctctgtcaaatgtcgtggtatggtctgatagatttaaatggtactagactaatgtacc
10801  tataataagaccttcttgtaactgattgttgccctttcgctttttttttttgtttgtttgt
10861  ttgtttttttttgagatggggtctcactctgttgcccaggctggagtgcagtgatgcaat
10921  cttggctcactgcaacctccacctccaaaggctcaagctatcctcccacttcagcctcct
10981  gagtagctgggactacaggcgcatgccaccacacccggttaatttttttgtggttttatag
11041  agatggggtttcaccatgttaccgaggctggtctcaaactcctggactcaagcagtctgc
11101  ccacttcagcctcccaaagtgctgcagttacaggcttgagccactgtgcctggcctgccc
11161  tttacttttaattggtgtatttgtgtttcatcttttacctactggtttttaaatataggg
11221  agtggtaagtctgtagatagaacagagtattaagtagacttaatggccagtaatctttag
11281  agtacatcagaaccagttttctgatggccaatctgcttttaattcactcttagacgttag
11341  agaaataggtgtggtttctgcatagggaaaattctgaaattaavvvvvvvvvvvvvgatc
11401  ctaagtggaaataatctaggtaaataggaattaaatgaaagagtatgagctacatcttca
11461  gtatacttggtagtttatgaggttagtttctctaatatagccagttggttgatttccacc
11521  tccaaggtgtatgaagtatgtatttttttaatgacaattcagtttttgagtaccttgtta
11581  tttttgtatattttcagCTGCTTGTGAATTTTCTGAGACGGATGTAACAAATACTGAACA
11641  TCATCAACCCAGTAATAATGATTTGAACACCACTGAGAAGCGTGCAGCTGAGAGGCATCC
11701  AGAAAAGTATCAGGGTAGTTCTGTTTCAAACTTGCATGTGGAGCCATGTGGCACAAATAC
11761  TCATGCCAGCTCATTACAGCATGAGAACAGCAGTTTATTACTCACTAAAGACAGAATGAA
11821  TGTAGAAAAGGCTGAATTCTGTAATAAAAGCAAACAGCCTGGCTTAGCAAGGAGCCAACA
11881  TAACAGATGGGCTGGAAGTAAGGAAACATGTAATGATAGGCGGACTCCCAGCACAGAAAA
11941  AAAGGTAGATCTGAATGCTGATCCCCTGTGTGAGAGAAAAGAATGGAATAAGCAGAAACT
12001  GCCATGCTCAGAGAATCCTAGAGATACTGAAGATGTTCCTTGGATAACACTAAATAGCAG
12061  CATTCAGAAAGTTAATGAGTGGTTTTCCAGAAGTGATGAACTGTTAGGTTCTGATGACTC
12121  ACATGATGGGGAGTCTGAATCAAATGCCAAAGTAGCTGATGTATTGGACGTTCTAAATGA
12181  GGTAGATGAATATTCTGGTTCTTCAGAGAAAATAGACTTACTGGCCAGTGATCCTCATGA
```

FIG. 10D

GeneDX 1009, pg. 369

```
12241  GGCTTTAATATGTAAAAGTGAAAGAGTTCACTCCAAATCAGTAGAGAGTAATATTGAAGG
12301  CCAAATATTTGGGAAAACCTATCGGAAGAAGGCAAGCCTCCCCAACTTAAGCCATGTAAC
12361  TGAAAATCTAATTATAGGAGCATTTGTTACTGAGCCACAGATAATACAAGAGCGTCCCCT
12421  CACAAATAAATTAAAGCGTAAAAGGAGACCTACATCAGGCCTTCATCCTGAGGATTTTAT
12481  CAAGAAAGCAGATTTGGCAGTTCAAAAGACTCCTGAAATGATAAATCAGGGAACTAACCA
12541  AACGGAGCAGAATGGTCAAGTGATGAATATTACTAATAGTGGTCATGAGAATAAAACAAA
12601  AGGTGATTCTATTCAGAATGAGAAAAATCCTAACCCAATAGAATCACTCGAAAAAGAATC
12661  TGCTTTCAAAACGAAAGCTGAACCTATAAGCAGCAGTATAAGCAATATGGAACTCGAATT
12721  AAATATCCACAATTCAAAAGCACCTAAAAAGAATAGGCTGAGGAGGAAGTCTTCTACCAG
12781  GCATATTCATGCGCTTGAACTAGTAGTCAGTAGAAATCTAAGCCCACCTAATTGTACTGA
12841  ATTGCAAATTGATAGTTGTTCTAGCAGTGAAGAGATAAAGAAAAAAAAGTACAACCAAAT
12901  GCCAGTCAGGCACAGCAGAAACCTACAACTCATGGAAGGTAAAGAACCTGCAACTGGAGC
12961  CAAGAAGAGTAACAAGCCAAATGAACAGACAAGTAAAAGACATGACAGCGATACTTTCCC
13021  AGAGCTGAAGTTAACAAATGCACCTGGTTCTTTTACTAAGTGTTCAAATACCAGTGAACT
13081  TAAAGAATTTGTCAATCCTAGCCTTCCAAGAGAAGAAAAAGAAGAGAAACTAGAAACAGT
13141  TAAAGTGTCTAATAATGCTGAAGACCCCAAAGATCTCATGTTAAGTGGAGAAAGGGTTTT
13201  GCAAACTGAAAGATCTGTAGAGAGTAGCAGTATTTCATTGGTACCTGGTACTGATTATGG
13261  CAGTCAGGAAAGTATCTCGTTACTGGAAGTTAGCACTCTAGGGAAGGCAAAAACAGAACC
13321  AAATAAATGTGTGAGTCAGTGTGCAGCATTTGAAAACCCCAAGGGACTAATTCATGGTTG
13381  TTCCAAAGATAATAGAAATGACACAGAAGGCTTTAAGTATCCATTGGGACATGAAGTTAA
13441  CCACAGTCGGGAAACAAGCATAGAAATGGAAGAAAGTGAACTTGATGCTCAGTATTTGCA
13501  GAATACATTCAAGGTTTCAAAGCGCCAGTCATTTGCTCCGTTTTCAAATCCAGGAAATGC
13561  AGAAGAGGAATGTGCAACATTCTCTGCCCACTCTGGGTCCTTAAAGAAACAAAGTCCAAA
13621  AGTCACTTTTGAATGTGAACAAAAGGAAGAAAATCAAGGAAAGAATGAGTCTAATATCAA
13681  GCCTGTACAGACAGTTAATATCACTGCAGGCTTTCCTGTGGTTGGTCAGAAAGATAAGCC
13741  AGTTGATAATGCCAAATGTAGTATCAAAGGAGGCTCTAGGTTTTGTCTATCATCTCAGTT
13801  CAGAGGCAACGAAACTGGACTCATTACTCCAAATAAACATGGACTTTTACAAAACCCATA
13861  TCGTATACCACCACTTTTTCCCATCAAGTCATTTGTTAAAACTAAATGTAAGAAAAATCT
13921  GCTAGAGGAAAACTTTGAGGAACATTCAATGTCACCTGAAAGAGAAATGGGAAATGAGAA
13981  CATTCCAAGTACAGTGAGCACAATTAGCCGTAATAACATTAGAGAAAATGTTTTTAAAGA
14041  AGCCAGCTCAAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGGCTCCAGTAT
14101  TAATGAAATAGGTTCCAGTGATGAAAACATTCAAGCAGAACTAGGTAGAAACAGAGGGCC
14161  AAAATTGAATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGTCTATAAACAAAGTCT
14221  TCCTGGAAGTAATTGTAAGCATCCTGAAATAAAAAAGCAAGAATATGAAGAAGTAGTTCA
14281  GACTGTTAATACAGATTTCTCTCCATATCTGATTTCAGATAACTTAGAACAGCCTATGGG
14341  AAGTAGTCATGCATCTCAGGTTTGTTCTGAGACACCTGATGACCTGTTAGATGATGGTGA
14401  AATAAAGGAAGATACTAGTTTTGCTGAAAATGACATTAAGGAAAGTTCTGCTGTTTTTAG
14461  CAAAAGCGTCCAGAAAGGAGAGCTTAGCAGGAGTCCTAGCCCTTTCACCCATACACATTT
14521  GGCTCAGGGTTACCGAAGAGGGGCCAAGAAATTAGAGTCCTCAGAAGAGAACTTATCTAG
14581  TGAGGATGAAGAGCTTCCCTGCTTCCAACACTTGTTATTTGGTAAAGTAAACAATATACC
14641  TTCTCAGTCTACTAGGCATAGCACCGTTGCTACCGAGTGTCTGTCTAAGAACACAGAGGA
14701  GAATTTATTATCATTGAAGAATAGCTTAAATGACTGCAGTAACCAGGTAATATTGGCAAA
14761  GGCATCTCAGGAACATCACCTTAGTGAGGAAACAAAATGTTCTGCTAGCTTGTTTTCTTC
14821  ACAGTGCAGTGAATTGGAAGACTTGACTGCAAATACAAACACCCAGGATCCTTTCTTGAT
14881  TGGTTCTTCCAAACAAATGAGGCATCAGTCTGAAAGCCAGGGAGTTGGTCTGAGTGACAA
14941  GGAATTGGTTTCAGATGATGAAGAAAGAGGAACGGGCTTGGAAGAAAATAATCAAGAAGA
15001  GCAAAGCATGGATTCAAACTTAGgtattggaaccaggtttttgtgtttgccccagtctat
15061  ttatagaagtgagctaaatgtttatgcttttggggagcacattttacaaatttccaagta
15121  tagttaaaggaactgcttcttaaacttgaaacatgttcctcctaaggtgcttttcataga
15181  aaaaagtccttcacacagctaggacgtcatctttgactgaatgagctttaacatcctaat
15241  tactggtggacttacttctggtttcattttataaagcaaatccCggtgtcccaaagcaag
```

FIG. 10E

GeneDX 1009, pg. 370

15301 gaatttaatcattttgtgtgacatgaaagtaaatccagtcctgccaatgagaagaaaaag
15361 acacagcaagttgcagcgtttatagtctgcttttacatctgaacctctgtttttgttatt
15421 taagGTGAAGCAGCATCTGGGTGTGAGAGTGAAACAAGCGTCTCTGAAGACTGCTCAGGG
15481 CTATCCTCTCAGAGTGACATTTTAACCACTCaggtaaaaagcgtgtgtgtgtgtgcacat
15541 gcgtgtgtgtggtgtcctttgcattcagtagtatgtatcccacattcttaggtttgctga
15601 catcatctctttgaattaatggcacaattgtttgtggttcattgtcvvvvvvvvvvvvvn
15661 gngaatgtaatcctaatatttcncnccnacttaaaagaataccactccaanggcatcnca
15721 atacatcaatcaattggggaattgggattttccctcnctaacatcantggaataatttca
15781 tggcattaattgcatgaatgtggttagattaaaaggtgttcatgctagaacttgtagttc
15841 catactaggtgatttcaattcctgtgctaaaattaatttgtatgatatattntcatttaa
15901 tggaaagcttctcaaagtatttcattttcttggtaccatttatcgtttttgaAGCAGAGG
15961 GATACCATGCAACATAACCTGATAAAGCTCCAGCAGGAAATGGCTGAACTAGAAGCTGTG
16021 TTAGAACAGCATGGGAGCCAGCCTTCTAACAGCTACCCTTCCATCATAAGTGACTCTTCT
16081 GCCCTTGAGGACCTGCGAAATCCAGAACAAAGCACATCAGAAAAAGgtgtgtattgttgg
16141 ccaaacactgatatcttaagcaaaattctttccttcccctttatctccttctgaagagta
16201 aggacctagctccaacattttatgatccttgctcagcacatgggtaattatggagccttg
16261 gttcttgtccctgctcacaactaatataccagtcagagggacccaaggcagtcattcatg
16321 ttgtcatctgagatacctacaacaagtagatgctatggggagcccatggvvvvvvvvvvv
16381 vvccattggtgctagcatctgtctgttgcattgcttgtgtttataaaattctgcctgata
16441 tacttgttaaaaaccaatttgtgtatcatagattgatgcttttgaaaaaaatcagtattc
16501 taacctgaattatcactatcagaacaaagcagtaaagtagatttgttttctcattccatt
16561 taaagCAGTATTAACTTCACAGAAAAGTAGTGAATACCCTATAAGCCAGAATCCAGAAGG
16621 CCTTTCTGCTGACAAGTTTGAGGTGTCTGCAGATAGTTCTACCAGTAAAAATAAAGAACC
16681 AGGAGTGGAAAGgtaagaaacatcaatgtaaagatgctgtggtatctgacatctttattt
16741 atattgaactctgattgttaatttttttcaccatactttctccagttttttgcatacag
16801 gcatttatacacttttattgctctaggatacttcttttgtttaatcctatataggvvvvv
16861 vvvvvvvvggataagntcaagagatattttgataggtgatgcagtgatnaattgngaaaa
16921 tttnctgcctgcttttaatcttccccgttctttcttcctncctccctccttcctncct
16981 cccgtccttnccttttcctttcccctccttccncctttctttccntctntctttccttttctt
17041 tcctgtctaccttttctttccttcctccctttcctttttctttttctttctttcctttcctttt
17101 ctttcctttctttcctttcctttctttcttgacagagtcttgctctgtcactcaggctgg
17161 agtgcagtggcgtgatctcgnctcactgcaacctctgtctcccaggttcaagcaattttc
17221 ctgcctcagcctcccgagtagctgagattacaggcgccagccaccacacccagctactga
17281 cctgcttttvvvvvvvvvvvvvvaaacagctgggagatatggtgcctcagaccaacccat
17341 gttatatgtcaaccctgacatattggcaggcaacatgaatccagacttctaggctgtcat
17401 gcgggctctttttgccagtcatttctgatctctctgacatgagctgtttcatttatgct
17461 ttggctgcccagcaagtatgatttgtcctttcacaattggtggcgatggttttctccttc
17521 catttatctttctagGTCATCCCCTTCTAAATGCCCATCATTAGATGATAGGTGGTACAT
17581 GCACAGTTGCTCTGGGAGTCTTCAGAATAGAAACTACCCATCTCAAGAGGAGCTCATTAA
17641 GGTTGTTGATGTGGAGGAGCAACAGCTGGAAGAGTCTGGGCCACACGATTTGACGGAAAC
17701 ATCTTACTTGCCAAGGCAAGATCTAGgtaatatttcatctgctgtattggaacaaacact
17761 ytgattttactctgaatcctacataaagatattctggttaaccaacttttagatgtacta
17821 gtctatcatggacacttttgttatacttaattaagcccactttagaaaaatagctcaagt
17881 gttaatcaaggtttacttgaaaattattgaaactgttaatccatctatattttaattaat
17941 ggtttaactaatgattttgaggatgwgggagtcktggtgtactctamatgtattatttca
18001 ggccaggcatagtggctcacgcctggtaatcccagtayycmrgagcccgaggcaggtgga
18061 gccagctgaggtcaggagttcaagacctgtcttggccaacatgggngaaaccctgtcttc
18121 ttcttaaaaaanacaaaaaaaattaactgggttgtgcttaggtgnatgccccgnatccta
18181 gttnttcttgnggttgagggaggagatcacnttggaccccggaggggnggtgggggng
18241 agcaggncaaaacacngacccagctggggtggaagggaagcccactcnaaaaaannttnv
18301 vvvvvvvvvvvvttttaggaaacaagctactttggatttccaccaacacctgtattcat

FIG. 10F

ATTACH TO APPL. NO. BY PLACE SUBCLASS DRAFTSMAN

```
18361  gtacccatttttctcttaacctaactttattggtctttttaattcttaacagagaccaga
18421  actttgtaattcaacattcatcgttgtgtaaattaaacttctcccattcctttcagAGGG
18481  AACCCCTTACCTGGAATCTGGAATCAGCCTCTTCTCTGATGACCCTGAATCTGATCCTTC
18541  TGAAGACAGAGCCCCAGAGTCAGCTCGTGTTGGCAACATACCATCTTCAACCTCTGCATT
18601  GAAAGTTCCCCAATTGAAAGTTGCAGAATCTGCCCAGAGTCCAGCTGCTGCTCATACTAC
18661  TGATACTGCTGGGTATAATGCAATGGAAGAAAGTGTGAGCAGGGAGAAGCCAGAATTGAC
18721  AGCTTCAACAGAAAGGGTCAACAAAAGAATGTCCATGGTGGTGTCTGGCCTGACCCCAGA
18781  AGAATTTgtgagtgtatccatatgtatctccctaatgactaagacttaacaacattctgg
18841  aaagagttttatgtaggtattgtcaattaataacctagaggaagaaatctagaaaacaat
18901  cacagttctgtgtaatttaatttcgattactaatttctgaaaatttagaayvvvvvvvvv
18961  vvvvncccnncccccnaatctgaaatgggggtaacccccccccaaccganacntgggtng
19021  cntagaganttttaatggcccnttctgaggnacanaagcttaagccaggngacgtggancn
19081  atgngttgtttnttgtttggttacctccagcctgggtgacagagcaagactctgtctaaa
19141  aaaaaaaaaaaaaaaaatcgactttaaatagttccaggacacgtgtagaacgtgcaggat
19201  tgctacgtaggtaaacatatgccatggtgggataactagtattctgagctgtgtgctaga
19261  ggtaactcatgataatggaatatttgatttaatttcagATGCTCGTGTACAAGTTTGCCA
19321  GAAAACACCACATCACTTTAACTAATCTAATTACTGAAGAGACTACTCATGTTGTTATGA
19381  AAACAGgtataccaagaacctttacagaataccttgcatctgctgcataaaaccacatga
19441  ggcgaggcacggtggcgcatgcctgtaatcgcagcactttgggaggccgaggcgggcaga
19501  tcacgagattaggagatcgagaccatcctggccagcatggtgaaaccccgtctctactan
19561  naaatggnaaaattanctgggtgtggtcgcgtgcncctgtagtcccagctactcgtgagg
19621  ctgaggcaggagaatcacttgaaccggggaaatggaggtttcagtgagcagagatcatnc
19681  ccctncattccagcctggcgacagagcaaggctccgtcnccnaaaaaatanaaaaaaaacg
19741  tgaacaaataagaatatttgttgagcatagcatggatgatagtcttctaatagtcaatca
19801  attactttatgaaagacaaataatagttttgctgcttccttacctccttttgttttgggt
19861  taagatttggagtgtgggccaggcacvvvvvvvvvvvvvgatctatagctagccttggcg
19921  tctagaagatgggtgttgagaagagggagtggaaagatatttcctctggtcttaacttca
19981  tatcagcctcccctagacttccaaatatccatacctgctggttataattagtggtgtttt
20041  cagcctctgattctgtcaccaggggttttagaatcataaatccagattgatcttgggagt
20101  gtaaaaaactgaggctctttagcttcttaggacagcagttcctgattttgttttcaactt
20161  ctaatcctttgagtgtttttcattctgcagATGCTGAGTTTGTGTGTGAACGGACACTGA
20221  AATATTTTCTAGGAATTGCGGGAGGAAAATGGGTAGTTAGCTATTTCTGtaagtataata
20281  ctatttctcccctcctccctttaacacctcagaattgcatttttacacctaacatttaac
20341  acctaaggtttttgctgatgctgagtctgagttaccaaaaggtctttaaattgtaatact
20401  aaactacttttatctttaatatcactttgttcaagataagctggtgatgctgggaaaatg
20461  ggtctcttttataactaataggacctaatctgctcctagcaatgttagcatatgagctag
20521  ggatttatttaatagtcggcaggaatccatgtgcarcagncaaacttataatgtttaaat
20581  taaacatcaactctgtctccagaaggaaactgctgctacaagccttattaaagggctgtg
20641  gctttagagggaaggacctctcctctgtcattcttcctgtgctcttttgtgaatcgctga
20701  cctctctatctccgtgaaaagagcacgttcttctgctgtatgtaacctgtctttttctatg
20761  atctctvvvvvvvvvvvvvvnaaaaacggggnnggganttggccttaaanccaaagggcna
20821  actccccaaccattnaaaaantgacnggggattattaaaancggcgggaaacatttcacn
20881  gcccaactaatattgttaaattaaaaccaccaccnctgcnccaaggagggaaactgctgc
20941  tacaagccttattaaagggctgtggctttagagggaaggacctctcctctgtcattcttc
21001  ctgtgctcttttgtgaatcgctgacctctctatgtccgtgaaaagagcacgttcttcgtc
21061  tgtatgtaacctgtcttttctatgatctctttagGGGTGACCCAGTCTATTAAAGAAAGA
21121  AAAATGCTGAATGAGgtaagtacttgatgttacaaactaaccagagatattcattcagtc
21181  atatagttaaaaatgtatttgcttccttccatcaatgcaccactttccttaacaatgcac
21241  aaattttccatgataatgaggatcatcaagaattatgcaggcctgcactgtggctcatac
21301  ctataatcccagcgctttgggaggctgaggcgcttggatcvvvvvvvvvvvvvaattttt
21361  tgtatttttagtagagatgaggttcaccatgttggtctagatctggtgtcgaacgtcctg
```

FIG. 10G

GeneDX 1009, pg. 372



```
21421  acctcaagtgatctgccagcctcagtctcccaaagtgctaggattacaggggtgagccac
21481  tgcgcctggcctgaatgcctaaaatatgacgtgtctgctccacttccattgaaggaagct
21541  tctctttctcttatcctgatgggttgtgtttggtttctttcagCATGATTTTGAAGTCAG
21601  AGGAGATGTGGTCAATGGAAGAAACCACCAAGGTCCAAAGCGAGCAAGAGAATCCCAGGA
21661  CAGAAAGgtaaagctccctccctcaagttgacaaaaatctcaccccaccactctgtattc
21721  cactcccctttgcagagatgggccgcttcattttgtaagacttattacatacatacacag
21781  tgctagatactttcacacaggttctttttcactcttccatcccaaccacataaataagt
21841  attgtctctactttatgaatgataaaactaagagatttagagaggctgtgtaatttggat
21901  tcccgtctcgggttcagatcvvvvvvvvvvvvvttggcctgattggtgacaaaagtgaga
21961  tgctcagtccttgaatgacaaagaatgcctgtagagttgcaggtccaactacatatgcac
22021  ttcaagaagatcttctgaaatctagtagtgttctggacattggactgcttgtccctggga
22081  agtagcagcagaaatgatcggtggtgaacagaagaaaaagaaaagctcttcctttttgaa
22141  agtctgtttttgaataaaagccaatattcttttataactagattttccttctctccatt
22201  cccctgtccctctctcttcctctcttcttccagATCTTCAGGGGGCTAGAAATCTGTTGC
22261  TATGGGCCCTTCACCAACATGCCCACAGgtaagagcctgggagaaccccagagttccagc
22321  accagcctttgtcttacatagtggagtattataagcaaggtcccacgatgggggttcctc
22381  agattgctgaaatgttctagaggctattctatttctctaccactctccaaacaaaacagc
22441  acctaaatgttatcctatggcaaaaaaaaactataccttgtccccttctcaagagcatg
22501  aaggtggttaatagttaggattcagtatgttatgtgttcagatggcgttgagctgctgtt
22561  agtgccvvvvvvvvvvvvvvtttgagagactatcaaaccttataccaagtggccttatgga
22621  gactgataaccagagtacatggcatatcagtggcaaattgacttaaaatccataccccta
22681  ctattttaagaccattgtcctttggagcagagagacagactctcccattgagaggtcttg
22741  ctataagccttcatccggagagtgtagggtagagggcctgggttaagtatgcagattact
22801  gcagtgattttacatgtaaatgtccattttagATCAACTGGAATGGATGGTACAGCTGTG
22861  TGGTGCTTCTGTGGTGAAGGAGCTTTCATCATTCACCCTTGGCACAgtaagtattgggtg
22921  ccctgtcagtgtgggaggacacaatattctctcctgtgagcaagactggcacctgtcagt
22981  ccctatggatgcccctactgtagcctcagaagtcttctctgcccacatacctgtgccaaa
23041  agactccatvvvvvvvvvvvvvvggtggtacgtgtctgtagttccagctacttgggaggct
23101  gagatggaaggattgcttgagcccaggaggcagaggtggnannttacgctgagatcacac
23161  cactgcactccagcctgggtgacagagcaagaccctgtctcaaaaacaaacaaaaaaaat
23221  gatgaagtgacagttccagtagtcctactttgacactttgaatgctctttccttcctggg
23281  gatccagGGTGTCCACCCAATTGTGGTTGTGCAGCCAGATGCCTGGACAGAGGACAATGG
23341  CTTCCATGgtaaggtgcctcgcatgtacctgtgctattagtggggtccttgtgcatgggt
23401  ttggtttatcactcattacctggtgcttgagtagcacagttcttggcacatttttaaata
23461  tttgttgaatgaatggctaaaatgtctttttgatgttttattgttatttgttttatatt
23521  gtaaaagtaatacatgaactgtttccatggggtgggagtaagatatgaatgttcatcacv
23581  vvvvvvvvvvvvcagtaatcctnagaactcatacgaccgggcccctggagtcgntgnttn
23641  gagcctagtccnggagaatgaattgacactaatctctgcttgtgttctctgtctccagCA
23701  ATTGGGCAGATGTGTGAGGCACCTGTGGTGACCCGAGAGTGGGTGTTGGACAGTGTAGCA
23761  CTCTACCAGTGCCAGGAGCTGGACACCTACCTGATACCCCAGATCCCCCACAGCCACTAC
23821  TGACTGCAGCCAGCCACAGGTACAGAGCCACAGGACCCCAAGAATGAGCTTACAAAGTGG
23881  CCTTTCCAGGCCCTGGGAGCTCCTCTCACTCTTCAGTCCTTCTACTGTCCTGGCTACTAA
23941  ATATTTTATGTACATCAGCCTGAAAAGGACTTCTGGCTATGCAAGGGTCCCTTAAAGATT
24001  TTCTGCTTGAAGTCTCCCTTGGAAAT
```

FIG. 10H

SUE FEE TRANSMITTAL

MAILING INSTRUCTIONS: Th... ...ISSUE FEE. Blocks 2 through 6 should ... completed where appropriate. All futher correspondence including the Issue Fee Receipt, the ... ...cation of maintenance fees will be mailed to addresses entered in Block 1 unless you direct otherwise, by: (a) specifying a new correspondence address in Block ..., or (b) providing the PTO with a separate "FEE ADDRESS" for maintenance fee notifications with the payment of issue Fee or thereafter. **See reverse for Certificate of Mailing, below.**

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending on the needs of the individual case. Any comments on the amount of time required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, D.C. 20231.
DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Issue Fee, Assistant Commissioner for Patents, Washington D.C. 20231

1. CORRESPONDENCE ADDRESS

18M2/0522
VENABLE BAETJER HOWARD AND CIVILETTI
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

2. INVENTOR(S) ADDRESS CHANGE (Complete only if there is a change)

INVENTOR'S NAME
Street Address
City, State and Zip Code
CO-INVENTOR'S NAME
Street Address
City, State and Zip Code
☐ Check if additional changes are enclosed

RECEIVED
Publishing Division
JUL 14 1997
04

APPLICATION NO.	FILING DATE	TOTAL CLAIMS	EXAMINER AND GROUP ART UNIT		DATE MAILED
08/488,011	01/05/96	037	REES, D	1807	05/22/97

First Named Applicant: SKOLNICK, MARK H.

TITLE OF INVENTION: 17Q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

ERROR CORRECTED

ATTY'S DOCKET NO.	CLASS-SUBCLASS	BATCH NO.	APPLN. TYPE	SMALL ENTITY	FEE DUE	DATE DUE
1 24884-109347	435-006.000	N28	UTILITY	NO	$1290.00	08/22/97

3. Correspondence address change (Complete only if there is a change)

06/27/1997 LBERGER 00000173 DAN:220261 08488011
01 FC:142 1290.00 CH
02 FC:561 30.00 CH
~~02 FC:122 180.00 CH~~

4. For printing on the patent front page, list the names of not more than 3 registered patent attorneys or agents OR, alternatively, the name of a firm having as a member a registered attorney or agent. If no name is listed, no name will be printed.

1 VENABLE, BAETJER, HOWARD & CIVILETTI, LLP
2
3

5. ASSIGNMENT DATA TO BE PRINTED ON THE PATENT (print or type)

(1) NAME OF ASSIGNEE: Myriad Genetics, Inc.; University of Utah Research Foundation; and United States of America, as * represented by the Secretary of Health and Human Services,

(2) ADDRESS: (CITY & STATE OR COUNTRY) Salt Lake City, Utah; Salt Lake City, Utah; and Rockville, Maryland

A. ☐ This application is NOT assigned.
☒ Assignment previously submitted to the Patent and Trademark Office.
☐ Assignment is being submitted under separate cover. Assignment should be directed to Box ASSIGNMENTS.
PLEASE NOTE: Unless an assignee is identified in Block 5, no assignee data will appear on the patent. Inclusion of assignee data is only appropriate when an assignment has been perviously submitted to the PTO or is being submitted under separate cover. Completion of this form is NOT a substitute for filing an assignment.

6a. The following fees are enclosed:
☐ Issue Fee ☐ Advance Order - # of Copies
6b. The following fees should be charged to:
DEPOSIT ACCOUNT NUMBER 22-0261
(ENCLOSE A COPY OF THIS FORM)
☒ Issue Fee ☒ Advance Order - # of Copies ten (10)
☐ Any Deficiencies in Enclosed Fees

The COMMISSIONER OF PATENTS AND TRADEMARKS is requested to apply the Issue Fee to the application identified above.

(Authorized Signature)	(Date)
Jeffrey L. Ihnen, Reg. 28,957	7-12-97

NOTE: The Issue Fee will not be accepted from anyone other than the applicant; a registered attorney or agent; or the assignee or other party in interest as shown by the records of the Patent and Trademark Office.

Certificate of Mailing

Note: If this certificate of mailing is used, it can be used to transmit the Issue Fee. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing.

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficent postage as first class mail in an envelope addressed to: **Box ISSUE FEE**
Assistant Commissioner for Patents
Washington, D.C. 20231

on: _____ **(Date)**
_____ **(Name of person making deposit)**
_____ **(Signature)**
_____ **(Date)**

GeneDX 1009, pg. 374

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. patent application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011 Group Art Unit: 1807

Filed: 07 June 1995 **BATCH NO. Z 99**

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

RECEIVED
Publishing Division
JUL 14 1997
04

PAYMENT OF ISSUE FEE AND TRANSMITTAL OF FORMAL DRAWINGS

Hon. Commissioner of Patents and Trademarks
Washington, D.C. 20231

Sir:

In accordance with the Notice of Allowance/Notice of Allowability mailed for the above application on 22 May 1997, submitted herewith are the Issue Fee Transmittal and twenty-six (26) sheets (10 figures) of formal drawing (the photographic representations of Figures 7, 9A. 9B and 9C are submitted in triplicate, for a total of 26 pages of drawing). Please substitute these drawings for those originally filed with the application.

If any fees are due (other than the issue fee/advance copies charged to our Deposit Account and indicated on the Issue Fee Transmittal), the same should be charged to Deposit Account No. 22-0261. This transmittal letter is submitted in triplicate.

Respectfully submitted,

Jeffrey L. Ihnen, Reg. No. 28,957

08488011 122 130.00

VENABLE, BAETJER, HOWARD & CIVILETTI, *LLP*
1201 New York Avenue, N.W., #1000
Washington, D.C. 20005
(202) 962-4810

Atty. Docket No. 24884-109347-09
Dated: 12 July 1997
7209.01

AO 120 (Rev. 2/97)

SOLICITOR

TO: Commissioner of Patents and Trademarks
Arlington, VA 22202

NOV 27 1998

U.S. PATENT & TRADEMARK OFFICE

REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK

In Compliance with 35 § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court m Distrid of Utah on the following XX Patents or ☐ Trademarks:

DOCKET NO.	DATE FILED	U.S. DISTRICT COURT
2:98cv829 S	11/19/98	District of Utah

PLAINTIFF	DEFENDANT
Myrian Genetics, Inc.	University of Pennsylvania

	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1	5,693,473		Myriad Genetics
2	5,707,999		"
3	5,747,282		"
4	5,753,441		"
5			

In the above-entitled case, the following patent(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading		
	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1			
2			
3			
4			
5			

In the above-entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Commissioner Cpy 3—Upon termination of action, mail this copy to Commissioner
Copy 2—Upon filing document adding patent(s), mail this copy to Commissioner Copy 4—Case file copy

210 MAR 19 1998
PAT & TRADEMARK OFF.

2318-123A
JLI:anb

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of	)	
	)	
Mark H. SKOLNICK et al.	)	
	)	
Serial Number. 08/488,011	)	Batch No. Z99
	)	
Filed: June 7, 1995	)	Group Art Unit: 1807
	)	
For: 17q LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE	)	

STATUS INQUIRY

Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Dear Sir:

Applicant respectfully inquires as to the status of the Issue Notification in the above-identified application. Applicant submitted its formal drawings and issue fee on July 14, 1997. No Issue Notification has been received.

Respectfully Submitted,

By _____
Jeffrey L. Ihnen
Registration No. 28,957
Attorney for Applicants
ROTHWELL, FIGG, ERNST & KURZ, p.c.
Suite 701-E, 555 13th Street, N.W.
Washington, D.C. 20004
Date March 19, 1998 Telephone: (202) 783-6040

2318-123.sta

5,753,537
5/19/98

PATENT
4998/PD-2704

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:
TADAO AKAMINE ET AL.
Serial No: 08/449,655
Filed: May 24, 1995
For: IMPURITY DOPING METHOD WITH DIFFUSION SOURCE OF BORON-SILICIDE FILM

Issue Batch No. E05
Date of NOA: 09/11/96
Group Art Unit: 1107

RECEIVED
APR 1 X 1998
Publishing Division
Corres/Allowed Files (r

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to:
Assistant Commissioner for Patents
Washington D.C. 20231, on
March 25, 1998
Date of Deposit
William H. Wright
Name
Signature
3/25/98
Date

REQUEST FOR STATUS

Box ISSUE FEE
Assistant Commissioner for Patents
Washington, D.C. 20231

Dear Sir:

This application was filed May 24, 1995 and the filing receipt was received on June 23, 1995. The Issue Fee, Formal Drawings and Priority Documents were filed December 5, 1996. A copy of the postcard stamped "RECEIVED Publishing Division DEC 09 1996 GP" is enclosed. To date, no further papers have been received in this case.

Please send the undersigned the original Letters Patent and ten soft copies paid for or a Status Report on this application.

Respectfully submitted,
LOEB & LOEB LLP

RECEIVED
PUBLISHING DIVISION
MAR 30 1998
00

Date: March 25, 1998

By: _____
William H. Wright
Registration No. 36,312
Attorney for Applicant(s)

10100 Santa Monica Blvd., Suite 2200
Los Angeles, CA 90067-4164
Telephone: (310) 282-2000
Facsimile: (310) 282-2192

H:\DF\004998\PD2704\STATREQ.WPD

S.N. 08/449,655 File No. PD-2704 Date Mailed 12-5-96 By: Margie

Title: Impurity Doping Method with Diffusion Source...

(Client Name) Seiko Inst.

The Following, due 12-11-96 in the U.S. Patent & Trademark Office was received in the Patent & Trademark Office on the date stamped hereon:

___ Amendment
___ TM Appln., Including Specimens
___ PCT Application Including
___ Spec. ___ Request
___ Application for Patent Including
___ Pages Spec.; No. of Claims
___ Declaration, Affidavit or Oath
___ Assign; Ck. No. _____ for $ _____
___ Verified Statement
✓ Letter of Transmittal
___ Maintenance Fee Transmittal
✓ Check No. 001412 for $ 1290.-
___ Check No. _____ for $ _____
Priority: _____
___ _____

RECEIVED
Publishing Division
DEC 09 1996

GP

✓ Drawings; # of Sheets ___
✓ Formal ___ Informal
✓ Issue Fee Transmittal
___ Letter Re _____
___ Notice of Appeal
___ Petition for _____
___ Advance soft copy order:
✓ Ck. No. 001411 for $ 30.-
✓ Certified Copy; # 4* of Docs.
___ Power of Attorney
___ IDS
* Japanese Patent Appln. Nos.
4-129537 filed 4-21-92;
5-060865 filed 2-25-93;
5-045446 filed 3-5-93 and
5-060521 filed 3-19-93.

GeneDX 1009, pg. 379

UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office

Address : COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

SERIAL NUMBER	FILING DATE	FIRST NAMED APPLICANT	ATTORNEY DOCKET NO.
08/485,011	01/05/96	SKOLNICK M	24004-1097

4102/0128

VENABLE BAETJER HOWARD AND CIVILETTI
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

EXAMINER	
REES,D	
ART UNIT	PAPER NUMBER
1807	

DATE MAILED: 01/28/98

The decision on the petition filed in the above entitled application is as follows:

☐ Delay in Prosecution Held Unavoidable (35 U.S.C. 133);

Petition Granted _____

☐ Delayed Payment of Issue Fee Accepted (35 U.S.C. 151),

Petition Granted _____

☑ Petition Granted Accepted Photographs

☐ Petition Denied _____

☐ Petition Dismissed _____

By direction of the Deputy
Assistant Commissioner for Patents

_____

HELEN CHILDS
CHIEF, SUPERVISORY LEGAL
INSTRUMENTS EXAMINER
GROUP 1800

GeneDX 1009, pg. 380

MAIL ROOM
DEC
1
1995

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011

Filed: 07 June 1995

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

STATEMENT PURSUANT TO 37 CFR §1.821(f)

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

In the matter of the above-identified application, presented herewith is a computer-readable copy of the Sequence Listing pursuant to 37 CFR §1.821(f). It is hereby certified that these copies are identical.

Respectfully submitted,

Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI
1201 New York Avenue, N.W., #1000
Washington, D.C. 20005
(202) 962-4810

Atty. Docket No. 24884-109347-B
Dated: 1 December 1995

70121 U.S. PTO
04/30/97
Initial Review
BOX AF

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re: Patent Application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011 Group Art Unit: 1807

Filed: 07 June 1995 Examiner: D. Rees

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

RECEIVED
MAY 02 1997
GROUP 1800

STATEMENT PURSUANT TO 37 CFR §1.821(f)

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

In the matter of the above-identified application, presented herewith pursuant to 37 CFR §1.821(f) is a computer-readable copy of the Sequence Listing submitted with a Final Amendment being filed concurrently herewith. It is hereby certified that the paper and computer readable copies are identical.

Respectfully submitted,

Jeffrey L. Ihnen
Registration No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, *LLP*
1201 New York Avenue, N.W., #1000
Washington, D.C. 20005
(202) 962-4810

Atty. Docket No. 24884-1092347-09
Dated: 30 April 1997
1935.01

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. patent application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011

Filed: 07 June 1995

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

ATTN: ***Customer Correction Branch***
Application Processing Division

Batch No. Z 99

Group: 1807 (D. Rees)

REQUEST FOR EXPEDITED ISSUANCE OF CORRECTED OFFICIAL FILING RECEIPT

RECEIVED
JUL 14 1997
04

Hon. Commissioner of Patents and Trademarks
Washington, D.C. 20231

Sir:

It has come to the attention of the undersigned that an error still exists in the records of the Patent and Trademark Office with regard to the filing date of the above application. Applicants filed on 10 September 1996 a *Request for Corrected Filing Receipt* which detailed the error in question, and how it is believed to have occurred. A copy of that Request, and the date-stamped Patent Office receipt which accompanied its filing, are attached hereto.

Since Applicants are submitting payment of the Issue Fee and formal drawings for this application concurrrently with this Request, it is urged that a corrected Official Filing Receipt, reflecting the correct filing date of 07 June 1995, be issued on an expedited basis in order that issuance of this patent will not be needlessly delayed and the correct filing date of 07 June 1995 will be reflected on the printed patent.

Respectfully submitted,

Jeffrey L. Ihnen, Reg. No. 28,957

VENABLE, BAETJER, HOWARD & CIVILETTI, *LLP*
1201 New York Avenue, N.W., #1000
Washington, D.C. 20005
(202) 962-4810

Atty. Docket No. 24884-109347-09
Dated: 12 July 1997
45696.01/DC2

PTO-103X
(Rev. 8-95)

FILING RECEIPT

UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office
ASSISTANT SECRETARY AND COMMISSIONER
OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

APPLICATION NUMBER	FILING DATE	GRP ART UNIT	FIL FEE REC'D	ATTORNEY DOCKET NO.	DRWGS	TOT CL	IND CL
08/488,011	01/05/96	1803	$1,034.00	24884-109347	19	27	2

VENABLE BAETJER HOWARD AND ~~CIVILETTI~~
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

06/07/95

Receipt is acknowledged of this nonprovisional Patent Application. It will be considered in its order and you will be notified as to the results of the examination. Be sure to provide the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION when inquiring about this application. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. If an error is noted on this Filing Receipt, please write to the Application Processing Division's Customer Correction Branch within 10 days of receipt. Please provide a copy of the Filing Receipt with the changes noted thereon.

Applicant(s)

MARK H. SKOLNICK, SALT LAKE CITY, UT; DAVID E. GOLDGAR, SALT LAKE CITY, UT; YOSHIO MIKI, SALT LAKE CITY, UT; JEFF SWENSON, SALT LAKE CITY, UT; ALEXANDER KAMB, SALT LAKE CITY, UT; KEITH D. HARSHMAN, SALT LAKE CITY, UT; DONNA M. SHATTUCK-EIDENS, SALT LAKE CITY, UT; SEAN V. TAVTIGIAN, SALT LAKE CITY, UT; ROGER W. WISEMAN, DURHAM, NC; P. ANDREW FUTREAL, DURHAM, NC.

CONTINUING DATA AS CLAIMED BY APPLICANT-
THIS APPLN IS A CIP OF 08/409,305 03/24/95
WHICH IS A CIP OF 08/348,824 11/29/94
WHICH IS A CIP OF 08/308,104 09/16/94
WHICH IS A CIP OF 08/300,266 09/02/94 ABN
WHICH IS A CIP OF 08/289,221 08/12/94

FOREIGN FILING LICENSE GRANTED 04/05/96
TITLE
17Q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

PRELIMINARY CLASS: 530

RECEIVED
JUN - 6 1996
By

(see reverse)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011

Filed: 07 June 1995

ATTN: Application Processing Division's Customer Correction Branch

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

REQUEST FOR CORRECTED FILING RECEIPT

RECEIVED
Publishing
JUL 14 1997
04

Hon. Commissioner of Patents
and Trademarks
Washington, DC 20231

Sir:

It has just come to the attention of the undersigned that there is an error in the Official Filing Receipt issued for the above-mentioned application.

While the Official Filing Receipt lists the filing date of the application as 05 January 1996, the application was actually filed on 07 June 1995. This 1995 filing date was confirmed in early correspondence received from the Patent Office, such as the *Notice to File Missing Parts* mailed on 18 August 1995, which correctly lists the filing date of the application as 07 June 1995.

Formal documents for the application were filed on 01 December 1995, and it is likely that they were processed at the Patent Office on 05 January 1996, which most likely led to the listing of this latter date as the filing date of the application.

COPY

The filing date of the application should be 07 June 1995, which is the date that the application was filed at the Patent Office. Please correct the filing date and mail a corrected Official Filing Receipt to the undersigned.

Respectfully submitted,

Stephen A. Saxe

Stephen A. Saxe
Registration No. 38,609

VENABLE, BAETJER, HOWARD & CIVILETTI, *LLP*
1201 New York Avenue, N.W., Suite 1000
Washington, DC 20005
(202) 962-4848

Attorney Docket No.: 24884-109347-09
Dated: 10 September 1996
29038.01

COPY



(PTO Receipt Stamp Here)

Receipt for Documents Filed

IN RE. [xx] PATENT [ ] TM

INVENTOR/APPLICANT/OWNER: Mark H. SKOLNICK et al.

SERIAL NO. 08,488,011 REG. NO.

COUNTRY: PAT. NO.

APPLICATION FILING DATE 07 June 1995

FOR: [ ] MARK [xx] TITLE 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

DOCUMENTS FILED:

==========

REQUEST FOR CORRECTED FILING RECEIPT

==========

DATE FILED: 10 September 1996

CLIENT/ MATTER NO. 24884-109347-09

DUE DATE: (NONE)

CLIENT NAME: Myriad

ATTY/SEC. SAS:ch

Venable, Baetjer, Howard & Civiletti, LLP - 1201 New York Avenue, N.W., Suite 1000, - Washington, D.C. 20005
Telephone No. (202) 962-4800 - Facsimile: (202) 962-8300
18407.01-DC2DOCS

COPY

Application No 08/488011

NOTICE TO COMPLY WITH REQUIREMENTS FOR PATENT APPLICATIONS CONTAINING NUCLEOTIDE SEQUENCE AND/OR AMINO ACID SEQUENCE DISCLOSURES

The nucleotide and/or amino acid sequence disclosure contained in this application does not comply with the requirements for such a disclosure as set forth in 37 CFR 1.821 - 1.825 for the following reason(s):

☒ 1. This application clearly fails to comply with the requirements of 37 CFR 1.821 - 1.825. Applicant's attention is directed to these regulations, published at 1114 OG 29, May 15, 1990 and at 55 FR 18230, May 1, 1990.

☐ 2. This application does not contain, as a separate part of the disclosure on paper copy, a "Sequence Listing" as required by 37 CFR 1.821(c).

☐ 3. A copy of the "Sequence Listing" in computer readable form has not been submitted as required by 37 CFR 1.821(e).

☐ 4. A copy of the "Sequence Listing" in computer readable form has been submitted. However, the content of the computer readable form does not comply with the requirements of 37 CFR 1.822 and/or 1.823, as indicated on the attached marked-up copy of the "Raw Sequence Listing."

☐ 5. The computer readable form that has been filed with this application has been found to be damaged and/or unreadable as indicated on the attached CRF Diskette Problem Report. A substitute computer readable form must be submitted as required by 37 CFR 1.825(d).

☐ 6. The paper copy of the "Sequence Listing" is not the same as the computer readable form of the "Sequence Listing" as required by 37 CFR 1.821(e).

☐ 7. Other: _____

Applicant must provide:

☒ An initial or substitute computer readable form (CRF) copy of the "Sequence Listing"

☒ An initial or substitute paper copy of the "Sequence Listing", as well as an amendment directing its entry into the specification

☒ A statement that the content of the paper and computer readable copies are the same and, where applicable, include no new matter, as required by 37 CFR 1.821(e) or 1.821(f) or 1.821(g) or 1.825(b) or 1.825(d)

For questions regarding compliance with these requirements, please contact:

For Rules Interpretation, call (703) 308-1123
For CRF submission help, call (703) 308-4212
For PatentIn software help, call (703) 308-6856

Please return a copy of this notice with your response.

PTO UTILITY GRANT

Paper Number 26

The United States of America

The Commissioner of Patents and Trademarks

Has received an application for a patent for a new and useful invention. The title and description of the invention are enclosed. The requirements of law have been complied with, and it has been determined that a patent on the invention shall be granted under the law.

Therefore, this

United States Patent

Grants to the person(s) having title to this patent the right to exclude others from making, using, offering for sale, or selling the invention throughout the United States of America or importing the invention into the United States of America for the term set forth below, subject to the payment of maintenance fees as provided by law.

If this application was filed prior to June 8, 1995, the term of this patent is the longer of seventeen years from the date of grant of this patent or twenty years from the earliest effective U.S. filing date of the application, subject to any statutory extension.

If this application was filed on or after June 8, 1995, the term of this patent is twenty years from the U.S. filing date, subject to an statutory extension. If the application contains a specific reference to an earlier filed application or applications under 35 U.S.C. 120, 121 or 365(c), the term of the patent is twenty years from the date on which the earliest application was filed, subject to any statutory extension.

Bruce Lehman
Commissioner of Patents and Trademarks

Attest

Form PTO-1584 (Rev. 2/97)

FPI-LOM

(RIGHT INSIDE)

DATE: 4-14-97

FROM: CUSTOMER CORRECTIONS
APPLICATION PROCESSING DIVISION
Loc. 0380

SUBJ: APPLICATION FILES NEEDED FOR CORRECTION/UPDATE

GROUP ART UNIT: 1800

APPLICATION NUMBER 08/488011 IS NEEDED IMMEDIATELY FOR CORRECTION.

PLEASE ATTACH THIS FORM TO THE ABOVE APPLICATION AND RETURN IT TO THE APPLICATION PROCESSION DIVISION, CUSTOMER'S CORRECTION CP2- ROOM 6C17.

IF YOU ARE UNABLE TO LOCATE THE APPLICATION OR HAVE A QUESTION, PLEASE CALL ME AT 308-3618.

THANK YOU FOR YOUR ASSISTANCE

Norma J. White

NORMA WHITE
SUPERVISOR

U.S. PATENT AND TRADEMARK OFFICE
SERIAL NO. 08/488011

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Mark H. SKOLNICK et al.

SEP 10 1996

Serial No. 08/488,011

Filed: 07 June 1995

ATTN: Application Processing
Division's Customer
Correction Branch

For: 17q-LINKED BREAST AND OVARIAN
CANCER SUSCEPTIBILITY GENE

REQUEST FOR CORRECTED FILING RECEIPT

Hon. Commissioner of Patents
and Trademarks
Washington, DC 20231

Sir:

It has just come to the attention of the undersigned that there is an error in the Official Filing Receipt issued for the above-mentioned application.

While the Official Filing Receipt lists the filing date of the application as 05 January 1996, the application was actually filed on 07 June 1995. This 1995 filing date was confirmed in early correspondence received from the Patent Office, such as the *Notice to File Missing Parts* mailed on 18 August 1995, which correctly lists the filing date of the application as 07 June 1995.

Formal documents for the application were filed on 01 December 1995, and it is likely that they were processed at the Patent Office on 05 January 1996, which most likely led to the listing of this latter date as the filing date of the application.

The filing date of the application should be 07 June 1995, which is the date that the application was filed at the Patent Office. Please correct the filing date and mail a corrected Official Filing Receipt to the undersigned.

Respectfully submitted,

Stephen A. Saxe

Stephen A. Saxe
Registration No. 38,609

VENABLE, BAETJER, HOWARD & CIVILETTI, *LLP*
1201 New York Avenue, N.W., Suite 1000
Washington, DC 20005
(202) 962-4848

Attorney Docket No.: 24884-109347-09
Dated: 10 September 1996
29038.01

OIPE JC
FEB 15 2002
PATENT & TRADEMARK OFFICE

RTC C of C

#27
JW

2318-123A
JLI:smt

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

RECEIVED
MAY 14 2002
TECH CENTER 1600/2900

In re Patent of)
)
Mark H. SKOLNICK et al.)
)
Patent No. 5,753,441)
)
Issued: May 19, 1998)
)
For: 17Q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE)

REQUEST FOR CERTIFICATE OF CORRECTION
OF PATENT AND TRADEMARK OFFICE ERRORS

APPROVED
JUL 16 2002
FOR THE DIRECTOR OF USPTO

Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Dear Sir:

Under the provisions of 35 U.S.C. 254, Applicants submit herewith duly completed form PTO 1050 and requests that a certificate of correction be issued in connection with the above-identified patent. The error listed on the attached form occurred on the part of the Patent and Trademark Office.

The Commissioner is requested to issue and forward a Certificate of Correction to the undersigned attorney for

CERTIFICATE
FEB 22 2002
OF CORRECTION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Mark H. SKOLNICK et al.

Serial No. 08/488,011

Filed: 07 June 1995

ATTN: Application Processing
Division's Customer
Correction Branch

For: 17q-LINKED BREAST AND OVARIAN
CANCER SUSCEPTIBILITY GENE

OIPE JC10
FEB 15 2002
PATENT & TRADEMARK OFFICE

REQUEST FOR CORRECTED FILING RECEIPT

Hon. Commissioner of Patents
and Trademarks
Washington, DC 20231

Sir:

It has just come to the attention of the undersigned that there is an error in the Official Filing Receipt issued for the above-mentioned application.

While the Official Filing Receipt lists the filing date of the application as 05 January 1996, the application was actually filed on 07 June 1995. This 1995 filing date was confirmed in early correspondence received from the Patent Office, such as the *Notice to File Missing Parts* mailed on 18 August 1995, which correctly lists the filing date of the application as 07 June 1995.

Formal documents for the application were filed on 01 December 1995, and it is likely that they were processed at the Patent Office on 05 January 1996, which most likely led to the listing of this latter date as the filing date of the application.

COPY

The filing date of the application should be 07 June 1995, which is the date that the application was filed at the Patent Office. Please correct the filing date and mail a corrected Official Filing Receipt to the undersigned.

Respectfully submitted,

Stephen A. Saxe
Registration No. 38,609

VENABLE, BAETJER, HOWARD & CIVILETTI, *LLP*
1201 New York Avenue, N.W., Suite 1000
Washington, DC 20005
(202) 962-4848

Attorney Docket No.: 24884-109347-09
Dated: 10 September 1996
29038.01

COPY

PTO-103X
(Rev. 8-95)

FILING RECEIPT

OIPE
FEB 15 2002
PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office
ASSISTANT SECRETARY AND COMMISSIONER
OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

APPLICATION NUMBER	FILING DATE	GRP ART UNIT	FIL FEE REC'D	ATTORNEY DOCKET NO.	DRWGS	TOT CL	IND CL
08/488,011	01/05/96	1803	$1,034.00	24884-109347	19	27	2

VENABLE BAETJER HOWARD AND ~~CIVILETTI~~ 06/07/95
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

Receipt is acknowledged of this nonprovisional Patent Application. It will be considered in its order and you will be notified as to the results of the examination. Be sure to provide the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION when inquiring about this application. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. If an error is noted on this Filing Receipt, please write to the Application Processing Division's Customer Correction Branch within 10 days of receipt. Please provide a copy of the Filing Receipt with the changes noted thereon.

Applicant(s)

MARK H. SKOLNICK, SALT LAKE CITY, UT; DAVID E. GOLDGAR, SALT LAKE CITY, UT; YOSHIO MIKI, SALT LAKE CITY, UT; JEFF SWENSON, SALT LAKE CITY, UT; ALEXANDER KAMB, SALT LAKE CITY, UT; KEITH D. HARSHMAN, SALT LAKE CITY, UT; DONNA M. SHATTUCK-EIDENS, SALT LAKE CITY, UT; SEAN V. TAVTIGIAN, SALT LAKE CITY, UT; ROGER W. WISEMAN, DURHAM, NC; P. ANDREW FUTREAL, DURHAM, NC.

CONTINUING DATA AS CLAIMED BY APPLICANT-
THIS APPLN IS A CIP OF 08/409,305 03/24/95
WHICH IS A CIP OF 08/348,824 11/29/94
WHICH IS A CIP OF 08/308,104 09/16/94
WHICH IS A CIP OF 08/300,266 09/02/94 ABN
WHICH IS A CIP OF 08/289,221 08/12/94

FOREIGN FILING LICENSE GRANTED 04/05/96
TITLE
17Q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

PRELIMINARY CLASS: 530

RECEIVED
MAY 14 2002
TECH CENTER 1600/2900

RECEIVED
JUN - 6 1996
By

(see reverse)

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,753,441

DATED : May 19, 1998

INVENTOR(S) : Mark H. SKOLNICK et al.

It is certified that an error appears in the above-identified patent and that said Letters Patent are hereby corrected as shown below:

On the front sheet, number [22] "Jan. 5, 1996" should be --June 7, 1995--.

MAILING ADDRESS OF SENDER:

Rothwell, Figg, Ernst & Manbeck
555 13th Street, Suite 701-East
Washington, D.C. 20004

PATENT NO. 5,753,441

No. of additional copies _____

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 5,753,441

DATED : May 19, 1998

INVENTOR(S) : Mark H. SKOLNICK et al.

It is certified that an error appears in the above-identified patent and that said Letters Patent are hereby corrected as shown below:

Title page, item
~~On the front sheet, number~~ [22] "Jan. 5, 1996" should be --June 7, 1995--.

RECEIVED
MAY 14 2002
TECH CENTER 1600/2900

MAILING ADDRESS OF SENDER:

Rothwell, Figg, Ernst & Manbeck
555 13th Street, Suite 701-East
Washington, D.C. 20004

PATENT NO. 5,753,441

No. of additional copies _____

08/488011

NOTICE RE: CERTIFICATES OF CORRECTION

DATE : 05/10/02

RECEIVED
MAY 14 2002
TECH CENTER 1600/2900

Paper No.: 28

TO : **Supervisor, Art Unit** 1600

SUBJECT : Certificate of Correction Request in Patent No.: 5753441

A response to the following question is requested with respect to the accompanying request for a certifi of correction.

With respect to the change(s) requested, correcting Office and/or Applicant's errors, should the patent read as shown in the certificate of correction? No new matter should be introduced, nor should t scope or meaning of the claims be changed.

PLEASE COMPLETE THIS FORM AND
RETURN WITH FILE, WITHIN 7 DAYS,
TO CERTIFICATES OF CORRECTION BRANCH - PK 3-915/922
PALM LOCATION 7580 - TEL. N0. 305-8309

THANK YOU FOR YOUR ASSISTANCE!

Note your decision, regarding the changes requested in the Request for Certificate of Correction, placing a check mark (+) in the box that reflects your decision, which corresponds to the question check above.

☑ **YES** ☐ **NO** ☐ **Comments below**

☐ **Comments:**

Supervisor **Art Unit** 1634

PTOL-306 (REV. 2/02) **U.S. DEPARTMENT OF COMMERCE Patent and Trademark Offi**

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 5,753,441
DATED : May 19, 1998
INVENTOR(S) : Mark H. Skolnick et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Title page,
Item [22], "Jan. 5, 1996" should be -- June 7, 1995 --.

Signed and Sealed this

Thirtieth Day of July, 2002

Attest:

Attesting Officer

JAMES E. ROGAN
Director of the United States Patent and Trademark Office

AO 120 (Rev. 2/97)

TO:
Commissioner of Patents and Trademarks
Arlington, VA 22202

SOLICITOR
NOV 27 1998
U.S. PATENT & TRADEMARK OFFICE

REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK

In Compliance with 35 § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court m District of Utah on the following XX Patents or ☐ Trademarks:

DOCKET NO.	DATE FILED	U.S. DISTRICT COURT
2:98cv829 S	11/19/98	District of Utah

PLAINTIFF	DEFENDANT
Myrian Genetics, Inc.	University of Pennsylvania

	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1	5,693,473		Myriad Genetics
2	5,707,999		"
3	5,747,282		"
4	5,753,441		"
5			

In the above-entitled case, the following patent(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading		
	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1			
2			
3			
4			
5			

In the above-entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Commissioner Cpy 3—Upon termination of action, mail this copy to Commissioner
Copy 2—Upon filing document adding patent(s), mail this copy to Commissioner Copy 4—Case file copy

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	PATENT NUMBER	GROUP ART UNIT	FILE WRAPPER LOCATION
08/488,011	5753441	1634	9200 4 10011 1046 105

Change of Address/Power of Attorney

The following fields have been set to Customer Number 6449 on

- Correspondence Address
- Maintenance Fee Address

The address of record for Customer Number 6449 is:
ROTHWELL, FIGG, ERNST & MANBECK, P.C.
1425 K STREET, N.W.
SUITE 800
WASHINGTON, DC 20005

The Practitioners of record for Customer Number 6449 are:

PTO INSTRUCTIONS:

Please take the following action when the correspondence address has been changed to a customer number:
1) Add 'ADDRESS CHANGE TO CUSTOMER NUMBER' on the next available content line of the File Jacket.
2) Put a line through the old address on the File Jacket and enter the Customer Number as the new address.
3) File this Notice in the File Jacket.

Please take the following action when the correspondence address has NOT been changed:
1) File this Notice in the File Jacket

UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office

ASSISTANT SECRETARY AND COMMISSIONER
OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

CHANGE OF ADDRESS/POWER OF ATTORNEY

FILE LOCATION 9200 SERIAL NUMBER 08488011 PATENT NUMBER 5753441

THE CORRESPONDENCE ADDRESS HAS BEEN CHANGED TO CUSTOMER # 6449

THE PRACTITIONERS OF RECORD HAVE BEEN CHANGED TO CUSTOMER # 6449

THE FEE ADDRESS HAS BEEN CHANGED TO CUSTOMER # 6449

ON 04/23/03 THE ADDRESS OF RECORD FOR CUSTOMER NUMBER 6449 IS:

ROTHWELL, FIGG, ERNST & MANBECK, P.C.
1425 K STREET, N.W.
SUITE 800
WASHINGTON DC 20005

AND THE PRACTITIONERS OF RECORD FOR CUSTOMER NUMBER 6449 ARE:

17348	18125	22720	22980	27195	28957	30377	30649	31414	32408
34627	35377	35447	35745	36105	36631	37889	39881	40315	41276
42013	44066	44122	44492	47948	51613				

PTO INSTRUCTIONS: PLEASE TAKE THE FOLLOWING ACTION WHEN THE CORRESPONDENCE ADDRESS HAS BEEN CHANGED TO CUSTOMER NUMBER: RECORD, ON THE NEXT AVAILABLE CONTENTS LINE OF THE FILE JACKET, 'ADDRESS CHANGE TO CUSTOMER NUMBER'. LINE THROUGH THE OLD ADDRESS ON THE FILE JACKET LABEL AND ENTER ONLY THE 'CUSTOMER NUMBER' AS THE NEW ADDRESS. FILE THIS LETTER IN THE FILE JACKET. WHEN ABOVE CHANGES ARE ONLY TO FEE ADDRESS AND/OR PRACTITIONERS OF RECORD, FILE LETTER IN THE FILE JACKET.
THIS FILE IS ASSIGNED TO GAU 1634.

PTO-FMD
TALBOT-1/97

PATENT APPLICATION FEE DETERMINATION RECORD

Effective October 1, 1994

Application or Docket Number: 488011

CLAIMS AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA
BASIC FEE		
TOTAL CLAIMS	27 minus 20 =	* 7
INDEPENDENT CLAIMS	2 minus 3 =	* —
MULTIPLE DEPENDENT CLAIM PRESENT		

* If the difference in column 1 is less than zero, enter "0" in column 2

SMALL ENTITY RATE	SMALL ENTITY FEE	OR	OTHER THAN SMALL ENTITY RATE	OTHER THAN SMALL ENTITY FEE
	365.00	OR		730.00
x$11=		OR	x$22=	154
x38=		OR	x76=	
+120=		OR	+240=	
TOTAL		OR	TOTAL	884

CLAIMS AS AMENDED - PART II

AMENDMENT A

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA
Total	*	Minus	**	=
Independent	*	Minus	***	=
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM				

SMALL ENTITY RATE	SMALL ENTITY ADDITIONAL FEE	OR	OTHER THAN SMALL ENTITY RATE	OTHER THAN SMALL ENTITY ADDITIONAL FEE
x$11=		OR	x$22=	
x38=		OR	x76=	
+120=		OR	+240=	
TOTAL ADDIT. FEE		OR	TOTAL ADDIT. FEE	

AMENDMENT B

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA
Total	*	Minus	**	=
Independent	*	Minus	***	=
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM				

SMALL ENTITY RATE	SMALL ENTITY ADDITIONAL FEE	OR	OTHER THAN SMALL ENTITY RATE	OTHER THAN SMALL ENTITY ADDITIONAL FEE
x$11=		OR	x$22=	
x38=		OR	x76=	
+120=		OR	+240=	
TOTAL ADDIT. FEE		OR	TOTAL ADDIT. FEE	

AMENDMENT C

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA
Total	*	Minus	**	=
Independent	*	Minus	***	=
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM				

SMALL ENTITY RATE	SMALL ENTITY ADDITIONAL FEE	OR	OTHER THAN SMALL ENTITY RATE	OTHER THAN SMALL ENTITY ADDITIONAL FEE
x$11=		OR	x$22=	
x38=		OR	x76=	
+120=		OR	+240=	
TOTAL ADDIT. FEE		OR	TOTAL ADDIT. FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20."
*** If the Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3."
The Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

Form PTO 1130
(REV 2/94)

U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

PACE DATA ENTRY CODING SHEET

1ST EXAMINER	D Wilson	DATE 8/16/95
2ND EXAMINER	Russell	DATE 1-5-96

APPLICATION NUMBER	TYPE APPL	FILING DATE (MONTH DAY YEAR)	SPECIAL HANDLING	GROUP ART UNIT	CLASS	SHEETS OF DRAWING
08/488011	1	06 07 95	0	1803	530	19

TOTAL CLAIMS	INDEPENDENT CLAIMS	SMALL ENTITY?	FILING FEE	FOREIGN LICENSE	ATTORNEY DOCKET NUMBER
27	2	0	1034	1	24854-1093-47

CONTINUITY DATA

CONT CODE	STATUS CODE	PARENT APPLICATION SERIAL NUMBER	PCT APPLICATION SERIAL NUMBER	PARENT PATENT NUMBER	PARENT FILING DATE (MONTH DAY YEAR)
03	2	08/490,935	PCT/		03 24 95
13	2	08/348,824	PCT/		11 29 94
13	2	08/308,104	PCT/		09 16 94
13	2	08/300,266	PCT/		09 02 94
13	2	08/289,221	PCT/		08 12 94

PCT/FOREIGN APPLICATION DATA

FOREIGN PRIORITY CLAIMED	COUNTRY CODE	PCT/FOREIGN APPLICATION SERIAL NUMBER	FOREIGN FILING DATE (MONTH DAY YEAR)

☆U.S.G.P.O.: 1994 - 384-021

CODE SHEET FOR CONTINUING DATA

Line	Code	Serial No.	Filing Date	Status	Document No.	Issue Date
104	72	409,305	3-24-95	03		
105	82	348,824	11-29-94	03		
106	82	308,104	9-16-94	03		
107	82	300,266	9-2-94	03		
108	82	289,221	8-12-94	03		
109						
110						
111						
112						
113						
114						
115						
116						
117						

Condition and Status Codes for Continuing Data

CONDITION CODE

71	Continuation of Ser. No.
81	which is a continuation of Ser. No.
91	and a continuation of Ser. No.
72	Continuation-in-part of Ser. No.
82	which is a continuation-in-part of Ser. No.
75	and a continuation-in-part of Ser. No.
74	Division of Ser. No.
84	which is a division of Ser. No.
76	and a division of Ser. No.
86	, said Ser. No.
89	Ser. No.
90	and Ser. No.
92	each

STATUS CODE

01	Patent No.
03	abandoned
04	Defensive Publication No.
05	Published Application No.
06	Reissue Patent No.

NOTE I: When the codes 86 and 92 are used, they must be followed by 81, 82 or 84 -- conditions beginning with "which is"

NOTE II: Codes 71, 72 and 74 may be used only on the first line; one of them must be used on the first line.

GeneDX 1009, pg. 406

US005753441A

United States Patent [19]

Skolnick et al.

[11] Patent Number: 5,753,441

[45] Date of Patent: May 19, 1998

[54] **17Q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE**

[75] Inventors: **Mark H. Skolnick**; **David E. Goldgar**; **Yoshio Miki**; **Jeff Swenson**; **Alexander Kamb**; **Keith D. Harshman**; **Donna M. Shattuck-Eidens**; **Sean V. Tavtigian**, all of Salt Lake City, Utah; **Roger W. Wiseman**; **P. Andrew Futreal**, both of Durham, N.C.

[73] Assignees: **Myriad Genetics, Inc.**; **University of Utah Research Foundation**, both of Salt Lake City, Utah; **The United States of America as represented by the Department of Health and Human Services**, Washington, D.C.

[21] Appl. No.: **488,011**

[22] Filed: **Jan. 5, 1996**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 409,305, Mar. 24, 1995, abandoned, which is a continuation-in-part of Ser. No. 348,824, Nov. 29, 1994, abandoned, which is a continuation-in-part of Ser. No. 308,104, Sep. 16, 1994, abandoned, which is a continuation-in-part of Ser. No. 300,266, Sep. 2, 1994, abandoned, which is a continuation-in-part of Ser. No. 289,221, Aug. 12, 1994, abandoned.

[51] **Int. Cl.**[6] **C12Q 1/68**; C12P 19/34; C07H 21/02; C07H 21/04

[52] **U.S. Cl.** **435/6**; 435/91.1; 435/91.2; 435/7.1; 435/7.2; 435/7.9; 435/4; 436/548; 436/500; 530/387.2; 530/388.1; 424/1.11; 424/88; 536/23.1; 536/24.3; 536/24.33

[58] **Field of Search** 435/6, 91.1, 91.2, 435/7.1–7.9, 4; 536/23.1, 24.3–24.33, 548; 436/500; 424/1.11, 88; 530/387.2, 388.1

[56] **References Cited**

U.S. PATENT DOCUMENTS

5,236,844 8/1993 Basset et al. .

FOREIGN PATENT DOCUMENTS

0 518 650	12/1992	European Pat. Off. .
91/09964	7/1991	WIPO .
92/00311	1/1992	WIPO .
92/11874	7/1992	WIPO .
94/00764	1/1994	WIPO .
95/19369	7/1995	WIPO .

OTHER PUBLICATIONS

Friedman, L.S. et al. (1994). "Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families," *Nature Genetics* 8:399–404.

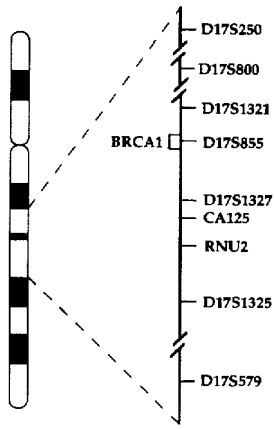
(List continued on next page.)

Primary Examiner—W. Gary Jones
Assistant Examiner—Dianne Rees
Attorney, Agent, or Firm—Venable, Baetjer, Howard & Civiletti, LLP

[57] **ABSTRACT**

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast and ovarian cancer predisposing gene (BRCA1), some mutant alleles of which cause susceptibility to cancer, in particular breast and ovarian cancer. More specifically, the invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The present invention further relates to somatic mutations in the BRCA1 gene in human breast and ovarian cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

37 Claims, 18 Drawing Sheets



OTHER PUBLICATIONS

Liang, P. et al. (1992). "Differential Display and Cloning of Messenger RNAs from Human Breast Cancer versus Mammary Epithelial Cells," *Cancer Research* 52:6966–6968.

Neuhausen, S.L. et al. (1994). "A P1–based physical map of the region from D17S776 to D17378 containing the breast cancer susceptibility gene BRCA1," *Hum. Mol. Gen.* 3:1919–1926.

Stampfer, M.R. et al. (1993). "Culture Systems for Study of Human Mammary Epithelial Cell Proliferation, Differentiation and Transformation," *Cancer Surveys* 18:7–34.

Albertsen, H., et al. (1994). "Genetic Mapping of the BRCA1 Region on Chromosome 17q21," *Am. J. Hum. Genet.* 54:516–525.

Anderson, D.E. (1972). "A Genetic Study of Human Breast Cancer," *J. Natl. Cancer Inst.* 48:1029–1034.

Arason, A., et al. (1993). "Linkage Analysis of Chromosome 17q Markers and Breast–Ovarian Cancer in Icelandic Families, and Possible Relationshp to Prostatic Cancer," *Am. J. Hum. Genet.* 52:711–717.

Bishop, T.D. and Gardner, E.J. (1980). "Analysis of the Genetic Predisposition to Cancer in Individual Pedigrees," Banbury Report 4: *Cancer Incidence in Defined Populations*, Cairns et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 389–408.

Bishop, D.T., et al. (1988). "Segregation and Linkage Analysis of Nine Utah Breast Cancer Pedigrees," *Genet. Epidemiol.* 5:151–169.

Black, D.M., et al. (1993). "A Somatic Cell Hybrid Map of the Long Arm of Human Chromosome 17, Containing the Familial Breast Cancer Locus (BRCA1)," *Am. J. Hum. Genet.* 52:702–710.

Bowcock, A.M., et al. (1993). "THRA1and D17S183 Flank an Interval of <4cM for the Breast–Ovarian Cancer Gene (BRCA1) on Chromosome 17q21," *Am. J. Hum. Genet.* 52:718–722.

Chamberlain, J.S., et al. (1993). "BRCA1 Maps Proximal to D17S579 on Chromosome 17q21 by Genetic Analysis," *Am. J. Hum. Genet.* 52:792–798.

Claus, E.B., et al. (1991). "Genetic Analysis of Breast Cancer in the Cancer and Steroid Hormone Study," *Am. J. Hum. Genet.* 48:232–242.

Cohen, B.B., et al. (1993). "Linkage of a Major Breast Cancer Gene to Chromosome 17q12–21: Results from 15 Edinburgh Families," *Am. J. Hum. Genet.* 52:723–729.

Devilee, P., et al. (1993). "Linkage to Markers for the Chromosome Region 17q12–q21 in 13 Dutch Breast Cancer Kindreds," *Am. J. Hum. Genet.* 52:730–735.

Dunphy, W.G. and Newport, J.W. (1989). "Fission Yeast p13 Blocks Mitotic Activation and Tyrosine Dephosphorylation of the Xenopus cdc2 Protein Kinase," *Cell* 58:181–191.

Easton, D.F., et al. (1993). "Genetic Linkage Analysis in Familial Breat and Ovarian Cancer: Results from 214 Families," *Am. J. Hum. Genet.* 52:678–701.

Fain, P.R. (1992). "Third international workshop on human chromosome 17 mapping," *Cytogen. Cell Genet.* 60:178–186.

Feunteun, J., et al. (1993). "A Breast–Ovarian Cancer Susceptibility Gene Maps to Chromosome 17q21," *Am. J. Hum. Genet.* 52:736–742:

Ford, D., et al. (1993). "The risks of cancer in BRCA1 mutation carriers," *Am. J. Hum. Genet. 53(supplement)*:abstract No. 298.

Futreal, P.A., et al. (1994). "BRCA1 Mutations in Primary Breast and Ovarian Carcinomas," *Science* 266:120–122.

Go, R.C.P., et al. (1983). "Genetic Epidemiology of Breast Cancer and Associated Cancers in High–Risk Families. I. Segregation Analysis," *J. Natl. Cancer Inst.* 71:455–461.

Goldgar, D.E., et al. (1993). "Chromosome 17q Linkage Studies of 18 Utah Breast Cancer Kindreds," *Am. J. Hum. Genet.* 52:743–748.

Gould, K.L. and Nurse, P. (1989). "Tyrosine phosphorylation of the fission yeast cdc2$^+$ protein kinase regulates entry into mitosis," *Nature* 342:39–45.

Hall, J.M., et al. (1990). "Linkage of Early–Onset Familial Breast Cancer to Chromosome 17q21," *Science* 250:1684–1689.

Hall, J.M., et al. (1992). "Closing in on a Breast Cancer Gene on Chromosome 17q," *Am. J. Hun. Genet.* 50:1235–1241.

Hosking, L., et al. (1995). "A somatic BRCA1 mutation in an ovarian tumour," *Nature Genetics* 9:343–344.

Ishibashi, T., et al. (1992). "Expression cloning of a human dual–specificity phosphatase," *Proc. Natl. Acad. Sci. USA* 89:12170–12174.

Kamb A., et al. (1994). "Localization of the VHR Phosphatase Gene and Its Analysis as a Candidate for BRCA1," *Genomics* 23:163–167.

Kumagai, A. and Dunphy, W.G. (1991). "The cdc25 Protein Controls Tyrosine Dephosphorylation of the cdc2 Protein in a Cell–Free System," *Cell* 64:903–914.

Lindblom, A., et al. (1993). "Linkage Analysis with Markers on 17q in 29 Swedish Breast Cancer Families," *Am. J. Hum. Genet.* 52:749–753.

Margaritte, P., et al. (1992). "Linkage of Familial Breast Cancer to Chromosome 17q21 May Not be Restricted to Early–Onset Disease," *Am. J. Hum. Genet.* 50:1231–1234.

Mazoyer, S., et al. (1993). "Linkage Analysis of 19 French Breast Cancer Families, with Five Chromosome 17q Markers," *Am. J. Hum. Genet.* 52:754–760.

Miki, Y., et al. (1994). "A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1," *Science* 266:66–71.

Milner, B.J., et al. (1993). "Linkage Studies with 17q and 18q Markers in a Breast/Ovarian Cancer Family," *Am. J. Hum. Genet.* 52:761–766.

Narod, S.A., et al. (1991). "Familial breast–ovarian cancer locus on chromosome 17q12–q23," *Lancet* 338:82–83.

Newman, B. et al. (1988). "Inheritance of human breast cancer: Evidence for autosomal dominant transmission in high–risk families," *Proc. Natl. Acad. Sci. USA* 85:3044–3048.

O'Connell, P., et al. (1994). "A Radiation Hybrid Map of the BRCA1 Region," *Am. J. Hum. Genet.* 54:526–534.

Roberts, L. (1993). "Zeroing In on a Breast Cancer Susceptibility Gene," *Science* 259:622–625.

Shattuck–Eidens, D., et al. (1995). "A Collaborative Survey of 80 Mutations in the BRCA1 Breast and Ovarian Cancer Susceptibility Gene," JAMA 273:535–541.

Simard, J., et al. (1994). "Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families," *Nature Genetics* 8:392–398.

Smith, S.A., et al. (1992). "Allele losses in the region 17q12–21 in familial breast and ovarian cancer involve the wild–type chromosome," *Nature Genetics* 2:128–131.

Smith, S.A., et al. (1993). "Genetic Heterogeneity and Localization of a Familial Breast–Ovarian Cancer Gene on Chromosome 17q12–q21," *Am. J. Hum. Genet.* 52:767–776.

Spurr, N.K., et al. (1993). "Linkage Analysis of Early–Onset Breast and Ovarian Cancer Families, with Markers on the Long Arm of Chromosome 17." *Am. J. Hum. Genet.* 52:777–785.

Teare, M.D., et al. (1993). "A Linkage Study in Seven Breast Cancer Families." *Am. J. Hum. Genet.* 52: 786–788.

Thompson, M.E., et al. (1995). "Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression." *Nature Genetics* 9:444–450.

Ullrich, A. and Schlessinger, J. (1990). "Signal Transduction by Receptors with Tyrosine Kinase Activity." *Cell* 61:203–212.

Williams, W.R. and Anderson, D.E. (1984). "Genetic Epidemiology of Breast Cancer: Segregation Analysis of 200 Danish Pedigrees." *Genet. Epidemiol.* 1:7–20.

Zimmermann, W., et al. (1993). "Linkage Analysis in German Breast Cancer Families with Early Onset of the Disease, Using Highly Polymorphic Markers from the Chromosome 17q11–q24 Region." *Am. J. Hum. Genet.* 52:789–791.

Goldberg et al. Clinical Chemistry 39: 2360–2374, 1993.

Wooster et al., Science 265:2088–2090, 1994.

Sato et al., Cancer Res. 52: 1643–1646, 1992.

Malkin et al.,Science 250: 1233–1238, 1990.

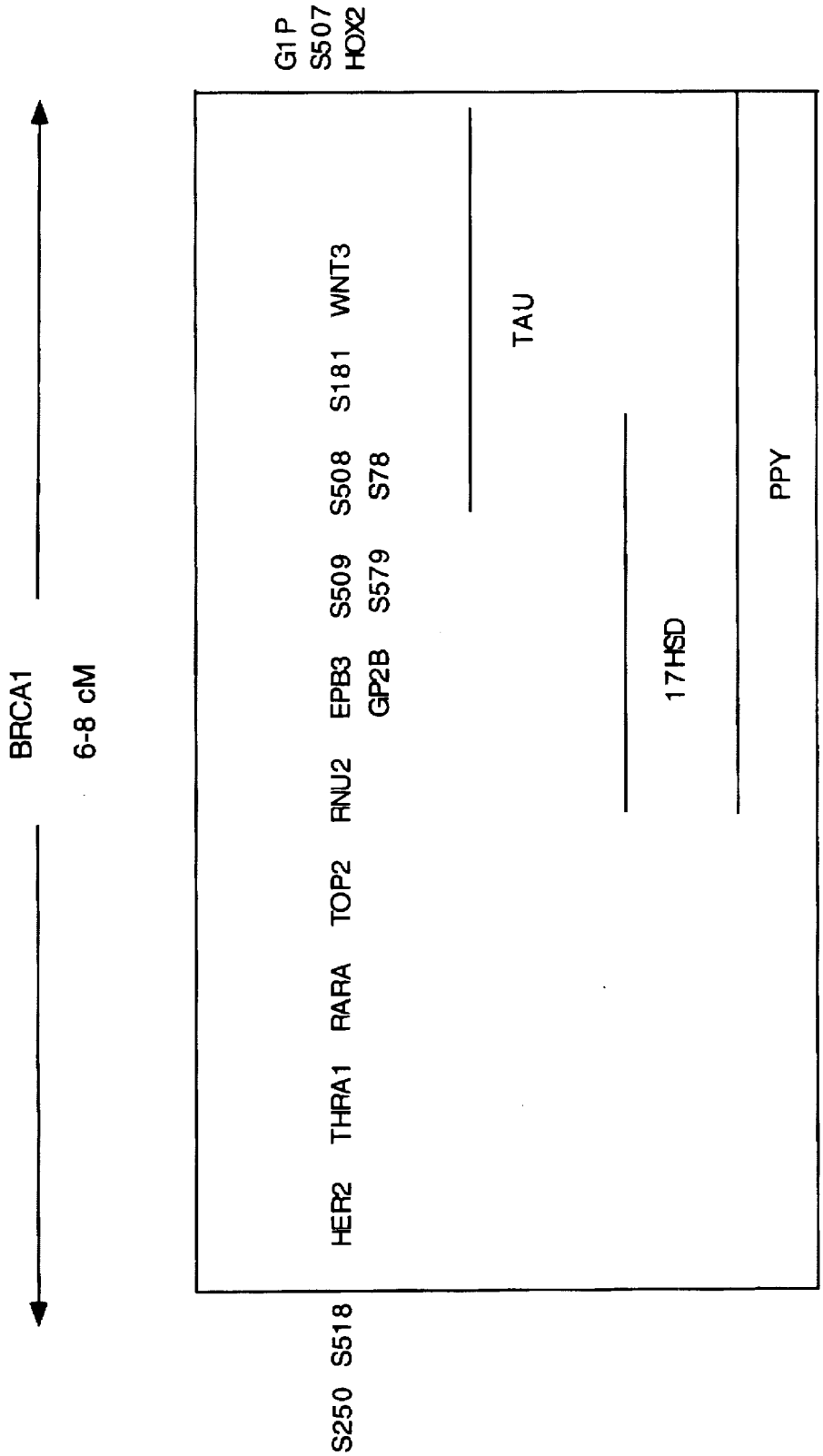
Boyd, Nature Genetics 9: 335–336, 1995.

Merajver et al., Nature Genetics 9:439–443, 1995.

Castilla et al Nature Genetics 8: 387–391, 1994.

Friedman et al., Nature Genetics 8: 1–6, 1994.

Goldgar et al. J. of Nat. Cancer Inst, 86: 200–209,1994.



Map of the early onset breast and ovarian cancer region (BRCA1)

FIG. 1

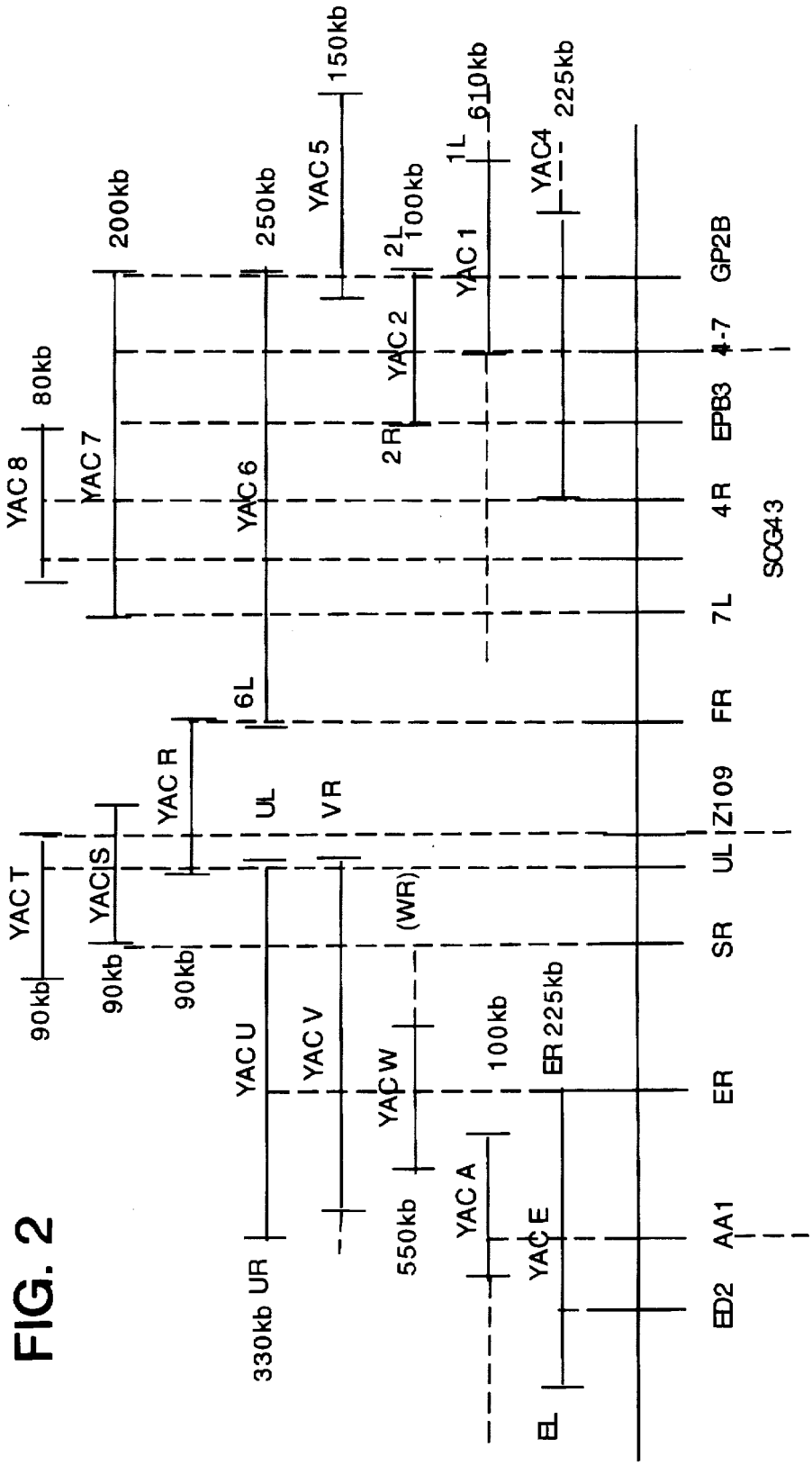

YAC T
90kb
YAC S
90kb
YAC R
90kb
YAC U
330kb UR
YAC V
YAC W
550kb
YAC A
YAC E
EL
100kb
ER 225kb
UL
VR
(WR)
6L
YAC 6
250kb
YAC 7
200kb
YAC 8
80kb
2R
YAC 2
100kb
YAC 5
150kb
1L
YAC 1
610kb
YAC4
225kb
ED2
AA1
ER
SR
UL
Z109
FR
7L
4R
SCG43
EPB3
4-7
GP2B

FIG. 2

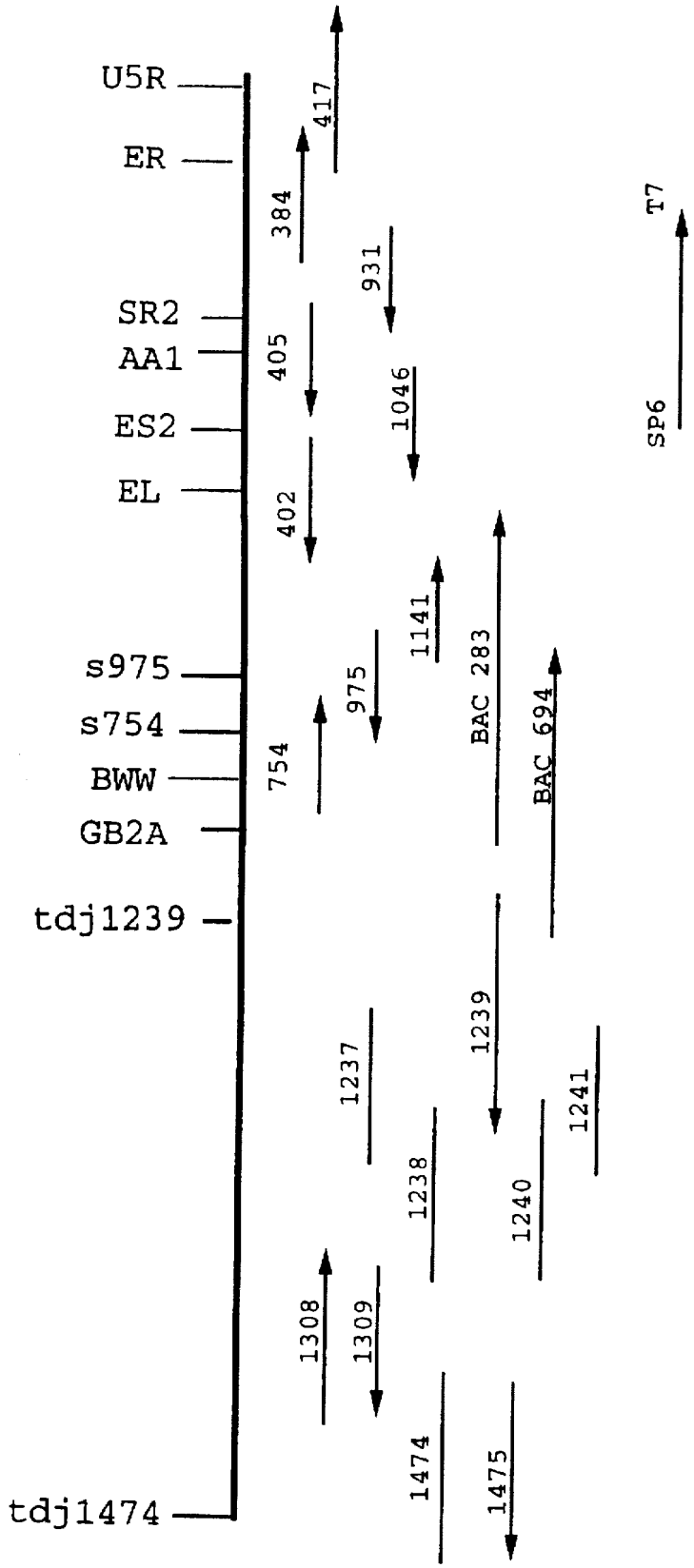
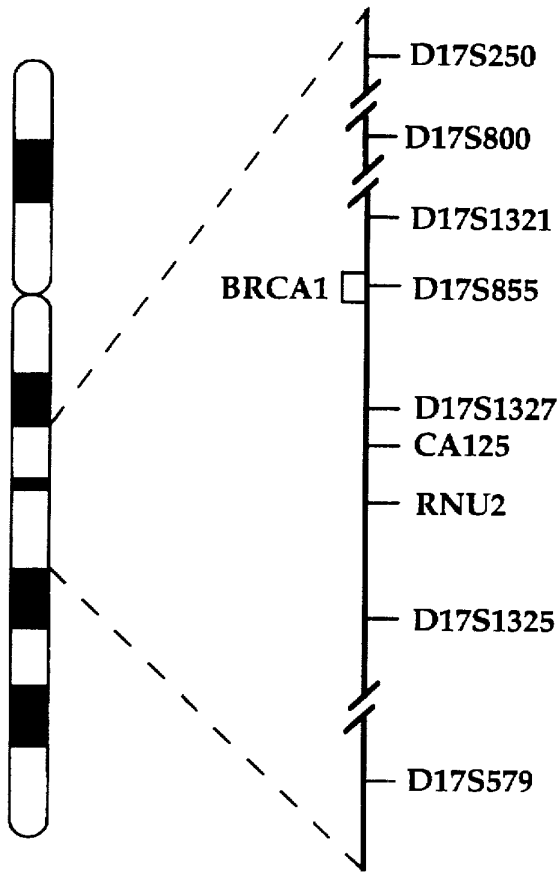

U5R
ER
SR2
AA1
ES2
EL
s975
s754
BWW
GB2A
tdj1239
tdj1474
417
384
931
405
1046
402
1141
975
754
BAC 283
BAC 694
1239
1237
1241
1238
1240
1308
1309
1474
1475
T7
SP6

FIG. 3

FIG. 4



SEQ. ID NO:		
82	BRCA1	CPICLELIKEPVSTK-CDHIFCKFCMLKLLNQKK---GPSQCPLCK
83	RPT1	CPICLELLKEPVSAD-CNHSFCRACITLNYESNRNTDGKGNCPVCR
84	RIN1	CPICLDMLKNTMTTKECLHRFCSDCIVTALRS-----GNKECPTCR
85	RFP1	CPVCLQYFAEPMMLD-CGHNICCACLARCWGTAC---TNVSCPQCR
	C3HC4 motif	C--C----------C-H--C--C----------------C--C

FIG. 5

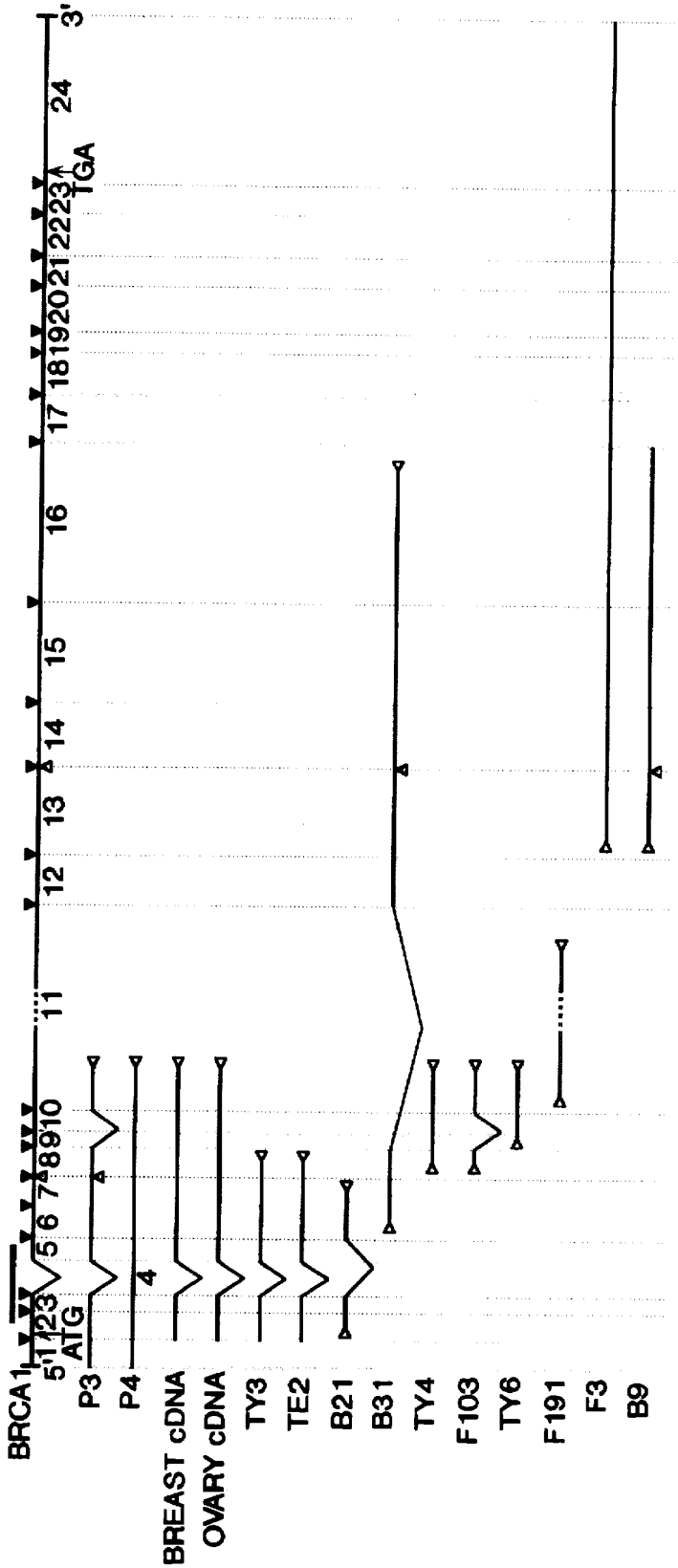

BRCA1
5'
ATG
1 2 3
4
5 6 7 8 9 10
11
12
13
14
15
16
17 18 19 20 21 22 23
TGA
24
3'
P3
P4
BREAST cDNA
OVARY cDNA
TY3
TE2
B21
B31
TY4
F103
TY6
F191
F3
B9

FIG. 6

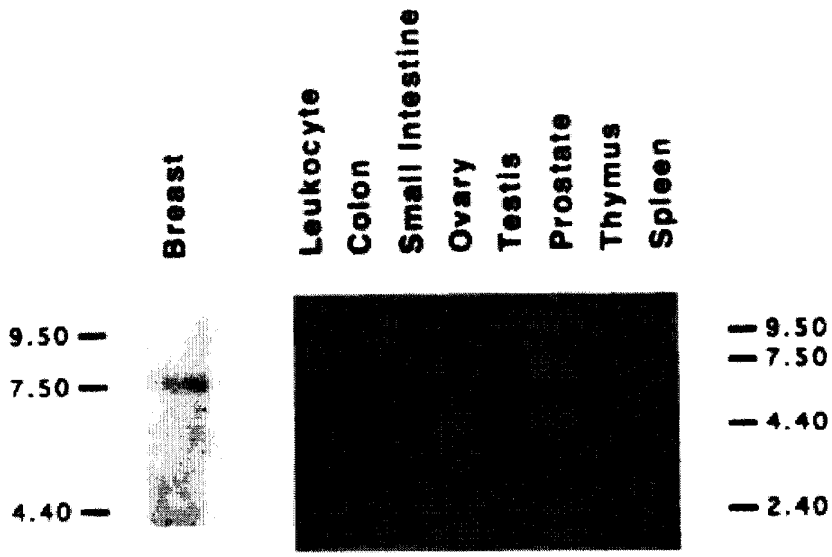

Breast
Leukocyte
Colon
Small Intestine
Ovary
Testis
Prostate
Thymus
Spleen
9.50
7.50
4.40
9.50
7.50
4.40
2.40

FIG. 7

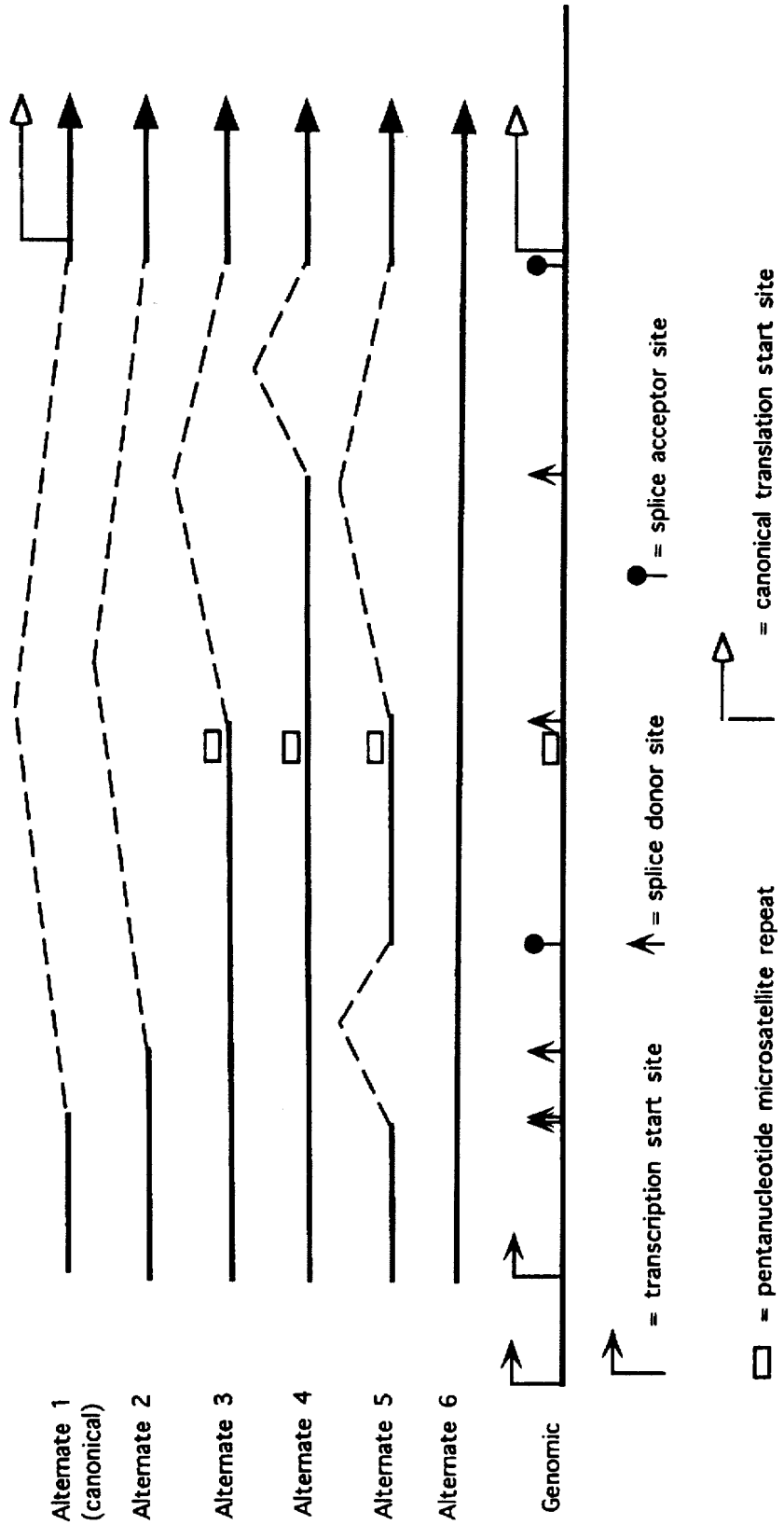

Alternate 1 (canonical)
Alternate 2
Alternate 3
Alternate 4
Alternate 5
Alternate 6
Genomic
= transcription start site
= splice donor site
= splice acceptor site
= canonical translation start site
= pentanucleotide microsatellite repeat

FIG. 8

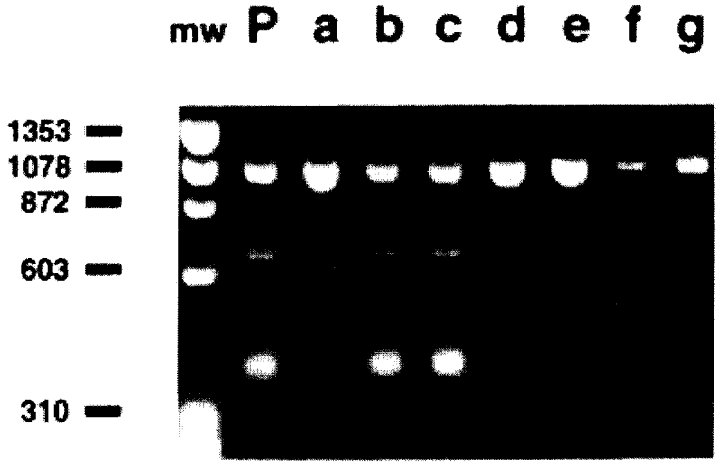


FIG. 9A

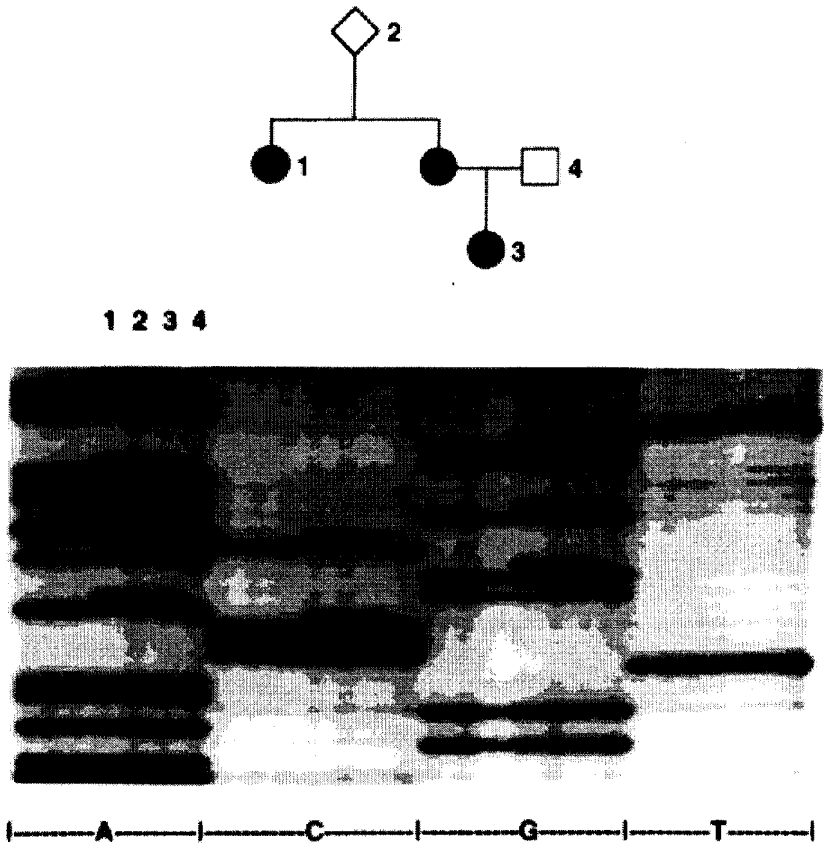


FIG. 9B

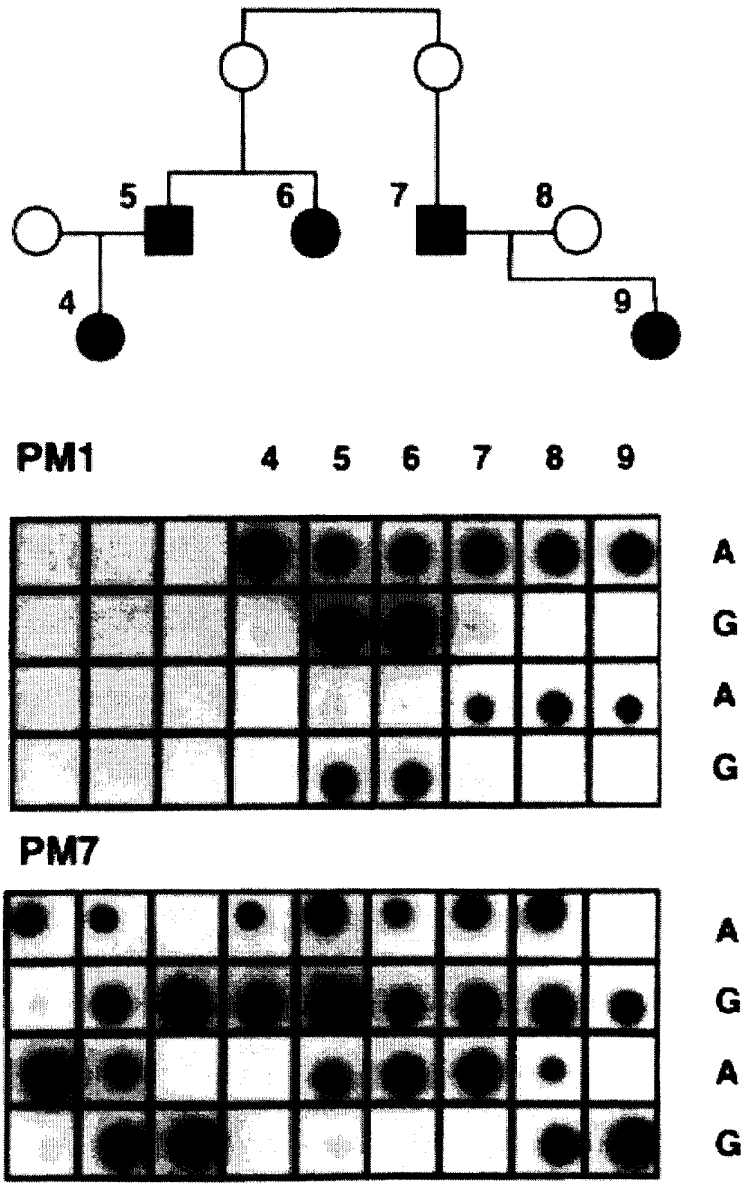


FIG. 9C

```
   1  gaggctagagggcaggcactttatggcaaactcaggtagaattcttcctcttccgtctct
  61  ttccttttacgtcatcggggagactgggtggcaatcgcagcccgagagacgcatggctct
 121  ttctgccctccatcctctgatgtaccttgatttcgtattctgagaggctgctgcttagcg
 181  gtagccccttggtttccgtggcaacggaaaagcgcgggaattacagataaattaaaactg
 241  cgactgcgcggcgtgAGCTCGCTGAGACTTCCTGGACCCCGCACCAGGCTGTGGGGTTTC
 301  TCAGATAACTGGGCCCCTGCGCTCAGGAGGCCTTCACCCTCTGCTCTGGGTAAAGgtagt
 361  agagtcccgggaaagggacaggggcccaagtgatgctctggggtactggcgtgggagag
 421  tggatttccgaagctgacagatgggtattctttgacgggggggtaggggcggaacctgaga
 481  ggcgtaaggcgttgtgaaccctggggagggggcagtttgtaggtcgcgagggaagcgct
 541  gaggatcaggaaggggcactgagtgtccgtgggggaatcctcgtgataggaactggaat
 601  atgccttgaggggacactatgtctttaaaaacgtcggctggtcatgaggtcaggagttc
 661  cagaccagcctgaccaacgtggtgaaactccgtctctactaaaaatacnaaaattagccg
 721  ggcgtggtgccgctccagctactcaggaggctgaggcaggagaatcgctagaacccggga
 781  ggcggaggttgcagtgagccgagatcgcgccattgcactccagcctgggcgacagagcga
 841  gactgtctcaaaacaaaacaaaacaaaacaaaacaaaaaacaccggctggtatgtatgag
 901  aggatgggaccttgtggaagaagaggtgccaggaatatgtctgggaaggggaggagacag
 961  gattttgtgggagggagaacttaagaactggatccatttgcgccattgagaaagcgcaag
1021  agggaagtagaggagcgtcagtagtaacagatgctgccggcagggatgtgcttgaggagg
1081  atccagagatgagagcaggtcactgggaaaggttaggggcggggaggccttgattggtgt
1141  tggtttggtcgttgttgattttggttttatgcaagaaaaagaaaacaaccagaaacattg
1201  gagaaagctaaggctaccaccacctacccggtcagtcactcctctgtagctttctctttc
1261  ttggagaaaggaaaagacccaaggggttggcagcgatatgtgaaaaaattcagaatttat
1321  gttgtctaattacaaaaagcaacttctagaatctttaaaaataaaggacgttgtcattag
1381  ttcttctggtttgtattattctaaaaccttccaaatcttcaaatttactttatttttaaaa
1441  tgataaaatgaagttgtcattttataaacctttaaaaagatatatatatatgtttttct
1501  aatgtgttaaagTTCATTGGAACAGAAAGAAATGGATTTATCTGCTCTTCGCGTTGAAGA
1561  AGTACAAAATGTCATTAATGCTATGCAGAAAATCTTAGAGTGTCCCATCTGgtaagtcag
1621  cacaagagtgtattaatttgggattcctatgattatctcctatgcaaatgaacagaattg
1681  accttacatactagggaagaaaagacatgtctagtaagattaggctattgtaattgctga
1741  ttttcttaactgaagaactttaaaaatatagaaaatgattccttgttctccatccactct
1801  gcctctcccactcctctccttttcaacacaatcctgtggtccgggaaagacagggctctg
1861  tcttgattggttctgcactgggcaggatctgttagatactgcatttgctttctccagctc
1921  taaavvvvvvvvvvvvvaaatgctgatgatagtatagagtattgaagggatcaatataat
1981  tctgttttgatatctgaaagctcactgaaggtaaggatcgtattctctgctgtattctca
2041  gttcctgacacagcagacatttaataaatattgaacgaacttgaggccttatgttgactc
2101  agtcataacagctcaaagttgaacttattcactaagaatagctttatttttaaataaatt
2161  attgagcctcatttatttttcttttttctccccccccctaccctgctagTCTGGAGTTGATCA
2221  AGGAACCTGTCTCCACAAAGTGTGACCACATATTTTGCAAgtaagtttgaatgtgttatg
2281  tggctccattattagcttttgtttttgtccttcataacccaggaaacacctaactttata
2341  gaagctttacttttcttcaattaagtgagaacgaaaatccaactccatttcattctttctc
2401  agagagtatatagttatcaaaagttggttgtaatcatagttcctggtaaagttttgacat
2461  atattatctttttttttttttgagacaagtctcgctctgtcgcccaggctggagtgcagt
2521  ggcatgaggcttgctcactgcacctccgcccccgagttcagcgactctvvvvvvvvvvvv
2581  vtgagatctagaccacatggtcaaagagatagaatgtgagcaataaatgaaccttaaatt
2641  tttcaacagctacttttttttttttttttgagacagGGKCTTACTCTGTTGTCCCAGCT
2701  GGAGTACAGWGTGCGATCATGAGGCTTACTGTTGCTTGACTCCTAGGCTCAAGCGATCCT
2761  ATCACCTCAGTCTCCAAGTAGCTGGACTgtaagtgcacaccaccatatccagctaaattt
2821  tgtgttttctgtagagacggggtttcgccatgtttcccaggctggtcttgaactttgggc
2881  ttaacccgtctgcccacctaggcatcccaaagtgctaggattacaggtgtgagtcatcat
2941  gcctggccagtattttagttagctctgtcttttcaagtcatatacaagttcattttcttt
3001  taagtttagttaacaaccttatatcatgtattcttttctagcataaagaaagattcgagg
```

FIG. 10A

```
3061  ccvvvvvvvvvvvvvvtgtgatcataacagtaagccatatgcatgtaagttcagttttcat
3121  agatcattgcttatgtagtttaggtttttgcttatgcagcatccaaaaacaattaggaaa
3181  ctattgcttgtaattcacctgccattacttttttaaatggctcttaagggcagttgtgaga
3241  ttatcttttcatggctatttgccttttgagtattctttctacaaaaggaagtaaattaaa
3301  ttgttctttctttctttataatttatagATTTTGCATGCTGAAACTTCTCAACCAGAAGA
3361  AAGGGCCTTCACAGTGTCCTTTATGTAAGAATGATATAACCAAAAGgtatataatttggt
3421  aatgatgctaggttggaagcaaccacagtaggaaaaagtagaaattatttaataacatag
3481  cgttcctataaaaccattcatcagaaaaatttataaaagagtttttagcacacagtaaat
3541  tatttccaaagttattttcctgaaagttttatgggcatctgccttatacaggtattgvvv
3601  vvvvvvvvvvggtaggcttaaatgaatgacaaaaagttactaaatcactgccatcacacg
3661  gtttatacagatgtcaatgatgtattgattatagaggttttctactgttgctgcatctta
3721  tttttatttgtttacatgtcttttcttattttagtgtccttaaaaggttgataatcactt
3781  gctgagtgtgtttctcaaacaatttaatttcagGAGCCTACAAGAAAGTACGAGATTTAG
3841  TCAACTTGTTGAAGAGCTATTGAAAATCATTTGTGCTTTTCAGCTTGACACAGGTTTGGA
3901  GTgtaagtgttgaatatcccaagaatgacactcaagtgctgtccatgaaaactcaggaag
3961  tttgcacaattactttctatgacgtggtgataagacctttttagtctaggttaattttagt
4021  tctgtatctgtaatctattttaaaaaattactcccactggtctcacaccttatttvvvvv
4081  vvvvvvvvaaaaaatcacaggtaaccttaatgcattgtcttaacacaacaaagagcatac
4141  atagggtttctcttggtttctttgattataattcatacatttttctctaactgcaaacat
4201  aatgttttcccttgtattttacagATGCAAACAGCTATAATTTTGCAAAAAAGGAAAATA
4261  ACTCTCCTGAACATCTAAAAGATGAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACC
4321  GTGCCAAAAGACTTCTACAGAGTGAACCCGAAAATCCTTCCTTGgtaaaaccatttgttt
4381  tcttcttcttcttcttcttcttttcttttttttttcttttttttttttgagatggagtctt
4441  gctctgtggcccaggctagaagcagtcctcctgccttagccncccttagtagctgggatta
4501  caggcacgcgcaccatgccaggctaatttttgtatttttagtagagacggggtttcatca
4561  tgttggccaggctggtctcgaactcctaacctcaggtgatcvvvvvvvvvvvvvvatgatg
4621  gagatcttaaaaagtaatcattctggggctgggcgtagtagcttgcacctgtaatcccag
4681  cacttcgggaggctgaggcaggcagataatttgaggtcaggagtttgagaccagcctggc
4741  caacatggtgaaacccatctctactaaaaatacaaaaattagctgggtgtggtggcacgt
4801  acctgtaatcccagctactcgggaggcggaggcacaagaattgcttgaacctaggacgcg
4861  gaggttgcagcgagccaagatcgcgccactgcactccagcctgggccgtagagtgagact
4921  ctgtctcaaaaaagaaaaaaaagtaattgttctagctgggcgcagtggctcttgcctgta
4981  atcccagcactttgggaggccaaggcgggtggatctcgagtcctagagttcaagaccagc
5041  ctaggcaatgtggtgaaaccccatcgctacaaaaaatacaaaaattagccaggcatggtg
5101  gcgtgcgcatgtagtcccagctccttgggaggctgaggtgggaggatcacttgaacccag
5161  gagacagaggttgcagtgaaccgagatcacgccaccacgctccagcctgggcaacagaac
5221  aagactctgtctaaaaaaatacaaataaaataaaagtagttctcacagtaccagcattca
5281  tttttcaaaagatatagagctaaaaaggaaggaaaaaaaaagtaatgttgggcttttaaa
5341  tactcgttcctatactaaatgttcttaggagtgctggggttttattgtcatcatttatcc
5401  tttttaaaaatgttattggccaggcacggtggctcatggctgtaatcccagcactttggg
5461  aggccgaggcaggcagatcacctgaggtcaggagtgtgagaccagcctggccaacatggc
5521  gaaacctgtctctactaaaaatacaaaaattaactaggcgtggtggtgtacgcctgtagt
5581  cccagctactcgggaggctgaggcaggagaatcaactgaaccagggaggtggaggttgca
5641  gtgtgccgagatcacgccactgcactctagcctggcaacagagcaagattctgtctcaaa
5701  aaaaaaaaacatatatacacatatatcccaaagtgctgggattacatatatatatatata
5761  tatatattatatatatatatatatatatgtgatatatatgtgatatatatataacata
5821  tatatatgtaatatatatgtgatatatatataatatatatatgtaatatatatgtgatat
5881  atatatatacacacacacacacatatatatgtatgtgtgtgtacacacacacacacaaat
5941  tagccaggcatagttgcacacgcttggtagacccagctactcaggaggctgagggaggag
6001  aatctcttgaacttaggaggcggaggttgcagtgagctgagattgcgccactgcactcca
6061  gcctgggtgacagagcaggactctgtacacccccaaaacaaaaaaaaaagttatcagat
```

FIG. 10B

```
6121  gtgattggaatgtatatcaagtatcagcttcaaaatatgctatattaatacttcaaaaat
6181  tacacaaataatacataatcaggtttgaaaaatttaagacaacmsaaraaaaaawycmaa
6241  tcacamatatcccacacattttattattmctmctmcwattattttgwagagmctgggtct
6300  cacycykttgctwatgctggtctttgaacyccykgccycaarcartcctsctccabcctc
6361  ccaargtgctggggatwataggcatgarctaaccgcacccagccccagacattttagtgt
6421  gtaaattcctgggcatttttcaaggcatcatacatgttagctgactgatgatggtcaat
6481  ttattttgtccatggtgtcaagtttctcttcaggaggaaaagcacagaactggccaagaa
6541  ttgcttgactgttctttaccatactgtttagCAGGAAACCAGTCTCAGTGTCCAACTCTC
6601  TAACCTTGGAACTGTGAGAACTCTGAGGACAAAGCAGCGGATACAACCTCAAAAGACGTC
6661  TGTCTACATTGAATTGGgtaagggtctcaggtttttaagtatttaataataattgctgg
6721  attccttatcttatagttttgccaaaaatcttggtcataatttgtatttgtggtaggcag
6781  ctttgggaagtgaatttttatgagccctatggtgagttataaaaaaatgtaaaagacgcagt
6841  tcccaccttgaagaatcttactttaaaaagggagcaaaagaggccaggcatggtggctca
6901  cacctgtaatcccagcactttgggaggccaaagtgggtggatcacctgaggtcgggagtt
6961  cgagaccagcctagccaacatggagaaactctgtctgtaccaaaaaataaaaaattagcc
7021  aggtgtggtggcacataactgtaatcccagctactcgggaggctgaggcaggagaatcac
7081  ttgaacccgggaggtggaggttgcggtgaaccgagatcgcaccattgcactccagcctgg
7141  gcaaaaatagcgaaactccatctaaaaaaaaaaaagagagcaaaagaaagamtmtctggt
7201  tttaamtmtgtgtaaatatgtttttggaaagatggagagtagcaataagaaaaaacatga
7261  tggattgctacagtatttagttccaagataaattgtactagatgaggaagccttttaaga
7321  agagctgaattgccaggcgcagtggctcacgcctgtaatcccagcactttgggaggccga
7381  ggtgggcggatcacctgaggtcgggagttcaagaccagcctgaccaacatggagaaaccc
7441  catctctactaaaaaaaaaaaaaaaaaaattagccggggtggtggcttatgcctgtaatc
7501  ccagctactcaggaggctgaggcaggagaatcgcttgaacccaggaagcagaggttgcag
7561  tgagccaagatcgcaccattgcactccagcctaggcaacaagagtgaaactccatctcaa
7621  aaaaaaaaaaaaagagctgaatcttggctgggcaggatggctcgtgcctgtaatcctaac
7681  gctttggaagaccgaggcagaaggattggttgagtccacgagtttaagaccagcctggcc
7741  aacataggggaaccctgtctctatttttaaaataataatacatttttggccggtgcggtg
7801  gctcatgcctgtaatcccaatactttgggaggctgaggcaggtagatcacctgaggtcag
7861  agttcgagaccagcctggataacctggtgaaacccctctttactaaaaatacaaaaaaaa
7921  aaaaaaattagctgggtgtggtagcacatgcttgtaatcccagctacttgggaggctgag
7981  gcaggagaatcgcttgaaccagggaggcggaggttacaatgagccaacactacaccactg
8041  cactccagcctgggcaatagagtgagactgcatctcaaaaaaataataatttttaaaaat
8101  aataaatttttttaagcttataaaaagaaaagttgaggccagcatagtagctcacatctg
8161  taatctcagcagtggcagaggattgcttgaagccaggagtttgagaccagcctgggcaac
8221  atagcaagacctcatctctacaaaaaaatttctttttttaaattagctgggtgtggtggtg
8281  tgcatctgtagtcccagctactcaggaggcagaggtgagtggatacattgaacccaggag
8341  tttgaggctgtagtgagctatgatcatgccactgcactccaacctgggtgacagagcaag
8401  acctccaaaaaaaaaaaaaaaaagagctgctgagctcagaattcaaactgggctctcaaat
8461  tggattttcttttagaatatatttataattaaaaaggatagccatcttttgagctcccag
8521  gcaccaccatctatttatcataacacttactgttttcccccccttatgatcataaattcct
8581  agacaacaggcattgtaaaaatagttatagtagttgatatttaggagcacttaactatat
8641  tccaggcactattgtgcttttcttgtataactcattagatgcttgtcagacctctgagat
8701  tgttcctattatacttattttacagatgagaaaattaaggcacagagaagttatgaaatt
8761  tttccaaggtattaaacctagtaagtggctgagccatgattcaaacctaggaagttagat
8821  gtcagagcctgtgctttttttttgttttgttttttgttttcagtagaaacgggggtctca
8881  ctttgttggccaggctggtcttgaactcctaacctcaaataatccacccatctcggcctc
8941  ctcaagtgctgggattacaggtgagagccactgtgcctggcgaagcccatgcctttaacc
9001  acttctctgtattacatactagcttaactagcattgtacctgccacagtagatgctcagt
9061  aaatatttctagttgaatatctgtttttcaacaagtacatttttttaaccctttaatta
9121  agaaaacttttattgatttatttttttgggggggaaatttttttagGATCTGATTCTTCTGAA
```

FIG. 10C

```
 9181  GATACCGTTAATAAGGCAACTTATTGCAGgtgagtcaaagagaaccttgtctatgaagc
 9241  tggtattttcctatttagttaatattaaggattgatgtttctctcttttaaaaatattt
 9301  taacttttattttaggttcagggatgtatgtgcagtttgttatataggtaaacacacgac
 9361  ttgggatttggtgtatagatttttttcatcatccgggtactaagcataccccacagtttt
 9421  ttgtttgctttctttctgaatttctccctcttcccaccttcctccctcaagtaggctggt
 9481  gtttctccagactagaatcatggtattggaagaaaccttagagatcatctagtttagttc
 9541  tctcattttatagtggaggaaataccctttttgtttgttggatttagttattagcactgt
 9601  ccaaaggaatttaggataacagtagaactctgcacatgcttgcttctagcagattgttct
 9661  ctaagttcctcatatacagtaatattgacacagcagtaattgtgactgatgaaaatgttc
 9721  aaggacttcattttcaactctttctttcctctgttccttatttccacatatctctcaagc
 9781  tttgtctgtatgttatataataaactacaagcaaccccaactatgttacctaccttcctt
 9841  aggaattattgcttgacccaggtttttttttttttttttttggagacggggtcttgccct
 9901  gttgccaggatggagtgtagtggcgccatctcggctcactgcaatctccaactccctggt
 9961  tcaagcgattctcctgtctcaatctcacgagtagctgggactacaggtatacaccaccac
10021  gcccggttaattgaccattccatttctttctttctctctttttttttttttttttgaga
10081  cagagtcttgctctgttgcccaggctggagtacagaggtgtgatctcacctctccgcaac
10141  gtctgcctcccaggttgaagccatactcctgcctcagcctctctagtagctgggactaca
10201  ggcgcgcgccaccacacccggctaatttttgtatttttagtagagatggggtttcaccat
10261  gttggccaggctggtcttgaactcatgacctcaagtggtccacccgcctcagcctcccaa
10321  agtgctggaattacaggcttgagccaccgtgcccagcaaccatttcatttcaactagaag
10381  tttctaaaggagagagcagctttcactaactaaataagattggtcagctttctgtaatcg
10441  aaagagctaaaatgtttgatcttggtcatttgacagttctgcatacatgtaactagtgtt
10501  tcttattaggactctgtctttttccctatagTGTGGGAGATCAAGAATTGTTACAAATCAC
10561  CCCTCAAGGAACCAGGGATGAAATCAGTTTGGATTCTGCAAAAAAGGgtaatggcaaagt
10621  ttgccaacttaacaggcactgaaaagagagtgggtagatacagtactgtaattagattat
10681  tctgaagaccatttgggacctttacaacccacaaaatctcttggcagagttagagtatca
10741  ttctctgtcaaatgtcgtggtatggtctgatagatttaaatggtactagactaatgtacc
10801  tataataagaccttcttgtaactgattgttgccctttcgctttttttttgtttgtttgt
10861  ttgtttttttttgagatggggtctcactctgttgcccaggctggagtgcagtgatgcaat
10921  cttggctcactgcaacctccacctccaaaggctcaagctatcctcccacttcagcctcct
10981  gagtagctgggactacaggcgcatgccaccacacccggttaatttttgtggttttatag
11041  agatggggtttcaccatgttaccgaggctggtctcaaactcctggactcaagcagtctgc
11101  ccacttcagcctcccaaagtgctgcagttacaggcttgagccactgtgcctggcctgccc
11161  tttacttttaattggtgtatttgtgtttcatcttttacctactggtttttaaatataggg
11221  agtggtaagtctgtagatagaacagagtattaagtagacttaatggccagtaatctttag
11281  agtacatcagaaccagttttctgatggccaatctgcttttaattcactcttagacgttag
11341  agaaataggtgtggtttctgcatagggaaaattctgaaattaavvvvvvvvvvvvvvgatc
11401  ctaagtggaaataatctaggtaaataggaattaaatgaaagagtatgagctacatcttca
11461  gtatacttggtagtttatgaggttagtttctctaatatagccagttggttgatttccacc
11521  tccaaggtgtatgaagtatgtatttttttaatgacaattcagtttttgagtaccttgtta
11581  tttttgtatattttcagCTGCTTGTGAATTTTCTGAGACGGATGTAACAAATACTGAACA
11641  TCATCAACCCAGTAATAATGATTTGAACACCACTGAGAAGCGTGCAGCTGAGAGGCATCC
11701  AGAAAAGTATCAGGGTAGTTCTGTTTCAAACTTGCATGTGGAGCCATGTGGCACAAATAC
11761  TCATGCCAGCTCATTACAGCATGAGAACAGCAGTTTATTACTCACTAAAGACAGAATGAA
11821  TGTAGAAAAGGCTGAATTCTGTAATAAAAGCAAACAGCCTGGCTTAGCAAGGAGCCAACA
11881  TAACAGATGGGCTGGAAGTAAGGAAACATGTAATGATAGGCGGACTCCCAGCACAGAAAA
11941  AAAGGTAGATCTGAATGCTGATCCCCTGTGTGAGAGAAAAGAATGGAATAAGCAGAAACT
12001  GCCATGCTCAGAGAATCCTAGAGATACTGAAGATGTTCCTTGGATAACACTAAATAGCAG
12061  CATTCAGAAAGTTAATGAGTGGTTTTCCAGAAGTGATGAACTGTTAGGTTCTGATGACTC
12121  ACATGATGGGGAGTCTGAATCAAATGCCAAAGTAGCTGATGTATTGGACGTTCTAAATGA
12181  GGTAGATGAATATTCTGGTTCTTCAGAGAAAATAGACTTACTGGCCAGTGATCCTCATGA
```

FIG. 10D

```
12241  GGCTTTAATATGTAAAAGTGAAAGAGTTCACTCCAAATCAGTAGAGAGTAATATTGAAGG
12301  CCAAATATTTGGGAAAACCTATCGGAAGAAGGCAAGCCTCCCCAACTTAAGCCATGTAAC
12361  TGAAAATCTAATTATAGGAGCATTTGTTACTGAGCCACAGATAATACAAGAGCGTCCCCT
12421  CACAAATAAATTAAAGCGTAAAAGGAGACCTACATCAGGCCTTCATCCTGAGGATTTTAT
12481  CAAGAAAGCAGATTTGGCAGTTCAAAAGACTCCTGAAATGATAAATCAGGGAACTAACCA
12541  AACGGAGCAGAATGGTCAAGTGATGAATATTACTAATAGTGGTCATGAGAATAAAACAAA
12601  AGGTGATTCTATTCAGAATGAGAAAAATCCTAACCCAATAGAATCACTCGAAAAAGAATC
12661  TGCTTTCAAAACGAAAGCTGAACCTATAAGCAGCAGTATAAGCAATATGGAACTCGAATT
12721  AAATATCCACAATTCAAAAGCACCTAAAAAGAATAGGCTGAGGAGGAAGTCTTCTACCAG
12781  GCATATTCATGCGCTTGAACTAGTAGTCAGTAGAAATCTAAGCCCACCTAATTGTACTGA
12841  ATTGCAAATTGATAGTTGTTCTAGCAGTGAAGAGATAAAGAAAAAAAAGTACAACCAAAT
12901  GCCAGTCAGGCACAGCAGAAACCTACAACTCATGGAAGGTAAAGAACCTGCAACTGGAGC
12961  CAAGAAGAGTAACAAGCCAAATGAACAGACAAGTAAAAGACATGACAGCGATACTTTCCC
13021  AGAGCTGAAGTTAACAAATGCACCTGGTTCTTTTACTAAGTGTTCAAATACCAGTGAACT
13081  TAAAGAATTTGTCAATCCTAGCCTTCCAAGAGAAGAAAAAGAAGAGAAACTAGAAACAGT
13141  TAAAGTGTCTAATAATGCTGAAGACCCCAAAGATCTCATGTTAAGTGGAGAAAGGGTTTT
13201  GCAAACTGAAAGATCTGTAGAGAGTAGCAGTATTTCATTGGTACCTGGTACTGATTATGG
13261  CACTCAGGAAAGTATCTCGTTACTGGAAGTTAGCACTCTAGGGAAGGCAAAAACAGAACC
13321  AAATAAATGTGTGAGTCAGTGTGCAGCATTTGAAAACCCCAAGGGACTAATTCATGGTTG
13381  TTCCAAAGATAATAGAAATGACACAGAAGGCTTTAAGTATCCATTGGGACATGAAGTTAA
13441  CCACAGTCGGGAAACAAGCATAGAAATGGAAGAAAGTGAACTTGATGCTCAGTATTTGCA
13501  GAATACATTCAAGGTTTCAAAGCGCCAGTCATTTGCTCCGTTTTCAAATCCAGGAAATGC
13561  AGAAGAGGAATGTGCAACATTCTCTGCCCACTCTGGGTCCTTAAAGAAACAAAGTCCAAA
13621  AGTCACTTTTGAATGTGAACAAAAGGAAGAAAATCAAGGAAAGAATGAGTCTAATATCAA
13681  GCCTGTACAGACAGTTAATATCACTGCAGGCTTTCCTGTGGTTGGTCAGAAAGATAAGCC
13741  AGTTGATAATGCCAAATGTAGTATCAAAGGAGGCTCTAGGTTTTGTCTATCATCTCAGTT
13801  CAGAGGCAACGAAACTGGACTCATTACTCCAAATAAACATGGACTTTTACAAAACCCATA
13861  TCGTATACCACCACTTTTTCCCATCAAGTCATTTGTTAAAACTAAATGTAAGAAAAATCT
13921  GCTAGAGGAAAACTTTGAGGAACATTCAATGTCACCTGAAAGAGAAATGGGAAATGAGAA
13981  CATTCCAAGTACAGTGAGCACAATTAGCCGTAATAACATTAGAGAAAATGTTTTTAAAGA
14041  AGCCAGCTCAAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGGCTCCAGTAT
14101  TAATGAAATAGGTTCCAGTGATGAAAACATTCAAGCAGAACTAGGTAGAAACAGAGGGCC
14161  AAAATTGAATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGTCTATAAACAAAGTCT
14221  TCCTGGAAGTAATTGTAAGCATCCTGAAATAAAAAAGCAAGAATATGAAGAAGTAGTTCA
14281  GACTGTTAATACAGATTTCTCTCCATATCTGATTTCAGATAACTTAGAACAGCCTATGGG
14341  AAGTAGTCATGCATCTCAGGTTTGTTCTGAGACACCTGATGACCTGTTAGATGATGGTGA
14401  AATAAAGGAAGATACTAGTTTTGCTGAAAATGACATTAAGGAAAGTTCTGCTGTTTTTAG
14461  CAAAAGCGTCCAGAAAGGAGAGCTTAGCAGGAGTCCTAGCCCTTTCACCCATACACATTT
14521  GGCTCAGGGTTACCGAAGAGGGGCCAAGAAATTAGAGTCCTCAGAAGAGAACTTATCTAG
14581  TGAGGATGAAGAGCTTCCCTGCTTCCAACACTTGTTATTTGGTAAAGTAAACAATATACC
14641  TTCTCAGTCTACTAGGCATAGCACCGTTGCTACCGAGTGTCTGTCTAAGAACACAGAGGA
14701  GAATTTATTATCATTGAAGAATAGCTTAAATGACTGCAGTAACCAGGTAATATTGGCAAA
14761  GGCATCTCAGGAACATCACCTTAGTGAGGAAACAAAATGTTCTGCTAGCTTGTTTTCTTC
14821  ACAGTGCAGTGAATTGGAAGACTTGACTGCAAATACAAACACCCAGGATCCTTTCTTGAT
14881  TGGTTCTTCCAAACAAATGAGGCATCAGTCTGAAAGCCAGGGAGTTGGTCTGAGTGACAA
14941  GGAATTGGTTTCAGATGATGAAGAAAGAGGAACGGGCTTGGAAGAAAATAATCAAGAAGA
15001  GCAAAGCATGGATTCAAACTTAGgtattggaaccaggtttttgtgtttgccccagtctat
15061  ttatagaagtgagctaaatgtttatgcttttggggagcacattttacaaatttccaagta
15121  tagttaaaggaactgcttcttaaacttgaaacatgttcctcctaaggtgcttttcataga
15181  aaaaagtccttcacacagctaggacgtcatctttgactgaatgagctttaacatcctaat
15241  tactggtggacttacttctggtttcattttataaagcaaatccCggtgtcccaaagcaag
```

FIG. 10E

15301 gaatttaatcattttgtgtgacatgaaagtaaatccagtcctgccaatgagaagaaaaag
15361 acacagcaagttgcagcgtttatagtctgcttttacatctgaacctctgtttttgttatt
15421 taagGTGAAGCAGCATCTGGGTGTGAGAGTGAAACAAGCGTCTCTGAAGACTGCTCAGGG
15481 CTATCCTCTCAGAGTGACATTTTAACCACTCaggtaaaaagcgtgtgtgtgtgtgcacat
15541 gcgtgtgtgtggtgtcctttgcattcagtagtatgtatcccacattcttaggtttgctga
15601 catcatctctttgaattaatggcacaattgtttgtggttcattgtcvvvvvvvvvvvvvvn
15661 gngaatgtaatcctaatatttcncnccnacttaaaagaataccactccaanggcatcnca
15721 atacatcaatcaattggggaattgggattttccctcnctaacatcantggaataatttca
15781 tggcattaattgcatgaatgtggttagattaaaaggtgttcatgctagaacttgtagttc
15841 catactaggtgatttcaattcctgtgctaaaattaatttgtatgatatattntcatttaa
15901 tggaaagcttctcaaagtatttcattttcttggtaccatttatcgtttttgaAGCAGAGG
15961 GATACCATGCAACATAACCTGATAAAGCTCCAGCAGGAAATGGCTGAACTAGAAGCTGTG
16021 TTAGAACAGCATGGGAGCCAGCCTTCTAACAGCTACCCTTCCATCATAAGTGACTCTTCT
16081 GCCCTTGAGGACCTGCGAAATCCAGAACAAAGCACATCAGAAAAAGgtgtgtattgttgg
16141 ccaaacactgatatcttaagcaaaattctttccttcccctttatctccttctgaagagta
16201 aggacctagctccaacattttatgatccttgctcagcacatgggtaattatggagccttg
16261 gttcttgtccctgctcacaactaatataccagtcagagggacccaaggcagtcattcatg
16321 ttgtcatctgagatacctacaacaagtagatgctatggggagcccatggvvvvvvvvvvv
16381 vvccattggtgctagcatctgtctgttgcattgcttgtgtttataaaattctgcctgata
16441 tacttgttaaaaaccaatttgtgtatcatagattgatgcttttgaaaaaaatcagtattc
16501 taacctgaattatcactatcagaacaaagcagtaaagtagatttgttttctcattccatt
16561 taaagCAGTATTAACTTCACAGAAAAGTAGTGAATACCCTATAAGCCAGAATCCAGAAGG
16621 CCTTTCTGCTGACAAGTTTGAGGTGTCTGCAGATAGTTCTACCAGTAAAAATAAAGAACC
16681 AGGAGTGGAAAGgtaagaaacatcaatgtaaagatgctgtggtatctgacatctttattt
16741 atattgaactctgattgttaatttttttcaccatactttctccagttttttgcatacag
16801 gcatttatacacttttattgctctaggatacttcttttgtttaatcctatataggvvvvv
16861 vvvvvvvvggataagntcaagagatattttgataggtgatgcagtgatnaattgngaaaa
16921 tttnctgcctgcttttaatcttcccccgttctttcttcctncctccctccctcctnccct
16981 cccgtccttncctttcctttccctccctccttccncctttctttccntctntctttcctttctt
17041 tcctgtctaccttttcttttccttcctccctttcctttttctttttctttctttcctttcctttt
17101 ctttccttttctttccttttcctttctttcttgacagagtcttgctctgtcactcaggctgg
17161 agtgcagtggcgtgatctcgnctcactgcaaccctctgtctcccaggttcaagcaattttc
17221 ctgcctcagcctcccgagtagctgagattacaggcgccagccaccacacccagctactga
17281 cctgcttttvvvvvvvvvvvvvvaaacagctgggagatatggtgcctcagaccaaccccat
17341 gttatatgtcaaccctgacatattggcaggcaacatgaatccagacttctaggctgtcat
17401 gcgggctctttttgccagtcatttctgatctctctgacatgagctgtttcatttatgct
17461 ttggctgcccagcaagtatgatttgtcctttcacaattggtggcgatggttttctccttc
17521 catttatctttctagGTCATCCCCTTCTAAATGCCCATCATTAGATGATAGGTGGTACAT
17581 GCACAGTTGCTCTGGGAGTCTTCAGAATAGAAACTACCCATCTCAAGAGGAGCTCATTAA
17641 GGTTGTTGATGTGGAGGAGCAACAGCTGGAAGAGTCTGGGCCACACGATTTGACGGAAAC
17701 ATCTTACTTGCCAAGGCAAGATCTAGgtaatatttcatctgctgtattggaacaaacact
17761 ytgattttactctgaatcctacataaagatattctggttaaccaacttttagatgtacta
17821 gtctatcatggacacttttgttatacttaattaagcccactttagaaaaatagctcaagt
17881 gttaatcaaggtttacttgaaaattattgaaactgttaatccatctatattttaattaat
17941 ggtttaactaatgattttgaggatgwgggagtcktggtgtactctamatgtattatttca
18001 ggccaggcatagtggctcacgcctggtaatcccagtayycmrgagcccgaggcaggtgga
18061 gccagctgaggtcaggagttcaagacctgtcttggccaacatgggngaaaccctgtcttc
18121 ttcttaaaaaanacaaaaaaaattaactgggttgtgcttaggtgnatgccccgnatccta
18181 gttnttcttgnggggttgagggaggagatcacnttggaccccggaggggnggggtgggggng
18241 agcaggncaaaacacngacccagctggggtggaagggaagcccactcnaaaaaannttnv
18301 vvvvvvvvvvvvttttaggaaacaagctactttggatttccaccaacacctgtattcat

FIG. 10F

```
18361  gtacccatttttctcttaacctaactttattggtctttttaattcttaacagagaccaga
18421  actttgtaattcaacattcatcgttgtgtaaattaaacttctcccattcctttcagAGGG
18481  AACCCCTTACCTGGAATCTGGAATCAGCCTCTTCTCTGATGACCCTGAATCTGATCCTTC
18541  TGAAGACAGAGCCCCAGAGTCAGCTCGTGTTGGCAACATACCATCTTCAACCTCTGCATT
18601  GAAAGTTCCCCAATTGAAAGTTGCAGAATCTGCCCAGAGTCCAGCTGCTGCTCATACTAC
18661  TGATACTGCTGGGTATAATGCAATGGAAGAAAGTGTGAGCAGGGAGAAGCCAGAATTGAC
18721  AGCTTCAACAGAAAGGGTCAACAAAAGAATGTCCATGGTGGTGTCTGGCCTGACCCCAGA
18781  AGAATTTgtgagtgtatccatatgtatctccctaatgactaagacttaacaacattctgg
18841  aaagagttttatgtaggtattgtcaattaataacctagaggaagaaatctagaaaacaat
18901  cacagttctgtgtaatttaatttcgattactaatttctgaaaatttagaayvvvvvvvvv
18961  vvvvncccnncccccnaatctgaaatgggggtaacccccccccaaccganacntgggtng
19021  cntagaganttttaatggcccnttctgaggnacanaagcttaagccaggngacgtggancn
19081  atgngttgtttnttgtttggttacctccagcctgggtgacagagcaagactctgtctaaa
19141  aaaaaaaaaaaaaaaaatcgactttaaatagttccaggacacgtgtagaacgtgcaggat
19201  tgctacgtaggtaaacatatgccatggtgggataactagtattctgagctgtgtgctaga
19261  ggtaactcatgataatggaatatttgatttaatttcagATGCTCGTGTACAAGTTTGCCA
19321  GAAAACACCACATCACTTTAACTAATCTAATTACTGAAGAGACTACTCATGTTGTTATGA
19381  AAACAGgtataccaagaacctttacagaataccttgcatctgctgcataaaaccacatga
19441  ggcgaggcacggtggcgcatgcctgtaatcgcagcactttgggaggccgaggcgggcaga
19501  tcacgagattaggagatcgagaccatcctggccagcatggtgaaaccccgtctctactan
19561  naaatggnaaaattanctgggtgtggtcgcgtgcncctgtagtcccagctactcgtgagg
19621  ctgaggcaggagaatcacttgaaccggggaaatggaggtttcagtgagcagagatcatnc
19681  ccctncattccagcctggcgacagagcaaggctccgtcnccnaaaaaataaaaaaaaacg
19741  tgaacaaataagaatatttgttgagcatagcatggatgatagtcttctaatagtcaatca
19801  attactttatgaaagacaaataatagttttgctgcttccttacctccttttgttttgggt
19861  taagatttggagtgtgggccaggcacvvvvvvvvvvvvvgatctatagctagccttggcg
19921  tctagaagatgggtgttgagaagagggagtggaaagatatttcctctggtcttaacttca
19981  tatcagcctcccctagacttccaaatatccatacctgctggttataattagtggtgtttt
20041  cagcctctgattctgtcaccaggggttttagaatcataaatccagattgatcttgggagt
20101  gtaaaaaactgaggctctttagcttcttaggacagcacttcctgattttgttttcaactt
20161  ctaatcctttgagtgtttttcattctgcagATGCTGAGTTTGTGTGTGAACGGACACTGA
20221  AATATTTTCTAGGAATTGCGGGAGGAAAATGGGTAGTTAGCTATTTCTgtaagtataata
20281  ctatttctcccctcctccctttaacacctcagaattgcatttttacacctaacatttaac
20341  acctaaggtttttgctgatgctgagtctgagttaccaaaaggtctttaaattgtaatact
20401  aaactacttttatctttaatatcactttgttcaagataagctggtgatgctgggaaaatg
20461  ggtctcttttataactaataggacctaatctgctcctagcaatgttagcatatgagctag
20521  ggatttatttaatagtcggcaggaatccatgtgcarcagncaaacttataatgtttaaat
20581  taaacatcaactctgtctccagaaggaaactgctgctacaagccttattaaagggctgtg
20641  gctttagagggaaggacctctcctctgtcattcttcctgtgctcttttgtgaatcgctga
20701  cctctctatctccgtgaaaagagcacgttcttctgctgtatgtaacctgtcttttctatg
20761  atctctvvvvvvvvvvvvvvnaaaaacggggnngggantgggccttaaanccaaagggcna
20821  actccccaaccattnaaaaantgacnggggattattaaaancggcgggaaacatttcacn
20881  gcccaactaatattgttaaattaaaaccaccaccnctgcnccaaggagggaaactgctgc
20941  tacaagccttattaaagggctgtggctttagagggaaggacctctcctctgtcattcttc
21001  ctgtgctcttttgtgaatcgctgacctctctatgtccgtgaaaagagcacgttcttcgtc
21061  tgtatgtaacctgtcttttctatgatctctttagGGGTGACCCAGTCTATTAAAGAAAGA
21121  AAAATGCTGAATGAGgtaagtacttgatgttacaaactaaccagagatattcattcagtc
21181  atatagttaaaaatgtatttgcttccttccatcaatgcaccactttccttaacaatgcac
21241  aaattttccatgataatgaggatcatcaagaattatgcaggcctgcactgtggctcatac
21301  ctataatcccagcgctttgggaggctgaggcgcttggatcvvvvvvvvvvvvvaattttt
21361  tgtatttttagtagagatgaggttcaccatgttggtctagatctggtgtcgaacgtcctg
```

FIG. 10G

```
21421  acctcaagtgatctgccagcctcagtctcccaaagtgctaggattacaggggtgagccac
21481  tgcgcctggcctgaatgcctaaaatatgacgtgtctgctccacttccattgaaggaagct
21541  tctctttctcttatcctgatgggttgtgtttggtttctttcagCATGATTTTGAAGTCAG
21601  AGGAGATGTGGTCAATGGAAGAAACCACCAAGGTCCAAAGCGAGCAAGAGAATCCCAGGA
21661  CAGAAAGgtaaagctccctccctcaagttgacaaaaatctcaccccaccactctgtattc
21721  cactccccttgcagagatgggccgcttcattttgtaagacttattacatacatacacag
21781  tgctagatactttcacacaggttctttttcactcttccatcccaaccacataaataagt
21841  attgtctctactttatgaatgataaaactaagagatttagagaggctgtgtaatttggat
21901  tcccgtctcgggttcagatcvvvvvvvvvvvvvttggcctgattggtgacaaaagtgaga
21961  tgctcagtccttgaatgacaaagaatgcctgtagagttgcaggtccaactacatatgcac
22021  ttcaagaagatcttctgaaatctagtagtgttctggacattggactgcttgtccctggga
22081  agtagcagcagaaatgatcggtggtgaacagaagaaaaagaaaagctcttcctttttgaa
22141  agtctgtttttgaataaaagccaatattcttttataactagattttccttctctccatt
22201  cccctgtccctctctcttcctctcttcttccagATCTTCAGGGGGCTAGAAATCTGTTGC
22261  TATGGGCCCTTCACCAACATGCCCACAGgtaagagcctgggagaaccccagagttccagc
22321  accagcctttgtcttacatagtggagtattataagcaaggtcccacgatgggggttcctc
22381  agattgctgaaatgttctagaggctattctatttctctaccactctccaaacaaaacagc
22441  acctaaatgttatcctatggcaaaaaaaaactataccttgtccccttctcaagagcatg
22501  aaggtggttaatagttaggattcagtatgttatgtgttcagatggcgttgagctgctgtt
22561  agtgccvvvvvvvvvvvvvvtttgagagactatcaaaccttataccaagtggccttatgga
22621  gactgataaccagagtacatggcatatcagtggcaaattgacttaaaatccataccccta
22681  ctattttaagaccattgtcctttggagcagagagacagactctcccattgagaggtcttg
22741  ctataagccttcatccggagagtgtagggtagagggcctgggttaagtatgcagattact
22801  gcagtgattttacatgtaaatgtccattttagATCAACTGGAATGGATGGTACAGCTGTG
22861  TGGTGCTTCTGTGGTGAAGGAGCTTTCATCATTCACCCTTGGCACAgtaagtattgggtg
22921  ccctgtcagtgtgggaggacacaatattctctcctgtgagcaagactggcacctgtcagt
22981  ccctatggatgcccctactgtagcctcagaagtcttctctgcccacatacctgtgccaaa
23041  agactccatvvvvvvvvvvvvvvggtggtacgtgtctgtagttccagctacttgggaggct
23101  gagatggaaggattgcttgagcccaggaggcagaggtggnannttacgctgagatcacac
23161  cactgcactccagcctgggtgacagagcaagaccctgtctcaaaaacaaacaaaaaaaat
23221  gatgaagtgacagttccagtagtcctactttgacactttgaatgctctttccttcctggg
23281  gatccagGGTGTCCACCCAATTGTGGTTGTGCAGCCAGATGCCTGGACAGAGGACAATGG
23341  CTTCCATGgtaaggtgcctcgcatgtacctgtgctattagtggggtccttgtgcatgggt
23401  ttggtttatcactcattacctggtgcttgagtagcacagttcttggcacatttttaaata
23461  tttgttgaatgaatggctaaaatgtctttttgatgttttattgttatttgttttatatt
23521  gtaaaagtaatacatgaactgtttccatggggtgggagtaagatatgaatgttcatcacv
23581  vvvvvvvvvvvvcagtaatcctnagaactcatacgaccgggcccctggagtcgntgnttn
23641  gagcctagtccnggagaatgaattgacactaatctctgcttgtgttctctgtctccagCA
23701  ATTGGGCAGATGTGTGAGGCACCTGTGGTGACCCGAGAGTGGGTGTTGGACAGTGTAGCA
23761  CTCTACCAGTGCCAGGAGCTGGACACCTACCTGATACCCCAGATCCCCCACAGCCACTAC
23821  TGACTGCAGCCAGCCACAGGTACAGAGCCACAGGACCCCAAGAATGAGCTTACAAAGTGG
23881  CCTTTCCAGGCCCTGGGAGCTCCTCTCACTCTTCAGTCCTTCTACTGTCCTGGCTACTAA
23941  ATATTTTATGTACATCAGCCTGAAAAGGACTTCTGGCTATGCAAGGGTCCCTTAAAGATT
24001  TTCTGCTTGAAGTCTCCCTTGGAAAT
```

FIG. 10H

170-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of application Ser. No. 08/409,305 filed on 24 March 1995 now abandoned, which is a continuation-in-part of application Ser. No. 08/348,824 filed on 29 November 1994 now abandoned, which is a continuation-in-part of application Ser. No. 08/308,104 filed on 16 September 1994 now abandoned, which is a continuation-in-part of application Ser. No. 08/300,266, filed on 2 September 1994 now abandoned, which is a continuation-in-part of application Ser. No. 08/289,221, filed on 12 August 1994 now abandoned, all incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast and ovarian cancer predisposing gene (BRCA1), some mutant alleles of which cause susceptibility to cancer, in particular, breast and ovarian cancer. More specifically, the invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The present invention further relates to somatic mutations in the BRCA1 gene in human breast and ovarian cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

The publications and other materials used herein to illuminate the background of the invention, and in particular, cases to provide additional details respecting the practice, are incorporated herein by reference, and for convenience, are referenced by author and date in the following text and respectively grouped in the appended List of References.

BACKGROUND OF THE INVENTION

The genetics of cancer is complicated, involving multiple dominant, positive regulators of the transformed state (oncogenes) as well as multiple recessive, negative regulators (tumor suppressor genes). Over one hundred oncogenes have been characterized. Fewer than a dozen tumor suppressor genes have been identified, but the number is expected to increase beyond fifty (Knudson, 1993).

The involvement of so many genes underscores the complexity of the growth control mechanisms that operate in cells to maintain the integrity of normal tissue. This complexity is manifest in another way. So far, no single gene has been shown to participate in the development of all, or even the majority of human cancers. The most common oncogenic mutations are in the ll-ras gene, found in 10–15% of all solid tumors (Anderson et al., 1992). The most frequently mutated tumor suppressor genes are the TP53 gene, homozygously deleted in roughly 50% of all tumors, and

CDKN2, which was homozygously deleted in 46% of tumor cell lines examined (Kamb et al., 1994). Without a target that is common to all transformed cells, the dream of a "magic bullet" that can destroy or revert cancer cells while leaving normal tissue unharmed is improbable. The hope for a new generation of specifically targeted antitumor drugs may rest on the ability to identify tumor suppressor genes or oncogenes that play general roles in control of cell division.

The tumor suppressor genes which have been cloned and characterized influence susceptibility to: 1) Retinoblastoma (RB1); 2) Wilms' tumor (WT1); 3) Li-Fraumeni (TP53); 4) Familial adenomatous polyposis (APC); 5) Neurofibromatosis type 1 (NF1); 6) Neurofibromatosis type 2 (NF2); 7) von Hippel-Lfindau syndrome (VHL); 8) Multiple endocrine neoplasia type 2A (MEN2A); and 9) Melanoma (CDKN2).

Tumor suppressor loci that have been mapped genetically but not yet isolated include genes for:

Multiple endocrine neoplasia type 1 (MEN1); Lynch cancer family syndrome 2 (LCFS2); Neuroblastoma (NB); Basal cell nevus syndrome (BCNS); Beckwith-Wiedemann syndrome (BWS); Renal cell carcinoma (RCC); Tuberous sclerosis 1 (TSC1); and Tuberous sclerosis 2 (TSC2). The tumor suppressor genes that have been characterized to date encode products with similarities to a variety of protein types, including DNA binding proteins (WT1), ancillary transcription regulators (RB1), GTPase activating proteins or GAPs (NF1), cytoskeletal components (NF2), membrane bound receptor kinases (MEN2A), cell cycle regulators (CDKN2) and others with no obvious similarity to known proteins (APC and VHL).

In many cases, the tumor suppressor gene originally identified through genetic studies has been shown to be lost or mutated in some sporadic tumors. This result suggests that regions of chromosomal aberration may signify the position of important tumor suppressor genes involved both in genetic predisposition to cancer and in sporadic cancer.

One of the hallmarks of several tumor suppressor genes characterized to date is that they are deleted at high frequency in certain tumor types. The deletions often involve loss of a single allele, a so-called loss of heterozygosity (LOH), but may also involve homozygous deletion of both alleles. For LOH, the remaining allele is presumed to be nonfunctional, either because of a preexisting inherited mutation, or because of a secondary sporadic mutation.

Breast cancer is one of the most significant diseases that affects women. At the current rate, American women have a 1 in 8 risk of developing breast cancer by age 95 (American Cancer Society, 1992). Treatment of breast cancer at later stages is often futile and disfiguring, making early detection a high priority in medical management of the disease. Ovarian cancer, although less frequent than breast cancer is often rapidly fatal and is the fourth most common cause of cancer mortality in American women. Genetic factors contribute to an ill-defined proportion of breast cancer incidence, estimated to be about 5% of all cases but approximately 25% of cases diagnosed before age 40 (Claus et al., 1991). Breast cancer has been subdivided into two types, early-age onset and late-age onset, based on an inflection in the age-specific incidence curve around age 50. Mutation of one gene, BRCA1, is thought to account for approximately 45% of familial breast cancer, but at least 80% of families with both breast and ovarian cancer (Easton et al., 1993).

Intense efforts to isolate the BRCA I gene have proceeded since it was first mapped in 1990 (Hall et al., 1990; Narod et al., 1991). A second locus, BRCA2, has recently been

mapped to chromosome 13q (Wooster et al., 1994) and appears to account for a proportion of early-onset breast cancer roughly equal to BRCA1, but confers a lower risk of ovarian cancer. The remaining susceptibility to early-onset breast cancer is divided between as yet unmapped genes for familial cancer, and rarer germline mutations in genes such as TP53 (Malkin et al., 1990). It has also been suggested that heterozygote carriers for defective forms of the Ataxia-Telangectasia gene are at higher risk for breast cancer (Swift et al., 1976; Swift et al., 1991). Late-age onset breast cancer is also often familial although the risks in relatives are not as high as those for early-onset breast cancer (Cannon-Albright et al., 1994; Mettlin et al., 1990). However, the percentage of such cases due to genetic susceptibility is unknown.

Breast cancer has long been recognized to be, in part, a familial disease (Anderson, 1972). Numerous investigators have examined the evidence for genetic inheritance and concluded that the data are most consistent with dominant inheritance for a major susceptibility locus or loci (Bishop and Gardner, 1980; Go et al., 1983; Willams and Anderson, 1984; Bishop et al., 1988; Newman et al., 1988; Claus et al., 1991). Recent results demonstrate that at least three loci exist which convey susceptibility to breast cancer as well as other cancers. These loci are the TP53 locus on chromosome 17p (Malkin et al., 1990), a 17q-linked susceptibility locus known as BRCA1 (Hall et al., 1990), and one or more loci responsible for the unmapped residual. Hall et al. (1990) indicated that the inherited breast cancer susceptibility in kindreds with early age onset is linked to chromosome 17q21; although subsequent studies by this group using a more appropriate genetic model partially refuted the limitation to early onset breast cancer (Margaritte et al., 1992).

Most strategies for cloning the 17q-linked breast cancer predisposing gene (BRCA1) require precise genetic localization studies. The simplest model for the functional role of BRCA1 holds that alleles of BRCA1 that predispose to cancer are recessive to wild type alleles; that is, cells that contain at least one wild type BRCA1 allele are not cancerous. However, cells that contain one wild type BRCA1 allele and one predisposing allele may occasionally suffer loss of the wild type allele either by random mutation or by chromosome loss during cell division (nondisjunction). All the progeny of such a mutant cell lack the wild type function of BRCA1 and may develop into tumors. According to this model, predisposing alleles of BRCA1 are recessive, yet susceptibility to cancer is inherited in a dominant fashion: women who possess one predisposing allele (and one wild type allele) risk developing cancer, because their mammary epithelial cells may spontaneously lose the wild type BRCA1 allele. This model applies to a group of cancer susceptibility loci known as tumor suppressors or antioncogenes, a class of genes that includes the retinoblastoma gene and neurofibromatosis gene. By inference this model may also explain the BRCA1 function, as has recently been suggested (Smith et al., 1992).

A second possibility is that BRCA1 predisposing alleles are truly dominant; that is, a wild type allele of BRCA1 cannot overcome the tumor forming role of the predisposing allele. Thus, a cell that carries both wild type and mutant alleles would not necessarily lose the wild type copy of BRCA1 before giving rise to malignant cells. Instead, mammary cells in predisposed individuals would undergo some other stochastic change(s) leading to cancer.

If BRCA1 predisposing alleles are recessive, the BRCA1 gene is expected to be expressed in normal mammary tissue but not functionally expressed in mammary tumors. In contrast, if BRCA1 predisposing alleles are dominant, the wild type BRCA1 gene may or may not be expressed in normal mammary tissue. However, the predisposing allele will likely be expressed in breast tumor cells.

The 17q linkage of BRCA1 was independently confirmed in three of five kindreds with both breast cancer and ovarian cancer (Narod et al., 1991). These studies claimed to localize the gene within a very large region, 15 centiMorgans (cM), or approximately 15 million base pairs, to either side of the linked marker pCMM86 (D17S74). However, attempts to define the region further by genetic studies, using markers surrounding pCMMS6, proved unsuccessful. Subsequent studies indicated that the gene was considerably more proximal (Easton et al., 1993) and that the original analysis was flawed (Margaritte et al., 1992). Hall et al., (1992) recently localized the BRCA1 gene to an approximately 8 cM interval (approximately 8 million base pairs) bounded by Mfd15 (D17S250) on the proximal side and the human GIP gene on the distal side. A slightly narrower interval for the BRCA1 locus, based on publicly available data, was agreed upon at the Chromosome 17 workshop in March of 1992 (Fain, 1992). The size of these regions and the uncertainty associated with them has made it exceedingly difficult to design and implement physical mapping and/or cloning strategies for isolating the BRCA1 gene.

Identification of a breast cancer susceptibility locus would permit the early detection of susceptible individuals and greatly increase our ability to understand the initial steps which lead to cancer. As susceptibility loci are often altered during tumor progression, cloning these genes could also be important in the development of better diagnostic and prognostic products, as well as better cancer therapies.

SUMMARY OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast cancer predisposing gene (BRCA1), some alleles of which cause susceptibility to cancer, in particular breast and ovarian cancer. More specifically, the present invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The invention further relates to somatic mutations in the BRCA1 gene in human breast cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a diagram showing the order of loci neighboring BRCA1 as determined by the chromosome 17 workshop. FIG. 1 is reproduced from Fain, 1992.

FIG. 2 is a schematic map of YACs which define part of Mfd15-Mfd188 region.

FIG. 3 is a schematic map of STSs, P1s and BACs in the BRCA1 region.

FIG. 4 is a schematic map of human chromosome 17. The pertinent region containing BRCA1 is expanded to indicate

the relative positions of two previously identified genes, CA125 and RNU2. BRCA1 spans the marker D17S855.

FIG. **5** shows alignment of the BRCA1 zinc-finger domain with 3 other zinc-finger domains that scored highest in a Smith-Waterman alignment. RPT1 encodes a protein that appears to be a negative regulator of the IL-2 receptor in mouse. RIN 1 encodes a DNA-binding protein that includes a RING-finger motif related to the zinc-finger. RFP1 encodes a putative transcription factor that is the N-terminal domain of the RET oncogene product. The bottom line contains the C3HC4 consensus zinc-finger sequence showing the positions of cysteines and one histidine that form the zinc ion binding pocket.

FIG. **6** is a diagram of BRCA1 mRNA showing the locations of introns and the variants of BRCA1 mRNA produced by alternative splicing. Intron locations are shown by dark triangles and the exons are numbered below the line representing the CDNA. The top cDNA is the composite used to generate the peptide sequence of BRCA1. Alternative forms identified as cDNA clones or hybrid selection clones are shown below.

FIG. **7** shows the tissue expression pattern of BRCA1. The blot was obtained from Clontech and contains RNA from the indicated tissues. Hybridization conditions were as recommended by the manufacturer using a probe consisting of nucleotide positions 3631 to 3930 of BRCA1. Note that both breast and ovary are heterogeneous tissues and the percentage of relevant epithelial cells can be variable. Molecular weight standards are in kilobases.

FIG. **8** is a diagram of the 5' untranslated region plus the beginning of the translated region of BRCA1 showing the locations of introns and the variants of BRCA1 mRNA produced by alternative splicing. Intron locations are shown by broken dashed lines. Six alternate splice forms are shown.

FIG. **9**A shows a nonsense mutation in Kindred 2082. P indicates the person originally screened, b and c are haplotype carriers, a, d, e, f, and g do not carry the BRCA1 haplotype. The C to T mutation results in a stop codon and creates a site for the restriction enzyme AvrII. PCR amplification products are cut with this enzyme. The carriers are heterozygous for the site and therefore show three bands. Non-carriers remain uncut.

FIG. **9**B shows a mutation and cosegregation analysis in BRCA1 kindreds. Carrier individuals are represented as filled circles and squares in the pedigree diagrams. Frameshift mutation in Kindred 1910. The first three lanes are control, noncarrier samples. Lanes labeled 1–3 contain sequences from carrier individuals. Lane 4 contains DNA from a kindred member who does not carry the BRCA1 mutation. The diamond is used to prevent identification of the kindred. The frameshift resulting from the additional C is apparent in lanes labeled 1, 2, and 3.

FIG. **9**C shows a mutation and cosegregation analysis in BRCA1 kindreds. Carrier individuals are represented as filled circles and squares in the pedigree diagrams. Inferred regulatory mutation in Kindred 2035. ASO analysis of carriers and noncarriers of 2 different polymorphisms (PM1 and PM7) which were examined for heterozygosity in the germline and compared to the heterozygosity of lymphocyte mRNA. The top 2 rows of each panel contain PCR products amplified from genomic DNA and the bottom 2 rows contain PCR products amplified from CDNA. "A" and "G" are the two alleles detected by the ASO. The dark spots indicate that a particular allele is present in the sample. The first three lanes of PM7 represent the three genotypes in the general population.

FIGS. **10**A–**10**H show genomic sequence of BRCA1. The lower case letters denote intron sequence while the upper case letters denote exon sequence. Indefinite intervals within introns are designated with vvvvvvvvvvvvv. Known polymorphic sites are shown as underlined and boldface type.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast cancer predisposing gene (BRCA1), some alleles of which cause susceptibility to cancer, in particular breast and ovarian cancer. More specifically, the present invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The invention further relates to somatic mutations in the BRCA1 gene in human breast cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

The present invention provides an isolated polynucleotide comprising all, or a portion of the BRCA1 locus or of a mutated BRCA1 locus, preferably at least eight bases and not more than about 100 kb in length. Such polynucleotides may be antisense polynucleotides. The present invention also provides a recombinant construct comprising such an isolated polynucleotide, for example, a recombinant construct suitable for expression in a transformed host cell.

Also provided by the present invention are methods of detecting a polynucleotide comprising a portion of the BRCA1 locus or its expression product in an analyte. Such methods may further comprise the step of amplifying the portion of the BRCA1 locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the BRCA1 locus. The method is useful for either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer.

The present invention also provides isolated antibodies, preferably monoclonal antibodies, which specifically bind to an isolated polypeptide comprised of at least five amino acid residues encoded by the BRCA1 locus.

The present invention also provides kits for detecting in an analyte a polynucleotide comprising a portion of the BRCA1 locus, the kits comprising a polynucleotide complementary to the portion of the BRCA1 locus packaged in a suitable container, and instructions for its use.

The present invention further provides methods of preparing a polynucleotide comprising polymerizing nucleotides to yield a sequence comprised of at least eight consecutive nucleotides of the BRCA1 locus; and methods of preparing a polypeptide comprising polymerizing amino acids to yield a sequence comprising at least five amino acids encoded within the BRCA1locus.

The present invention further provides methods of screening the BRCA1 gene to identify mutations. Such methods may further comprise the step of amplifying a portion of the BRCA1 locus, and may further include a step of providing

a set of polynucleotides which are primers for amplification of said portion of the BRCA1 locus. The method is useful for identifying mutations for use in either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer.

The present invention further provides methods of screening suspected BRCA1 mutant alleles to identify mutations in the BRCA1 gene.

In addition, the present invention provides methods of screening drugs for cancer therapy to identify suitable drugs for restoring BRCA1 gene product function.

Finally, the present invention provides the means necessary for production of gene-based therapies directed at cancer cells. These therapeutic agents may take the form of polynucleotides comprising all or a portion of the BRCA1 locus placed in appropriate vectors or delivered to target cells in more direct ways such that the function of the BRCA1 protein is reconstituted. Therapeutic agents may also take the form of polypeptides based on either a portion of, or the entire protein sequence of BRCA1. These may functionally replace the activity of BRCA1 in vivo.

It is a discovery of the present invention that the BRCA1locus which predisposes individuals to breast cancer and ovarian cancer, is a gene encoding a BRCA1 protein, which has been found to have no significant homology with known protein or DNA sequences. This gene is termed BRCA1 herein. It is a discovery of the present invention that mutations in the BRCA1 locus in the germline are indicative of a predisposition to breast cancer and ovarian cancer. Finally, it is a discovery of the present invention that somatic mutations in the BRCA1 locus are also associated with breast cancer, ovarian cancer and other cancers, which represents an indicator of these cancers or of the prognosis of these cancers. The mutational events of the BRCA1 locus can involve deletions, insertions and point mutations within the coding sequence and the non-coding sequence.

Starting from a region on the long arm of human chromosome 17 of the human genome, 17q, which has a size estimated at about 8 million base pairs, a region which contains a genetic locus, BRCA1, which causes susceptibility to cancer, including breast and ovarian cancer, has been identified.

The region containing the BRCA1 locus was identified using a variety of genetic techniques. Genetic mapping techniques initially defined the BRCA1 region in terms of recombination with genetic markers. Based upon studies of large extended families ("kindreds") with multiple cases of breast cancer (and ovarian cancer cases in some kindreds), a chromosomal region has been pinpointed that contains the BRCA1 gene as well as other putative susceptibility alleles in the BRCA1 locus. Two meiotic breakpoints have been discovered on the distal side of the BRCA1 locus which are expressed as recombinants between genetic markers and the disease, and one recombinant on the proximal side of the BRCA1 locus. Thus, a region which contains the BRCA1 locus is physically bounded by these markers.

The use of the genetic markers provided by this invention allowed the identification of clones which cover the region from a human yeast artificial chromosome (YAC) or a human bacterial artificial chromosome (BAC) library. It also allowed for the identification and preparation of more easily manipulated cosmid, P1 and BAC clones from this region and the construction of a from a subset of the clones. These cosmids, P1s, YACs and BACs provide the basis for cloning the BRCA1 locus and provide the basis for developing reagents effective, for example, in the diagnosis and treatment of breast and/or ovarian cancer. The BRCA1 gene and other potential susceptibility genes have been isolated from this region. The isolation was done using software trapping (a computational method for identifying sequences likely to contain coding exons, from contiguous or discontinuous genomic DNA sequences), hybrid selection techniques and direct screening, with whole or partial cDNA inserts from cosmids, Pls and BACs, in the region to screen cDNA libraries. These methods were used to obtain sequences of loci expressed in breast and other tissue. These candidate loci were analyzed to identify sequences which confer cancer susceptibility. We have discovered that there are mutations in the coding sequence of the BRCA1 locus in kindreds which are responsible for the 17q-linked cancer susceptibility known as BRCA1. This gene was not known to be in this region. The present invention not only facilitates the early detection of certain cancers, so vital to patient survival, but also permits the detection of susceptible individuals before they develop cancer.

Population Resources

Large, well-documented Utah kindreds are especially important in providing good resources for human genetic studies. Each large kindred independently provides the power to detect whether a BRCA1 susceptibility allele is segregating in that family. Recombinants informative for localization and isolation of the BRCA1 locus could be obtained only from kindreds large enough to confirm the presence of a susceptibility allele. Large sibships are especially important for studying breast cancer, since penetrance of the BRCA1 susceptibility allele is reduced both by age and sex, making informative sibships difficult to find. Furthermore, large sibships are essential for constructing haplotypes of deceased individuals by inference from the haplotypes of their close relatives.

While other populations may also provide beneficial information, such studies generally require much greater effort, and the families are usually much smaller and thus less informative. Utah's age-adjusted breast cancer incidence is 20% lower than the average U.S. rate. The lower incidence in Utah is probably due largely to an early age at first pregnancy, increasing the probability that cases found in Utah kindreds carry a genetic predisposition.

Genetic Mapping

Given a set of informative families, genetic markers are essential for linking a disease to a region of a chromosome. Such markers include restriction fragment length polymorphisms (RFLPs) (Botstein et al., 1980), markers with a variable number of tandem repeats (VNTRs) (Jeffreys et al., 1985; Nakamura et al., 1987), and an abundant class of DNA polymorphisms based on short tandem repeats (STRs), especially repeats of CpA (Weber and May, 1989; Litt et al., 1989). To generate a genetic map, one selects potential genetic markers and tests them using DNA extracted from members of the kindreds being studied.

Genetic markers useful in searching for a genetic locus associated with a disease can be selected on an ad hoc basis, by densely covering a specific chromosome, or by detailed analysis of a specific region of a chromosome. A preferred method for selecting genetic markers linked with a disease involves evaluating the degree of informativeness of kindreds to determine the ideal distance between genetic markers of a given degree of polymorphism, then selecting markers from known genetic maps which are ideally spaced for maximal efficiency. Informativeness of kindreds is measured by the probability that the markers will be heterozygous in unrelated individuals. It is also most efficient to use STR markers which are detected by amplification of the

target nucleic acid sequence using PCR; such markers are highly informative, easy to assay (Weber and May, 1989), and can be assayed simultaneously using multiplexing strategies (Skolnick and Wallace, 1988), greatly reducing the number of experiments required.

Once linkage has been established, one needs to find markers that flank the disease locus, i.e., one or more markers proximal to the disease locus, and one or more markers distal to the disease locus. Where possible, candidate markers can be selected from a known genetic map. Where none is known, new markers can be identified by the STR technique, as shown in the Examples.

Genetic mapping is usually an iterative process. In the present invention, it began by defining flanking genetic markers around the BRCA1 locus, then replacing these flanking markers with other markers that were successively closer to the BRCA1 locus. As an initial step, recombination events, defined by large extended kindreds, helped specifically to localize the BRCA1 locus as either distal or proximal to a specific genetic marker (Goldgar et al., 1994).

The region surrounding BRCA1, until the disclosure of the present invention, was not well mapped and there were few markers. Therefore, short repetitive sequences on cosmids subcloned from YACs, which had been physically mapped, were analyzed in order to develop new genetic markers. Using this approach, one marker of the present invention, 42D6, was discovered which replaced pCMM86 as the distal flanking marker for the BRCA1 region. Since 42D6 is approximately 14 cM from pCMM86, the BRCA1 region was thus reduced by approximately 14 centiMorgans (Easton et al., 1993). The present invention thus began by finding a much more closely linked distal flanking marker of the BRCA1 region. BRCA1 was then discovered to be distal to the genetic marker Mfd15. Therefore, BRCA1 was shown to be in a region of 6 to 10 million bases bounded by Mfd15 and 42D6. Marker Mfd191 was subsequently discovered to be distal to Mfd15 and proximal to BRCA1. Thus, Mfd15 was replaced with Mfd191 as the closest proximal genetic marker. Similarly, it was discovered that genetic marker Mfd188 could replace genetic marker 42D6, narrowing the region containing the BRCA1 locus to approximately 1.5 million bases. Then the marker Mfd191 was replaced with tdj 1474 as the proximal marker and Mfd188 was replaced with U5R as the distal marker, further narrowing the BRCA1 region to a small enough region to allow isolation and characterization of the BRCA1 locus (see FIG. 3), using techniques known in the art and described herein.

Physical Mapping

Three distinct methods were employed to physically map the region. The first was the use of yeast artificial chromosomes (YACs) to clone the region which is flanked by tdj 1474 and U5R. The second was the creation of a set of P1, BAC and cosmid clones which cover the region containing the BRCA1 locus.

Yeast Artificial Chromosomes (YACs).

Once a sufficiently small region containing the BRCA1 locus was identified, physical isolation of the DNA in the region proceeded by identifying a set of overlapping YACs which covers the region. Useful YACs can be isolated from known libraries, such as the St. Louis and CEPH YAC libraries, which are widely distributed and contain approximately 50,000 YACs each. The YACs isolated were from these publicly accessible libraries and can be obtained from a number of sources including the Michigan Genome Center. Clearly, others who had access to these YACs, without the disclosure of the present invention, would not have known the value of the specific YACs we selected since they would not have known which YACs were within, and which YACs outside of, the smallest region containing the BRCA1 locus.

Cosmid, P1 and BAC Clones.

In the present invention, it is advantageous to proceed by obtaining cosmid, P1, and BAC clones to cover this region. The smaller size of these inserts, compared to YAC inserts, makes them more useful as specific hybridization probes. Furthermore, having the cloned DNA in bacterial cells, rather than in yeast cells, greatly increases the ease with which the DNA of interest can be manipulated, and improves the signal-to-noise ratio of hybridization assays. For cosmid subclones of YACs, the DNA is partially digested with the restriction enzyme Sau3A and cloned into the BamHI site of the pWE15 cosmid vector (Stratagene, cat. #1251201). The cosmids containing human sequences are screened by hybridization with human repetitive DNA (e.g., Gibco/BRL, Hfuman C_0t–1 DNA, cat. 5279SA), and then fingerprinted by a variety of techniques, as detailed in the Examples.

P1 and BAC clones are obtained by screening libraries constructed from the total human genome with specific sequence tagged sites (STSs) derived from the YACs, cosmids or P1s and BACs, isolated as described herein.

These P1, BAC and cosmid clones can be compared by interspersed repetitive sequence (IRS) PCR and/or restriction enzyme digests followed by gel electrophoresis and comparison of the resulting DNA fragments ("fingerprints") (Maniatis et al., 1982). The clones can also be characterized by the presence of STSs. The fingerprints are used to define an overlapping contiguous set of clones which covers the region but is not excessively redundant, referred to herein as a "minimum tiling path". Such a minimum tiling path forms the basis for subsequent experiments to identify cDNAs which may originate from the BRCA1 locus.

Coverage of the Gap with P1 and BAC Clones.

To cover any gaps in the BRCA1 contig between the identified cosmids with genomic clones, clones in P1 and BAC vectors which contain inserts of genomic DNA roughly twice as large as cosmids for P1s and still greater for BACs (Stemberg, 1990; Sternberg et al., 1990; Pierce et al., 1992; Shizuya et al., 1992) were used. P1 clones were isolated by Genome Sciences using PCR primers provided by us for screening. BACs were provided by hybridization techniques in Dr. Mel Simon's laboratory. The strategy of using P1 clones also permitted the covering of the genomic region with an independent set of clones not derived from YACs. This guards against the possibility of other deletions in YACs that have not been detected. These new sequences derived from the P1 clones provide the material for further screening for candidate genes, as described below.

Gene Isolation

There are many techniques for testing genomic clones for the presence of sequences likely to be candidates for the coding sequence of a locus one is attempting to isolate, including but not limited to:

a. zoo blots

b. identifying HTF islands

c. exon trapping

d. hybridizing cDNA to cosmids or YACs.

e. screening CDNA libraries.

(a) Zoo blots.

The first technique is to hybridize cosmids to Southern blots to identify DNA sequences which are evolutionarily conserved, and which therefore give positive hybridization signals with DNA from species of varying degrees of

relationship to humans (such as monkey, cow, chicken, pig, mouse and rat). Southern blots containing such DNA from a variety of species are commercially available (Clonetech, Cat. 7753-1).

(b) Identifying HTF islands.

The second technique involves finding regions rich in the nucleotides C and G, which often occur near or within coding sequences. Such sequences are called HTF (HpaI tiny fragment) or CpG islands, as restriction enzymes specific for sites which contain CpG dimers cut frequently in these regions (Lindsay et al, 1987).

(c) Exon trapping.

The third technique is exon trapping, a method that identifies sequences in genomic DNA which contain splice junctions and therefore are likely to comprise coding sequences of genes. Exon amplification (Buckler et al., 1991) is used to select and amplify exons from DNA clones described above. Exon amplification is based on the selection of RNA sequences which are flanked by functional 5' and/or 3' splice sites. The products of the exon amplification are used to screen the breast cDNA libraries to identify a manageable number of candidate genes for further study. Exon trapping can also be performed on small segments of sequenced DNA using computer programs or by software trapping.

(d) Hybridizing cDNA to Cosmids, P1s, BACs or YACs.

The fourth technique is a modification of the selective enrichment technique which utilizes hybridization of cDNA to cosmids, P1s, BACs or YACs and permits transcribed sequences to be identified in, and recovered from cloned genomic DNA (Kandpal et al., 1990). The selective enrichment technique, as modified for the present purpose, involves binding DNA from the region of BRCA1 present in a YAC to a column matrix and selecting cDNAs from the relevant libraries which hybridize with the bound DNA, followed by amplification and purification of the bound DNA, resulting in a great enrichment for cDNAs in the region represented by the cloned genomic DNA.

(e) Identification of cDNAs.

The fifth technique is to identify cDNAs that correspond to the BRCA1 locus. Hybridization probes containing putative coding sequences, selected using any of the above techniques, are used to screen various libraries, including breast tissue CDNA libraries, ovarian CDNA libraries, and any other necessary libraries.

Another variation on the theme of direct selection of CDNA was also used to find candidate genes for BRCA1 (Lovett et al, 1991; Futreal, 1993). This method uses cosmid, P1 or BAC DNA as the probe. The probe DNA is digested with a blunt cutting restriction enzyme such as HaeIII. Double stranded adapters are then ligated onto the DNA and serve as binding sites for primers in subsequent PCR amplification reactions using biotinylated primers. Target CDNA is generated from mRNA derived from tissue samples, e.g., breast tissue, by synthesis of either random primed or oligo(dT) primed first strand followed by second strand synthesis. The CDNA ends are rendered blunt and ligated onto double-stranded adapters. These adapters serve as amplification sites for PCR. The target and probe sequences are denatured and mixed with human C_0t–1 DNA to block repetitive sequences. Solution hybridization is carried out to high C_0t–½ values to ensure hybridization of rare target cDNA molecules. The annealed material is then captured on avidin beads, washed at high stringency and the retained cDNAs are eluted and amplified by PCR. The selected cDNA is subjected to further rounds of enrichment before cloning into a plasmid vector for analysis.

Testing the cDNA for Candidacy

Proof that the cDNA is the BRCA1 locus is obtained by finding sequences in DNA extracted from affected kindred members which create abnormal BRCA1 gene products or abnormal levels of BRCA1 gene product. Such BRCA1 susceptibility alleles will co-segregate with the disease in large kindreds. They will also be present at a much higher frequency in non-kindred individuals with breast and ovarian cancer then in individuals in the general population. Finally, since tumors often mutate somatically at loci which are in other instances mutated in the germline, we expect to see normal germline BRCA1 alleles mutated into sequences which are identical or similar to BRCA1 susceptibility alleles in DNA extracted from tumor tissue. Whether one is comparing BRCA1 sequences from tumor tissue to BRCA1 alleles from the germline of the same individuals, or one is comparing germline BRCA1 alleles from cancer cases to those from unaffected individuals, the key is to find mutations which are serious enough to cause obvious disruption to the normal function of the gene product. These mutations can take a number of forms. The most severe forms would be frame shift mutations or large deletions which would cause the gene to code for an abnormal protein or one which would significantly alter protein expression. Less severe disruptive mutations would include small in-frame deletions and nonconservative base pair substitutions which would have a significant effect on the protein produced, such as changes to or from a cysteine residue, from a basic to an acidic amino acid or vice versa, from a hydrophobic to hydrophilic amino acid or vice versa, or other mutations which would affect secondary, tertiary or quaternary protein structure. Silent mutations or those resulting in conservative amino acid substitutions would not generally be expected to disrupt protein function.

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type BRCA1 locus is detected. In addition, the method can be performed by detecting the wild-type BRCA1 locus and confirming the lack of a predisposition to cancer at the BRCA1 locus. "Alteration of a wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or of only a portion of the gene. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those which occur only in certain tissues, e.g., in the tumor tissue, and are not inherited in the germline. Geimline mutations can be found in any of a body's tissues and are inherited. If only a single allele is somatically mutated, an early neoplastic state is indicated. However, if both alleles are somatically mutated, then a late neoplastic state is indicated. The finding of BRCA1 mutations thus provides both diagnostic and prognostic information. A BRCA1 allele which is not deleted (e.g., found on the sister chromosome to a chromosome carrying a BRCA1 deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. It is believed that many mutations found in tumor tissues will be those leading to decreased expression of the BRCA1 gene product. However, mutations leading to nonfunctional gene products would also lead to a cancerous state. Point mutational events may occur in regulatory regions, such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the BRCA1 gene product, or to a decrease in mRNA stability or translation efficiency.

Useful diagnostic techniques include, but are not limited to fluorescent in situ hybridization (FISH), direct DNA

sequencing, PFGE analysis, Southern blot analysis, single stranded conformation analysis (SSCA), RNase protection assay, allele-specific oligonucleotide (ASO), dot blot analysis and PCR-SSCP, as discussed in detail further below.

Predisposition to cancers, such as breast and ovarian cancer, and the other cancers identified herein, can be ascertained by testing any tissue of a human for mutations of the BRCA1 gene. For example, a person who has inherited a germline BRCA1 mutation would be prone to develop cancers. This can be determined by testing DNA from any tissue of the person's body. Most simply, blood can be drawn and DNA extracted from the cells of the blood. In addition, prenatal diagnosis can be accomplished by testing fetal cells, placental cells or amniotic cells for mutations of the BRCA1 gene. Alteration of a wild-type BRCA1 allele, whether, for example, by point mutation or deletion, can be detected by any of the means discussed herein.

There are several methods that can be used to detect DNA sequence variation. Direct DNA sequencing, either manual sequencing or automated fluorescent sequencing can detect sequence variation. For a gene as large as BRCA1, manual sequencing is very labor-intensive, but under optimal conditions, mutations in the coding sequence of a gene are rarely missed. Another approach is the single-stranded conformation polymorphism assay (SSCA) (Orita et al., 1989). This method does not detect all sequence changes, especially if the DNA fragment size is greater than 200 bp, but can be optimized to detect most DNA sequence variation. The reduced detection sensitivity is a disadvantage, but the increased throughput possible with SSCA makes it an attractive, viable alternative to direct sequencing for mutation detection on a research basis. The fragments which have shifted mobility on SSCA gels are then sequenced to determine the exact nature of the DNA sequence variation. Other approaches based on the detection of mismatches between the two complementary DNA strands include clamped denaturing gel electrophoresis (CDGE) (Sheffield et al., 1991), heteroduplex analysis (HA) (White et al., 1992) and chemical mismatch cleavage (CMC) (Grompe et al., 1989). None of the methods described above will detect large deletions, duplications or insertions, nor will they detect a regulatory mutation which affects transcription or translation of the protein. Other methods which might detect these classes of mutations such as a protein truncation assay or the asymmetric assay, detect only specific types of mutations and would not detect missense mutations. A review of currently available methods of detecting DNA sequence variation can be found in a recent review by Grompe (1993). Once a mutation is known, an allele specific detection approach such as allele specific oligonucleotide (ASO) hybridization can be utilized to rapidly screen large numbers of other samples for that same mutation.

In order to detect the alteration of the wild-type BRCA1 gene in a tissue, it is helpful to isolate the tissue free from surrounding normal tissues. Means for enriching tissue preparation for tumor cells are known in the art. For example, the tissue may be isolated from paraffin or cryostat sections. Cancer cells may also be separated from normal cells by flow cytometry. These techniques, as well as other techniques for separating tumor cells from normal cells, are well known in the art. If the tumor tissue is highly contaminated with normal cells, detection of mutations is more difficult.

A rapid preliminary analysis to detect polymorphisms in DNA sequences can be performed by looking at a series of Southern blots of DNA cut with one or more restriction enzymes, preferably with a large number of restriction enzymes. Each blot contains a series of normal individuals and a series of cancer cases, tumors, or both. Southern blots displaying hybridizing fragments (differing in length from control DNA when probed with sequences near or including the BRCA1 locus) indicate a possible mutation. If restriction enzymes which produce very large restriction fragments are used, then pulsed field gel electrophoresis (PFGE) is employed.

Detection of point mutations may be accomplished by molecular cloning of the BRCA1 allele(s) and sequencing the allele(s) using techniques well known in the art. Alternatively, the gene sequences can be amplified directly from a genomic DNA preparation from the tumor tissue, using known techniques. The DNA sequence of the amplified sequences can then be determined.

There are six well known methods for a more complete, yet still indirect, test for confirming the presence of a susceptibility allele: 1) single stranded conformation analysis (SSCA) (Orita et al., 1989); 2) denaturing gradient gel electrophoresis (DGGE) (Wartell et al., 1990; Sheffield et al., 1989); 3) RNase protection assays (Finkelstein et al., 1990; Kinszler el aL, 1991); 4) allele-specific oligonucleotides (ASOs) (Conner et al., 1983); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. Coli* mutS protein (Modrich, 1991); and 6) allele-specific PCR (Rano & Kidd, 1989). For allele-specific PCR, primers are used which hybridize at their 3' ends to a particular BRCA1 mutation. If the particular BRCA1 mutation is not present, an amplification product is not observed. Amplification Refractory Mutation System (ARMS) can also be used, as disclosed in European Patent Application Publication No. 0332435 and in Newton et al., 1989. Insertions and deletions of genes can also be detected by cloning, sequencing and amplification. In addition, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to score alteration of an allele or an insertion in a polymorphic fragment. Such a method is particularly useful for screening relatives of an affected individual for the presence of the BRCA1 mutation found in that individual. Other techniques for detecting insertions and deletions as known in the art can be used.

In the first three methods (SSCA, DGGE and RNase protection assay), a new electrophoretic band appears. SSCA detects a band which migrates differentially because the sequence change causes a difference in single-strand, intramolecular base pairing. RNase protection involves cleavage of the mutant polynucleotide into two or more smaller fragments. DGGE detects differences in migration rates of mutant sequences compared to wild-type sequences, using a denaturing gradient gel. In an allele-specific oligonucleotide assay, an oligonucleotide is designed which detects a specific sequence, and the assay is performed by detecting the presence or absence of a hybridization signal. In the mutS assay, the protein binds only to sequences that contain a nucleotide mismatch in a heteroduplex between mutant and wild-type sequences.

Mismatches, according to the present invention, are hybridized nucleic acid duplexes in which the two strands are not 100% complementary. Lack of total homology may be due to deletions, insertions, inversions or substitutions. Mismatch detection can be used to detect point mutations in the gene or in its mRNA product. While these techniques are less sensitive than sequencing, they are simpler to perform on a large number of tumor samples. An example of a mismatch cleavage technique is the RNase protection method. In the practice of the present invention, the method involves the use of a labeled riboprobe which is comple-

mentary to the human wild-type BRCA1 gene coding sequence. The riboprobe and either mRNA or DNA isolated from the tumor tissue are annealed (hybridized) together and subsequently digested with the enzyme RNase A which is able to detect some mismatches in a duplex RNA structure. If a mismatch is detected by RNase A, it cleaves at the site of the mismatch. Thus, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch has been detected and cleaved by RNase A, an RNA product will be seen which is smaller than the full length duplex RNA for the riboprobe and the mRNA or DNA. The riboprobe need not be the full length of the BRCA1 mRNA or gene but can be a segment of either. If the riboprobe comprises only a segment of the BRCA1 mRNA or gene, it will be desirable to use a number of these probes to screen the whole mRNA sequence for mismatches.

In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. See, e.g., Cotton et al., 1988; Shenk et al., 1975; Novack et al., 1986. Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. See, e.g., Cariello, 1988. With either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR (see below) before hybridization. Changes in DNA of the BRCA1 gene can also be detected using Southern hybridization, especially if the changes are gross rearrangements, such as deletions and insertions.

DNA sequences of the BRCA1 gene which have been amplified by screened using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the BRCA1 gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length, corresponding to a portion of the BRCA1 gene sequence. By use of a battery of such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously identified mutation in the BRCA1 gene. Hybridization of allele-specific probes with amplified BRCA1 sequences can be performed, for example, on a nylon filter. Hybridization to a particular probe under stringent hybridization conditions indicates the presence of the same mutation in the tumor tissue as in the allele-specific probe.

The most definitive test for mutations in a candidate locus is to directly compare genomic BRCA1 sequences from cancer patients with those from a control population. Alternatively, one could sequence messenger RNA after amplification, e.g., by PCR, thereby eliminating the necessity of determining the exon structure of the candidate gene.

Mutations from cancer patients falling outside the coding region of BRCA1 can be detected by examining the non-coding regions, such as introns and regulatory sequences near or within the BRCA1 gene. An early indication that mutations in noncoding regions are important may come from Northern blot experiments that reveal messenger RNA molecules of abnormal size or abundance in cancer patients as compared to control individuals.

Alteration of BRCA1 mRNA expression can be detected by any techniques known in the art. These include Northern blot analysis, PCR amplification and RNase protection. Diminished mRNA expression indicates an alteration of the wild-type BRCA1 gene. Alteration of wild-type BRCA1 genes can also be detected by screening for alteration of wild-type BRCA1 protein. For example, monoclonal antibodies immunoreactive with BRCA1 can be used to screen a tissue. Lack of cognate antigen would indicate a BRCA1 mutation. Antibodies specific for products of mutant alleles could also be used to detect mutant BRCA1 gene product. Such immunological assays can be done in any convenient formats known in the art. These include Western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered BRCA1 protein can be used to detect alteration of wild-type BRCA1 genes. Functional assays, such as protein binding determinations, can be used. In addition, assays can be used which detect BRCA1 biochemical function. Finding a mutant BRCA1 gene product indicates alteration of a wild-type BRCA1 gene.

Mutant BRCA1 genes or gene products can also be detected in other human body samples, such as serum, stool, urine and sputum. The same techniques discussed above for detection of mutant BRCA1 genes or gene products in tissues can be applied to other body samples. Cancer cells are sloughed off from tumors and appear in such body samples. In addition, the BRCA1 gene product itself may be secreted into the extracellular space and found in these body samples even in the absence of cancer cells. By screening such body samples, a simple early diagnosis can be achieved for many types of cancers. In addition, the progress of chemotherapy or radiotherapy can be monitored more easily by testing such body samples for mutant BRCA1 genes or gene products.

The methods of diagnosis of the present invention are applicable to any tumor in which BRCA1 has a role in tumorigenesis. The diagnostic method of the present invention is useful for clinicians, so they can decide upon an appropriate course of treatment.

The primer pairs of the present invention are useful for determination of the nucleotide sequence of a particular BRCA1 allele using PCR. The pairs of single-stranded DNA primers can be annealed to sequences within or surrounding the BRCA1 gene on chromosome 17q21 in order to prime amplifying DNA synthesis of the BRCA1 gene itself. A complete set of these primers allows synthesis of all of the nucleotides of the BRCA1 gene coding sequences, i.e., the exons. The set of primers preferably allows synthesis of both intron and exon sequences. Allele-specific primers can also be used. Such primers anneal only to particular BRCA1 mutant alleles, and thus will only amplify a product in the presence of the mutant allele as a template.

In order to facilitate subsequent cloning of amplified sequences, primers may have restriction enzyme site sequences appended to their 5' ends. Thus, all nucleotides of the primers are derived from BRCA1 sequences or sequences adjacent to BRCA1, except for the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art. The primers themselves can be synthesized using techniques which are well known in the art. Generally, the primers can be made using oligonucleotide synthesizing machines which are commercially available. Given the sequence of the BRCA1 open reading frame shown in SEQ ID NO:1, design of particular primers is well within the skill of the art.

The nucleic acid probes provided by the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA and in the RNase protection method for detecting point mutations already discussed above. The probes can be used to detect PCR amplification products. They may also be used to detect mismatches with the BRCA1 gene or mRNA using other techniques.

It has been discovered that individuals with the wild-type BRCA1 gene do not have cancer which results from the BRCA1 allele. However, mutations which interfere with the function of the BRCA1 protein are involved in the patho-

genesis of cancer. Thus, the presence of an altered (or a mutant) BRCA1 gene which produces a protein having a loss of function, or altered function, directly correlates to an increased risk of cancer. In order to detect a BRCA1 gene mutation, a biological sample is prepared and analyzed for a difference between the sequence of the BRCA1 allele being analyzed and the sequence of the wild-type BRCA1 allele. Mutant BRCA1 alleles can be initially identified by any of the techniques described above. The mutant alleles are then sequenced to identify the specific mutation of the particular mutant allele. Alternatively, mutant BRCA1 alleles can be initially identified by identifying mutant (altered) BRCA1 proteins, using conventional techniques. The mutant alleles are then sequenced to identify the specific mutation for each allele. The mutations, especially those which lead to an altered function of the BRCA1 protein, are then used for the diagnostic and prognostic methods of the present invention.

Definitions

The present invention employs the following definitions:

"Amplification of Polynucleotides" utilizes methods such as the polymerase chain reaction (PCR), ligation amplification (or ligase chain reaction, LCR) and amplification methods based on the use of Q-beta replicase. These methods are well known and widely practiced in the art. See, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202 and Innis et al., 1990 (for PCR); and Wu et al., 1989a (for LCR). Reagents and hardware for conducting PCR are commercially available. Primers useful to amplify sequences from the BRCA1 region are preferably complementary to, and hybridize specifically to sequences in the BRCA1 region or in regions that flank a target region therein. BRCA1 sequences generated by amplification may be sequenced directly. Alternatively, but less desirably, the amplified sequence(s) may be cloned prior to sequence analysis. A method for the direct cloning and sequence analysis of enzymatically amplified genomic segments has been described by Scharf, 1986.

"Analyte polynucleotide" and "analyte strand" refer to a single- or double-stranded polynucleotide which is suspected of containing a target sequence, and which may be present in a variety of types of samples, including biological samples.

"Antibodies." The present invention also provides polyclonal and/or monoclonal antibodies and fragments thereof, and immunologic binding equivalents thereof, which are capable of specifically binding to the BRCA1 polypeptides and fragments thereof or to polynucleotide sequences from the BRCA1 region, particularly from the BRCA1 locus or a portion thereof. The term "antibody" is used both to refer to a homogeneous molecular entity, or a mixture such as a serum product made up of a plurality of different molecular entities. Polypeptides may be prepared synthetically in a peptide synthesizer and coupled to a carrier molecule (e.g., keyhole limpet hemocyanin) and injected over several months into rabbits. Rabbit sera is tested for immunoreactivity to the BRCA1 polypeptide or fragment. Monoclonal antibodies may be made by injecting mice with the protein polypeptides, fusion proteins or fragments thereof. Monoclonal antibodies will be screened by ELISA and tested for specific immunoreactivity with BRCA1 polypeptide or fragments thereof. See, Harlow & Lane, 1988. These antibodies will be useful in assays as well as pharmaceuticals.

Once a sufficient quantity of desired polypeptide has been obtained, it may be used for various purposes. A typical use is the production of antibodies specific for binding. These antibodies may be either polyclonal or monoclonal, and may be produced by in vitro or in vivo techniques well known in the art. For production of polyclonal antibodies, an appropriate target immune system, typically mouse or rabbit, is selected. Substantially purified antigen is presented to the immune system in a fashion determined by methods appropriate for the animal and by other parameters well known to immunologists. Typical sites for injection are in footpads, intramuscularly, intraperitoneally, or intradermally. Of course, other species may be substituted for mouse or rabbit. Polyclonal antibodies are then purified using techniques known in the art, adjusted for the desired specificity.

An immunological response is usually assayed with an immunoassay. Normally, such immunoassays involve some purification of a source of antigen, for example, that produced by the same cells and in the same fashion as the antigen. A variety of immunoassay methods are well known in the art. See, e.g., Harlow & Lane, 1988, or Goding, 1986.

Monoclonal antibodies with affinities of 10^{-8} M^{-1} or preferably 10^{-9} to 10^{-10} to M^{-1} or stronger will typically be made by standard procedures as described, e.g., in Harlow & Lane, 1988 or Goding, 1986. Briefly, appropriate animals will be selected and the desired immunization protocol followed. After the appropriate period of time, the spleens of such animals are excised and individual spleen cells fused, typically, to immortalized myeloma cells under appropriate selection conditions. Thereafter, the cells are clonally separated and the supernatants of each clone tested for their production of an appropriate antibody specific for the desired region of the antigen.

Other suitable techniques involve in vitro exposure of lymphocytes to the antigenic polypeptides, or alternatively, to selection of libraries of antibodies in phage or similar vectors. See Huse et al., 1989. The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, magnetic particles and the like. Patents teaching the use of such labels include U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced (see U.S. Pat. No. 4,816,567).

"Binding partner" refers to a molecule capable of binding a ligand molecule with high specificity, as for example, an antigen and an antigen-specific antibody or an enzyme and its inhibitor. In general, the specific binding partners must bind with sufficient affinity to immobilize the analyte copy/complementary strand duplex (in the case of polynucleotide hybridization) under the isolation conditions. Specific binding partners are known in the art and include, for example, biotin and avidin or streptavidin, IgG and protein A, the numerous, known receptor-ligand couples, and complementary polynucleotide strands. In the case of complementary polynucleotide binding partners, the partners are normally at least about 15 bases in length, and may be at least 40 bases in length. The polynucleotides may be composed of DNA, RNA, or synthetic nucleotide analogs.

A "biological sample" refers to a sample of tissue or fluid suspected of containing an analyte polynucleotide or polypeptide from an individual including, but not limited to, e.g., plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, blood cells, tumors, organs, tissue and samples of in vitro cell culture constituents.

As used herein, the terms "diagnosing" or "prognosing," as used in the context of neoplasia, are used to indicate 1) the classification of lesions as neoplasia, 2) the determination of the severity of the neoplasia, or 3) the monitoring of the disease progression, prior to, during and after treatment.

"Encode". A polynucleotide is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the mRNA for and/or the polypeptide or a fragment thereof. The anti-sense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

"Isolated" or "substantially pure". An "isolated" or "substantially pure" nucleic acid (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components which naturally accompany a native human sequence or protein, e.g., ribosomes, polymerases, many other human genome sequences and proteins. The term embraces a nucleic acid sequence or protein which has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogs or analogs biologically synthesized by heterologous systems.

"BRCA1 Allele" refers to normal alleles of the BRCA1 locus as well as alleles carrying variations that predispose individuals to develop cancer of many sites including, for example, breast, ovarian, colorectal and prostate cancer. Such predisposing alleles are also called "BRCA1 susceptibility alleles".

"BRCA1 Locus," "BRCA1 Gene," "BRCA1 Nucleic Acids" or "BRCA1 Polynucleotide" each refer to polynucleotides, all of which are in the BRCA1 region, that are likely to be expressed in normal tissue, certain alleles of which predispose an individual to develop breast, ovarian, colorectal and prostate cancers. Mutations at the BRCA1 locus may be involved in the initiation and/or progression of other types of tumors. The locus is indicated in part by mutations that predispose individuals to develop cancer. These mutations fall within the BRCA1 region described infra. The BRCA1 locus is intended to coding sequences, intervening sequences and regulatory elements controlling transcription and/or translation. The BRCA1 locus is intended to include all allelic variations of the DNA sequence.

These terms, when applied to a nucleic acid, refer to a nucleic acid which encodes a BRCA1 polypeptide, fragment, homolog or variant, including, e.g., protein fusions or deletions. The nucleic acids of the present invention will possess a sequence which is either derived from, or substantially similar to a natural BRCA1-encoding gene or one having substantial homology with a natural BRCA1-encoding gene or a portion thereof. The coding sequence for a BRCA1 polypeptide is shown in SEQ ID NO:1, with the amino acid sequence shown in SEQ ID NO:2.

The polynucleotide compositions of this invention include RNA, cDNA, genomic DNA, synthetic forms, and mixed polymers, both sense and antisense strands, and may be chemically or biochemically modified or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those skilled in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.), charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids, etc.). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The present invention provides recombinant nucleic acids comprising all or part of the BRCA1 region. The recombinant construct may be capable of replicating autonomously in a host cell. Alternatively, the recombinant construct may become integrated into the chromosomal DNA of the host cell. Such a recombinant polynucleotide comprises a polynucleotide of genomic, cDNA, semi-synthetic, or synthetic origin which, by virtue of its origin or manipulation, 1) is not associated with all or a portion of a polynucleotide with which it is associated in nature; 2) is linked to a polynucleotide other than that to which it is linked in nature; or 3) does not occur in nature.

Therefore, recombinant nucleic acids comprising sequences otherwise not naturally occurring are provided by this invention. Although the wild-type sequence may be employed, it will often be altered, e.g., by deletion, substitution or insertion.

cDNA or genomic libraries of various types may be screened as natural sources of the nucleic acids of the present invention, or such nucleic acids may be provided by amplification of sequences resident in genomic DNA or other natural sources, e.g., by PCR. The choice of cDNA libraries normally corresponds to a tissue source which is abundant in mRNA for the desired proteins. Phage libraries are normally preferred, but other types of libraries may be used. Clones of a library are spread onto plates, transferred to a substrate for screening, denatured and probed for the presence of desired sequences.

The DNA sequences used in this invention will usually comprise at least about five codons (15 nucleotides), more usually at least about 7–15 codons, and most preferably, at least about 35 codons. One or more introns may also be present. This number of nucleotides is usually about the minimal length required for a successful probe that would hybridize specifically with a BRCA1-encoding sequence.

Techniques for nucleic acid manipulation are described generally, for example, in Sambrook et al., 1989 or Ausubel et al., 1992. Reagents useful in applying such techniques, such as restriction enzymes and the like, are widely known in the art and commercially available from such vendors as New England BioLabs, Boehringer Mannheim, Amersham, Promega Biotec, U. S. Biochemicals, New England Nuclear, and a number of other sources. The recombinant nucleic acid sequences used to produce fusion proteins of the present invention may be derived from natural or synthetic sequences. Many natural gene sequences are obtainable from various cDNA or from genomic libraries using appropriate probes. See, GenBank, National Institutes of Health.

"BRCA1 Region" refers to a portion of human chromosome 17q21 bounded by the markers tdj1474 and U5R. This region contains the BRCA1 locus, including the BRCA1 gene.

As used herein, the terms "BRCA1 locus," "BRCA1 allele" and "BRCA1 region" all refer to the double-stranded DNA comprising the locus, allele, or region, as well as either of the single-stranded DNAs comprising the locus, allele or region.

As used herein, a "portion" of the BRCA1 locus or region or allele is defined as having a minimal size of at least about

eight nucleotides, or preferably about 15 nucleotides, or more preferably at least about 25 nucleotides, and may have a minimal size of at least about 40 nucleotides.

"BRCA1 protein" or "BRCA1 polypeptide" refer to a protein or polypeptide encoded by the BRCA1 locus, variants or fragments thereof. The term "polypeptide" refers to a polymer of amino acids and its equivalent and does not refer to a specific length of the product; thus, peptides, oligopeptides and proteins are included within the definition of a polypeptide. This term also does not refer to, or exclude modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations, and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages as well as other modifications known in the art, both naturally and non-naturally occurring. Ordinarily, such polypeptides will be at least about 50% homologous to the native BRCA1 sequence, preferably in excess of about 90%, and more preferably at least about 95% homologous. Also included are proteins encoded by DNA which hybridize under high or low stringency conditions, to BRCA1-encoding nucleic acids and closely related polypeptides or proteins retrieved by antisera to the BRCA1 protein (s).

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acids, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression.

"Probes". Polynucleotide polymorphisms associated with BRCA1 alleles which predispose to certain cancers or are associated with most cancers are detected by hybridization with a polynucleotide probe which forms a stable hybrid with that of the target sequence, under stringent to moderately stringent hybridization and wash conditions. If it is expected that the probes will be perfectly complementary to the target sequence, stringent conditions will be used. Hybridization stringency may be lessened if some mismatching is expected, for example, if variants are expected with the result that the probe will not be completely complementary. Conditions are chosen which rule out nonspecific/adventitious bindings, that is, which minimize noise. Since such indications identify neutral DNA polymorphisms as well as mutations, these indications need further analysis to demonstrate detection of a BRCA1 susceptibility allele.

Probes for BRCA1 alleles may be derived from the sequences of the BRCA1 region or its cDNAs. The probes may be of any suitable length, which span all or a portion of the BRCA1 region, and which allow specific hybridization to the BRCA1 region. If the target sequence contains a sequence identical to that of the probe, the probes may be short, e.g., in the range of about 8–30 base pairs, since the hybrid will be relatively stable under even stringent conditions. If some degree of mismatch is expected with the probe, i.e., if it is suspected that the probe will hybridize to a variant region, a longer probe may be employed which hybridizes to the target sequence with the requisite specificity.

The probes will include an isolated polynucleotide attached to a label or reporter molecule and may be used to isolate other polynucleotide sequences, having sequence similarity by standard methods. For techniques for preparing and labeling probes see, e.g., Sambrook et al., 1989 or Ausubel et al., 1992. Other similar polynucleotides may be selected by using homologous polynucleotides. Alternatively, polynucleotides encoding these or similar polypeptides may be synthesized or selected by use of the redundancy in the genetic code. Various codon substitutions may be introduced, e.g., by silent changes (thereby producing various restriction sites) or to optimize expression for a particular system. Mutations may be introduced to modify the properties of the polypeptide, perhaps to change ligand-binding affinities, interchain affinities, or the polypeptide degradation or turnover rate.

Probes comprising synthetic oligonucleotides or other polynucleotides of the present invention may be derived from naturally occurring or recombinant single- or double-stranded polynucleotides, or be chemically synthesized. Probes may also be labeled by nick translation, Klenow fill-in reaction, or other methods known in the art.

Portions of the polynucleotide sequence having at least about eight nucleotides, usually at least about 15 nucleotides, and fewer than about 6 kb, usually fewer than about 1.0 kb, from a polynucleotide sequence encoding BRCA1 are preferred as probes. The probes may also be used to determine whether mRNA encoding BRCA1 is present in a cell or tissue.

"Protein modifications or fragments" are provided by the present invention for BRCA1 polypeptides or fragments thereof which are substantially homologous to primary structural sequence but which include, e.g., in vivo or in vitro chemical and biochemical modifications or which incorporate unusual amino acids. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, e.g., with radionuclides, and various enzymatic modifications, as will be readily appreciated by those well skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{32}P, ligands which bind to labeled antiligands (e.g., antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods of labeling polypeptides are well known in the art. See, e.g., Sambrook et al., 1989 or Ausubel et al., 1992.

Besides substantially full-length polypeptides, the present invention provides for biologically active fragments of the polypeptides. Significant biological activities include ligand-binding, immunological activity and other biological activities characteristic of BRCA1 polypeptides. Immunological activities include both immunogenic function in a target immune system, as well as sharing of immunological epitopes for binding, serving as either a competitor or substitute antigen for an epitope of the BRCA1 protein. As used herein, "epitope" refers to an antigenic determinant of a polypeptide. An epitope could comprise three amino acids in a spatial conformation which is unique to the epitope. Generally, an epitope consists of at least five such amino acids, and more usually consists of at least 8–10 such amino acids. Methods of determining the spatial conformation of such amino acids are known in the art.

For immunological purposes, tandem-repeat polypeptide segments may be used as immunogens, thereby producing highly antigenic proteins. Alternatively, such polypeptides

will serve as highly efficient competitors for specific binding. Production of antibodies specific for BRCA1 polypeptides or fragments thereof is described below.

The present invention also provides for fusion polypeptides, comprising BRCA1 polypeptides and fragments. Homologous polypeptides may be fusions between two or more BRCA1 polypeptide sequences or between the sequences of BRCA1 and a related protein. Likewise, heterologous fusions may be constructed which would exhibit a combination of properties or activities of the derivative proteins. For example, ligand-binding or other domains may be "swapped" between different new fusion polypeptides or fragments. Such homologous or heterologous fusion polypeptides may display, for example, altered strength or specificity of binding. Fusion partners include immunoglobulins, bacterial β-galactosidase, trpE, protein A, β-lactamase, alpha amylase, alcohol dehydrogenase and yeast alpha mating factor. See, e.g., Godowski et al., 1988.

Fusion proteins will typically be made by either recombinant nucleic acid methods, as described below, or may be chemically synthesized. Techniques for the synthesis of polypeptides are described, for example, in Merrifield, 1963.

"Protein purification" refers to various methods for the isolation of the BRCA1 polypeptides from other biological material, such as from cells transformed with recombinant nucleic acids encoding BRCA1, and are well known in the art. For example, such polypeptides may be purified by immunoaffinity chromatography employing, e.g., the antibodies provided by the present invention. Various methods of protein purification are well known in the art, and include those described in Deutscher, 1990 and Scopes, 1982.

The terms "isolated", "substantially pure", and "substantially homogeneous" are used interchangeably to describe a protein or polypeptide which has been separated from components which accompany it in its natural state. A monomeric protein is substantially pure when at least about 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure protein will typically comprise about 60 to 90% W/W of a protein sample, more usually about 95%, and preferably will be over about 99% pure. Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art which are utilized for purification.

A BRCA1 protein is substantially free of naturally associated components when it is separated from the native contaminants which accompany it in its natural state. Thus, a polypeptide which is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be substantially free from its naturally associated components. A protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art.

A polypeptide produced as an expression product of an isolated and manipulated genetic sequence is an "isolated polypeptide," as used herein, even if expressed in a homologous cell type. Synthetically made forms or molecules expressed by heterologous cells are inherently isolated molecules.

"Recombinant nucleic acid" is a nucleic acid which is not naturally occurring, or which is made by the artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. Such is usually done to replace a codon with a redundant codon encoding the same or a conservative amino acid, while typically introducing or removing a sequence recognition site. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions.

"Regulatory sequences" refers to those sequences normally within 100 kb of the coding region of a locus, but they may also be more distant from the coding region, which affect the expression of the gene (including transcription of the gene, and translation, splicing, stability or the like of the messenger RNA).

"Substantial homology or similarity". A nucleic acid or fragment thereof is "substantially homologous" ("or substantially similar") to another if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95–98% of the nucleotide bases.

Alternatively, substantial homology or (similarity) exists when a nucleic acid or fragment thereof will hybridize to another nucleic acid (or a complementary strand thereof) under selective hybridization conditions, to a strand, or to its complement. Selectivity of hybridization exists when hybridization which is substantially more selective than total lack of specificity occurs. Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 14 nucleotides, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90%. See, Kanehisa, 1984. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will often be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions will generally include temperatures in excess of 30° C., typically in excess of 37° C., and preferably in excess of 45° C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Wetmur & Davidson, 1968.

Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

The terms "substantial homology" or "substantial identity", when referring to polypeptides, indicate that the polypeptide or protein in question exhibits at least about 30% identity with an entire naturally-occurring protein or a portion thereof, usually at least about 70% identity, and preferably at least about 95% identity.

"Substantially similar function" refers to the function of a modified nucleic acid or a modified protein, with reference

to the wild-type BRCA1 nucleic acid or wild-type BRCA1 polypeptide. The modified polypeptide will be substantially homologous to the wild-type BRCA1 polypeptide and will have substantially the same function. The modified polypeptide may have an altered amino acid sequence and/or may contain modified amino acids. In addition to the similarity of function, the modified polypeptide may have other useful properties, such as a longer half-life. The similarity of function (activity) of the modified polypeptide may be substantially the same as the activity of the wild-type BRCA1 polypeptide. Alternatively, the similarity of function (activity) of the modified polypeptide may be higher than the activity of the wild-type BRCA1 polypeptide. The modified polypeptide is synthesized using conventional techniques, or is encoded by a modified nucleic acid and produced using conventional techniques. The modified nucleic acid is prepared by conventional techniques. A nucleic acid with a function substantially similar to the wild-type BRCA1 gene function produces the modified protein described above.

Homology, for polypeptides, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wis. 53705. Protein analysis software matches similar sequences using measure of homology assigned to various substitutions, deletions and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

A polypeptide "fragment," "portion" or "segment" is a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids.

The polypeptides of the present invention, if soluble, may be coupled to a solid-phase support, e.g., nitrocellulose, nylon, column packing materials (e.g., Sepharose beads), magnetic beads, glass wool, plastic, metal, polymer gels, cells, or other substrates. Such supports may take the form, for example, of beads, wells, dipsticks, or membranes.

"Target region" refers to a region of the nucleic acid which is amplified and/or detected. The term "target sequence" refers to a sequence with which a probe or primer will form a stable hybrid under desired conditions.

The practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, genetics, and immunology. See, e.g., Maniatis et al., 1982; Sambrook et al., 1989; Ausubel et al., 1992; Glover, 1985; Anand, 1992; Guthrie & Fink, 1991. A general discussion of techniques and materials for human gene mapping, including mapping of human chromosome 17q, is provided, e.g., in White and Lalouel, 1988.

Preparation of recombinant or chemically synthesized nucleic acids; vectors, transformation, host cells

Large amounts of the polynucleotides of the present invention may be produced by replication in a suitable host cell. Natural or synthetic polynucleotide fragments coding for a desired fragment will be incorporated into recombinant polynucleotide constructs, usually DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell. Usually the polynucleotide constructs will be suitable for replication in a unicellular host, such as yeast or bacteria, but may also be intended for introduction to (with and without integration within the genome) cultured mammalian or plant or other eukaryotic cell lines. The purification of nucleic acids produced by the methods of the present invention is described, e.g., in Sambrook et al., 1989 or Ausubel et al., 1992.

The polynucleotides of the present invention may also be produced by chemical synthesis, e.g., by the phosphoramidite method described by Beaucage & Carruthers, 1981 or the triester method according to Matteucci and Caruthers, 1981, and may be performed on commercial, automated oligonucleotide synthesizers. A double-stranded fragment may be obtained from the single-stranded product of chemical synthesis either by synthesizing the complementary strand and annealing the strands together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Polynucleotide constructs prepared for introduction into a prokaryotic or eukaryotic host may comprise a replication system recognized by the host, including the intended polynucleotide fragment encoding the desired polypeptide, and will preferably also include transcription and translational initiation regulatory sequences operably linked to the polypeptide encoding segment. Expression vectors may include, for example, an origin of replication or autonomously replicating sequence (ARS) and expression control sequences, a promoter, an enhancer and necessary processing information sites, such as ribosome-binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, and mRNA stabilizing sequences. Secretion signals may also be included where appropriate, whether from a native BRCA1 protein or from other receptors or from secreted polypeptides of the same or related species, which allow the protein to cross and/or lodge in cell membranes, and thus attain its functional topology, or be secreted from the cell. Such vectors may be prepared by means of standard recombinant techniques well known in the art and discussed, for example, in Sambrook et al., 1989 or Ausubel et al. 1992.

An appropriate promoter and other necessary vector sequences will be selected so as to be functional in the host, and may include, when appropriate, those naturally associated with BRCA1 genes. Examples of workable combinations of cell lines and expression vectors are described in Sambrook et al., 1989 or Ausubel et al., 1992; see also, e.g., Metzger et al., 1988. Many useful vectors are known in the art and may be obtained from such vendors as Stratagene, New England Biolabs, Promega Biotech, and others. Promoters such as the trp, lac and phage promoters, tRNA promoters and glycolytic enzyme promoters may be used in prokaryotic hosts. Useful yeast promoters include promoter regions for metallothionein, 3-phosphoglycerate kinase or other glycolytic enzymes such as enolase or glyceraldehyde-3-phosphate dehydrogenase, enzymes responsible for maltose and galactose utilization, and others. Vectors and promoters suitable for use in yeast expression are further described in Hitzeman et al., EP 73,675A. Appropriate non-native mammalian promoters might include the early and late promoters from SV40 (Fiers et al., 1978) or promoters derived from murine Moloney leukemia virus, mouse tumor virus, avian sarcoma viruses, adenovirus II, bovine papilloma virus or polyoma. In addition, the construct may be joined to an amplifiable gene (e.g., DHFR) so that multiple copies of the gene may be made. For appropriate enhancer and other expression control sequences, see also *Enhancers and Eukaryotic Gene Expression*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1983).

While such expression vectors may replicate autonomously, they may also replicate by being inserted into the genome of the host cell, by methods well known in the art.

Expression and cloning vectors will likely contain a selectable marker, a gene encoding a protein necessary for survival or growth of a host cell transformed with the vector. The presence of this gene ensures growth of only those host cells which express the inserts. Typical selection genes encode proteins that a) confer resistance to antibiotics or other toxic substances, e.g. ampicillin, neomycin, methotrexate, etc.; b) complement auxotrophic deficiencies, or c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for Bacilli. The choice of the proper selectable marker will depend on the host cell, and appropriate markers for different hosts are well known in the art.

The vectors containing the nucleic acids of interest can be transcribed in vitro, and the resulting RNA introduced into the host cell by well-known methods, e.g., by injection (see, Kubo et al., 1988), or the vectors can be introduced directly into host cells by methods well known in the art, which vary depending on the type of cellular host, including electroporation; transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; infection (where the vector is an infectious agent, such as a retroviral genome); and other methods. See generally, Sambrook et al., 1989 and Ausubel et al., 1992. The introduction of the polynucleotides into the host cell by any method known in the art, including, inter alia, those described above, will be referred to herein as "transformation." The cells into which have been introduced nucleic acids described above are meant to also include the progeny of such cells.

Large quantities of the nucleic acids and polypeptides of the present invention may be prepared by expressing the BRCA1 nucleic acids or portions thereof in vectors or other expression vehicles in compatible prokaryotic or eukaryotic host cells. The most commonly used prokaryotic hosts are strains of *Escherichia coli*, although other prokaryotes, such as *Bacillus subtilis* or Pseudomonas may also be used.

Mammalian or other eukaryotic host cells, such as those of yeast, filamentous fungi, plant, insect, or amphibian or avian species, may also be useful for production of the proteins of the present invention. Propagation of mammalian cells in culture is per se well known. See, Jakoby and Pastan, 1979. Examples of commonly used mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cells, and WI38, BHK, and COS cell lines, although it will be appreciated by the skilled practitioner that other cell lines may be appropriate, e.g., to provide higher expression, desirable glycosylation patterns, or other features.

Clones are selected by using markers depending on the mode of the vector construction. The marker may be on the same or a different DNA molecule, preferably the same DNA molecule. In prokaryotic hosts, the transformant may be selected, e.g., by resistance to ampicillin, tetracycline or other antibiotics. Production of a particular product based on temperature sensitivity may also serve as an appropriate marker.

Prokaryotic or eukaryotic cells transformed with the polynucleotides of the present invention will be useful not only for the production of the nucleic acids and polypeptides of the present invention, but also, for example, in studying the characteristics of BRCA1 polypeptides.

Antisense polynucleotide sequences are useful in preventing or diminishing the expression of the BRCA1 locus, as will be appreciated by those skilled in the art. For example, polynucleotide vectors containing all or a portion of the BRCA1 locus or other sequences from the BRCA1 region (particularly those flanking the BRCA1 locus) may be placed under the control of a promoter in an antisense orientation and introduced into a cell. Expression of such an antisense construct within a cell will interfere with BRCA1 transcription and/or translation and/or replication.

The probes and primers based on the BRCA1 gene sequences disclosed herein are used to identify homologous BRCA1 gene sequences and proteins in other species. These BRCA1 gene sequences and proteins are used in the diagnostic/prognostic, therapeutic and drug screening methods described herein for the species from which they have been isolated.

Methods of Use: Nucleic Acid Diagnosis and Diagnostic Kits

In order to detect the presence of a BRCA1 allele predisposing an individual to cancer, a biological sample such as blood is prepared and analyzed for the presence or absence of susceptibility alleles of BRCA1. In order to detect the presence of neoplasia, the progression toward malignancy of a precursor lesion, or as a prognostic indicator, a biological sample of the lesion is prepared and analyzed for the presence or absence of mutant alleles of BRCAL. Results of these tests and interpretive information are returned to the health care provider for communication to the tested individual. Such diagnoses may be performed by diagnostic laboratories, or, alternatively, diagnostic kits are manufactured and sold to health care providers or to private individuals for self-diagnosis.

Initially, the screening method involves amplification of the relevant BRCA1 sequences. In another preferred embodiment of the invention, the screening method involves a non-PCR based strategy. Such screening methods include two-step label amplification methodologies that are well known in the art. Both PCR and non-PCR based screening strategies can detect target sequences with a high level of sensitivity.

The most popular method used today is target amplification. Here, the target nucleic acid sequence is amplified with polymerases. One particularly preferred method using polymerase-driven amplification is the polymerase chain reaction (PCR). The polymerase chain reaction and other polymerase-driven amplification assays can achieve over a million-fold increase in copy number through the use of polymerase-driven amplification cycles. Once amplified, the resulting nucleic acid can be sequenced or used as a substrate for DNA probes.

When the probes are used to detect the presence of the target sequences (for example, in screening for cancer susceptibility), the biological sample to be analyzed, such as blood or serum, may be treated, if desired, to extract the nucleic acids. The sample nucleic acid may be prepared in various ways to facilitate detection of the target sequence; e.g. denaturation, restriction digestion, electrophoresis or dot blotting. The targeted region of the analyte nucleic acid usually must be at least partially single-stranded to form hybrids with the targeting sequence of the probe. If the sequence is naturally single-stranded, denaturation will not be required. However, if the sequence is double-stranded, the sequence will probably need to be denatured. Denaturation can be carried out by various techniques known in the art.

Analyte nucleic acid and probe are incubated under conditions which promote stable hybrid formation of the target sequence in the probe with the putative targeted sequence in the analyte. The region of the probes which is used to bind to the analyte can be made completely complementary to the targeted region of human chromosome 17q.

Therefore, high stringency conditions are desirable in order to prevent false positives. However, conditions of high stringency are used only if the probes are complementary to regions of the chromosome which are unique in the genome. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, base composition, probe length, and concentration of formamide. These factors are outlined in, for example, Maniatis et al., 1982 and Sambrook et al., 1989. Under certain circumstances, the formation of higher order hybrids, such as triplexes, quadraplexes, etc., may be desired to provide the means of detecting target sequences.

Detection, if any, of the resulting hybrid is usually accomplished by the use of labeled probes. Alternatively, the probe may be unlabeled, but may be detectable by specific binding with a ligand which is labeled, either directly or indirectly. Suitable labels, and methods for labeling probes and ligands are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (e.g., nick translation, random priming or kinasing), biotin, fluorescent groups, chemiluminescent groups (e.g., dioxetanes, particularly triggered dioxetanes), enzymes, antibodies and the like. Variations of this basic scheme are known in the art, and include those variations that facilitate separation of the hybrids to be detected from extraneous materials and/or that amplify the signal from the labeled moiety. A number of these variations are reviewed in, e.g., Matthews & Kricka, 1988; Landegren et al., 1988; Mittlin, 1989; U.S. Patent No. 4,868,105, and in EPO Publication No. 225,807.

As noted above, non-PCR based screening assays are also contemplated in this invention. An exemplary non-PCR based procedure is provided in Example 11. This procedure hybridizes a nucleic acid probe (or an analog such as a methyl phosphonate backbone replacing the normal phosphodiester), to the low level DNA target. This probe may have an enzyme covalently linked to the probe, such that the covalent linkage does not interfere with the specificity of the hybridization. This enzyme-probe-conjugate-target nucleic acid complex can then be isolated away from the free probe enzyme conjugate and a substrate is added for enzyme detection. Enzymatic activity is observed as a change in color development or luminescent output resulting in a 10^3–10^6 increase in sensitivity. For an example relating to the preparation of oligodeoxynucleotide-alkaline phosphatase conjugates and their use as hybridization probes see Jablonski et al., 1986.

Two-step label amplification methodologies are known in the art. These assays work on the principle that a small ligand (such as digoxigenin, biotin, or the like) is attached to a nucleic acid probe capable of specifically binding BRCA1. Exemplary probes are provided in Table 9 of this patent application and additionally include the nucleic acid probe corresponding to nucleotide positions 3631 to 3930 of SEQ ID NO:1. Allele specific probes are also contemplated within the scope of this example and exemplary allele specific probes include probes encompassing the predisposing mutations summarized in Tables 11 and 12 of this patent application.

In one example, the small ligand attached to the nucleic acid probe is specifically recognized by an antibody-enzyme conjugate. In one embodiment of this example, digoxigenin is attached to the nucleic acid probe. Hybridization is detected by an antibody-alkaline phosphatase conjugate which turns over a chemiluminescent substrate. For methods for labeling nucleic acid probes according to this embodiment see Martin et al., 1990. In a second example, the small ligand is recognized by a second ligand-enzyme conjugate that is capable of specifically complexing to the first ligand. A well known embodiment of this example is the biotin-avidin type of interactions. For methods for labeling nucleic acid probes and their use in biotin-avidin based assays see Rigby et al., 1977 and Nguyen et al., 1992.

It is also contemplated within the scope of this invention that the nucleic acid probe assays of this invention will employ a cocktail of nucleic acid probes capable of detecting BRCA1. Thus, in one example to detect the presence of BRCA1 in a cell sample, more than one probe complementary to BRCA1 is employed and in particular the number of different probes is alternatively 2, 3, or 5 different nucleic acid probe sequences. In another example, to detect the presence of mutations in the BRCA1 gene sequence in a patient, more than one probe complementary to BRCA1 is employed where the cocktail includes probes capable of binding to the allele-specific mutations identified in populations of patients with alterations in BRCA1. In this embodiment, any number of probes can be used, and will preferably include probes corresponding to the major gene mutations identified as predisposing an individual to breast cancer. Some candidate probes contemplated within the scope of the invention include probes that include the allele-specific mutations identified in Tables 11 and 12 and those that have the BRCA1 regions corresponding to SEQ ID NO:1 both 5' and 3' to the mutation site.

Methods of Use: Peptide Diagnosis and Diagnostic Kits

The neoplastic condition of lesions can also be detected on the basis of the alteration of wild-type BRCA1 polypeptide. Such alterations can be determined by sequence analysis in accordance with conventional techniques. More preferably, antibodies (polyclonal or monoclonal) are used to detect differences in, or the absence of BRCA1 peptides. The antibodies may be prepared as discussed above under the heading "Antibodies" and as further shown in Examples 12 and 13. Other techniques for raising and purifying antibodies are well known in the art and any such techniques may be chosen to achieve the preparations claimed in this invention. In a preferred embodiment of the invention, antibodies will immunoprecipitate BRCA1 proteins from solution as well as react with BRCA1 protein on Western or immunoblots of polyacrylamide gels. In another preferred embodiment, antibodies will detect BRCA1 proteins in paraffin or frozen tissue sections, using immunocytochemical techniques.

Preferred embodiments relating to methods for detecting BRCA1 or its mutations include enzyme linked immunosorbent assays (ELISA), radioimmunoassays (RIA), immunoradiometric assays (IRMA) and immunoenzymatic assays (IEMA), including sandwich assays using monoclonal and/or polyclonal antibodies. Exemplary sandwich assays are described by David et al. in U.S. Pat. Nos. 4,376,110 and 4,486,530, hereby incorporated by reference, and exemplified in Example 14.

Methods of Use: Drug Screening

This invention is particularly useful for screening compounds by using the BRCA1 polypeptide or binding fragment thereof in any of a variety of drug screening techniques.

The BRCA1 polypeptide or fragment employed in such a test may either be free in solution, affixed to a solid support, or borne on a cell surface. One method of drug screening utilizes eucaryotic or procaryotic host cells which are stably transformed with recombinant polynucleotides expressing the polypeptide or fragment, preferably in competitive binding assays. Such cells, either in viable or fixed form, can be

used for standard binding assays. One may measure, for example, for the formation of complexes between a BRCA1 polypeptide or fragment and the agent being tested, or examine the degree to which the formation of a complex between a BRCA1 polypeptide or fragment and a known ligand is interfered with by the agent being tested.

Thus, the present invention provides methods of screening for drugs comprising contacting such an agent with a BRCA1 polypeptide or fragment thereof and assaying (i) for the presence of a complex between the agent and the BRCA1 polypeptide or fragment, or (ii) for the presence of a complex between the BRCA1 polypeptide or fragment and a ligand, by methods well known in the art. In such competitive binding assays the BRCA1 polypeptide or fragment is typically labeled. Free BRCA1 polypeptide or fragment is separated from that present in a protein:protein complex, and the amount of free (i.e., uncomplexed) label is a measure of the binding of the agent being tested to BRCA1 or its interference with BRCA1:ligand binding, respectively.

Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to the BRCA1 polypeptides and is described in detail in Geysen, PCT published application WO 84/03564, published on Sep. 13, 1984. Briefly stated, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with BRCA1 polypeptide and washed. Bound BRCA1 polypeptide is then detected by methods well known in the art.

Purified BRCA1 can be coated directly onto plates for use in the aforementioned drug screening techniques. However, non-neutralizing antibodies to the polypeptide can be used to capture antibodies to immobilize the BRCA1 polypeptide on the solid phase.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of specifically binding the BRCA1 polypeptide compete with a test compound for binding to the BRCA1 polypeptide or fragments thereof. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants of the BRCA1 polypeptide.

A further technique for drug screening involves the use of host eukaryotic cell lines or cells (such as described above) which have a nonfunctional BRCA1 gene. These host cell lines or cells are defective at the BRCA1 polypeptide level. The host cell lines or cells are grown in the presence of drug compound. The rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of BRCA1 defective cells.

Methods of Ujse: Rational Drug Design

The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which, e.g., enhance or interfere with the function of a polypeptide in vivo. See, e.g., Hodgson, 1991. In one approach, one first determines the three-dimensional structure of a protein of interest (e.g., BRCA1 polypeptide) or, for example, of the BRCA1-receptor or ligand complex, by x-ray crystallography, by computer modeling or most typically, by a combination of approaches. Less often, useful information regarding the structure of a polypeptide may be gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson et al., 1990). In addition, peptides (e.g., BRCA1 polypeptide) are analyzed by an alanine scan (Wells, 1991). In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

It is also possible to isolate a target-specific antibody, selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idio-typic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original receptor. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

Thus, one may design drugs which have, e.g., improved BRCA1 polypeptide activity or stability or which act as inhibitors, agonists, antagonists, etc. of BRCA1 polypeptide, activity. By virtue of the availability of cloned BRCA1 sequences, sufficient amounts of the BRCA1 polypeptide may be made available to perform such analytical studies as x-ray crystallography. In addition, the knowledge of the BRCA1 protein sequence provided herein will guide those employing computer modeling techniques in place of, or in addition to x-ray crystallography.

Methods of Use: Gene Therapy

According to the present invention, a method is also provided of supplying wild-type BRCA1 function to a cell which carries mutant BRCA1 alleles. Supplying such a function should suppress neoplastic growth of the recipient cells. The wild-type BRCA1 gene or a part of the gene may be introduced into the cell in a vector such that the gene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. If a gene fragment is introduced and expressed in a cell carrying a mutant BRCA1 allele, the gene fragment should encode a part of the BRCA1 protein which is required for non-neoplastic growth of the cell. More preferred is the situation where the wild-type BRCA1 gene or a part thereof is introduced into the mutant cell in such a way that it recombines with the endogenous mutant BRCA1 gene present in the cell. Such recombination requires a double recombination event which results in the correction of the BRCA1 gene mutation. Vectors for introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector may be used. Methods for introducing DNA into cells such as electroporation, calcium phosphate co-precipitation and viral transduction are known in the art, and the choice of method is within the competence of the routineer. Cells transformed with the wild-type BRCA1 gene can be used as model systems to study cancer remission and drug treatments which promote such remission.

As generally discussed above, the BRCA1 gene or fragment, where applicable, may be employed in gene therapy methods in order to increase the amount of the expression products of such genes in cancer cells. Such gene therapy is particularly appropriate for use in both cancerous and pre-cancerous cells, in which the level of BRCA1 polypeptide is absent or diminished compared to normal cells. It may also be useful to increase the level of expression of a given BRCA1 gene even in those tumor cells in which the mutant gene is expressed at a "normal" level, but the gene product is not fully functional.

Gene therapy would be carried out according to generally accepted methods, for example, as described by Friedman, 1991. Cells from a patient's tumor would be first analyzed by the diagnostic methods described above, to ascertain the production of BRCA1 polypeptide in the tumor cells. A virus or plasmid vector (see further details below), containing a copy of the BRCA1 gene linked to expression control elements and capable of replicating inside the tumor cells, is prepared. Suitable vectors are known, such as disclosed in U.S. Pat. No. 5,252,479 and PCT published application WO 93/07282. The vector is then injected into the patient, either locally at the site of the tumor or systemically (in order to reach any tumor cells that may have metastasized to other sites). If the transfected gene is not permanently incorporated into the genome of each of the targeted tumor cells, the treatment may have to be repeated periodically.

Gene transfer systems known in the art may be useful in the practice of the gene therapy methods of the present invention. These include viral and nonviral transfer methods. A number of viruses have been used as gene transfer vectors, including papovaviruses, e.g., SV40 (Madzak et al., 1992), adenovirus (Berkner, 1992; Berkner et al., 1988; Gorziglia and Kapikian, 1992; Quantin et al., 1992; Rosenfeld et al., 1992; Wilkinson et al., 1992; Stratford-Perricaudet et al., 1990), vaccinia virus (Moss, 1992), adeno-associated virus (Muzyczka, 1992; Ohi et al., 1990), herpesviruses including HSV and EBV (Margolskee, 1992; Johnson et al., 1992; Fink et al., 1992; Breakfield and Geller, 1987; Freese et al., 1990), and retroviruses of avian (Brandyopadhyay and Temin, 1984; Petropoulos et al., 1992), murine (Miller, 1992; Miller el al. 1985; Sorge et al., 1984; Mann and Baltimore, 1985; Miller et al., 1988), and human origin (Shimada et al., 1991; Helseth et al., 1990; Page et al., 1990; Buchschacher and Panganiban, 1992). Most human gene therapy protocols have been based on disabled murine retroviruses.

Nonviral gene transfer methods known in the art include chemical techniques such as calcium phosphate coprecipitation (Graham and van der Eb, 1973; Pellicer et al., 1980); mechanical techniques, for example microinjection (Anderson et al., 1980; Gordon et al., 1980; Brinster et al., 1981; Constantini and Lacy, 1981); membrane fusion-mediated transfer via liposomes (Felgner et al., 1987; Wang and Huang, 1989; Kaneda et al, 1989; Stewart et al., 1992; Nabel et al., 1990; Lim et al., 1992); and direct DNA uptake and receptor-mediated DNA transfer (Wolff et al., 1990; Wu et al., 1991; Zenke et al., 1990; Wu et al., 1989b; Wolff et al., 1991; Wagner et al., 1990; Wagner et al., 1991; Cotten et al., 1990; Curiel et al., 1991a; Curiel et al., 1991b). Viral-mediated gene transfer can be combined with direct in vivo gene transfer using liposome delivery, allowing one to direct the viral vectors to the tumor cells and not into the surrounding nondividing cells. Alternatively, the retroviral vector producer cell line can be injected into tumors (Culver et al., 1992). Injection of producer cells would then provide a continuous source of vector particles. This technique has been approved for use in humans with inoperable brain tumors.

In an approach which combines biological and physical gene transfer methods, plasmid DNA of any size is combined with a polylysine-conjugated antibody specific to the adenovirus hexon protein, and the resulting complex is bound to an adenovirus vector. The trimolecular complex is then used to infect cells. The adenovirus vector permits efficient binding, internalization, and degradation of the endosome before the coupled DNA is damaged.

Liposome/DNA complexes have been shown to be capable of mediating direct in vivo gene transfer. While in standard liposome preparations the gene transfer process is nonspecific, localized in vivo uptake and expression have been reported in tumor deposits, for example, following direct in situ administration (Nabel, 1992).

Gene transfer techniques which target DNA directly to breast and ovarian tissues, e.g., epithelial cells of the breast or ovaries, is preferred. Receptor-mediated gene transfer, for example, is accomplished by the conjugation of DNA (usually in the form of covalently closed supercoiled plasmid) to a protein ligand via polylysine. Ligands are chosen on the basis of the presence of the corresponding ligand receptors on the cell surface of the target cell/tissue type. One appropriate receptor/ligand pair may include the estrogen receptor and its ligand, estrogen (and estrogen analogues). These ligand-DNA conjugates can be injected directly into the blood if desired and are directed to the target tissue where receptor binding and internalization of the DNA-protein complex occurs. To overcome the problem of intracellular destruction of DNA, coinfection with adenovirus can be included to disrupt endosome function.

The therapy involves two steps which can be performed singly or jointly. In the first step, prepubescent females who carry a BRCA1 susceptibility allele are treated with a gene delivery vehicle such that some or all of their mammary ductal epithelial precursor cells receive at least one additional copy of a functional normal BRCA1 allele. In this step, the treated individuals have reduced risk of breast cancer to the extent that the effect of the susceptible allele has been countered by the presence of the normal allele. In the second step of a preventive therapy, predisposed young females, in particular women who have received the proposed gene therapeutic treatment, undergo hormonal therapy to mimic the effects on the breast of a full term pregnancy.

Methods of Use: Peptide Therapy

Peptides which have BRCA1 activity can be supplied to cells which carry mutant or missing BRCA1 alleles. The sequence of the BRCA1 protein is disclosed (SEQ ID NO:2). Protein can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors. Alternatively, BRCA1 polypeptide can be extracted from BRCA1-producing mammalian cells. In addition, the techniques of synthetic chemistry can be employed to synthesize BRCA1 protein. Any of such techniques can provide the preparation of the present invention which comprises the BRCA1 protein. The preparation is substantially free of other human proteins. This is most readily accomplished by synthesis in a microorganism or in vitro.

Active BRCA1 molecules can be introduced into cells by microinjection or by use of liposomes, for example. Alternatively, some active molecules may be taken up by cells, actively or by diffusion. Extracellular application of the BRCA1 gene product may be sufficient to affect tumor growth. Supply of molecules with BRCA1 activity should lead to partial reversal of the neoplastic state. Other molecules with BRCA1 activity (for example, peptides, drugs or organic compounds) may also be used to effect such a reversal. Modified polypeptides having substantially similar function are also used for peptide therapy.

Methods of Use: Transformed Hosts

Similarly, cells and animals which carry a mutant BRCA1 allele can be used as model systems to study and test for substances which have potential as therapeutic agents. The cells are typically cultured epithelial cells. These may be isolated from individuals with BRCA1 mutations, either somatic or germline. Alternatively, the cell line can be engineered to carry the mutation in the BRCA1 allele, as described above. After a test substance is applied to the cells,

the neoplastically transformed phenotype of the cell is determined. Any trait of neoplastically transformed cells can be assessed, including anchorage-independent growth, tumorigenicity in nude mice, invasiveness of cells, and growth factor dependence. Assays for each of these traits are known in the art.

Animals for testing therapeutic agents can be selected after mutagenesis of whole animals or after treatment of germline cells or zygotes. Such treatments include insertion of mutant BRCA1 alleles, usually from a second animal species, as well as insertion of disrupted homologous genes. Alternatively, the endogenous BRCA1 gene(s) of the animals may be disrupted by insertion or deletion mutation or other genetic alterations using conventional techniques (Capecchi, 1989; Valancius and Smithies, 1991; Hasty et al., 1991; Shinkai et al., 1992; Mombaerts et al., 1992; Philpott et al., 1992; Snouwaert et al., 1992; Donehower et al., 1992). After test substances have been administered to the animals, the growth of tumors must be assessed. If the test substance prevents or suppresses the growth of tumors, then the test substance is a candidate therapeutic agent for the treatment of the cancers identified herein. These animal models provide an extremely important testing vehicle for potential therapeutic products.

The present invention is described by reference to the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described below were utilized.

EXAMPLE 1

Ascertain and Study Kindreds Likely to Have a 1 7q-Linked Breast Cancer Susceptibility Locus

Extensive cancer prone kindreds were ascertained by our University of Utah collabrators from a defined population providing a large set of extended kindreds with multiple cases of breast cancer and many relatives available to study. The large number of meioses present in these large kindreds provided the power to detect whether the BRCA1 locus was segregating, and increased the opportunity for informative recombinants to occur within the small region being investigated. This vastly improved the chances of establishing linkage to the BRCA1 region, and greatly facilitated the reduction of the BRCA1 region to a manageable size, which permits identification of the BRCA1 locus itself.

Each kindred was extended through all available connecting relatives by our collabrators, and to all informative first degree relatives of each proband or cancer case. For these kindreds, additional breast cancer cases and individuals with cancer at other sites of interest (e.g. ovarian) who also appeared in the kindreds were identified through the tumor registry linked files. All breast cancers reported in the kindred which were not confirmed in the Utah Cancer Registry were researched. Medical records or death certificates were obtained for confirmation of all cancers. Each key connecting individual and all informative individuals were invited by our collaborators to participate by providing a blood sample from which DNA was extracted. They also sampled spouses and relatives of deceased cases so that the genotype of the deceased cases could be inferred from the genotypes of their relatives.

Ten kindreds which had three or more cancer cases with inferable genotypes were selected for linkage studies to 17q markers from a set of 29 kindreds originally ascertained for a study of proliferative breast disease and breast cancer (Skolnick et al., 1990). The criterion for selection of these kindreds was the presence of two sisters or a mother and her daughter with breast cancer. Additionally, two kindreds which have been studied since 1980 as part of their breast cancer linkage studies (K1001, K9018), six kindreds ascertained from for the presence of clusters of breast and/or ovarian cancer (K2019, K2073, K2079, K2080, K2039, K2082) and a self-referred kindred with early onset breast cancer (K2035) were included. These kindreds were investigated and expanded in our collaborators clinic in the manner described above. Table 1 displays the characteristics of these 19 kindreds which are the subject of subsequent exameach kin Table 1, for each kindred the total number of individuals in our database, the number of typed individuals, and the minimum, median, and maximum age at diagnosis of breast/ovarian cancer are reported. Kindreds are sorted in ascending order of median age at diagnosis of breast cancer. Four women diagnosed with both ovarian and breast cancer are counted in both categories.

TABLE 1

Description of the 19 Kindreds

	No. of Individuals		Breast				Ovarian			
				Age at Dx				Age at Dx		
KINDRED	Total	Sample	# Aff.	Min.	Med.	Max.	# Aff.	Min.	Med.	Max.
1910	15	10	4	27	34	49	—	—	—	—
1001	133	98	13	28	37	64	—	—	—	—
2035	42	25	8	28	37	45	1	—	60	—
2027	21	11	4	34	38	41	—	—	—	—
9018	54	17	9	30	40	72	2	46	48	50
1925	50	27	4	39	42	53	—	—	—	—
1927	49	29	5	32	42	51	—	—	—	—
1911	28	21	7	28	42	76	—	—	—	—
1929	16	11	4	34	43	73	—	—	—	—
1901	35	19	10	31	44	76	—	—	—	—
2082	180	105	20	27	47	67	10	45	52	66
2019	42	19	10	42	53	79	—	—	—	—

TABLE 1-continued

Description of the 19 Kindreds										
	No. of Individuals		Breast				Ovarian			
				Age at Dx				Age at Dx		
KINDRED	Total	Sample	# Aff.	Min.	Med.	Max.	# Aff.	Min.	Med.	Max.
1900	70	23	8	45	55	70	1	—	78	—
2080	264	74	22+	27	55	92	4	45	53	71
2073	57	29	9	35	57	80	—	—	—	—
1917	16	6	4	43	58	61	—	—	—	—
1920	22	14	3	62	63	68	—	—	—	—
2079	136	18	14	38	66	84	4	52	59	65
2039	87	40	14	44	68	88	4	41	51	75

+Includes one case of male breast cancer.

EXAMPLE 2

Selection of Kindreds Which are Linked to Chromosome 17q and Localization of BRCA1 to the Interval Mfd15–Mfd188

For each sample collected in these 19 kindreds, DNA was extracted from blood (or in two cases from paraffin-embedded tissue blocks) using standard laboratory protocols. Genotyping in this study was restricted to short tandem repeat (STR) markers since, in general, they have high heterozygosity and PCR methods offer rapid turnaround while using very small amounts of DNA. To aid in this effort, four such STR markers on chromosome 17 were developed by screening a chromosome specific cosmid library for CA positive clones. Three of these markers localized to the long arm: (46E6, Easton et al., 1993); (42D6, Easton et al., 1993); 26C2 (D17S514, Oliphant et al., 1991), while the other, 12G6 (D17S513, Oliphant et al., 1991), localized to the short arm near the p53 tumor suppressor locus. Two of these, 42D6 and 46E6, were submitted to the Breast Cancer Linkage Consortium for typing of breast cancer families by investigators worldwide. Oligonucleotide sequences for markers not developed in our laboratory were obtained from published reports, or as part of the Breast Cancer Linkage Consortium, or from other investigators. All genotyping films were scored blindly with a standard lane marker used to maintain consistent coding of alleles. Key samples in the four kindreds presented here underwent duplicate typing for all relevant markers. All 19 kindreds have been typed for two polymorphic CA repeat markers: 42D6 (D17S588), a CA repeat isolated in our laboratory, and Mfd15 (D17S250), a CA repeat provided by J. Weber (Weber et al., 1990). Several sources of probes were used to create genetic markers on chromosome 17, specifically chromosome 17 cosmid and lambda phage libraries created from sorted chromosomes by the Los Alamos National Laboratories (van Dilla et al., 1986).

LOD scores for each kindred with these two markers (42D6, Mfd15) and a third marker, Mfd188 (D17S579, Hall et al., 1992), located roughly midway between these two markers, were calculated for two values of the recombination fraction, 0.001 and 0.1. (For calculation of LOD scores, see Oh, 1985). Likelihoods were computed under the model derived by Claus et al., 1991, which assumes an estimated gene frequency of 0.003, a lifetime risk in gene carriers of about 0.80, and population based age-specific risks for breast cancer in non-gene carriers. Allele frequencies for the three markers used for the LOD score calculations were calculated from our own laboratory typings of unrelated individuals in the CEPH panel (White and Lalouel, 1988). Table 2 shows the results of the pairwise linkage analysis of each kindred with the three markers 42D6, Mfd188, and Mfd15.

TABLE 2

Pairwise Linkage Analysis of Kindreds						
	Mfd15 (D17S250) Recombination		Mfd188 (D17S579) Recombination		42D6 (D17S588) Recombination	
KINDRED	0.001	0.1	0.001	0.1	0.001	0.1
1910	0.06	0.30	0.06	0.30	0.06	0.30
1001	−0.30	−0.09	NT	NT	−0.52	−0.19
2035	2.34	1.85	0.94	0.90	2.34	1.82
2027	−1.22	−0.33	−1.20	−0.42	−1.16	−0.33
9018	−0.54	−0.22	−0.17	−0.10	0.11	0.07
1925	1.08	0.79	0.55	0.38	−0.11	−0.07
1927	−0.41	0.01	−0.35	0.07	−0.44	−0.02
1911	−0.27	−0.13	−0.43	−0.23	0.49	0.38
1929	−0.49	−0.25	NT	NT	−0.49	−0.25
1901	1.50	1.17	0.78	0.57	0.65	0.37
2082	4.25	3.36	6.07	5.11	2.00	3.56
2019	−0.10	−0.01	−0.11	−0.05	−0.18	−0.10

TABLE 2-continued

	Pairwise Linkage Analysis of Kindreds					
	Mfd15 (D17S250) Recombination		Mfd188 (D17S579) Recombination		42D6 (D17S588) Recombination	
KINDRED	0.001	0.1	0.001	0.1	0.001	0.1
1900	−0.14	−0.11	NT	NT	−0.12	−0.05
2080	−0.16	−0.04	0.76	0.74	−1.25	−0.58
2073	−0.41	−0.29	0.63	0.49	−0.23	−0.13
1917	−0.02	−0.02	NT	NT	−0.01	0.00
1920	−0.03	−0.02	NT	NT	0.00	0.00
2079	0.02	0.01	−0.01	−0.01	0.01	0.01
2039	−1.67	−0.83	0.12	0.59	−1.15	0.02

NT - Kindred not typed for Mfd188.

Using a criterion for linkage to 17q of a LOD score>1.0 for at least one locus under the CASH model (Claus et al., 1991), four of the 19 kindreds appeared to be linked to 17q (K1901, K1925, K2035, K2082). A number of additional kindreds showed some evidence of linkage but at this time could not be definitively assigned to the linked category. These included kindreds K1911, K2073, K2039, and K2080. Three of the 17q-linked kindreds had informative recombinants in this region and these are detailed below.

Kindred 2082 is the largest 17q-linked breast cancer family reported to date by any group. The kindred contains 20 cases of breast cancer, and ten cases of ovarian cancer. Two cases have both ovarian and breast cancer. The evidence of linkage to 17q for this family is overwhelming; the LOD score with the linked haplotype is over 6.0, despite the existence of three cases of breast cancer which appear to be sporadic, i.e., these cases share no part of the linked haplotype between Mfd15 and 42D6. These three sporadic cases were diagnosed with breast cancer at ages 46, 47, and 54. In smaller kindreds, sporadic cancers of this type greatly confound the analysis of linkage and the correct identification of key recombinants. The key recombinant in the 2082 kindred is a woman who developed ovarian cancer at age 45 whose mother and aunt had ovarian cancer at ages 58 and 66, respectively. She inherited the linked portion of the haplotype for both Mfd188 and 42D6 while inheriting unlinked alleles at Mfd15; this recombinant event placed BRCA1 distal to Mfd15.

K1901 is typical of early-onset breast cancer kindreds. The kindred contains 10 cases of breast cancer with a median age at diagnosis of 43.5 years of age; four cases were diagnosed under age 40. The LOD score for this kindred with the marker 42D6 is 1.5, resulting in a posterior probability of 17q-linkage of 0.96. Examination of haplotypes in this kindred identified a recombinant haplotype in an obligate male carrier and his affected daughter who was diagnosed with breast cancer at age 45. Their linked allele for marker Mfd15 differs from that found in all other cases in the kindred (except one case which could not be completely inferred from her children). The two haplotypes are identical for Mfd188 and 42D6. Accordingly, data from Kindred 1901 would also place the BRCA1 locus distal to Mfd15.

Kindred 2035 is similar to K1901 in disease phenotype. The median age of diagnosis for the eight cases of breast cancer in this kindred is 37. One case also had ovarian cancer at age 60. The breast cancer cases in this family descend from two sisters who were both unaffected with breast cancer until their death in the eighth decade. Each branch contains four cases of breast cancer with at least one case in each branch having markedly early onset. This kindred has a LOD score of 2.34 with Mfd15. The haplotypes segregating with breast cancer in the two branches share an identical allele at Mfd15 but differ for the distal loci Mfd188 and NM23 (a marker typed as part of the consortium which is located just distal to 42D6 (Hall et al., 1992)). Although the two haplotypes are concordant for marker 42D6, it is likely that the alleles are shared identical by state (the same allele but derived from different ancestors), rather than identical by descent (derived from a common ancestor) since the shared allele is the second most common allele observed at this locus. By contrast the linked allele shared at Mfd15 has a frequency of 0.04. This is a key recombinant in our dataset as it is the sole recombinant in which BRCA1 segregated with the proximal portion of the haplotype, thus setting the distal boundary to the BRCA1 region. For this event not to be a key recombinant requires that a second mutant BRCA1 gene be present in a spouse marrying into the kindred who also shares the rare Mfd15 allele segregating with breast cancer in both branches of the kindred. This event has a probability of less than one in a thousand. The evidence from this kindred therefore placed the BRCA1 locus proximal to Mfd188.

EXAMPLE 3

Creation of a Fine Structure Map and Refinement of the BRCA1 Region to Mfd191–Mfd188 using Additional STR Polymorphisms

In order to improve the characterization of our recombinants and define closer flanking markers, a dense map of this relatively small region on chromosome 17q was required. The chromosome 17 workshop has produced a consensus map of this region (FIG. 1) based on a combination of genetic and physical mapping studies (Fain, 1992). This map contains both highly polymorphic STR polymorphisms, and a number of nonpolymorphic expressed genes. Because this map did not give details on the evidence for this order nor give any measure of local support for inversions in the order of adjacent loci, we viewed it as a rough guide for obtaining resources to be used for the development of new markers and construction of our own detailed genetic and physical map of a small region containing BRCA1. Our approach was to analyze existing STR markers provided by other investigators and any newly developed markers from our laboratory with respect to both a panel of meiotic (genetic) breakpoints identified using DNA from the CEPH reference families and a panel of somatic cell hybrids (physical breakpoints) constructed for this region. These markers included 26C2 developed in our laboratory which maps proximal to Mfd15, Mfd191 (provided by James Weber),

THRA1 (Futreal et al., 1992a), and three polymorphisms kindly provided to us by Dr. Donald Black, NM23 (Hall et al. 1992), SCG40 (D17S181), and 6C1 (D17S293).

Genetic localization of markers.

In order to localize new markers genetically within the region of interest, we have identified a number of key meiotic breakpoints within the region, both in the CEPH reference panel and in our large breast cancer kindred (K2082). Given the small genetic distance in this region, they are likely to be only a relatively small set of recombinants which can be used for this purpose, and they are likely to group markers into sets. The orders of the markers within each set can only be determined by physical mapping. However the number of genotypings necessary to position a new marker is minimized. These breakpoints are illustrated in Tables 3 and 4. Using this approach we were able to genetically order the markers THRA1, 6C1, SCG40, and Mfd191. As can be seen from Tables 3 and 4, THRA1 and MFD191 both map inside the Mfd15–Mfd188 region we had previously identified as containing the BRCA1 locus. In Tables 3 and 4, M/P indicates a maternal or paternal recombinant. A "1" indicates inherited allele is of grandpaternal origin, while a "0" indicates grandmaternal origin, and "-" indicates that the locus was untyped or uninformative.

TABLE 3

CEPH Recombinants

Family	ID	M/P	Mfd15	THRA1	Mfd191	Mfd188	SCG40	6C1	42D6
13292	4	M	1	1	1	0	0	0	0
13294	4	M	1	1	1	0	0	0	0
13294	6	M	0	0	1	1	—	—	—
1334	3	M	1	1	1	1	1	0	0
1333	4	M	1	1	1	0	—	—	0
1333	6	M	0	0	1	1	—	—	1
1333	8	P	1	0	0	0	—	—	0
1377	8	M	0	—	0	0	0	0	1

TABLE 4

Kindred 2082 Recombinants

Family	ID	M/P	Mfd15	Mfd191	Mfd188	SCG40	6C1	42D6
75		M	0	1	1	1	—	—
63		M	0	0	1	1	—	1
125		M	1	1	1	0	—	0
40		M	1	1	0	0	—	0

Analysis of markers Mfd15, Mfd188, Mfd191, and THRA1 in our recombinant families.

Mfd15, Mfd188, Mfd191 and THRA1 were typed in our recombinant families and examined for additional information to localize the BRCA1 locus. In kindred 1901, the Mfd15 recombinant was recombinant for THRA1 but uninformative for Mfd191, thus placing BRCA1 distal to THRA1. In K2082, the recombinant with Mfd15 also was recombinant with Mfd191, thus placing the BRCA1 locus distal to Mfd191 (Goldgar et al., 1994). Examination of THRA1 and Mfd191 in kindred K2035 yielded no further localization information as the two branches were concordant for both markers. However, SCG40 and 6C1 both displayed the same pattern as Mfd188, thus increasing our confidence in the localization information provided by the Mfd188 recombinant in this family. The BRCA1 locus, or at least a portion of it, therefore lies within an interval bounded by Mfd191 on the proximal side and Mfd188 on the distal side.

EXAMPLE 4

Development of Genetic and Physical Resources in the Region of Interest

To increase the number of highly polymorphic loci in the Mfd191–Mfd188 region, we developed a number of STR markers in our laboratory from cosmids and YACs which physically map to the region. These markers allowed us to further refine the region.

STSs were identified from genes known to be in the desired region to identify YACs which contained these loci, which were then used to identify subclones in cosmids, P1s or BACs. These subclones were then screened for the presence of a CA tandem repeat using a $(CA)_n$ oligonucleotide (Pharmacia). Clones with a strong signal were selected preferentially, since they were more likely to represent CA-repeats which have a large number of repeats and/or are of near-perfect fidelity to the $(CA)_n$ pattern. Both of these characteristics are known to increase the probability of polymorphism (Weber, 1990). These clones were sequenced directly from the vector to locate the repeat. We obtained a unique sequence on one side of the CA-repeat by using one of a set of possible primers complementary to the end of a CA-repeat, such as $(GT)_{10}T$. Based on this unique sequence, a primer was made to sequence back across the repeat in the other direction, yielding a unique sequence for design of a second primer flanking the CA-repeat. STRs were then screened for polymorphism on a small group of unrelated individuals and tested against the hybrid panel to confirm their physical localization. New markers which satisfied these criteria were then typed in a set of 40 unrelated individuals from the Utah and CEPH families to obtain allele frequencies appropriate for the study population. Many of the other markers reported in this study were tested in a smaller group of CEPH unrelated individuals to obtain similarly appropriate allele frequencies.

Using the procedure described above, a total of eight polymorphic STRs was found from these YACS. Of the loci identified in this manner, four were both polymorphic and localized to the BRCA1 region. Four markers did not localize to chromosome 17, reflecting the chimeric nature of the YACs used. The four markers which were in the region were denoted AA1, ED2, 4–7, and YM29. AA1 and ED2 were developed from YACs positive for the RNU2 gene, 4–7 from an EPB3 YAC and YM29 from a cosmid which localized to the region by the hybrid panel. A description of the number of alleles, heterozygosity and source of these

four and all other STR polymorphisms analyzed in the breast cancer kindreds is given below in Table 5.

TABLE 5

Polymorphic Short Tandem Repeat Markers Used for Fine Structure Mapping of the BRCA1 Locus

Clone	Gene	Na**	Hetero-zygosity	Allele* Frequency (%) 1	2	3	4	5	6
Mfd15	D17S250	10	0.82	26	22	15	7	7	23
THRA1	THRA1	5							
Mfd191	D17S776	7	0.55	48	20	11	7	7	7
ED2	D17S1327	12	0.55	62	9	8	5	5	11
AA1	D17S1326	7	0.83	28	28	25	8	6	5
CA375	D17S184	10	0.75	26	15	11	9	9	20
4-7	D17S1183	9	0.50	63	15	8	6	4	4
YM29	—	9	0.62	42	24	12	7	7	8
Mfd188	D17S579	12	0.92	33	18	8	8	8	25
SCG40	D17S181	14	0.90	20	18	18	10	8	35
42D6	D17S588	11	0.86	21	17	11	10	9	32
6C1	D17S293	7	0.75	30	30	11	11	9	9
Z109	D17S750	9	0.70	33	27	7	7	7	19
tdj1475	D17S1321	13	0.84	21	16	11	11	8	33
CF4	D17S1320	6	0.63	50	27	9	7	4	3
tdj1239	D17S1328	10	0.80	86	10	9	7	4	14
U5	D17S1325	13	0.83	19	16	12	10	9	34

*Allele codes 1–5 are listed in decreasing frequency; allele numbers do not correspond to fragment sizes. Allele 6 frequency is the joint frequency of all other alleles for each locus.
**Number of alleles seen in the genetically independent DNA samples used for calculating allele frequencies.

The four STR polymorphisms which mapped physically to the region (4–7, ED2, AA1, YM29) were analyzed in the meiotic, breakpoint panel shown initially in Tables 3 and 4. Tables 6 and 7 contain the relevant CEPH data and Kindred 2082 data for localization of these four markers. In the tables, M/P indicates a maternal or paternal recombinant. A "1" indicates inherited allele is of grandpaternal origin, while a "0" indicates grandmaternal origin, and "-" indicates that the locus was untyped or uninformative.

From CEPH 1333-04, we see that AA1 and YM29 must lie distal to Mfd191. From 13292, it can be inferred that both AA1 and ED2 are proximal to 4–7, YM29, and Mfd188. The recombinants found in K2082 provide some additional ordering information. Three independent observations (individual numbers 22, 40, & 63) place AA1, ED2, 4–7, and YM29, and Mfd188 distal to Mfd191, while ID 125 places 4–7, YM29, and Mfd188 proximal to SCG40. No genetic information on the relative ordering within the two clusters of markers AA1/ED2 and 4–7/YM29/Mfd188 was obtained from the genetic recombinant analysis. Although ordering loci with respect to hybrids which are known to contain "holes" in which small pieces of interstitial human DNA may be missing is problematic, the hybrid patterns indicate that 4–7 lies above both YM29 and Mfd188.

EXAMPLE 5

Genetic Analyses of Breast Cancer Kindreds with Markers AA1, 4–7, ED2, and YM29

In addition to the three kindreds containing key recombinants which have been discussed previously, kindred K2039 was shown through analysis of the newly developed STR markers to be linked to the region and to contain a useful recombinant.

Table 8 defines the haplotypes (shown in coded form) of the kindreds in terms of specific marker alleles at each locus and their respective frequencies. In Table 8, alleles are listed in descending order of frequency; frequencies of alleles 1–5 for each locus are given in Table 5. Haplotypes coded H are BRCA1 associated haplotypes, P designates a partial H haplotype, and an R indicates an observable recombinant haplotype. As evident in Table 8, not all kindreds were typed for all markers; moreover, not all individuals within a kindred were typed for an identical set of markers, especially in K2082. With one exception, only haplotypes inherited from affected or at-risk kindred members are shown; haplotypes from spouses marrying into the kindred are not

TABLE 6

Key Recombinants Used for Genetic Ordering of New STR Loci Developed in Our Laboratory Within the BRCA1 Region of 17q

CEPH Family	ID	M/P	Mfd15	THRA1	Mfd191	ED2	AA1	Z109	4-7	YM29	Mfd188	SCG40	42D6
13292	4	M	1	1	1	1	1	0	0	0	0	0	0
13294	4	M	1	0	0	—	0	—	—	—	0	—	—
13294	6	M	0	0	1	—	1	—	—	—	1	—	—
1333	4	M	1	1	1	—	0	—	—	0	0	—	0
1333	6	M	0	0	1	—	1	—	—	1	1	—	1
1333	3	M	0	0	1	—	—	—	1	1	1	—	1

TABLE 7

Kindred 2082 Recombinants

ID	M/P	Mfd15	Mfd191	ED2	AA1	4-7	YM29	Mfd188	SCG40	42D6
63	M	0	0	1	—	1	1	1	1	1
125	M	1	1	1	—	1	1	1	0	0
40	M	1	1	0	—	0	—	0	0	0
22	P	0	0	1	1	1	1	1	1	1

described. Thus in a given sibship, the appearance of haplotypes X and Y indicates that both haplotypes from the affected/at-risk individual were seen and neither was a breast cancer associated haplotype.

affected daughter. In this family the odds in favor of affected individuals carrying BRCA1 susceptibility alleles are extremely high; the only possible interpretations of the data are that BRCA1 is distal to Mfd191 or alternatively that the

TABLE 8

Breast Cancer Linked Haplotypes Found in the Three Kindreds

Kin.	HAP	Mfd 15	THRA1	Mfd 191	tdj 1475	ED2	AA1	Z109	CA375	4-7	YM29	Mfd 188	SCG40	6C1	42D6
1901	H1	1	5	5	3	1	4	NI	NI	1	1	3	NI	NI	1
	R2	9	2	5	6	1	4	NI	NI	1	1	3	NI	NI	1
2082	H1	3	NI	4	6	6	1	NI	NI	2	1	4	2	NI	1
	P1	3	NI	4	NI	NI	NI	NI	NI	NI	NI	4	2	NI	1
	P2	3	NI	NI	NI	NI	NI	NI	NI	NI	NI	4	NI	NI	NI
	R1	6	NI	1	5	6	1	NI	NI	2	1	4	2	NI	1
	R2	6	NI	4	6	6	1	NI	NI	2	1	4	2	NI	1
	R3	3	NI	4	NI	6	1	NI	NI	2	1	4	1	NI	7
	R4	7	NI	1	NI	1	5	NI	NI	4	6	1	2	NI	1
	R5	3	NI	4	NI	NI	NI	NI	NI	NI	2	1	NI	NI	NI
	R6	3	NI	4	3	1	2	NI	NI	1	2	2	6	NI	6
	R7	3	NI	4	3	7	1	NI	NI	1	1	3	7	NI	4
2035	HI	8	2	1	NI	5	1	1	4	3	1	6	8	2	4
	H2	8	2	1	NI	5	1	1	2	1	1	2	3	1	4
	R2	8	2	1	NI	5	1	1	2	1	1	2	3	6	1

In kindred K1901, the new markers showed no observable recombination with breast cancer susceptibility, indicating that the recombination event in this kindred most likely took place between THRA1 and ED2. Thus, no new BRCA1 localization information was obtained based upon studying the four new markers in this kindred. In kindred 2082 the key recombinant individual has inherited the linked alleles for ED2, 4–7, AA1, and YM29, and was recombinant for tdj1474 indicating that the recombination event occurred in this individual between tdj1474 and ED2/AA1.

There are three haplotypes of interest in kindred K2035, H1, H2, and R2 shown in Table 8. H1 is present in the four cases and one obligate male carrier descendant from individual 17 while H2 is present or inferred in two cases and two obligate male carriers in descendants of individual 10. R2 is identical to H2 for loci between and including Mfd15 and SCG40, but has recombined between SCG40 and 42D6. Since we have established that BRCA1 is proximal to 42D6, this H2/R2 difference adds no further localization information. H1 and R2 share an identical allele at Mfd15, THRA1, AA1, and ED2 but differ for loci presumed distal to ED2, i.e., 4–7, Mfd188, SCG40, and 6C1. Although the two haplotypes are concordant for the 5th allele for marker YM29, a marker which maps physically between 4–7 and Mfd188, it is likely that the alleles are shared identical by state rather than identical by descent since this allele is the most common allele at this locus with a frequency estimated in CEPH parents of 0.42. By contrast, the linked alleles shared at the Mfd15 and ED2 loci have frequencies of 0.04 and 0.09, respectively. They also share more common alleles at Mfd191 (frequency=0.52), THRA1, and AA1 (frequency=0.28). This is the key recombinant in the set as it is the sole recombinant in which breast cancer segregated with the proximal portion of the haplotype, thus setting the distal boundary. The evidence from this kindred therefore places the BRCA1 locus proximal to 4–7.

The recombination event in kindred 2082 which places BRCA1 distal to tdj1474 is the only one of the four events described which can be directly inferred; that is, the affected mother's genotype can be inferred from her spouse and offspring, and the recombinant haplotype can be seen in her purported recombinant is a sporadic case of ovarian cancer at age 44. Rather than a directly observable or inferred recombinant, interpretation of kindred 2035 depends on the observation of distinct 17q-haplotypes segregating in different and sometimes distantly related branches of the kindred. The observation that portions of these haplotypes have alleles in common for some markers while they differ at other markers places the BRCA1 locus in the shared region. The confidence in this placement depends on several factors: the relationship between the individuals carrying the respective haplotypes, the frequency of the shared allele, the certainty with which the haplotypes can be shown to segregate with the BRCA1 locus, and the density of the markers in the region which define the haplotype. In the case of kindred 2035, the two branches are closely related, and each branch has a number of early onset cases which carry the respective haplotype. While two of the shared alleles are common, (Mfd191, THRA1), the estimated frequencies of the shared alleles at Mfd15, AA1, and ED2 are 0.04, 0.28, and 0.09, respectively. It is therefore highly likely that these alleles are identical by descent (derived from a common ancestor) rather than identical by state (the same allele but derived from the general population).

EXAMPLE 6

Refined Physical Mapping Studies Place the BRCA1 Gene in a Region Flanked by tdj1474 and U5R

Since its initial localization to chromosome 17q in 1990 (Hall et al., 1990) a great deal of effort has gone into localizing the BRCA1 gene to a region small enough to allow implementation of effective positional cloning strategies to isolate the gene. The BRCA1 locus was first localized to the interval Mfd15 (D17S250)-42D6 (D17S588) by multipoint linkage analysis (Easton et al., 1993) in the collaborative Breast Cancer Linkage Consortium dataset consisting of 214 families collected worldwide. Subsequent refinements of the localization have been based upon individual recombinant events in specific families. The region THRA1-D17S183 was defined by Bowcock et al., 1993; and the region THRA1-D17S78 was defined by Simard et al., 1993.

We further showed that the BRCA1 locus must lie distal to the marker Mfd191 (D17S776) (Goldgar et al., 1994). This marker is known to lie distal to THRA1 and RARA. The smallest published region for the BRCA1 locus is thus between D17S776 and D17S78. This region still contains approximately 1.5 million bases of DNA, making the isolation and testing of all genes in the region a very difficult task. We have therefore undertaken the tasks of constructing a physical map of the region, isolating a set of polymorphic STR markers located in the region, and analyzing these new markers in a set of informative families to refine the location of the BRCA1 gene to a manageable interval.

Four families provide important genetic evidence for localization of BRCA1 to a sufficiently small region for the application of positional cloning strategies. Two families (K2082, K1901) provide data relating to the proximal boundary for BRCA1 and the other two (K2035, K1813) fix the distal boundary. These families are discussed in detail below. A total of 15 Short Tandem Repeat markers assayable by PCR were used to refine this localization in the families studied. These markers include DS17S7654, DS17S975, tdj1474, and tdj1239. Primer sequences for these markers are provided in SEQ ID NO:3 and SEQ ID NO:4 for DS17S754; in SEQ ID NO:5 and SEQ ID NO:6 for DS17S975; in SEQ ID NO:7 and SEQ ID NO:8 for tdj1474; and, in SEQ ID NO:9 and SEQ ID NO:10 for tdj1239.

Kindred 2082

Kindred 2082 is the largest BRCA1-linked breast/ovarian cancer family studied to date. It has a LOD score of 8.6, providing unequivocal evidence for 17q linkage. This family has been previously described and shown to contain a critical recombinant placing BRCA1 distal to MFD191 (D17S776). This recombinant occurred in a woman diagnosed with ovarian cancer at age 45 whose mother had ovarian cancer at age 63. The affected mother was deceased; however, from her children, she could be inferred to have the linked haplotype present in the 30 other linked cases in the family in the region between Mfd15 and Mfd188. Her affected daughter received the linked allele at the loci ED2, 4–7, and Mfd188, but received the allele on the non-BRCA1 chromosome at Mfd15 and Mfd191. In order to further localize this recombination breakpoint, we tested DNA from the key members of this family for the following markers derived from physical mapping resources: tdj1474, tdj1239, CF4, D17S855. For the markers tdj1474 and CF4, the affected daughter did not receive the linked allele. For the STR locus tdj1239, however, the mother could be inferred to be informative and her daughter did receive the BRCA1-associated allele. D17S855 was not informative in this family. Based on this analysis, the order is 17q centromere—Mfd191—17HSD—CF4—tdj1474—tdj1239—D17S855—ED2—4–7—Mfd188—17q telomere. The recombinant described above therefore places BRCA1 distal to tdj1474, and the breakpoint is localized to the interval between tdj1474 and tdj1239. The only alternative explanation for the data in this family other than that of BRCA1 being located distal to tdj1474, is that the ovarian cancer present in the recombinant individual is caused by reasons independent of the BRCA1 gene. Given that ovarian cancer diagnosed before age 50 is rare, this alternate explanation is exceedingly unlikely.

Kindred 1901

Kindred 1901 is an early-onset breast cancer family with 7 cases of breast cancer diagnosed before 50, 4 of which were diagnosed before age 40. In addition, there were three cases of breast cancer diagnosed between the ages of 50 and 70. One case of breast cancer also had ovarian cancer at age 61. This family currently has a LOD score of 1.5 with D17S855. Given this linkage evidence and the presence of at lease one ovarian cancer case, this family has a posterior probability of being due to BRCA1 of over 0.99. In this family, the recombination comes from the fact that an individual who is the brother of the ovarian cancer case from which the majority of the other cases descend, only shares a portion of the haplotype which is cosegregating with the other cases in the family. However, he passed this partial haplotype to his daughter who developed breast cancer at age 44. If this case is due to the BRCA1 gene, then only the part of the haplotype shared between this brother and his sister can contain the BRCA1 gene. The difficulty in interpretation of this kind of information is that while one can be sure of the markers which are not shared and therefore recombinant, markers which are concordant can either be shared because they are non-recombinant, or because their parent was homozygous. Without the parental genotypic data it is impossible to discriminate between these alternatives. Inspection of the haplotype in K1901, shows that he does not share the linked allele at Mfd15 (D17S250), THRA1, CF4 (D17S1320), and tdj1474 (17DS1321). He does share the linked allele at Mfd191 (D17S776), ED2 (D17S1327), tdj1239 (D17S1328), and Mfd188 (D17S579). Although the allele shared at Mfd191 is relatively rare (0.07), we would presume that the parent was homozygous since they are recombinant with markers located nearby on either side, and a double recombination event in this region would be extremely unlikely. Thus the evidence in this family would also place the BRCA1 locus distal to tdj1474. However, the lower limit of this breakpoint is impossible to determine without parental genotype information. It is intriguing that the key recombinant breakpoint in this family confirms the result in Kindred 2082. As before, the localization information in this family is only meaningful if the breast cancer was due to the BRCA1 gene. However, her relatively early age at diagnosis (44) makes this seem very likely since the risk of breast cancer before age 45 in the general population is low (approximately 1%).

Kindred 2035

This family is similar to K1901 in that the information on the critical recombinant events is not directly observed but is inferred from the observation that the two haplotypes which are cosegregating with the early onset breast cancer in the two branches of the family appear identical for markers located in the proximal portion of the 17q BRCA1 region but differ at more distal loci. Each of these two haplotypes occurs in at least four cases of early-onset or bilateral breast cancer. The overall LOD score with ED2 in this family is 2.2, and considering that there is a case of ovarian cancer in the family (indicating a prior probability of BRCA1 linkage of 80%), the resulting posterior probability that this family is linked to BRCA1 is 0.998. The haplotypes are identical for the markers Mfd15, THRA1, Mfd191, ED2, AA1, D17S858 and D17S902. The common allele at Mfd15 and ED2 are both quite rare, indicating that this haplotype is shared identical by descent. The haplotypes are discordant, however, for CA375, 4–7, and Mfd188, and several more distal markers. This indicates that the BRCA1 locus must lie above the marker CA-375. This marker is located approximately 50 kb below D17S78, so it serves primarily as additional confirmation of this previous lower boundary as reported in Simard et al. (1993).

Kindred 1813

Kindred 1813 is a small family with four cases of breast cancer diagnosed at very early ages whose mother also had breast cancer diagnosed at an early age and ovarian cancer

some years later. This situation is somewhat complicated by the fact the four cases appear to have three different fathers, only one of whom has been genotyped. However, by typing a number of different markers in the BRCA1 region as well as highly polymorphic markers elsewhere in the genome, the paternity of all children in the family has been determined with a high degree of certainty. This family yields a maximum multipoint LOD score of 0.60 with 17q markers and, given that there is at least one case of ovarian cancer, results in a posterior probability of being a BRCA1 linked family of 0.93. This family contains a directly observable recombination event in individual 18 (see FIG. **5** in Simard et al., *Human Mol. Genet.* 2:1193–1199 (1993)), who developed breast cancer at age 34. The genotype of her affected mother at the relevant 17q loci can be inferred from her genotypes, her affected sister's genotypes, and the genotypes of three other unaffected siblings. Individual 18 inherits the BRCA1-linked alleles for the following loci: Mfd15, THRA1, D17S800, D17S855, AA1, and D17S931. However, for markers below D17S931, i.e., U5R, vrs31, D17S858, and D17S579, she has inherited the alleles located on the non-disease bearing chromosome. The evidence from this family therefore would place the BRCA1 locus proximal to the marker U5R. Because of her early age at diagnosis (34) it is extremely unlikely that the recombinant individual's cancer is not due to the gene responsible for the other cases of breast/ovarian cancer in this family; the uncertainty in this family comes from our somewhat smaller amount of evidence that breast cancer in this family is due to BRCA1 rather than a second, as yet unmapped, breast cancer susceptibility locus.

Size of the region containing BRCA1

Based on the genetic data described in detail above, the BRCA1 locus must lie in the interval between the markers tdj1474 and U5R, both of which were isolated in our laboratory. Based upon the physical maps shown in FIGS. **2** and **3**, we can try to estimate the physical distance between these two loci. It takes approximately 14 P1 clones with an average insert size of approximately 80 kb to span the region. However, because all of these P1s overlap to some unknown degree, the physical region is most likely much smaller than 14 times 80 kb. Based on restriction maps of the clones covering the region, we estimate the size of the region containing BRCA1 to be approximately 650 kb.

EXAMPLE 7

Identification of Candidate cDNA Clones for the BRCA1 Locus by Genomic Analysis of the Contig Region

Complete screen of the plausible region.

The first method to identify candidate cDNAs, although labor intensive, used known techniques. The method comprised the screening of cosmids and P1 and BAC clones in the contig to identify putative coding sequences. The clones containing putative coding sequences were then used as probes on filters of cDNA libraries to identify candidate cDNA clones for future analysis. The clones were screened for putative coding sequences by either of two methods.

Zoo blots.

The first method for identifying putative coding sequences was by screening the cosmid and P1 clones for sequences conserved through evolution across several species. This technique is referred to as "zoo blot analysis" and is described by Monaco, 1986. Specifically, DNAs from cow, chicken, pig, mouse and rat were digested with the restriction enzymes EcoRI and HindIII (8 μg of DNA per enzyme). The digested DNAs were separated overnight on an 0.7% gel at 20 volts for 16 hours (14 cm gel), and the DNA transferred to Nylon membranes using standard Southern blot techniques. For example, the zoo blot filter was treated at 65° C. in 0.1×SSC, 0.5% SDS, and 0.2M Tris, pH 8.0, for 30 minutes and then blocked overnight at 42° C. in 5× SSC, 10% PEG 8000, 20 mM $NaPO_4$ pH 6.8, 100 μg/ml Salmon Sperm DNA, 1× Denhardt's, 50% formamide, 0.1% SDS, and 2 μg/ml C_0t–1DNA.

The cosmid and P1 clones to be analyzed were digested with a restriction enzyme to release the human DNA from the vector DNA. The DNA was separated on a 14 cm, 0.5% agarose gel run overnight at 20 volts for 16 hours. The human DNA bands were cut out of the gel and electroeluted from the gel wedge at 100 volts for at least two hours in 0.5× Tris Acetate buffer (Maniatis et al., 1982). The eluted Not I digested DNA (~15 kb to 25 kb) was then digested with EcoRI restriction enzyme to give smaller fragments (~0.5 kb to 5.0 kb) which melt apart more easily for the next step of labeling the DNA with radionucleotides. The DNA fragments were labeled by means of the hexamer random prime labeling method (Boehringer-Mannheim, Cat. #1004760). The labeled DNA was spermine precipitated (add 100 μl TE, 5 μl 0.1M spermine, and 5 μl of 10 mg/ml salmon sperm DNA) to remove unincorporated radionucleotides. The labeled DNA was then resuspended in 100 μl TE, 0.5M NaCl at 65° C. for 5 minutes and then blocked with Human C_0t–1 DNA for 2–4 hrs. as per the manufacturer's instructions (Gibco/BRL, Cat. #5279SA). The C_0t–1 blocked probe was incubated on the zoo blot filters in the blocking solution overnight at 42° C. The filters were washed for 30 minutes at room temperature in 2×SSC, 0.1% SDS, and then in the same buffer for 30 minutes at 55° C. The filters were then exposed 1 to 3 days at –70° C. to Kodak XAR-5 film with an intensifying screen. Thus, the zoo blots were hybridized with either the pool of Eco-R1 fragments from the insert, or each of the fragments individually.

HTF island analysis.

The second method for identifying cosmids to use as probes on the cDNA libraries was HTF island analysis. Since the pulsed-field map can reveal HTF islands, cosmids that map to these HTF island regions were analyzed with priority. HTF islands are segments of DNA which contain a very high frequency of unmethylated CpG dinucleotides (Tonolio et al., 1990) and are revealed by the clustering of restriction sites of enzymes whose recognition sequences include CpG dinucleotides. Enzymes known to be useful in HTF-island analysis are AscI, NotI, BssHII, EagI, SacII, NaeI, NarI, SmaI, and MluI (Anand, 1992). A pulsed-field map was created using the enzymes NotI, NruI, EagI, SacII, and SalI, and two HTF islands were found. These islands are located in the distal end of the region, one being distal to the GP2B locus, and the other being proximal to the same locus, both outside the BRCA1 region. The cosmids derived from the YACs that cover these two locations were analyzed to identify those that contain these restriction sites, and thus the HTF islands.

cDNA screening.

Those clones that contain HTF islands or show hybridization to other species DNA besides human are likely to contain coding sequences. The human DNA from these clones was isolated as whole insert or as EcoRI fragments and labeled as described above. The labeled DNA was used to screen filters of various cDNA libraries under the same conditions as the zoo blots except that the CDNA filters

undergo a more stringent wash of 0.1×SSC, 0.1% SDS at 65° C. for 30 minutes twice.

Most of the CDNA libraries used to date in our studies (libraries from normal breast tissue, breast tissue from a woman in her eighth month of pregnancy and a breast malignancy) were prepared at Clonetech, Inc. The CDNA library generated from breast tissue of an 8 month pregnant woman is available from Clonetech (Cat. #HL1037a) in the Lambda gt-10 vector, and is grown in C600OHfl bacterial host cells. Normal breast tissue and malignant breast tissue samples were isolated from a 37 year old Caucasian female and one-gram of each tissue was sent to Clonetech for mRNA processing and cDNA library construction. The latter two libraries were generated using both random and oligo-dT priming, with size selection of the final products which were then cloned into the Lambda Zap II vector, and grown in XL1-blue strain of bacteria as described by the manufacturer. Additional tissue-specific CDNA libraries include human fetal brain (Stratagene, Cat. 936206), human testis (Clonetech Cat. HL3024), human thymus (Clonetech Cat. HL1127n), human brain (Clonetech Cat. HL11810), human placenta (Clonetech Cat 1075b), and human skeletal muscle (Clonetech Cat. HL 1124b).

The CDNA libraries were plated with their host cells on NZCYM plates, and filter lifts are made in duplicate from each plate as per Maniatis et al. (1982). Insert (human) DNA from the candidate genomic clones was purified and radioactively labeled to high specific activity. The radioactive DNA was then hybridized to the CDNA filters to identify those cDNAs which correspond to genes located within the candidate cosmid clone. cDNAs identified by this method were picked, replated, and screened again with the labeled clone insert or its derived EcoRl fragment DNA to verify their positive status. Clones that were positive after this second round of screening were then grown up and their DNA purified for Southern blot analysis and sequencing. Clones were either purified as plasmid through in vivo excision of the plasmid from the Lambda vector as described in the protocols from the manufacturers, or isolated from the Lambda vector as a restriction fragment and subdloned into plasmid vector.

The Southern blot analysis was performed in duplicate, one using the original genomic insert DNA as a probe to verify that cDNA insert contains hybridizing sequences. The second blot was hybridized with cDNA insert DNA from the largest cDNA clone to identify which clones represent the same gene. All cDNAs which hybridize with the genomic clone and are unique were sequenced and the DNA analyzed to determine if the sequences represent known or unique genes. All cDNA clones which appear to be unique were further analyzed as candidate BRCA1 loci. Specifically, the clones are hybridized to Northern blots to look for breast specific expression and differential expression in normal versus breast tumor RNAs. They are also analyzed by PCR on clones in the BRCA1 region to verify their location. To map the extent of the locus, full length cDNAs are isolated and their sequences used as PCR probes on the YACs and the clones surrounding and including the original identifying clones. Intron-exon boundaries are then further defined through sequence analysis.

We have screened the normal breast, 8 month pregnant breast and fetal brain cDNA libraries with zoo blot-positive Eco R1 fragments from cosmid BAC and P1 clones in the region. Potential BRCA1 cDNA clones were identified among the three libraries. Clones were picked, replated, and screened again with the original probe to verify that they were positive.

Analysis of hybrid-selected cDNA.

cDNA fragments obtained from direct selection were checked by Southern blot hybridization against the probe DNA to verify that they originated from the contig. Those that passed this test were sequenced in their entirety. The set of DNA sequences obtained in this way were then checked against each other to find independent clones that overlapped. For example, the clones 694-65, 1240-1 and 1240-33 were obtained independently and subsequently shown to derive from the same contiguous cDNA sequence which has been named EST:489:1.

Analysis of candidate clones.

One or more of the candidate genes generated from above were sequenced and the information used for identification and classification of each expressed gene. The DNA sequences were compared to known genes by nucleotide sequence comparisons and by translation in all frames followed by a comparison with known amino acid sequences. This was accomplished using Genetic Data Environment (GDE) version 2.2 software and the Basic Local Alignment Search Tool (Blast) series of client/server software packages (e.g., BLASTN 1.3.13 MP), for sequence comparison against both local and remote sequence databases (e.g., GenBank), running on Sun SPARC workstations. Sequences reconstructed from collections of cDNA clones identified with the cosmids and P1s have been generated. All candidate genes that represented new sequences were analyzed further to test their candidacy for the putative BRCA1 locus.

Mutation screening.

To screen for mutations in the affected pedigrees, two different approaches were followed. First, genomic DNA isolated from family members known to carry the susceptibility allele of BRCA1 was used as a template for amplification of candidate gene sequences by PCR. If the PCR primers flank or overlap an intron/exon boundary, the amplified fragment will be larger than predicted from the cDNA sequence or will not be present in the amplified mixture. By a combination of such amplification experiments and sequencing of P1, BAC or cosmid clones using the set of designed primers it is possible to establish the intron/exon structure and ultimately obtain the DNA sequences of genomic DNA from the pedigrees.

A second approach that is much more rapid if the intron/exon structure of the candidate gene is complex involves sequencing fragments amplified from pedigree lymphocyte cDNA. cDNA synthesized from lymphocyte mRNA extracted from pedigree blood was used as a substrate for PCR amplification using the set of designed primers. If the candidate gene is expressed to a significant extent in lymphocytes, such experiments usually produce amplified fragments that can be sequenced directly without knowledge of intron/exon junctions.

The products of such sequencing reactions were analyzed by gel electrophoresis to determine positions in the sequence that contain either mutations such as deletions or insertions, or base pair substitutions that cause amino acid changes or other detrimental effects.

Any sequence within the BRCA1 region that is expressed in breast is considered to be a candidate gene for BRCA1. Compelling evidence that a given candidate gene corresponds to BRCA1 comes from a demonstration that pedigree families contain defective alleles of the candidate.

EXAMPLE 8

Identification of BRCA1

Identification of BRCA1.

Using several strategies, a detailed map of transcripts was developed for the 600 kb region of 17q21 between D17S1321 and D17S1324. Candidate expressed sequences were defined as DNA sequences obtained from: 1) direct screening of breast, fetal brain, or lymphocyte CDNA libraries, 2) hybrid selection of breast, lymphocyte or ovary cDNAs, or 3) random sequencing of genomic DNA and prediction of coding exons by XPOUND (Thomas and Skolnick, 1994). These expressed sequences in many cases were assembled into contigs composed of several independently identified sequences. Candidate genes may comprise more than one of these candidate expressed sequences. Sixty-five candidate expressed sequences within this region were identified by hybrid selection, by direct screening of CDNA libraries, and by random sequencing of P1 subclones. Expressed sequences were characterized by transcript size, DNA sequence, database comparison, expression pattern, genomic structure, and, most importantly, DNA sequence analysis in individuals from kindreds segregating 17q-linked breast and ovarian cancer susceptibility.

Three independent contigs of expressed sequence, 1141:1 (649 bp), 694:5 (213 bp) and 754:2 (1079 bp) were isolated and eventually shown to represent portions of BRCA1. When ESTs for these contigs were used as hybridization probes for Northern analysis, a single transcript of approximately 7.8 kb was observed in normal breast mRNA, suggesting that they encode different portions of a single gene. Screens of breast, fetal brain, thymus, testes, lymphocyte and placental cDNA libraries and PCR experiments with breast mRNA linked the 1141:1, 694:5 and 754:2 contigs. 5' RACE experiments with thymus, testes, and breast mRNA extended the contig to the putative 5' end, yielding a composite full length sequence. PCR and direct sequencing of P1s and BACs in the region were used to identify the location of introns and allowed the determination of splice donor and acceptor sites. These three expressed sequences were merged into a single transcription unit that proved in the final analysis to be BRCA1. This transcription unit is located adjacent to D17S855 in the center of the 600 kb region (FIG. 4).

Combination of sequences obtained from cDNA clones, hybrid selection sequences, and amplified PCR products allowed construction of a composite full length BRCA1 cDNA (SEQ ID NO:1). The sequence of the BRCA1 cDNA (up through the stop codon) has also been deposited with GenBank and assigned accession number U-14680. This deposited sequence is incorporated herein by reference. The cDNA clone extending farthest in the 3' direction contains a poly(A) tract preceded by a polyadenylation signal. Conceptual translation of the cDNA revealed a single long open reading frame of 208 kilodaltons (amino acid sequence: SEQ ID NO:2) with a potential initiation codon flanked by sequences resembling the Kozak consensus sequence (Kozak, 1987). Smith-Waterman (Smith and Waterman, 1981) and BLAST (Altschul et al., 1990) searches identified a sequence near the amino terminus with considerable homology to zinc-finger domains (FIG. 5). This sequence contains cysteine and histidine residues present in the consensus C3HC4 zinc-finger motif and shares multiple other residues with zinc-finger proteins in the databases. The BRCA1 gene is composed of 23 coding exons arrayed over more than 100 kb of genomic DNA (FIG. 6). Northern blots using fragments of the BRCA1 cDNA as probes identified a single transcript of about 7.8 kb, present most abundantly in breast, thymus and testis, and also present in ovary (FIG. 7). Four alternatively spliced products were observed as independent cDNA clones; 3 of these were detected in breast and 2 in ovary mRNA (FIG. 6). A PCR survey from tissue cDNAs further supports the idea that there is considerable heterogeneity near the 5' end of transcripts from this gene; the molecular basis for the heterogeneity involves differential choice of the first splice donor site, and the changes detected all alter the transcript in the region 5' of the identified start codon. We have detected six potential alternate splice donors in this 5' untranslated region, with the longest deletion being 1,155 bp. The predominant form of the BRCA1 protein in breast and ovary lacks exon 4. The nucleotide sequence for BRCA1 exon 4 is shown in SEQ ID NO:11, with the predicted amino acid sequence shown in SEQ ID NO:12.

Additional 5' sequence of BRCA1 genomic DNA is set forth in SEQ ID NO:13. The G at position 1 represents the potential start site in testis. The A in position 140 represents the potential start site in somatic tissue. There are six alternative splice forms of this 5' sequence as shown in FIG. **8**. The G at position 356 represents the canonical first splice donor site. The G at position 444 represents the first splice donor site in two clones (testis 1 and testis 2). The G at position 889 represents the first splice donor site in thymus 3. A fourth splice donor site is the G at position 1230. The T at position 1513 represents the splice acceptor site for all of the above splice donors. A fifth alternate splice form has a first splice donor site at position 349 with a first acceptor site at position 591 and a second splice donor site at position 889 and a second acceptor site at position 1513. A sixth alternate form is unspliced in this 5' region. The A at position 1532 is the canonical start site, which appears at position 120 of SEQ ID NO:1. Partial genomic DNA sequences determined for BRCA1 are set forth in FIGS. **10A–10H** and SEQ ID Numbers:14–34. The lower case letters (in FIGS. **10A–10H**) denote intron sequence while the upper case letters denote exon sequence. Indefinite intervals within introns are designated with vvvvvvvvvvvvv in FIGS. **10A–10H**. The intron/exon junctions are shown in Table 9. The CAG found at the 5' end of exons 8 and 14 is found in some cDNAs but not in others. Known polymorphic sites are shown in FIGS. **10A–10H** in boldface type and are underlined.

TABLE 9

Exon	Base Position*			Intron Borders	
No.	5'	3'	Length	5'	3'
e1	1	100	100	GATAAATTAAAACTGCGACTGCGCGGCGTG[35]*	GTAGTAGAGTCCCGGGAAAGGGACAGGGGG[36]
e2	101	199	99	ATATATATATGTTTTTCTAATGTGTTAAAG[37]	GTAAGTCAGCACAAGAGTGTATTAATTTGG[38]
e3	200	253	54	TTTCTTTTTCTCCCCCCCCTACCCTGCTAG[39]	GTAAGTTTGAATGTGTTATGTGGCTCCATT[40]

TABLE 9-continued

Exon	Base Position*			Intron Borders	
No.	5'	3'	Length	5'	3'
e4	***	***	111	AGCTACTTTTTTTTTTTTTTTTTGAGACAG[41]	GTAAGTGCACACCACCATATCCAGCTAAAT[42]
e5	254	331	78	AATTGTTCTTTCTTTCTTTATAATTTATAG[43]	GTATATAATTTGGTAATGATGCTAGGTTGG[44]
e6	332	420	89	GAGTGTGTTTCTCAAACAATTTAATTTCAG[45]	GTAAGTGTTGAATATCCCAAGAATGACACT[46]
e7	421	560	140	AAACATAATGTTTTCCCTTGTATTTTACAG[47]	GTAAAACCATTTGTTTTCTTCTTCTTCTTC[48]
e8	561	666	106	TGCTTGACTGTTCTTTACCATACTGTTTAG[49]	GTAAGGGTCTCAGGTTTTTTAAGTATTTAA[50]
e9	667	712	46	TGATTTATTTTTTGGGGGGAAATTTTTTAG[51]	GTGAGTCAAAGAGAACCTTTGTCTATGAAG[52]
e10	713	789	77	TCTTATTAGGACTCTGTCTTTTCCCTATAG[53]	GTAATGGCAAAGTTTGCCAACTTAACAGGC[54]
e11	790	4215	3426	GAGTACCTTGTTATTTTTGTATATTTTCAG[55]	GTATTGGAACCAGGTTTTTGTGTTTGCCCC[56]
e12	4216	4302	87	ACATCTGAACCTCTGTTTTTGTTATTTAAG[57]	AGGTAAAAAGCGTGTGTGTGTGTGCACATG[58]
e13	4303	4476	174	CATTTTCTTGGTACCATTTATCGTTTTTGA[59]	GTGTGTATTGTTGGCCAAACACTGATATCT[60]
e14	4477	4603	127	AGTAGATTTGTTTTCTCATTCCATTTAAAG[61]	GTAAGAAACATCAATGTAAAGATGCTGTGG[62]
e15	4604	4794	191	ATGGTTTTCTCCTTCCATTTATCTTTCTAG[63]**	GTAATATTTCATCTGCTGTATTGGAACAAA[64]
e16	4795	5105	311	TGTAAATTAAACTTCTCCCATTCCTTTCAG[65]	GTGAGTGTATCCATATGTATCTCCCTAATG[66]
e17	5106	5193	88	ATGATAATGGAATATTTGATTTAATTTCAG[67]	GTATACCAAGAACCTTTACAGAATACCTTG[68]
e18	5194	5271	78	CTAATCCTTTGAGTGTTTTTCATTCTGCAG[69]	GTAAGTATAATACTATTTCTCCCCTCCTCC[70]
e19	5272	5312	41	TGTAACCTGTCTTTTCTATGATCTCTTTAG[71]	GTAAGTACTTGATGTTACAAACTAACCAGA[72]
e20	5313	5396	84	TCCTGATGGGTTGTGTTTGGTTTCTTTCAG[73]	GTAAAGCTCCCTCCCTCAAGTTGACAAAAA[74]
e21	5397	5451	55	CTGTCCCTCTCTCTTCCTCTCTTCTTCCAG[75]	GTAAGAGCCTGGGAGAACCCCAGAGTTCCA[76]
e22	5452	5525	74	AGTGATTTTACATGTAAATGTCCATTTTAG[77]	GTAAGTATTGGGTGCCCTGTCAGTGTGGGA[78]
e23	5526	5586	61	TTGAATGCTCTTTCCTTCCTGGGGATCCAG[79]	GTAAGGTGCCTCGCATGTACCTGTGCTATT[80]
e24	5587	5914	328	CTAATCTCTGCTTGTGTTCTCTGTCTCCAG[81]	

*Base numbers in SEQ ID NO: 1.
**Numbers in superscript refer to SEQ ID NOS.
***e4 from SEQ ID NO: 11.

Low stringency blots in which genomic DNA from organisms of diverse phylogenetic background were probed with BRCA1 sequences that lack the zinc-finger region revealed strongly hybridizing fragments in human, monkey, sheep and pig, and very weak hybridization signals in rodents. This result indicates that, apart from the zinc-finger domain, BRCA1 is conserved only at a moderate level through evolution.

Germline BRCA1 mutations in 17q-linked kindreds.

The most rigorous test for BRCA1 candidate genes is to search for potentially disruptive mutations in carrier individuals from kindreds that segregate 17q-linked susceptibility to breast and ovarian cancer. Such individuals must contain BRCA1 alleles that differ from the wildtype sequence. The set of DNA samples used in this analysis consisted of DNA from individuals representing 8 different BRCA1 kindreds (Table 10).

TABLE 10

KINDRED DESCRIPTIONS AND ASSOCIATED LOD SCORES

	Cases (n)			Sporadic	LOD	
Kindred	Br	Br < 50	Ov	Cases[1] (n)	Score	Marker(s)
2082	31	20	22	7	9.49	D17S1327
2099	22	14	2*	0	2.36	D17S800/D17S855[2]
2035	10	8	1*	0	2.25	D17S1327
1901	10	7	1*	0	1.50	D17S855
1925	4	3	0	0	0.55	D17S579
1910	5	4	0	0	0.36	D17S579/D17S250[2]
1927	5	4	0	1	−0.44	D17S250
1911	8	5	0	2	−0.20	D17S250

[1]Number of women with breast cancer (diagnosed under age 50) or ovarian cancer (diagnosed at any age) who do not share the BRCA1-linked haplotype segregating in the remainder of the cases in the kindred.
[2]Multipoint LOD score calculated using both markers
*kindred contains one individual who had both breast and ovarian cancer; this individual is counted as a breast cancer case and as an ovarian cancer case.

The logarithm of the odds (LOD) scores in these kindreds range from 9.49 to −0.44 for a set of markers in 17q21. Four of the families have convincing LOD scores for linkage, and 4 have low positive or negative LOD scores. The latter kindreds were included because they demonstrate haplotype sharing at chromosome 17q21 for at least 3 affected members. Furthermore, all kindreds in the set display early age of breast cancer onset and 4 of the kindreds include at least one case of ovarian cancer, both hallmarks of BRCA1 kindreds. One kindred, 2082, has nearly equal incidence of breast and ovarian cancer, an unusual occurrence given the relative rarity of ovarian cancer in the population. All of the kindreds except two were ascertained in Utah. K2035 is from the midwest. K2099 is an African-American kindred from the southern U.S.A.

In the initial screen for predisposing mutations in BRCA1, DNA from one individual who carries the predisposing haplotype in each kindred was tested. The 23 coding exons and associated splice junctions were amplified either from genomic DNA samples or from CDNA prepared from lymphocyte mRNA. When the amplified DNA sequences were compared to the wildtype sequence, 4 of the 8 kindred samples were found to contain sequence variants (Table 11).

TABLE 11

PREDISPOSING MUTATIONS

Kindred Number	Mutation	Coding Effect	Location*
2082	C→T	Gln→Stop	4056
1910	extra C	frameshift	5385
2099	T→G	Met→Arg	5443
2035	?	loss of transcript	
1901	11 bp deletion	frameshift	189

*In Sequence ID NO: 1

All four sequence variants are heterozygous and each appears in only one of the kindreds. Kindred 2082 contains a nonsense mutation in exon 11 (FIG. 9A). Kindred 1910 contains a single nucleotide insertion in exon 20 (FIG. 9B), and Kindred 2099 contains a missense mutation in exon 21, resulting in a Met→Arg substitution. The frameshift and

GeneDX 1009, pg. 455

nonsense mutations are likely disruptive to the function of the BRCA1 product. The peptide encoded by the frameshift allele in Kindred 1910 would contain an altered amino acid sequence beginning 108 residues from the wildtype C-terminus. The peptide encoded by the frameshift allele in Kindred 1901 would contain an altered amino acid sequence beginning with the 24th residue from the wildtype N-terminus. The mutant allele in Kindred 2082 would encode a protein missing 551 residues from the C-terminus. The missense substitution observed in Kindred 2099 is potentially disruptive as it causes the replacement of a small hydrophobic amino acid (Met), by a large charged residue (Arg). Eleven common polymorphisms were also identified, 8 in coding sequence and 3 in introns.

The individual studied in Kindred 2035 evidently contains a regulatory mutation in BRCA1. In her cDNA, a polymorphic site (A→G at base 3667) appeared homozygous, whereas her genomic DNA revealed heterozygosity at this position (FIG. 9C). A possible explanation for this observation is that mRNA from her mutated BRCA1 allele is absent due to a mutation that affects its production or stability. This possibility was explored further by examining 5 polymorphic sites in the BRCA1 coding region, which are separated by as much as 3.5 kb in the BRCA1 transcript. In all cases where her genomic DNA appeared heterozygous for a polymorphism, cDNA appeared homozygous. In individuals from other kindreds and in non-haplotype carriers in Kindred 2035, these polymorphic sites could be observed as heterozygous in cDNA, implying that amplification from cDNA was not biased in favor of one allele. This analysis indicates that a BRCA1 mutation in Kindred 2035 either prevents transcription or causes instability or aberrant splicing of the BRCA1 transcript.

Cosegregation of BRCA1 mutations with BRCA1 haplotypes and population frequency analysis.

In addition to potentially disrupting protein function, two criteria must be met for a sequence variant to qualify as a candidate predisposing mutation. The variant must: 1) be present in individuals from the kindred who carry the predisposing BRCA1 haplotype and absent in other members of the kindred, and 2) be rare in the general population.

Each mutation was tested for cosegregation with BRCA1. For the frameshift mutation in Kindred 1910, two other haplotype carriers and one non-carrier were sequenced (FIG. 9B). Only the carriers exhibited the frameshift mutation. The C to T change in Kindred 2082 created a new AvrII restriction site. Other carriers and non-carriers in the kindred were tested for the presence of the restriction site (FIG. 9A). An allele-specific oligonucleotide (ASO) was designed to detect the presence of the sequence variant in Kindred 2099. Several individuals from the kindred, some known to carry the haplotype associated with the predisposing allele, and others known not to carry the associated haplotype, were screened by ASO for the mutation previously detected in the kindred. In each kindred, the corresponding mutant allele was detected in individuals carrying the BRCA1-associated haplotype, and was not detected in noncarriers. In the case of the potential regulatory mutation observed in the individual from Kindred 2035, cDNA and genomic DNA from carriers in the kindred were compared for heterozygosity at polymorphic sites. In every instance, the extinguished allele in the cDNA sample was shown to lie on the chromosome that carries the BRCA1 predisposing allele (FIG. 9C).

To exclude the possibility that the mutations were simply common polymorphisms in the population, ASOs for each mutation were used to screen a set of normal DNA samples. Gene frequency estimates in Caucasians were based on random samples from the Utah population. Gene frequency estimates in African-Americans were based on 39 samples provided by M. Peracek-Vance which originate from African-Americans used in her linkage studies and 20 newborn Utah African-Americans. None of the 4 potential predisposing mutations was found in the appropriate control population, indicating that they are rare in the general population. Thus, two important requirements for BRCA1 susceptibility alleles were fulfilled by the candidate predisposing mutations: 1) cosegregation of the mutant allele with disease, and 2) absence of the mutant allele in controls, indicating a low gene frequency in the general population.

Phenotypic Expression of BRCA1 Mutations.

The effect of the mutations on the BRCA1 protein correlated with differences in the observed phenotypic expression in the BRCA1 kindreds. Most BRCA1 kindreds have a moderately increased ovarian cancer risk, and a smaller subset have high risks of ovarian cancer, comparable to those for breast cancer (Easton et al., 1993). Three of the four kindreds in which BRCA1 mutations were detected fall into the former category, while the fourth (K2082) falls into the high ovarian cancer risk group. Since the BRCA1 nonsense mutation found in K2082 lies closer to the amino terminus than the other mutations detected, it might be expected to have a different phenotype. In fact, Kindred K2082 mutation has a high incidence of ovarian cancer, and a later mean age at diagnosis of breast cancer cases than the other kindreds (Goldgar et al., 1994). This difference in age of onset could be due to an ascertainment bias in the smaller, more highly penetrant families, or it could reflect tissue-specific differences in the behavior of BRCA1 mutations. The other 3 kindreds that segregate known BRCA1 mutations have, on average, one ovarian cancer for every 10 cases of breast cancer, but have a high proportion of breast cancer cases diagnosed in their late 20's or early 30's. Kindred 1910, which has a frameshift mutation, is noteworthy because three of the four affected individuals had bilateral breast cancer, and in each case the second tumor was diagnosed within a year of the first occurrence. Kindred 2035, which segregates a potential regulatory BRCA1 mutation, might also be expected to have a dramatic phenotype. Eighty percent of breast cancer cases in this kindred occur under age 50. This figure is as high as any in the set, suggesting a BRCA1 mutant allele of high penetrance (Table 10).

Although the mutations described above clearly are deleterious, causing breast cancer in women at very young ages, each of the four kindreds with mutations includes at least one woman who carries the mutation who lived until age 80 without developing a malignancy. It will be of utmost importance in the studies that follow to identify other genetic or environmental factors that may ameliorate the effects of BRCA1 mutations.

In four of the eight putative BRCA1-linked kindreds, potential predisposing mutations were not found. Three of these four have LOD scores for BRCA1-linked markers of less than 0.55. Thus, these kindreds may not in reality segregate BRCA1 predisposing alleles. Alternatively, the mutations in these four kindreds may lie in regions of BRCA1 that, for example, affect the level of transcript and therefore have thus far escaped detection.

Role of BRCA1 in Cancer.

Most tumor suppressor genes identified to date give rise to protein products that are absent, nonfunctional, or reduced in function. The majority of TP53 mutations are missense;

some of these have been shown to produce abnormal p53 molecules that interfere with the function of the wildtype product (Shaulian et al., 1992; Srivastava et al., 1993). A similar dominant negative mechanism of action has been proposed for some adenomatous polyposis coli (APC) alleles that produce truncated molecules (Su et al., 1993), and for point mutations in the Wilms' tumor gene (WT1) that alter DNA binding of the protein (Little et al., 1993). The nature of the mutations observed in the BRCA1 coding sequence is consistent with production of either dominant negative proteins or nonfunctional proteins. The regulatory mutation inferred in Kindred 2035 cannot be a dominant negative; rather, this mutation likely causes reduction or complete loss of BRCA1 expression from the affected allele.

The BRCA1 protein contains a C_3HC_4 zinc-finger domain, similar to those found in numerous DNA binding proteins and implicated in zinc-dependent binding to nucleic acids. The first 180 amino acids of BRCA1 contain five more basic residues than acidic residues. In contrast, the remainder of the molecule is very acidic, with a net excess of 70 acidic residues. The excess negative charge is particularly concentrated near the C-terminus. Thus, one possibility is that BRCA1 encodes a transcription factor with an N-terminal DNA binding domain and a C-terminal transactivational "acidic blob" domain. Interestingly, another familial tumor suppressor gene, WT1, also contains a zinc-finger motif (Haber et al., 1990). Many cancer predisposing mutations in WT1 alter zinc-finger domains (Little et al., 1993; Haber et al., 1990; Little et al., 1992). WT1 encodes a transcription factor, and alternative splicing of exons that encode parts of the zinc-finger domain alter the DNA binding properties of WT1 (Bickmore et al., 1992). Some alternatively spliced forms of WT1 mRNA generate molecules that act as transcriptional repressors (Drummond et al., 1994). Some BRCA1 splicing variants may alter the zinc-finger motif, raising the possibility that a regulatory mechanism similar to that which occurs in WT1 may apply to BRCA1.

EXAMPLE 9

Analysis of Tumors for BRCA1 Mutations

To focus the analysis on tumors most likely to contain BRCA1 mutations, primary breast and ovarian carcinomas were typed for LOH in the BRCA1 region. Three highly polymorphic, simple tandem repeat markers were used to assess LOH: D17S1323 and D17S855, which are intragenic to BRCA1, and D17S1327, which lies approximately 100 kb distal to BRCA1. The combined LOH frequency in informative cases (i.e., where the germline was heterozygous) was 32/72 (44%) for the breast carcinomas and 12/21 (57%) for the ovarian carcinomas, consistent with previous measurements of LOH in the region (Futreal et al., 1992b; Jacobs et al, 1993; Sato et al., 1990; Eccles et al., 1990; Cropp et al., 1994). The analysis thus defined a panel of 32 breast tumors and 12 ovarian tumors of mixed race and age of onset to be examined for BRCA mutations. The complete 5,589 bp coding region and intron/exon boundary sequences of the gene were screened in this tumor set by direct sequencing alone or by a combination of single-strand conformation analysis (SSCA) and direct sequencing.

A total of six mutations (of which two are identical) was found, one in an ovarian tumor, four in breast tumors and one in a male unaffected haplotype carrier (Table 12). One mutation, Glu1541Ter, introduced a stop codon that would create a truncated protein missing 323 amino acids at the carboxy terminus. In addition, two missense mutations were identified. These are Ala1708Glu and Met1775Arg and involve substitutions of small, hydrophobic residues by charged residues. Patients 17764 and 19964 are from the same family. In patient OV24 nucleotide 2575 is deleted and in patients 17764 and 19964 nucleotides 2993–2996 are deleted.

TABLE 12

		Predisposing Mutations			
Patient	Codon	Nucleotide Change	Amino Acid Change	Age of Onset	Family History
BT098	1541	GAG → TAG	Glu → Stop	39	–
OV24	819	1 bp deletion	frameshift	44	–
BT106	1708	GCG → GAG	Ala → Glu	24	+
MC44	1775	ATG → AGG	Met → Arg	42	+
17764	958	4 bp deletion	frameshift	31	+
19964	958	4 bp deletion	frameshift		+*

*Unaffected haplotype carrier, male

Several lines of evidence suggest that all five mutations represent BRCA1 susceptibility alleles:

(i) all mutations are present in the germline;

(ii) all are absent in appropriate control populations, suggesting they are not common polymorphisms;

(iii) each mutant allele is retained in the tumor, as is the case in tumors from patients belonging to kindreds that segregate BRCA1 susceptibility alleles (Smith et al., 1992; Kelsell et al., 1993) (if the mutations represented neutral polymorphisms, they should be retained in only 50% of the cases);

(iv) the age of onset in the four breast cancer cases with mutations varied between 24 and 42 years of age, consistent with the early age of onset of breast cancer in individuals with BRCA1 susceptibility; similarly, the ovarian cancer case was diagnosed at 44, an age that falls in the youngest 13% of all ovarian cancer cases; and finally,

(v) three of the five cases have positive family histories of breast or ovarian cancer found retrospectively in their medical records, although the tumor set was not selected with regard to this criterion.

BT106 was diagnosed at a very early age with breast cancer. Her mother had ovarian cancer, her father had melanoma, and her paternal grandmother also had breast cancer. Patient MC44, an African-American, had bilateral breast cancer at an early age. This patient had a sister who died of breast cancer at a very early age, another sister who died of lymphoma, and a brother who died of lung cancer. Her mutation (Met1775Arg) had been detected previously in Kindred 2099, an African-American family that segregates a BRCA1 susceptibility allele, and was absent in African-American and Caucasian controls. Patient MC44, to our knowledge, is unrelated to Kindred 2099. The detection of a rare mutant allele, once in a BRCA1 kindred and once in the germline of an apparently unrelated early-onset breast cancer case, suggests that the Met1775Arg change may be a common predisposing mutation in African-Americans. Collectively, these observations indicate that all four BRCA1 mutations in tumors represent susceptibility alleles; no somatic mutations were detected in the samples analyzed.

The paucity of somatic BRCA1 mutations is unexpected, given the frequency of LOH on 17q, and the usual role of susceptibility genes as tumor suppressors in cancer progression. There are three possible explanations for this result: (i) some BRCA1 mutations in coding sequences were missed

by our screening procedure; (ii) BRCA1 somatic mutations fall primarily outside the coding exons; and (iii) LOH events in 17q do not reflect BRCA1 somatic mutations.

If somatic BRCA1 mutations truly are rare in breast and ovary carcinomas, this would have strong implications for the biology of BRCA1. The apparent lack of somatic BRCA1 mutations implies that there may be some fundamental difference in the genesis of tumors in genetically predisposed BRCA1 carriers, compared with tumors in the general population. For example, mutations in BRCA1 may have an effect only on tumor formation at a specific stage early in breast and ovarian development. This possibility is consistent with a primary function for BRCA1 in premenopausal breast cancer. Such a model for the role of BRCA1 in breast and ovarian cancer predicts an interaction between reproductive hormones and BRCA1 function. However, no clinical or pathological differences in familial versus sporadic breast and ovary tumors, other than age of onset, have been described (Lynch et al., 1990). On the other hand, the recent finding of increased TP53 mutation and microsatellite instability in breast tumors from patients with a family history of breast cancer (Glebov et al., 1994) may reflect some difference in tumors that arise in genetically predisposed persons. The involvement of BRCA1 in this phenomenon can now be addressed directly. Alternatively, the lack of somatic BRCA1 mutations may result from the existence of multiple genes that function in the same pathway of tumor suppression as BRCA1, but which collectively represent a more favored target for mutation in sporadic tumors. Since mutation of a single element in a genetic pathway is generally sufficient to disrupt the pathway, BRCA1 might mutate at a rate that is far lower than the sum of the mutational rates of the other elements.

EXAMPLE 10

Analysis of the BRCA1 Gene

The structure and function of BRCA1 gene are determined according to the following methods.

Biological Studies.

Mammalian expression vectors containing BRCA1 cDNA are constructed and transfected into appropriate breast carcinoma cells with lesions in the gene. Wild-type BRCA1 cDNA as well as altered BRCA1 cDNA are utilized. The altered BRCA1 cDNA can be obtained from altered BRCA1 alleles or produced as described below. Phenotypic reversion in cultures (e.g., cell morphology, doubling time, anchorage-independent growth) and in animals (e.g., tumorigenicity) is examined. The studies will employ both wild-type and mutant forms (Section B) of the gene.

Molecular Genetics Studies.

In vitro mutagenesis is performed to construct deletion mutants and missense mutants (by single base-pair substitutions in individual codons and cluster charged→alanine scanning mutagenesis). The mutants are used in biological, biochemical and biophysical studies.

Mechanism Studies.

The ability of BRCA1 protein to bind to known and unknown DNA sequences is examined. Its ability to transactivate promoters is analyzed by transient reporter expression systems in mammalian cells. Conventional procedures such as particle-capture and yeast two-hybrid system are used to discover and identify any functional partners. The nature and functions of the partners are characterized. These partners in turn are targets for drug discovery.

Structural Studies.

Recombinant proteins are produced in *E. coli*, yeast, insect and/or mammalian cells and are used in crystallographical and NMR studies. Molecular modeling of the proteins is also employed. These studies facilitate structure-driven drug design.

EXAMPLE 11

Two Step Assay to Detect the Presence of BRCA1 in a Sample

Patient sample is processed according to the method disclosed by Antonarakis et al. (1985), separated through a 1% agarose gel and transferred to nylon membrane for Southern blot analysis.

Membranes are UV cross linked at 150 mJ using a GS Gene Linker (Bio-Rad). BRCA1 probe corresponding to nucleotide positions 3631–3930 of SEQ ID NO:1 is subcloned into pTZ18U. The phagemids are transformed into *E. coli* MV1190 infected with M13KO7 helper phage (Bio-Rad, Richmond, Calif.). Single stranded DNA is isolated according to standard procedures (see Sambrook et al., 1989).

Blots are prehybridized for 15–30 min at 65° C. in 7% sodium dodecyl sulfate (SDS) in 0.5M $NaPO_4$. The methods follow those described by Nguyen et al., 1992. The blots are hybridized overnight at 65° C. in 7% SDS, 0.5M $NaPO_4$ with 25–50 ng/ml single stranded probe DNA. Post-hybridization washes consist of two 30 min washes in 5% SDS, 40 mM $NaPO_4$ at 65° C., followed by two 30 min washes in 1% SDS, 40 mM $NaPO_4$ at 65° C.

Next the blots are rinsed with phosphate buffered saline (pH 6.8) for 5 min at room temperature and incubated with 0.2% casein in PBS for 30–60 min at room temperature and rinsed in PBS for 5 min. The blots are then preincubated for 5–10 minutes in a shaking water bath at 45° C. with hybridization buffer consisting of 6M urea, 0.3M NaCl, and 5× Denhardt's solution (see Sambrook, et al., 1989). The buffer is removed and replaced with 50–75 μl/cm^2 fresh hybridization buffer plus 2.5 nM of the covalently cross-linked oligonucleotide-alkaline phosphatase conjugate with the nucleotide sequence complementary to the universal primer site (UP-AP, Bio-Rad). The blots are hybridized for 20–30 min at 45° C. and post hybridization washes are incubated at 45° C. as two 10 min washes in 6M urea, 1× standard saline citrate (SSC), 0.1% SDS and one 10 min wash in 1× SSC, 0.1% TRITON®X-100, detergent. The blots are rinsed for 10 min at room temperature with 1× SSC.

Blots are incubated for 10 min at room temperature with shaking in the substrate buffer consisting of 0.1M diethanolamine, 1 mM $MgCl_2$, 0.02% sodium azide, pH 10.0. Individual blots are placed in heat sealable bags with substrate buffer and 0.2 mM AMPPD (3-(2'-spiroadamantane)-4-methoxy-4-(3'-phosphoryloxy)phenyl-1,2-dioxetane, disodium salt, Bio-Rad). After a 20 min incubation at room temperature with shaking, the excess AMPPD solution is removed. The blot is exposed to X-ray film overnight. Positive bands indicate the presence of BRCA1.

EXAMPLE 12

Generation of Polyclonal Antibody against BRCA1

Segments of BRCA1 coding sequence were expressed as fusion protein in *E. coli*. The overexpressed protein was purified by gel elution and used to immunize rabbits and mice using a procedure similar to the one described by

Harlow and Lane, 1988. This procedure has been shown to generate Abs against various other proteins (for example, see Kraemer et al., 1993).

Briefly, a stretch of BRCA1 coding sequence was cloned as a fusion protein in plasmid PET5A (Novagen, Inc., Madison, Wis.). The BRCA1 incorporated sequence includes the amino acids corresponding to #1361–1554 of SEQ ID NO:2. After induction with IPTG, the overexpression of a fusion protein with the expected molecular weight was verified by SDS/PAGE. Fusion protein was purified from the gel by electroelution. The identification of the protein as the BRCA1 fusion product was verified by protein sequencing at the N-terminus. Next, the purified protein was used as immunogen in rabbits. Rabbits were immunized with 100 μg of the protein in complete Freund's adjuvant and boosted twice in 3 week intervals, first with 100 μg of immunogen in incomplete Freund's adjuvant followed by 100 μg of immunogen in PBS. Antibody containing serum is collected two weeks thereafter.

This procedure is repeated to generate antibodies against the mutant forms of the BRCA1 gene. These antibodies, in conjunction with antibodies to wild type BRCA1, are used to detect the presence and the relative level of the mutant forms in various tissues and biological fluids.

EXAMPLE 13

Generation of Monoclonal Antibodies Specific for BRCA1

Monoclonal antibodies are generated according to the following protocol. Mice are immunized with immunogen comprising intact BRCA1 or BRCA1 peptides (wild type or mutant) conjugated to keyhole limpet hemocyanin using glutaraldehyde or EDC as is well known.

The immunogen is mixed with an adjuvant. Each mouse receives four injections of 10 to 100 μg of immunogen and after the fourth injection blood samples are taken from the mice to determine if the serum contains antibody to the immunogen. Serum titer is determined by ELISA or RIA. Mice with sera indicating the presence of antibody to the immunogen are selected for hybridoma production.

Spleens are removed from immune mice and a single cell suspension is prepared (see Harlow and Lane, 1988). Cell fusions are performed essentially as described by Kohler and Milstein, 1975. Briefly, P3.65.3 myeloma cells (American Type Culture Collection, Rockville, Md.) are fused with immune spleen cells using polyethylene glycol as described by Harlow and Lane, 1988. Cells are plated at a density of 2×10^5 cells/well in 96 well tissue culture plates. Individual wells are examined for growth and the supernatants of wells with growth are tested for the presence of BRCA1 specific antibodies by ELISA or RIA using wild type or mutant BRCA1 target protein. Cells in positive wells are expanded and subcloned to establish and confirm monoclonality.

Clones with the desired specificities are expanded and grown as ascites in mice or in a hollow fiber system to produce sufficient quantities of antibody for characterization and assay development.

EXAMPLE 14

Sandwich Assay for BRCA1

Monoclonal antibody is attached to a solid surface such as a plate, tube, bead, or particle. Preferably, the antibody is attached to the well surface of a 96-well ELISA plate. 100 μl sample (e.g., serum, urine, tissue cytosol) containing the BRCA1 peptide/protein (wild-type or mutant) is added to the solid phase antibody. The sample is incubated for 2 hrs at room temperature. Next the sample fluid is decanted, and the solid phase is washed with buffer to remove unbound material. 100 μl of a second monoclonal antibody (to a different determinant on the BRCA1 peptide/protein) is added to the solid phase. This antibody is labeled with a detector molecule (e.g., ^{125}I, enzyme, fluorophore, or a chromophore) and the solid phase with the second antibody is incubated for two hrs at room temperature. The second antibody is decanted and the solid phase is washed with buffer to remove unbound material.

The amount of bound label, which is proportional to the amount of BRCA1 peptide/protein present in the sample, is quantitated. Separate assays are performed using monoclonal antibodies which are specific for the wild-type BRCA1 as well as monoclonal antibodies specific for each of the mutations identified in BRCA1.

Industrial Utility

As previously described above, the present invention provides materials and methods for use in testing BRCA1 alleles of an individual and an interpretation of the normal or predisposing nature of the alleles. Individuals at higher than normal risk might modify their lifestyles appropriately. In the case of BRCA1, the most significant non-genetic risk factor is the protective effect of an early, full term pregnancy. Therefore, women at risk could consider early childbearing or a therapy designed to simulate the hormonal effects of an early full-term pregnancy. Women at high risk would also strive for early detection and would be more highly motivated to learn and practice breast self examination. Such women would also be highly motivated to have regular mammograms, perhaps starting at an earlier age than the general population. Ovarian screening could also be undertaken at greater frequency. Diagnostic methods based on sequence analysis of the BRCA1 locus could also be applied to tumor detection and classification. Sequence analysis could be used to diagnose precursor lesions. With the evolution of the method and the accumulation of information about BRCA1 and other causative loci, it could become possible to separate cancers into benign and malignant.

Women with breast cancers may follow different surgical procedures if they are predisposed, and therefore likely to have additional cancers, than if they are not predisposed. Other therapies may be developed, using either peptides or small molecules (rational drug design). Peptides could be the missing gene product itself or a portion of the missing gene product. Alternatively, the therapeutic agent could be another molecule that mimics the deleterious gene's function, either a peptide or a nonpeptidic molecule that seeks to counteract the deleterious effect of the inherited locus. The therapy could also be gene based, through introduction of a normal BRCA1 allele into individuals to make a protein which will counteract the effect of the deleterious allele. These gene therapies may take many forms and may be directed either toward preventing the tumor from forming, curing a cancer once it has occurred, or stopping a cancer from metastasizing.

It will be appreciated that the methods and compositions of the instant invention can be incorporated in the form of a variety of embodiments, only a few of which are disclosed herein. It will be apparent to the artisan that other embodiments exist and do not depart from the spirit of the invention. Thus, the described embodiments are illustrative and should not be construed as restrictive.

LIST OF REFERENCES

Altschul, S. F. et al. (1990). *J. Mol. Biol.* 215: 195–197.

American Cancer Society, Cancer Facts & Figures—1992. (American Cancer Society, Atlanta, Ga.).
Anand, R. (1992). *Techniques for the Analysis of Complex Genomes*, (Academic Press).
Anderson, et al. (1980). *Proc. Natl. Acad. Sci. USA* 77:5399–5403.
Anderson, D. E. (1972). *J. Natl. Cancer Inst.* 48:1029–1034.
Anderson, J. A., et al. (1992). *J. Otolaryngology* 21:321.
Antonarakis, S. E., et al. (1985). *New Eng J. Med.* 313:842–848.
Ausubel, F. M., et al. (1992). *Current Protocols in Molecular Biology*, (J. Wiley and Sons, N.Y.)
Beaucage & Carruthers (1981). *Tetra. Letts.* 22:1859–1862.
Berkner (1992). *Curr. Top. Microbiol. Immunol.* 158:39–61.
Berkner, et al. (1988). *BioTechniques* 6:616–629.
Bickmore, W. A., et al. (1992). *Science* 257:235–7.
Bishop, D. T., et al. (1988). *Genet. Epidemiol.* 5:151–169.
Bishop, D. T. and Gardner, E. J. (1980). In: *Banbury Report* 4: *Cancer Incidence in Defined Populations* (J. Cairns, J. L. Lyon, M. Skolnick, eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 309–408.
Botstein, et al. (1980). *Am. J Hum. Genet.* 32:314–331.
Bowcock, A. M., et al. (1993). *Am. J. Hum. Genet.* 52:718.
Brandyopadhyay and Temin (1984). *Mol. Cell. Biol.* 4:749–754.
Breakfield and Geller (1987). *Mol. Neurobiol.* 1:337–371.
Brinster, et al. (1981). *Cell* 27:223–231.
Buchschacher and Panganiban (1992). *J. Virol.* 66:2731–2739.
Buckler, et al. (1991). *Proc. Natl. Acad. Sci. USA* 88:4005–4009.
Cannon-Albright, L., et al. (1994). *Cancer Research* 54:2378–2385.
Capecchi, M. R. (1989). *Science* 244:1288.
Cariello (1988). *Human Genetics* 42:726.
Claus, E., et al (1991). *Am. J. Hum. Genet.* 48:232–242.
Conner, B. J., et al. (1983). *Proc. Natl. Acad. Sci. USA* 80:278–282.
Constantini and Lacy (1981). *Nature* 294:92–94.
Cotten, et al. (1990). *Proc. Natl. Acad. Sci. USA* 87:4033–4037.
Cotton, et al. (1988). *Proc. Natl. Acad. Sci. USA* 85:4397.
Cropp, C. S., et al. (1994). *Concer Res.* 54:2548–2551.
Culver, et al. (1992). *Science* 256:1550–1552.
Curiel, et al. (1991a). *Proc. Natl. Acad. Sci. USA* 88:8850–8854.
Curiel, et al. (1991b). *Hum. Gene Ther.* 3:147–154.
Deutscher, M. (1990). *Meth. Enzymology* 182 (Academic Press, San Diego, Calif.).
Donehower, L. A., et al. (1992). *Nature* 356:215.
Drummond, I. A., et al. (1994). *Mol. Cell Biol.* 14:3800–9.
Easton, D., et al. (1993). *Am. J. Hum. Genet.* 52:678–701.
Eccles, D. M., et al (1990). *Oncogene* 5:1599–1601.
Enhancers and Eurkaryotic Gene Expression, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1983).
Erickson, J. et al., (1990). *Science* 249:527–533.
Fain, P. R. (1992). *Cytogen. Cell Genet.* 60:178.
Felgner, et al. (1987). *Proc. Natl Acad. Sci. USA* 84:7413–7417.
Fiers, et al. (1978). *Nature* 273:113.
Fink, et al. (1992). *Hum. Gene Ther.* 3:11–19.
Finkelstein, J., et al. (1990). *Genomics* 7:167–172.
Freese, et al. (1990). *Biochem. Pharmacol.* 40:2189–2199.
Friedman, T. (1991). *In Therapy for Genetic Diseases*, T. Friedman, ed., Oxford University Press, pp. 105–121.
Futreal (1993). Ph.D. Thesis, University of North Carolina, Chapel Hill.
Futreal, A., et al. (1 992a). *Hum. Molec. Genet.* 1:66.
Futreal, P. A., et al. (1992b). *Cancer Res.* 52:2624–2627.
Glebov, O. K., et al. (1994). *Cancer Res.* 54:3703–3709.
Glover, D. (1985). *DNA Cloning*, I and II (Oxford Press).
Go, R. C. P., et al (1983). *J. Natl. Cancer Inst.* 71:455–461.
Goding (1986). *Monoclonal Antibodies: Principles and Practice*, 2d ed. (Academic Press, N.Y.).
Godowski, et al. (1988). *Science* 241:812–816.
Goldgar, D. E., et al. (1994). *J Natl. Can. Inst.* 86:3:200–209.
Gordon, et al. (1980). *Proc. Natl. Acad. Sci. USA* 77:7380–7384.
Gorziglia and Kapikian (1992). *J Virol.* 66:4407–4412.
Graham and van der Eb (1973). *Virology* 52:456–467.
Grompe, M., (1993). *Nature Genetics* 5:111–117.
Grompe, M., et al., (1989). *Proc. Natl. Acad. Sci. USA* 86:5855–5892.
Guthrie, G. & Fink, G. R. (1991). *Guide to Yeast Genetics and Molecular Biology* (Academic Press).
Haber, D. A., et al. (1990). *Cell* 61:1257–69.
Hall, J. M., et al. (1990). *Science* 250:1684–1689.
Hall, J. M., et al. (1992). *Am. J. Hum. Genet.* 50:1235–1241.
Harlow & Lane (1988). *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
Hasty, P., K., et al. (1991). *Nature* 350:243.
Helseth, et al. (1990). *J. Virol.* 64:2416–2420.
Hodgson, J. (1991). *Bio/Technology* 9:19–21.
Huse, et al. (1989). *Science* 246:1275–1281.
Innis et al. (1990). *PCR Protocols. A Guide to Methods and Applications* (Academic Press, San Diego, Calif.).
Jablonski, E., et al. (1986). *Nuc. Acids Res.* 14:6115–6128.
Jacobs, I. J., et al. (1993). i Cancer Res. 53:1218–1221.
Jakoby, W. B. and Pastan, I. H. (eds.) (1979). *Cell Culture. Methods in Enzymology*, volume 58 (Academic Press, Inc., Harcourt Brace Jovanovich (New York)).
Jeffreys, et al. (1985). *Nature* 314:67–73.
Johnson, et al. (1992). *J. Virol.* 66:2952–2965.
Kamb, A. et al. (1994). *Science* 264:436–440.
Kandpal, et al. (1990). *Nucl. Acids Res.* 18:1789–1795.
Kaneda, et al. (1989). *J. Biol. Chem.* 264:12126–12129.
Kanehisa (1984). *Nucl. Acids Res.* 12:203–213.
Kelsell, D. P., et al. (1993). *Human Mol. Genet.* 2:1823–1828.
Kinszler, K. W., et al. (1991). *Science* 251:1366–1370.
Knudson, A. G. (1993). *Nature Genet.* 5:103.
Kohler, G. and Milstein, C. (1975). *Nature* 256:495–497.
Kozak, M. (1987). *Nucleic Acids Res.* 15:8125–8148.
Kraemer, F. B. et al. (1993). *J. Lipid Res.* 34:663–672.
Kubo, T., et al. (1988). *FEBS Letts.* 241:119.
Landegren, et al. (1988). *Science* 242:229.
Lim, et al. (1992). *Circulation* 83:2007–2011.
Lindsay, S., et al. (1987). *Nature* 327:336–368.
Litt, et al. (1989). *Am. J. Hum. Genet.* 44:397–401.
Little, M. H., et al. (1992). *Proc. Natl. Acad. Sci. USA* 89:4791.
Little, M. H., et al. (1993). *Hum. Mol. Genet.* 2:259.
Lovett, et al. (1991). *Proc. Natl. Acad. Sci. USA* 88:9628–9632.
Lynch, H. T., et al. (1990). *Gynecol. Oncol.* 36:48–55.
Madzak, et al. (1992). *J. Gen. Virol.* 73:1533–1536.
Malkin, D., et al. (1990). *Science* 250:1233–1238.
Maniatis, T., et al. (1982). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).
Mann and Baltimore (1985). *J. Virol.* 54:401–407.
Margaritte, et al. (1992). *Am. J. Hum. Genet.* 50:1231–1234.

Margolskee (1992). *Curr. Top. Mvicrobiol. Immunol.* 158:67–90.
Martin, R., et al. (1990). *BioTechniques* 9:762–768.
Matteucci, M. D. and Caruthers, M. H. (1981). *J Am. Chem. Soc.* 103:3185.
Matthews & Kricka (1988). *Anal. Biochem.* 169:1.
Merrifield (1963). *J. Am. Chem. Soc.* 85:2149–2156.
Mettlin, C., et al. (1990). *American Journal of Epidemiology* 131:973–983.
Metzger, et al. (1988). *Nature* 334:31–36.
Miller (1992). *Curr. Top. Microbiol. Immunol.* 158:1–24.
Miller, et al. (1985). *Mol. Cell. Biol.* 5:431–437.
Miller, et al. (1988). *J. Virol.* 62:4337–4345.
Mittlin (1989). *Clinical Chem.* 35:1819.
Modrich, P. (1991). *Ann. Rev. Genet.* 25:229–253.
Mombaerts, P., et al. (1992). *Cell* 68:869.
Monaco, et al. (1986). *Nature* 323:646.
Moss (1992). *Curr. Top. Alicrobiol. Immunol.* 158:25–38.
Muzyczka (1992). *Curr. Top. Microbiol. Immunol.* 158:97–123.
Nabel (1992). *Hum. Gene Ther.* 3:399–410.
Nabel, et al. (1990). *Science* 249:1285–1288.
Nakamura, et al. (1987). *Science* 235:1616–1622.
Narod, S. A., et al. (1991). *The Lancet* 338:82–83.
Newman, B., et al. (1988). *Proc. Natl. Acad. Sci. USA* 85:3044–3048.
Newton, C. R., Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., Smith, J. C., and Markham, A. F. (1989). *Nucl. Acids Res.* 17:2503–2516.
Nguyen, Q., et al. (1992). *BioTechniques* 13:116–123.
Novack, et al. (1986). *Proc. Natl. Acad. Sci. USA* 83:586.
Oh, J. (1985). *Analysis of Human Genetic Linkage*, Johns Hopkins University Press, Baltimore, Md., pp. 1–216.
Ohi, et al. (1990). *Gene* 89:279–282.
Oliphant, A., et al. (1991). *Nucleic Acid Res.* 19:4794.
Oliphant, A., et at. (1991). *Nucleic Acid Res.* 19:4795.
Orita, et al. (1989). *Proc. Natl. Acad. Sci. USA* 86:2776–2770.
Page, et al. (1990). *J. Virol.* 64:5370–5276.
Pellicer, et al. (1980). *Science* 209:1414–1422.
Petropoulos, et al. (I 992). *J. Virol.* 66:3391–3397.
Philpott, K. L., et al. (1992). *Science* 256:1448.
Pierce, et al. (1992). *Proc. Natl. Acad. Sci. USA* 89:2056–2060.
Quantin, et al. (1992). *Proc. Natl. Acad Sci. USA* 89:2581–2584.
Rano & Kidd (1989). *Nucl. Acids Res.* 17:8392.
Rigby, P. W. J., et al. (1977). *J. Mol. Biol.* 113:237–251.
Rosenfeld, et al. (1992). *Cell* 68:143–155.
Sambrook, J., et al. (1989). *Molecular Cloning. A Laboratory Manual*, 2nd Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).
Sato, T., et al (1990). *Cancer Res.* 50:7184–7189.
Scharf (1986). *Science* 233:1076.
Scopes, R. (1982). *Protein Purification: Principles and Practice*, (Springer-Verlag, N.Y.).
Shaulian, E., et al. (1992). *Mol. Cell Biol.* 12:5581–92.
Sheffield, V. C., et al. (1989). *Proc. Natl. Acad Sci. USA* 86:232–236.
Sheffield, V. C., et al. (1991). *Am. J. Hum. Genet.* 49:699–706.
Shenk, et al. (1975). *Proc. Natl. Acad. Sci. USA* 72:989.
Shimada, et al. (1991). *J. Clin. Invest.* 88:1043–1047.
Shinkai, Y., et al. (1992). *Cell* 68:855.
Shizuya, H., et al. (1992). *Proc. Natl. Acad Sci. USA* 89:8794–8797.
Simard, J., et al. (1993). *Human Mol. Genet.* 2:1193–1199.
Skolnick, M. H. and Wallace, B. R. (1988). *Genomics* 2:273–279.
Skolnick, M. H., et al. (1990). *Science* 250:1715–1720.
Smith, S. A., et al. (1992). *Nature Genetics* 2:128–131.
Smith, T. F. and Waterman, M. S. (1981). *J. Mol. Biol.* 147:195–197.
Snouwaert, J. N., et al. (1992). *Science* 257:1083.
Sorge, et al. (1984). *Mol. Cell. Biol.* 4:1730–1737.
Srivastava, S., et al. (1993). *Cancer Res.* 53:4452–5.
Sternberg (1990). *Proc. Natl. Acad. Sci. USA* 87:103–107.
Sternberg, et al. (1990). *The New Biologist* 2:151–162.
Stewart, et al (1992). *Hum. Gene Ther.* 3:267–275.
Stratford-Perricaudet, et al. (1990). *Hum. Gene Ther.* 1:241–256.
Swift, M., et al. (1991). *N. Fngl. J. Med.* 325:1831–1836.
Swift, M., et al. (1976). *Cancer Res.* 36:209–215.
Su, L. K., et al. (1993). *Cancer Res.* 53:2728–31.
Thomas, A. and Skolnick, M. H. (1994). *IMA Journal of Mathematics Applied in Medicine and Biology* (in press).
Tonolio, D., et al. (1990). Cold Spring Harbor Conference.
Valancius, V. & Smithies, O. (1991). *Mol. Cell Biol.* 11:1402.
van Dilla, et al. (1986). *Biotechnology* 4:537–552.
Wagner, et al. (1990). *Proc. Natl. Acad. Sci. USA* 87:3410–3414.
Wagner, et al. (1991). *Proc. Natl. Acad Sci. USA* 88:4255–4259.
Wang and Huang (1989). *Biochemistry* 28:9508–9514.
Wartell, R. M., et al. (1990). *Nucl. Acids Res.* 18:2699–2705.
Weber, J. L. (1990). *Genomics* 7:524–530.
Weber and May (1989). *Am. .J. Hum. Genet.* 44:388–396.
Weber, J. L., et al. (1990). *Nucleic Acid Res.* 18:4640.
Wells, J. A. (1991). *Methods in Enzymol* 202:390–411.
Wetmur & Davidson (1968). *J. Mol. Biol.* 31:349–370.
White, M. B., et al., (1992). *Genomics* 12:301–306.
White and Lalouel (1988). *Ann. Rev. Genet.* 22:259–279.
Wilkinson, et al. (1992). *Nucleic Acids Res.* 20:2233–2239.
Willams and Anderson (1984). *Genet. Epidemiol.* 1:7–20.
Wolff, et al. (1990). *Science* 247:1465–1468.
Wolff, et al. (1991). *BioTechniques* 11:474–485.
Wooster, R., et al. (1994). *Science* 265:2088.
Wu, et al. (1989a). *Genomics* 4:560–569.
Wu, et al. (1989b). *J. Biol. Chem.* 264:16985–16987.
Wu, et al. (1991). *J. Biol. Chem.* 266:14338–14342.
Zenke, et al. (1990). *Proc. Natl. Acad. Sci. USA* 87:3655–3659.

List of Patents and Patent Applications:

U.S. Pat. No. 3,817,837
U.S. Pat. No. 3,850,752
U.S. Pat. No. 3,939,350
U.S. Pat. No. 3,996,345
U.S. Pat. No. 4,275,149
U.S. Pat. No. 4,277,437
U.S. Pat. No. 4,366,241
U.S. Pat. No. 4,376,110
U.S. Pat. No. 4,486,530
U.S. Pat. No. 4,683,195
U.S. Pat. No. 4,683,202
U.S. Pat. No. 4,816,567
U.S. Pat. No. 4,868,105
U.S. Pat. No. 5,252,479
EPO Publication No. 225,807
European Patent Application Publication No. 0332435
Geysen, H., PCT published application WO 84/03564, published 13 Sep. 1984
Hitzeman et al., EP 73,675A
PCT published application WO 93/07282

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i i i) NUMBER OF SEQUENCES: 85

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5914 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: cDNA

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 120..5708

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
AGCTCGCTGA GACTTCCTGG ACCCCGCACC AGGCTGTGGG GTTTCTCAGA TAACTGGGCC      60

CCTGCGCTCA GGAGGCCTTC ACCCTCTGCT CTGGGTAAAG TTCATTGGAA CAGAAAGAA      119

ATG GAT TTA TCT GCT CTT CGC GTT GAA GAA GTA CAA AAT GTC ATT AAT        167
Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
  1               5                  10                  15

GCT ATG CAG AAA ATC TTA GAG TGT CCC ATC TGT CTG GAG TTG ATC AAG       215
Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
            20                  25                  30

GAA CCT GTC TCC ACA AAG TGT GAC CAC ATA TTT TGC AAA TTT TGC ATG       263
Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
        35                  40                  45

CTG AAA CTT CTC AAC CAG AAG AAA GGG CCT TCA CAG TGT CCT TTA TGT       311
Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
    50                  55                  60

AAG AAT GAT ATA ACC AAA AGG AGC CTA CAA GAA AGT ACG AGA TTT AGT       359
Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
65                  70                  75                  80

CAA CTT GTT GAA GAG CTA TTG AAA ATC ATT TGT GCT TTT CAG CTT GAC       407
Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
                85                  90                  95

ACA GGT TTG GAG TAT GCA AAC AGC TAT AAT TTT GCA AAA AAG GAA AAT       455
Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
            100                 105                 110

AAC TCT CCT GAA CAT CTA AAA GAT GAA GTT TCT ATC ATC CAA AGT ATG       503
Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
        115                 120                 125

GGC TAC AGA AAC CGT GCC AAA AGA CTT CTA CAG AGT GAA CCC GAA AAT       551
Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
    130                 135                 140

CCT TCC TTG CAG GAA ACC AGT CTC AGT GTC CAA CTC TCT AAC CTT GGA       599
Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
145                 150                 155                 160

ACT GTG AGA ACT CTG AGG ACA AAG CAG CGG ATA CAA CCT CAA AAG ACG       647
Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
                165                 170                 175

TCT GTC TAC ATT GAA TTG GGA TCT GAT TCT TCT GAA GAT ACC GTT AAT       695
```

-continued

```
Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
                180                 185                 190

AAG GCA ACT TAT TGC AGT GTG GGA GAT CAA GAA TTG TTA CAA ATC ACC      743
Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
        195                 200                 205

CCT CAA GGA ACC AGG GAT GAA ATC AGT TTG GAT TCT GCA AAA AAG GCT      791
Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
    210                 215                 220

GCT TGT GAA TTT TCT GAG ACG GAT GTA ACA AAT ACT GAA CAT CAT CAA      839
Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
225                 230                 235                 240

CCC AGT AAT AAT GAT TTG AAC ACC ACT GAG AAG CGT GCA GCT GAG AGG      887
Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg
                245                 250                 255

CAT CCA GAA AAG TAT CAG GGT AGT TCT GTT TCA AAC TTG CAT GTG GAG      935
His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu
            260                 265                 270

CCA TGT GGC ACA AAT ACT CAT GCC AGC TCA TTA CAG CAT GAG AAC AGC      983
Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser
        275                 280                 285

AGT TTA TTA CTC ACT AAA GAC AGA ATG AAT GTA GAA AAG GCT GAA TTC     1031
Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe
    290                 295                 300

TGT AAT AAA AGC AAA CAG CCT GGC TTA GCA AGG AGC CAA CAT AAC AGA     1079
Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg
305                 310                 315                 320

TGG GCT GGA AGT AAG GAA ACA TGT AAT GAT AGG CGG ACT CCC AGC ACA     1127
Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr
                325                 330                 335

GAA AAA AAG GTA GAT CTG AAT GCT GAT CCC CTG TGT GAG AGA AAA GAA     1175
Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu
            340                 345                 350

TGG AAT AAG CAG AAA CTG CCA TGC TCA GAG AAT CCT AGA GAT ACT GAA     1223
Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu
        355                 360                 365

GAT GTT CCT TGG ATA ACA CTA AAT AGC AGC ATT CAG AAA GTT AAT GAG     1271
Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu
    370                 375                 380

TGG TTT TCC AGA AGT GAT GAA CTG TTA GGT TCT GAT GAC TCA CAT GAT     1319
Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp
385                 390                 395                 400

GGG GAG TCT GAA TCA AAT GCC AAA GTA GCT GAT GTA TTG GAC GTT CTA     1367
Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu
                405                 410                 415

AAT GAG GTA GAT GAA TAT TCT GGT TCT TCA GAG AAA ATA GAC TTA CTG     1415
Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu
            420                 425                 430

GCC AGT GAT CCT CAT GAG GCT TTA ATA TGT AAA AGT GAA AGA GTT CAC     1463
Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Glu Arg Val His
        435                 440                 445

TCC AAA TCA GTA GAG AGT AAT ATT GAA GAC AAA ATA TTT GGG AAA ACC     1511
Ser Lys Ser Val Glu Ser Asn Ile Glu Asp Lys Ile Phe Gly Lys Thr
    450                 455                 460

TAT CGG AAG AAG GCA AGC CTC CCC AAC TTA AGC CAT GTA ACT GAA AAT     1559
Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn
465                 470                 475                 480

CTA ATT ATA GGA GCA TTT GTT ACT GAG CCA CAG ATA ATA CAA GAG CGT     1607
Leu Ile Ile Gly Ala Phe Val Thr Glu Pro Gln Ile Ile Gln Glu Arg
                485                 490                 495

CCC CTC ACA AAT AAA TTA AAG CGT AAA AGG AGA CCT ACA TCA GGC CTT     1655
```

-continued

```
Pro Leu Thr Asn Lys Leu Lys Arg Lys Arg Arg Pro Thr Ser Gly Leu
                500                 505                 510

CAT CCT GAG GAT TTT ATC AAG AAA GCA GAT TTG GCA GTT CAA AAG ACT      1703
His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr
            515                 520                 525

CCT GAA ATG ATA AAT CAG GGA ACT AAC CAA ACG GAG CAG AAT GGT CAA      1751
Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln
        530                 535                 540

GTG ATG AAT ATT ACT AAT AGT GGT CAT GAG AAT AAA ACA AAA GGT GAT      1799
Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp
545                 550                 555                 560

TCT ATT CAG AAT GAG AAA AAT CCT AAC CCA ATA GAA TCA CTC GAA AAA      1847
Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys
                565                 570                 575

GAA TCT GCT TTC AAA ACG AAA GCT GAA CCT ATA AGC AGC AGT ATA AGC      1895
Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser
            580                 585                 590

AAT ATG GAA CTC GAA TTA AAT ATC CAC AAT TCA AAA GCA CCT AAA AAG      1943
Asn Met Glu Leu Glu Leu Asn Ile His Asn Ser Lys Ala Pro Lys Lys
        595                 600                 605

AAT AGG CTG AGG AGG AAG TCT TCT ACC AGG CAT ATT CAT GCG CTT GAA      1991
Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Ala Leu Glu
    610                 615                 620

CTA GTA GTC AGT AGA AAT CTA AGC CCA CCT AAT TGT ACT GAA TTG CAA      2039
Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln
625                 630                 635                 640

ATT GAT AGT TGT TCT AGC AGT GAA GAG ATA AAG AAA AAA AAG TAC AAC      2087
Ile Asp Ser Cys Ser Ser Ser Glu Glu Ile Lys Lys Lys Lys Tyr Asn
                645                 650                 655

CAA ATG CCA GTC AGG CAC AGC AGA AAC CTA CAA CTC ATG GAA GGT AAA      2135
Gln Met Pro Val Arg His Ser Arg Asn Leu Gln Leu Met Glu Gly Lys
            660                 665                 670

GAA CCT GCA ACT GGA GCC AAG AAG AGT AAC AAG CCA AAT GAA CAG ACA      2183
Glu Pro Ala Thr Gly Ala Lys Lys Ser Asn Lys Pro Asn Glu Gln Thr
        675                 680                 685

AGT AAA AGA CAT GAC AGC GAT ACT TTC CCA GAG CTG AAG TTA ACA AAT      2231
Ser Lys Arg His Asp Ser Asp Thr Phe Pro Glu Leu Lys Leu Thr Asn
    690                 695                 700

GCA CCT GGT TCT TTT ACT AAG TGT TCA AAT ACC AGT GAA CTT AAA GAA      2279
Ala Pro Gly Ser Phe Thr Lys Cys Ser Asn Thr Ser Glu Leu Lys Glu
705                 710                 715                 720

TTT GTC AAT CCT AGC CTT CCA AGA GAA GAA AAA GAA GAG AAA CTA GAA      2327
Phe Val Asn Pro Ser Leu Pro Arg Glu Glu Lys Glu Glu Lys Leu Glu
                725                 730                 735

ACA GTT AAA GTG TCT AAT AAT GCT GAA GAC CCC AAA GAT CTC ATG TTA      2375
Thr Val Lys Val Ser Asn Asn Ala Glu Asp Pro Lys Asp Leu Met Leu
            740                 745                 750

AGT GGA GAA AGG GTT TTG CAA ACT GAA AGA TCT GTA GAG AGT AGC AGT      2423
Ser Gly Glu Arg Val Leu Gln Thr Glu Arg Ser Val Glu Ser Ser Ser
        755                 760                 765

ATT TCA TTG GTA CCT GGT ACT GAT TAT GGC ACT CAG GAA AGT ATC TCG      2471
Ile Ser Leu Val Pro Gly Thr Asp Tyr Gly Thr Gln Glu Ser Ile Ser
    770                 775                 780

TTA CTG GAA GTT AGC ACT CTA GGG AAG GCA AAA ACA GAA CCA AAT AAA      2519
Leu Leu Glu Val Ser Thr Leu Gly Lys Ala Lys Thr Glu Pro Asn Lys
785                 790                 795                 800

TGT GTG AGT CAG TGT GCA GCA TTT GAA AAC CCC AAG GGA CTA ATT CAT      2567
Cys Val Ser Gln Cys Ala Ala Phe Glu Asn Pro Lys Gly Leu Ile His
                805                 810                 815

GGT TGT TCC AAA GAT AAT AGA AAT GAC ACA GAA GGC TTT AAG TAT CCA      2615
```

-continued

```
Gly Cys Ser Lys Asp Asn Arg Asn Asp Thr Glu Gly Phe Lys Tyr Pro
                820                 825                 830

TTG GGA CAT GAA GTT AAC CAC AGT CGG GAA ACA AGC ATA GAA ATG GAA      2663
Leu Gly His Glu Val Asn His Ser Arg Glu Thr Ser Ile Glu Met Glu
            835                 840                 845

GAA AGT GAA CTT GAT GCT CAG TAT TTG CAG AAT ACA TTC AAG GTT TCA      2711
Glu Ser Glu Leu Asp Ala Gln Tyr Leu Gln Asn Thr Phe Lys Val Ser
        850                 855                 860

AAG CGC CAG TCA TTT GCT CCG TTT TCA AAT CCA GGA AAT GCA GAA GAG      2759
Lys Arg Gln Ser Phe Ala Pro Phe Ser Asn Pro Gly Asn Ala Glu Glu
865                 870                 875                 880

GAA TGT GCA ACA TTC TCT GCC CAC TCT GGG TCC TTA AAG AAA CAA AGT      2807
Glu Cys Ala Thr Phe Ser Ala His Ser Gly Ser Leu Lys Lys Gln Ser
                885                 890                 895

CCA AAA GTC ACT TTT GAA TGT GAA CAA AAG GAA GAA AAT CAA GGA AAG      2855
Pro Lys Val Thr Phe Glu Cys Glu Gln Lys Glu Glu Asn Gln Gly Lys
            900                 905                 910

AAT GAG TCT AAT ATC AAG CCT GTA CAG ACA GTT AAT ATC ACT GCA GGC      2903
Asn Glu Ser Asn Ile Lys Pro Val Gln Thr Val Asn Ile Thr Ala Gly
        915                 920                 925

TTT CCT GTG GTT GGT CAG AAA GAT AAG CCA GTT GAT AAT GCC AAA TGT      2951
Phe Pro Val Val Gly Gln Lys Asp Lys Pro Val Asp Asn Ala Lys Cys
    930                 935                 940

AGT ATC AAA GGA GGC TCT AGG TTT TGT CTA TCA TCT CAG TTC AGA GGC      2999
Ser Ile Lys Gly Gly Ser Arg Phe Cys Leu Ser Ser Gln Phe Arg Gly
945                 950                 955                 960

AAC GAA ACT GGA CTC ATT ACT CCA AAT AAA CAT GGA CTT TTA CAA AAC      3047
Asn Glu Thr Gly Leu Ile Thr Pro Asn Lys His Gly Leu Leu Gln Asn
                965                 970                 975

CCA TAT CGT ATA CCA CCA CTT TTT CCC ATC AAG TCA TTT GTT AAA ACT      3095
Pro Tyr Arg Ile Pro Pro Leu Phe Pro Ile Lys Ser Phe Val Lys Thr
            980                 985                 990

AAA TGT AAG AAA AAT CTG CTA GAG GAA AAC TTT GAG GAA CAT TCA ATG      3143
Lys Cys Lys Lys Asn Leu Leu Glu Glu Asn Phe Glu Glu His Ser Met
        995                 1000                1005

TCA CCT GAA AGA GAA ATG GGA AAT GAG AAC ATT CCA AGT ACA GTG AGC      3191
Ser Pro Glu Arg Glu Met Gly Asn Glu Asn Ile Pro Ser Thr Val Ser
    1010                1015                1020

ACA ATT AGC CGT AAT AAC ATT AGA GAA AAT GTT TTT AAA GAA GCC AGC      3239
Thr Ile Ser Arg Asn Asn Ile Arg Glu Asn Val Phe Lys Glu Ala Ser
1025                1030                1035                1040

TCA AGC AAT ATT AAT GAA GTA GGT TCC AGT ACT AAT GAA GTG GGC TCC      3287
Ser Ser Asn Ile Asn Glu Val Gly Ser Ser Thr Asn Glu Val Gly Ser
                1045                1050                1055

AGT ATT AAT GAA ATA GGT TCC AGT GAT GAA AAC ATT CAA GCA GAA CTA      3335
Ser Ile Asn Glu Ile Gly Ser Ser Asp Glu Asn Ile Gln Ala Glu Leu
            1060                1065                1070

GGT AGA AAC AGA GGG CCA AAA TTG AAT GCT ATG CTT AGA TTA GGG GTT      3383
Gly Arg Asn Arg Gly Pro Lys Leu Asn Ala Met Leu Arg Leu Gly Val
        1075                1080                1085

TTG CAA CCT GAG GTC TAT AAA CAA AGT CTT CCT GGA AGT AAT TGT AAG      3431
Leu Gln Pro Glu Val Tyr Lys Gln Ser Leu Pro Gly Ser Asn Cys Lys
    1090                1095                1100

CAT CCT GAA ATA AAA AAG CAA GAA TAT GAA GAA GTA GTT CAG ACT GTT      3479
His Pro Glu Ile Lys Lys Gln Glu Tyr Glu Glu Val Val Gln Thr Val
1105                1110                1115                1120

AAT ACA GAT TTC TCT CCA TAT CTG ATT TCA GAT AAC TTA GAA CAG CCT      3527
Asn Thr Asp Phe Ser Pro Tyr Leu Ile Ser Asp Asn Leu Glu Gln Pro
                1125                1130                1135

ATG GGA AGT AGT CAT GCA TCT CAG GTT TGT TCT GAG ACA CCT GAT GAC      3575
```

-continued

```
Met Gly Ser Ser His Ala Ser Gln Val Cys Ser Glu Thr Pro Asp Asp
                1140                1145                1150

CTG TTA GAT GAT GGT GAA ATA AAG GAA GAT ACT AGT TTT GCT GAA AAT      3623
Leu Leu Asp Asp Gly Glu Ile Lys Glu Asp Thr Ser Phe Ala Glu Asn
          1155                1160                1165

GAC ATT AAG GAA AGT TCT GCT GTT TTT AGC AAA AGC GTC CAG AAA GGA      3671
Asp Ile Lys Glu Ser Ser Ala Val Phe Ser Lys Ser Val Gln Lys Gly
     1170                1175                1180

GAG CTT AGC AGG AGT CCT AGC CCT TTC ACC CAT ACA CAT TTG GCT CAG      3719
Glu Leu Ser Arg Ser Pro Ser Pro Phe Thr His Thr His Leu Ala Gln
1185                1190                1195                1200

GGT TAC CGA AGA GGG GCC AAG AAA TTA GAG TCC TCA GAA GAG AAC TTA      3767
Gly Tyr Arg Arg Gly Ala Lys Lys Leu Glu Ser Ser Glu Glu Asn Leu
                1205                1210                1215

TCT AGT GAG GAT GAA GAG CTT CCC TGC TTC CAA CAC TTG TTA TTT GGT      3815
Ser Ser Glu Asp Glu Glu Leu Pro Cys Phe Gln His Leu Leu Phe Gly
          1220                1225                1230

AAA GTA AAC AAT ATA CCT TCT CAG TCT ACT AGG CAT AGC ACC GTT GCT      3863
Lys Val Asn Asn Ile Pro Ser Gln Ser Thr Arg His Ser Thr Val Ala
     1235                1240                1245

ACC GAG TGT CTG TCT AAG AAC ACA GAG GAG AAT TTA TTA TCA TTG AAG      3911
Thr Glu Cys Leu Ser Lys Asn Thr Glu Glu Asn Leu Leu Ser Leu Lys
1250                1255                1260

AAT AGC TTA AAT GAC TGC AGT AAC CAG GTA ATA TTG GCA AAG GCA TCT      3959
Asn Ser Leu Asn Asp Cys Ser Asn Gln Val Ile Leu Ala Lys Ala Ser
1265                1270                1275                1280

CAG GAA CAT CAC CTT AGT GAG GAA ACA AAA TGT TCT GCT AGC TTG TTT      4007
Gln Glu His His Leu Ser Glu Glu Thr Lys Cys Ser Ala Ser Leu Phe
                1285                1290                1295

TCT TCA CAG TGC AGT GAA TTG GAA GAC TTG ACT GCA AAT ACA AAC ACC      4055
Ser Ser Gln Cys Ser Glu Leu Glu Asp Leu Thr Ala Asn Thr Asn Thr
          1300                1305                1310

CAG GAT CCT TTC TTG ATT GGT TCT TCC AAA CAA ATG AGG CAT CAG TCT      4103
Gln Asp Pro Phe Leu Ile Gly Ser Ser Lys Gln Met Arg His Gln Ser
     1315                1320                1325

GAA AGC CAG GGA GTT GGT CTG AGT GAC AAG GAA TTG GTT TCA GAT GAT      4151
Glu Ser Gln Gly Val Gly Leu Ser Asp Lys Glu Leu Val Ser Asp Asp
1330                1335                1340

GAA GAA AGA GGA ACG GGC TTG GAA GAA AAT AAT CAA GAA GAG CAA AGC      4199
Glu Glu Arg Gly Thr Gly Leu Glu Glu Asn Asn Gln Glu Glu Gln Ser
1345                1350                1355                1360

ATG GAT TCA AAC TTA GGT GAA GCA GCA TCT GGG TGT GAG AGT GAA ACA      4247
Met Asp Ser Asn Leu Gly Glu Ala Ala Ser Gly Cys Glu Ser Glu Thr
                1365                1370                1375

AGC GTC TCT GAA GAC TGC TCA GGG CTA TCC TCT CAG AGT GAC ATT TTA      4295
Ser Val Ser Glu Asp Cys Ser Gly Leu Ser Ser Gln Ser Asp Ile Leu
          1380                1385                1390

ACC ACT CAG CAG AGG GAT ACC ATG CAA CAT AAC CTG ATA AAG CTC CAG      4343
Thr Thr Gln Gln Arg Asp Thr Met Gln His Asn Leu Ile Lys Leu Gln
     1395                1400                1405

CAG GAA ATG GCT GAA CTA GAA GCT GTG TTA GAA CAG CAT GGG AGC CAG      4391
Gln Glu Met Ala Glu Leu Glu Ala Val Leu Glu Gln His Gly Ser Gln
1410                1415                1420

CCT TCT AAC AGC TAC CCT TCC ATC ATA AGT GAC TCT TCT GCC CTT GAG      4439
Pro Ser Asn Ser Tyr Pro Ser Ile Ile Ser Asp Ser Ser Ala Leu Glu
1425                1430                1435                1440

GAC CTG CGA AAT CCA GAA CAA AGC ACA TCA GAA AAA GCA GTA TTA ACT      4487
Asp Leu Arg Asn Pro Glu Gln Ser Thr Ser Glu Lys Ala Val Leu Thr
                1445                1450                1455

TCA CAG AAA AGT AGT GAA TAC CCT ATA AGC CAG AAT CCA GAA GGC CTT      4535
```

-continued

```
Ser Gln Lys Ser Ser Glu Tyr Pro Ile Ser Gln Asn Pro Glu Gly Leu
                1460                1465                1470

TCT GCT GAC AAG TTT GAG GTG TCT GCA GAT AGT TCT ACC AGT AAA AAT      4583
Ser Ala Asp Lys Phe Glu Val Ser Ala Asp Ser Ser Thr Ser Lys Asn
        1475                1480                1485

AAA GAA CCA GGA GTG GAA AGG TCA TCC CCT TCT AAA TGC CCA TCA TTA      4631
Lys Glu Pro Gly Val Glu Arg Ser Ser Pro Ser Lys Cys Pro Ser Leu
    1490                1495                1500

GAT GAT AGG TGG TAC ATG CAC AGT TGC TCT GGG AGT CTT CAG AAT AGA      4679
Asp Asp Arg Trp Tyr Met His Ser Cys Ser Gly Ser Leu Gln Asn Arg
1505                1510                1515                1520

AAC TAC CCA TCT CAA GAG GAG CTC ATT AAG GTT GTT GAT GTG GAG GAG      4727
Asn Tyr Pro Ser Gln Glu Glu Leu Ile Lys Val Val Asp Val Glu Glu
                1525                1530                1535

CAA CAG CTG GAA GAG TCT GGG CCA CAC GAT TTG ACG GAA ACA TCT TAC      4775
Gln Gln Leu Glu Glu Ser Gly Pro His Asp Leu Thr Glu Thr Ser Tyr
            1540                1545                1550

TTG CCA AGG CAA GAT CTA GAG GGA ACC CCT TAC CTG GAA TCT GGA ATC      4823
Leu Pro Arg Gln Asp Leu Glu Gly Thr Pro Tyr Leu Glu Ser Gly Ile
        1555                1560                1565

AGC CTC TTC TCT GAT GAC CCT GAA TCT GAT CCT TCT GAA GAC AGA GCC      4871
Ser Leu Phe Ser Asp Asp Pro Glu Ser Asp Pro Ser Glu Asp Arg Ala
    1570                1575                1580

CCA GAG TCA GCT CGT GTT GGC AAC ATA CCA TCT TCA ACC TCT GCA TTG      4919
Pro Glu Ser Ala Arg Val Gly Asn Ile Pro Ser Ser Thr Ser Ala Leu
1585                1590                1595                1600

AAA GTT CCC CAA TTG AAA GTT GCA GAA TCT GCC CAG AGT CCA GCT GCT      4967
Lys Val Pro Gln Leu Lys Val Ala Glu Ser Ala Gln Ser Pro Ala Ala
                1605                1610                1615

GCT CAT ACT ACT GAT ACT GCT GGG TAT AAT GCA ATG GAA GAA AGT GTG      5015
Ala His Thr Thr Asp Thr Ala Gly Tyr Asn Ala Met Glu Glu Ser Val
            1620                1625                1630

AGC AGG GAG AAG CCA GAA TTG ACA GCT TCA ACA GAA AGG GTC AAC AAA      5063
Ser Arg Glu Lys Pro Glu Leu Thr Ala Ser Thr Glu Arg Val Asn Lys
        1635                1640                1645

AGA ATG TCC ATG GTG GTG TCT GGC CTG ACC CCA GAA GAA TTT ATG CTC      5111
Arg Met Ser Met Val Val Ser Gly Leu Thr Pro Glu Glu Phe Met Leu
    1650                1655                1660

GTG TAC AAG TTT GCC AGA AAA CAC CAC ATC ACT TTA ACT AAT CTA ATT      5159
Val Tyr Lys Phe Ala Arg Lys His His Ile Thr Leu Thr Asn Leu Ile
1665                1670                1675                1680

ACT GAA GAG ACT ACT CAT GTT GTT ATG AAA ACA GAT GCT GAG TTT GTG      5207
Thr Glu Glu Thr Thr His Val Val Met Lys Thr Asp Ala Glu Phe Val
                1685                1690                1695

TGT GAA CGG ACA CTG AAA TAT TTT CTA GGA ATT GCG GGA GGA AAA TGG      5255
Cys Glu Arg Thr Leu Lys Tyr Phe Leu Gly Ile Ala Gly Gly Lys Trp
            1700                1705                1710

GTA GTT AGC TAT TTC TGG GTG ACC CAG TCT ATT AAA GAA AGA AAA ATG      5303
Val Val Ser Tyr Phe Trp Val Thr Gln Ser Ile Lys Glu Arg Lys Met
        1715                1720                1725

CTG AAT GAG CAT GAT TTT GAA GTC AGA GGA GAT GTG GTC AAT GGA AGA      5351
Leu Asn Glu His Asp Phe Glu Val Arg Gly Asp Val Val Asn Gly Arg
    1730                1735                1740

AAC CAC CAA GGT CCA AAG CGA GCA AGA GAA TCC CAG GAC AGA AAG ATC      5399
Asn His Gln Gly Pro Lys Arg Ala Arg Glu Ser Gln Asp Arg Lys Ile
1745                1750                1755                1760

TTC AGG GGG CTA GAA ATC TGT TGC TAT GGG CCC TTC ACC AAC ATG CCC      5447
Phe Arg Gly Leu Glu Ile Cys Cys Tyr Gly Pro Phe Thr Asn Met Pro
                1765                1770                1775

ACA GAT CAA CTG GAA TGG ATG GTA CAG CTG TGT GGT GCT TCT GTG GTG      5495
```

-continued

```
Thr Asp Gln Leu Glu Trp Met Val Gln Leu Cys Gly Ala Ser Val Val
                1780                1785                1790

AAG GAG CTT TCA TCA TTC ACC CTT GGC ACA GGT GTC CAC CCA ATT GTG      5543
Lys Glu Leu Ser Ser Phe Thr Leu Gly Thr Gly Val His Pro Ile Val
          1795                1800                1805

GTT GTG CAG CCA GAT GCC TGG ACA GAG GAC AAT GGC TTC CAT GCA ATT      5591
Val Val Gln Pro Asp Ala Trp Thr Glu Asp Asn Gly Phe His Ala Ile
     1810                1815                1820

GGG CAG ATG TGT GAG GCA CCT GTG GTG ACC CGA GAG TGG GTG TTG GAC      5639
Gly Gln Met Cys Glu Ala Pro Val Val Thr Arg Glu Trp Val Leu Asp
1825                1830                1835                1840

AGT GTA GCA CTC TAC CAG TGC CAG GAG CTG GAC ACC TAC CTG ATA CCC      5687
Ser Val Ala Leu Tyr Gln Cys Gln Glu Leu Asp Thr Tyr Leu Ile Pro
                1845                1850                1855

CAG ATC CCC CAC AGC CAC TAC TGA CTGCAGCCAG CCACAGGTAC AGAGCCACAG      5741
Gln Ile Pro His Ser His Tyr
          1860

GACCCCAAGA ATGAGCTTAC AAAGTGGCCT TTCCAGGCCC TGGGAGCTCC TCTCACTCTT    5801

CAGTCCTTCT ACTGTCCTGG CTACTAAATA TTTTATGTAC ATCAGCCTGA AAAGGACTTC    5861

TGGCTATGCA AGGGTCCCTT AAAGATTTTC TGCTTGAAGT CTCCCTTGGA AAT           5914
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1863 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
  1               5                  10                  15

Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
            20                  25                  30

Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
        35                  40                  45

Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
    50                  55                  60

Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
65                  70                  75                  80

Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
                85                  90                  95

Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
            100                 105                 110

Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
        115                 120                 125

Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
    130                 135                 140

Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
145                 150                 155                 160

Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
                165                 170                 175

Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
            180                 185                 190

Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
        195                 200                 205
```

-continued

Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
210 215 220

Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
225 230 235 240

Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg
245 250 255

His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu
260 265 270

Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser
275 280 285

Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe
290 295 300

Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg
305 310 315 320

Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr
325 330 335

Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu
340 345 350

Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu
355 360 365

Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu
370 375 380

Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp
385 390 395 400

Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu
405 410 415

Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu
420 425 430

Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Glu Arg Val His
435 440 445

Ser Lys Ser Val Glu Ser Asn Ile Glu Asp Lys Ile Phe Gly Lys Thr
450 455 460

Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn
465 470 475 480

Leu Ile Ile Gly Ala Phe Val Thr Glu Pro Gln Ile Ile Gln Glu Arg
485 490 495

Pro Leu Thr Asn Lys Leu Lys Arg Lys Arg Arg Pro Thr Ser Gly Leu
500 505 510

His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr
515 520 525

Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln
530 535 540

Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp
545 550 555 560

Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys
565 570 575

Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser
580 585 590

Asn Met Glu Leu Glu Leu Asn Ile His Asn Ser Lys Ala Pro Lys Lys
595 600 605

Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Ala Leu Glu
610 615 620

Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln
625 630 635 640

-continued

```
Ile Asp Ser Cys Ser Ser Ser Glu Glu Ile Lys Lys Lys Lys Tyr Asn
                645                 650                 655

Gln Met Pro Val Arg His Ser Arg Asn Leu Gln Leu Met Glu Gly Lys
            660                 665                 670

Glu Pro Ala Thr Gly Ala Lys Lys Ser Asn Lys Pro Asn Glu Gln Thr
        675                 680                 685

Ser Lys Arg His Asp Ser Asp Thr Phe Pro Glu Leu Lys Leu Thr Asn
    690                 695                 700

Ala Pro Gly Ser Phe Thr Lys Cys Ser Asn Thr Ser Glu Leu Lys Glu
705                 710                 715                 720

Phe Val Asn Pro Ser Leu Pro Arg Glu Glu Lys Glu Glu Lys Leu Glu
                725                 730                 735

Thr Val Lys Val Ser Asn Asn Ala Glu Asp Pro Lys Asp Leu Met Leu
            740                 745                 750

Ser Gly Glu Arg Val Leu Gln Thr Glu Arg Ser Val Glu Ser Ser Ser
        755                 760                 765

Ile Ser Leu Val Pro Gly Thr Asp Tyr Gly Thr Gln Glu Ser Ile Ser
    770                 775                 780

Leu Leu Glu Val Ser Thr Leu Gly Lys Ala Lys Thr Glu Pro Asn Lys
785                 790                 795                 800

Cys Val Ser Gln Cys Ala Ala Phe Glu Asn Pro Lys Gly Leu Ile His
                805                 810                 815

Gly Cys Ser Lys Asp Asn Arg Asn Asp Thr Glu Gly Phe Lys Tyr Pro
            820                 825                 830

Leu Gly His Glu Val Asn His Ser Arg Glu Thr Ser Ile Glu Met Glu
        835                 840                 845

Glu Ser Glu Leu Asp Ala Gln Tyr Leu Gln Asn Thr Phe Lys Val Ser
    850                 855                 860

Lys Arg Gln Ser Phe Ala Pro Phe Ser Asn Pro Gly Asn Ala Glu Glu
865                 870                 875                 880

Glu Cys Ala Thr Phe Ser Ala His Ser Gly Ser Leu Lys Lys Gln Ser
                885                 890                 895

Pro Lys Val Thr Phe Glu Cys Glu Gln Lys Glu Glu Asn Gln Gly Lys
            900                 905                 910

Asn Glu Ser Asn Ile Lys Pro Val Gln Thr Val Asn Ile Thr Ala Gly
        915                 920                 925

Phe Pro Val Val Gly Gln Lys Asp Lys Pro Val Asp Asn Ala Lys Cys
    930                 935                 940

Ser Ile Lys Gly Gly Ser Arg Phe Cys Leu Ser Ser Gln Phe Arg Gly
945                 950                 955                 960

Asn Glu Thr Gly Leu Ile Thr Pro Asn Lys His Gly Leu Leu Gln Asn
                965                 970                 975

Pro Tyr Arg Ile Pro Pro Leu Phe Pro Ile Lys Ser Phe Val Lys Thr
            980                 985                 990

Lys Cys Lys Lys Asn Leu Leu Glu Glu Asn Phe Glu Glu His Ser Met
        995                 1000                1005

Ser Pro Glu Arg Glu Met Gly Asn Glu Asn Ile Pro Ser Thr Val Ser
    1010                1015                1020

Thr Ile Ser Arg Asn Asn Ile Arg Glu Asn Val Phe Lys Glu Ala Ser
1025                1030                1035                1040

Ser Ser Asn Ile Asn Glu Val Gly Ser Ser Thr Asn Glu Val Gly Ser
                1045                1050                1055

Ser Ile Asn Glu Ile Gly Ser Ser Asp Glu Asn Ile Gln Ala Glu Leu
```

-continued

1060 1065 1070

Gly Arg Asn Arg Gly Pro Lys Leu Asn Ala Met Leu Arg Leu Gly Val
1075 1080 1085

Leu Gln Pro Glu Val Tyr Lys Gln Ser Leu Pro Gly Ser Asn Cys Lys
1090 1095 1100

His Pro Glu Ile Lys Lys Gln Glu Tyr Glu Glu Val Val Gln Thr Val
1105 1110 1115 1120

Asn Thr Asp Phe Ser Pro Tyr Leu Ile Ser Asp Asn Leu Glu Gln Pro
1125 1130 1135

Met Gly Ser Ser His Ala Ser Gln Val Cys Ser Glu Thr Pro Asp Asp
1140 1145 1150

Leu Leu Asp Asp Gly Glu Ile Lys Glu Asp Thr Ser Phe Ala Glu Asn
1155 1160 1165

Asp Ile Lys Glu Ser Ser Ala Val Phe Ser Lys Ser Val Gln Lys Gly
1170 1175 1180

Glu Leu Ser Arg Ser Pro Ser Pro Phe Thr His Thr His Leu Ala Gln
1185 1190 1195 1200

Gly Tyr Arg Arg Gly Ala Lys Lys Leu Glu Ser Ser Glu Glu Asn Leu
1205 1210 1215

Ser Ser Glu Asp Glu Glu Leu Pro Cys Phe Gln His Leu Leu Phe Gly
1220 1225 1230

Lys Val Asn Asn Ile Pro Ser Gln Ser Thr Arg His Ser Thr Val Ala
1235 1240 1245

Thr Glu Cys Leu Ser Lys Asn Thr Glu Glu Asn Leu Leu Ser Leu Lys
1250 1255 1260

Asn Ser Leu Asn Asp Cys Ser Asn Gln Val Ile Leu Ala Lys Ala Ser
1265 1270 1275 1280

Gln Glu His His Leu Ser Glu Glu Thr Lys Cys Ser Ala Ser Leu Phe
1285 1290 1295

Ser Ser Gln Cys Ser Glu Leu Glu Asp Leu Thr Ala Asn Thr Asn Thr
1300 1305 1310

Gln Asp Pro Phe Leu Ile Gly Ser Ser Lys Gln Met Arg His Gln Ser
1315 1320 1325

Glu Ser Gln Gly Val Gly Leu Ser Asp Lys Glu Leu Val Ser Asp Asp
1330 1335 1340

Glu Glu Arg Gly Thr Gly Leu Glu Glu Asn Asn Gln Glu Glu Gln Ser
1345 1350 1355 1360

Met Asp Ser Asn Leu Gly Glu Ala Ala Ser Gly Cys Glu Ser Glu Thr
1365 1370 1375

Ser Val Ser Glu Asp Cys Ser Gly Leu Ser Ser Gln Ser Asp Ile Leu
1380 1385 1390

Thr Thr Gln Gln Arg Asp Thr Met Gln His Asn Leu Ile Lys Leu Gln
1395 1400 1405

Gln Glu Met Ala Glu Leu Glu Ala Val Leu Glu Gln His Gly Ser Gln
1410 1415 1420

Pro Ser Asn Ser Tyr Pro Ser Ile Ile Ser Asp Ser Ser Ala Leu Glu
1425 1430 1435 1440

Asp Leu Arg Asn Pro Glu Gln Ser Thr Ser Glu Lys Ala Val Leu Thr
1445 1450 1455

Ser Gln Lys Ser Ser Glu Tyr Pro Ile Ser Gln Asn Pro Glu Gly Leu
1460 1465 1470

Ser Ala Asp Lys Phe Glu Val Ser Ala Asp Ser Ser Thr Ser Lys Asn
1475 1480 1485

-continued

```
Lys Glu Pro Gly Val Glu Arg Ser Ser Pro Ser Lys Cys Pro Ser Leu
     1490                1495                1500

Asp Asp Arg Trp Tyr Met His Ser Cys Ser Gly Ser Leu Gln Asn Arg
1505                1510                1515                1520

Asn Tyr Pro Ser Gln Glu Glu Leu Ile Lys Val Val Asp Val Glu Glu
                1525                1530                1535

Gln Gln Leu Glu Glu Ser Gly Pro His Asp Leu Thr Glu Thr Ser Tyr
            1540                1545                1550

Leu Pro Arg Gln Asp Leu Glu Gly Thr Pro Tyr Leu Glu Ser Gly Ile
        1555                1560                1565

Ser Leu Phe Ser Asp Asp Pro Glu Ser Asp Pro Ser Glu Asp Arg Ala
     1570                1575                1580

Pro Glu Ser Ala Arg Val Gly Asn Ile Pro Ser Ser Thr Ser Ala Leu
1585                1590                1595                1600

Lys Val Pro Gln Leu Lys Val Ala Glu Ser Ala Gln Ser Pro Ala Ala
                1605                1610                1615

Ala His Thr Thr Asp Thr Ala Gly Tyr Asn Ala Met Glu Glu Ser Val
            1620                1625                1630

Ser Arg Glu Lys Pro Glu Leu Thr Ala Ser Thr Glu Arg Val Asn Lys
        1635                1640                1645

Arg Met Ser Met Val Val Ser Gly Leu Thr Pro Glu Glu Phe Met Leu
     1650                1655                1660

Val Tyr Lys Phe Ala Arg Lys His His Ile Thr Leu Thr Asn Leu Ile
1665                1670                1675                1680

Thr Glu Glu Thr Thr His Val Val Met Lys Thr Asp Ala Glu Phe Val
                1685                1690                1695

Cys Glu Arg Thr Leu Lys Tyr Phe Leu Gly Ile Ala Gly Gly Lys Trp
            1700                1705                1710

Val Val Ser Tyr Phe Trp Val Thr Gln Ser Ile Lys Glu Arg Lys Met
        1715                1720                1725

Leu Asn Glu His Asp Phe Glu Val Arg Gly Asp Val Val Asn Gly Arg
     1730                1735                1740

Asn His Gln Gly Pro Lys Arg Ala Arg Glu Ser Gln Asp Arg Lys Ile
1745                1750                1755                1760

Phe Arg Gly Leu Glu Ile Cys Cys Tyr Gly Pro Phe Thr Asn Met Pro
                1765                1770                1775

Thr Asp Gln Leu Glu Trp Met Val Gln Leu Cys Gly Ala Ser Val Val
            1780                1785                1790

Lys Glu Leu Ser Ser Phe Thr Leu Gly Thr Gly Val His Pro Ile Val
        1795                1800                1805

Val Val Gln Pro Asp Ala Trp Thr Glu Asp Asn Gly Phe His Ala Ile
     1810                1815                1820

Gly Gln Met Cys Glu Ala Pro Val Val Thr Arg Glu Trp Val Leu Asp
1825                1830                1835                1840

Ser Val Ala Leu Tyr Gln Cys Gln Glu Leu Asp Thr Tyr Leu Ile Pro
                1845                1850                1855

Gln Ile Pro His Ser His Tyr
            1860
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

-continued

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: s754 A

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CTAGCCTGGG CAACAAACGA 20

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: s754 B

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GCAGGAAGCA GGAATGGAAC 20

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: s975 A

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TAGGAGATGG ATTATTGGTG 20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: s975 B

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGGCAACTTT GCAATGAGTG 20

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: tdj1474 A

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CAGAGTGAGA CCTTGTCTCA AA 22

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: tdj1474 B

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TTCTGCAAAC ACCTTAAACT CAG 23

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: tdj1239 A

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:9:

AACCTGGAAG GCAGAGGTTG 20

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

-continued

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: tdj1239 B

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
TCTGTACCTG CTAAGCAGTG G                                                21
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 111 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: cDNA

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..111

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
G GKC TTA CTC TGT TGT CCC AGC TGG AGT ACA GWG TGC GAT CAT GAG          46
  Xaa Leu Leu Cys Cys Pro Ser Trp Ser Thr Xaa Cys Asp His Glu
  1865                1870                1875

GCT TAC TGT TGC TTG ACT CCT AGG CTC AAG CGA TCC TAT CAC CTC AGT        94
Ala Tyr Cys Cys Leu Thr Pro Arg Leu Lys Arg Ser Tyr His Leu Ser
1880                1885                1890                1895

CTC CAA GTA GCT GGA  CT                                              111
Leu Gln Val Ala Gly
                1900
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Xaa Leu Leu Cys Cys Pro Ser Trp Ser Thr Xaa Cys Asp His Glu Ala
  1               5                  10                  15

Tyr Cys Cys Leu Thr Pro Arg Leu Lys Arg Ser Tyr His Leu Ser Leu
            20                  25                  30

Gln Val Ala Gly
        35
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1534 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

-continued

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
GAGGCTAGAG GGCAGGCACT TTATGGCAAA CTCAGGTAGA ATTCTTCCTC TTCCGTCTCT      60
TTCCTTTTAC GTCATCGGGG AGACTGGGTG GCAATCGCAG CCCGAGAGAC GCATGGCTCT     120
TTCTGCCCTC CATCCTCTGA TGTACCTTGA TTTCGTATTC TGAGAGGCTG CTGCTTAGCG     180
GTAGCCCCTT GGTTTCCGTG GCAACGGAAA AGCGCGGGAA TTACAGATAA ATTAAAACTG     240
CGACTGCGCG GCGTGAGCTC GCTGAGACTT CCTGGACCCC GCACCAGGCT GTGGGGTTTC     300
TCAGATAACT GGGCCCCTGC GCTCAGGAGG CCTTCACCCT CTGCTCTGGG TAAAGGTAGT     360
AGAGTCCCGG GAAAGGGACA GGGGGCCCAA GTGATGCTCT GGGGTACTGG CGTGGGAGAG     420
TGGATTTCCG AAGCTGACAG ATGGGTATTC TTTGACGGGG GGTAGGGGCG GAACCTGAGA     480
GGCGTAAGGC GTTGTGAACC CTGGGGAGGG GGGCAGTTTG TAGGTCGCGA GGGAAGCGCT     540
GAGGATCAGG AAGGGGGCAC TGAGTGTCCG TGGGGGAATC CTCGTGATAG GAACTGGAAT     600
ATGCCTTGAG GGGGACACTA TGTCTTTAAA AACGTCGGCT GGTCATGAGG TCAGGAGTTC     660
CAGACCAGCC TGACCAACGT GGTGAAACTC CGTCTCTACT AAAAATACNA AAATTAGCCG     720
GGCGTGGTGC CGCTCCAGCT ACTCAGGAGG CTGAGGCAGG AGAATCGCTA GAACCCGGGA     780
GGCGGAGGTT GCAGTGAGCC GAGATCGCGC CATTGCACTC CAGCCTGGGC GACAGAGCGA     840
GACTGTCTCA AAACAAAACA AAACAAAACA AAACAAAAAA CACCGGCTGG TATGTATGAG     900
AGGATGGGAC CTTGTGGAAG AAGAGGTGCC AGGAATATGT CTGGGAAGGG GAGGAGACAG     960
GATTTTGTGG GAGGGAGAAC TTAAGAACTG GATCCATTTG CGCCATTGAG AAAGCGCAAG    1020
AGGGAAGTAG AGGAGCGTCA GTAGTAACAG ATGCTGCCGG CAGGGATGTG CTTGAGGAGG    1080
ATCCAGAGAT GAGAGCAGGT CACTGGGAAA GGTTAGGGGC GGGGAGGCCT TGATTGGTGT    1140
TGGTTTGGTC GTTGTTGATT TTGGTTTTAT GCAAGAAAAA GAAAACAACC AGAAACATTG    1200
GAGAAAGCTA AGGCTACCAC CACCTACCCG GTCAGTCACT CCTCTGTAGC TTTCTCTTTC    1260
TTGGAGAAAG GAAAAGACCC AAGGGGTTGG CAGCGATATG TGAAAAAATT CAGAATTTAT    1320
GTTGTCTAAT TACAAAAAGC AACTTCTAGA ATCTTTAAAA ATAAAGGACG TTGTCATTAG    1380
TTCTTCTGGT TTGTATTATT CTAAAACCTT CCAAATCTTC AAATTTACTT TATTTTAAAA    1440
TGATAAAATG AAGTTGTCAT TTTATAAACC TTTTAAAAAG ATATATATAT ATGTTTTTCT    1500
AATGTGTTAA AGTTCATTGG AACAGAAAGA AATG                                1534
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1924 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAGGCTAGAG GGCAGGCACT TTATGGCAAA CTCAGGTAGA ATTCTTCCTC TTCCGTCTCT 60

TTCCTTTTAC GTCATCGGGG AGACTGGGTG GCAATCGCAG CCCGAGAGAC GCATGGCTCT 120

TTCTGCCCTC CATCCTCTGA TGTACCTTGA TTTCGTATTC TGAGAGGCTG CTGCTTAGCG 180

GTAGCCCCTT GGTTTCCGTG GCAACGGAAA AGCGCGGGAA TTACAGATAA ATTAAAACTG 240

CGACTGCGCG GCGTGAGCTC GCTGAGACTT CCTGGACCCC GCACCAGGCT GTGGGGTTTC 300

TCAGATAACT GGGCCCCTGC GCTCAGGAGG CCTTCACCCT CTGCTCTGGG TAAAGGTAGT 360

AGAGTCCCGG GAAAGGGACA GGGGGCCCAA GTGATGCTCT GGGGTACTGG CGTGGGAGAG 420

TGGATTTCCG AAGCTGACAG ATGGGTATTC TTTGACGGGG GGTAGGGGCG GAACCTGAGA 480

GGCGTAAGGC GTTGTGAACC CTGGGGAGGG GGGCAGTTTG TAGGTCGCGA GGGAAGCGCT 540

GAGGATCAGG AAGGGGGCAC TGAGTGTCCG TGGGGGAATC CTCGTGATAG GAACTGGAAT 600

ATGCCTTGAG GGGGACACTA TGTCTTTAAA AACGTCGGCT GGTCATGAGG TCAGGAGTTC 660

CAGACCAGCC TGACCAACGT GGTGAAACTC CGTCTCTACT AAAAATACNA AAATTAGCCG 720

GGCGTGGTGC CGCTCCAGCT ACTCAGGAGG CTGAGGCAGG AGAATCGCTA GAACCCGGGA 780

GGCGGAGGTT GCAGTGAGCC GAGATCGCGC CATTGCACTC CAGCCTGGGC GACAGAGCGA 840

GACTGTCTCA AAACAAAACA AAACAAAACA AAACAAAAAA CACCGGCTGG TATGTATGAG 900

AGGATGGGAC CTTGTGGAAG AAGAGGTGCC AGGAATATGT CTGGGAAGGG GAGGAGACAG 960

GATTTTGTGG GAGGGAGAAC TTAAGAACTG GATCCATTTG CGCCATTGAG AAAGCGCAAG 1020

AGGGAAGTAG AGGAGCGTCA GTAGTAACAG ATGCTGCCGG CAGGGATGTG CTTGAGGAGG 1080

ATCCAGAGAT GAGAGCAGGT CACTGGGAAA GGTTAGGGGC GGGGAGGCCT TGATTGGTGT 1140

TGGTTTGGTC GTTGTTGATT TTGGTTTTAT GCAAGAAAAA GAAAACAACC AGAAACATTG 1200

GAGAAAGCTA AGGCTACCAC CACCTACCCG GTCAGTCACT CCTCTGTAGC TTTCTCTTTC 1260

TTGGAGAAAG GAAAAGACCC AAGGGGTTGG CAGCGATATG TGAAAAAATT CAGAATTTAT 1320

GTTGTCTAAT TACAAAAAGC AACTTCTAGA ATCTTTAAAA ATAAAGGACG TTGTCATTAG 1380

TTCTTCTGGT TTGTATTATT CTAAAACCTT CCAAATCTTC AAATTTACTT TATTTTAAAA 1440

TGATAAAATG AAGTTGTCAT TTTATAAACC TTTTAAAAAG ATATATATAT ATGTTTTTCT 1500

AATGTGTTAA AGTTCATTGG AACAGAAAGA AATGGATTTA TCTGCTCTTC GCGTTGAAGA 1560

AGTACAAAAT GTCATTAATG CTATGCAGAA AATCTTAGAG TGTCCCATCT GGTAAGTCAG 1620

CACAAGAGTG TATTAATTTG GGATTCCTAT GATTATCTCC TATGCAAATG AACAGAATTG 1680

ACCTTACATA CTAGGGAAGA AAAGACATGT CTAGTAAGAT TAGGCTATTG TAATTGCTGA 1740

TTTTCTTAAC TGAAGAACTT TAAAAATATA GAAAATGATT CCTTGTTCTC CATCCACTCT 1800

GCCTCTCCCA CTCCTCTCCT TTTCAACACA ATCCTGTGGT CCGGGAAAGA CAGGGCTCTG 1860

TCTTGATTGG TTCTGCACTG GGCAGGATCT GTTAGATACT GCATTTGCTT TCTCCAGCTC 1920

TAAA 1924

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 631 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

-continued

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
AAATGCTGAT GATAGTATAG AGTATTGAAG GGATCAATAT AATTCTGTTT TGATATCTGA      60
AAGCTCACTG AAGGTAAGGA TCGTATTCTC TGCTGTATTC TCAGTTCCTG ACACAGCAGA     120
CATTTAATAA ATATTGAACG AACTTGAGGC CTTATGTTGA CTCAGTCATA ACAGCTCAAA     180
GTTGAACTTA TTCACTAAGA ATAGCTTTAT TTTTAAATAA ATTATTGAGC CTCATTTATT     240
TTCTTTTTCT CCCCCCCCTA CCCTGCTAGT CTGGAGTTGA TCAAGGAACC TGTCTCCACA     300
AAGTGTGACC ACATATTTTG CAAGTAAGTT TGAATGTGTT ATGTGGCTCC ATTATTAGCT     360
TTTGTTTTTG TCCTTCATAA CCCAGGAAAC ACCTAACTTT ATAGAAGCTT TACTTTCTTC     420
AATTAAGTGA GAACGAAAAT CCAACTCCAT TTCATTCTTT CTCAGAGAGT ATATAGTTAT     480
CAAAAGTTGG TTGTAATCAT AGTTCCTGGT AAAGTTTTGA CATATATTAT CTTTTTTTTT     540
TTTTGAGACA AGTCTCGCTC TGTCGCCCAG GCTGGAGTGC AGTGGCATGA GGCTTGCTCA     600
CTGCACCTCC GCCCCCGAGT TCAGCGACTC T                                    631
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 481 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
TGAGATCTAG ACCACATGGT CAAAGAGATA GAATGTGAGC AATAAATGAA CCTTAAATTT      60
TTCAACAGCT ACTTTTTTTT TTTTTTTTTG AGACAGGGKC TTACTCTGTT GTCCCAGCTG     120
GAGTACAGWG TGCGATCATG AGGCTTACTG TTGCTTGACT CCTAGGCTCA AGCGATCCTA     180
TCACCTCAGT CTCCAAGTAG CTGGACTGTA AGTGCACACC ACCATATCCA GCTAAATTTT     240
GTGTTTTCTG TAGAGACGGG GTTTCGCCAT GTTTCCCAGG CTGGTCTTGA ACTTTGGGCT     300
TAACCCGTCT GCCCACCTAG GCATCCCAAA GTGCTAGGAT TACAGGTGTG AGTCATCATG     360
CCTGGCCAGT ATTTTAGTTA GCTCTGTCTT TTCAAGTCAT ATACAAGTTC ATTTTCTTTT     420
AAGTTTAGTT AACAACCTTA TATCATGTAT TCTTTTCTAG CATAAAGAAA GATTCGAGGC     480
C                                                                     481
```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 522 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

-continued

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
TGTGATCATA ACAGTAAGCC ATATGCATGT AAGTTCAGTT TTCATAGATC ATTGCTTATG      60
TAGTTTAGGT TTTTGCTTAT GCAGCATCCA AAAACAATTA GGAAACTATT GCTTGTAATT     120
CACCTGCCAT TACTTTTTAA ATGGCTCTTA AGGGCAGTTG TGAGATTATC TTTTCATGGC     180
TATTTGCCTT TTGAGTATTC TTTCTACAAA AGGAAGTAAA TTAAATTGTT CTTTCTTTCT     240
TTATAATTTA TAGATTTTGC ATGCTGAAAC TTCTCAACCA GAAGAAAGGG CCTTCACAGT     300
GTCCTTTATG TAAGAATGAT ATAACCAAAA GGTATATAAT TTGGTAATGA TGCTAGGTTG     360
GAAGCAACCA CAGTAGGAAA AAGTAGAAAT TATTTAATAA CATAGCGTTC CTATAAAACC     420
ATTCATCAGA AAAATTTATA AAAGAGTTTT TAGCACACAG TAAATTATTT CCAAAGTTAT     480
TTTCCTGAAA GTTTTATGGG CATCTGCCTT ATACAGGTAT TG                        522
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 465 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
GGTAGGCTTA AATGAATGAC AAAAAGTTAC TAAATCACTG CCATCACACG GTTTATACAG      60
ATGTCAATGA TGTATTGATT ATAGAGGTTT TCTACTGTTG CTGCATCTTA TTTTTATTTG     120
TTTACATGTC TTTTCTTATT TTAGTGTCCT TAAAAGGTTG ATAATCACTT GCTGAGTGTG     180
TTTCTCAAAC AATTTAATTT CAGGAGCCTA CAAGAAAGTA CGAGATTTAG TCAACTTGTT     240
GAAGAGCTAT TGAAAATCAT TTGTGCTTTT CAGCTTGACA CAGGTTTGGA GTGTAAGTGT     300
TGAATATCCC AAGAATGACA CTCAAGTGCT GTCCATGAAA ACTCAGGAAG TTTGCACAAT     360
TACTTTCTAT GACGTGGTGA TAAGACCTTT TAGTCTAGGT TAATTTTAGT TCTGTATCTG     420
TAATCTATTT TAAAAAATTA CTCCCACTGG TCTCACACCT TATTT                     465
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 513 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:19:

AAAAAATCAC AGGTAACCTT AATGCATTGT CTTAACACAA CAAAGAGCAT ACATAGGGTT 60

TCTCTTGGTT TCTTTGATTA TAATTCATAC ATTTTTCTCT AACTGCAAAC ATAATGTTTT 120

CCCTTGTATT TTACAGATGC AAACAGCTAT AATTTTGCAA AAAAGGAAAA TAACTCTCCT 180

GAACATCTAA AAGATGAAGT TTCTATCATC CAAAGTATGG GCTACAGAAA CCGTGCCAAA 240

AGACTTCTAC AGAGTGAACC CGAAAATCCT TCCTTGGTAA AACCATTTGT TTTCTTCTTC 300

TTCTTCTTCT TCTTTTCTTT TTTTTTTCTT TTTTTTTTTG AGATGGAGTC TTGCTCTGTG 360

GCCCAGGCTA GAAGCAGTCC TCCTGCCTTA GCCNCCTTAG TAGCTGGGAT TACAGGCACG 420

CGCACCATGC CAGGCTAATT TTTGTATTTT TAGTAGAGAC GGGGTTTCAT CATGTTGGCC 480

AGGCTGGTCT CGAACTCCTA ACCTCAGGTG ATC 513

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6769 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATGATGGAGA TCTTAAAAAG TAATCATTCT GGGGCTGGGC GTAGTAGCTT GCACCTGTAA 60

TCCCAGCACT TCGGGAGGCT GAGGCAGGCA GATAATTTGA GGTCAGGAGT TTGAGACCAG 120

CCTGGCCAAC ATGGTGAAAC CCATCTCTAC TAAAAATACA AAAATTAGCT GGGTGTGGTG 180

GCACGTACCT GTAATCCCAG CTACTCGGGA GGCGGAGGCA CAAGAATTGC TTGAACCTAG 240

GACGCGGAGG TTGCAGCGAG CCAAGATCGC GCCACTGCAC TCCAGCCTGG GCCGTAGAGT 300

GAGACTCTGT CTCAAAAAAG AAAAAAAAGT AATTGTTCTA GCTGGGCGCA GTGGCTCTTG 360

CCTGTAATCC CAGCACTTTG GGAGGCCAAG GCGGGTGGAT CTCGAGTCCT AGAGTTCAAG 420

ACCAGCCTAG GCAATGTGGT GAAACCCCAT CGCTACAAAA AATACAAAAA TTAGCCAGGC 480

ATGGTGGCGT GCGCATGTAG TCCCAGCTCC TTGGGAGGCT GAGGTGGGAG GATCACTTGA 540

ACCCAGGAGA CAGAGGTTGC AGTGAACCGA GATCACGCCA CCACGCTCCA GCCTGGGCAA 600

CAGAACAAGA CTCTGTCTAA AAAAATACAA ATAAAATAAA AGTAGTTCTC ACAGTACCAG 660

CATTCATTTT TCAAAAGATA TAGAGCTAAA AAGGAAGGAA AAAAAAAGTA ATGTTGGGCT 720

TTTAAATACT CGTTCCTATA CTAAATGTTC TTAGGAGTGC TGGGGTTTTA TTGTCATCAT 780

TTATCCTTTT TAAAAATGTT ATTGGCCAGG CACGGTGGCT CATGGCTGTA ATCCCAGCAC 840

TTTGGGAGGC CGAGGCAGGC AGATCACCTG AGGTCAGGAG TGTGAGACCA GCCTGGCCAA 900

CATGGCGAAA CCTGTCTCTA CTAAAAATAC AAAAATTAAC TAGGCGTGGT GGTGTACGCC 960

TGTAGTCCCA GCTACTCGGG AGGCTGAGGC AGGAGAATCA ACTGAACCAG GGAGGTGGAG 1020

GTTGCAGTGT GCCGAGATCA CGCCACTGCA CTCTAGCCTG GCAACAGAGC AAGATTCTGT 1080

CTCAAAAAAA AAAAACATAT ATACACATAT ATCCCAAAGT GCTGGGATTA CATATATATA 1140

TATATATATA TATTATATAT ATATATATAT ATATATGTGA TATATATGTG ATATATATAT 1200

-continued

```
AACATATATA TATGTAATAT ATATGTGATA TATATATAAT ATATATATGT AATATATATG   1260
TGATATATAT ATATACACAC ACACACACAT ATATATGTAT GTGTGTGTAC ACACACACAC   1320
ACAAATTAGC CAGGCATAGT TGCACACGCT TGGTAGACCC AGCTACTCAG GAGGCTGAGG   1380
GAGGAGAATC TCTTGAACTT AGGAGGCGGA GGTTGCAGTG AGCTGAGATT GCGCCACTGC   1440
ACTCCAGCCT GGGTGACAGA GCAGGACTCT GTACACCCCC CAAAACAAAA AAAAAAGTTA   1500
TCAGATGTGA TTGGAATGTA TATCAAGTAT CAGCTTCAAA ATATGCTATA TTAATACTTC   1560
AAAAATTACA CAAATAATAC ATAATCAGGT TTGAAAAATT TAAGACAACM SAARAAAAAA   1620
WYCMAATCAC AMATATCCCA CACATTTTAT TATTMCTMCT MCWATTATTT TGWAGAGMCT   1680
GGGTCTCACY CYKTTGCTWA TGCTGGTCTT TGAACYCCYK GCCYCAARCA RTCCTSCTCC   1740
ABCCTCCCAA RGTGCTGGGG ATWATAGGCA TGARCTAACC GCACCCAGCC CCAGACATTT   1800
TAGTGTGTAA ATTCCTGGGC ATTTTTTCAA GGCATCATAC ATGTTAGCTG ACTGATGATG   1860
GTCAATTTAT TTTGTCCATG GTGTCAAGTT TCTCTTCAGG AGGAAAAGCA CAGAACTGGC   1920
CAACAATTGC TTGACTGTTC TTTACCATAC TGTTTAGCAG GAAACCAGTC TCAGTGTCCA   1980
ACTCTCTAAC CTTGGAACTG TGAGAACTCT GAGGACAAAG CAGCGGATAC AACCTCAAAA   2040
GACGTCTGTC TACATTGAAT TGGGTAAGGG TCTCAGGTTT TTTAAGTATT TAATAATAAT   2100
TGCTGGATTC CTTATCTTAT AGTTTTGCCA AAAATCTTGG TCATAATTTG TATTTGTGGT   2160
AGGCAGCTTT GGGAAGTGAA TTTTATGAGC CCTATGGTGA GTTATAAAAA ATGTAAAAGA   2220
CGCAGTTCCC ACCTTGAAGA ATCTTACTTT AAAAAGGGAG CAAAAGAGGC CAGGCATGGT   2280
GGCTCACACC TGTAATCCCA GCACTTTGGG AGGCCAAAGT GGGTGGATCA CCTGAGGTCG   2340
GGAGTTCGAG ACCAGCCTAG CCAACATGGA GAAACTCTGT CTGTACCAAA AAATAAAAAA   2400
TTAGCCAGGT GTGGTGGCAC ATAACTGTAA TCCCAGCTAC TCGGGAGGCT GAGGCAGGAG   2460
AATCACTTGA ACCCGGGAGG TGGAGGTTGC GGTGAACCGA GATCGCACCA TTGCACTCCA   2520
GCCTGGGCAA AAATAGCGAA ACTCCATCTA AAAAAAAAAA AGAGAGCAAA AGAAAGAMTM   2580
TCTGGTTTTA AMTMTGTGTA AATATGTTTT TGGAAAGATG GAGAGTAGCA ATAAGAAAAA   2640
ACATGATGGA TTGCTACAGT ATTTAGTTCC AAGATAAATT GTACTAGATG AGGAAGCCTT   2700
TTAAGAAGAG CTGAATTGCC AGGCGCAGTG GCTCACGCCT GTAATCCCAG CACTTTGGGA   2760
GGCCGAGGTG GGCGGATCAC CTGAGGTCGG GAGTTCAAGA CCAGCCTGAC CAACATGGAG   2820
AAACCCCATC TCTACTAAAA AAAAAAAAAA AAAAATTAGC CGGGGTGGTG GCTTATGCCT   2880
GTAATCCCAG CTACTCAGGA GGCTGAGGCA GGAGAATCGC TTGAACCCAG GAAGCAGAGG   2940
TTGCAGTGAG CCAAGATCGC ACCATTGCAC TCCAGCCTAG GCAACAAGAG TGAAACTCCA   3000
TCTCAAAAAA AAAAAAAAAG AGCTGAATCT TGGCTGGGCA GGATGGCTCG TGCCTGTAAT   3060
CCTAACGCTT TGGAAGACCG AGGCAGAAGG ATTGGTTGAG TCCACGAGTT TAAGACCAGC   3120
CTGGCCAACA TAGGGGAACC CTGTCTCTAT TTTTAAAATA ATAATACATT TTTGGCCGGT   3180
GCGGTGGCTC ATGCCTGTAA TCCCAATACT TTGGGAGGCT GAGGCAGGTA GATCACCTGA   3240
GGTCAGAGTT CGAGACCAGC CTGGATAACC TGGTGAAACC CCTCTTTACT AAAAATACAA   3300
AAAAAAAAAA AAATTAGCTG GGTGTGGTAG CACATGCTTG TAATCCCAGC TACTTGGGAG   3360
GCTGAGGCAG GAGAATCGCT TGAACCAGGG AGGCGGAGGT TACAATGAGC CAACACTACA   3420
CCACTGCACT CCAGCCTGGG CAATAGAGTG AGACTGCATC TCAAAAAAAT AATAATTTTT   3480
AAAAATAATA AATTTTTTTA AGCTTATAAA AAGAAAAGTT GAGGCCAGCA TAGTAGCTCA   3540
CATCTGTAAT CTCAGCAGTG GCAGAGGATT GCTTGAAGCC AGGAGTTTGA GACCAGCCTG   3600
```

-continued

```
GGCAACATAG CAAGACCTCA TCTCTACAAA AAAATTTCTT TTTTAAATTA GCTGGGTGTG   3660
GTGGTGTGCA TCTGTAGTCC CAGCTACTCA GGAGGCAGAG GTGAGTGGAT ACATTGAACC   3720
CAGGAGTTTG AGGCTGTAGT GAGCTATGAT CATGCCACTG CACTCCAACC TGGGTGACAG   3780
AGCAAGACCT CCAAAAAAAA AAAAAAAAGA GCTGCTGAGC TCAGAATTCA AACTGGGCTC   3840
TCAAATTGGA TTTTCTTTTA GAATATATTT ATAATTAAAA AGGATAGCCA TCTTTTGAGC   3900
TCCCAGGCAC CACCATCTAT TTATCATAAC ACTTACTGTT TTCCCCCCTT ATGATCATAA   3960
ATTCCTAGAC AACAGGCATT GTAAAAATAG TTATAGTAGT TGATATTTAG GAGCACTTAA   4020
CTATATTCCA GGCACTATTG TGCTTTTCTT GTATAACTCA TTAGATGCTT GTCAGACCTC   4080
TGAGATTGTT CCTATTATAC TTATTTTACA GATGAGAAAA TTAAGGCACA GAGAAGTTAT   4140
GAAATTTTTC CAAGGTATTA AACCTAGTAA GTGGCTGAGC CATGATTCAA ACCTAGGAAG   4200
TTAGATGTCA GAGCCTGTGC TTTTTTTTTG TTTTTGTTTT TGTTTTCAGT AGAAACGGGG   4260
GTCTCACTTT GTTGGCCAGG CTGGTCTTGA ACTCCTAACC TCAAATAATC CACCCATCTC   4320
GGCCTCCTCA AGTGCTGGGA TTACAGGTGA GAGCCACTGT GCCTGGCGAA GCCCATGCCT   4380
TTAACCACTT CTCTGTATTA CATACTAGCT TAACTAGCAT TGTACCTGCC ACAGTAGATG   4440
CTCAGTAAAT ATTTCTAGTT GAATATCTGT TTTTCAACAA GTACATTTTT TTAACCCTTT   4500
TAATTAAGAA AACTTTTATT GATTTATTTT TTGGGGGGAA ATTTTTTAGG ATCTGATTCT   4560
TCTGAAGATA CCGTTAATAA GGCAACTTAT TGCAGGTGAG TCAAAGAGAA CCTTTGTCTA   4620
TGAAGCTGGT ATTTTCCTAT TTAGTTAATA TTAAGGATTG ATGTTTCTCT CTTTTTAAAA   4680
ATATTTTAAC TTTTATTTTA GGTTCAGGGA TGTATGTGCA GTTTGTTATA TAGGTAAACA   4740
CACGACTTGG GATTTGGTGT ATAGATTTTT TTCATCATCC GGGTACTAAG CATACCCCAC   4800
AGTTTTTTGT TTGCTTTCTT TCTGAATTTC TCCCTCTTCC CACCTTCCTC CCTCAAGTAG   4860
GCTGGTGTTT CTCCAGACTA GAATCATGGT ATTGGAAGAA ACCTTAGAGA TCATCTAGTT   4920
TAGTTCTCTC ATTTTATAGT GGAGGAAATA CCCTTTTTGT TTGTTGGATT TAGTTATTAG   4980
CACTGTCCAA AGGAATTTAG GATAACAGTA GAACTCTGCA CATGCTTGCT TCTAGCAGAT   5040
TGTTCTCTAA GTTCCTCATA TACAGTAATA TTGACACAGC AGTAATTGTG ACTGATGAAA   5100
ATGTTCAAGG ACTTCATTTT CAACTCTTTC TTTCCTCTGT TCCTTATTTC CACATATCTC   5160
TCAAGCTTTG TCTGTATGTT ATATAATAAA CTACAAGCAA CCCCAACTAT GTTACCTACC   5220
TTCCTTAGGA ATTATTGCTT GACCCAGGTT TTTTTTTTTT TTTTTTGGA GACGGGGTCT   5280
TGCCCTGTTG CCAGGATGGA GTGTAGTGGC GCCATCTCGG CTCACTGCAA TCTCCAACTC   5340
CCTGGTTCAA GCGATTCTCC TGTCTCAATC TCACGAGTAG CTGGGACTAC AGGTATACAC   5400
CACCACGCCC GGTTAATTGA CCATTCCATT TCTTTCTTTC TCTCTTTTTT TTTTTTTTTT   5460
TTGAGACAGA GTCTTGCTCT GTTGCCCAGG CTGGAGTACA GAGGTGTGAT CTCACCTCTC   5520
CGCAACGTCT GCCTCCCAGG TTGAAGCCAT ACTCCTGCCT CAGCCTCTCT AGTAGCTGGG   5580
ACTACAGGCG CGCGCCACCA CACCCGGCTA ATTTTTGTAT TTTTAGTAGA GATGGGGTTT   5640
CACCATGTTG GCCAGGCTGG TCTTGAACTC ATGACCTCAA GTGGTCCACC CGCCTCAGCC   5700
TCCCAAAGTG CTGGAATTAC AGGCTTGAGC CACCGTGCCC AGCAACCATT TCATTTCAAC   5760
TAGAAGTTTC TAAAGGAGAG AGCAGCTTTC ACTAACTAAA TAAGATTGGT CAGCTTTCTG   5820
TAATCGAAAG AGCTAAAATG TTTGATCTTG GTCATTTGAC AGTTCTGCAT ACATGTAACT   5880
AGTGTTTCTT ATTAGGACTC TGTCTTTTCC CTATAGTGTG GGAGATCAAG AATTGTTACA   5940
AATCACCCCT CAAGGAACCA GGGATGAAAT CAGTTTGGAT TCTGCAAAAA AGGGTAATGG   6000
```

-continued

```
CAAAGTTTGC CAACTTAACA GGCACTGAAA AGAGAGTGGG TAGATACAGT ACTGTAATTA   6060
GATTATTCTG AAGACCATTT GGGACCTTTA CAACCCACAA AATCTCTTGG CAGAGTTAGA   6120
GTATCATTCT CTGTCAAATG TCGTGGTATG GTCTGATAGA TTTAAATGGT ACTAGACTAA   6180
TGTACCTATA ATAAGACCTT CTTGTAACTG ATTGTTGCCC TTTCGCTTTT TTTTTTGTTT   6240
GTTTGTTTGT TTTTTTTTGA GATGGGGTCT CACTCTGTTG CCCAGGCTGG AGTGCAGTGA   6300
TGCAATCTTG GCTCACTGCA ACCTCCACCT CCAAAGGCTC AAGCTATCCT CCCACTTCAG   6360
CCTCCTGAGT AGCTGGGACT ACAGGCGCAT GCCACCACAC CCGGTTAATT TTTTGTGGTT   6420
TTATAGAGAT GGGGTTTCAC CATGTTACCG AGGCTGGTCT CAAACTCCTG GACTCAAGCA   6480
GTCTGCCCAC TTCAGCCTCC CAAAGTGCTG CAGTTACAGG CTTGAGCCAC TGTGCCTGGC   6540
CTGCCCTTTA CTTTTAATTG GTGTATTTGT GTTTCATCTT TTACCTACTG GTTTTTAAAT   6600
ATAGGGAGTG GTAAGTCTGT AGATAGAACA GAGTATTAAG TAGACTTAAT GGCCAGTAAT   6660
CTTTAGAGTA CATCAGAACC AGTTTTCTGA TGGCCAATCT GCTTTTAATT CACTCTTAGA   6720
CGTTAGAGAA ATAGGTGTGG TTTCTGCATA GGGAAAATTC TGAAATTAA               6769
```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4249 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```
GATCCTAAGT GGAAATAATC TAGGTAAATA GGAATTAAAT GAAAGAGTAT GAGCTACATC     60
TTCAGTATAC TTGGTAGTTT ATGAGGTTAG TTTCTCTAAT ATAGCCAGTT GGTTGATTTC    120
CACCTCCAAG GTGTATGAAG TATGTATTTT TTTAATGACA ATTCAGTTTT TGAGTACCTT    180
GTTATTTTTG TATATTTTCA GCTGCTTGTG AATTTTCTGA GACGGATGTA ACAAATACTG    240
AACATCATCA ACCCAGTAAT AATGATTTGA ACACCACTGA GAAGCGTGCA GCTGAGAGGC    300
ATCCAGAAAA GTATCAGGGT AGTTCTGTTT CAAACTTGCA TGTGGAGCCA TGTGGCACAA    360
ATACTCATGC CAGCTCATTA CAGCATGAGA ACAGCAGTTT ATTACTCACT AAAGACAGAA    420
TGAATGTAGA AAAGGCTGAA TTCTGTAATA AAAGCAAACA GCCTGGCTTA GCAAGGAGCC    480
AACATAACAG ATGGGCTGGA AGTAAGGAAA CATGTAATGA TAGGCGGACT CCCAGCACAG    540
AAAAAAAGGT AGATCTGAAT GCTGATCCCC TGTGTGAGAG AAAAGAATGG AATAAGCAGA    600
AACTGCCATG CTCAGAGAAT CCTAGAGATA CTGAAGATGT TCCTTGGATA ACACTAAATA    660
GCAGCATTCA GAAAGTTAAT GAGTGGTTTT CCAGAAGTGA TGAACTGTTA GGTTCTGATG    720
ACTCACATGA TGGGGAGTCT GAATCAAATG CCAAAGTAGC TGATGTATTG GACGTTCTAA    780
ATGAGGTAGA TGAATATTCT GGTTCTTCAG AGAAAATAGA CTTACTGGCC AGTGATCCTC    840
ATGAGGCTTT AATATGTAAA AGTGAAAGAG TTCACTCCAA ATCAGTAGAG AGTAATATTG    900
AAGGCCAAAT ATTTGGGAAA ACCTATCGGA AGAAGGCAAG CCTCCCCAAC TTAAGCCATG    960
TAACTGAAAA TCTAATTATA GGAGCATTTG TTACTGAGCC ACAGATAATA CAAGAGCGTC   1020
```

-continued

```
CCCTCACAAA TAAATTAAAG CGTAAAAGGA GACCTACATC AGGCCTTCAT CCTGAGGATT   1080
TTATCAAGAA AGCAGATTTG GCAGTTCAAA AGACTCCTGA AATGATAAAT CAGGGAACTA   1140
ACCAAACGGA GCAGAATGGT CAAGTGATGA ATATTACTAA TAGTGGTCAT GAGAATAAAA   1200
CAAAAGGTGA TTCTATTCAG AATGAGAAAA ATCCTAACCC AATAGAATCA CTCGAAAAAG   1260
AATCTGCTTT CAAAACGAAA GCTGAACCTA TAAGCAGCAG TATAAGCAAT ATGGAACTCG   1320
AATTAAATAT CCACAATTCA AAAGCACCTA AAAAGAATAG GCTGAGGAGG AAGTCTTCTA   1380
CCAGGCATAT TCATGCGCTT GAACTAGTAG TCAGTAGAAA TCTAAGCCCA CCTAATTGTA   1440
CTGAATTGCA AATTGATAGT TGTTCTAGCA GTGAAGAGAT AAAGAAAAAA AAGTACAACC   1500
AAATGCCAGT CAGGCACAGC AGAAACCTAC AACTCATGGA AGGTAAAGAA CCTGCAACTG   1560
GAGCCAAGAA GAGTAACAAG CCAAATGAAC AGACAAGTAA AAGACATGAC AGCGATACTT   1620
TCCCAGAGCT GAAGTTAACA AATGCACCTG GTTCTTTTAC TAAGTGTTCA AATACCAGTG   1680
AACTTAAAGA ATTTGTCAAT CCTAGCCTTC CAAGAGAAGA AAAAGAAGAG AACTAGAAAC   1740
AGTTAAAGTG TCTAATAATG CTGAAGACCC CAAAGATCTC ATGTTAAGTG GAGAAAGGGT   1800
TTTGCAAACT GAAAGATCTG TAGAGAGTAG CAGTATTTCA TTGGTACCTG GTACTGATTA   1860
TGGCACTCAG GAAAGTATCT CGTTACTGGA AGTTAGCACT CTAGGGAAGG CAAAAACAGA   1920
ACCAAATAAA TGTGTGAGTC AGTGTGCAGC ATTTGAAAAC CCCAAGGGAC TAATTCATGG   1980
TTGTTCCAAA GATAATAGAA ATGACACAGA AGGCTTTAAG TATCCATTGG GACATGAAGT   2040
TAACCACAGT CGGGAAACAA GCATAGAAAT GGAAGAAAGT GAACTTGATG CTCAGTATTT   2100
GCAGAATACA TTCAAGGTTT CAAAGCGCCA GTCATTTGCT CCGTTTTCAA ATCCAGGAAA   2160
TGCAGAAGAG GAATGTGCAA CATTCTCTGC CCACTCTGGG TCCTTAAAGA AACAAAGTCC   2220
AAAAGTCACT TTTGAATGTG AACAAAAGGA AGAAAATCAA GGAAAGAATG AGTCTAATAT   2280
CAAGCCTGTA CAGACAGTTA ATATCACTGC AGGCTTTCCT GTGGTTGGTC AGAAAGATAA   2340
GCCAGTTGAT AATGCCAAAT GTAGTATCAA AGGAGGCTCT AGGTTTTGTC TATCATCTCA   2400
GTTCAGAGGC AACGAAACTG GACTCATTAC TCCAAATAAA CATGGACTTT TACAAAACCC   2460
ATATCGTATA CCACCACTTT TTCCCATCAA GTCATTTGTT AAAACTAAAT GTAAGAAAAA   2520
TCTGCTAGAG GAAAACTTTG AGGAACATTC AATGTCACCT GAAAGAGAAA TGGGAAATGA   2580
GAACATTCCA AGTACAGTGA GCACAATTAG CCGTAATAAC ATTAGAGAAA ATGTTTTTAA   2640
AGAAGCCAGC TCAAGCAATA TTAATGAAGT AGGTTCCAGT ACTAATGAAG TGGGCTCCAG   2700
TATTAATGAA ATAGGTTCCA GTGATGAAAA CATTCAAGCA GAACTAGGTA GAAACAGAGG   2760
GCCAAAATTG AATGCTATGC TTAGATTAGG GGTTTTGCAA CCTGAGGTCT ATAAACAAAG   2820
TCTTCCTGGA AGTAATTGTA AGCATCCTGA AATAAAAAAG CAAGAATATG AAGAAGTAGT   2880
TCAGACTGTT AATACAGATT TCTCTCCATA TCTGATTTCA GATAACTTAG AACAGCCTAT   2940
GGGAAGTAGT CATGCATCTC AGGTTTGTTC TGAGACACCT GATGACCTGT TAGATGATGG   3000
TGAAATAAAG GAAGATACTA GTTTTGCTGA AAATGACATT AAGGAAAGTT CTGCTGTTTT   3060
TAGCAAAAGC GTCCAGAAAG GAGAGCTTAG CAGGAGTCCT AGCCCTTTCA CCCATACACA   3120
TTTGGCTCAG GGTTACCGAA GAGGGGCCAA GAAATTAGAG TCCTCAGAAG AGAACTTATC   3180
TAGTGAGGAT GAAGAGCTTC CCTGCTTCCA ACACTTGTTA TTTGGTAAAG TAAACAATAT   3240
ACCTTCTCAG TCTACTAGGC ATAGCACCGT TGCTACCGAG TGTCTGTCTA AGAACACAGA   3300
GGAGAATTTA TTATCATTGA AGAATAGCTT AAATGACTGC AGTAACCAGG TAATATTGGC   3360
AAAGGCATCT CAGGAACATC ACCTTAGTGA GGAAACAAAA TGTTCTGCTA GCTTGTTTTC   3420
```

-continued

```
TTCACAGTGC AGTGAATTGG AAGACTTGAC TGCAAATACA AACACCCAGG ATCCTTTCTT   3480
GATTGGTTCT TCCAAACAAA TGAGGCATCA GTCTGAAAGC CAGGGAGTTG GTCTGAGTGA   3540
CAAGGAATTG GTTTCAGATG ATGAAGAAAG AGGAACGGGC TTGGAAGAAA ATAATCAAGA   3600
AGAGCAAAGC ATGGATTCAA ACTTAGGTAT TGGAACCAGG TTTTTGTGTT TGCCCCAGTC   3660
TATTTATAGA AGTGAGCTAA ATGTTTATGC TTTTGGGGAG CACATTTTAC AAATTTCCAA   3720
GTATAGTTAA AGGAACTGCT TCTTAAACTT GAAACATGTT CCTCCTAAGG TGCTTTTCAT   3780
AGAAAAAAGT CCTTCACACA GCTAGGACGT CATCTTTGAC TGAATGAGCT TTAACATCCT   3840
AATTACTGGT GGACTTACTT CTGGTTTCAT TTTATAAAGC AAATCCCGGT GTCCCAAAGC   3900
AAGGAATTTA ATCATTTTGT GTGACATGAA AGTAAATCCA GTCCTGCCAA TGAGAAGAAA   3960
AAGACACAGC AAGTTGCAGC GTTTATAGTC TGCTTTTACA TCTGAACCTC TGTTTTTGTT   4020
ATTTAAGGTG AAGCAGCATC TGGGTGTGAG AGTGAAACAA GCGTCTCTGA AGACTGCTCA   4080
GGGCTATCCT CTCAGAGTGA CATTTTAACC ACTCAGGTAA AAAGCGTGTG TGTGTGTGCA   4140
CATGCGTGTG TGTGGTGTCC TTTGCATTCA GTAGTATGTA TCCCACATTC TTAGGTTTGC   4200
TGACATCATC TCTTTGAATT AATGGCACAA TTGTTTGTGG TTCATTGTC               4249
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 710 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
NGNGAATGTA ATCCTAATAT TTCNCNCCNA CTTAAAAGAA TACCACTCCA ANGGCATCNC     60
AATACATCAA TCAATTGGGG AATTGGGATT TTCCCTCNCT AACATCANTG GAATAATTTC    120
ATGGCATTAA TTGCATGAAT GTGGTTAGAT TAAAAGGTGT TCATGCTAGA ACTTGTAGTT    180
CCATACTAGG TGATTTCAAT TCCTGTGCTA AAATTAATTT GTATGATATA TTNTCATTTA    240
ATGGAAAGCT TCTCAAAGTA TTTCATTTTC TTGGTACCAT TTATCGTTTT TGAAGCAGAG    300
GGATACCATG CAACATAACC TGATAAAGCT CCAGCAGGAA ATGGCTGAAC TAGAAGCTGT    360
GTTAGAACAG CATGGGAGCC AGCCTTCTAA CAGCTACCCT TCCATCATAA GTGACTCTTC    420
TGCCCTTGAG GACCTGCGAA ATCCAGAACA AAGCACATCA GAAAAAGGTG TGTATTGTTG    480
GCCAAACACT GATATCTTAA GCAAAATTCT TTCCTTCCCC TTTATCTCCT TCTGAAGAGT    540
AAGGACCTAG CTCCAACATT TTATGATCCT TGCTCAGCAC ATGGGTAATT ATGGAGCCTT    600
GGTTCTTGTC CCTGCTCACA ACTAATATAC CAGTCAGAGG GACCCAAGGC AGTCATTCAT    660
GTTGTCATCT GAGATACCTA CAACAAGTAG ATGCTATGGG GAGCCCATGG               710
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 473 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

-continued

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCATTGGTGC TAGCATCTGT CTGTTGCATT GCTTGTGTTT ATAAAATTCT GCCTGATATA 60

CTTGTTAAAA ACCAATTTGT GTATCATAGA TTGATGCTTT TGAAAAAAAT CAGTATTCTA 120

ACCTGAATTA TCACTATCAG AACAAAGCAG TAAAGTAGAT TTGTTTTCTC ATTCCATTTA 180

AAGCAGTATT AACTTCACAG AAAAGTAGTG AATACCCTAT AAGCCAGAAT CCAGAAGGCC 240

TTTCTGCTGA CAAGTTTGAG GTGTCTGCAG ATAGTTCTAC CAGTAAAAAT AAAGAACCAG 300

GAGTGGAAAG GTAAGAAACA TCAATGTAAA GATGCTGTGG TATCTGACAT CTTTATTTAT 360

ATTGAACTCT GATTGTTAAT TTTTTTCACC ATACTTTCTC CAGTTTTTTT GCATACAGGC 420

ATTTATACAC TTTTATTGCT CTAGGATACT TCTTTTGTTT AATCCTATAT AGG 473

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 421 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GGATAAGNTC AAGAGATATT TTGATAGGTG ATGCAGTGAT NAATTGNGAA AATTTNCTGC 60

CTGCTTTTAA TCTTCCCCCG TTCTTTCTTC CTNCCTCCCT CCCTTCCTNC CTCCCGTCCT 120

TNCCTTTCCT TTCCCTCCCT TCCNCCTTCT TTCCNTCTNT CTTTCCTTTC TTTCCTGTCT 180

ACCTTTCTTT CCTTCCTCCC TTCCTTTTCT TTTCTTTCTT TCCTTTCCTT TTCTTTCCTT 240

TCTTTCCTTT CCTTTCTTTC TTGACAGAGT CTTGCTCTGT CACTCAGGCT GGAGTGCAGT 300

GGCGTGATCT CGNCTCACTG CAACCTCTGT CTCCCAGGTT CAAGCAATTT TCCTGCCTCA 360

GCCTCCCGAG TAGCTGAGAT TACAGGCGCC AGCCACCACA CCCAGCTACT GACCTGCTTT 420

T 421

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 997 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

-continued

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:25:

AAACAGCTGG GAGATATGGT GCCTCAGACC AACCCCATGT TATATGTCAA CCCTGACATA 60

TTGGCAGGCA ACATGAATCC AGACTTCTAG GCTGTCATGC GGGCTCTTTT TTGCCAGTCA 120

TTTCTGATCT CTCTGACATG AGCTGTTTCA TTTATGCTTT GGCTGCCCAG CAAGTATGAT 180

TTGTCCTTTC ACAATTGGTG GCGATGGTTT TCTCCTTCCA TTTATCTTTC TAGGTCATCC 240

CCTTCTAAAT GCCCATCATT AGATGATAGG TGGTACATGC ACAGTTGCTC TGGGAGTCTT 300

CAGAATAGAA ACTACCCATC TCAAGAGGAG CTCATTAAGG TTGTTGATGT GGAGGAGCAA 360

CAGCTGGAAG AGTCTGGGCC ACACGATTTG ACGGAAACAT CTTACTTGCC AAGGCAAGAT 420

CTAGGTAATA TTTCATCTGC TGTATTGGAA CAAACACTYT GATTTTACTC TGAATCCTAC 480

ATAAAGATAT TCTGGTTAAC CAACTTTTAG ATGTACTAGT CTATCATGGA CACTTTTGTT 540

ATACTTAATT AAGCCCACTT TAGAAAAATA GCTCAAGTGT TAATCAAGGT TTACTTGAAA 600

ATTATTGAAA CTGTTAATCC ATCTATATTT TAATTAATGG TTTAACTAAT GATTTTGAGG 660

ATGWGGGAGT CKTGGTGTAC TCTAMATGTA TTATTTCAGG CCAGGCATAG TGGCTCACGC 720

CTGGTAATCC CAGTAYYCMR GAGCCCGAGG CAGGTGGAGC CAGCTGAGGT CAGGAGTTCA 780

AGACCTGTCT TGGCCAACAT GGGNGAAACC CTGTCTTCTT CTTAAAAAAN ACAAAAAAAA 840

TTAACTGGGT TGTGCTTAGG TGNATGCCCC GNATCCTAGT TNTTCTTGNG GGTTGAGGGA 900

GGAGATCACN TTGGACCCCG GAGGGGNGGG TGGGGGNGAG CAGGNCAAAA CACNGACCCA 960

GCTGGGGTGG AAGGGAAGCC CACTCNAAAA AANNTTN 997

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 639 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TTTTTAGGAA ACAAGCTACT TTGGATTTCC ACCAACACCT GTATTCATGT ACCCATTTTT 60

CTCTTAACCT AACTTTATTG GTCTTTTTAA TTCTTAACAG AGACCAGAAC TTTGTAATTC 120

AACATTCATC GTTGTGTAAA TTAAACTTCT CCCATTCCTT TCAGAGGGAA CCCCTTACCT 180

GGAATCTGGA ATCAGCCTCT TCTCTGATGA CCCTGAATCT GATCCTTCTG AAGACAGAGC 240

CCCAGAGTCA GCTCGTGTTG GCAACATACC ATCTTCAACC TCTGCATTGA AAGTTCCCCA 300

ATTGAAAGTT GCAGAATCTG CCCAGAGTCC AGCTGCTGCT CATACTACTG ATACTGCTGG 360

GTATAATGCA ATGGAAGAAA GTGTGAGCAG GGAGAAGCCA GAATTGACAG CTTCAACAGA 420

AAGGGTCAAC AAAAGAATGT CCATGGTGGT GTCTGGCCTG ACCCCAGAAG AATTTGTGAG 480

TGTATCCATA TGTATCTCCC TAATGACTAA GACTTAACAA CATTCTGGAA AGAGTTTTAT 540

GTAGGTATTG TCAATTAATA ACCTAGAGGA AGAAATCTAG AAAACAATCA CAGTTCTGTG 600

TAATTTAATT TCGATTACTA ATTTCTGAAA ATTTAGAAY 639

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 922 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```
NCCCNNCCCC CNAATCTGAA ATGGGGGTAA CCCCCCCCCA ACCGANACNT GGGTNGCNTA      60
GAGANTTTAA TGGCCCNTTC TGAGGNACAN AAGCTTAAGC CAGGNGACGT GGANCNATGN     120
GTTGTTTNTT GTTTGGTTAC CTCCAGCCTG GGTGACAGAG CAAGACTCTG TCTAAAAAAA     180
AAAAAAAAAA AAATCGACTT TAAATAGTTC CAGGACACGT GTAGAACGTG CAGGATTGCT     240
ACGTAGGTAA ACATATGCCA TGGTGGGATA ACTAGTATTC TGAGCTGTGT GCTAGAGGTA     300
ACTCATGATA ATGGAATATT TGATTTAATT TCAGATGCTC GTGTACAAGT TTGCCAGAAA     360
ACACCACATC ACTTTAACTA ATCTAATTAC TGAAGAGACT ACTCATGTTG TTATGAAAAC     420
AGGTATACCA AGAACCTTTA CAGAATACCT TGCATCTGCT GCATAAAACC ACATGAGGCG     480
AGGCACGGTG GCGCATGCCT GTAATCGCAG CACTTTGGGA GGCCGAGGCG GGCAGATCAC     540
GAGATTAGGA GATCGAGACC ATCCTGGCCA GCATGGTGAA ACCCCGTCTC TACTANNAAA     600
TGGNAAAATT ANCTGGGTGT GGTCGCGTGC NCCTGTAGTC CCAGCTACTC GTGAGGCTGA     660
GGCAGGAGAA TCACTTGAAC CGGGGAAATG GAGGTTTCAG TGAGCAGAGA TCATNCCCCT     720
NCATTCCAGC CTGGCGACAG AGCAAGGCTC CGTCNCCNAA AAAATAAAAA AAAACGTGAA     780
CAAATAAGAA TATTTGTTGA GCATAGCATG GATGATAGTC TTCTAATAGT CAATCAATTA     840
CTTTATGAAA GACAAATAAT AGTTTTGCTG CTTCCTTACC TCCTTTTGTT TTGGGTTAAG     900
ATTTGGAGTG TGGGCCAGGC AC                                              922
```

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 867 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```
GATCTATAGC TAGCCTTGGC GTCTAGAAGA TGGGTGTTGA GAAGAGGGAG TGGAAAGATA      60
TTTCCTCTGG TCTTAACTTC ATATCAGCCT CCCCTAGACT TCCAAATATC CATACCTGCT     120
GGTTATAATT AGTGGTGTTT TCAGCCTCTG ATTCTGTCAC CAGGGGTTTT AGAATCATAA     180
ATCCAGATTG ATCTTGGGAG TGTAAAAAAC TGAGGCTCTT TAGCTTCTTA GGACAGCACT     240
```

-continued

```
TCCTGATTTT GTTTTCAACT TCTAATCCTT TGAGTGTTTT TCATTCTGCA GATGCTGAGT     300
TTGTGTGTGA ACGGACACTG AAATATTTTC TAGGAATTGC GGGAGGAAAA TGGGTAGTTA     360
GCTATTTCTG TAAGTATAAT ACTATTTCTC CCCTCCTCCC TTTAACACCT CAGAATTGCA     420
TTTTTACACC TAACATTTAA CACCTAAGGT TTTTGCTGAT GCTGAGTCTG AGTTACCAAA     480
AGGTCTTTAA ATTGTAATAC TAAACTACTT TTATCTTTAA TATCACTTTG TTCAAGATAA     540
GCTGGTGATG CTGGGAAAAT GGGTCTCTTT TATAACTAAT AGGACCTAAT CTGCTCCTAG     600
CAATGTTAGC ATATGAGCTA GGGATTTATT TAATAGTCGG CAGGAATCCA TGTGCARCAG     660
NCAAACTTAT AATGTTTAAA TTAAACATCA ACTCTGTCTC CAGAAGGAAA CTGCTGCTAC     720
AAGCCTTATT AAAGGGCTGT GGCTTTAGAG GGAAGGACCT CTCCTCTGTC ATTCTTCCTG     780
TGCTCTTTTG TGAATCGCTG ACCTCTCTAT CTCCGTGAAA AGAGCACGTT CTTCTGCTGT     840
ATGTAACCTG TCTTTTCTAT GATCTCT                                         867
```

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 561 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```
NAAAAACGGG GNNGGGANTG GGCCTTAAAN CCAAAGGGCN AACTCCCCAA CCATTNAAAA      60
ANTGACNGGG GATTATTAAA ANCGGCGGGA AACATTTCAC NGCCCAACTA ATATTGTTAA     120
ATTAAAACCA CCACCNCTGC NCCAAGGAGG GAAACTGCTG CTACAAGCCT TATTAAAGGG     180
CTGTGGCTTT AGAGGGAAGG ACCTCTCCTC TGTCATTCTT CCTGTGCTCT TTTGTGAATC     240
GCTGACCTCT CTATGTCCGT GAAAAGAGCA CGTTCTTCGT CTGTATGTAA CCTGTCTTTT     300
CTATGATCTC TTTAGGGGTG ACCCAGTCTA TTAAAGAAAG AAAAATGCTG AATGAGGTAA     360
GTACTTGATG TTACAAACTA ACCAGAGATA TTCATTCAGT CATATAGTTA AAAATGTATT     420
TGCTTCCTTC CATCAATGCA CCACTTTCCT TAACAATGCA CAAATTTTCC ATGATAATGA     480
GGATCATCAA GAATTATGCA GGCCTGCACT GTGGCTCATA CCTATAATCC CAGCGCTTTG     540
GGAGGCTGAG GCGCTTGGAT C                                               561
```

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 567 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```
AATTTTTTGT ATTTTTAGTA GAGATGAGGT TCACCATGTT GGTCTAGATC TGGTGTCGAA      60
CGTCCTGACC TCAAGTGATC TGCCAGCCTC AGTCTCCCAA AGTGCTAGGA TTACAGGGGT     120
GAGCCACTGC GCCTGGCCTG AATGCCTAAA ATATGACGTG TCTGCTCCAC TTCCATTGAA     180
GGAAGCTTCT CTTTCTCTTA TCCTGATGGG TTGTGTTTGG TTTCTTTCAG CATGATTTTG     240
AAGTCAGAGG AGATGTGGTC AATGGAAGAA ACCACCAAGG TCCAAAGCGA GCAAGAGAAT     300
CCCAGGACAG AAAGGTAAAG CTCCCTCCCT CAAGTTGACA AAAATCTCAC CCCACCACTC     360
TGTATTCCAC TCCCCTTTGC AGAGATGGGC CGCTTCATTT TGTAAGACTT ATTACATACA     420
TACACAGTGC TAGATACTTT CACACAGGTT CTTTTTTCAC TCTTCCATCC CAACCACATA     480
AATAAGTATT GTCTCTACTT TATGAATGAT AAAACTAAGA GATTTAGAGA GGCTGTGTAA     540
TTTGGATTCC CGTCTCGGGT TCAGATC                                         567
```

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 633 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```
TTGGCCTGAT TGGTGACAAA AGTGAGATGC TCAGTCCTTG AATGACAAAG AATGCCTGTA      60
GAGTTGCAGG TCCAACTACA TATGCACTTC AAGAAGATCT TCTGAAATCT AGTAGTGTTC     120
TGGACATTGG ACTGCTTGTC CCTGGGAAGT AGCAGCAGAA ATGATCGGTG GTGAACAGAA     180
GAAAAAGAAA AGCTCTTCCT TTTTGAAAGT CTGTTTTTTG AATAAAAGCC AATATTCTTT     240
TATAACTAGA TTTTCCTTCT CTCCATTCCC CTGTCCCTCT CTCTTCCTCT CTTCTTCCAG     300
ATCTTCAGGG GGCTAGAAAT CTGTTGCTAT GGGCCCTTCA CCAACATGCC CACAGGTAAG     360
AGCCTGGGAG AACCCCAGAG TTCCAGCACC AGCCTTTGTC TTACATAGTG GAGTATTATA     420
AGCAAGGTCC CACGATGGGG GTTCCTCAGA TTGCTGAAAT GTTCTAGAGG CTATTCTATT     480
TCTCTACCAC TCTCCAAACA AAACAGCACC TAAATGTTAT CCTATGGCAA AAAAAAACTA     540
TACCTTGTCC CCCTTCTCAA GAGCATGAAG GTGGTTAATA GTTAGGATTC AGTATGTTAT     600
GTGTTCAGAT GGCGTTGAGC TGCTGTTAGT GCC                                  633
```

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 470 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

-continued

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TTTGAGAGAC TATCAAACCT TATACCAAGT GGCCTTATGG AGACTGATAA CCAGAGTACA 60

TGGCATATCA GTGGCAAATT GACTTAAAAT CCATACCCCT ACTATTTTAA GACCATTGTC 120

CTTTGGAGCA GAGAGACAGA CTCTCCCATT GAGAGGTCTT GCTATAAGCC TTCATCCGGA 180

GAGTGTAGGG TAGAGGGCCT GGGTTAAGTA TGCAGATTAC TGCAGTGATT TTACATGTAA 240

ATGTCCATTT TAGATCAACT GGAATGGATG GTACAGCTGT GTGGTGCTTC TGTGGTGAAG 300

GAGCTTTCAT CATTCACCCT TGGCACAGTA AGTATTGGGT GCCCTGTCAG TGTGGGAGGA 360

CACAATATTC TCTCCTGTGA GCAAGACTGG CACCTGTCAG TCCCTATGGA TGCCCCTACT 420

GTAGCCTCAG AAGTCTTCTC TGCCCACATA CCTGTGCCAA AAGACTCCAT 470

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 517 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGTGGTACGT GTCTGTAGTT CCAGCTACTT GGGAGGCTGA GATGGAAGGA TTGCTTGAGC 60

CCAGGAGGCA GAGGTGGNAN NTTACGCTGA GATCACACCA CTGCACTCCA GCCTGGGTGA 120

CAGAGCAAGA CCCTGTCTCA AAAACAAACA AAAAAAATGA TGAAGTGACA GTTCCAGTAG 180

TCCTACTTTG ACACTTTGAA TGCTCTTTCC TTCCTGGGGA TCCAGGGTGT CCACCCAATT 240

GTGGTTGTGC AGCCAGATGC CTGGACAGAG GACAATGGCT TCCATGGTAA GGTGCCTCGC 300

ATGTACCTGT GCTATTAGTG GGGTCCTTGT GCATGGGTTT GGTTTATCAC TCATTACCTG 360

GTGCTTGAGT AGCACAGTTC TTGGCACATT TTTAAATATT TGTTGAATGA ATGGCTAAAA 420

TGTCTTTTTG ATGTTTTTAT TGTTATTTGT TTTATATTGT AAAAGTAATA CATGAACTGT 480

TTCCATGGGG TGGGAGTAAG ATATGAATGT TCATCAC 517

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 434 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CAGTAATCCT NAGAACTCAT ACGACCGGGC CCCTGGAGTC GNTGNTTNGA GCCTAGTCCN 60

-continued

GGAGAATGAA TTGACACTAA TCTCTGCTTG TGTTCTCTGT CTCCAGCAAT TGGGCAGATG 120

TGTGAGGCAC CTGTGGTGAC CCGAGAGTGG GTGTTGGACA GTGTAGCACT CTACCAGTGC 180

CAGGAGCTGG ACACCTACCT GATACCCCAG ATCCCCCACA GCCACTACTG ACTGCAGCCA 240

GCCACAGGTA CAGAGCCACA GGACCCCAAG AATGAGCTTA CAAAGTGGCC TTTCCAGGCC 300

CTGGGAGCTC CTCTCACTCT TCAGTCCTTC TACTGTCCTG GCTACTAAAT ATTTTATGTA 360

CATCAGCCTG AAAAGGACTT CTGGCTATGC AAGGGTCCCT TAAAGATTTT CTGCTTGAAG 420

TCTCCCTTGG AAAT 434

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GATAAATTAA AACTGCGACT GCGCGGCGTG 30

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GTAGTAGAGT CCCGGGAAAG GGACAGGGGG 30

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:37:

ATATATATAT GTTTTTCTAA TGTGTTAAAG 30

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GTAAGTCAGC ACAAGAGTGT ATTAATTTGG 30

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TTTCTTTTTC TCCCCCCCCT ACCCTGCTAG 30

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GTAAGTTTGA ATGTGTTATG TGGCTCCATT 30

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AGCTACTTTT TTTTTTTTTT TTTGAGACAG 30

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GTAAGTGCAC ACCACCATAT CCAGCTAAAT 30

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:43:

AATTGTTCTT TCTTTCTTTA TAATTTATAG 30

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GTATATAATT TGGTAATGAT GCTAGGTTGG 30

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GAGTGTGTTT CTCAAACAAT TTAATTTCAG 30

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GTAAGTGTTG AATATCCCAA GAATGACACT 30

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:47:

AAACATAATG TTTTCCCTTG TATTTTACAG 30

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GTAAAACCAT TTGTTTTCTT CTTCTTCTTC 30

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:49:

TGCTTGACTG TTCTTTACCA TACTGTTTAG 30

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GTAAGGGTCT CAGGTTTTTT AAGTATTTAA 30

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:51:

TGATTTATTT TTTGGGGGGA AATTTTTTAG 30

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GTGAGTCAAA GAGAACCTTT GTCTATGAAG 30

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:53:

TCTTATTAGG ACTCTGTCTT TTCCCTATAG 30

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GTAATGGCAA AGTTTGCCAA CTTAACAGGC 30

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GAGTACCTTG TTATTTTTGT ATATTTTCAG 30

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GTATTGGAAC CAGGTTTTTG TGTTTGCCCC 30

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:57:

ACATCTGAAC CTCTGTTTTT GTTATTTAAG 30

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:58:

AGGTAAAAAG CGTGTGTGTG TGTGCACATG 30

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CATTTTCTTG GTACCATTTA TCGTTTTTGA 30

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:60:

GTGTGTATTG TTGGCCAAAC ACTGATATCT 30

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:61:

AGTAGATTTG TTTTCTCATT CCATTTAAAG 30

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GTAAGAAACA TCAATGTAAA GATGCTGTGG 30

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:63:

ATGGTTTTCT CCTTCCATTT ATCTTTCTAG 30

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GTAATATTTC ATCTGCTGTA TTGGAACAAA 30

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:65:

TGTAAATTAA ACTTCTCCCA TTCCTTTCAG 30

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GTGAGTGTAT CCATATGTAT CTCCCTAATG 30

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:67:

ATGATAATGG AATATTTGAT TTAATTTCAG 30

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GTATACCAAG AACCTTTACA GAATACCTTG 30

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CTAATCCTTT GAGTGTTTTT CATTCTGCAG 30

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:70:

GTAAGTATAA TACTATTTCT CCCCTCCTCC 30

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:71:

TGTAACCTGT CTTTTCTATG ATCTCTTTAG 30

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:72:

GTAAGTACTT GATGTTACAA ACTAACCAGA 30

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:73:

TCCTGATGGG TTGTGTTTGG TTTCTTTCAG 30

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:74:

GTAAAGCTCC CTCCCTCAAG TTGACAAAAA 30

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:75:

CTGTCCCTCT CTCTTCCTCT CTTCTTCCAG 30

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:76:

GTAAGAGCCT GGGAGAACCC CAGAGTTCCA 30

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:77:

AGTGATTTTA CATGTAAATG TCCATTTTAG 30

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:78:

GTAAGTATTG GGTGCCCTGT CAGTGTGGGA 30

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:79:

TTGAATGCTC TTTCCTTCCT GGGGATCCAG 30

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:80:

GTAAGGTGCC TCGCATGTAC CTGTGCTATT 30

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:81:

CTAATCTCTG CTTGTGTTCT CTGTCTCCAG 30

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 amino acids

-continued

(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

```
Cys Pro Ile Cys Leu Glu Leu Ile Lys Glu Pro Val Ser Thr Lys Cys
1               5                  10                  15

Asp His Ile Phe Cys Lys Phe Cys Met Leu Lys Leu Leu Asn Gln Lys
            20                  25                  30

Lys Gly Pro Ser Gln Cys Pro Leu Cys Lys
        35                  40
```

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

```
Cys Pro Ile Cys Leu Glu Leu Leu Lys Glu Pro Val Ser Ala Asp Cys
1               5                  10                  15

Asn His Ser Phe Cys Arg Ala Cys Ile Thr Leu Asn Tyr Glu Ser Asn
            20                  25                  30

Arg Asn Thr Asp Gly Lys Gly Asn Cys Pro Val Cys Arg
        35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

```
Cys Pro Ile Cys Leu Asp Met Leu Lys Asn Thr Met Thr Thr Lys Glu
1               5                  10                  15

Cys Leu His Arg Phe Cys Ser Asp Cys Ile Val Thr Ala Leu Arg Ser
            20                  25                  30

Gly Asn Lys Glu Cys Pro Thr Cys Arg
        35                  40
```

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

-continued

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

```
Cys Pro Val Cys Leu Gln Tyr Phe Ala Glu Pro Met Met Leu Asp Cys
1               5                  10                  15

Gly His Asn Ile Cys Cys Ala Cys Leu Ala Arg Cys Trp Gly Thr Ala
            20                  25                  30

Cys Thr Asn Val Ser Cys Pro Gln Cys Arg
        35                  40
```

What is claimed is:

1. A method for screening germline of a human subject for an alteration of a BRCA1 gene which comprises comparing germline sequence of a BRCA1 gene or BRCA1 RNA from a tissue sample from said subject or a sequence of BRCA1 cDNA made from mRNA from said sample with germline sequences of wild-type BRCA1 gene, wild-type BRCA1 RNA or wild-type BRCA1 cDNA, wherein a difference in the sequence of the BRCA1 gene, BRCA1 RNA or BRCA1 cDNA of the subject from wild-type indicates an alteration in the BRCA1 gene in said subject.

2. The method of claim 1 wherein the wild-type BRCA1 gene has the sequence set forth in SEQ ID NO:1.

3. The method of claim 1 wherein the nucleic acid sequence of BRCA1 RNA from the subject is compared to nucleic acid sequences of wild-type BRCA1 gene, BRCA1 RNA orBRCA1 cDNA.

4. The method of claim 3 wherein the nucleic acid sequence is compared by hybridizing a BRCA1 gene probe which specifically hybridizes to a BRCA1 allele to RNA isolated from said subject and detecting of the presence of a hybridization product, wherein a presence of said product indicates the presence of said allele in the subject.

5. The method of claim 1 wherein a regulatory region of the BRCA1 gene from said subject is compared with a regulatory region of wild-type BRCA1 gene sequences, said regulatory region corresponding to nucleotides 1-1531 of SEQ ID NO:13.

6. The method of claim 1 wherein a germline nucleic acid sequence is compared by obtaining a first BRCA1 gene fragment from a BRCA1 gene from a human sample and a second BRCA1 gene fragment from a wild-1ype BRCA1 gene, said second fragment corresponding to said first fragment forming single-stranded DNA from said first BRCA1 gene fragment and from said second BRCA1 gene fragment, electrophoresing said single-stranded DNAs on a non-denaturing polyacrylamide gel, comparing the mobility of said single-stranded DNAs on said gel to determine if said single-stranded DNA from said first BRCA1 gene fragment is shifted relative to said second BRCA1 gene fragment and sequencing said single-stranded DNA from said first BRCA1 gene fragment having a shift in electrophoretic mobility.

7. The method of claim 1 wherein a germline nucleic acid sequence is compared by hybridizing a BRCA1 gene probe which specifically hybridizes to a BRCA1 allele to genomic DNA isolated from said sample and detecting the presence of a hybridization product wherein a presence of said product indicates the presence of said allele in the subject.

8. The method of claim 1 wherein a germline nucleic acid sequence is compared by amplifying all or part of a BRCA1 gene from said sample using a set of primers to produce amplified nucleic acids and sequencing the amplified nucleic acids.

9. The method of claim 1 wherein a germline nucleic acid sequence is compared by amplifying all or part of a BRCA1 gene using a primer specific for a specific BRCA1 mutant allele and detecting the presence of an amplified product, wherein the presence of said product indicates the presence of said specific allele.

10. The method of claim 1 wherein a germline nucleic acid sequence is compared by molecularly cloning all or part of a BRCA1 gene from said sample to produce a cloned nucleic acid and sequencing the cloned nucleic acid.

11. The method of claim 1 wherein a germline nucleic acid sequence is compared by obtaining a first BRCA1 gene fragment from (a) BRCA1 gene genomic DNA isolated from said sample, (b) BRCA1 RNA isolated from said sample or (c) BRCA1 CDNA made from mRNA isolated from said sample, obtaining a second BRCA1 gene fragment from (a) wild-type BRCA1 genomic DNA, (b) wild-type BRCA1 RNA or (c) wild-type cDNA made from wild-type mRNA, said second BRCA1 gene fragment corresponding to said first BRCA1 gene fragment, forming single-stranded DNA from said first BRCA1 gene fragment and from said second BRCA1 gene fragment, forming a heteroduplex consisting of single-stranded DNA from said BRCA1 gene fragment and single-stranded DNA from said second BRCA1 gene fragment, analyzing the heteroduplex to determine if said single-stranded DNA from said first BRCA1 gene fragment has a mismatch relative to said single-stranded DNA from said second BRCA1 gene fragment and sequencing said first single-stranded DNA from said first BRCA1 gene fragment having a mismatch.

12. The method of claim 1 wherein a germline nucleic acid sequence is compared by amplifying BRCA1 nucleic acids from said sample to produce amplified nucleic acids, hybridizing the amplified nucleic acids to a BRCA1 DNA probe specific for a BRCA1 allele and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said allele in the subject.

13. The method of claim 1 wherein a germline nucleic acid sequence is compared by analyzing BRCA1 nucleic acids in said sample for a deletion mutation.

14. The method of claim 1 wherein a germline nucleic acid sequence is compared by analyzing BRCA1 nucleic acids in said sample for a point mutation.

15. The method of claim 1 wherein a germline nucleic acid sequence is compared by analyzing BRCA1 nucleic acids in said sample for an insertion mutation.

16. The method of claim 1 wherein a germline nucleic acid sequence is compared by hybridizing the tissue sample in situ with a nucleic acid probe specific for a BRCA1 allele and detecting the presence of a hybridization product, wherein a presence of said product indicates the presence of said allele in the subject.

17. The method of claim 1 wherein a nucleic acid of a germline BRCA1 cDNA made from mRNA from said

sample is compared to nucleic acid sequences of wild-type BRCA1 gene, BRCA1 RNA or BRCA1 cDNA.

18. The method of claim **1** wherein a nucleic acid sequence of a germline BRCA1 gene from the subject is compared to nucleic acid sequences of wild-type BRCA1 gene, BRCA1 RNA or BRCA1 CDNA.

19. The method of claim **1** wherein said difference is selected from the group consisting of missense mutations within the zinc finger motif, deletions, insertions, frameshift mutations, nonsense mutations and splice site mutations.

20. A method for detecting a germline alteration in a BRCA1 gene, said alteration selected from the group consisting of the alterations set forth in Tables 11 and 12 which comprises analyzing a sequence of the BRCA1 gene or BRCA1 RNA from a human sample or analyzing the sequence of BRCA1 CDNA made from mRNA from said sample.

21. The method of claim **20** wherein a germline alteration is detected by hybridizing a BRCA1 gene probe which specifically hybridizes to an allele of one of said alterations to RNA isolated from said human sample and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said allele in the sample.

22. The method of claim **20** wherein a germline alteration is detected by hybridizing a BRCA1 gene probe which specifically hybridizes to one of said alterations to genomic DNA isolated from said sample and detecting of the presence of a hybridization product, wherein the presence of said product indicates the presence of said alteration in the sample.

23. The method of claim **20** wherein a germline alteration is detected by amplifying all or part of a BRCA1 gene in said sample using a set of primers to produce amplified nucleic acids and sequencing the amplified nucleic acids.

24. The method of claim **20** wherein a germline alteration is detected by amplifying part of a BRCA1 gene in said sample using a primer specific for an allele having one of said alterations and detecting the presence of an amplified product, wherein the presence of said product indicates the presence of said allele in the sample.

25. The method of claim **20** wherein a germline alteration is detected by molecularly cloning all or part of a BRCA1 gene in said sample to produce a cloned nucleic acid and sequencing the cloned nucleic acid.

26. The method of claim **20** wherein a germline alteration is detected by amplifying BRCA1 gene nucleic acids in said sample, hybridizing the amplified nucleic acids to a BRCA1 DNA probe specific for one of said alterations and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said alteration.

27. A method for screening for a germline alteration in a BRCA1 gene in a human subject which comprises analyzing a BRCA1 polypeptide from a tissue sample from said subject for an altered BRCA1 polypeptide by (i) detecting either a full length BRCA1 polypeptide or a truncated BRCA1 polypeptide or (ii) contacting an antibody which specifically binds to an epitope of an altered BRCA1 polypeptide to the BRCA1 polypeptide from said sample and detecting bound antibody, wherein the presence of a truncated protein or bound antibody indicates the presence of a germline alteration in the BRCA1 gene.

28. The method of claim **27** wherein a BRCA1 polypeptide is analyzed by detecting a truncated BRCA1 polypeptide.

29. The method of claim **27** wherein a BRCA1 polypeptide is analyzed by contacting an antibody which specifically binds to an epitope of an altered BRCA1 polypeptide to the BRCA1 polypeptide from said sample.

30. The method of claim **20** wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

31. The method of claim **21** wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

32. The method of claim **22** wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

33. The method of claim **23** wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

34. The method of claim **24** wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

35. The method of claim **25** wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

36. The method of claim **26** wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

37. A kit for screening for an alteration in a BRCA1 gene in a human subject which comprises at least one antibody (i) which specifically binds to wild-type BRCA1 polypeptide but not a truncated BRCA1 polypeptide or (ii) which specifically binds to an epitope of an altered BRCA1 polypeptide.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 5,753,441
DATED : May 19, 1998
INVENTOR(S) : Mark H. Skolnick et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Title page,
Item [22], "Jan. 5, 1996" should be -- June 7, 1995 --.

Signed and Sealed this

Thirtieth Day of July, 2002

Attest:

Attesting Officer

JAMES E. ROGAN
Director of the United States Patent and Trademark Office

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 5,753,441
APPLICATION NO. : 08/488011
DATED : May 19, 1998
INVENTOR(S) : Skolnick et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Col. 155, Claim 3, line 31, please change "orBRCA1" to --or BRCA1--.

Col. 155, Claim 6, line 45, please change "wild-1ype" to --wild-type--.

Col. 155, Claim 6, line 47, please change "fragment" to --fragment,--.

Col. 156, Claim 7, line 61, please change "product" to --product,--.

Col. 156, Claim 11, line 29, please change "CDNA" to --cDNA--.

Col. 156, Claim 11, line 37, please change "BRCA1" to --first BRCA1--.

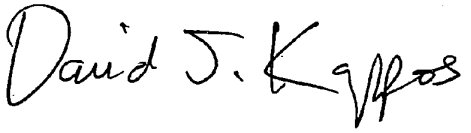
Col. 156, Claim 16, line 62, please change "in situ" to --*in situ*--.

Col. 157, Claim 18, line 6, please change "CDNA" to --cDNA--.

Col. 157, Claim 20, line 16, please change "CDNA" to --cDNA--.

Signed and Sealed this

First Day of December, 2009



David J. Kappos
Director of the United States Patent and Trademark Office

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number : **0 518 650 A2**

(12) **EUROPEAN PATENT APPLICATION**

(21) Application number : **92305333.4**

(22) Date of filing : **10.06.92**

(51) Int. Cl.[5] : **G01N 33/68,** G01N 33/53, G01N 33/569, C12Q 1/68, C07H 21/04, A61K 48/00, G01N 33/574

(30) Priority : **14.06.91 US 715182**
31.03.92 US 860758

(43) Date of publication of application :
16.12.92 Bulletin 92/51

(84) Designated Contracting States :
AT BE CH DE DK ES FR GB GR IT LI LU MC NL PT SE

(71) Applicant : **THE JOHNS HOPKINS UNIVERSITY**
720 Rutland Avenue
Baltimore, MD 21205 (US)

(71) Applicant : **PHARMAGENICS, INC.**
4 Pearl Court
Allendale, New Jersey 07401 (US)

(72) Inventor : **Vogelstein, Bert**
3700 Breton Way
Baltimore, Maryland 21208 (US)
Inventor : **Kinzler, Kenneth W.**
1348 Halstead Road
Baltimore, Maryland 21234 (US)
Inventor : **Sherman, Michael I.**
314 Forest Avenue
Glen Ridge, New Jersey 07028 (US)

(74) Representative : **Turner, Peter Charles et al**
Withers & Rogers 4 Dyer's Buildings, Holborn
London EC1N 2JT (GB)

(54) **Sequence specific DNA binding by P53.**

(57) Specific sequences in the human genome are the sites of strong binding of wild-type p53 protein, but not mutant forms of the protein. These sequences are used diagnostically to detect cells in which the These sequences are used diagnostically to detect cells in which the amount of wild-type p53 is diminished. The sequences can also be used to screen for agents which correct for loss of wild-type p53 to DNA in cancer cells.

EP 0 518 650 A2

GeneDX 1009, pg. 509

This work was partially supported by the U.S. government under NIH grants CA06973, CA09243, CA35494, CA09071 and CA43460. The U.S. government retains certain rights in this invention.

TECHNICAL FIELD OF THE INVENTION

This invention relates to diagnostic and therapeutic methods for detecting and alleviating cancers. Specifically it relates to the identification of DNA sequences specifically bound by tumor suppressor p53.

BACKGROUND OF THE INVENTION

The gene for the nuclear phosphoprotein p53 is the most commonly mutated gene yet identified in human cancers (Vogelstein, B., Nature, 348:681 (1990)). Missense mutations occur in tumors of the colon, lung, breast, ovary, bladder, and several other organs (S.J. Baker, et al., Science, 244:217 (1989); J.M. Nigro, et al., Nature, 342:705 (1989); T. Takahashi, et al., Science, 246:491 (1989); Romano, et al., Oncogene, 4:1483 (1989), Menon, Proc. Natl Acad. Sci. USA, 87:5435 (1990); Iggo, et al., Lancet ii, 675 (1990); T. Takahashi, et al., J. Clin. Invest. 86:363 (1990); Mulligan, Proc. Natl Acad. Sci. USA, 87:5863 (1990); Bartek, et al., Oncogene, 5:893 (1990); Stratton et al., Oncogene, 5:1297 (1990)). One of the important challenges of current cancer research is the elucidation of the biochemical properties of the p53 gene product and the way in which mutations of the p53 gene affect these properties.

Although some biological characteristics of p53 have been defined, such as its ability to suppress the growth of in vitro transformed murine cells (Eliyahu, et al., Proc. Natl Acad. Sci. USA 86:8763 (1989); Finlay, et al., Cell, 57:1083 (1989)) or human cancer cells (Baker, et al., Science, 249:912 (1990); Mercer, et al., Proc. Natl Acad. Sci, USA, 87:6166 (1990); Diller et al., Mol. Cell Biol. 10:5772 (1990)), the biochemical basis of this suppression remains largely unknown. As a step towards understanding such properties, we have attempted to determine whether p53 binds to specific DNA sequences within the human genome.

SUMMARY OF THE INVENTION

It is an object of the invention to provide a method for detecting the presence of wild-type p53 protein in a cell.

It is another object of the invention to provide a method for providing the physiological effect of wild-type p53 protein to a cell.

It is yet another object of the invention to provide a double-stranded DNA fragment which contains a p53-specific DNA binding site.

It is yet another object of the invention to provide a single-stranded oligonucleotide or oligonucleotide containing nucleotide analogs which can specifically complex to a p53-specific DNA binding site.

It is still another object of the invention to provide methods for identifying compounds which specifically bind to p53-specific DNA binding sequences.

It is still another object of the invention to provide methods for identifying compounds which restore the ability of mutant p53 proteins to bind to specific DNA binding sequences.

These and other object of the invention are provided by one or more of the embodiments described below. In one embodiment a method is provided for detecting the presence of wild-type p53 protein in a cell, comprising the steps of: contacting a p53-specific-binding DNA fragment with a cell lysate from a tissue of a human to bind the DNA fragment to wild-type p53 present in the cell lysate; and detecting the binding of the p53-specific-binding DNA fragment to wild-type p53.

In another embodiment of the invention a method is disclosed for providing the physiological effect of wild-type p53 protein to cell, comprising the steps of: providing to a cell a compound which is able to specifically complex with a p53-specific binding site.

In yet another embodiment a double-stranded DNA fragment is provided which comprises a p53-specific-DNA binding site, wherein the fragment comprises more than one monomer repeat of the sequence 5′-RRRCWWGYYY-3′ and wherein the fragment is covalently attached to an insoluble polymeric support.

In another embodiment of the invention a single-stranded oligonucleotide containing natural nucleotides and/or nucleotide analogs is provided which is able to complex specifically with a p53-specific binding site, said binding site comprising more than one monomer of the sequence 5′-RRRCWWGYYY-3′.

In yet another embodiment of the invention a method is provided for identifying compounds which specifically bind to p53-specific DNA binding sequences, comprising the steps of: contacting a p53-specific DNA binding fragment immobilized on a solid support with a test compound to bind the test compound to the DNA fragment; and determining the amount of test compound which is bound to the DNA fragment.

GeneDX 1009, pg. 510

In even another embodiment of the invention a method is provided for identifying compounds which specifically bind to p53-specific-DNA binding sequences, comprising the steps of: contacting a p53-binding DNA fragment immobilized on a solid support with both a test compound and wild-type p53 protein to bind the wild-type p53 protein to the DNA fragment; determining the amount of wild-type p53 protein which is bound to the DNA fragment, inhibition of binding of wild-type p53 protein by the test compound suggesting binding of the test compound to the p53-specific DNA binding sequences.

In still another embodiment a method of prescreening agents for use in cancer therapy is provided comprising: measuring the amount of binding of a p53 protein which is encoded by a mutant gene found in cancer cells of a patient to a DNA molecule which comprises more than one monomer of RRRCWWGYYY; measuring the amount of binding of said p53 protein to said DNA molecule in the presence of a test substance; and comparing the amount of binding of the p53 protein in the presence of said test substance to the amount of binding of the p53 protein in the absence of said test substance, a test substance which increases the amount of binding being a candidate for use in cancer therapy.

In another embodiment of the invention a method is provided for prescreening agents for use in cancer therapy comprising: contacting a transfected cell with a test substance, said transfected cell containing a p53 protein which is encoded by a mutant gene found in cancer cells of a patient and a reporter gene construct comprising a reporter gene which encodes an assayable product and a sequence which conforms to the p53 consensus binding site having more than one monomer of RRRCWWGYYY, wherein said sequence is upstream from and adjacent to said reporter gene; and determining whether the amount of expression of said reporter gene is altered by the test substance, a test substance which alters the amount of expression of said reporter gene being a candidate for use in cancer therapy.

In still another embodiment a method of prescreening agents for use in cancer therapy is provided comprising: adding RNA polymerase and ribonucleotides to a transcription construct, said transcription construct comprising a reporter gene which encodes an assayable product and a sequence which conforms to the p53 consensus binding site having more than one monomer of RRRCWWGYYY, said sequence being upstream from and adjacent to said reporter gene, said step of adding being effected in the presence and absence of a test substance; determining whether the amount of transcription of said reporter gene is altered by the presence of said test substance, a test substance which alters the amount of transcription of said reporter gene being a candidate for use in cancer therapy.

In a further embodiment a DNA construct is provided comprising: a reporter gene which encodes an assayable product; and a sequence which conforms to the p53 consensus binding site having more than one monomer of RRRCWWGYYY upstream from and adjacent to said reporter gene; wherein said DNA construct is selected from the group consisting of a recombinant plasmid, a viral vector or an isolated molecule of DNA.

In another embodiment of the invention a method is provided of diagnosing tumor-inducing or hyperplastia-inducing strains of human papilloma virus (HPV) comprising: contacting cells or cell extracts of patients suspected of being infected by HPV with a p53-specific binding DNA fragment; and detecting the amount of wild-type p53 in said cells or cell extract which binds to said DNA fragment, absence of bound p53 indicating infection by strains of HPV which sequester p53.

These and other embodiments of the invention provide the art with new tools for detecting and remedying the loss of function of the most commonly mutated gene identified in human cancers. Further it provides methods for identifying agents which can suppress neoplastic growth.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A. Screening for fragments bound by p53 using an immunoprecipitation assay. Panel 1 contains the hFosAva2 clone; panel 2, 772 C_{BE}; panel 3, Lambda 5R; panel 4, a pool of clones with inserts of randomly cloned human genomic sequences. 772 C_{BE} and Lambda 5R contain HinfI fragments (259 and 190 bp, respectively) which bound p53 relatively strongly (arrowheads). "C" - control lane, containing 2% of the labelled DNA used in the binding reactions. "B" - bound DNA recovered from the immunoprecipitate. Figure 1B. Tests for dependence on p53 and specific antibody. Cell lysates were produced by infection with vaccinia virus that did (+) or did not (-) contain an insert of wild-type p53 cDNA. Immunoprecipitation was performed with anti-p53 monoclonal antibodies (+) or normal mouse IgG (-).

Figure 2. Relative abilities of wild-type and mutant p53 to precipitate fragment A. "C" - control lanes, containing 2% of the labelled DNA used in the binding reaction, "B" - bound DNA recovered from the immunoprecipitate. Figure 2A. Increasing quantities of wild-type and mutant 273^{his} p53, affinity-purified from a baculovirus expression system, were used to precipitate labelled C_{BE} fragments. Figure 2B. Lysates from a vaccinia virus system (Vac) producing the wild-type (wt), mutant (175^{his}), or no p53 protein (-), were used to immunoprecipitate labelled C_{BE} fragments. Equivalent quantities of p53 were present in the wild-type and mutant p53 lysates, as

GeneDX 1009, pg. 511

assessed by Western blot. In the "Bac" lane, affinity-purified p53 produced in baculovirus-infected insect cells was used in place of the vaccinia-infected lysates.

Figure 3A. Fragment A and adjacent vector sequences in the 10d subclone of 772 C_{BE}. This differs from the published 772 C_{BE} sequence (Genbank M25718) in the number of CTT repeats (bp 173-229) and in the presence of A instead of C at bp 116. Figure 3B. Fragment B and adjacent vector sequences in the 8a subclone of Lambda 5R. A related sequence on file (Genbank X05913) varied somewhat from the Lambda 5R subclone studied.

Figure 4. Binding of various subfragments of fragments A and B to p53 from vaccinia-infected cell lysates. Figure 4A. Subfragments of fragment A (subclone 10d) were assayed by immunoprecipitation for their ability to bind wild-type p53 from vaccinia-infected cell lysates. Binding of at least 2% of the DNA added to the reaction was judged as a positive (+) result; lesser but significant binding was recorded as "+/-". Double Lines (=) denote fragment A sequences. Single lines (-) denote polylinker sequences of the vector, not originally present in fragment A (Fig. 1). Fragment 5mut1 had a G to T transversion at bp 120; 5mut2 had G to T transversions at bp 120 to 122. Figure 4B. The fragment A (panels 1-4) and fragment B (panel 5) subfragments illustrated in Figure 4A are labelled to the left of the bands. The "v" band in panel 4 corresponds to the 2.9 kb vector into which subfragments 6 was cloned. Subfragment 8 (panel 5) contained bp 104-238 of fragment B (see Fig. 3B). Control lanes (C) contained 2% of the labelled fragments used in the binding assays (B).

Figure 5. Effects of methylation and point mutations in fragment A on DNA-binding. Figure 5A. Methylation interference assay. "B" - bound DNA recovered from the immunoprecipitate. "C" - equivalent amount of control DNA fragment, not subject to binding reaction. Bound and control DNA samples were cleaved at methylated guanines and equal amount separated by electrophoresis on a 6% denaturing gel. Dots represent methylation-sensitive sites (open for partial, solid for strong interference); some variation in band intensities occurred between assays, and only the reproducible changes are marked. Figure 5B. Binding of "mutant" subfragments of fragment A (5mut1 and 5mut2) to purified baculovirus-produced p53 is compared to that of the normal subfragment 5 sequence. 5mut1 contains a T instead of G at bp 120, and 5mut2 contains Ts in place of Gs at bp 120, 121, and 122.

Figure 6A shows RNase protection analysis of transfected clonal lines. A labeled antisense p53 probe was hybridized with total cellular RNA, and digested with RNase A. Endogenous RNA included all sequences represented in the labeled probe. Exogenous p53 RNA produced from the expression vectors extended only about 2/3 of the length of the probe.

Figure 6B shows Southern blot analysis of transfected clonal lines. The exogenous p53 gene was present on a 1.8 kb BamHI fragment. The endogenous p53 gene gave rise to a 7.8 kb BamHI fragment. Other sized fragments presumably arose by rearrangements.

Figure 7A shows expression analysis of pooled clones; the analysis was as described in Figure 6A.

Figure 7B shows a Southern blot analysis of SW480 pooled clones.

Figure 8. Isolation of human genomic sequences which bound to p53.

Figure 8A. Experimental strategy used for isolation and analysis of human genomic DNA fragments which bound to p53.

Figure 8B. Immunoprecipitation (IP) assays of cloned fragments. Clones of amplified and selected (AS) DNA were tested for the presence of p53-binding fragments by IP. For each clone, the bound DNA is shown in the B lane, adjacent to a control (c) lane containing 2% of the total end-labeled DNA used in the binding assay. In this representative experiment, eight binding fragments were identified, representing six unique genomic fragments. The inserts from the clones in lanes labeled, 2, 3, 5, 9, 10, and 11 contained p53-binding fragment, while the other lanes contained none. The clones in lanes 2 and 5 each contained two binding fragments.

Figure 9. Mapping of p53-binding sites by DNase I protection (DP) and methylation interference (MI). For each footprint, the first and fourth lanes contain control samples of the total labeled DNA, whereas the middle two lanes contain an equivalent amount of p53-bound DNA. DNA sequences corresponding to the p53-binding regions on strand 1 are shown in Fig. 10.

Figure 10. Definition of a consensus binding site for p53. The p53 binding site of 18 cloned human genomic DNA fragments, determined by footprinting methods are displayed along the central axis of symmetry which separates the two 10 bp consensus monomers. Nucleotides in capital letters represent identity of a genomic sequence to the consensus, whereas lower case letters identify disparity with the consensus. Sequences surrounding the consensus or separating the two 10 bp monomers are also shown in lower case. The ten synthetic oligonucleotides investigated for the ability to be bound by p53 are shown at the bottom. Oligonucleotides No. 6 to 10 were tested after cloning into plasmid vectors. Lower case letters represent vector-derived sequences. Combined nucleotide usage (%) within the two monomers of the consensus binding site is shown in the middle.

Figure 11. Binding of synthetic oligonucleotides to wild-type (wt) and mutant p53 proteins.

Figure 11A. The 10 bp consensus monomer was insufficient for binding, whereas dimers in various orien-

GeneDX 1009, pg. 512

tations or multimers of the 10 bp consensus bound strongly to p53. For each sample, the control lane (C) contained 2% of the total DNA used in the binding reaction, and was composed of two fragments: 2.9 kb vector DNA fragment, and a fragment of 40 to 80 bp containing no insert (lane 9; XhoI plus Pst1 digest of pBluescript II SK+), the 10 bp consensus monomer sequence 5′-AGGCATGTCT-3′(lane 5), or multimers of this sequence arranged as indicated (lanes 1 to 4, 6, to 8). Bound DNA from the IP is shown in the B lanes.

Figure 11B. Comparison of the ability of wild-type and mutant p53 to bind to the consensus dimer. In vitro translated p53 proteins were tested for the ability to bind the consensus dimer by IP. Two percent of the total DNA used for binding is shown in lane 1. Lane 7 shows binding to baculovirus-produced human wild-type p53 protein. Lanes 2 to 6 show binding of in vitro translated wild-type and mutant p53 proteins. The mutant p53 proteins contained changes at codon 143 (val to ala), 175 (arg to his), 248 (arg to trp), and 273 (arg to his).

Figure 12. Reporter and expression construct used in transfections are shown.

Figure 12A. Reporter constructs in pBluescript II SK+ vector (Stratagene). PG_n, concatemers of n copies of the p53 binding sequence PG. MG_n, concatemers of the mutated sequence, which does not bind p53. CAT, the chloramphenicol acetyltransferase coding sequence. LacZ, the β-galactosidase coding sequence. Py, the early gene promoter from polyomavirus. CYC, the yeast cytochrome c gene promoter.

Figure 12B. Expression vectors. CMV, the cytomegalovirus promoter from the parent vector pCMVneoBam. Gal, the galactose-inducible promoter from yeast.

Figure 13. Correlation of DNA-binding and transactivation is shown.

Figure 13A. Relative DNA-binding abilities of various length concatemers of a p53-binding sequence (PG_n series), using an immunoprecipitation assay. Clones were cleaved by restriction endonucleases to extricate the concatemers, end-labelled, incubated with purified baculovirus-produced wild-type human p53, immunoprecipitated with anti-p53 and protein A-Sepharose, and bound fragments recovered and separated on a non-denaturing polyacrylamide gel. C, control lane, containing 2% of the labeled DNA used in the binding reactions. B, bound DNA recovered from the binding reactions.

Figure 13B. Transactivation efficiencies of reporters containing the various PG_n concatemers compared by chloramphenicol acetyltransferase (CAT) assay. 1.7 μg of the expression vector p53-wt were transfected into HCT 116 cells. Reporters had one orientation of the PG_n sequence (...TGCCT...Py...CAT...), except for PG_2-CAT and PG_{13}-CAT, which had the opposite orientation (...AGGCA...Py...CAT...). Results are expressed relative to the CAT activity in lane 7, which was arbitrarily set at 100.

Figure 14. Comparison of the ability of wild-type and mutant p53 to bind to a concatemerized binding sequence (PG_{16}), and lack of DNA-binding ability to MG_{15} are shown. Each p53 form (as a rabbit reticulocyte preparation (Promega) made essentially as described in Kern, et al., Oncogene, 6:131-136 (1991); equalized for p53 quantity by immunoblot analysis, data not shown) was used to immunoprecipitate end-labeled DNA. C, control lanes, containing 2% of the labeled DNA used in the binding reaction. B, bound DNA recovered from the immunoprecipitate.

Figure 15. Effects of varying the location of the binding sequence relative to the promoter are shown.

Figure 15A. Efficiencies of reporters with PG_{13} separated by varying distances from the polyomavirus promoter. Various lengths of the non-binding sequence (MG_n series) were used to provide inert spacers of defined length (see Fig. 8A). CAT assays are shown.

Figure 15B. Transactivation efficiency of a reporter with PG_{16} downstream of CAT (CAT-PG_{16}). 1.7 μg each of expression and reporter construct were transfected. The MG_{15}-CAT served as a negative control. CAT assays are shown.

Figure 16. Relative abilities of wild-type and mutant p53 to activate transcription are shown.

Figure 16A. Representative CAT assays (Exp. 2 from Table 1).

Figure 16B. Western blot analysis of p53 expression in lysates of transfected cells, showing levels of expression at least as high as that of wild-type p53 were obtained with the mutant p53 clones. In one case (143^{ala}), the level of expression of p53 (lane 3) was slightly lower than with the other constructs, for unknown reasons. When 2.55 μg of vector was used in the transfection, however, the level produced by the p53-143 vector (lane 7) was at least as high as that for the p53-wt vector, but no transactivation was observed (Table I).

Figure 17. Effect of co-expressed 175^{his} mutant p53 on transactivation by wild-type p53 is shown. 1.7 μg of PG_{13}-CAT reporter was used with the stated amount of p53 expression constructs. The balance of transfected plasmid was pCMVneoBam (for expressor) and pBluescript II SK+ (for reporter) to a total of 5.1 μg. CAT assays are shown.

Figure 18. Dominant-negative effect of various mutant p53 proteins are demonstrated. 0.85 μg of p53-wt was used in all transfections, without or with the addition of 0.85 μg or 2.55 μg mutant p53 construct, or with an additional 2.55 μg p53-wt. 1.7 μg PG_{13}-CAT reporter was used in each case. The composite results shown are representative of at least two transfections done on separate days.

Figure 19. A biochemical model for the effect of p53 mutations is depicted. Cellular p53 exists as oligomers.

GeneDX 1009, pg. 513

Oligomeric p53 binds its recognition sequence of DNA (shaded thick line) and activates the transcription of adjacent genes (filled box). The depiction of p53 dimers is for illustrative purposes only: tetramers and other forms likely occur. Hetero-oligomerization with mutant proteins inactivates the transactivation function of the participating wild-type p53 molecules, producing the dominant-negative effect of missense p53 mutations. The remaining activity provided by the residual wild-type homo-oligomers is lost upon deletion of the remaining wild-type allele, a frequent event in the progression of human neoplasms.

DETAILED DESCRIPTION OF THE INVENTION

It is a finding of this invention that wild-type p53 protein binds specific fragment of human chromosomal DNA. Each of the fragment contains more than one monomer of the double-stranded motif 5′-RRRCWWGYYY-3′ separated by 0 to 13 bp. Some of these sequences are found near origins of replication of certain animal viruses and animal cells. See Jelinek et al, Proc. Natl. Acad. Sci. USA, vol. 77, pp. 1398-1402 (1980). Four mutant forms of p53 protein which are commonly found in human tumors do not have the ability to bind to these sequences. Thus, a function of p53 may be mediated by its ability to bind to specific DNA sequences in the human genome.

Wild-type p53 protein binds specifically to certain DNA sequences. Two previously isolated DNA fragments, called herein fragment A and fragment B, have been identified which contain residues which appear responsible for the binding. These residues are located within DNA segments which contain nucleotides 103-134 of SEQ ID NO:1 and nucleotides 104-198 of SEQ ID NO:2.

Fragment A contains sequences near a putative replication origin of the ribosomal gene cluster, while fragment B contains sequences that may allow adjacent sequences to replicate as extrachromosomal circles in HeLa cells (Sylvester, et al., Gene 84, 193, 1989). The TGCCT repeat present in the DNA binding region of both of these fragments has been observed in other potential replication origins. Thus p53 may be involved in the regulation of initiation of DNA synthesis by virtue of its binding.

It has been found that p53 will specifically bind to other sequences in the human genome with similar sequence motifs. Using a strategy coupling immunoprecipitation to "whole-genome PCR" (Kinzler, et al., Nucleic Acids Research, 17:3645-3653 (1989)), nineteen human DNA fragments that bind to p53 have been identified. Each of the fragments contain a sequence which conforms to a dimer of the double-stranded motif 5′-RRRCWWGYYY-3′, separated by 0 to 13 bp. These dimers directly mediate binding, as assessed by DNase I protection and methylation interference assays. The consensus dimers contain a striking symmetry, with four 5′-RRRCW-3′units oriented in alternating directions. A synthetic monomer containing the 10 bp consensus sequence is insufficient for binding, while the combination of two or more monomers bind strongly to wt p53, but negligibly to p53 mutants. Thus, more than one monomer appears to be required for binding. The spacing between monomers may be from 0 to 40 nucleotides, although all natural binding sites isolated have spacings of less than 15 nucleotides. The symmetry of the four half-sites within the consensus dimers suggests that p53 interacts with DNA as a tetrameric protein. The eighteen unique clones shown in Figure 10 allow the identification of adjacent genes which may be regulated by p53 and may mediate its growth-suppressive action.

A sequence which conforms to the consensus sequence need not have the exact sequence of the dimer. As shown in Figure 10, p53-binding sequences can have nucleotides which are different than those designated. There can also be extra nucleotides inserted within the dimer. Nucleotides can also be missing from the dimer. Typically p53-binding sequences will vary from the consensus sequence at no more than 5 nucleotide positions.

It is an additional finding of the present invention that wild-type p53 can activate the expression of genes adjacent to a specific binding site. Moreover, the level of in vivo transactivation is proportional to the in vitro strength of DNA-binding. Mutant p53 encoded by oncogenic p53 genes (i.e., those found in cancer cells of human patients), completely lose the ability to transactivate. In addition the mutant p53 proteins exert a dominant-negative effect, dramatically reducing the transactivating activity of wild-type p53.

Based on the sequence information of the p53 specific-DNA-binding fragment, a number of diagnostic and therapeutic methods have been devised. According to one such method, cell lysates are tested for the presence or absence of wild-type p53 by virtue of its specific DNA binding ability. As it is known for various cancers and stages of cancers that one or both of the p53 alleles in tumor tissues can be mutant, testing for the presence or absence of wild-type p53 protein can provide diagnostic and prognostic information regarding a tumor and the patient. The cells to be tested are typically isolated from a tissue suspected of being neoplastic. Preferably the tissues are carefully prepared and isolated so that non-neoplastic tissues are not mixed with the neoplastic tissues, which can confound the analysis. Means for separating neoplastic tissues from non-neoplastic tissues are known in the art and include dissection of paraffin or cryostat sections, as well as use of flow cytometry. A cell lysate can be prepared from the tumor tissue according to any method known in the art. The cell lysate is then incubated with DNA fragments which are known to bind the wild-type p53 protein, under conditions which

GeneDX 1009, pg. 514

are conducive to such DNA/protein interactions. Alternatively, a histological sample can be analyzed by incubation with DNA fragments, as described for cell lysates.

It is known that p53 also binds non-specifically to DNA. Specific binding can be distinguished from non-specific binding by any means known in the art. For example, specific binding interactions are stronger than non-specific binding interactions. Thus the incubation mixture can be subjected to any agent or condition which destabilizes protein/DNA interactions such that the specific binding reaction is the predominant one detected. Alternatively, as taught more specifically below, a non-specific competitor, such as dI-dC, can be added to the incubation mixture. If the DNA containing the specific binding sites is labelled and the competitor is unlabeled, then the specific binding reactions will be the ones predominantly detected upon measuring labelled DNA.

According to one embodiment of the invention the DNA which is bound to p53 is separated from unbound DNA by immunoprecipitation with antibodies which are specific for p53. Use of two different monoclonal anti-p53 antibodies may result in more complete immunoprecipitation than either one alone. Unbound DNA remains in suspension. The amount of DNA which is in the immunoprecipitate can be quantitated by any means known in the art. According to one aspect of the invention, the DNA fragment is labelled with a detectable moiety, such as a radioactive moiety, a colorimetric moiety or a fluorescent moiety. Techniques for so labelling DNA are well known in the art. According to other embodiments of the invention, p53 which binds to the specific DNA sequence of the present invention can be detected by gel shift assays (Tan, Cell, 62:367, 1990), nuclease protection assays (see Example 9, infra), or methylase interference assays (see Example 9, infra).

According to another embodiment of the invention, after incubation of p53 with specific binding DNA fragments all components of the cell lysate which do not bind to the DNA fragments are removed. This can be accomplished, among other ways, by employing DNA fragment which are attached to an insoluble polymeric support such as agarose, cellulose and the like. After binding, all non-binding components can be washed away, leaving p53 bound to the DNA/solid support. The p53 can be quantitated by any means known in the art. It can be determined using an immunological assay, such as an ELISA, RIA or Western blotting.

The diagnostic assay of the present invention has applicability not only with regard to cancers which are known to involve mutation of p53, but also with regard to human viruses such as human papilloma virus (HPV). HPV protein E6 binds tightly to wild-type but not mutant p53. See Werness et al., Science, 248, 76-69 (1990). This tight binding is likely to block the interaction of p53 with it specific DNA binding sequences. By testing cells or cell extracts suspected of being infected with potentially tumor-inducing or hyperplastia-inducing strains of HPV or possibly other viruses, infected cells can be identified, because the E6 protein of the infected cells will have sequestered the wild-type p53, rendering it unable to bind to its specific DNA binding sequences. Such assays may be performed on cell extracts or on histological specimens.

According to the present invention a method is also provided of supplying wild-type p53 function to a cell which carries mutant p53 alleles. The wild-type p53 gene or a part of the gene may be introduced into the cell in a vector such that the gene remains extrachromosomal. In such a situation the gene will be expressed by the cell from the extrachromosomal location. If the mutant p53 genes present in the cell are expressed, then the wild-type p53 gene or gene portion should be expressed to a higher level than that of the mutant gene. This is because the mutant forms of the protein are thought to oligomerize with wild-type forms of the protein. (Eliyahu et al., Oncogene, vol. 3, p. 313, 1988.) If a gene portion is introduced and expressed in a cell carrying a mutant p53 allele, the gene portion should encode a part of the p53 protein which is required for non-neoplastic growth of the cell. More preferred is the situation where the wild-type p53 gene or a part of it is introduced into the mutant cell in such a way that it recombines with the endogenous mutant p53 gene present in the cell. Such recombination would require a double recombination event which would result in the correction of the p53 gene mutation. Vectors for introduction of genes both for recombination and for extrachromosomal maintenance are known in the art and any suitable vector may be used.

Polypeptides or other molecules which have p53 activity may be supplied to cells which carry mutant p53 alleles. The active molecules can be introduced into the cell by microinjection or by liposomes, for example. Alternatively, some such active molecules may be taken up by the cells, actively or by diffusion. Supply of such active molecules will effect a less aggressive state of neoplasia.

According to the present invention compounds which have p53 activity are those which specifically complex with a p53-specific DNA binding site. Wild-type p53 is one such compound, but portions of p53 which retain the ability to bind to p53-specific binding sites may also be used. Oligonucleotides and oligonucleotide containing nucleotide analogs are also contemplated among those compounds which are able to complex with a p53-specific DNA binding site. Although applicants do not wish to be bound by any particular theory, it is believed that oligonucleotides bind double-stranded DNA to form triplexes. Such triplexes have been shown to block transcription of certain genes, as well as protect the DNA binding sites from the action of enzymes such as DNA methylases. Although originally such oligonucleotides were thought to require only or predominantly pyrimidines (cytosine and thymine), purines have also successfully been incorporated into triplex forming oli-

GeneDX 1009, pg. 515

gonucleotides. Particular oligonucleotides which may be used include: nucleotides 140-162 of SEQ ID NO:2, nucleotides 128-158 of SEQ ID NO: 1, nucleotides 114-123 of SEQ IQ NO: 1, or portions thereof having at least ten nucleotides.

Oligonucleotides containing nucleotide analogs or otherwise modified oligonucleotides may also be useful in the formation of complexes with double-stranded DNA. For example, certain modifications to the 3′-terminus can be employed to reduce the susceptibility of the oligonucleotide to nuclease degradation. Moieties which may be appended to the 3′- or 5′-termini include a substituted or unsubstituted amino group, polyethylene glycol, polylysine, acridine, dodecanol, and cholesterol. Oligonucleotides and oligonucleotides containing nucleotide analogs which may be used include methylphosphonates, aminomethylphosphonates, phosphorothioates, phosphorodithioates, substituted or unsubstituted phosphoramidates, oligoribonucleotides, oligodeoxyribonucleotides, alpha-oligonucleotides and mixtures thereof. Other modifications to oligonucleotides may be desirable to increase the uptake by cells or nuclei of the oligonucleotides or to decrease nuclease sensitivity. All such modifications are within the contemplation of the invention.

Switchback linkers may also be incorporated into the midst of an oligonucleotide or analog. Such linkers are taught by Riordan and Martin (Nature, 350, 452, 1991). They are designed by molecular modeling to provide the proper spacing between portions of an oligonucleotide which are to interact with different strands of a double-stranded DNA molecule. Examples of oligonucleotides having such linkers include the following: TTGCCTTGCCT-switchback linker-CCT-switchback linker-CTTGCCT (corresponding to nucleotides 105-125 of the double-stranded sequence represented by SEQ ID NO:1) or portions thereof.

Single-tranded, linear or circular oligonucleotides containing nucleotide analogs which are able to complex specifically with all or part of a p53-specific binding site as described above are also contemplated as part of the invention. (See Kool, J. Am. Chem. Soc., vol. 113:625-626, 1991, with regard to circular oligonucleotides.) Such oligonucleotides containing nucleotide analogs should comprise at least about ten nucleotides in length in order to have the requisite specificity with respect to the entire human genome. Oligonucleotides or oligonucleotides containing nucleotide analogs which are able to complex specifically with part of a p53 binding site as well as with adjacent sequences are also contemplated as part of the invention. The oligonucleotides or oligonucleotides containing nucleotide analogs will preferably bind to the identified p53 binding regions. However, other binding regions in the human genome may well be found which are also suitable target for the oligonucleotides or analogs of the present invention. These other binding sites may well be the primary targets of p53; complexation of these sites may inhibit the unregulated growth which characterizes neoplastic cells.

Double-stranded DNA fragments which comprise a p53-specific DNA binding site and are attached to an insoluble polymeric support are also contemplated by this invention. The support may be agarose, cellulose, polycarbonate, polystyrene and the like. Such supported fragments may be used in screens to identify compounds which bind to p53-specific DNA binding sites. Similarly, such supported fragments may be used to perform diagnostic tests on cell lysates from suspected tumor tissues. They may also be used in assays used to screen potential chemotherapeutic agents, as discussed infra.

Although any method can be employed which utilizes the p53-specific DNA binding sites of the present invention, two particular methods are disclosed for screening for additional compounds that bind to p53-specific DNA binding sites. According to one method a test compound is incubated with a supported DNA fragment, as described above. The amount of test compound which binds to the supported DNA fragment is determined. This determination can be performed according to any means which is convenient. For example, the amount of a compound which can be removed after incubation with the supported fragment can be compared to the amount originally applied. Alternatively, the test compound can be labelled and the amount which binds to the supported fragment can be assayed directly. In order to render this screening method more specific, soluble DNA fragments which do not contain the p53 DNA binding sequence can be added to the incubation mixture. The soluble fragment would not have the ability to specifically bind to p53 wild-type protein.

According to another screening method for compounds to simulate the specific DNA binding activity of p53, test compounds are incubated with supported DNA fragments as described above. However, in this method wild-type p53 protein is also added to the incubation mixture. The amount of p53 protein which binds to the DNA fragment is measured using methods as described above. The amount of p53 protein bound is compared to the amount which binds in the absence of the test compound. Any diminution of p53 binding which result from the presence of the test compound is presumptively due to the competition of the test compound with p53 for the specific DNA binding sites of the supported fragments. Direct binding of the test compound to the binding site fragments can be confirmed using the assay described above.

The presence or absence of p53 genes or gene products can also be detected in body samples, such as, serum, stool, or other body fluids, such as urine and sputum. The same techniques discussed above for detection of the presence or absence of gene products in tissues can be applied to other body samples. By screening such body samples, a simple early diagnosis can be achieved for many types of cancers. In addition, the

GeneDX 1009, pg. 516

progress of chemotherapy or radiotherapy can be monitored more easily by testing such body samples for the presence or absence p53 genes or gene products.

The methods of the present invention for diagnosis and therapy of neoplastic tissue is applicable across a broad range of tumors. These include lung, breast, brain, colorectal, bladder, mesenchyme, prostate, liver as well as stomach tumors. In addition the method may be used in leukemias and osteosarcomas. It thus appears that the p53 gene has a role in the development of a broad range of tumors. The methods of diagnosis and therapy of the present invention are applicable to any tumor in which p53 has a role in tumorigenesis. The diagnostic method of the present invention is useful for clinicians so that they can decide upon an appropriate course of treatment. For example, a tumor displaying loss of both p53 alleles suggests a more aggressive therapeutic regimen than a tumor displaying loss of only one p53 allele.

Compounds which have p53-specific DNA-binding activity, including wild-type p53 protein, polypeptides corresponding to portions of wild-type p53 protein, oligonucleotides and oligonucleotide containing nucleotide analogues, as well as other organic molecules can also be administered to humans and animals as a pharmaceutical and therapeutic composition. Effective amounts will be administered to cause neoplastic cells to become less aggressively neoplastic or even to stop the growth of the neoplastic cells entirely. Generally, such amounts will be in the range of 10 ng to 10 μg per dose per person or other animal. The therapeutic compounds can be prepared in any conventional pharmaceutical excipient, such as physiological saline or other physiologically compatible aqueous buffer. Typically, the compounds will be administered by injection, either intravenous or intramuscular. However, other administration methods as are known in the art and may be used to administer the compounds of the present invention.

As a result of the discoveries of the present invention, screening methods can be devised to isolate chemical agents which may have use in cancer therapy. Specifically, agents can be screened for the ability to affect the structure of mutant p53 molecules so that their ability to bind and/or transactivate at specific-DNA-binding sites is restored. The necessary components for such a screening method are provided by this invention and include DNA molecules which contain more than one monomer of the sequence RRRCWWGYYY, and mutant p53 proteins which are found in tumors.1/

One such prescreening method is a binding assay in which the amount of binding of a p53 mutant protein to a DNA molecule which comprises the consensus binding site (or a conforming sequence) is measured. The amount of binding is also measured for a p53 mutant protein in the presence of a test substance. If the test substance increases the amount of p53 binding, then the test substance is a candidate for use in anti-tumor therapy. Further testing will be desirable before use in humans is attempted.

Methods for measuring the amount of binding can be any which are known in the art. See, e.g., Tan, et al., Cell, 62:367-377 (1990). One particular method employs immunoprecipitation. Briefly, purified p53 or a lysate of a cell expressing p53 is incubated with radiolabeled DNA and anti-p53 antibodies under conditions where proteins bind to DNA. Protein A-Sepharose and poly-dIdC-poly-dIdC are then added for an additional incubation. A pellet is formed and washed and the proteins are removed by digestion with a protease, and DNA is obtained by phenol extraction. The extracted DNA is then analyzed by electrophoresis and quantified. Quantitation of the DNA can be by autoradiography, for example. The amount of DNA immunoprecipitated is proportional to the amount of binding of the p53 protein to the DNA.

According to another method, the ability of a mutant p53 protein from a tumor to transactivate transcription in vitro is assessed with and without a test substance. If the test substance increases the amount of transcription activated by the p53, then the test substance is a candidate for use in anti-tumor therapy. Transcriptional activation is measured using a transcription construct which comprises a reporter gene encoding a convenient assayable enzyme activity, such as chloramphenicol acetyltransferase or β-galactosidase, and an upstream p53 consensus binding site (or a conforming sequence). The binding site must be upstream, although the distance from the start of transcription is not critical. The binding site, which is adjacent to the reporter gene, may be from 0 to 1 kb upstream. In vitro transcription assay systems are well known in the art. See, e.g., Lue, Science, 246, 661-664 (1989).

According to still another method, transcriptional activation can be measured in a cell containing a mutant p53 protein which has been transfected with a reporter gene construct comprising a gene encoding an assayable product, such as an enzyme or antigen, with a p53 consensus binding site (or conforming sequence) adjacent and upstream therefrom. The transfected cells are treated with a test substance. If the amount of transactivation caused by the mutant p53 is enhanced by the test substance, then the substance is a candidate for anti-tumor therapy presumably due to its interaction with the mutant protein. A particular transactivation assay

1/ Not all mutations in p53 destroy specific-DNA-binding ability. For example, mutations in phosphorylation sites of p53 have been made and tested; they retain binding activity. Such mutations have never been found in tumors. Mutations in p53 which are found in tumors are termed oncogenic herein.

GeneDX 1009, pg. 517

(transient expression assay) is described in Example 12, infra. Particular reporter constructs are shown in Figure 12A. Others may be used within the scope of the invention. Alternatively, if the transfected cell contains wild-type p53 diminished activation would indicate that the test substance competes with p53 for binding sites.

In another embodiment of the invention, oligonucleotides can be isolated which restore to mutant p53 proteins the ability to bind to the consensus binding sequence or conforming sequences. Mutant p53 protein and random oligonucleotides are added to a solid support on which p53-specific-binding DNA fragments are immobilized. Oligonucleotides which bind to the solid support are recovered and analyzed. Those whose binding to the solid support is dependent on the presence of the mutant p53 protein are presumptively binding the support by binding to and restoring the conformation of the mutant protein.

Transient expression constructs are conveniently made on plasmids and viral vectors, so that they can be propagated. These can also be used in vitro for transcription assays in the presence of RNA polymerase, ribonucleotides, and other cofactors.

EXAMPLES

Example 1

This example demonstrates the screening methods used to identify a p53 sequence-specific binding site.

In an attempt to identify a possible sequence-specific binding site, numerous cloned DNA sequences were screened using an immunoprecipitation technique. The immunoprecipitation assay for DNA-binding was modified from R.D.G. McKay, J. Mol. Biol. 145. 471 (1981). Binding reactions included 5 μl of vaccinia-infected cell lysate or purified baculovirus p53 preparations, 95 μl binding buffer (20 mM Tris pH 7.2, 100 mM NaCl, 10% glycerol, 1% NP-40, 5 mM EDTA), 6-20 x 10^4 dpm ^{32}P-labelled DNA, 4 μl (0.4 μg) pAb421 and 4 μl (0.4 μg) pAb1801 anti-p53 purified monoclonal antibodies (Oncogene Science), at 4°C for 30 min. 1.5 mg protein A-Sepharose (Sigma) and 12.5 μg poly-dIdC-poly-dIdC (Pharmacia), in 25 μl binding buffer, were added and the reactions rotated end-over-end at 4°C for 30 min. The pellet was washed twice with binding buffer and the proteins digested with SDS-proteinase K, then extracted with phenol and chloroform. The DNA was ethanol-precipitated and dissolved in electrophoresis sample buffer. The fragments were separated on a Tris-borate non-denaturing polyacrylamide (7, 10, or 12%) gel. The gel was fixed and dried for autoradiography.

Recombinant vaccinia viruses expressing p53 from the 40K promoter were isolated and purified as described in J. Lyons et al., Infec. and Immun. 58, 4089 (1990).

Vaccinia-infected cell lysates were prepared as follows: the indicated cells (about 2.5 x 10^7) were infected (MOI 2), harvested at 24 hrs and lysed in 2 ml lysis buffer (PBS containing 5mM EDTA, 0.5% NP-40, 0.5 mM PMSF, 10 μg/ml TPCK, 1 ug/ml aprotinin, 10 μg/ml trypsin inhibitor, 1 μg/ml leupeptin). Cell debris was pelleted at 16,000xg and the supernatant UV-irradiated (0.5 joules). The supernatant was frozen in aliquots at -80°C and found to be stable for at least 6 months.

Baculovirus-produced p53 was purified as described in P.N. Friedman, et al., Proc. Natl. Acad. Sci. USA, 87, 9275 (1990); while the presence of an accessory cofactor in the purified preparation cannot be excluded, SDS-polyacrylamide gel electrophoresis of the preparation revealed only a single polypeptide migrating at 53 kD upon silver-staining.

Two classes of clones were tested. The first class consisted of 400 clones containing inserts of 300 to 1000 base-pairs obtained randomly from the human genome. The second class consisted of cosmid and plasmid clones chosen because they contained sequences which might be important in normal growth control. Random human genomic clones were prepared from a partial MboI digest of human DNA; 300-1000 bp fragments were purified and ligated to EcoRI linkers, then cloned into the EcoRI site of pBluescript II (Stratagene). Clones selected because they might be important in normal growth included SP65hFosAva2, containing regulatory sequences from the fos gene, obtained from T Curran; 772 C_{BE} (J.E. Sylvester, R. Petersen, R.D. Schmeckel, Gene, 84, 193 (1989)), from J. Sylvester; a 4.4 kb BglII subclone of Lambda 5R (R. Misra, et al., Nucl. Acids Res. 17:8327 (1989)), from C. Schmid; the c-myc clone was HSR-1 (K. Alitano, et al., Proc. Natl. Acad. Sci. USA, 80, 1707 (1983)) from J. Bishop. Cosmids containing the entire p53 genomic region were from Y. Nakamura; cosmids including the entire DHFR amplicon from hamster cells (B. Anachkova and J.L. Hamlin, Mol. Cell Biol. 9, 532 (1989)) were from J. Hamlin.

Each clone was digested with an appropriate restriction endonuclease, end-labelled with ^{32}P, and incubated with p53 protein from a lysate of cells infected with a recombinant vaccinia virus expressing p53 protein. Labelled DNA fragments which bound to p53 were then recovered by immunoprecipitation with monoclonal antibodies against p53. Of the more than 1400 restriction fragments tested, only two bound reproducibly to p53 under the experimental conditions used: a 259 basepair HinfI fragment (fragment A) of clone 772 C_{BE} (Panel 2, Fig. 1A), and a 190 basepair HinfI fragment (fragment B) of clone Lambda 5R (Panel 3, Fig. 1A); these frag-

GeneDX 1009, pg. 518

ments bound to a far greater extent than any of the other labelled fragments of larger or smaller size present in the same assay mixes.

Example 2

This example demonstrates that the immunoprecipitation of fragment A is dependent on both p53 protein and anti-p53 antibodies.

The immunoprecipitation assay was performed on fragment A as described in Example 1. Cell lysates were produced by infection with vaccinia virus that did or did not contain an insert of wild-type p53 DNA. Either anti-p53 antibodies were used or normal mouse IgG was used.

Lysates from cells infected with wild-type vaccinia virus (devoid of p53) were not able to specifically immunoprecipitate fragment A (Fig. 1B). Similarly, the detection of the precipitation of fragment A was dependent on the presence of anti-p53 antibodies (Fig. 1B). The binding was evident in lysates prepared from either human HeLa cells or monkey BSC40 cells infected with vaccinia virus and expressing wild-type p53 (Fig. 1B).

Affinity-purified baculovirus-produced wild-type p53 protein was substituted for the vaccinia-infected cell lysates in the immunoprecipitation assay and found to bind fragment A strongly (Fig. 2A). This suggested that the binding to fragment A was an intrinsic property of the p53 polypeptide and not dependent on other factors present in the vaccinia virus-infected cell lysates.

Example 3

The example demonstrates that p53 mutant proteins found in human tumors fail to bind to fragment A.

Increasing quantities of wild-type and mutant 273^{his} p53 protein, affinity purified from a baculovirus expression system, were used to immunoprecipitate labelled fragment from C_{BE}. See Fig. 2A. The proportion of fragment A bound to wild-type p53 protein increased in tandem with the amount of p53 added to the assay mixture. (Fig. 2A) In contrast, fragment A did not specifically bind to a mutant form of p53 (273^{his}) protein even at the highest p53 protein concentration used. The 273^{his} mutation is the most common p53 mutant identified in human tumors. Another p53 mutant (175^{his}) protein commonly found in human tumors also failed to bind to fragment A (Fig. 1B).

Example 4

This example defines the particular sequences within fragment A that allow it to bind to wild-type p53 protein.

Fragment A was subcloned, and the 259 bp insert from the subclone (10d, SEQ ID: 1 and Fig. 3A) bound to p53 as expected. A strategy based on the polymerase chain reaction (PCR) and restriction endonuclease digestion was used to generate subfragments of this clone.

One primer for each PCR was labelled with ^{32}P at the 5′ end with T4 polynucleotide kinase in a 5 µl reaction, and the kinase inactivated at 70°C for 5 min. PCR contained 350 ng of each of the appropriate primers and approximately 50 ng plasmid template in a 50 µl reaction, using 25 cycles and the PCR conditions specified in Baker SJ, et al., Cancer. Res., 50:7717 (1990). The products were extracted with phenol and chloroform, ethanol-precipitated, and dissolved in 3mM Tris, 0.2 mM EDTA prior to binding. Subfragment 1 contained bp 1 to 425 of subclone 10d of fragment A (Fig. 3A); subfragments 1a, 1b, 1c, 1d, and 1e were generated by digestion of subfragment 1 with BamHI, MboI, HindIII, HindIII, and BamHI, respectively, from fragment 1. Subfragment 2, contained bp 283 to 425. Subfragment 3a was generated by digestion of subfragment 3 (bp 106 to 294) with Hae III. Subfragment 4a was produced from subfragment 4 (gp 1 to 141) by Hind III digestion. Subfragments 5a and 5b were products of the HaeIII digestion of subfragment 5 (bp 87 to 141). "Mutant" subfragments 5mut1 and 5 mut2 were produced using primers P3ml (5′-GAAAGAAAAGGCAAGGCCAGGAAAGT-3′) and P3mut2 (5′-GAAAGAAAAGGCAAGGCCATTAAAGT-3′) and were identical to subfragment 5 except for the positions underlined in the primers. Subfragment 6 contained bp 106 to 138, and the insert was excised by restriction with HindIII and BamHI to generate 6a or with HindIII and EcoRI to generate 6b. Subfragment 3, including basepairs 106 to 294 (Fig. 4B, panel 2) bound well to p53 as did subfragment 4, containing basepairs 1 to 141 (Fig. 4B, panel 3). This and similar assays done with additional subfragments (Figs. 4A and 4B) localized the critical sequences to basepairs 106 to 141. This segment contained three repeats of the sequence TGCCT (Fig. 3A). Digestion of subfragment 3 with HaeIII (cleaving between bp 125-126 and removing two of the TGCCT repeats) greatly reduced this binding (Fig. 4B, subfragment 3A, panel 2), suggesting that a critical sequence lay at or near this restriction site and that a single TGCCT repeat was not sufficient for binding. Additional subfragments were tested (#5, bp 87 to 141, Figs. 4A and 5B; #6, bp 106 to 138, Figs. 4A and 4B, panel

GeneDX 1009, pg. 519

4), and it was established that a 33 bp insert containing three TGCCT repeats provided binding capability.

Example 5

This example demonstrates that certain G residues are critical for binding of p53 to fragment A.

We studied the requirements for binding at the single nucleotide level using a methylation interference assay. PCR products labelled at one end were generated using primers labelled with T4 polynucleotide kinase (US Biochemicals). The product of each PCR reaction was purified by polyacrylamide gel electrophoresis, eluted from a crushed gel slice in 500 mM ammonium acetate, extracted with phenol and chloroform, and precipitated with ethanol. 2 x 10^6 dpm of DNA was methylated at guanine residues using dimethylsulfate as described (T. Maniatis, E.F. Fritsch, J. Sambrook, Molecular Cloning; A Laboratory Manual, 1st ed., Cold Spring Harbor Laboratory, 1983, p. 477), ethanol-precipitated and dissolved in 10 µl 3 mM Tris, 0.2 mM EDTA. 0.5 µl was removed as the DNA control. 4.5 µl was added to a binding reaction containing baculovirus-produced p53 or vaccinia-infected cell lysates. The precipitated DNA was purified by SDS/proteinase K digestion, extracted with phenol-chloroform, and ethanol-precipitated. The control DNA and precipitates of bound DNA were cleaved with piperidine at the methylated sites. Equivalent amounts of labelled DNA were loaded and separated on a denaturing polyacrylamide (14.5%) sequencing gel, which was fixed and dried for autoradiography. Subfragment 5, demonstrating efficient binding (Figs. 4A and 5B) was methylated in vitro, and immunoprecipitated after binding to p53. The bound DNA was then cleaved with piperidine at methylated residues and separated by electrophoresis on a sequencing gel. Assay of one strand (Fig. 5A, right) demonstrated that methylation at the G at bp 120 significantly interfered with binding. On the opposite strand (Fig. 5A, left), the most effective interference was produced by methylation at G residues at bp 117, 121, and 122. Partial interference was also produced by methylation at nearby G sites (bp 110-112, 114, and 115). Thus, the methylation interference assay pinpointed one of the central TGCCT repeats (centered at basepair 121) and adjacent basepairs as critical for binding.

To obtain independent evidence of the specificity for the G residues identified by methylation interference, in vitro mutagenesis was used. A DNA fragment was generated that was identical to subfragment 5 except for the substitution of G at bp 120, 121, and 122 with T residues. This "mutant" subfragment (#5mut2) bound poorly to p53 (Fig. 5B). A fragment identical to subfragment 5 except for a single basepair (T substituted for G at bp 120) was then tested. This fragment (#5mut1) also did not bind appreciably (Fig. 5B).

Example 6

This example defines the region of fragment B which is important for p53 binding.

Fragment B was subcloned (sequence in Fig. 3B and SEQ ID NO:2). Interestingly, this fragment had two repeats of the TGCCT motif (centered at bp 135 and 152). A PCR strategy similar to that used for fragment A was used to demonstrate that a 95 basepair subfragment (bp 104 to 198), which contained both of these repeats, was sufficient for binding (Fig. 4B, panel 5).

Example 7

This example shows that expression of the wild-type p53 gene in human colorectal carcinoma cells dramatically inhibits their growth and that a mutant p53 gene cloned from a human colorectal carcinoma was incapable of exerting such inhibition.

The colorectal carcinoma lines SW480 and SW837, which are representative of 75% of colon carcinomas, have each lost one copy of chromosome 17p (including the p53 gene) and the remaining p53 allele is mutated (Baker, et al., Science 244, 217 (1989); Nigro et al, Nature 342, 705 (1989)). The SW837 line contains an arginine to tryptophan mutation at codon 248 (Nigro, supra). The SW480 line contains two point mutations, arginine to histidine at codon 273 and proline to serine at codon 309 (Nigro, supra.). The substitutions at codon 248 and 273 are typical of those observed in human tumors, occurring within two of the four mutation "hot spots" (Nigro, supra).

For the transfection studies, we constructed a vector, pCMV-Neo-Bam, engineered to contain two independent transcription units. The expression vector pCMV-Neo-Bam was derived from plasmid BCMGNeo-mIL2 (Karasuyama, et al., J. Exp. Med. 169, 13 (1989) by excision of the human beta globin sequences and bovine papilloma virus sequences with Bam HI and Not I. Next, the interleukin 2 (IL-2) sequences present at the unique Xho I site were removed, and the Xho I site was changed to a Bam HI site by linker addition. The vector included CMV promoter/enhancer sequences, which could drive expression of the insert at the Bam HI site, and splicing and polyadenylation sites derived from the rabbit beta globin gene, which ensured proper processing of the

GeneDX 1009, pg. 520

transcribed insert in the cells. A pBR322 origin of replication and β-lactamase gene facilitated growth of the plasmid in E. coli. The plasmid conferred geneticin resistance through expression of the neomycin resistance gene under separate control of an HSV thymidine kinase promoter. The first transcription unit comprised a cytomegalovirus (CMV) promoter/enhancer upstream of a site for insertion of the cDNA sequences to be expressed, and splice and polyadenylation sites to ensure appropriate processing. The second transcription unit included a herpes simplex virus (HSV) thymidine kinase promoter/enhancer upstream of the neomycin resistance gene, allowing for selection of transfected cells in geneticin.

A wild-type p53 cDNA was inserted into pCMV-Neo-Bam to produce pC53-SN3. Similarly, a vector, pC53-SCX3, expressing a mutant cDNA from human colorectal tumor CX3, was also constructed. The only difference between pC53-SN3 and pC53-SCX3 was a single nucleotide (C to T) resulting in a substitution of alanine for valine at p53 codon 143 in pC53-SCX3. The two constructs were made as follows: a 1.8-kb Xba I fragment, extending from nucleotide -130 to 1671 relative to the translation initiation site, was isolated from wild-type or CX3 cDNA clones. The fragment was blunt-ended with the Klenow fragment of DNA polymerase, ligated to Bam HI linkers, and cloned into the unique Bam HI site in the expression vector pCMV-Neo-Bam.

The constructs were transfected into SW837 and SW480 cells (obtained from the American Type Culture Collection, Rockville, Maryland), and geneticin-resistant colonies were counted 3 weeks later. Cells transfected with pC53-SN3 formed five-to tenfold fewer colonies than those transfected with pC53-SCX3 in both recipient cell types (Table 1).

Table 1. Colony formation after transfection with wild-type and mutant p53 expression vectors. For each experiment, one or two 75-cm^2 flasks were transfected, and the total colonies counted after 3 to 4 weeks of selection in geneticin (0.8 mg/ml). Exp. experiment.

Cell line	Exp.	No. of geneticin-resistant colonies formed	
		pC53-SCX3 (mutant)	pC53-SN3 (wild-type)
SW837	1	754	66
	2	817	62
SW480	1	449	79
	2	364	26
RKO	1	1858	190
	2	1825	166
VACO 235	1	18	16
	2	26	28

In both SW837 and SW480 cells, the number of colonies produced by the expression vector pCMV-Neo-Bam (without a p53 cDNA insert) was similar to that induced by the pC53-SCX3 construct.

These results suggest that the wild-type p53 gene inhibits the clonal growth of both the SW837 and SW480 cell lines; however, a significant number of colonies formed after transfection of the wild-type construct. If wild-type p53 expression were truly inhibitory to cell growth, one would expect that no colonies would form or that p53 expression in the colonies that did form would be reduced compared to that produced with the mutant p53 cDNA construct. To evaluate this issue, we expanded independent SW480 and SW837 colonies into lines, and ribonuclease (RNase) protection analysis was performed to determine the amount of p53 mRNA expressed from the exogenously introduced sequences. Twelve of 31 lines (38%) derived from transfection with the pC53-SCX3 construct were found to express the exogenous mutant p53 mRNA. This percentage was consistent with results expected in human cells transfected with a vector containing two independent transcription unit. Previous studies have shown that, in contrast to rodent cells, primate cells are able to integrate only a small amount of foreign DNA (approximately 6 kb), so that only 10 to 30% of clones selected for the expression of one tran-

GeneDX 1009, pg. 521

scription unit also contain the second unit in an intact form (F. Collabere-Garapin, et al., Gene 50, 279 (1986); Hoeijmakers, et al., Exp. Cell Res., 169, 111 (1987); Mayne et al., Gene 66, 65 (1988), Dean, et al., Exp. Cell Res., 183, 473 (1989). In contrast, no expression of exogenous p53 wild-type mRNA was seen in any of 21 clonal lines established from either SW480 or SW837 cells transfected with the pC53-SN3 vector (Fig. 6A). These RNase protection results were supported by analysis of the exogenous p53 DNA sequences within the clones. All of the p53-expressing clones derived from the pC53-SCX3 transfection contained an intact copy of the exogenous p53 gene (Fig. 6B). In contrast, in all the clones derived from the pC53-SN3 transfection, the exogenous p53 sequences were deleted or rearranged (Fig. 6B).

The results from individual clones were further supported by the analysis of pooled clones, in which numerous colonies could be simultaneously assessed. Forty or more clones from two to three separate transfection experiments were pooled and analyzed approximately 3 weeks after transfection. RNase protection studies showed substantial expression of wild-type sequences was not detectable (Fig. 7A). Result from Southern (DNA) blotting were consistent with the RNase protection studies, in that poled colonies from the wild-type transfectants had no detectable unrearranged exogenous p53 sequences, in contrast to the intact p53 sequences in colonies derived from the mutant p53 cDNA expression vector (Fig. 7B).

The conclusions made from the above experiments are dependent on the assumption that p53 protein was produced in the transfected cell lines. Clones containing exogenous mutant p53 sequences produced p53 mRNA at a concentration 1.5 to 3.5 times higher than that produced by the endogenous p53 gene (Figs. 6A and 7A). Immunoblot analysis showed that there was a concomitant small increase in p53 protein expression in the transfectants (1.5- to 3-fold) compared to the untransfected cells. However, this increase was difficult to measure quantitatively, since these cells produced significant amounts of endogenous p53 protein that (unlike endogenous p53 mRNA) could not be distinguished from that produced by the vectors. To confirm that transfected human cells expressed p53 protein from our constructs, we studied an additional colorectal carcinoma cell line (RKO). RKO cells were obtained through the generosity of M. Brattain. Although RKO cells did not contain a mutation within the susceptible p53 coding sequences, i.e., exons 5-9, they expressed low concentrations of p53 mRNA compared to normal colorectal mucosa or the other lines studied and did not produce detectable amount of protein.

Results of colony formation assays in transfected RKO cells were similar to those in SW480 and SW837 cells. Colony formation by wild-type p53 gene transfectants occurred with a tenfold decrease in efficiency compared to the mutant p53 construct (Table 1). Immunocytochemical detection of p53 protein in transfected RKO cells was done as follows: approximately 5×10^4 cells were cytocentrifuged onto polylysine-coated slides, fixed for 10 min in formalin, and permeabilized for 5 min in 0.5% Triton X-100. A mouse monoclonal antibody against human p53 protein (Ab1801) in combination with the ABC immunoperoxidase system (Vector Laboratories), was used for immunocytochemical detection of p53 protein (Banks, et al., Eur. J. Biochem. 159, 529 (1986)). Ten to 20 randomly selected microscopic fields were analyzed per slide. These observations are consistent with the greater stability of mutant compared to wild-type p53 protein noted previously (C.A. Finlay et al., Mol. Cell Biol. 8, 531 (1988)). However, transient mRNA expression was also significantly lower in the SN3 transfectants compared to the SCX3 transfectants at 48 and 96 hours, supporting the idea that RKO cells expressing wild-type p53 were at a selective disadvantage compared to those producing mutant p53 products.

To obtain additional evidence that cells expressing p53 were inhibited in their growth potential, we examined the effect of p53 gene expression on DNA synthesis in transfected RKO cells were labeled with [^{3}H] thymidine for 2 hours. The cells were subsequently fixed, immunocytochemically stained for the presence of p53 protein, and autoradiographed. The number of cells undergoing DNA replication was only slightly lower in cells producing exogenous mutant p53 protein than in cells that did not express any detectable p53 protein. Expression of the wild-type protein, however, dramatically inhibited the incorporation of thymidine (Table 2).

GeneDX 1009, pg. 522

Table 2. Immunocytochemistry and [^{3}H]thymidine incorporation of transfected RKO cells.

Plasmid	Percent of cells expressing p53 protein at				Percent of cells incorporating [^{3}H]thymidine in	
	6 hrs	24 hrs	48 hrs	96 hrs	p53 expressors	p53 non-expressors
pC53-SCX3	1.0	11	4.3	2.0	24	31
pC53-SN3	1.9	5.2	0.3	0.2	1.7	33

These results all suggested that wild-type p53 exerted an inhibitory effect on the growth of carcinoma cells in vitro. To evaluate whether this inhibitory effect was cell type-specific, we transfected colorectal epithelial cells derived from a benign tumor of the colon (the VACO 235 adenoma cell line). VACO 235 cells are described by J.K.V. Willson et al., Cancer Res., 47, 2704 (1987). Previous studies have shown that most adenomas contain two copies of chromosome 17p and express wild-type p53 mRNA at concentrations similar to that of normal colonic mucosa. Analogously, the p53 alleles of the VACO 235 cell line were sequenced (exons 5-9) and found to be wild-type and the expression of p53 mRNA was found to be similar to that of normal colorectal mucosa. In contrast to the results seen with SW480, SW837, and RKO cells, the pC53-SN3 and pC53-SCX3 constructs produced similar numbers of geneticin-resistant colonies after transfection of the VACO 235 line (Table 1). We considered, however, that the most definitive test for differential growth inhibition by wild-type versus mutant p53 genes involved analysis of exogenous p53 expression in pooled transfectants. Through such analysis, a large number of colonies could be examined simultaneously and the expression of exogenous mutant and wild-type p53 genes directly compared. Striking differences in the relative expression from the transfected genes were seen in all three carcinoma cell lines tested. VACO 235 transfectants, however, expressed similar amounts of exogenous p53 mRNA from either pC53-SN3 (wild-type) or pC53-SCX3 (mutant) p53 construct (Fig. 7A).

In summary, our results suggest that expression of the wild-type p53 gene in colorectal carcinoma cell lines was incompatible with proliferation. The inhibitory effects of wild-type p53 were specific in two ways. First, a single point mutation in a p53 gene construct abrogated it suppressive properties as measured by three separate assays (colony formation, exogenous p53 expression in transfected clones, and thymidine incorporation). The CX3 mutant provided a control for gene specificity as it contained only one conservative mutation, resulting in a substitution of one hydrophobic amino acid (alanine) for another (valine) at a single codon. Second, the growth-suppressive effect of the wild-type p53 construct was cell type-specific. Introduction of the wild-type vector into the VACO 235 adenoma cell line had no measurable inhibitory effect compared to the mutant p53 vector. There are several differences between the cell lines that could account for the differential effect of the introduced vectors. Regardless of the basis for the difference, the results with the VACO 235 cell line minimize the possibility that the wild-type p53 construct had some nonspecific, toxic effect on recipient cells; the effect was cell type-dependent.

The transfection and expression results of Table 1 and Fig. 7A suggest that cells at the premalignant stages of tumor progression (VACO 235) may be less sensitive to the inhibitory effect of wild-type p53 than malignant cells (SW480, SW837, and RKO). This hypothesis is consistent with previous results that suggest the wild-type p53 is less inhibitory to the growth of normal rat embryo fibroblasts than to their oncogene-transfected derivatives. Finlay et al., Cell 57, 1083 (1989); Eliyahu et al., Proc. Natl. Acad. Sci. USA, 86, 8763 (1989). This sensitivity may only be relative: expression of the wild-type gene at high concentrations might inhibit the growth of any cell type, including non-neoplastic cells, by overwhelming normal regulatory processes such as phosphorylation. Samad et al., Proc. Natl. Acad. Sci. USA, 83, 897 (1986); Meek et al., Mol. Cell Biol. 8, 461 (1988). Genetic alterations that occur during the progression of colorectal tumors may increase the sensitivity of cells to p53 inhibition, making wild-type p53 expression a key, rate-limiting factor for further tumor growth and expansion. At this point, and not before, mutations in the p53 gene would confer a selective growth advantage to cells in vivo, which would explain the frequent occurrence of p53 gene mutations and allelic loss only in the more advanced stages of colorectal tumorigenesis.

GeneDX 1009, pg. 523

Example 8

This example demonstrates the identification of human genomic fragments that can bond to wt p53 protein in vitro.

This schema in Figure 8A outlines the experimental strategy used for the isolation and analysis of p53 binding sites. Total human genomic DNA fragments were ligated to specially designed "catch" linkers (Kinzler, et al. (1989), Nucleic Acids Research, 17:3645-3653, and Kinzler, et al. (1990), Molec. Cell. Biol., 10:634-642) to allow for subsequent PCR amplification and cloning. The linked genomic DNA was then incubated with wt p53 and precipitated with anti-p53 antibodies. The bound DNA was subsequently amplified by PCR using primers complementary to the catch linkers, and the process repeated. After four rounds of sequential immunoprecipitation and PCR, the amplified and selected (AS) DNA was cloned. Clones were picked at random and tested for p53 binding first by immunoprecipitation (IP), and then by methylation interference (MI) and DNase I protection (DP).

Following the outline in Figure 8A., we tested the inserts of 530 clones for binding to p53. Restriction fragment of the clones were end-labeled and incubated with purified human wt p53 protein produced in baculovirus-infected cells.

Whole-genome PCR was performed as previously described, except that only one oligonucleotide (5′-GAGTAGAATTCTAATATCTC-3′) was used for amplification (Kinzler, et al. (1989), Nucleic Acids Research, 17:3645-3653, and Kinzler, et al. (1990), Molec. Cell. Biol., 10:634-642). Two hundred ng of "catch"-linked human genomic DNA were incubated with 100 ng of baculovirus-produced human wt p53 purified as described (Friedman, et al. (1990), Proc. Natl. Acad. Sci. U.S.A., 87:9275-9279), and immunoprecipitated as described below. After 4 rounds of IP and PCR, the AS DNA was cleaved with Eco RI and cloned into either the vector Lambda Zap II or pBluescript II SK+ (Stratagene). Individual clones were picked at random and tested for p53 binding. In panel B, cloned plasmid DNA samples were cleaved with Eco RI and end-labeled by Klenow fill-in. For IP (McKay, et al. (1981), J. Mol. Biol, 145:471-479), ten ng of DNA were incubated with 100 ng of baculovirus-produced human wt p53 and 100 ng of poly dI-dC at 4°C for 30 minutes in 100 µl of "DNA-binding buffer" containing 100 mM NaCl, 20 mM Tris, pH 7.0, 10% glycerol, 1 % NP40, and 5 mM DTT. DNA fragments bound to p53 were complexed to antibodies by the addition of 8 µl containing 400 ng each of anti-p53 antibodies pAb421 and pAb1801, both obtained from Oncogene Science, and incubated for 30 minutes at 4°C. The DNA-binding buffer containing 1.5 mg protein A precipitated following the addition of 26 µl of DNA-binding buffer containing 1.5 mg protein A Sepharose and 10 µg of poly dI-dC and mixing at 4°C for 30 minutes. After removal of the supernatant, the immunoprecipitate was washed twice with 1 ml of DNA-binding buffer. Bound DNA was purified by treatment with SDS and proteinase K at 48°C for 30 minutes, extracted with phenol and chloroform, precipitated with ethanol, separated by electrophoresis on a 10% nondenaturing polyacrylamide gel, and autoradiographed.

Twenty-three of the clones were found to contain fragments that bound to p53. Examples of the IP experiment are shown in Figure 8B. Clone S61 (lanes 11B,C) contains a single genomic fragment of 202 bp which bound to p53. Clone N2 contained five fragments, only one of which (357bp) bound to p53 (lanes 10B,C). Other examples of p53-binding fragments were obtained, and each of these was subcloned for further analysis. In contrast, we found that none of over 1000 clones containing unselected human DNA inserts of similar size bound to p53 using the IP assay. Thus, the whole-genome PCR procedure significantly enriched for p53-binding sequences.

Example 9

This example demonstrates the localization of p53 contacts with bound DNA fragments.

Localization of the regions bound by p53 was obtained by DP or MI assays using the subcloned DNA fragments as probes. For MI, the fragments were methylated at G residues and bound to p53 (Fig. 9). Methylation of G residues critical for p53 binding resulted in interference with IP. For example, methylation at nucleotides, 217, 22, 227 to 229, and 233 of the 248 bp insert from clone 11B3 completely interfered with the binding of this fragment to p53 (Fig. 9, footprint 2). When the opposite strand was analized, interference was observed at the G residues corresponding to nucleotides 219, 223, 224, 230, 235, and 236 (Fig. 9, footprint 1). For DP, labelled DNA fragments were first subject to IP, then incubated with various amounts of DNase I. For clone N22, p53 binding provided protection against DNase I cleavage at residues 187 to 211 (Fig. 9, footprint 9). MI showed interference by G residues only within the region protected by DNase I (Fig. 9, footprint 10). Other examples of DP and MI mapping are shown in Figure 9. p53-binding DNA fragments were subcloned and labeled on one end, gel-purified and subjected to DP or MI mapping. For MI, 10 ng of DNA were incubated in 200 µl of 50 mM Na-cacodylate, 1 mM EDTA, pH 8.0 and 5 µl of 10% dimethylsulfate/90% ethanol for 5 minutes at

GeneDX 1009, pg. 524

20°C to methylate G residues. Fifty μl containing 1.5 m Na-acetate, 1 M β-mercaptoethanol and 60 μg of glycogen were added. The mixture was ethanol-precipitated, washed, and resuspended in 5 μl of 3mM Tris, 0.2 mM EDTA, pH 7.5, and allowed to bind to wild-type p53 as described in the legend to Figure 1. After IP and DNA purification, the samples were incubated with 100 μl of 1 M piperidine at 90°C for 30 minutes. The samples were then dried under vacuum and separated electrophoretically on a 6% polyacrylamide sequencing gel. The control DNA samples were carried through all incubations except no p53 was added. For these control samples, the protein A Sepharose pellets were treated with SDS and proteinase K without removal of the supernatants (which contained the labeled DNA in the absence of p53).

For DP assays, end-labeled DNA fragments were immunoprecipitated as described in the legend to Figure 8. The protein A Sepharose pellet were incubated for two minutes at 25°C with 200 ng DNase I in 5 mM $MgCl_2$. After purification of the DNA, as described above, samples were separated by electrophoresis on sequencing gels and loaded as described above for MI. MI was performed on all 18 genomic DNA fragments which bound to p53. DP assays were performed on 13 fragments and the regions of protection uniformly coincided with those indicated by the MI assays.

Example 10

This example analyzes the sequences of p53-bound DNA regions.

We next compared the sequences of the twenty-three clones. The average insert sequence was 307 bp (range 139-470). We found that ten of the twenty-three clones were not unique, showing at least one hundred contiguous nucleotides identical to one other clone. Thus, the twenty-three clones represented only eighteen independent genomic DNA fragment. We attempted to find similarities among these eighteen fragments by computer methods, but found no significant relationships. However, when we aligned the regions involved in p53 binding (as assessed by MI and DP), a striking and consistent feature of the clones became apparent (Fig. 10). Each of the binding sites contained two copies of the 10 bp motif 5′-RRRCWWGYYY-3′, separated by 0 to 13 bp. One clone (S592) contained two separate areas of footprinting, and both regions contained a dimer of the 10 bp motif (Fig. 10). In all clones, the regions displaying DP and MI were always centered within the dimers, and G residues within the 10 bp motif strongly interfered with binding to p53 (examples in Fig. 9). The 10 bp consensus monomer contained an internal symmetry, with two oppositely oriented half-sites of the form 5′-RRRCW-3′. This symmetry was extended in the dimers, which contained four half-sites oriented in alternating directions, forming a pseudopalindromic structure, sometimes with an intervening loop. This consensus dimer was also recognized in the p53 binding sequence mapped within plasmid CBE10d (Fig. 10). Despite the remarkable symmetry noted for all p53 binding sequences, none of the genomic sites were palindromic.

Example 11

This example demonstrates that a dimer of the consensus motif is required to bind p53 and that mutant p53 proteins found in tumors do not bind to the consensus sequence.

To determine if the 10 bp consensus monomer could bind to p53, a synthetic oligonucleotide (5′-AGGCATGTCT-3′) containing the consensus sequence was studied. Oligonucleotide duplexes were tested either directly or after cloning into plasmid vectors. The monomer was found not to bind to p53, either alone (not shown) or flanked by 43 nucleotides of plasmid sequences (Fig. 11A, lane 5). In contrast, the dimers (composed of two copies of the monomer arranged in head-to-head, tail-to-tail, or head-to-tail orientation), each bound strongly to p53 protein (Fig. 11A, lanes 1 to 4, 6). Higher-order oligomers of the monomer did not bind any better than the dimer in the IP assay (Fig. 11A, lanes 6 to 8). A different monomer, still fitting the consensus sequence, but perfectly palindromic, also bound as a dimer but not as a monomer (Fig. 10, synthetic oligonucleotides 3,4). Two variants of the consensus motif were also tested for binding. In the first, the two critical G:C bp's at position 4 and 6 of the monomer were substituted with A:T bp's (Fig. 10, synthetic oligonucleotides 1,2). Although this sequence was perfectly symmetrical, it did not bind to p53 either as a monomer or as a dimer. We also tested direct repeats of the 5′-PuPuPuC(A/T)-3′ half-site, and found that these did not bind to p53 (Fig. 10, synthetic oligonucleotide 5). Thus, the mirror-image symmetry of the half-sites within the 10 bp consensus monomer was critical for its activity.

Finally, we tested p53 mutants representing each of the four "hot-spots" frequently altered in human cancers ((Nigro, et al. (1989), Nature, 342:705-708, and Holstein, et al. (1991), Science, 253:49-53) for ability to bind to the consensus dimer. None of the p53 mutants bound appreciably to this sequence (Fig. 11B) under conditions where the wt protein bound strongly. These experiments also showed that in vitro translated p53, as well as that purified from baculovirus-infected insect cells, had the capacity to bind DNA specifically.

In summary, a set of human genomic DNA sequences which could bind p53 were isolated and used to de-

GeneDX 1009, pg. 525

fine a consensus binding sequence for p53. The symmetry of the four half-sites within the consensus dimers suggests that p53 interacts with DNA as a tetrameric protein. This is consistent with studies suggesting that p53 assembles into homotetramers (Krais, et al. (1988), J. Virol., 62:4737-4744, and Weinberg (1991), Science, 254: 1138-1146).

Example 12

This example demonstrates that intact p53 can activate expression in human cells.

We first made reporter plasmids (PG_n-CAT series) containing part of the polyomavirus early promoter and the CAT gene located downstream of DNA sequences which could bind to p53 in vitro (Figure 8). For the CAT reporters, concatemers of the p53-binding region of C_{BE} were formed by ligation of complementary oligonucleotides, ligated into the EcoRV site of pBluescript II SK+ (Stratagene) to form the PG_n and MG_n series. The BglII-BamHI fragment of pPyOICAT (Murakami, et al. (1990) Oncogene, 5:5), containing the polyomavirus early promoter and the CAT gene coding region, was ligated into the BamHI site of the PG_n and MG_n series clones to form the PG_n-CAT and MG_n-CAT series, and the orientation of the inserts characterized by restriction enzyme analysis. The PG_9-MG_n-CAT and PG_{13}-MG_n-CAT series were formed by excising the HindIII-SalI fragments of PG_9-CAT and PG_{13}-CAT, blunt-ending, attaching XbaI linkers, and ligating into the XbaI site of the MG_n-CAT series plasmids (where n=1, 5, 10, and 15). For the yeast β-galactosidase reporter plasmids, PG and MG sequences were ligated as SalI-SmaI fragments to the SalI and filled-in XhoI sites of pCZ (Buchanan, et al. (1988), Mol. Cell Biol., 8:50806). The construction of the p53-wt expression construct has been described (Baker, et al. (1990), Science, 249:912); the mutant expression plasmids were constructed similarly from the previously described cDNA plasmids (Nigro, et al. (1989), Nature, 342:705, and Kern, et al. (1991), Oncogene, 6:131), or in the case of the engineered phosphorylation site mutants, by in vitro mutagenesis (Altered Sites, Promega) with verification by sequencing. The construction of the yeast p53 expression vectors based on pRS314 has been described (Nigro, et al., Mol. Cell Biol. (in press)).

For the p53 binding sequences, we used a series of concatemers of the oligonucleotide PG (5′-CCTGCCTGGACTTGCCTGG-3′). This contained the binding region of plasmid C_{BE}, previously shown to bind p53 in vitro. The reporter and an expression vector coding for the intact human wild-type protein (p53-wt) (Fig. 12B), were transfected together into the human colorectal cancer cell line HCT 116. This line makes low amount of apparently wild-type p53 protein.

Exons 5-8 of the p53 genes from HCT 116 cells were amplified by PCR and sequenced as described in Sidransky et al. (1991), Science, 252:706. Previously, these exons had been shown to contain over 90% of the mutations observed in human tumors. No mutations were observed from HCT 116 cells. Small amount of apparently wild-type protein could be detected in Western blots of HCT 116 protein (See Fig. 16B).

The intact wild-type p53 protein was indeed able to activate transcription (Fig. 13). We found that the level of transactivation of the CAT gene depended on the strength of binding to p53 of the upstream sequences. Thus, the longer the number of PG repeats, the greater the binding to p53 in vitro (Fig. 13A) and the higher the CAT expression in vivo (Fig. 13B, lanes 1-7). The level of transactivation also increased with the amount of p53 expression vector transfected (Table 3).

GeneDX 1009, pg. 526

Table 3. Transactivation in human cells

p53 Expression Vector[2/]	Reporter	Relative CAT Activity[1/] Exp. 1	Exp. 2 .85 µg	Exp. 2 2.55µg	Exp. 3
p53-wt	$PG_{13}CAT$	100	100	210	100
Tumor-derived mutations:					
p53-143	$PG_{13}CAT$	3	1	4	
p53-175	$PG_{13}CAT$	2	3	6	
p53-248	$PG_{13}CAT$	2	4	7	
p53-273	$PG_{13}CAT$	2	2	3	
Endogenous activities:					
p53-wt	$MG_{15}CAT$[3/]		<1		2
none[4/]	$PG_{13}CAT$		20		2

1/ All transfections used 1.7 µg reporter in HCT 116 cells. Transfection and CAT assays were performed as described (Sadransky, et al. (1991), Science, 252:706). Activities reflect the fraction of chloramphenicol converted to an acetylated form, expressed as relative values with one transfection of each experiment arbitrarily designated as having a value of 100. Three representative experiments (Exp.) shown.

2/ Experiment 2 used both 0.85 and 2.55 µg levels of expressor. Experiment 1 used 1.7 µg, and experiment 3 used 0.85 µg.

3/ MG_{15}-CAT provided a control for DNA-binding specificity and an estimate of background from promoter-independent "readthrough" transcription and/or basal promoter activity.

4/ The expressor-negative transfection provided an estimate of the activity from endogenous wild-type p53 of the HCT 116 cells. In various transfections, the endogenous activity was 2-20% of that following transfection with exogenous wild-type p53, depending on the efficiency of transfection. The activity from endogenous p53 was diminished by transfection of oncogenic mutant p53.

Specificity for the binding sequences was tested by replacing six GC basepairs in the 18-basepair oligonucleotide to generate a mutant form of PG, termed MG (5′-CCTTAATGGACTTTAATGG-3′). This sequence, when

GeneDX 1009, pg. 527

multimerized and placed upstream of the CAT reporter, did not bind to p53 in vitro (Fig. 14, lanes 7 and 8), nor did it activate CAT expression in vivo (Table 3; Fig. 13B, lane 8; Fig. 15A, lane 6, Fig. 15B, lane 2). The transactivation of CAT was independent of the orientation of the PG multimer upstream of the CAT gene (Fig. 13B). Placing an additional 59 to 333 basepairs between the PG multimer and the promoter also had little effect on transactivation (Fig. 15A, lanes 2-5). However, placement of the PG binding sequences downstream of the CAT gene did not allow transactivation (Fig. 15, lane 3). These observations indicated that the PG sequences act as an upstream activator element, although not as a classical enhancer.

CAT assay. Cultures of HCT 116 cells at 50-80% confluence in 25 cm^2 flasks were transfected using Lipofectin (BRL, Gaithersburg, MD) according to the manufacturer's instructions. All flasks within an experiment were transfected with the same total amount of plasmid using pCMVneoBam or pBluescript II SK+ as "filler". Cells were harvested at 20-24 hr. and the CAT activity of the lysates measured by acetylation of ^{14}C-labeled chloramphenicol (ICN) as previously described (Gorman, et al. (1982), Mol. Cell Biol., 2:1044). The Bio-Rad protein assay was used to assure equivalence of lysate protein. Percent conversion to the acetylated form of chloramphenicol was calculated after quantitation by scintillation counting of excised chromatographic spot. Results reported are representative of at least two transfections done on separate days.

Example 13

This example demonstrates that oncogenic mutant p53 genes consistently fail to transactivate.

The precise correlation between the binding of the PG multimer in vitro (Fig. 13A) and the expression from PG_n-CAT reporters in vivo (Fig. 13B) strongly suggested that this expression was a direct result of p53 binding to PG multimers in vivo. If such an activity of p53 were crucial to it tumor suppressor activity, one would expect that naturally occurring mutants of p53 would be defective in this function. p53 mutations generally occur in four different p53 "hot-spot" regions, representing four evolutionarily conserved domains of the p53 protein (Baker, et al. (1989), Science, 244:217, Hollstein, et al. (1991), Science, 253:49, and Soussi, et al. (1987), Oncogene, 1:71). Mutant p53 genes representing each of these hot-spot were transiently transfected into cells together with the PG_{13}-CAT reporter plasmid. The consistency of the result obtained with these mutant expression vectors was striking. All tumor-derived mutations examined had lost the ability to transactivate CAT, despite the wide range of positions (codons 143-273) in which these mutations were situated (Table 3, and Figure 16A). As might be expected for a cell line expressing low amounts of endogenous wild-type p53, there was often some innate transactivating ability associated with the HCT 116 cells themselves (Table 3; Fig. 13B, lane 1; Figure 16A, lane 1); interestingly, the levels of CAT observed with the mutants (Table 3 and Fig. 16A, lanes 4-7) were actually less than that observed with no exogenous p53 expression vector, suggesting that the mutant p53 products might be inhibiting the low amounts of expression mediated by the endogenous p53 (discussed below). The mutants studied included one which was typical of those found in the germline of Li-Fraumeni patients (248^{trp})(Malkin, et al. (1990), Science, 250:1233; Strivastava, et al. (1990), Nature, 348:747) and three found commonly in a variety of human tumors (143^{val}, 175^{his}, 273^{his})(Hollstein, et al. (1991), Science, 253:49). Immunoblots confirmed that mutant p53 proteins were expressed at levels comparable to those of wild-type p53 (Fig. 16B).

Western blots were performed as follows: 100 μg protein of each cell lysate was separated on a 10% polyacrylamide SDS gel, transferred onto a PVDF (Millipore, Bedford, MA) membrane, blocked with 5% nonfat milk, incubated with 1 μg/ml PAb1801 (Oncogene Sciences) then ^{125}I-labeled goat-anti-mouse Ab (NEN), and autoradiographed. The inability of p53 proteins having tumor-derived mutations to bind specifically to the p53 recognition sequences in this construct (Fig. 14, lanes 3-6) is thus consistently reflected as defects in transactivation.

Example 14

This example demonstrates that intact p53 transactivates in yeast.

We next sought to determine whether this activity of p53 was confined to mammalian cells. If the ability of p53 to bind to DNA sequences in vivo and activate the transcription of adjacent genes were an intrinsic feature of the protein, it was possible that this activity would be manifest in simpler eukaryotes. Indeed, the N-terminal acidic activation domain of p53 has been reported to function in S. cerevisiae when fused to the DNA-binding domain of GAL4 (Fields, et al. (1990), Science, 249:1046). We thus stably transfected yeast with the LacZ reporter gene placed downstream of PG multimers (PG_n-LacZ), as well as with a galactose-inducible p53 expression vector (Fig. 12). The addition of galactose to the medium resulted in p53 expression accompanied by a dramatic elevation in β-galactosidase expression (Table 4). When the MG multimer (non-p53 binding) was substituted for the PG multimer, no activation was observed (Table 4). Moreover, p53 mutants had no capacity

GeneDX 1009, pg. 528

to transactivate in yeast cells (Table 4). Thus, the results in S. cerevisiae were analogous to those observed in human cells.

Table 4. Activation of Gene Expression in Yeast by p53

p53 Expression Vector	Reporter	β-Gal Activity Exp. 1	Exp. 2
none	PG16-LacZ	3	2
Yp53	PG16-LacZ	13,000	8,000
Yp53-143	PG16-LacZ	150	85
Yp53-273	PG16-LacZ	5	2
Yp53	PG4-LacZ	15,000	11,000
Yp53	PG1-LacZ	2,600	3,000
Yp53	MG15-LacZ	15	4

Wild-type or mutant p53 expression vectors and the β-galactosidase reporter plasmids were transfected into S. cerevisiae and clones obtained. p53 expression was induced with galactose, and β-galactosidase activity was measured in units of nanomoles per minute per milligram of protein. Two independent clones (Exp. 1 and Exp. 2) were tested. Less than 1 unit was seen in the absence of galactose induction. The residual activity of the Yp53-143 mutant may have been due to a slight wild-type activity observable with valine to alanine substitution mutants at the relatively low temperature (30°C) used for yeast growth.

Example 15

This example demonstrates the dominant-negative effect of p53 mutations on wild-type p53's ability to bind DNA and enhance transcription.

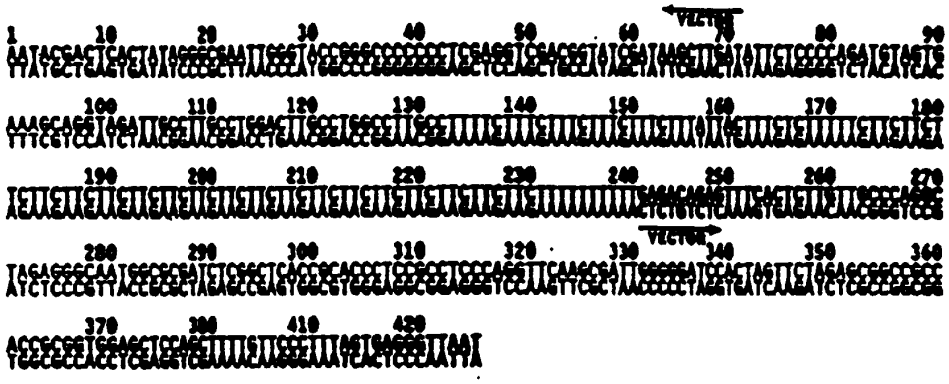
If the transcriptional activation of p53 were fundamental to its biological role, p53 mutants should be able to interfere with this activation, just as p53 mutants can interfere with the tumor-suppressor activity of wild-type p53 when both are expressed (Levine, et al. (1991), Nature, 351:453). To assess this possibility, we co-transfected the wild-type and mutant p53 genes with the PG_{13}-CAT reporter plasmid into HCT 116 cells (Figs. 17 and 18). When equal amounts of wild-type and mutant p53 expression vectors were used, the expression of CAT decreased by approximately 50% compared to that achieved by wild-type p53 alone. A three-fold ratio of mutant p53 to wild-type p53 produced an 84-95% reduction. When additional wild-type p53 was substituted for mutant p53 in the cotransfection experiment, the CAT expression increased rather than decreased, as expected (Fig. 18, "wt" lane). Thus, expression of the CAT reporter was dependent on the ratio of wild-type to mutant p53 in the cell.

The dominant-negative effect could be caused by a failure of the mutant/wild-type complexes to bind to DNA or to a failure to activate transcription once bound. To distinguish between these two possibilities, we co-translated wild-type p53 and the 175^{his} mutant of p53 in an experiment similar to that shown in Figure 14. The mutant p53 protein inhibited the ability of wild-type protein to bind DNA by over 90% when co-expressed at a 3:1 ratio.

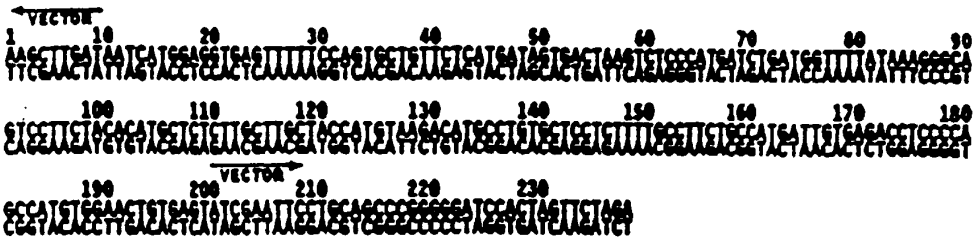
GeneDX 1009, pg. 529

SEQUENCE LISTING

SEQ ID NO: 1



SEQ ID NO: 2



Claims

1. A method for detecting the presence of wild-type p53 protein in a cell, comprising the steps of:
   contacting a p53-specific binding DNA fragment with a cell lysate from a tissue of a human, to bind the DNA fragment to wild-type p53 present in the cell lysate;
   detecting the presence of wild-type p53 protein in the cell by detecting binding of the DNA fragment to wild-type p53.

2. The method of claim 1 wherein the p53-specific binding DNA segment comprises nucleotides 103 to 134 as shown in SEQ ID No:1.

3. The method of claim 1 wherein the p53-specific binding DNA segment comprises nucleotides 104 to 123

GeneDX 1009, pg. 530

as shown in SEQ ID No:1.

4. The method of claim 1 wherein the p53-specific binding DNA fragment comprises more than one monomer of the sequence 5′-RRRCWWGYYY-3′.

5. The method of claim 1 wherein there are from 0 to 40 nucleotides between said monomers.

6. The method of claim 1 wherein the DNA fragment is labelled with a detectable moiety selected from the group consisting of:
a radioactive moiety, a colorimetric moiety, or a fluorescent moiety.

7. The method of claim 1 wherein the step of determining the amount of p53-specific binding DNA comprises:
immunoprecipitating p53 protein with anti-p53 monoclonal antibodies.

8. A method of detecting the presence of a wild-type p53 protein in a cell, comprising the steps of:
providing a histological section from a human;
incubating the section with a detectably-labeled p53-specific binding DNA fragment to bind said DNA fragment to wild-type p53 present in the histological sample;
removing unbound DNA fragment from the histological section; and
determining the amount of DNA fragment which is bound to the histological sample.

9. The method of claim 8 wherein the p53-specific binding DNA segment comprises nucleotides 103 to 134 as shown in SEQ ID NO:1.

10. The method of claim 8 wherein the p53-specific binding DNA segment comprises nucleotides 104 to 123 as shown in SEQ ID NO:1.

11. The method of claim 8 wherein the p53-specific binding DNA fragment comprises more than one monomer of the sequence 5′-RRRCWWGYYY-3′.

12. The method of claim 11 wherein there are between 0 and 40 nucleotides between said monomers.

13. The method of claim 8 wherein the DNA fragment is labelled with a detectable moiety selected from the group consisting of:
a radioactive moiety, a colorimetric moiety, or a fluorescent moiety.

14. A method of providing the physiological effect of wild-type p53 protein to a cell, comprising the steps of:
providing to a cell a compound which is is able to complex specifically with a p53-specific binding site.

15. The method of claim 14 wherein the compound comprises a single-stranded , linear or circular, oligonucleotice or oligonucleotide containing nucleotide analogs which can form a complex with a p53 specific DNA binding site.

16. The method of claim 14 wherein the compound comprises a polypeptide.

17. The method of claim 16 wherein the polypeptide comprises all or a part of human wild-type p53 protein.

18. The method of claim 15 wherein the oligonucleotide or oligonucleotide containing nucleotide analogs comprises the monomer sequence RRRCWWGYY or the complement thereof.

19. The method of claim 14 wherein the compound comprises at least a portion of the monomer sequence RRRCWWGYY as well as sequences adjacent to said monomer sequence in the human genome.

20. The method of claim 18 wherein the oligonucleotide or oligonucleotide containing nucleotide analogs comprises more than one monomer of said sequence.

21. The method of claim 18 wherein the oligonucleotide or oligonucleotide containing nucleotide analogs comprises between 0 and 40 nucleotides between said monomers.

22. A double-stranded DNA fragment which comprises a p53-specific DNA binding site, wherein the fragment

GeneDX 1009, pg. 531

comprises more than one monomer of the sequence RRRCWWGYYY and wherein the fragment is covalently attached to an insoluble polymeric support.

23. A linear or circular single-stranded oligonucleotide or oligonucleotide containing nucleotide analogs which is able to complex specifically with a p53-specific binding site, said binding site comprising more than one monomer of the sequence RRRCWWGYYY.

24. The oligonucleotide or oligonucleotide containing nucleotide analogs of claim 23 which contains one or more switchback linkers which allow the oligonucleotide to complex with both strands of the p53-specific binding site.

25. The oligonucleotide or oligonucleotide containing nucleotide analogs of claim 23 which is selected from the group consisting of a methylphosphonate, an aminomethylphosphonate, an aminomethylphosphonate, a phosphorothioate, a phosphorodithioate, a substituted or unsubstituted phosphoramidate, an oligoribonucleotide, an oligodeoxyribonucleotide, an alpha-oligonucleotide and mixtures thereof.

26. The oligonucleotide or oligonucleotide containing nucleotide analogs of claim 23 which is terminated at the 3′ or 5′ end with a moiety which reduces susceptibility to oligonucleotide degradation or facilitates uptake by the cells.

27. The oligonucleotide or oligonucleotide containing nucleotide analogs of claim 26 wherein the moiety is selected from the group consisting of: a substituted or unsubstituted amino moiety, polyethylene glycol, polylysine, acridine, dodecanol, and cholesterol.

28. A method of identifying compounds which specifically bind to p53-specific DNA binding sequences, comprising the steps of:

contacting a p53-specific-binding DNA fragment with a test compound to bind the test compound to the DNA fragment;

determining the amount of test compound which is bound to the DNA fragment.

29. The method of claim 28 wherein soluble DNA fragments are incubated with the test compound and the p53-specific-binding DNA fragment immobilized on a solid support, said soluble DNA fragments not having the ability to specifically bind wild-type p53 protein.

30. A method of identifying compounds which specifically bind to p53-specific DNA binding sequences, comprising the steps of:

contacting a p53-specific-binding DNA fragment immobilized on a solid support with both a test compound and wild-type p53 protein to bind the wild-type p53 protein to the DNA fragment;

determining the amount of wild-type p53 protein which is bound to the DNA fragment, inhibition of binding of wild-type p53 protein by the test compound indicating binding of the test compound to the p53-specific DNA binding sequences.

31. A method of supplying wild-type p53 gene function to a cell which has lost said gene function by virtue of mutation in a p53 gene, comprising:

introducing a wild-type p53 gene into a cell which has lost said gene function such that said gene is expressed in the cell.

32. The method of claim 31 wherein said wild-type p53 gene is expressed to a level higher than any mutant p53 gene present in the cell.

33. The method of claim 31 wherein the introduced wild-type p53 gene recombines with the endogenous mutant p53 gene present in the cell by a double recombination event to correct the p53 gene mutation.

34. A method of supplying a wild-type p53 gene function to a cell which has lost said gene function by virtue of a mutation in a p53 gene, comprising:

introducing a portion of a wild-type p53 gene into a cell which has lost said gene function such that said portion is expressed in the cell, said portion encoding a part of the p53 protein which is required for non-neoplastic growth of said cell.

35. A method of pre-screening agents for use in cancer therapy, comprising:

GeneDX 1009, pg. 532

measuring the amount of binding of a p53 protein encoded by a mutant gene found in cancer cells of a patient to a DNA molecule which conforms to the consensus binding site having more than one monomer of RRRCWWGYYY;

measuring the amount of binding of said p53 protein to said DNA molecule in the presence of a test substance; and

comparing the amount of binding of the p53 protein in the presence of said test substance to the amount of binding of the p53 protein in the absence of said test substance, a test substance which increases the amount of binding being a candidate for use in cancer therapy.

36. A method of pre-screening agents for use in cancer therapy, comprising:

contacting a transfected cell with a test substance, said transfected cell containing a p53 protein which is encoded by a mutant gene found in cancer cells of a patient and a reporter gene construct comprising a reporter gene which encodes an assayable product and a sequence which conforms to the p53 consensus binding site having more than one monomer of RRRCWWGYYY, wherein said sequence is upstream from and adjacent to said reporter gene; and

determining whether the amount of expression of said reporter gene is altered by the test substance, a test substance which alters the amount of expression of said reporter gene being a candidate for use in cancer therapy.

37. A method of pre-screening agents for use in cancer therapy, comprising:

adding RNA polymerase and ribonucleotides to a transcription construct, said transcription construct comprising a reporter gene which encodes an assayable product and a sequence which conforms to the p53 consensus binding site having at least two monomers of RRRCWWGYYY, said sequence being upstream from and adjacent to said reporter gene, said step of adding being effected in the presence and absence of a test substance; and

determining whether the amount of transcription of said reporter gene is altered by the presence of said test substance, a test substance which alters the amount of transcription of said reporter gene being a candidate for use in cancer therapy.

38. A DNA construct for use in screening potential chemotherapeutic agents, comprising:

a reporter gene which encodes an assayable product;

a sequence which conforms to the p53 consensus binding site having more than one monomer of RRRCWWGYYY upstream from and adjacent to said reporter gene, wherein said DNA construct is selected from the group consisting of a recombinant plasmid, a viral vector or an isolated molecule of DNA.

39. A method of diagnosing tumor-inducing or hyperplastic-inducing strains of human papilloma virus (HPV) comprising:

contacting cells or cell extracts of patients suspected of being infected by HPV with a p53-specific binding DNA fragment;

detecting the amount of wild-type p53 in said cells or cell extract which binds to said DNA fragment, absence of bound p53 indicating infection by strains of HPV which sequester p53.

40. A method of pre-screening agents for use in cancer therapy, comprising:

contacting a transfected cell with a test substance, said transfected cell containing a wild-type p53 protein and a reporter gene construct comprising a reporter gene which encodes an assayable product and a sequence which conforms to the p53 consensus binding site having more than one monomer of RRRCWWGYYY, wherein said sequence is upstream from and adjacent to said reporter gene; and

determining whether the amount of expression of said reporter gene is altered by the test substance, a test substance which alters the amount of expression of said reporter gene being a candidate for use in cancer therapy.

41. A method of pre-screening oligonucleotides for use in cancer therapy, comprising:

adding a p53 protein which is encoded by a mutant gene found in a cancer patient and a preparation of random oligonucleotides to a p53-specific-binding DNA fragment immobilized on a solid support;

recovering the olignucleotides which bound to the solid support.

GeneDX 1009, pg. 533

FIG. IA

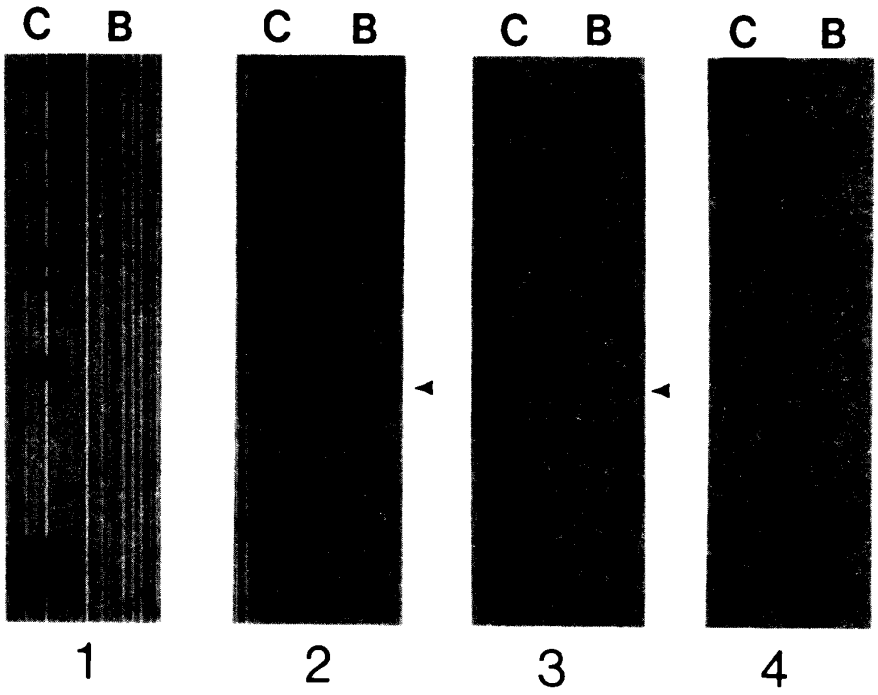
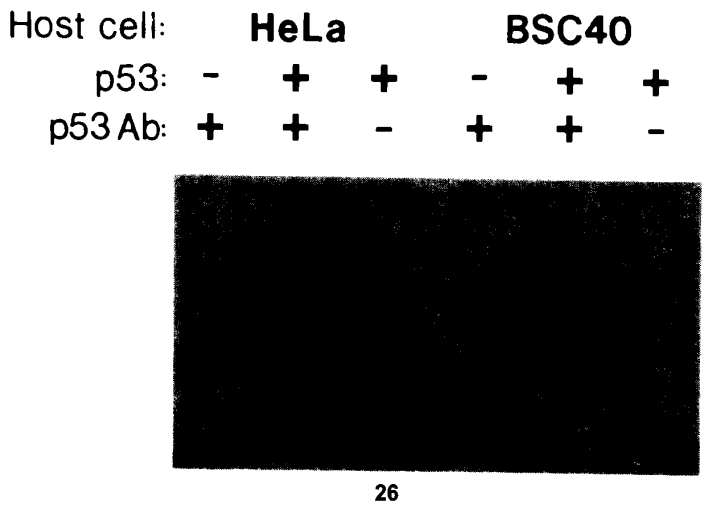


FIG. IB



GeneDX 1009, pg. 534

FIG. 2A

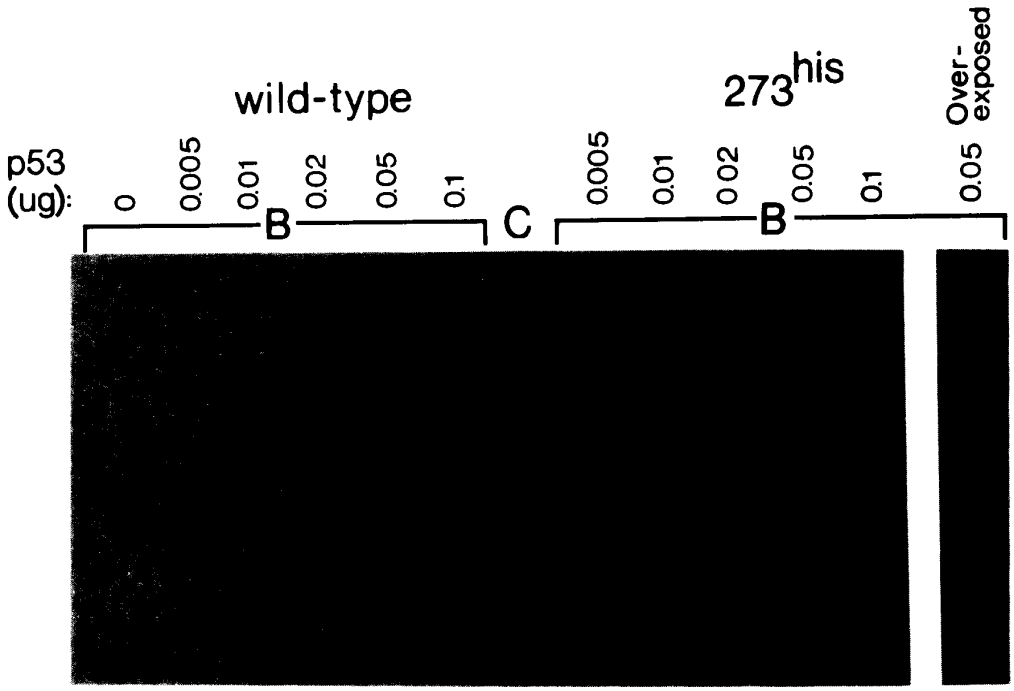
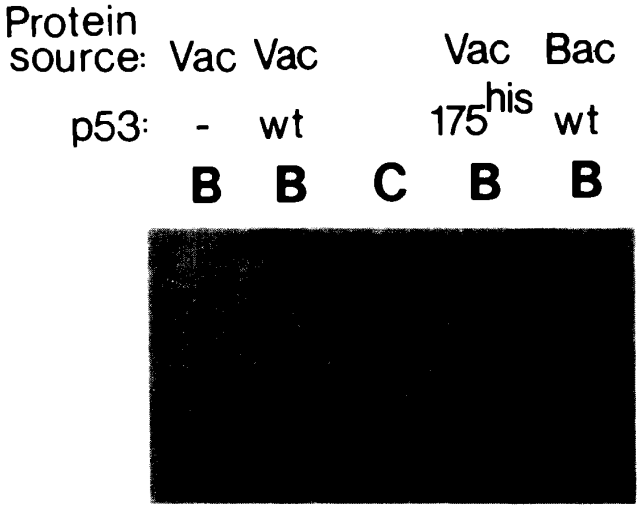
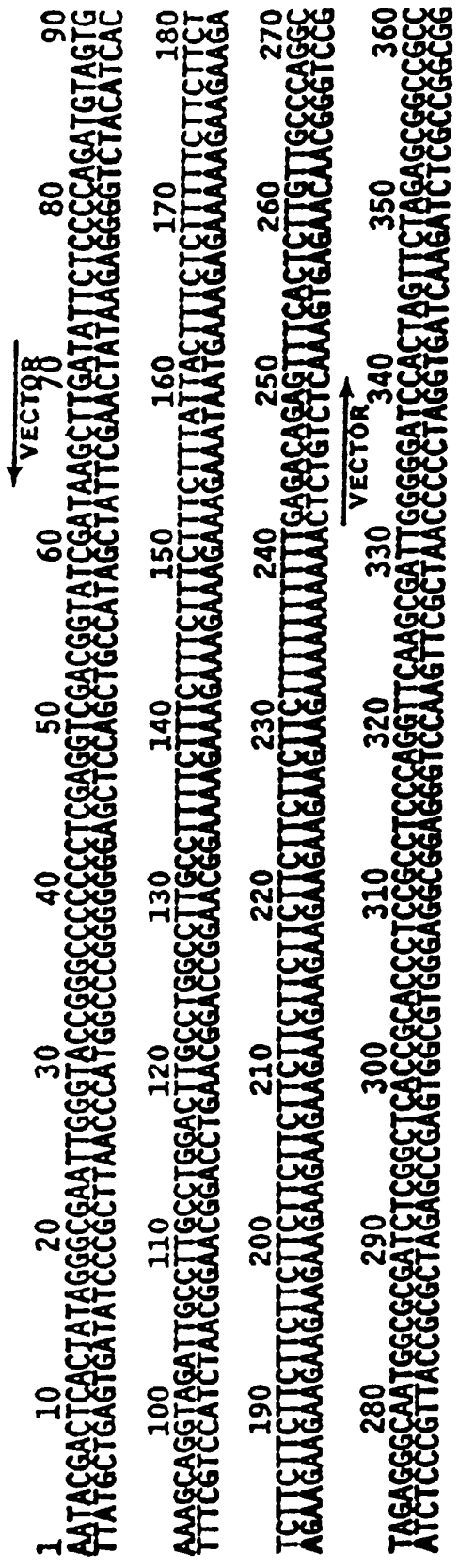


FIG. 2B



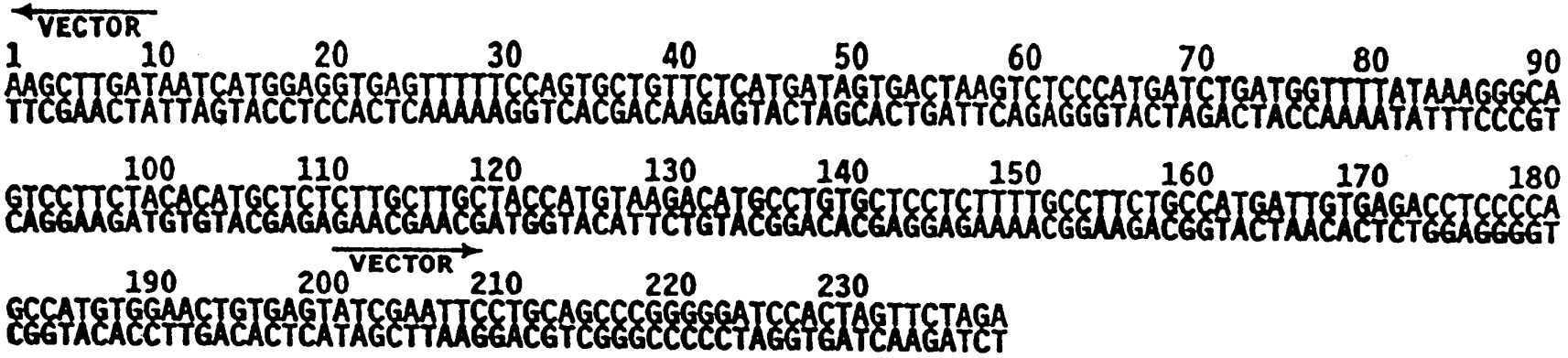
GeneDX 1009, pg. 535

FIG. 3A



GeneDX 1009, pg. 536

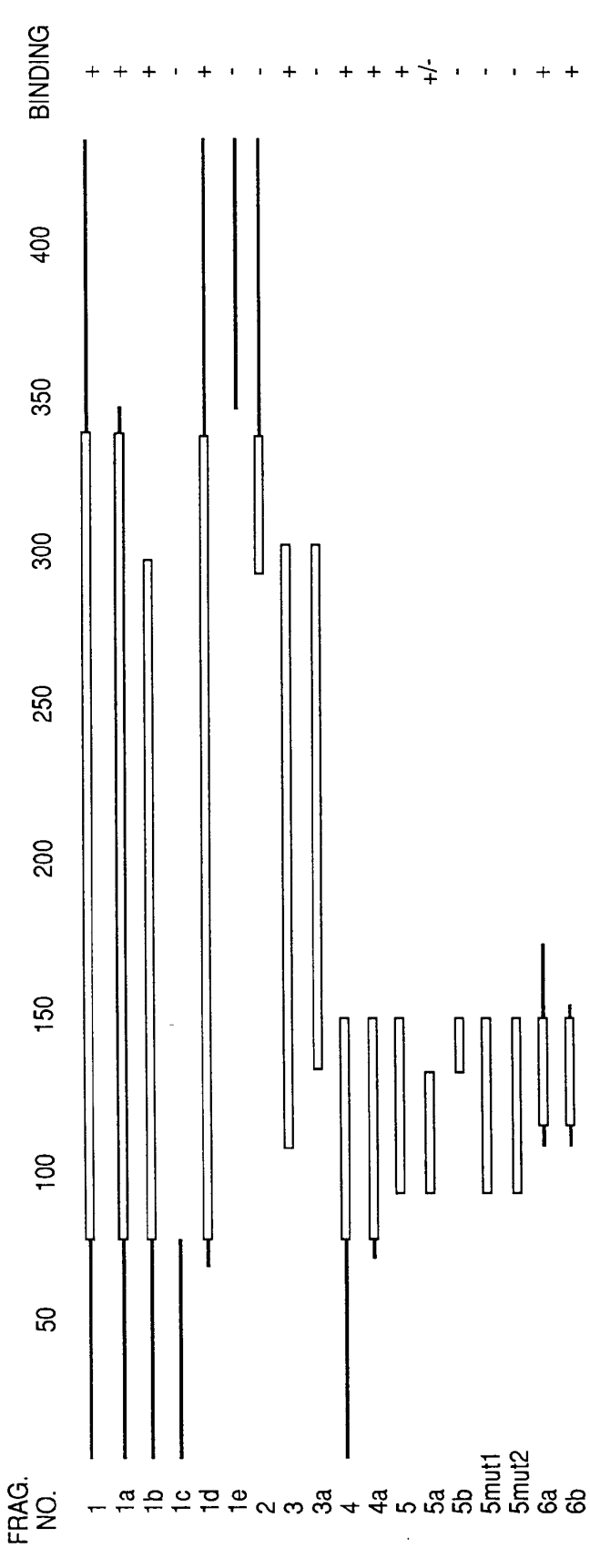
FIG. 3B



EP 0 518 650 A2

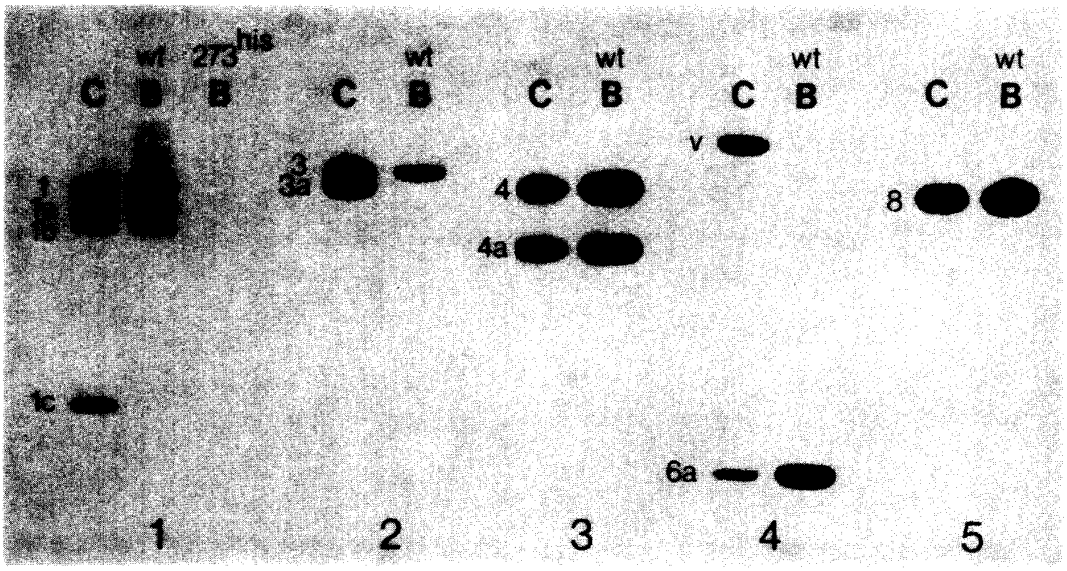
GeneDX 1009, pg. 537

FIG. 4A



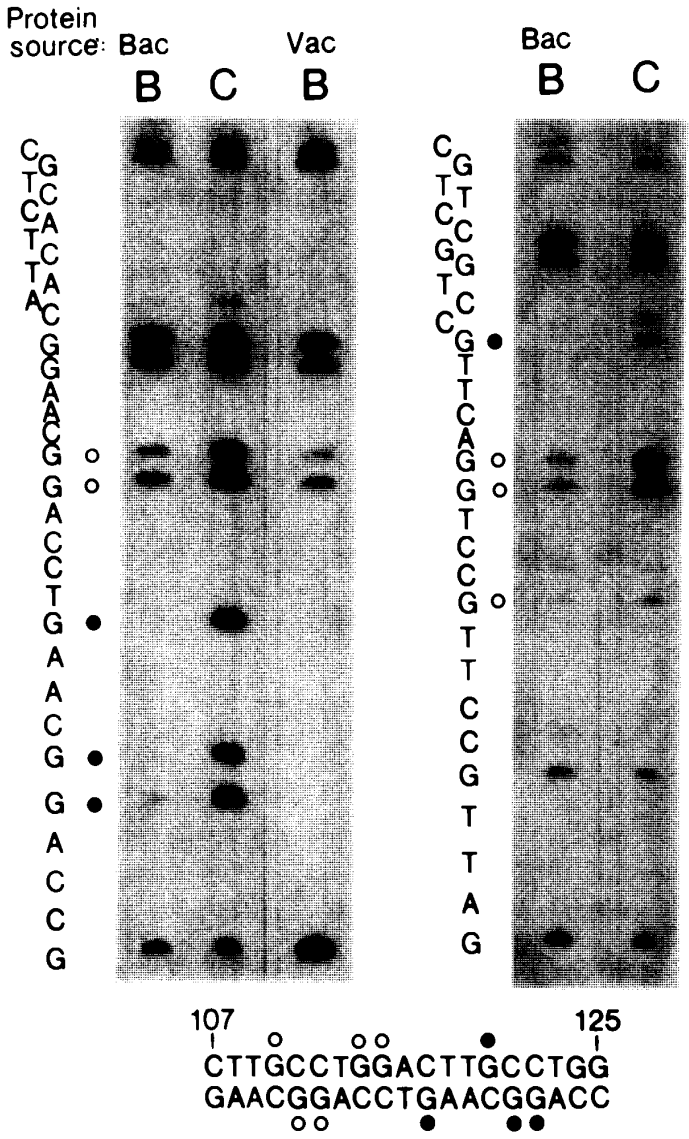
GeneDX 1009, pg. 538

FIG. 4B



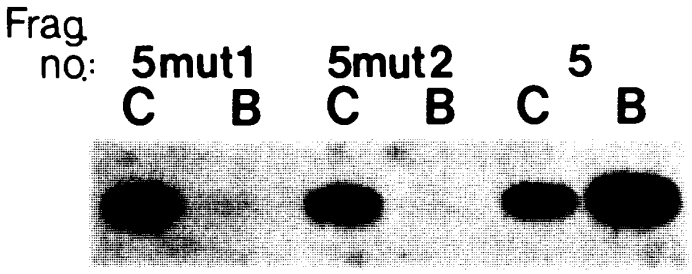
GeneDX 1009, pg. 539

FIG. 5A



GeneDX 1009, pg. 540

FIG. 5B



GeneDX 1009, pg. 541

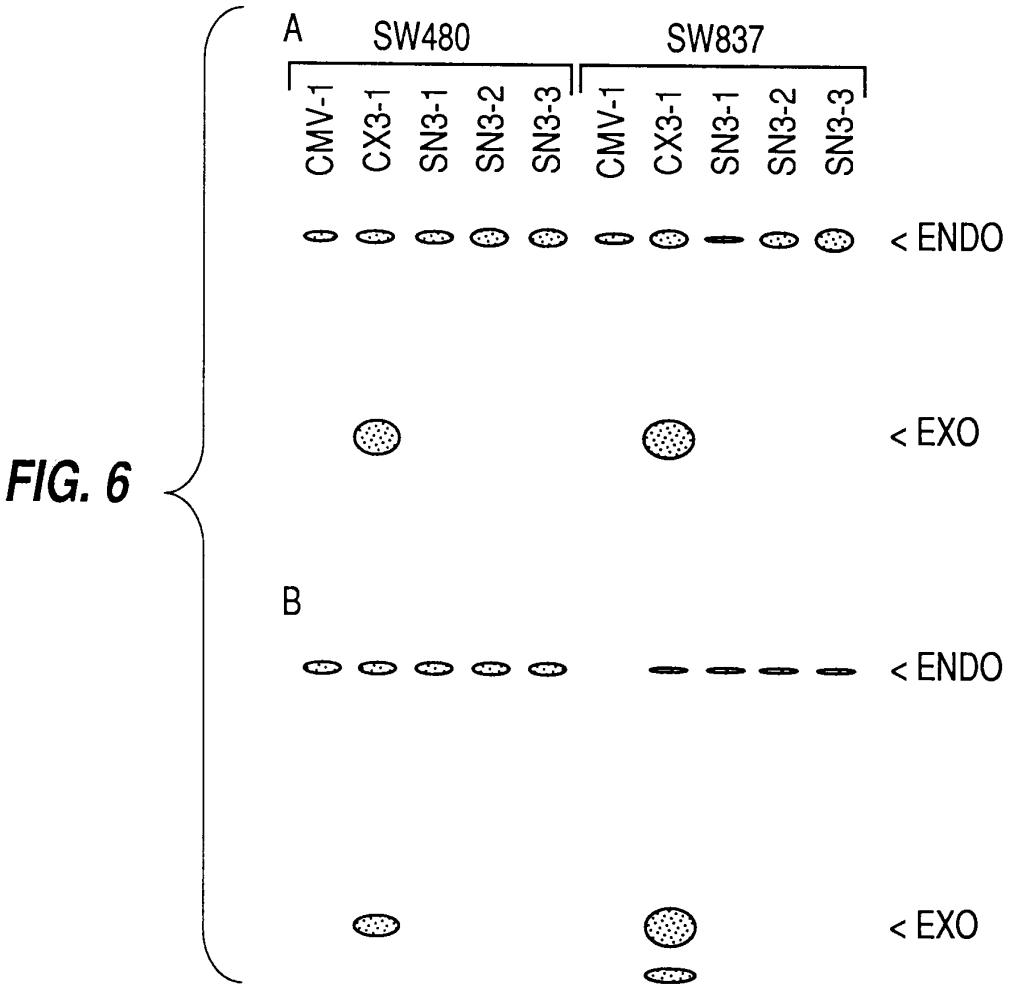


FIG. 6

GeneDX 1009, pg. 542

FIG. 7A

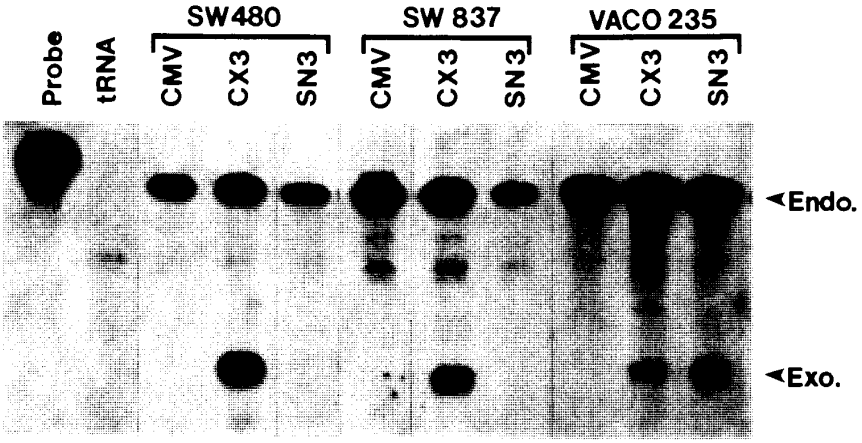
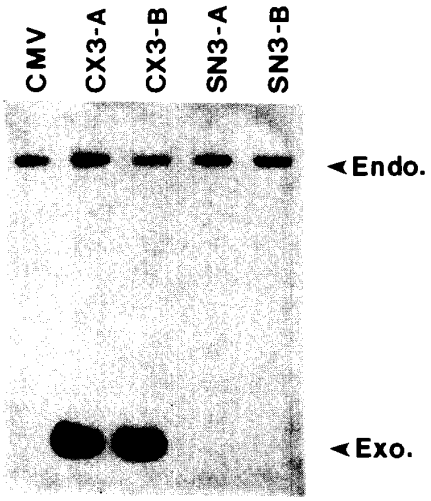
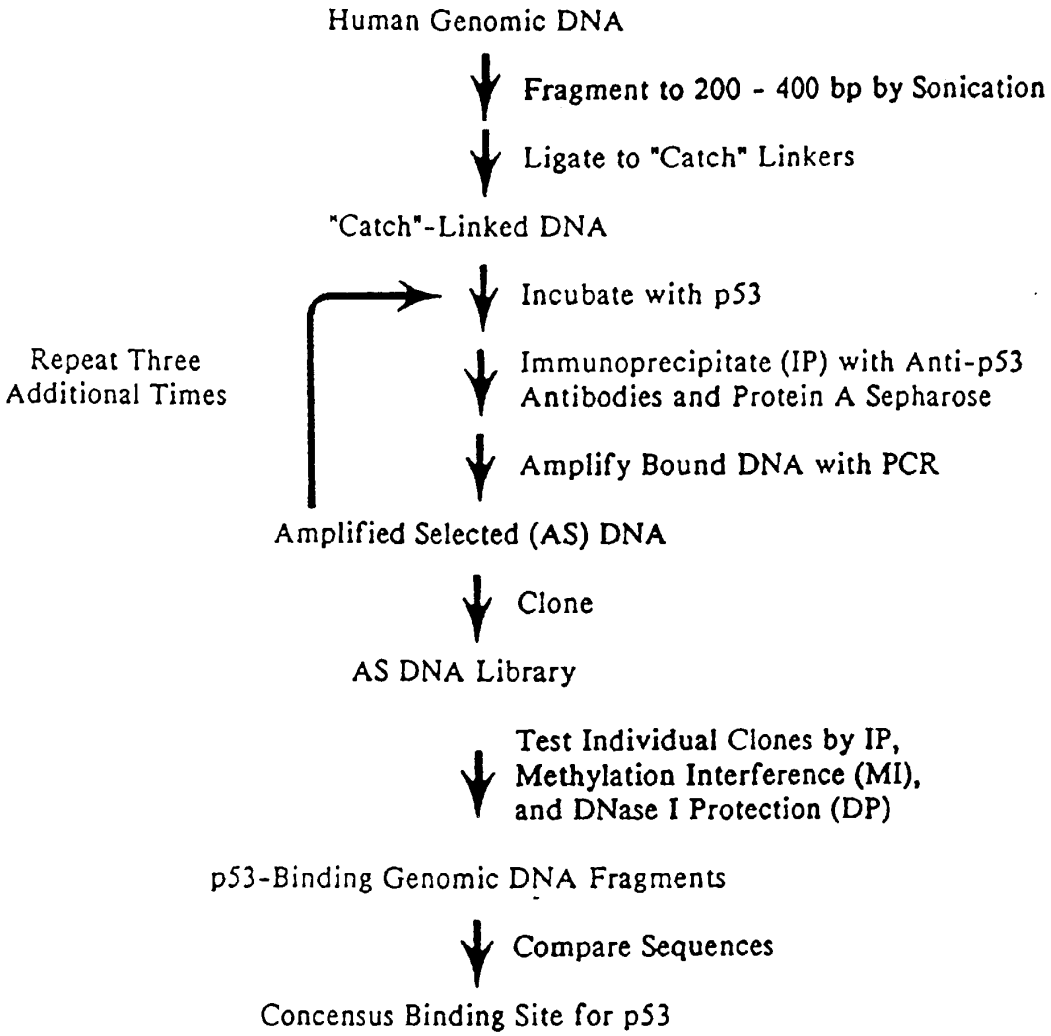


FIG. 7B



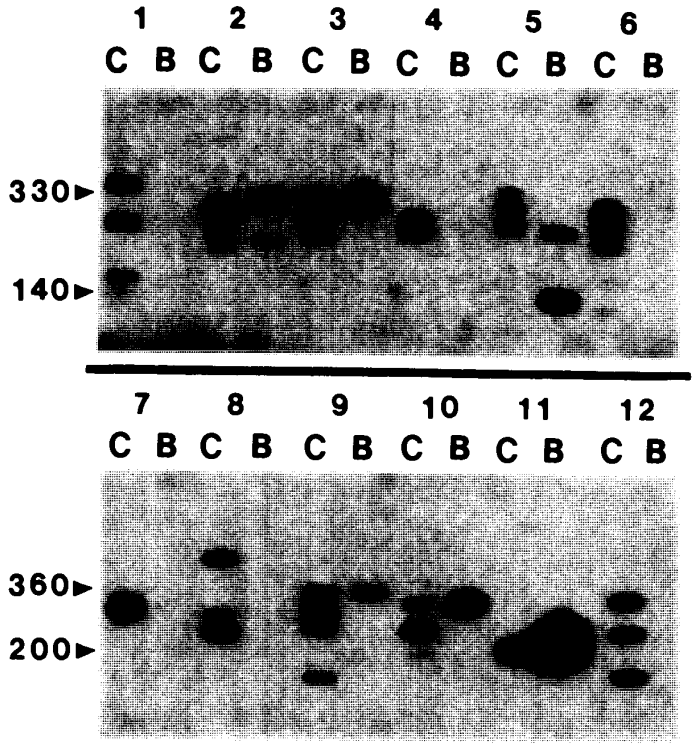
GeneDX 1009, pg. 543

FIG. 8A



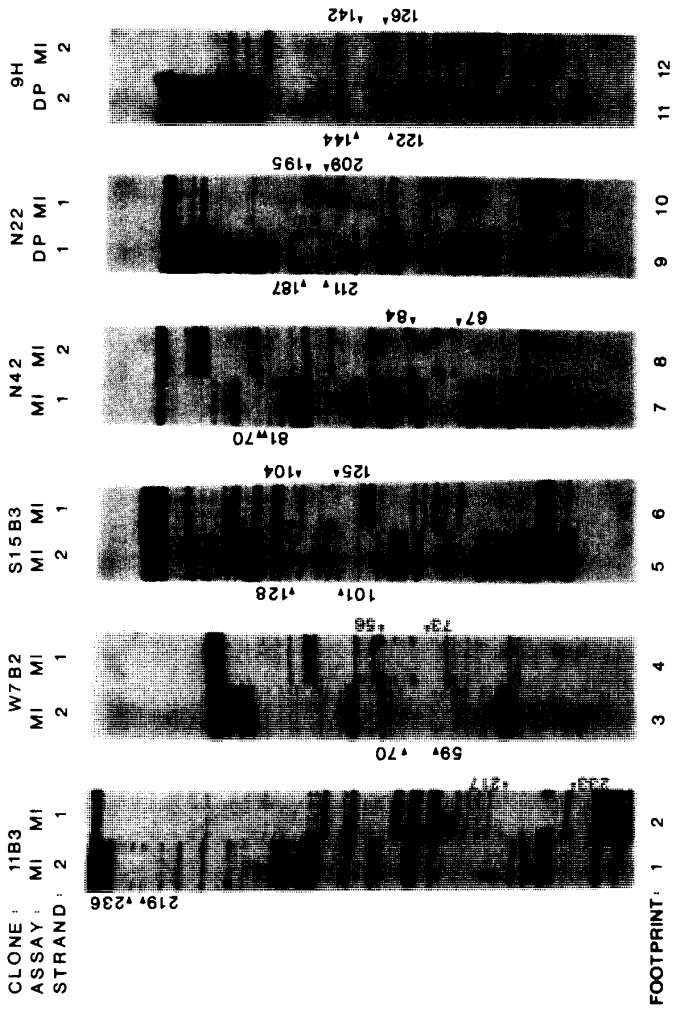
GeneDX 1009, pg. 544

FIG. 8B



GeneDX 1009, pg. 545

FIG. 9



GeneDX 1009, pg. 546

FIG. 10A

	CLONE	SIZE (bp)	5'-(bp)	nnnnnnnnnnnnnnn	RRRCWWGYYY	nnnnnnnnnnnnnn	RRRCWWGYYY	nnnnnnnnnnnnnn	3'-(bp)
1.	S57	295	144	cgacctgtcacaccg	GGGCcTGTCa		cAGCATGaCCT	acctgtcacaccggg	194
2.	N22	357	178	atttcaccatgctt	ctGCATGTCT		AGGCAAGTCa	cttctccactggcc	227
3.	11A2	387	317	ccccatcctccactg	AAACAaTGCCCC		AGACTTGTCT	ctccggcctgaatga	367
4.	W211	249	119	tttgtcctaccatcc	AGGCATGCCT		TTGCCT	cactcgttatttcct	164
5.	W7B2	139	41	tatctgtgcagctgt	GGGCATGTTT t		AGGCAAGCTT	cctgtgctagttccc	91
6.	3H	126	50	aactagatcctttttc	AGACATGTTa		tAACAAGTCa	gtacaagttttattt	99
7.	8A	483	445	gctggtgcacaagag	tGACATGTCC		cGACgTGTTT	tgtc	483
8.	532	335	229	catcatgccacctgc	AGGCATGTTC tggat		GGGC TGTCT	tgtgctttgttgttt	282
9.	64A2	349	120	caaaccagggtgtct	tGACTTGCCCT atcctgggaggt		tGACATGTTC	ctcccctttcccctc	181
10.	W7A1	264	124	gccaaacataaccac	cAGC TGCCCa		AGGCATGCa g	taccacgctcagccc	173
11.	S61	202	1	c	cAACTTGTCT attctgtgttgat		GGACATGTTC	ccgttttggctatt	49
12.	11B3	248	201	actgttgatgatgaa	AGACAAGCCTT a		GGGCAgGTCC	tgggggtgggg	248
13.	N42	248	49	gcagtgtggtggagg	AAACAAGCCCC a		GGAtgTGCCC	agggcaggctgggac	99
14.	S201	326	164	tgttcataccctgtcc	AcACTTGTCT		AtACCcTGCCT	acacctgtcttgttt	214
15.	S15B3	248	83	ctttaattcagttgt	AAACATGaCTT gttcattata		tGACATGTTC	aattacaattcgatt	143
16.	S592I	254	39	ctcagttctcagctg	GGACTTGCCC		tGGCCcAGCCC	tggggtcactgctgc	88
17.	S592II	254	130	tgcctcagcacctttc	AGGtTcTGCCC		GGGCTTGTTC	ctttcctttcagcat	179
18.	2HB	470	42	gccttttgttgtgccc	tGACTTGCCCC		AGACATGTTT	gggaatgtcttgtgc	91
19.	9H	467	108	gtattctctttttcct	AAGCATGCCT		tGACTTGTTC	tttcatctcctctga	157
20.	CBE10d	425	89	tgaaagcaggtagat	tGccCTTGCCT		GGACTTGCCT	ggccttgccttttct	138

GeneDX 1009, pg. 547

FIG. 10B

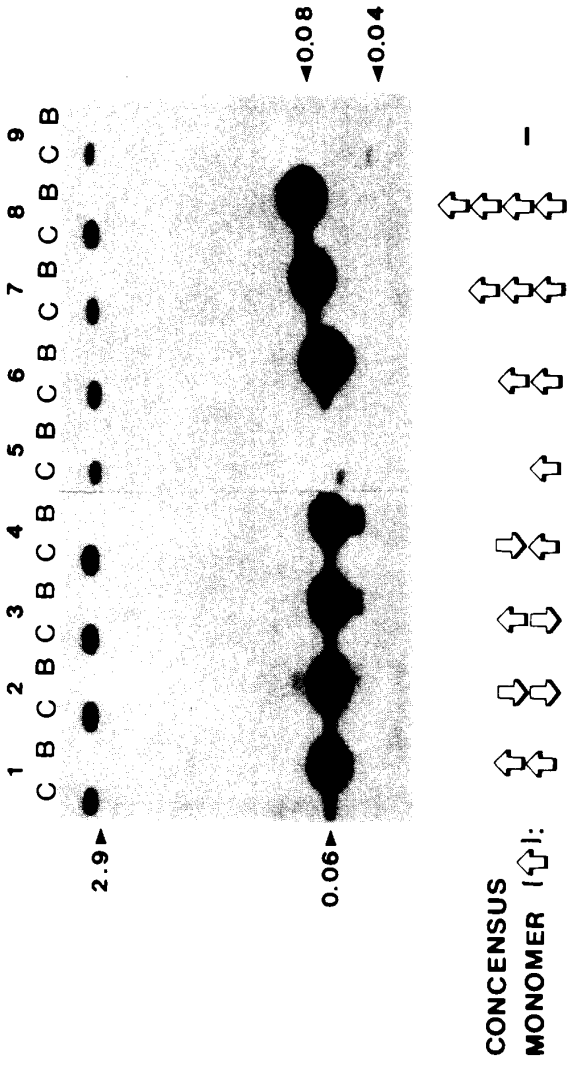
COMBINED NUCLEOTIDE USAGE (%) WITHIN THE TWO MONOMERS OF THE CONSENSUS BINDING SITE:

	5'- R	R	R	C	W	W	G	Y	Y	Y	-3'
A	40	20	55	0	53	15	0	0	0	12	A
C	13	3	3	93	8	0	0	50	68	35	C
G	23	70	40	0	8	3	100	0	0	3	G
T	23	5	0	5	30	82	0	50	30	48	T

SYNTHETIC OLIGONUCLEOTIDES: NO.	p53 BINDING				
1.	-		AGG aATt CCT		
2.	-		AGG aATt CCT	AGGaATtCCT	
3.	-		AGGCATGTCT		
4.	+		AGGCATGCCT	AGGCATGCCT	
5.	-		AGGCAAGgCa	AGGCAAGgCa	
6.	+	tgcaggaattcgat	AGGCATGTCT	AGGCATGTCT	atcaagcttatcgat
7.	+	tgcaggaattcgat	AGGCATGTCT	AGACATGCCT	atcaagcttatcgat
8.	+	tgcaggaattcgat	AGACATGCCT	AGGCATGTCT	atcaagcttatcgat
9.	-	tgcaggaattcgat	AGGCATGTCT		atcaagcttatcgat
10.	-	tgcaggaattcgat			atcaagcttatcgat

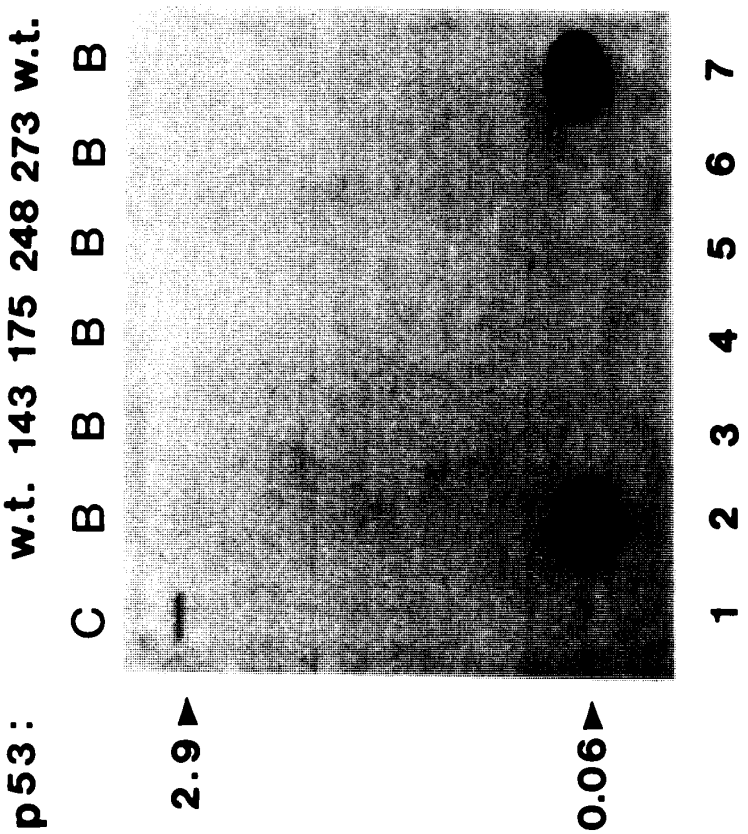
GeneDX 1009, pg. 548

FIG. IIA



GeneDX 1009, pg. 549

FIG. IIB



GeneDX 1009, pg. 550

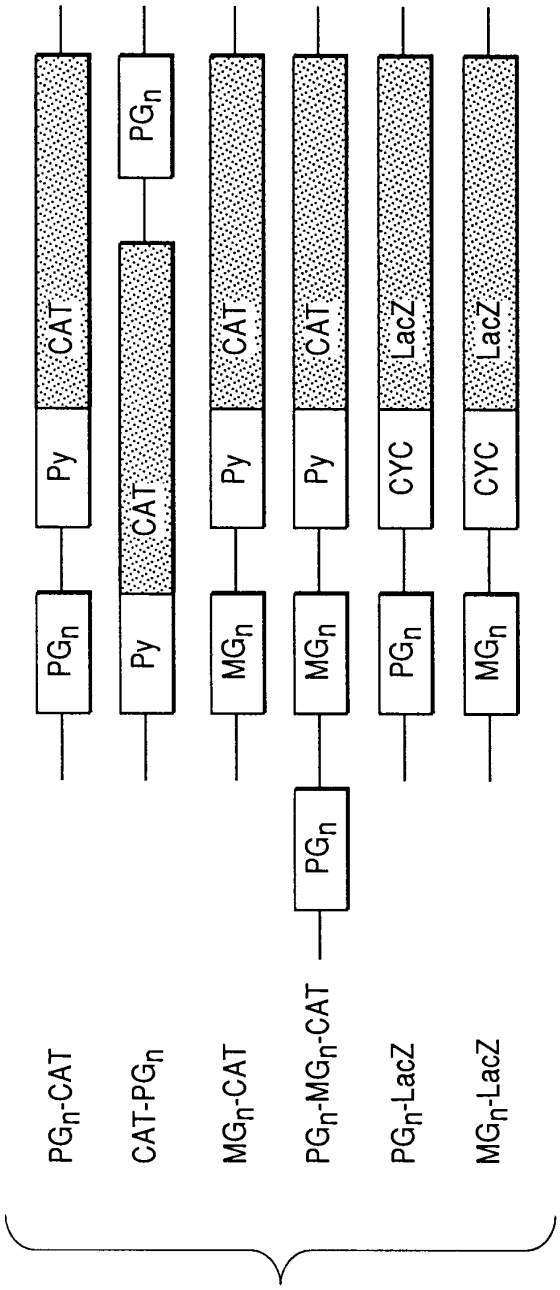
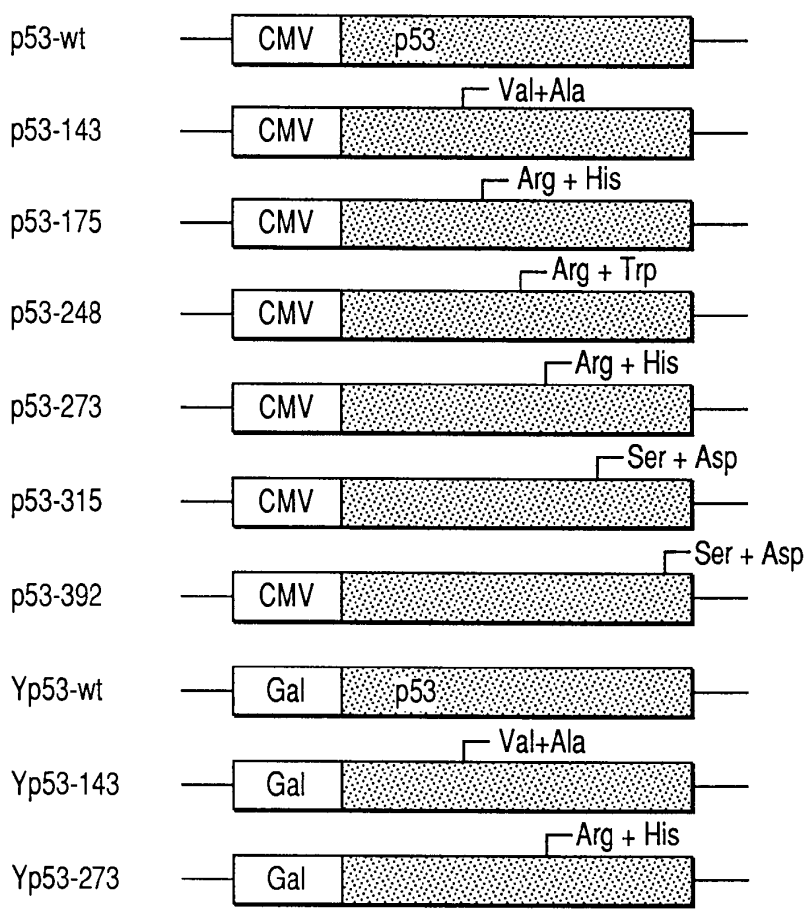


FIG. 12A

GeneDX 1009, pg. 551

FIG. 12B



GeneDX 1009, pg. 552

FIG. 13A

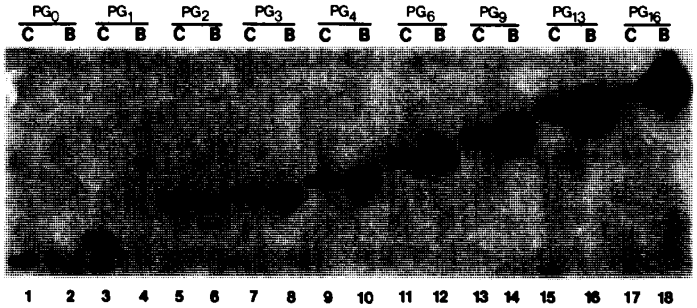
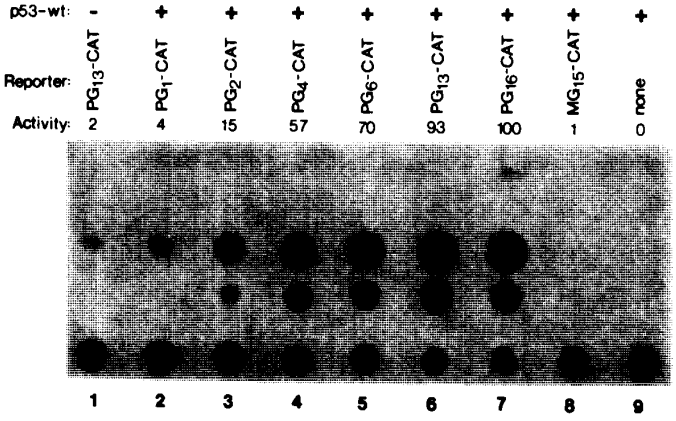
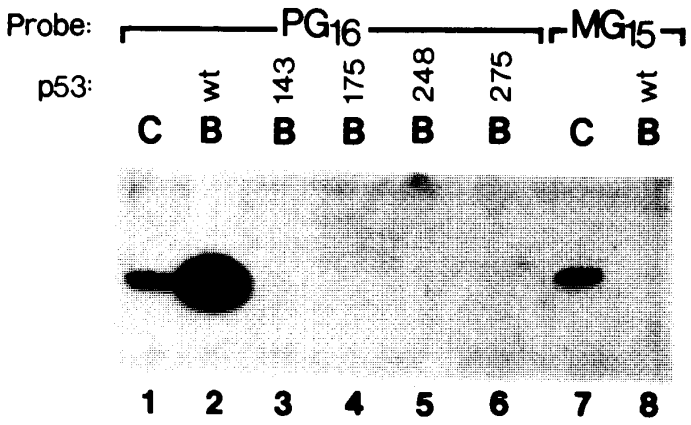


FIG. 13B



GeneDX 1009, pg. 553

FIG. 14



GeneDX 1009, pg. 554

FIG. 15A

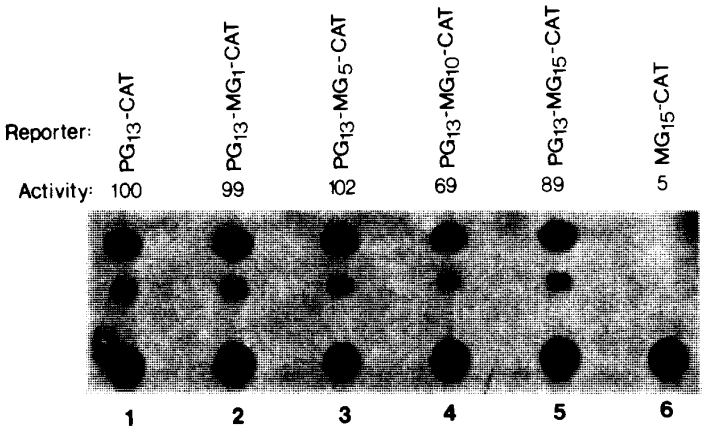
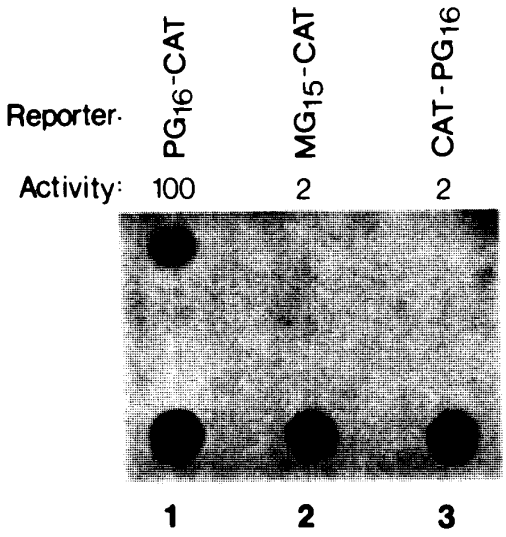


FIG. 15B



GeneDX 1009, pg. 555

FIG. 16A

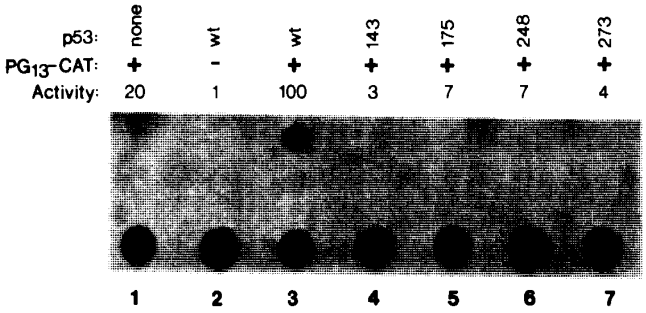
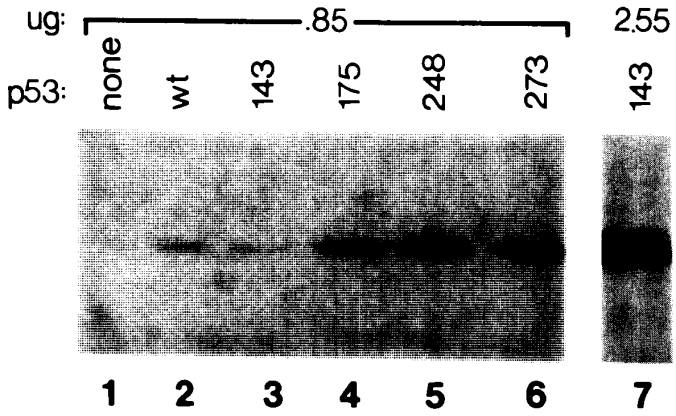
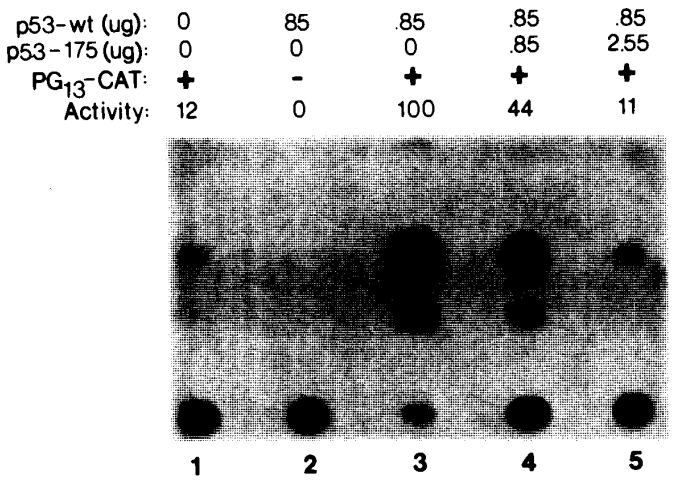


FIG. 16B



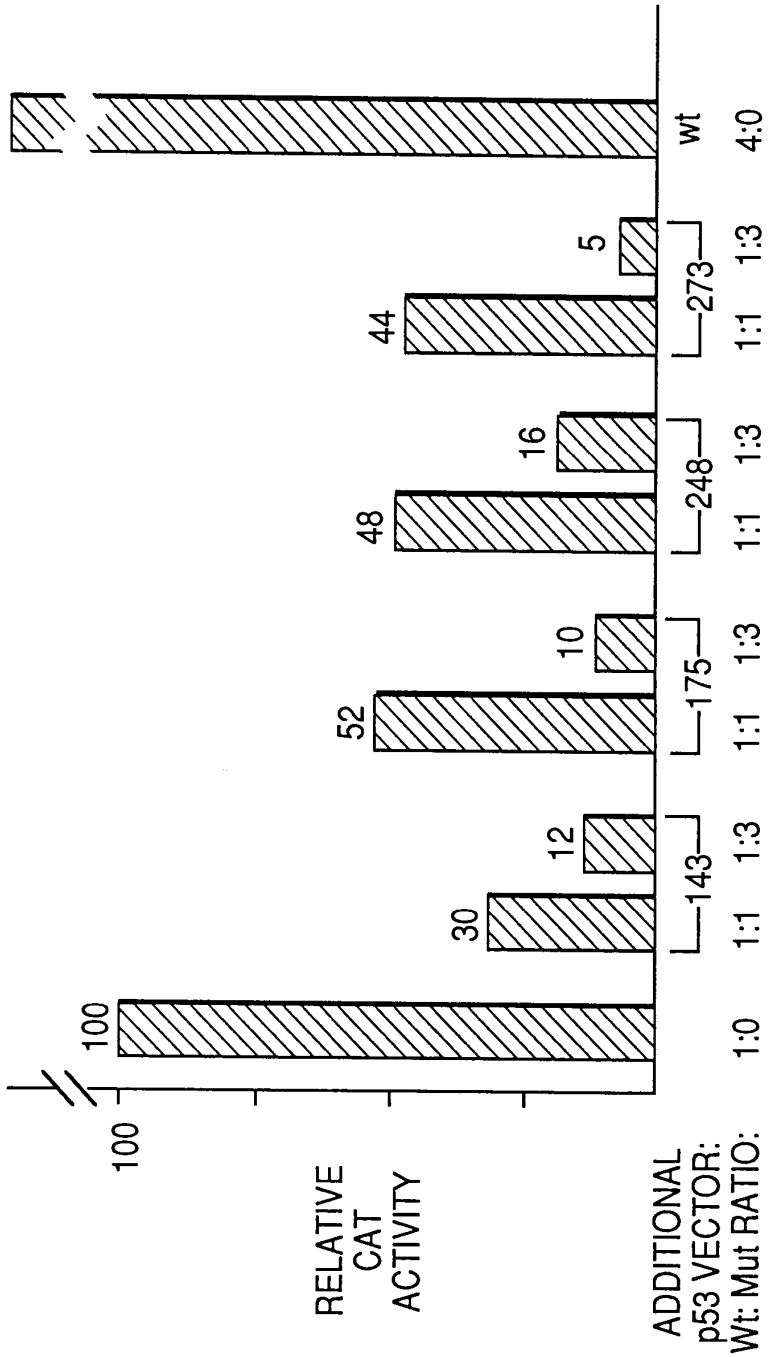
GeneDX 1009, pg. 556

FIG. 17



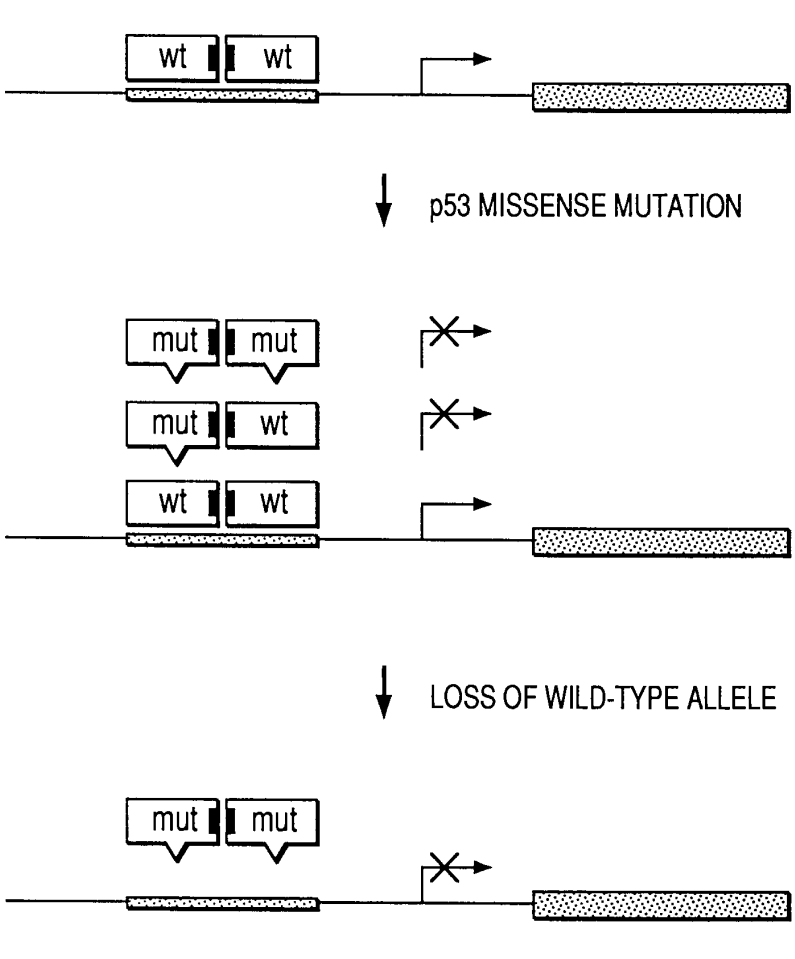
GeneDX 1009, pg. 557

FIG. 18



GeneDX 1009, pg. 558

FIG. 19



GeneDX 1009, pg. 559

US005236844A

United States Patent [19]

Basset et al.

[11] **Patent Number: 5,236,844**

[45] **Date of Patent: Aug. 17, 1993**

[54] **ANALYTICAL MARKERS FOR MALIGNANT BREAST CANCER**

[75] Inventors: **Paul Basset; Jean-Pierre Bellocq,** both of Strasbourg; **Pierre Chambon,** Blaesheim, all of France

[73] Assignees: **Institut Natl. de la Sante et de la Recherche Medicale; Centre National de la Rescherche Scientifiquue; Universite Louis Pasteur Strasbourg,** France

[21] Appl. No.: **794,393**

[22] Filed: **Nov. 21, 1991**

[30] **Foreign Application Priority Data**

Nov. 21, 1990 [GB] United Kingdom 9025326

[51] **Int. Cl.**[5] **C12N 15/70;** C07H 21/04

[52] **U.S. Cl.** **435/320.1;** 536/23.5

[58] **Field of Search** 435/6, 320.1; 536/27, 536/23.5; 530/350

[56] **References Cited**

PUBLICATIONS

Chen et al (1990, Sep.) Oncoqche 5(9) 1391-5.

Davis et al (1987) P.N.A.S. 81, 2194-2198.

Salomon et al., in *Breast Cancer: Cellular and Molecular Biology (eds., Lippman, M. E. and Dickson, R. B.), pp. 363-389 (Kluwer, Boston, (1988).*

Primary Examiner—Margaret Moskowitz

Assistant Examiner—Lisa Bennett

Attorney, Agent, or Firm—Sterne, Kessler, Goldstein & Fox

[57] **ABSTRACT**

The present invention relates to a gene encoding a new member of the metalloproteinase family which has been found to be specifically associated with invasive breast cancer, and to methods of diagnosis for such cancer comprising detection of the marker or its nucleotide sequence, and to treatment or prophylaxis by inhibiting, altering the activity of or binding the marker, or by interfering with its synthesis.

8 Claims, 18 Drawing Sheets

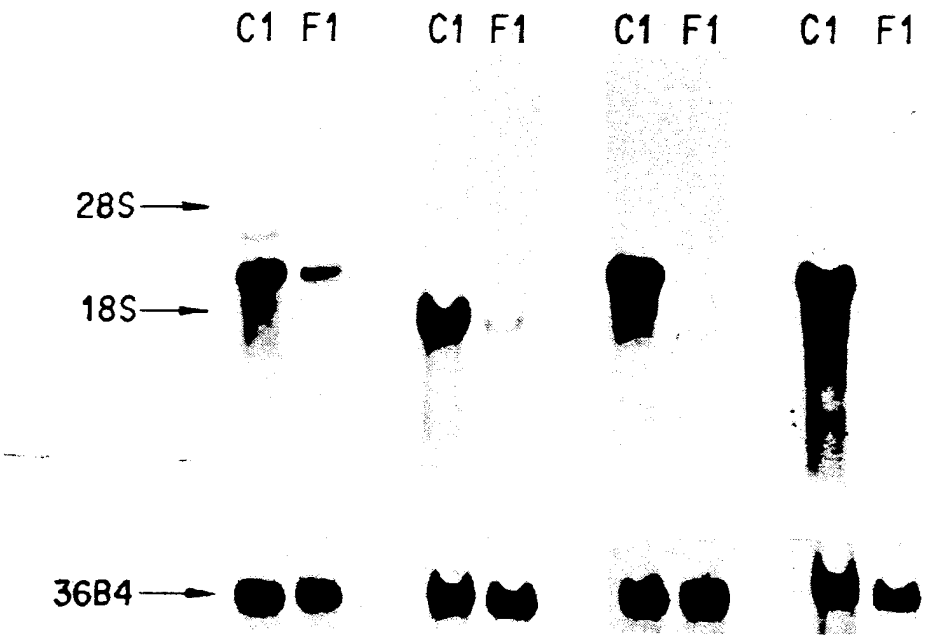


FIG. 1a FIG. 1b FIG. 1c FIG. 1d

```
CCGGGGCGG ATG GCT CCG GCC GCC TGG CTC CGC AGC GCG GCC GCG CGC          48
          Met Ala Pro Ala Ala Trp Leu Arg Ser Ala Ala Ala Arg
            1               5                  10

GCC CTC CTG CCC CCG ATG CTG CTG CTG CTG CTC CAG CCG CCG CCG CTG        96
Ala Leu Leu Pro Pro Met Leu Leu Leu Leu Leu Gln Pro Pro Pro Leu
     15                  20                  25

CTG GCC CGG GCT CTG CCG CCG GAC GTC CAC CAC CTC CAT GCC GAG AGG       144
Leu Ala↑Arg Ala↑Leu Pro Pro Asp Val His His Leu His Ala Glu Arg
 30                  35                  40                  45

AGG GGG CCA CAG CCC TGG CAT GCA GCC CTG CCC AGT AGC CCG GCA CCT       192
Arg Gly Pro Gln Pro Trp His Ala Ala Leu Pro Ser Ser Pro Ala Pro
                   50                  55                  60

GCC CCT GCC ACG CAG GAA GCC CCC CGG CCT GCC AGC AGC CTC AGG CCT       240
Ala Pro Ala Thr Gln Glu Ala Pro Arg Pro Ala Ser Ser Leu Arg Pro
               65                  70                  75

CCC CGC TGT GGC GTG CCC GAC CCA TCT GAT GGG CTG AGT GCC CGC AAC       288
Pro Arg Cys Gly Val Pro Asp Pro Ser Asp Gly Leu Ser Ala Arg Asn
          80                  85                  90

CGA CAG AAG AGG TTC GTG CTT TCT GGC GGG CGC TGG GAG AAG ACG GAC       336
Arg Gln Lys Arg Phe Val Leu Ser Gly Gly Arg Trp Glu Lys Thr Asp
     95                  100                 105
```

FIG. 2A

GeneDX 1009, pg. 562

```
CTC ACC TAC AGG ATC CTT CGG TTC CCA TGG CAG TTG GTG CAG GAG CAG        384
Leu Thr Tyr Arg Ile Leu Arg Phe Pro Trp Gln Leu Val Gln Glu Gln
110                 115                 120                 125

GTG CGG CAG ACG ATG GCA GAG GCC CTA AAG GTA TGG AGC GAT GTG ACG        432
Val Arg Gln Thr Met Ala Glu Ala Leu Lys Val Trp Ser Asp Val Thr
                130                 135                 140

CCA CTC ACC TTT ACT GAG GTG CAC GAG GGC CGT GCT GAC ATC ATG ATC        480
Pro Leu Thr Phe Thr Glu Val His Glu Gly Arg Ala Asp Ile Met Ile
            145                 150                 155

GAC TTC GCC AGG TAC TGG CAT GGG GAC GAC CTG CCG TTT GAT GGG CCT        528
Asp Phe Ala Arg Tyr Trp His Gly Asp Asp Leu Pro Phe Asp Gly Pro
        160                 165                 170

GGG GGC ATC CTG GCC CAT GCC TTC TTC CCC AAG ACT CAC CGA GAA GGG        576
Gly Gly Ile Leu Ala His Ala Phe Phe Pro Lys Thr His Arg Glu Gly
    175                 180                 185

GAT GTC CAC TTC GAC TAT GAT GAG ACC TGG ACT ATC GGG GAT GAC CAG        624
Asp Val His Phe Asp Tyr Asp Glu Thr Trp Thr Ile Gly Asp Asp Gln
190                 195                 200                 205

GGC ACA GAC CTG CTG CAG GTG GCA GCC CAT GAA TTT GGC CAC GTG CTG        672
Gly Thr Asp Leu Leu Gln Val Ala Ala His Glu Phe Gly His Val Leu
                210                 215                 220
```

FIG.2B

```
GGG CTG CAG CAC ACA ACA GCA GCC AAG GCC CTG ATG TCC GCC TTC TAC          720
Gly Leu Gln His Thr Thr Ala Ala Lys Ala Leu Met Ser Ala Phe Tyr
            225                 230                 235

ACC TTT CGC TAC CCA CTG AGT CTC AGC CCA GAT GAC TGC AGG GGC GTT          768
Thr Phe Arg Tyr Pro Leu Ser Leu Ser Pro Asp Asp Cys Arg Gly Val
        240                 245                 250

CAA CAC CTA TAT GGC CAG CCC TGG CCC ACT GTC ACC TCC AGG ACC CCA          816
Gln His Leu Tyr Gly Gln Pro Trp Pro Thr Val Thr Ser Arg Thr Pro
    255                 260                 265

GCC CTG GGC CCC CAG GCT GGG ATA GAC ACC AAT GAG ATT GCA CCG CTG          864
Ala Leu Gly Pro Gln Ala Gly Ile Asp Thr Asn Glu Ile Ala Pro Leu
270                 275                 280                 285

GAG CCA GAC GCC CCG CCA GAT GCC TGT GAG GCC TCC TTT GAC GCG GTC          912
Glu Pro Asp Ala Pro Pro Asp Ala Cys Glu Ala Ser Phe Asp Ala Val
                290                 295                 300

TCC ACC ATC CGA GGC GAG CTC TTT TTC TTC AAA GCG GGC TTT GTG TGG          960
Ser Thr Ile Arg Gly Glu Leu Phe Phe Phe Lys Ala Gly Phe Val Trp
            305                 310                 315

CGC CTC CGT GGG GGC CAG CTG CAG CCC GGC TAC CCA GCA TTG GCC TCT         1008
Arg Leu Arg Gly Gly Gln Leu Gln Pro Gly Tyr Pro Ala Leu Ala Ser
        320                 325                 330
```

FIG. 2C

GeneDX 1009, pg. 564

```
CGC CAC TGG CAG GGA CTG CCC AGC CCT GTG GAC GCT GCC TTC GAG GAT      1056
Arg His Trp Gln Gly Leu Pro Ser Pro Val Asp Ala Ala Phe Glu Asp
    335                 340                 345

GCC CAG GGC CAC ATT TGG TTC TTC CAA GGT GCT CAG TAC TGG GTG TAC      1104
Ala Gln Gly His Ile Trp Phe Phe Gln Gly Ala Gln Tyr Trp Val Tyr
350                 355                 360                 365

GAC GGT GAA AAG CCA GTC CTG GGC CCC GCA CCC CTC ACC GAG CTG GGC      1152
Asp Gly Glu Lys Pro Val Leu Gly Pro Ala Pro Leu Thr Glu Leu Gly
                370                 375                 380

CTG GTG AGG TTC CCG GTC CAT GCT GCC TTG GTC TGG GGT CCC GAG AAG      1200
Leu Val Arg Phe Pro Val His Ala Ala Leu Val Trp Gly Pro Glu Lys
            385                 390                 395

AAC AAG ATC TAC TTC TTC CGA GGC AGG GAC TAC TGG CGT TTC CAC CCC      1248
Asn Lys Ile Tyr Phe Phe Arg Gly Arg Asp Tyr Trp Arg Phe His Pro
        400                 405                 410

AGC ACC CGG CGT GTA GAC AGT CCC GTG CCC CGC AGG GCC ACT GAC TGG      1296
Ser Thr Arg Arg Val Asp Ser Pro Val Pro Arg Arg Ala Thr Asp Trp
    415                 420                 425

AGA GGG GTG CCC TCT GAG ATC GAC GCT GCC TTC CAG GAT GCT GAT GGC      1344
Arg Gly Val Pro Ser Glu Ile Asp Ala Ala Phe Gln Asp Ala Asp Gly
430                 435                 440                 445
```

FIG. 2D

```
TAT GCC TAC TTC CTG CGC GGC CGC CTC TAC TGG AAG TTT GAC CCT GTG      1392
Tyr Ala Tyr Phe Leu Arg Gly Arg Leu Tyr Trp Lys Phe Asp Pro Val
                450                 455                 460

AAG GTG AAG GCT CTG GAA GGC TTC CCC CGT CTC GTG GGT CCT GAC TTC      1440
Lys Val Lys Ala Leu Glu Gly Phe Pro Arg Leu Val Gly Pro Asp Phe
            465                 470                 475

TTT GGC TGT GCC GAG CCT GCC AAC ACT TTC CTC TGACCATGGC TTGGATGCCC    1493
Phe Gly Cys Ala Glu Pro Ala Asn Thr Phe Leu
        480                 485

TCAGGGGTGC TGACCCCTGC CAGGCCACGA ATATCAGGCT AGAGACCCAT GGCCATCTTT    1553

GTGGCTGTGG GCACCAGGCA TGGGACTGAG CCCATGTCTC CTGCAGGGGG ATGGGGTGGG    1613

GTACAACCAC CATGACAACT GCCGGGAGGG CCACGCAGGT CGTGGTCACC TGCCAGCGAC    1673

TGTCTCAGAC TGGGCAGGGA GGCTTTGGCA TGACTTAAGA GGAAGGGCAG TCTTGGGACC    1733

CGCTATGCAG GTCCTGGCAA ACCTGGCTGC CCTGTCTCAT CCCTGTCCCT CAGGGTAGCA    1793

CCATGGCAGG ACTGGGGGAA CTGGAGTGTC CTTGCTGTAT CCCTGTTGTG AGGTTCCTTC    1853

CAGGGGCTGG CACTGAAGCA AGGGTGCTGG GGCCCCATGG CCTTCAGCCC TGGCTGAGCA    1913

ACTGGGCTGT AGGGCAGGGC CACTTCCTGA GGTCAGGTCT TGGTAGGTGC CTGCATCTGT    1973
```

FIG. 2E

GeneDX 1009, pg. 566

```
CTGCCTTCTG GCTGACAATC CTGGAAATCT GTTCTCCAGA ATCCAGGCCA AAAAGTTCAC    2033
AGTCAAATGG GGAGGGGTAT TCTTCATGCA GGAGACCCCA GGCCCTGGAG GCTGCAACAT    2093
ACCTCAATCC TGTCCCAGGC CGGATCCTCC TGAAGCCCTT TTCGCAGCAC TGCTATCCTC    2153
CAAAGCCATT GTAAATGTGT GTACAGTGTG TATAAACCTT CTTCTTCTTT TTTTTTTTTA    2213
AACTGAGGAT TGTCATTAAA CACAGTTGTT TTCTAAAAAA AAA                      2256
```

FIG. 2F

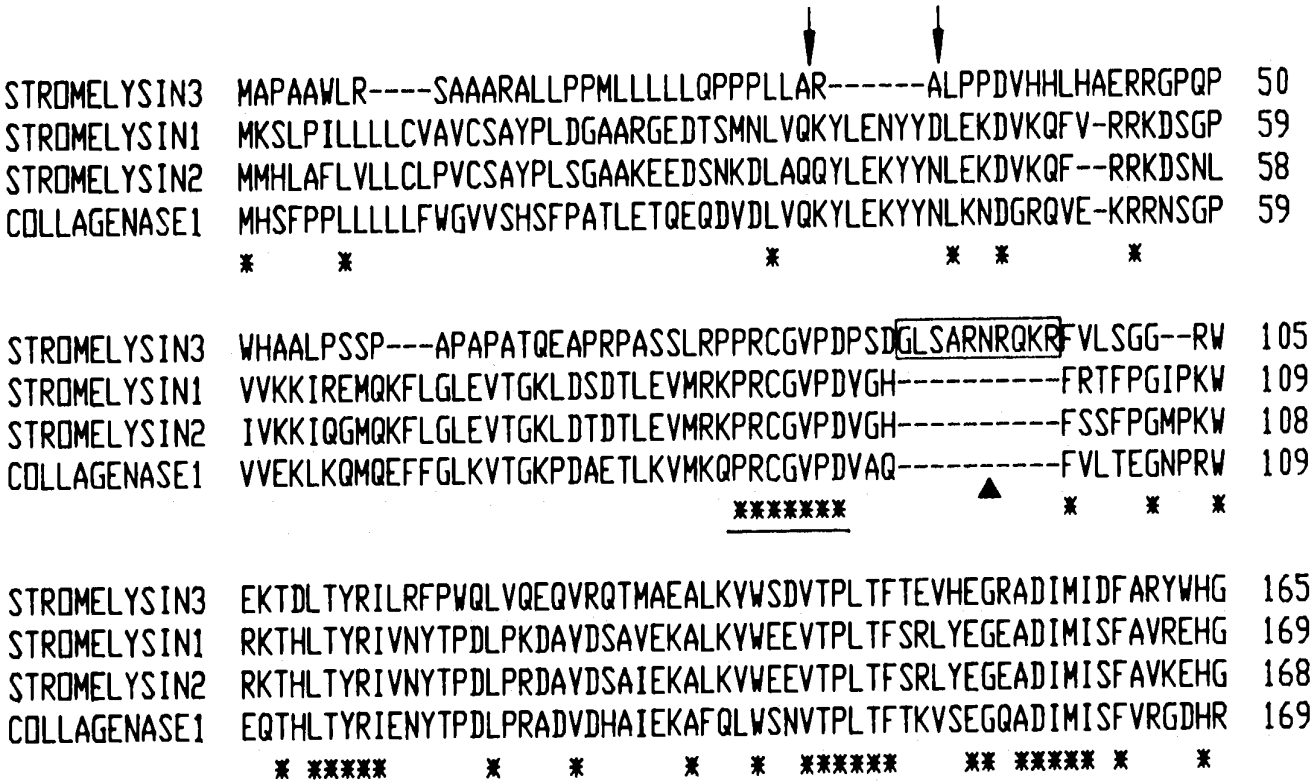

STROMELYSIN3 MAPAAWLR----SAAARALLPPMLLLLLQPPPLLAR------ALPPDVHHLHAERRGPQP 50
STROMELYSIN1 MKSLPILLLLCVAVCSAYPLDGAARGEDTSMNLVQKYLENYYDLEKDVKQFV-RRKDSGP 59
STROMELYSIN2 MMHLAFLVLLCLPVCSAYPLSGAAKEEDSNKDLAQQYLEKYYNLEKDVKQF--RRKDSNL 58
COLLAGENASE1 MHSFPPLLLLLFWGVVSHSFPATLETQEQDVDLVQKYLEKYYNLKNDGRQVE-KRRNSGP 59
STROMELYSIN3 WHAALPSSP---APAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVLSGG--RW 105
STROMELYSIN1 VVKKIREMQKFLGLEVTGKLDSDTLEVMRKPRCGVPDVGH----------FRTFPGIPKW 109
STROMELYSIN2 IVKKIQGMQKFLGLEVTGKLDTDTLEVMRKPRCGVPDVGH----------FSSFPGMPKW 108
COLLAGENASE1 VVEKLKQMQEFFGLKVTGKPDAETLKVMKQPRCGVPDVAQ----------FVLTEGNPRW 109
STROMELYSIN3 EKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRADIMIDFARYWHG 165
STROMELYSIN1 RKTHLTYRIVNYTPDLPKDAVDSAVEKALKVWEEVTPLTFSRLYEGEADIMISFAVREHG 169
STROMELYSIN2 RKTHLTYRIVNYTPDLPRDAVDSAIEKALKVWEEVTPLTFSRLYEGEADIMISFAVKEHG 168
COLLAGENASE1 EQTHLTYRIENYTPDLPRADVDHAIEKAFQLWSNVTPLTFTKVSEGQADIMISFVRGDHR 169

FIG.3A

```
STROMELYSIN3   DDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLGLQH  225
STROMELYSIN1   DFYPFDGPGNVLAHAYAPGPGINGDAHFDDDEQWT-KDTTGTNLFLVAAHEIGHSLGLFH  228
STROMELYSIN2   DFYSFDGPGHSLAHAYPPGPGLYGDIHFDDDEKWT-EDASGTNLFLVAAHELGHSLGLFH  227
COLLAGENASE1   DNSPFDGPGGNLAHAFQPGPGIGGDAHFDEHERWT-NNFTEYNLHRVAAHELGHSLGLSH  228
               *   ***** ****  *     ** ***  * **        *  ***** ** *** *

STROMELYSIN3   TTAAKALMSAFYTFRYPLS---LSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTNEI  282
STROMELYSIN1   SANTEALMYPLYHSLTDLTRFRLSQDDINGIQSLYGPP---PDSPETPLVPTEPV-----  280
STROMELYSIN2   SANTEALMYPLYNSFTELAQFRLSQDDVNGIQSLYGPP---PASTEEPLVPTKSV-----  279
COLLAGENASE1   STDIGALMYPSYTFSGDV---QLAQDDIDGIQAIYGRS---QNPVQ--------------  268
                    ***   *          *  **  * *  **

STROMELYSIN3   APLEPDAPPDACEA--SFDAVSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGLP  340
STROMELYSIN1   -PPEPGTPAN-CDPALSFDAVSTLRGEILIFKDRHFWRKSLRKLEPELH-LISSFWPSLP  337
STROMELYSIN2   -PSGSEMPAK-CDPALSFDAISTLRGEYLFFKDRYFWRRSHWNPEPEFH-LISAFWPSLP  336
COLLAGENASE1   -PIGPQTPKA-CDSKLTFDAITTIRGEVMFFKDRFYMRTNPFYPEVELN-FISVFWPQLP  325
                *     *   *     ***  * ***   **     *                * *  **
```

FIG.3B

GeneDX 1009, pg. 569

```
STROMELYSIN3   SPVDAAFE-DAQGHIWFFQGAQYWVYDGEKPVLG--PAPLTELGLVRFPVH-AALYWGPE  396
STROMELYSIN1   SGVDAAYEVTSKDLVFIFKGNQFWAIRGNEVRAGYPRGIHT-LGFPPTVRKIDAAISDKE  396
STROMELYSIN2   SYLDAAYEVNSRDTVFIFKGNEFWAIRGNEVQAGYPRGIHT-LGFPPTIRKIDAAVSDKE  395
COLLAGENASE1   NGLEAAYEFADRDEVRFFKGNKYWAVQGQNVLHGYPKDIYSSFGFPRTVKHIDAALSEEN  385
                  ** *        * *   *   *     *         *         *

STROMELYSIN3   KNKIYFFRGRDYWRFHPSTRRVDSPVPRR-ATDWRGVPSEIDAAFQDADGYAYFLRGRLY  455
STROMELYSIN1   KNKTYFFVEDKYWRFDEKRNSMEPGFPKQIAEDFPGIDSKIDAVFEEF-GFFYFFTGSSQ  455
STROMELYSIN2   KKKTYFFAADKYWRFDENSQSMEQGFPRLIADDFPGVEPKVDAVLQAF-GFFYFFSGSSQ  454
COLLAGENASE1   TGKTYFFVANKYWRYDEYKRSMDPGYPKMIAHDFPGIGHKVDAVFMKD-GFFYFFHGTRQ  444
                 * ***    ***           *   * *  *     **      *  **  *

STROMELYSIN3   WKFDPVKVKALEGFPRLVGPDFFGCAEPANTFL  488
STROMELYSIN1   LEFDPNAKKVTHT---LKSNSWLNC--------  477
STROMELYSIN2   FEFDPNARMVTHI---LKSNSWLHC--------  476
COLLAGENASE1   YKFDPKTKRILTL---QKANSWFNCRKN-----  469
                 ***                 *
```

FIG.3C

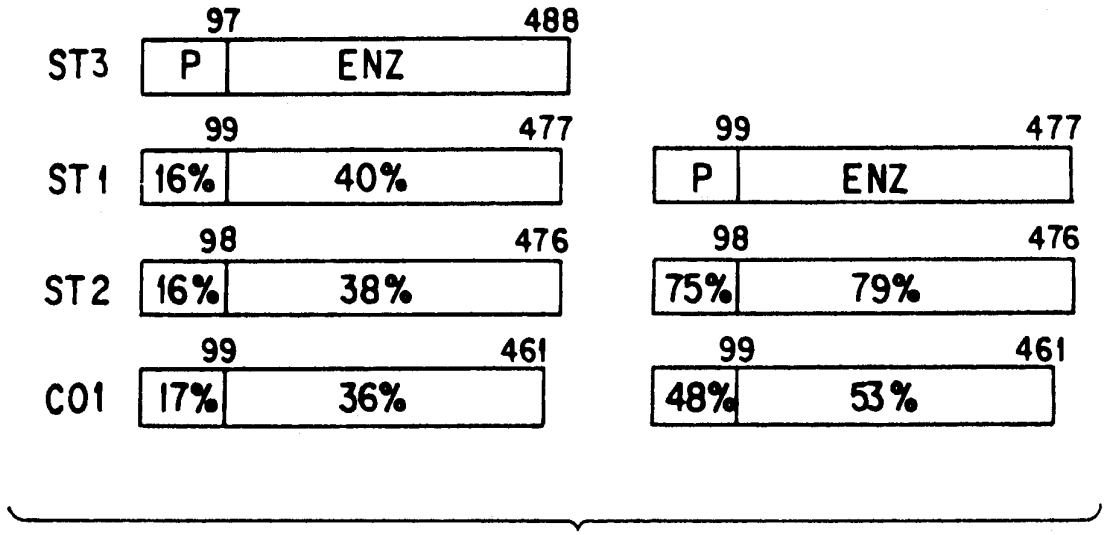

97
488
ST3
P
ENZ
99
477
ST1
16%
40%
98
476
ST2
16%
38%
99
461
C01
17%
36%
99
477
P
ENZ
98
476
75%
79%
99
461
48%
53%

FIG.3D

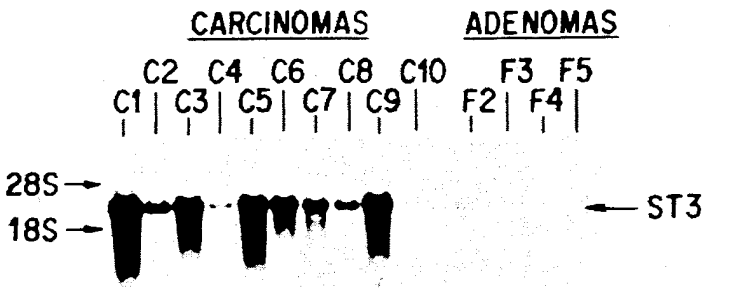


FIG. 4a

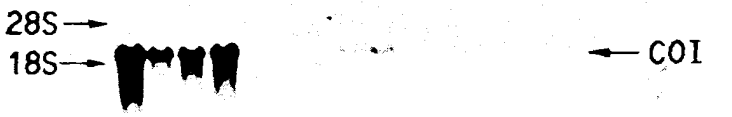


FIG. 4b



FIG. 4c



FIG. 4d



FIG. 4e

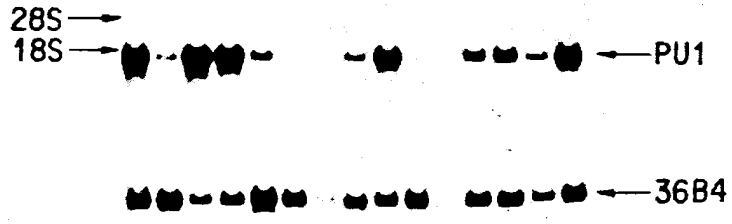


FIG. 4f

GeneDX 1009, pg. 572

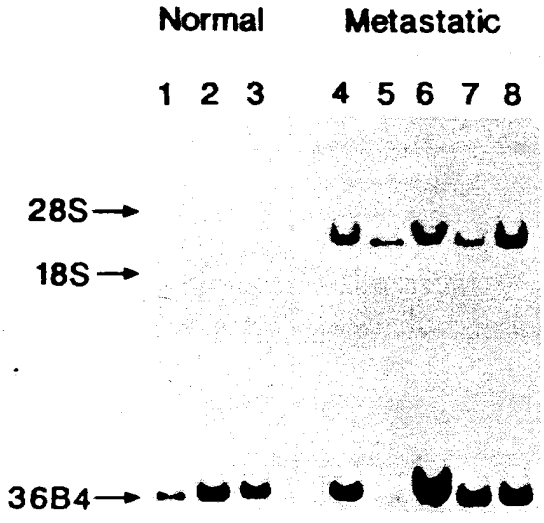


FIG. 5a

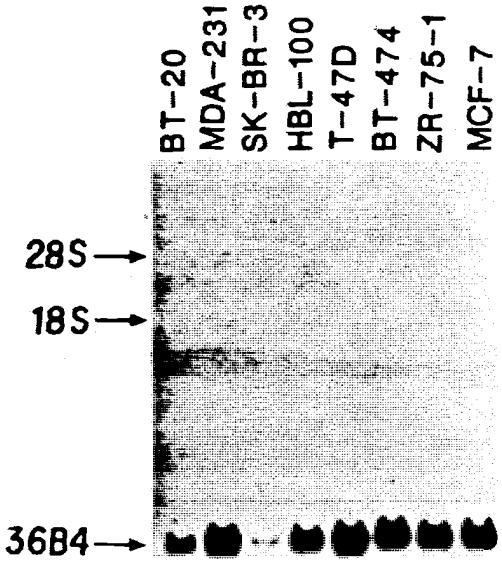


FIG. 5b

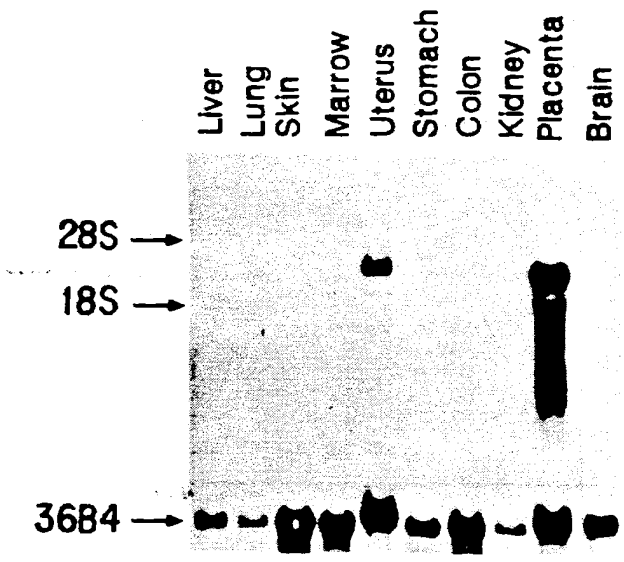


FIG. 5c

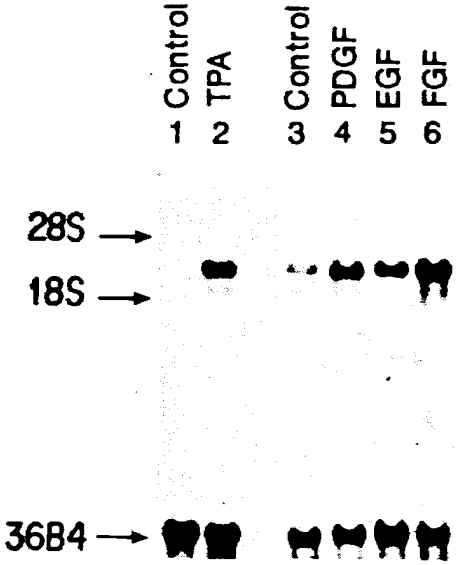


FIG. 5d

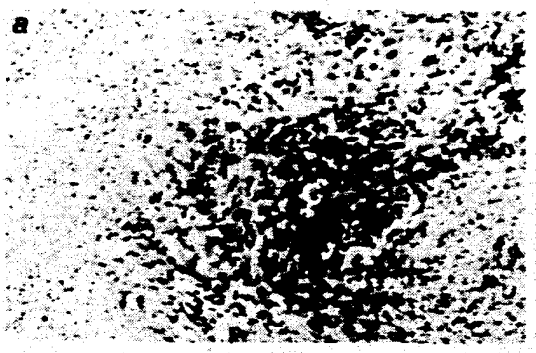


FIG. 6a

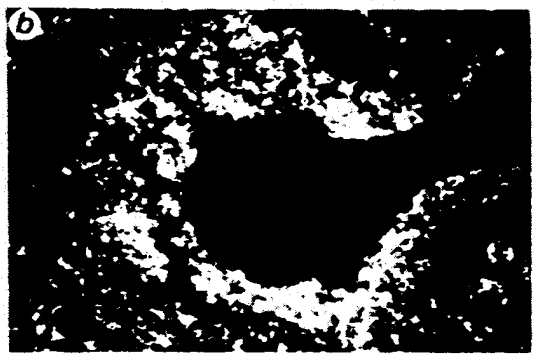


FIG. 6b

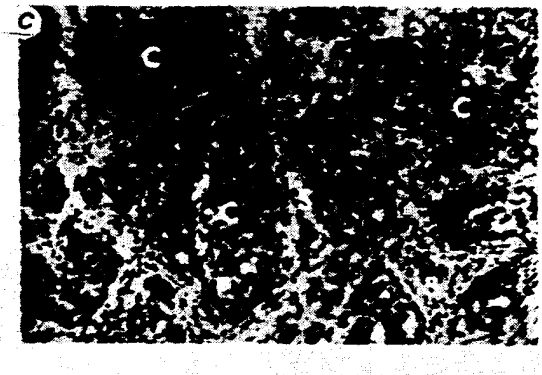


FIG. 6c

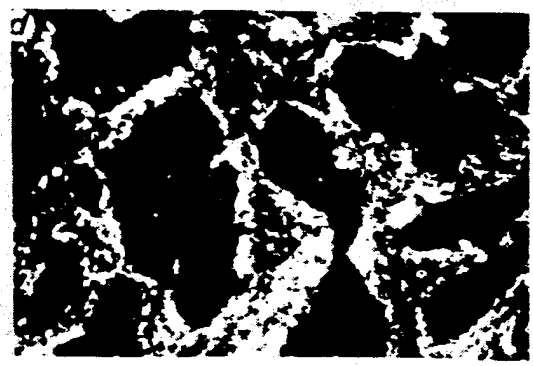


FIG. 6d



FIG. 6e

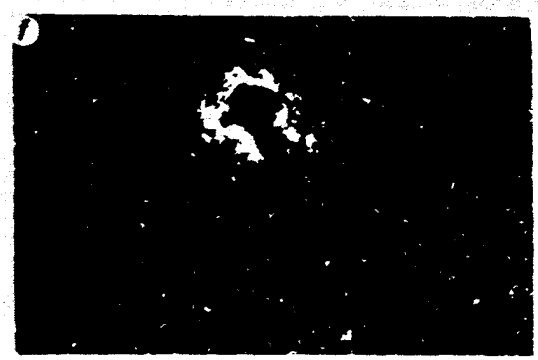


FIG. 6f

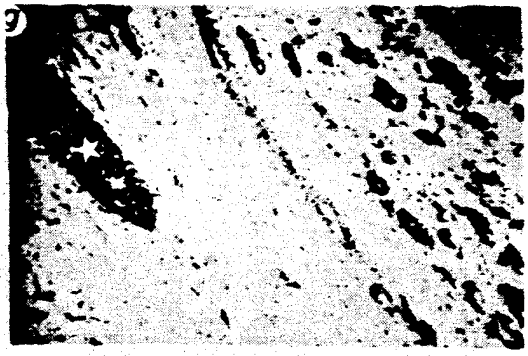


FIG. 6g

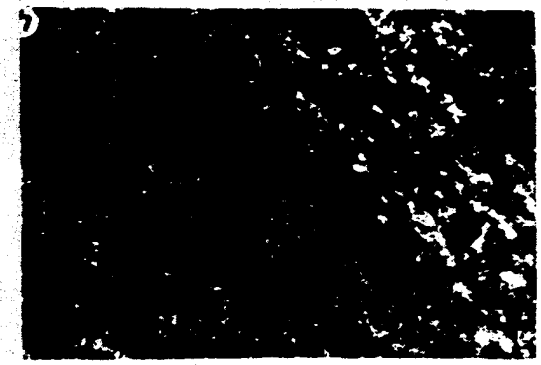


FIG. 6h

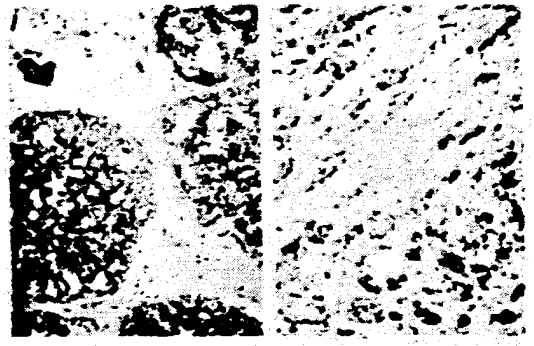


FIG. 6i

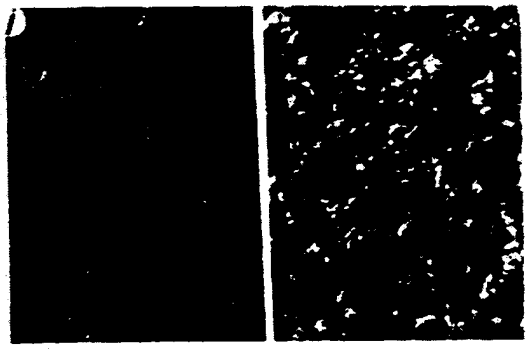


FIG. 6j

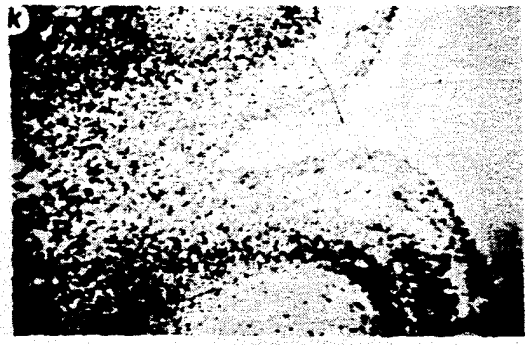


FIG. 6k



FIG. 6l

```
CCCGGGGCGGATGGCACGGGCCGCCTGTCTCCTCCGCGCGATTTCGGGGTGCCTCCTGCT  60
          M  A  R  A  A  C  L  L  R  A  I  S  G  C  L  L  L  17
CCCGCTGCCTCTGCTGCTCCTGTTGCTGCTGCTCCTGCCGTCGCCGCTGATGGCCCGGGC  120
  P  L  P  L  L  L  L  L  L  L  L  L  P  S  P  L  M  A  R  A  37
CAGGCCACCGGAGAGTCACCGTCATCACCCTGTGAAGAAAGGGCCTCGGCTCCTGCATGC  180
  R  P  P  E  S  H  R  H  H  P  V  K  K  G  P  R  L  L  H  A  57
AGCTCTGCCTAATACCTTGACATCTGTCCCCGCGTCTCATTGGGTCCCTAGTCCTGCCGG  240
  A  L  P  N  T  L  T  S  V  P  A  S  H  W  V  P  S  P  A  G  77
TAGCTCCAGGCCTCTACGATGTGGTGTGCCCGACCTGCCTGATGTACTGAATGCCCGGAA  300
  S  S  R  P  L  R  C  G  V  P  D  L  P  D  V  L  N  A  R  N  97
CCGACAGAAGCGCTTCGTCCTGTCAGGAGGACGCTGGGAGAAGACAGACCTCACCTATAG  360
  R  Q  K  R  F  V  L  S  G  G  R  W  E  K  T  D  L  T  Y  R  117
GATCCTCCGGTTCCCATGGCAGCTTGTAAGGGAGCAAGTCCGGCAGACAGTGGCAGAGGC  420
  I  L  R  F  P  W  Q  L  V  R  E  Q  V  R  Q  T  V  A  E  A  139
CCTCCAGGTATGGAGTGAAGTGACCCCACTCACTTTCACTGAGGTGCACGAGGGACGCGC  480
  L  Q  V  W  S  E  V  T  P  L  T  F  T  E  V  H  E  G  R  A  157
TGACATCATGATCGACTTCGCAAGGTACTGGGATGGTGACAACTTGCCGTTTGACGGGCC  540
  D  I  M  I  D  F  A  R  Y  W  D  G  D  N  L  P  F  D  G  P  177
TGGGGGCATCCTGGCCCATGGCTTCTTCCCTAAGACCCACCGAGAAGGGGATGTCCACTT  600
  G  G  I  L  A  H  G  F  F  P  K  T  H  R  E  G  D  V  H  F  197
TGACTATGATGAAACTTGGACTATTGGGGACAACCAGGGAACTGACCTGCTGCAAGTGGC  660
  D  Y  D  E  T  W  T  I  G  D  N  Q  G  T  D  L  L  Q  V  A  217
GGCTCATGAATTTGGCCATGTTCTGGGGCTACAACACACCACAGCAGCTAAGGCCCTCAT  720
  A  H  E  F  G  H  V  L  G  L  Q  H  T  T  A  A  K  A  L  M  237
GTCCCCTTTCTACACCTTCCGCTACCCTCTGAGCCTTAGCCCAGATGACCGAAGGGGCAT  780
  S  P  F  Y  T  F  R  Y  P  L  S  L  S  P  D  D  R  R  G  I  257
CCAGCACCTCTATGGGCGGCCCCAGATGACCCCCACCTCCCCCGCCCCAACTTTGAGCTC  840
  Q  H  L  Y  G  R  P  Q  M  T  P  T  S  P  A  P  T  L  S  S  277
CCAGGCTGGGACAGATACCAATGAGATTGCACTGCTGGAGCCGGAAACCCCGCCAGATGT  900
  Q  A  G  T  D  T  N  E  I  A  L  L  E  P  E  T  P  P  D  V  297
CTGTGAGACTTCCTTCGACGCGGTTTCCACCATCCGAGGAGAGCTCTTCTTCTTCAAGGC  960
  C  E  T  S  F  D  A  V  S  T  I  R  G  E  L  F  F  F  K  A  317
AGGCTTTGTGTGGAGGCTGCGCAGTGGGCGACTGCAGCCCGGGTATCCTGCTTTGGCCTC  1020
  G  F  V  W  R  L  R  S  G  R  L  Q  P  G  Y  P  A  L  A  S  337
```

FIG.7A

```
TCGGCACTGGCAAGGACTGCCCAGCCCTGTGGATGCAGCTTTTGAGGATGCCCAGGGCCA  1080
  R  H  W  Q  G  L  P  S  P  V  D  A  A  F  E  D  A  Q  G  Q   357
GATTTGGTTCTTCCAAGGTGCTCAGTACTGGGTATATGATGGTGAGAAGCCAGTCCTAGG  1140
  I  W  F  F  Q  G  A  Q  Y  W  V  Y  D  G  E  K  P  V  L  G   377
CCCTGCACCACTCTCCAAGCTGGGCCTGCAAGGGTCCCCAGTTCATGCCGCCTTGGTCTG  1200
  P  A  P  L  S  K  L  G  L  Q  G  S  P  V  H  A  A  L  V  W   397
GGGTCCTGAGAAGAACAAGATCTACTTCTTCCGAGGTGGAGACTATTGGCGTTTCCACCC  1260
  G  P  E  K  N  K  I  Y  F  F  R  G  G  D  Y  W  R  F  H  P   417
CAGAACCCAGCGAGTGGACAATCCCGTGCCCCGGCGCTCCACTGACTGGCGAGGGGTACC  1320
  R  T  Q  R  V  D  N  P  V  P  R  R  S  T  D  W  R  G  V  P   437
TTCTGAGATTGATGCTGCCTTCCAGGATGCTGAGGGCTATGCCTACTTCCTTCGTGGCCA  1380
  S  E  I  D  A  A  F  Q  D  A  E  G  Y  A  Y  F  L  R  G  H   457
TCTCTACTGGAAGTTTGATCCCGTGAAGGTGAAGGTCCTAGAAGGCTTTCCTCGCCCCGT  1440
  L  Y  W  K  F  D  P  V  K  V  K  V  L  E  G  F  P  R  P  V   477
AGGTCCTGACTTCTTTGACTGTGCTGAGCCTGCCAATACTTTCCGCTGACAACACTTTGG  1500
  G  P  D  F  F  D  C  A  E  P  A  N  T  F  R  -               492
ATGCATTCAGGGGTACTGACTCCTGCCAGGGCACTTAGATCATGTAAGAGACCCACAGCC  1560
ATATCTGTGGCTCTGGCTTCAGGCATGGGACAGACAGGGCCTATGTCTCCTCAGGGGAGT  1620
GGGTTGGGGTGCAGCCACTGTTTGTAGGAACGACCATGCTGTCATGTCACCTGCCAACAA  1680
TTGTCTCAGACTAGGCAAAGGCTTTGGTGTTACTTAAAAATAAGGGAGGTTTTGGGCTGG  1740
CAATATTTCAGCTACCAATAATCCACAGTCAGCCTGGTTGCCCAAGGTCTCCTATCTCTG  1800
TCCCTCAATGTAGAACCCCCACACAAACTCAGGAATCACCTGCAATGAGGTTCCTGTTGG  1860
GAGTGGTGTTGGTAATGAGATGCCCAGGGTACCATGCTGCCCCTGCTAAGCAACTGGACC  1920
AGTATCTTTCCTGGTAAGTCAGCTCTGGAGAGATAGTGAACTGATCATATTCTGGCAGGT  1980
GATTCAGACAAGTGCTTCCTGGAACTCAGGCCCCAAGGTACACAGCCAGCCAAGGAGGCA  2040
GCTGCTTCCTCCCAGAGACACGGAACCTCAAAGGCCCCACATACCTCACAGCCTTGCCCC  2100
AGGCCATTTCTTTCTGGGGCCCTCTTCCTAGCACAGGTACCCTCTAAGCCATGTACATGT  2160
GTATACAGTGTATAAAGACTTTTTTAAAAAAACAAAAAACCAAACCCCAAAAAAGCCAAG  2220
ACTGTCATTAAACATGAGTGTTTTCTAAAAAAAAAAAAA                        2260
```

FIG.7 B

ANALYTICAL MARKERS FOR MALIGNANT BREAST CANCER

FIELD OF THE INVENTION

The present invention relates to tumor-associated enzyme markers.

Utilizing DNA sequences encoding stromelysin-3, and antibodies capable of binding to stromelysin-3, the present invention provides methods for diagnosing cancer, specifically malignant breast cancer.

BACKGROUND OF THE INVENTION

The number of deaths around the world from cancer each year continues to be of major concern, with only a few treatments being available for specific types of cancer, and these having no absolute guarantee of success. Most treatments rely on a general "shotgun" approach, killing off rapidly growing cells in the hope that rapidly growing cancerous cells will succumb, either to the treatment, or at least be reduced in numbers to allow the body's system to eliminate the remainder.

The search for cures has been hampered by the discovery that different forms of cancer require different treatments. Given that virtually any part of the body can be affected by cancer, the task becomes enormous.

Nevertheless, despite their differences, cancers also share a number of similarities. Prime amongst these is the growth of undifferentiated tissue. However, even this is not 100% accurate, in that certain cancerous cells do exhibit a degree of differentiation, and this is shown in the sex cancers, such as those of breast and testicle, where tumors may be positive or negative for hormone receptors. Treatment of these tumors depends on the hormone state, and may be as simple as administration of the relevant hormone antagonist, such as tamoxifen.

Another factor which most cancers share is that, in order to be fatal, they must metastasize. Until such time as metastasis occurs, a tumor, although it may be malignant, is confined to one area of the body. This may cause discomfort and/or pain, or even lead to more serious symptoms, but if it can be located, it may be surgically removed and, if done with adequate care, cause no further problems.

However, once metastasis sets in, surgical resection may remove the parent tumor, but cancerous cells have invaded the body, and only chemotherapy, or some particular form of targeting therapy, stands any chance of success.

Thus, the ability to invade locally and to metastasize in organs distant from the primary tumor (tumor progression) is the lethal event in the course of most cancers. Alteration/degradation of the extracellular matrix (ECM) surrounding the primary tumor, and modifications of the tumor cell adhesive properties, are known to be crucial for dissociation of the metastatic cells from the primary tumor cells (Liotta, *Cancer Res.* 46:1–7 (1986); Hart et al., *Biochim. Biophys. Acta* 989:65–84 (1989)).

Tumor angiogenesis is essential for both primary tumor expansion and metastatic tumor spread, and angiogenesis itself requires ECM degradation (Blood et al., *Biochim. Biophys. Acta* 1032:89–118 (1990)). Thus, malignancy is a systemic disease in which interactions between the neoplastic cells and their environment play a crucial role during evolution of the pathological process (Fidler, I. J., *Cancer Metastasis Rev.* 5:29–49 (1986)).

Identifying the alterations in gene expression which are associated with malignant tumors, including those involved in tumor progression, is clearly a prerequisite not only for a full understanding of cancer, but also to develop new rational therapies against cancer. Mutations and/or abnormal control of expression of two groups of cellular genes (the proto-oncogenes and the tumor suppressor genes) have been shown to lead in a multistep process to the loss of normal growth control and to the acquisition of the transformed cell phenotype (Weinberg, R. A., *Cancer Res.* 49:3713–3721 (1989)). However, the molecular mechanisms which lead to tumor progression are much less clear (Nowell, P. C., *Cancer Res.* 46:2203–2207 (1986); Fidler, I. J., *Cytometry* 10:673–680 (1989)).

Thus, a further problem arises, in that the genes characteristic of cancerous cells are very often host genes being abnormally expressed. It is quite often the case that a particular protein marker for a given cancer is over-expressed in that cancer, but is also expressed elsewhere throughout the body, albeit at reduced levels.

Some of the proteins associated with cancers are enzymes which break down the extracellular matrix, which is important for maintaining cells in their proper relationship to each other. One such class is the metalloproteinases (MMPs) (Matrisian, L. M., *Trends Genet.* 6:121–125 (1990)), so called because they bind zinc. However, none has been found to be diagnostic, of cancer, or any particular tumors, although the presence of some may be indicative.

MMPs are involved in a number of physiological and pathological processes in which ECM remodelling and cell migration are implicated, e.g. morphogenesis and embryonic development, rheumatoid arthritis, and tumor invasion and metastasis. MMP inhibitors are known to be able to block tumor invasion and angiogenesis, which are crucial for tumor progression, in experimental models.

All members of the matrix metalloproteinase family are proteinases which degrade at least one component of ECM, are secreted in a latent form and require activation, such as proteolysis (e.g. by plasmin) to become active. Interstitial collagenases specifically attack connective tissue collagens (I to III), whereas type IV collagenases (72 kD and 92 kD) degrade collagens present in the basement membrane and fibronectin. Stromelysins (transins)-1 and -2, and also pump-1, have a much broader substrate specificity, degrading proteoglycans, laminin, fibronectin, and collagens (III to V).

In man, most of the malignant tumors are carcinomas, and among non-smokers, breast cancer is the leading cause of mortality by cancer in woman (Willett, W., *Nature* 338:389–394 (1989)). The expression of several oncogenes has been reported to be altered in malignant breast cells and tumors, but no particular pattern of oncogene/suppressor gene expression can be consistently associated with breast cancer (Gullick, W. J., *Prog. Growth Factor Res.* 2:1–13 (1990)).

However, the neoplastic cells of breast tumors are often embedded in an adipose and mesenchymal stroma, which may also be important in control of their proliferation and in their ability to metastasize. Indeed, it is known that stroma cells can modulate, both positively and negatively, the growth of normal mammary epithelium (Salomon et al., in *Breast Cancer: Cellular and Molecular Biology* (eds., Lippman, M. E. and Dickson,

R. B.), pp. 363–389 (Kluwer, Boston,. (1988)), and that interactions between the epithelial and stromal components can influence epithelial carcinogenesis in the mammary gland (DeOme et al., *Cancer Res.* 38:2103–2111 (1978)).

The existence of "activated" (Tremblay, *G. Exp. Mol. Pathol.* 31:248–260 (1979)) and/or abnormal (Grey et al., *Proc. Nat. Acad. Sci. USA* 86:2438–2442 (1989)) fibroblasts in malignant breast tumors has been postulated, and it has been proposed that breast cancer could represent a partial escape from dependence on a stromal requirement or an abnormally strong response to a stromal component.

Owing to the nature of cancerous tissue, it is usually relatively easy to set up a continuous culture, or cell line, of a given cancer, a process which makes it easy to study the effects of a given treatment regimen. A significant drawback to such systems lies in their very nature—while the test treatment will establish whether it can act directly against the cells, it is by no means certain what effect the treatment will have in vivo, and biochemical analysis of such lines is inavitably in the absence of the tissue normally surrounding the tumor in vivo.

SUMMARY OF THE INVENTION

It is an object of the invention to identify genes whose expression is increased in breast carcinomas, whereby breast carcinomas are considered as malignant epithelial cells interacting with their surrouding stroma.

We have now found that a previously uncharacterized protein is diagnostic of certain invasive cancers, especially breast carcinomas, head and neck squamous cell carcinomas and skin (squamous and basal cell types) carcinomas. The protein apparently belongs to the group of metalloproteinases, and is referred to as stromelysin-3 herein.

DESCRIPTION OF THE FIGURES

FIG. 1. Northern blot analysis of total RNA from C1 breast carcinoma and Fibroadenoma.

RNA isolated from C1 breast carcinoma and fibroadenoma cells was probed using four independently isolated cDNA probes as described in example 1.

FIG. 2. Nucleotide sequence of stromelysin-3 cDNA.

The nucleotide sequence of the cDNA and the deduced amino acid sequence of stromelysin-3 is presented. Starting from the 5′ end, the underlying nucleotide sequences correspond to: the punitive signal peptide; the PRCGBPD sequence characteristic of prometalloproteinase, the conserved histidine residues of the zinc binding domain and the poly(A+) signal sequence.

FIG. 3. Comparison of Metalloproteinase Sequences.

Amino acid sequences are aligned and compared for stromelysin-3, stromelysin-2, stromelysin-1 and collagenase-1, all putative metalloproteinase, as described in example 3.

FIG. 4. Northern Blot Analysis of Human Metalloproteinase

Total RNA was prepared from four oestrogen receptor negative breast carcinomas (C1, grade II; C2, C3 and C4, grade III), six oestrogen receptor positive breast carcinomas (C5, C8 and C9, grade II; C6 and C7, grade III; Cio, grade I) and four breast fibroadenomas (F2-F5).

The RNA was probed with (a) stromelysin-3 RNA, (b) type 1 collagenase RNA (COI) (c) 92 kd type 4 collagenase RNA (COIV 92k) (d) 72 kd type 4 collagenase RNA (COIV 72k) (e) stromelysin-1 and 2 RNA (ST1/2) and pump-1 RNA (PUI), as described in example 4.

FIG. 5. Northern Blot Analysis of Stromelysin-3 RNA From Various Cell Lines and Tissues.

(a) Three normal and five metastatic auxiliary lymph nodes from patients with breast cancer; (b) F o u r oestrogen receptive negative (BT-20, MDA-231, SK-BR-3, HBL-100) and four oestrogen receptor positive (T-47D, BT-474, ZR-75-1, MCF-7) breast carcinoma lines; (c) Ten normal human tissues; and (d) HFL-1 Human Foetal Deployed Fibroblasts (ATCC CCL 153) cultured in serum free medium (1 and 2), in the absence (1) or presence (2) of tPA are cultured in serum free media supplemented with 20 mg/ml insulin (3-6), in the absent (3) or presence (4) of PDGF, (5) of EGF, or (6) of bFGF, were probed with stromelysin-3 sequences, as described in example 5.

FIG. 6. Localization of Stromelysin-3 RNA Transcripts in Sections of Breast Carcinoma and Embryolimbuds.

Bright field micrographics of tissue sections (X 100) stained with hematoxylin (A,C,E,G,I and K); and dark field images of the same sections (still stained with hematoxylin) after in situ hybridization with anti-sense stromelysin-3 cRNA (B,D,F,H,J and L) as described in example 6.

FIG. 7. cDNA Sequence of Mouse ST3 cDNA.

cDNA sequence of mouse ST3 gene and comparison with human ST3 cDNA sequences.

DETAILED DESCRIPTION OF THE INVENTION

In a first aspect, the present invention provides a process for the diagnosis of invasive cancer, especially breast, head and neck, and skin carcinomas, comprising the detection of either stromelysin-3, or a nucleotide sequence encoding stromelysin-3.

In an alternative aspect, the present invention provides the use of an agent to interfer with the synthesis or activity of stromelysin-3 in the treatment or prophylaxis of invasive cancer, especially breast, head and neck, and skin carcinomas.

It will be appreciated that metastatic tumors are invasive, but that invasive tumors are not necessarily metastatic (for example basal cell skin carcinomas).

As expression of the stromelysin-3 gene is specific to regions of ECM degradation and apparently encodes a metalloproteinase, it is assumed that its ECM degrading activity is crucial to tumor progression into metastasis. Expression of stromelysin-3 by the stromal cells is likely to break down an important part of the ECM, thereby allowing cancerous cells to migrate away from the parent tumor.

Accordingly, any agent which can affect the activity of stromelysin-3 will have an effect on metastasis. Such agents will suitably be those which either prevent synthesis of the protein, prevent maturation of the protein, or alter the activity of the enzyme, either by blocking or by altering its activity.

Expression of the stromelysin-3 gene was found to be, in the first instance, diagnostic of breast cancer in the metastatic phase. In fact, this result was achieved by the detection of mRNA in a variety of resected tumors. Breast cancer was chosen, as this is responsible for the highest death rate, by cancer, in the non-smoking female population.

Stromelysin-3 is a novel protein almost certainly belonging to the MMP family, and is associated with invasive breast carcinomas, irrespective of their hormonal status.

The members of the MMP family require an activation step, which may be associated with removal of the pre- and pro-sequences, to become active. The amino acid sequence of pro- and mature stromelysin-3 is notably different from those of the previously characterized MMPs, and may exhibit distinct properties regarding maturation, activation and specificity for ECM components.

The stromelysin-3 gene is expressed by all primary invasive breast carcinomas, by some of their metastases, and in tissues in which extensive ECM remodelling is known to occur (uterus, placenta and limb bud) analyzed for such expression, but not in breast fibroadenomas and normal adult tissues, suggesting that the stromelysin-3 gene product plays an important role in breast cancer progression. Also in agreement with this concept, the stromelysin-3 gene is not expressed in most in situ breast carcinomas, with the exception of in situ carcinomas of the comedo type, which are usually considered as preinvasive lesions and are often associated with microinvasion. Thus the presence of stromelysin-3 RNA transcripts in other than the low concentrations found elsewhere in the body, other than uterus or placenta, is diagnostic of a metastatic cancer or of a cancer with a high risk of becoming invasive.

Stromelysin-3 may be involved in the lytic processes which are likely to be associated with invasive tumor growth. Alternatively, it is possible that stromelysin-3 could also play a role in the formation of desmoplasia, which is associated with most invasive breast cancer lesions, and may represent a host reaction to prevent further malignant cell spread (Ahmed, A., *Pathol. Annu.* 25(Pt2):237–286 (1990)). In such an instance, enhancement of stromelysin-3 activity would be advantageous.

Further, the restricted expression of the stromelysin-3 gene in stromal fibroblasts immediately surrounding the neoplastic cell islands is strikingly in contrast to collagenase IV, another metalloproteinase known to be associated with the malignant conversion of some tumorigenic cells, and cathepsin D, a lysosomal aspartyl protease whose expression is increased in breast carcinomas, both of which are expressed, not in the fibroblasts, but in the neoplastic epithelial cells of breast cancers (Monteagudo et al., *Am. J. Pathol.* 136:585–592 (1990); Garcia et al., *Steroid Biochem.* 27:439–445 (1987)).

To identify the novel breast cancer marker, a cDNA library was constructed, and substracted with poly (A+) RNA from a fibroadenoma source. By this process, the cDNA library was enriched for sequences characteristic of metastatic cancers.

A number of clones was grown up and screened using probes derived from poly(A+) RNA from metastatic tumors and from fibroadenomas. Those clones which bound more greatly to the probes derived from metastatic cancer poly(A+) RNA were then grown up further.

Of the clones generated in this manner, one was found to be differentially expressed to the extent that high rates of expression were only found in malignant breast and pharyngeal cancers, head, neck, and skin (squamous and basal cell type) carcinomas, as well as in the uterus and placenta, in all of which there is a breaking down of the ECM, which, when associated with cancer, allows cancerous cells to spread around the body (metastasis).

In the case of the uterus and the placenta, breakdown of the ECM occurs naturally, whilst the same event elsewhere is likely to be characteristic of tumor growth.

It is also interesting to note that expression of the stromelysin-3 gene was found in interdigital differentiation during limb budding in the foetus, which is associated with breakdown of the ECM.

Characterization of the cDNA sequence illustrated that there was an open reading frame. Comparison of the encoded protein sequence with a known library established that the protein belonged to a family known to break down the ECM. Although the sequence of stromelysin-3 bears less similarity to the other members of its family than any of the other members bear to each other, it does, nevertheless, present a number of characteristic regions which serve to identify the nature of the enzyme. Accordingly, the protein has been named stromelysin-3, although it may be a collagenase, or may break down a different constituent of the ECM altogether.

Construction of nucleotide probes to establish the occurrence of stromelysin-3 mRNA revealed a tissue distribution as described above, and also enabled photomicrographs to exactly locate the areas of expression of the stromelysin-3 gene by labelling.

An analysis of the photomicrographs generated by this method showed, somewhat suprisingly, that the stromelysin-3 gene was not expressed in the cancerous cells itself, but in the surrounding stroma. In addition, the stroma did not exhibit any evidence of stromelysin-3 mRNA when the basement membrane of the tumor was still intact (see FIG. 6). The stromelysin-3 gene is expressed by all primary invasive breast carcinomas, by some of their metastases nodes, and in tissues in which extensive ECM remodelling is known to occur (uterus, placenta and limb bud) analyzed for such expression, but not in breast fibroadenomas and normal adult tissues, suggesting that the stromelysin-3 gene product plays an important role in breast cancer progression. Also in agreement with this concept, the stromelysin-3 gene is not expressed in most in situ breast carcinomas, with the exception of in situ carcinomas of the comedo type, which are usually considered as preinvasive lesions and are often associated with microinvasion. Stromelysin-3 always occurs in the stroma of metastatic cancers, and does not occur in the stroma of in situ primary tumors (tumors still having a basement membrane and which are non-invasive). Thus the presence of stromelysin-3 RNA transcripts in other than the low concentrations found elsewhere in the body, other than uterus or placenta, is diagnostic of a metastatic cancer or of a cancer with a high risk of becoming invasive.

Furthermore, expression of the stromelysin-3 gene was not detected in any ER-positive or negative breast cancer cell lines, even though some of them are known to secrete and possess receptors for EGF/TGF-α and FGF (factors which are implicated in expression of the stromelysin-3 gene).

Accordingly, standard detection techniques applied to stromelysin-3, its precursors or its coding nucleotide sequences may be used to diagnose a metastatic cancer, or to confirm that a primary tumor has not yet reached the fatal metastatic phase.

Such techniques may include detection with nucleotide probes, such as in the manner described above, or

may comprise detection of the stromelysin-3 protein by, for example, antibodies or their equivalent.

The nucleotide probes may be any that will hybridize more strongly to the sequence shown in the accompanying FIG. 2 than to other naturally occurring sequences. Types of probe include cDNA, riboprobes, synthetic oligonucleotides and genomic probes. The type of probe used will generally be dictated by the particular situation, such as riboprobes for in situ hybridization, and cDNA for Northern blotting, for example. The most preferred probes are those which correspond to the negative strand of the cDNA of FIG. 2. It is also possible to provide probes which recognize introns located within the stromelysin-3 gene, but this is not necessarily as reliable as detecting RNA transcripts.

Detection of the stromelysin-3 encoding gene, per se, will generally serve no purpose in diagnosis, but other forms of assay to detect transcripts and other expression products will generally be useful. The probes may be as short as is required to differentially recognize stromelysin-3 mRNA transcripts, and may be as short as, for example, 15 bases.

The form of labelling of the probes may be any that is appropriate, such as the use of radioisotopes, for example, ^{32}P and ^{35}S. Labelling with radioisotopes may be achieved, whether the probe is synthesized chemically or biologically, by the use of suitably labelled bases. Other forms of labelling may include enzyme or antibody labelling such as is characteristic of ELISA, but detection of mRNA transcripts by labelled probes will generally be by way of X-radiography.

Detection of RNA transcripts may be achieved by Northern blotting, for example, wherein a preparation of RNA is run on a denaturing agarose gel, and transferred to a suitable support, such as activated cellulose, nitrocellulose or glass or nylon membranes. Radiolabelled cDNA or RNA is then hybridized to the preparation, washed and analyzed by autoradiography.

In situ hybridization visualization may also be employed (Example 6), wherein a [^{35}S]-labelled antisense cRNA probe is hybridized with a thin section of a biopsy sample, washed, cleaved with RNase and exposed to a sensitive emulsion for autoradiography. The samples may be stained with haematoxylin to demonstrate the histological composition of the sample, and dark field imaging with a suitable light filter shows up the developed emulsion.

Immunohistochemistry may be used to detect expression of stromelysin-3 in a biopsy sample. A suitable antibody is brought into contact with, for example, a thin layer of cells, washed, and then contacted with a second, labelled antibody. Labelling may be by enzyme, such as peroxidase, avidin or by radiolabelling. Chromogenic labels are generally preferable, as they can be detected under a microscope.

More generally preferred is to detect the protein by immunoassay, for example by ELISA or RIA, which can be extremely rapid. Thus, it is generally preferred to use antibodies, or antibody equivalents, to detect stromelysin-3, but use of a suitably labelled stromelysin-3 substrate may also be advantageous.

It may not be necessary to label the substrate, provided that the product of the enzymatic process is detectable and characteristic in its own right (such as hydrogen peroxide for example). However, if it is necessary to label the substrate, then this may also comprise enzyme labelling, labelling with radioisotopes, antibody labelling, fluorescent marker labelling or any other suitable form which will be readily apparent to those skilled in the art.

Most preferred for detecting stromelysin-3 expression is the use of antibodies. Antibodies may be prepared as described below, and used in any suitable manner to detect expression of stromelysin-3.

Antibody-based techniques include ELISA (enzyme linked immunosorbent assay) and RIA (radioimmunoassay). Any conventional procedures may be employed for such immunoassays. The procedures may suitably be conducted such that: a stromelysin-3 standard is labelled with a radioisotope such as ^{125}I or ^{35}S, or an assayable enzyme, such as horseradish peroxidase or alkaline phosphatase and, together with the unlabelled sample, is brought into contact with the corresponding antibody, whereon a second antibody is used to bind the first and radioactivity or the immobilized enzyme assayed (competitive assay); alternatively, stromelysin-3 in the sample is allowed to react with the corresponding immobilized antibody, radioisotope- or enzyme-labelled anti-stromelysin-3 antibody is allowed to react with the system and radioactivity or the enzyme assayed (ELISA-sandwich assay). Other conventional methods may also be employed as suitable.

The above techniques may be conducted essentially as a "one-step" or "two-step" assay. The "one-step" assay involves contacting antigen with immobilized antibody and, without washing, contacting the mixture with labeled antibody. The "two-step" assay involves washing before contacting the mixture with labeled antibody. Other conventional methods may also be employed as suitable.

Enzymatic and radio-labelling of stromelysin-3 and/or the antibodies may be effected by conventional means. Such means will generally include covalent linking of the enzyme to the antigen or the antibody in question, such as by glutaraldehyde, specifically so as not to adversely affect the activity of the enzyme, by which is meant that the enzyme must still be capable of interacting with its substrate, although it is not necessary for all of the enzyme to be active, provided that enough remains active to permit the assay to be effected. Indeed, some techniques for binding enzyme are non-specific (such as using formaldehyde), and will only yield a proportion of active enzyme.

It is usually desirable to immobilize one component of the assay system on a support, thereby allowing other components of the system to be brought into contact with the component and readily removed without laborious and time-consuming labor. It is possible for a second phase to be immobilized away from the first, but one phase is usually sufficient.

It is possible to immobilize the enzyme itself on a support, but if solid-phase enzyme is required, then this is generally best achieved by binding to antibody and affixing the antibody to a support, models and systems for which are well-known in the art. Simple polyethylene may provide a suitable support.

Enzymes employable for labelling are not particularly limited, but may be selected from the members of the oxidase group, for example. These catalyze the production of hydrogen peroxide by reaction with their substrates, and glucose oxidase is often used for its good stability, ease of availability and cheapness, as well as the ready availability of its substrate (glucose). Activity of the oxidase may be assayed by measuring the concentration of hydrogen peroxide formed after reaction of

the enzyme-labelled antibody with the substrate under controlled conditions well-known in the art.

Other techniques may be used to detect stromelysin-3 according to preference. One such is Western blotting (Towbin et al., *Proc. Nat. Acad. Sci.* 76:4350 (1979)), wherein a suitably treated sample is run on an SDS PAGE gel before being transferred to a solid support, such as a nitrocellulose filter. Anti-stromelysin-3 antibodies (unlabelled) are then brought into contact with the support and assayed by a secondary immunological reagent, such as labelled protein A or anti-immunoglobulin (suitable labels including ^{125}I, horseradish peroxidase and alkaline phosphatase).

Samples for diagnostic purposes may be obtained from any relevant site. A sample obtained direct from the tumor, such as the stroma or cytosol, may be ideal, but it may also be appropriate to obtain the sample from blood, for example. However, if the sample is derived from blood, highly sensitive assays may be required, as the amount of stromelysin-3 would then be diluted through the bloodstream. Such diagnosis may be of particular importance in monitoring progress of a patient, such as after surgery to remove a tumor. If a reference reading is taken after the operation, then another taken at regular intervals, any rise could be indicative of a relapse, or possibly a metastasis. The taking of such readings may need to take into account activity in the uterus, for example.

Anti-stromelysin-3 antibodies may also be used for imaging purposes. Besides enzymes, other suitable labels include radioisotopes, iodine (^{125}I, ^{121}I), carbon (^{14}C), salphee (^{35}S), tritium (^{3}H), indium (^{112}In), and technetium (^{99m}Tc), fluorescent labels, such as fluorescein and rhodamine, and biotin.

However, for in vivo imaging purposes, the position becomes more restrictive, as antibodies are not detectable, as such, from outside the body, and so must be labelled, or otherwise modified, to permit detection.

Markers for this purpose may be any that do not substantially interfere with the antibody binding, but which allow external detection. Suitable markers may include those that may be detected by X-radiography, NMR or ESR. For X-radiographic techniques, suitable markers include any radioisotope that emits detectable radiation but that is not overtly harmful to the patient, such as barium or caesium, for example. Suitable markers for NMR and ESR generally include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by suitable labelling of nutrients for the relevant hybridoma, for example.

In the case of in vivo imaging methods, an antibody or antibody fragment which has been labelled with an appropriate detectable imaging moiety, such as a radioisotope (for example, ^{131}I, ^{112}In, ^{99m}Tc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously or intraperitoneally) into the subject (such as a human) to be examined.

The size of the subject, and the imaging system used, will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of technetium-99m. the labelled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain stromelysin-3. The labelled antibody or antibody fragment can then be detected using known techniques.

For a general discussion of this technological area, see S. W. Burchiel et al., "Immunopharmacokinetics of Radiolabelled Antibodies and Their Fragments" (Chapter 13 in *Tumor Imaging: The Radiochemical Detection of Cancer*, eds., S. W. Burchiel and B. A. Rhodes, Masson Publishing Inc. (1982)).

The antibodies may be raised against either a peptide of stromelysin-3 or the whole molecule. Such a peptide may be presented together with a carrier protein, such as an albumin, to an animal system or, if it is long enough, say 25 amino acid residues, without a carrier. Human antibodies are unlikely to be able to recognize stromelysin-3, as this protein will represent a self protein.

As used herein, the term "peptide" means any molecule comprising 2 or more amino acids linked via a peptide bond. As such, the term includes oligopeptides, polypeptides and proteins.

Polyclonal antibodies generated by the above technique may be used direct, or suitable antibody producing cells may be isolated from the animal and used to form a hybridoma by known means (Kohler and Milstein, *Nature* 256:795 et seq. (1975)). Selection of an appropriate hybridoma will also be apparent to those skilled in the art, and the resulting antibody may be used in a suitable assay to identify stromelysin-3.

Antibodies, or their equivalents, may also be used in accordance with the present invention for the treatment or prophylaxis of metastatic cancers. Administration of a suitable dose of the antibody may serve to block production, or to block the effective activity of stromelysin-3, and this may provide a crucial time window in which to treat the malignant growth.

Prophylaxis may be appropriate even at very early stages of the disease, as it is not known what actually leads to metastasis in any given case. Thus, administration of the antibodies, their equivalents, or factors, such as TIMPs (naturally occurring compounds which regulate the MMPs-tissue inhibitors of metalloproteinases), which interfere with stromelysin-3 activity, may be effected as soon as cancer is diagnosed, and treatment continued for as long as is necessary, preferably until the threat of the disease has been removed.

A preferred form of treatment is to employ the so-called magic bullet technique, where a suitable toxin is attached to the antibodies which then target the area of the tumor. Such toxins are well known in the art, and may comprise toxic radioisotopes, heavy metals, enzymes and complement activators, as well as such natural toxins as ricin which are capable of acting at the level of only one or two molecules per cell. It may also be possible to use such a technique to deliver localized doses of hormone antagonists or any other suitable physiologically active compounds, which may be used, for example, to treat cancers.

It will be appreciated that antibodies for use in accordance with the present invention, whether for diagnostic or therapeutic applications, may be monoclonal or polyclonal as appropriate. Antibody equivalents of these may comprise: the Fab' fragments of the antibodies, such as Fab, Fab', $F(ab')_2$ and Fv; idiotopes; or the results of allotope grafting (where the recognition region of an animal antibody is grafted into the appropriate region of a human antibody to avoid an immune response in the patient), for example. Other suitable

GeneDX 1009, pg. 583

modifications and/or agents will be apparent to those skilled in the art.

In addition to using antibodies to inhibit or remove stromelysin-3, it may also be possible to use other forms of inhibitor. Such inhibitors may be general (for ECM degrading enzymes, for example), or specific for stromelysin-3. Tissue inhibitors of metalloproteinases (TIMPs) are known to exist, and it is extremely likely that there is a specific TIMP for stromelysin-3. Such a TIMP is easily identifiable by standard techniques.

Synthetic inhibitors of stromelysin-3 may also be manufactured, and these will generally correspond to the area of the substrate affected by the enzymatic activity. It is generally preferred that such inhibitors correspond to a frozen intermediate between the substrate and the cleavage products, but it is also possible to provide a sterically hindered version of the binding site, or a version of the binding site which will, itself, irreversibly bind to stromelysin-3. Other suitable inhibitors will be apparent to the skilled person.

Other methods for blocking stromelysin-3 activity may also be employed. These may constitute denaturing agents, for example, although these tend to be non-specific and could only be adequately employed if they could be targeted, such as by the use of specific antibodies. Other forms of stromelysin-3 blocking activity could be effected by blocking the progress from pre-proprotein through to protein. This process provides several target stages, and it is only necessary to identify a stage which can be independently blocked so as not to affect other vital enzymes, or which can again be targeted.

It may also be possible to use peptides or other small molecules to selectively recognize a tertiary structure on stromelysin-3, thereby blocking its enzymic activity. Such an activity blocker need not necessarily bind the active site, but may serve to alter or freeze the tertiary structure of stromelysin-3, destroying, suspending or altering its activity. The blocker also need not necessarily act by itself, but may be linked to another molecule for this purpose, or may serve as a recognition site for a suitable inactivating agent.

Our studies have demonstrated that the occurrence of type I collagenase and 92 kD type IV collagenase mRNAs is exclusively associated with malignant tumors, although the reverse does not always hold (i.e. tumors are not always associated with these proteins).

There is apparently a parallel between the expression of the stromelysin-3 gene and that of the tenascin gene, in invasive breast carcinomas. The ECM glycoprotein tenascin (Chiquet-Ehrismann et al., *Cell* 47:131–139 (1986)) appears to play an essential role in epithelial mesenchyme cell interactions and cell migration during normal development, including that of the mammary gland during organogenesis.

Tenascin has consistently been found to be over-expressed in the fibrous stroma of malignant breast tumors, and appears to be induced in a similar manner to stromelysin-3. When compared with fibronectin, tenascin is a poor substrate for attachment of mammary tumor epithelial cells, suggesting that it may allow them to become invasive.

Thus, stromelysin-3 may act in concert with tenascin during the invasive phase of breast cancer. Stromelysin-3 and tenascin may also be co-expressed during embryogenesis in the regions where epithelium-mesenchyme interactions are known to play an important role, and where cell migration is taking place.

Accordingly, the present invention also provides a process for the diagnosis of metastatic cancer as defined above, further comprising the detection of any of the foregoing proteins, or a nucleotide sequence encoding them.

The invention also provides a use in the treatment or prophylaxis of metastic cancer, further comprising the use of an agent to bind any of the foregoing proteins.

The present invention further provides a nucleotide sequence encoding all or part of stromelysin-3. The sequence of stromelysin-3 is preferably that shown in FIG. 2 of the accompanying drawings, whilst the nucleotide sequence is also preferably that shown in FIG. 2. However, it will be appreciated that the nucleotide sequence may be substantially different from that shown in the Figure, due to degeneracy in the genetic code, provided that it still encodes at least a part of stromelysin-3.

The necessary sequence may vary even further, according to the use to which it is to be put. If it is intended for use to detect RNA transcripts in biological samples, then it will usually be preferable that it more nearly corresponds to the sequence given in FIG. 2. However, the sequence may still vary, provided that hydbridization is possible under the selected conditions of stringency.

A probe may be reverse-engineered by one skilled in the art from the peptide sequence of FIG. 2. However use of such probes may be limited, as it will be appreciated that any one given reverse-engineered sequence will not necessarily hybridize well, or at all with any given complementary sequence reverse-engineered from the same peptide, owing to the degeneracy of the genetic code. This is a factor common in the calculations of those skilled in the art, and the degeneracy of any given sequence is frequently so broad as to yield a large number of probes for any one sequence.

If the nucleotide sequence is required for expression of a stromelysin-3 peptide or entire enzyme, then there may be a considerably greater leeway, both as described above with respect to the genetic code, and also to the fact that some amino acid sequence of stromelysin-3 may be varied without significant effect on the structure or function of the enzyme.

If such differences in sequence are contemplated, then it should be borne in mind that there will be critical areas on the molecule which determine activity. Such areas will usually comprise residues which make up the binding site, or which form tertiary structures which affect the binding site. In general, it is possible to replace residues which form the tertiary structure, provided that residues performing a similar function are used. In other instances, the type of residue may be completely unimportant.

Accordingly, the present invention also includes any variants and mutants on the sequence which still show substantial stromelysin-3 activity, or which exhibit characteristic regions of stromelysin-3 for use in generating antibodies, for example. Such variants and mutants include deletions, additions, insertions, inversions, repeats and type-substitutions (for example, substituting one hydrophilic residue for another, but not strongly hydrophilic for strongly hydrophobic as a rule). Small changes will generally have little effect on activity, unless they are an essential part of the molecule, and may be a side-product of genetic manipulation, for example, when generating extra restriction sites, should such be desired. Modification may also include replace-

ment of one or more of the residues with any other suitable residue, and such replacement may either be 1:1 or any other suitable ratio, greater or less than unity.

Spot mutations and other changes in the coding sequence may be effected to add or delete restriction sites, for example, to otherwise assist in genetic manipulation-/expression, or to enhance or otherwise conveniently to modify the stromelysin-3 molecule.

It will also be appreciated that a stromelysin-3 equivalent will be found in other animals, especially mammals, and sequence information from such sources can be of particular importance to elucidate the conserved regions of the stromelysin-3 molecule. For example, the corresponding sequence in the mouse is ≃80% conserved, including such as the 10 amino acid sequence in the prodomain characteristic of stromelysin-3. It will also be appreciated that animal sequences corresponding to human stromelysin-3 sequences will be readily detectable by methods known in the art and described above, and such sequences and their peptides, as well as mutants and variants thereof, form a part of the invention.

The sequences of the invention may also be engineered to provide restriction sites, if desired. This can be done so as not to interfere with the peptide sequence of the encoded stromelysin-3, or may interfere to any extent desired or necessary, provided that the final product has the properties desired.

As stated above, although hybridization can be an unreliable indication of sequence homology, preferred sequences will generally be those showing in excess of 50%, preferably 70% and more preferably 80% homology with the sequence of FIG. 2.

As with the other metalloproteinases, stromelysin-3 is originally expressed as a pre-proenzyme. Thus, two stages of cleavage are observed in vivo. Cleavage is not necessarily a requirement for in vitro expression, and it may be possible for *E. coli*, for example, to express the mature protein.

Where it is desired to express stromelysin-3 or a characteristic peptide thereof, any suitable system can be used. The general nature of suitable vectors, expression vectors and constructions therefor will be apparent to those skilled in the art.

By "characteristic" is meant any peptide which has a sequence unique to stromelysin-3. Such a sequence may be important to stromelysin-3 activity, or may just be a sequence not found in other peptides. However, sequences important to stromelysin-3 activity are generally preferred, as these are more likely to be conserved within a population.

Suitable expression vectors may be based on phages or plasmids, both of which are generally host-specific, although these can often be engineered for other hosts. Other suitable vectors include cosmids and retroviruses, and any other vehicles, which may or may not be specific for a given system. Again, control sequences, such as recognition, promoter, operator, inducer, terminator and other sequences essential and/or useful in the regulation of expression, will be readily apparent to those skilled in the art, and may be associated with the natural stromelysin-3 sequence or with the vector used, or may be derived from any other source as suitable. The vectors may be modified or engineered in any suitable manner.

Correct preparation of nucleotide sequences may be confirmed, for example, by the method of Sanger et al. (*Proc. Natl. Acad. Sci. USA* 74:5463–7 (1977)).

A cDNA fragment encoding the stromelysin-3 of the invention may easily be inserted into a suitable vector. Ideally, the receiving vector has suitable restriction sites for ease of insertion, but blunt-end ligation, for example, may also be used, although this may lead to uncertainty over reading frame and direction of insertion. In such an instance, it is a matter of course to test transformants for expression, 1 in 6 of which should have the correct reading frame. Suitable vectors may be selected as a matter of course by those skilled in the art according to the expression system desired.

By transforming a suitable organism or, preferably, eukaryotic cell line, such as HeLa, with the plasmid obtained, selecting the transformant with ampicillin or by other suitable means if required, and adding tryptophan or other suitable promoter-inducer (such as indoleacrylic acid) if necessary, the desired stromelysin-3 may be expressed. The extent of expression may be analyzed by SDS polyacrylamide gel electrophoresis—SDS-PAGE (Lemelli, *Nature* 227:680–685 (1970)).

Suitable methods for growing and transforming cultures etc. are usefully illustrated in, for example, Maniatis (*Molecular Cloning, A Laboratory Notebook*, Maniatis et al. (eds.), Cold Spring Harbor Labs, NY (1989)).

Cultures useful for production of stromelysin-3, or a peptide thereof, may suitably be cultures of any living cells, and may vary from prokaryotic expression systems up to eukaryotic expression systems. One preferred prokaryotic system is that of *E. coli*, owing to its ease of manipulation. However, it is also possible to use a higher system, such as a mammalian cell line, for expression of a eukaryotic protein. Currently preferred cell lines for transient expression are the HeLa and Cos cell lines. Other expression systems include the Chinese Hamster Ovary (CHO) cell line.

One valuable system is the baculovirus system, wherein butterfly cells are cotransfected with a DNA vector encoding stromelysin-3, or a suitable peptide, and baculovirus DNA. Recombination occurs within the cell, and suitable baculovirus recombinants may be selected by standard techniques. Thereafter, the recombinant may be used to infect the cell line as desired, stromelysin-3 or peptide being expressed on infection. A particular advantage of this system is the amount of protein produced, which can be in the range of about 1 to about 500 mg/liter.

Although such systems tend not to be as easy to use as the *E. coli* system, the advantage lies in the processing of the protein after primary synthesis. *E. coli*, for example, does not employ the same system for processing pre-proproteins as mammalian cells.

Other expression systems which may be employed include streptomycetes, for example, and yeasts, such as Saccharomyces spp., especially *S. cerevisiae*. Any system may be used as desired, generally depending on what is required by the operator. Suitable systems may also be used to amplify the genetic material, but it is generally convenient to use *E. coli* for this purpose where only proliferation of the DNA is required.

It may be advantageous to produce only the mature enzyme, for the purposes of raising antibodies, as the sequence of the mature enzyme is common to the pro- and prepro sequences also. However, it will be appreciated that cleavage of the pro and prepro portions may alter the tertiary configuration of the molecule, and so it is possible that an antibody raised against the mature enzyme will not detect the proenzyme, for example.

Antibodies raised to the enzyme in either of its earlier states and/or to the pre- or pro-peptides which are cleaved may also prove useful.

The peptide or nucleotide sequence may be any that is characteristic of stromelysin-3, having consideration to the purpose to which it is to be put. Ideally, the sequences would be completely characteristic of stromelysin-3, but the length of such sequences may vary according to the region of the stromelysin-3 molecule. The most preferred regions are those which are highly conserved, and which are not shared with other proteins, although it may be advantageous if the sequence is characteristic of the MMPs or, more particularly, those MMPs associated with invasive tumors.

The invention includes and relates to equivalents of the above peptide and nucleotide sequences, the term "equivalent" being used in the sense of the preceding description, that is to say, equivalents in the sense of sequences having substitutions at the C- or N-terminals, or anywhere else.

The invention also includes mutants of the sequences, the term "mutants" being used with reference to deletions, additions, insertions, inversions and replacement of amino acid residues or bases in the sequence subject to the restrictions described above.

The present invention further includes variants of the sequences, which term is used in relation to other naturally occurring stromelysin-3 which may be discovered from time to time and which shares essentially the same sequence as shown in FIG. 2, but which vary therefrom in a manner to be expected within a large population. Within this definition lie allelic variation and those peptides from other species showing a similar type of activity and having a related sequence. Also included, although less preferred, are animal sequences.

We have also discovered that stromelysin-3 expression can be stimulated by, for example, growth factors and tumor promoters. Typical examples of such factors include EGF FGF and PDGF and TPA. Thus, in conjunction with the foregoing processes, detection of any of these factors in a tumor sample may also help to diagnose the metastatic condition of a cancer.

Thus, the invention also provides the treatment of a metastatic cancer by altering the expression of the stromelysin-3 gene. This may be effected by interfering with the factor required to stimulate stromelysin-3 production, such as by directing specific antibodies against the factor, which antibodies may be further modified to achieve the desired result. It may also be possible to block the receptor for the factor, something which may be more easily achieved by localization of the necessary binding agent, which may be an antibody or synthetic peptide, for example.

Affecting stromelysin-3 gene expression may also be achieved more directly, such as by blocking of a site, such as the promoter, on the genomic DNA.

Where the present invention provides for the administration of, for example, antibodies to a patient, then this may be by any suitable route. If the tumor is still thought to be, or diagnosed as, localized, then an appropriate method of administration may be by injection direct to the site. If the target is breast cancer, then an injection to the breast may suffice, or an implant may be used. If TIMPs are to be administered, for example, then it may also be possible to employ a dermal patch for prolonged administration.

If the cancer is pharyngeal, then a further option may be oral administration, for example, by means of gargling.

In either instance, administration may instead, or additionally, be by injection, including subcutaneous, intramuscular, intravenous and intradermal injections.

Formulations may be any that are appropriate to the route of administration, and will be apparent to those skilled in the art. The formulations may contain a suitable carrier, such as saline, and may also comprise bulking agents, other medicinal preparations, adjuvants and any other suitable pharmaceutical ingredients.

Suitable preparations may also include vaccines comprising stromelysin-3 or a characteristic peptide thereof. Such vaccines may be active or passive, but passive is generally preferred as stromelysin-3 expression occurs in the uterus, and indefinite exposure to antistromelysin-3 antibodies may have undesirable effects. However, active vaccination may be advantageous, especially where a patient has had a hysterectomy, as no tissues will then normally express stromelysin-3. Other suitable vaccines include recombinant viruses containing a nucleotide sequence encoding a stromelysin-3 or a characteristic peptide thereof. One suitable such virus is the vaccinia virus.

The following Examples serve to illustrate the present invention, and are not intended to limit the invention in any manner.

EXAMPLE 1

Cloning of a breast cancer specific cDNA

A breast cancer cDNA library was constructed in the λgt10 vector using poly(A+) RNA from a surgical resection-sample (referred to as tumor C1) of a primary breast cancer. 50,000 plaques were differentially screened using (+) and (−) probes corresponding to cDNAs reverse-transcribed from C1-poly(A+) RNA and poly(A+) RNA from a breast fibroadenoma (referred to as F1), respectively.

FIG. 1 shows a Northern blot analysis of total RNA from C1-breast carcinoma and F1-fibroadenoma using cDNA probes of four genes (A-D) exhibiting higher levels of expression in the carcinoma than in the fibroadenoma. Each lane contained 8 μg of total RNA. The filters were reprobed using the 36B4 probe which corresponds to an ubiquitously expressed gene (Rio et al., *Proc. Nat. Acad. Sci. USA* 84:9243–9247 (1987)).

Specifically, total RNA was prepared (Chirgwin et al., *Biochemistry* 18:5294–5299 (1979)) from surgical specimens stored in liquid nitrogen, and poly(A+) RNA was selected by oligo(dT)-cellulose chromatography. A breast cancer-enriched cDNA library was constructed using cDNA prepared from an oestrogen receptor-negative, grade II, ductal carcinoma (referred to as C1), in which stromal cells represented approximately 50% of the total cell population.

Prior to cloning, the single-stranded cDNA was substracted with an excess of poly(A+) RNA from a breast fibroadenoma (referred to as F1), and the single-stranded enriched material was purified by hydroxyapatite chromatography (Davis et al., *Proc. Nat. Acad. Sci.* USA 81:2194–2198 (1984); Rhyner et al., *Neuroscience Res.* 16:167–181 (1986)).

The breast cancer-enriched cDNA was made double-stranded and cloned into the EcoRI site of the λgt10 vector. Three million recombinant phages were obtained, and ≃50,000 were differentially screened using

GeneDX 1009, pg. 586

replica nylon filters (Biodyne A, Pall Corporation) from plates containing ≈5,000 cDNA clones.

(+) and (−) probes were made using C1-breast cancer cDNA and F1-breast fibroadenoma cDNA, respectively. Both probes were substracted (Davis et al., *Proc. Nat. Acad. Sci. USA* 81:2194–2198 (1984); Rhyner et al., *Neuroscience Res.* 16:167–181 (1986)) with an excess of total human liver RNA before [^{32}P]-labeling using random priming synthesis.

Hybridizations were for two days under stringent conditions (50% formamide, 42° C.) and washing was in 2×SSC, 0.1% SDS, at 22° C., followed by 0.1×SSC, 0.1% SDS at 55° C. 130 differentially labelled plaques were selected for a second screening.

The cDNA inserts of five differential plaques taken at random were purified by PCR amplification, [^{32}P]-labelled, and hybridized to all of the differential plaques to identify related clones. This procedure was repeated several times with differential plaques taken at random, finally yielding four genes referred to as A to D, which exhibited higher levels of expression in C1-carcinoma than in F1-fibroadenoma. The Northern blots for C1-breast cancer and F1-breast fibroadenoma were prepared using total RNA (8 μg) separated by electrophoresis in 1% agarose gels containing formaldehyde and transferred to Hybond-N filters (Amersham).

The blots were stained with methylene blue before prehybridization to check for the integrity and amounts of transferred RNA. Hybridization (18 h) and washing were performed under standard conditions, as described above, using [^{32}P]-labeled cDNA inserts corresponding to A-D genes.

The genes A and B, which were also expressed in normal colon (not shown), were not examined further.

Although expressed in colon (not shown), the C gene was partially characterized because of its high level of differential expression (FIG. 1). It was also expressed in a variety of transformed epithelial cell lines and in normal human skin (not shown). Sequencing of the cDNA of one C clone indicated that the corresponding gene belongs to the keratin gene superfamily (data not shown).

Finally, the D gene (also referred to herein as the stromelysin-3 gene) was further studied, because of its marked differential expression between C1-carcinoma and F1-fibroadenoma (FIG. 1), and also because it was not expressed in normal human colon and in a number of other human tissues (infra).

EXAMPLE 2

Sequencing of Stromelysin-3 Gene and Encoded Protein

Several independent clones were isolated from a non-substracted C1-breast cancer λgt10 cDNA library using a D cDNA insert as probe, and sequenced. FIG. 2 shows the nucleotide sequence of the full length D cDNA and the corresponding protein sequence.

The cDNA open reading frame, encoding a 488 amino acid-long protein, is followed by a 714 base 3'-untranslated region containing a poly(A) addition signal located 14 bases upstream from the 3'-end of the RNA. A presumptive initiation methionine is located at nucleotide position 10–12. Although the corresponding AUG is not associated with and located in a sequence which conforms to the Kozak consensus motif, translation is probably initiated at this AUG, since the sequence immediately downstream corresponds to that for a hydrophobic leader peptide, an expected feature (infra).

In FIG. 2, which shows the nucleotide sequence of stromelysin-3 cDNA and deduced amino acid sequence, the nucleotide residues are numbered in the 5' to 3' direction and deduced amino acids in the open reading frame are designated by their one-letter codes. Starting from the 5'-end, the underlined nucleotide sequences correspond to: the putative signal peptide (two potential cleavage sites are marked by arrows); the PRCGVPD sequence characteristic of prometalloproteinases; the conserved histidine residues of the zinc-binding domain (Matrisian, L. M., *Trends Genet.* 6:21–125 (1990)); and the poly(A) addition signal sequence.

Specifically, a cDNA insert corresponding to the 3'-part of D cDNA [250 bp including a 19 bp poly(AT) region] was [^{32}P]-labeled by random priming synthesis and used to screen a non-substrated λgt10 cDNA library generated from C1-breast tumor poly(A+) RNA by the method of Gubler and Hoffmann (*Gene* 25:262–269 (1983)]. Several independent clones were identified and subcloned in M13 sequencing vector. DNA sequence was determined by the dideoxy method using sequenase and the deaza-dGTP reagent kit from US Biochemical. The sequence was analyzed using the PC/GENE software package.

EXAMPLE 3

Stomelysin-3, a putative metalloproteinase

FIG. 3 shows a comparison of the predicted amino-acid sequences of human stromelysins and human type I collagenase.

(a) Amino-acid sequences were aligned using a multialignment program (Higgins et al., *Gene* 73:237–244 (1988)). Amino-acid residues identical in all of the four sequences are marked by stars. The arrows denote putative signal peptide eleavage sites of stromelysin-3. The arrowhead points to the cleavage which occurs on activation of type I procollagenase and prostromelysins. The 10 amino-acid residues specific to stromelysin-3 at the level of this cleavage site are boxed. The PRCGVPD sequence and the conserved residues of the putative zincbinding domain are underlined.

(b) Left, regions of similarity (in percent amino-acid identity) between stromelysin-3, stromelysin-1 (ST1, Whitham et al., *Biochem. J.* 240:913–916 (1986)), stromelysin-2 (ST2, Muller et al., *Biochem. J.* 253:187–192 (1988)) and type I collagenase (COI, Whitham et al., *Biochem. J.* 240:913–916 (1986));

(b) Right, regions of similarity between ST1, ST2 and COI; P indicates the signal peptide and the pro-domain; ENZ indicates the domain corresponding to the mature active enzymes.

Thus, comparison of the derived protein sequence with the Swissprot data library (release 14) showed that the new protein belongs to the family of secreted matrix metalloproteinases (MMPs) (FIG. 3*a*). Accordingly, the new protein possesses an hydrophobic N-terminal leader sequence candidate (underlined in FIG. 2), and exhibits the highly conserved sequence PRCGVPD (amino-acid residues 78–84), which is characteristic of the prodomain of the MMPs, as well as having the zinc binding site of MMPs (amino-acid residues 212–225-FIG. 3*a*) (Matrisian, L. M., *Trends Genet.* 6:121–125 (1990)).

GeneDX 1009, pg. 587

By analogy with the other members of the family, the N-terminal amino acid of the mature protein is likely to correspond to phenylalanine 98 of the pre-proprotein (Whitham et al., *Biochem. J.* 240:913–916 (1986)) (FIG. **3***a*). After optimal alignments, the similarity between the putative mature protein is 40% with stromelysin-1 (Whitham et al., *Biochem. J.* 240:913–916 (1986)), 38% with stromelysin-2 (Muller et al., *Biochem. J.* 253:187–192 (1988)) and 36% with type I collagenase (Goldberg et al., *J. Biol. Chem.* 261:6600–6605 (1986)) (FIG. **3***b*).

The substrate specificity of the new protein is not known. Herein, it is referred to as stromelysin-3, although its similarity with stromelysin-1 (40%) is clearly much below that existing between stromelysin-1 and stromelysin-2 (79%), and even lower than the similarity existing between type I collagenase and stromelysin-1 (53%) (FIG. **3***b*). Thus, while the protein is an MMP, the cognomen "stromelysin" is not necessarily strictly accurate, but is convenient.

In addition, upstream of the PRCGVPD sequence, there is no significant similarity between stromelysin-3 and the other MMPs with which it is has been compared (FIG. **3**). However, stromelysin-3 has a unique short sequence (amino-acid residues 88–97) at a position corresponding substantially precisely with the proprotein cleavage site of type I collagenase and the stromelysins (Whitham et al., supra). Further, stromelysin-3, as with type I collagenase and the other stromelysins, does not exhibit the fibronectin-like domain characteristic of type IV collagenases (Wilhelm et al., *J. Biol. Chem.* 264:17213–17221 (1989)).

EXAMPLE 4

Over-expression in Breast Carcinomas

The occurrence of stromelysin-3 RNA transcripts was studied in resected samples of 30 breast carcinomas and five breast fibroadenomas.

FIG. **4** shows Northern blot analyses of human metalloproteinase RNAs in breast tumors:
(a) stromelysin-3 RNA;
(b) type I collagenase RNA (COI);
(c) 92-kD type IV collagenase RNA (COIV 92K);
(d) 72-kD type IV collagenase RNA (COIV 72K);
(e) stromelysin-1 and -2 RNA's (ST1/2); and
(f) pump-1 RNA (PUI).

Total RNA was prepared from four oestrogen receptor-negative breast carcinomas (C1, grade II; C2, C3 and C4, grade III), six oestrogen receptor-positive breast carcinomas (C5, C8 and C9, grade II; C6 and C7 grade III; CIO, grade I) and four breast fibroadenomas (F2–F5). Each lane contained 8 μg of RNA. The 36B4 signal corresponds to the RNA of a control gene (FIG. **1**).

Specifically, several Northern blots were prepared in parallel with identical RNA samples, as for FIG. **1**, and hybridized with either of the following cDNA probes: (a) 1.6 kb insert covering the 3' -part of stromelysin-3 cDNA, (b) COI cDNA, (e) ST2 cDNA (which cross-hybridizes with ST1 RNA), (f) PUI cDNA (COI, ST2 and PU1 probes kindly provided by R. Breathnach, Muller et al., *Biochem. J.* 253:187–192 (1988)), or 80 mer antisense oligonucleotide probes corresponding to (c) COIV 92K (nucleotides 2144–2223, Wilhelm et al., *J. Biol. Chem.* 2641:7213–17221 (1989)) and (d) COIV 72K (nucleotides 1937–2016, Collier et al., *J. Biol. Chem.* 263:6579–6587 (1988)).

The cDNA probes were (^{32}P)-labeled using random priming synthesis ($\simeq 5\times10^8$ cpm/μg) and the oligonucleotides were labeled using 5'-end kination $\simeq 10_8$ cpm/μg). Hybridizations were carried out under stringent conditions (42° C., 50% formamide) with $\simeq 10^6$ cpm/ml. The filters were then washed in 2×SSC, 0.1% SDS, at 22° C., followed by 0.1×SSC, 0.1% SDS at 55° C. Autoradiography was for (a), 18 h, (b), 20 h, (c), (d) and (e), 4 days, (f), 2 days, at −80° C. with an intensifying screen.

Stromelysin-3 mRNA was found in all of the breast carcinomas, regardless of whether they were oestradiol receptor (ER) positive (C5–C10) or negative (C1–C4) (FIG. **4***a*), but not in the fibroadenoma samples, with one exception (F2) where the level of expression was similar to the lowest level observed in breast carcinomas.

The occurrence of RNA transcripts of the other members of the MMP gene family was also investigated in the same samples (FIGS. **4***b–f*). These other members of the MMP family can clearly be separated into two classes, according to their pattern of expression in human breast tumors. The first class includes the 72 kD type IV collagenase (COIV 72K, FIG. **4***d*), stromelysin-1 and -2 (ST1/2, FIG. **4***e*) and pump-1 (PU1, FIG. **4***f*), all of which genes were expressed in both malignant and benign tumors. By contrast, the second class, which includes stromelysin-3 (FIG. **4***a*), type I collagenase (COI, FIG. **4***b*) and the 92 kD type IV collagenase (COIV 92K, FIG. **4***c*) genes, shows over-expression only in breast carcinomas, although only stromelysin-3 was consistently associated therewith.

The patterns of expression were not identical for the three genes of the second class. Type I collagenase RNA transcripts were not detected in the C5, C6, C7 and CIO carcinomas, and the 92 kD type IV collagenase RNA transcripts were not seen in the C7 and CIO samples, but the stromelysin-3 RNA transcripts were clearly detected in all tumors.

Thus, stromelysin-3 appears to be diagnostic of invasive breast carcinomas, while type I collagenase and the 92 kD type IV collagenase may also be specifically involved in breast cancer progression in some cases.

EXAMPLE 5

Expression in Cells from Various Sources

FIG. **5** shows Northern blot analyses of stromelysin-3 RNA in various cell lines and tissues.

(a) Three normal and five metastatic auxillary lymph nodes from patients with breast cancers;

(b) four oestrogen receptor-negative (BT-20, MDA-231, SK-BR3, HBL-100) and four oestrogen receptor-positive (T-47D, BT-474, ZR75-1, MCF-7) breast cancer cell lines;

(c) 10 normal human tissues,

(d) HFL-1 human foetal diploid fibroblasts (ATCC CCL 153) cultured in serum-free medium (1 and 2), in the absence (1) or presence (2) of TPA (10 ng/ml) or cultured in serum-free medium supplemented with 20 μg/ml insulin (3 to 6), in the absence (3) or presence (4) of PDGF (20 ng/ml, British Biotechnology), (5) of EGF (20 ng/ml, Collaborative Research) or (6) of bFGF (10 ng/ml, kindly provided by Pettmann (*FEBS Lett.* 189:102–108 (1985))).

In (a), each lane contained 10 μg of total RNA with the exception of lane 5 (2 μg) and lane 6 (20 μg). In (b)

GeneDX 1009, pg. 588

and (c), each lane contained 8 μg of total RNA, and in (d), each line contained 5 μg of cytoplasmic RNA.

Specifically, in (a), (b) and (c) the blots were made and processed as indicated in FIG. 4 for stromelysin-3. In (d), confluent HFL-1 fibroblasts were kept in serum-free DMEM culture medium. After 24 hrs, fresh medium was added and supplemented or not with TPA or growth factors, as indicated above. After 24 hrs of culture, the cells were harvested and cytoplasmic RNA prepared (Gough, N.M., *Analyt. Biochem.* 173:93–95 (1988).

The blots were then prepared and processed as indicated in FIG. 4 for stromelysin-3, but the autoradiography was for three days.

92 kD collagenase IV RNA transcripts were found in 3 normal and 5 breast cancer metastatic lymph nodes, whereas stromelysin-3 RNA transcripts were detected only in the metastatic nodes (FIG. 5*a*, and data not shown).

In contrast to the results obtained with primary malignant breast tumors and metastatic lymph nodes, no stromelysin-3 RNA transcripts could be detected under similar conditions in eight human breast cancer cell lines, irrespective of their ER status (FIG. 5*b*). Similarly, stromelysin-3 RNA transcripts could not be detected in a number of normal human adult tissues (FIG. 5*c*), with two notable exceptions, uterus and placenta.

Stromelysin-3 is not apparently associated with all cancers, and only low levels of stromelysin-3 RNA transcripts were found in RNA samples from colon, ovary, kidney and lung cancers. However, high levels of expression, comparable to those found in breast cancers, were observed in larynx cancer RNA samples (data not shown).

EXAMPLE 6

Specific Expression in Stromal Cells of Invasive Tumors

The expression of the stromelysin-3 gene in primitive breast carcinomas, but not in a number of established breast cancer cell lines, suggested that the gene was expressed in the stromal cells surrounding the tumor, rather than in the neoplastic cells themselves.

In situ hybridization experiments using a [^{35}S]-labelled stromelysin-3 antisense riboprobe were performed using sections from six carcinomas (tumors C1, C3, C5, C9, C10, referred to as for FIG. 4, and tumor C11, an ER-positive carcinoma not shown in FIG. 4).

Specifically, in situ hybridization was carried out as described by Cox et al. (*Dev. Biol.* 101:485–502 (1984)). Deparaffinised and acid-treated sections (6 μm thick) were proteinase K-treated and hybridized overnight with [^{35}S]-labelled antisense transcripts from a stromelysin-3 cDNA insert (467 bp extending from nucleotides 1128 to 1594) subcloned in Bluescript II (Stratagene). Hybridization was followed by RNase treatment (20 μg/ml, 30 min, 37° C.) and stringent washing (2×SSC, 50% formamide, 60° C., 2 h), prior to autoradiography using NTB2 emulsion (Kodak). Autoradiography was for 15 days. No significant labeling above background was observed under similar conditions using a sense riboprobe (not shown).

FIG. 6 shows the presence of stromelysin-3 RNA transcripts in sections of breast carcinomas and embryo limb bud. a, c, e, g, i and k: bright fields of tissue sections (x100) stained with haematoxylin; b, d, f, h, j and l: the same sections (still stained with haematoxylin) after in situ hybridization with an antisense stromelysin-3 cRNA probe and dark field imaging.

a and b, grade II ductal breast carcinoma (tumor C1, see FIG. 4): infiltrating cancer cells (C) are surrounded by a stroma rich in fusiform cells (S); stromelysin-3 RNA transcripts are most abundant in the stromal cells immediately surrounding the neoplastic epithelial cells. c and d, grade III ductal breast carcinoma (tumor C3, see FIG. 4): multiple islands of infiltrating breast cancer cells (C) are surrounded by stromal cells; the expression of the stromelysin-3 gene is weaker in the central part of most of the stromal trabeculae (S) i.e. in the region which is the farthest away from the neoplastic cells. e and f: ductal carcinoma, (tumor C3, see FIG. 4) together with two normal lobules (N); stromelysin-3 RNA transcripts were detected exclusively in the stroma apposing the infiltrating cancer cells (C), with the exception of a small area rich in lymphocytes (arrow). g and h, ductal carcinoma (tumor C10, see FIG. 4): stromelysin-3 RNA transcripts can be detected above background in the stromal cells surrounding the infiltrating (upper corner, right) but not the in situ (star) breast cancer cells. i and j, ductal carcinoma (tumor C11, ER-positive, grade II, carcinoma); left: carcinoma in situ (stars), no stromelysin-3 RNA transcripts can be detected in the stromal cells; right: infiltrating neoplastic cells surrounded by stromal cells expressing the stromelysin-3 gene. k and l, interdigital region of an 8-week-old human embryo limb bud: stromelysin-3 RNA transcripts are detected in the mesoderm underlying the primitive epiderm, most notably in the interdigital area (M); note that the primitive epiderm (arrows), the cartilage in formation (PC), and the surrounding mesoderm are not labelled.

In all cases, stromelysin-3 RNA transcripts were detected only in the stromal cells surrounding the islands of malignant epithelial cells which formed the invasive component of the tumors (FIG. 6: panels a and b, for tumor C1; panels c, d, e and f for tumor C3; panels g and h for tumor C10; panels i and j, right hand side, for tumor C11; and data not shown for tumors C5 and C9).

In metastatic lymph nodes (same source as C5) the expression of the stromelysin-3 gene was also restricted to the stromal cells surrounding the metastatic epithelial cells (data not shown).

It is particularly notable that, in all cases, the malignant epithelial cells themselves were not labelled, and that the highest levels of expression were observed in the stromal cells in apposition to the malignant cells. In marked contrast, no significant expression could be detected in the stromal cells surrounding in situ carcinoma lesions still surrounded by a basement membrane (panels g and h for tumor C10; panels i and j, left hand side, for tumor G11), while the labelling of stromal cells could be clearly observed in the invasive component of the same tumors (panels g and h for tumor CIO; panels i and j, right hand side, for tumor C11).

Also, no significant expression could be detected in stromal cells located at a distance from the cancerous cells nor in the stromal cells surrounding normal ducts and ductules (e.g. panels e and f). No discrete foci of stromelysin-3 transcripts were detected in sections of the F2 fibroadenoma weakly positive for stromelysin-3 RNA on Northern blots (FIG. 4*a* and not shown).

Both fibroblasts and myofibroblasts are known to be present in the stroma of invasive breast carcinomas (Ahmed, A., *Pathol. Annu.* 25(Pt2):237–286 (1990)).

Using our in situ hybridization technique, it was not possible to determine whether only one or both of these cell types expressed stromelysin-3 gene.

EXAMPLE 7

Stimulation by Growth Factors

The above results indicate that expression of the stromelysin-3 gene in stromal cells is likely to be induced by a diffusible factor secreted by the neoplastic cells. Growth factors such as EGF, FGF and PDGF, as well as some cytokines (IL-1α,β, and TNF-α), and tumor promoters (e.g. TPA) are known to activate the transcription of MMP genes (Kerr et al., *Science* 242:1424–1427 (1988)). It has been also reported that tumor cells from several sources produce a factor(s) which stimulates collagenase I production by human fibroblasts (Lippman et al., *Recent Prog. Hormone Res.* 45:383–440 (1990)). PDGF, FGF and TGF-α activities are known to be secreted by breast cancer cells in vitro (Salomon et al., in *Breast cancer: cellular and molecular biology* (eds., Lippman, M. E. and Dickson, R. B.), pp. 363–389 (Kluwer, Boston, (1988)).

To investigate whether stromelysin-3 gene expression could be modulated by exogenous stimuli, human foetal diploid fibroblasts were grown in the absence or presence of either PDGF, bFGF and EGF, and also of the tumor promoter TPA. Addition of either of these growth factors or of TPA resulted in an increase of stromelysin-3 RNA transcripts in the fibroblasts, the strongest stimulation being observed with bFGF (FIG. 5*d*).

EXAMPLE 8

Stromelysin-3 Expression in the Embryo

As the stromelysin-3 gene was expressed in response to stimulation by growth factors in foetal fibroblasts, we investigated whether the gene might be normally expressed during embryo development.

Stromelysin-3 transcripts were detected in several discrete regions of an 8-week-old human embryo, notably in the interdigital region of the limb buds (FIG. 6, panels k and l, and data not shown), an area known to be associated with programmed cell death at this stage of embryogenesis (Milaire, J., *Organogenesis* (eds., De Haan, R. L. and Ursprung, H.), pp. 283–300 (Holt, Rinehart and Winston, N.Y., 1965)).

The labelling was observed in the embryonic mesoderm underlying the primitive epiderm which remained unlabelled. Notably, the mesenchymal cells located at a distance from the epiblast were also mRNA-negative.

Thus, the finding that the stromelysin-3 gene is expressed during normal embryonic development in an area where tissue remodelling is well documented suggests that expression of stromelysin-3 in breast tumors plays a role in the ECM remodelling processes associated with cancer progression.

EXAMPLE 9

Cloning of mouse cDNA encoding ST3

A probe containing human cDNA encoding ST3 was used to screen a mouse placental cDNA library using standard procedures (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989)). The screening resulted in obtaining a full length mouse cDNA encoding ST3 (FIG. 7).

An analysis of the two sequences revealed a 89% homology in the amino acid sequence of the mature form of ST3 and approximately a 55% homology at the pre and prodomains (FIG. 7).

The pattern of expression in various mouse cells was determined using the methods described in Examples 4–8.

The pattern of ST3 expression in mouse was found to be identical with respect to tissue specificity as that found with human tissue. The highest level of expression was found in placental and uteran tissue.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i i i) NUMBER OF SEQUENCES: 4

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 10..1473

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CCGGGGCGG ATG GCT CCG GCC GCC TGG CTC CGC AGC GCG GCC GCG CGC          48
          Met Ala Pro Ala Ala Trp Leu Arg Ser Ala Ala Ala Arg
            1               5                  10

GCC CTC CTG CCC CCG ATG CTG CTG CTG CTG CTC CAG CCG CCG CCG CTG        96
Ala Leu Leu Pro Pro Met Leu Leu Leu Leu Leu Gln Pro Pro Pro Leu
      15                  20                  25

CTG GCC CGG GCT CTG CCG CCG GAC GTC CAC CAC CTC CAT GCC GAG AGG       144
```

-continued

```
Leu Ala Arg Ala Leu Pro Pro Asp Val His His Leu His Ala Glu Arg
 30                  35                  40                  45

AGG GGG CCA CAG CCC TGG CAT GCA GCC CTG CCC AGT AGC CCG GCA CCT      192
Arg Gly Pro Gln Pro Trp His Ala Ala Leu Pro Ser Ser Pro Ala Pro
                50                  55                  60

GCC CCT GCC ACG CAG GAA GCC CCC CGG CCT GCC AGC AGC CTC AGG CCT      240
Ala Pro Ala Thr Gln Glu Ala Pro Arg Pro Ala Ser Ser Leu Arg Pro
            65                  70                  75

CCC CGC TGT GGC GTG CCC GAC CCA TCT GAT GGG CTG AGT GCC CGC AAC      288
Pro Arg Cys Gly Val Pro Asp Pro Ser Asp Gly Leu Ser Ala Arg Asn
        80                  85                  90

CGA CAG AAG AGG TTC GTG CTT TCT GGC GGG CGC TGG GAG AAG ACG GAC      336
Arg Gln Lys Arg Phe Val Leu Ser Gly Gly Arg Trp Glu Lys Thr Asp
    95                  100                 105

CTC ACC TAC AGG ATC CTT CGG TTC CCA TGG CAG TTG GTG CAG GAG CAG      384
Leu Thr Tyr Arg Ile Leu Arg Phe Pro Trp Gln Leu Val Gln Glu Gln
110                 115                 120                 125

GTG CGG CAG ACG ATG GCA GAG GCC CTA AAG GTA TGG AGC GAT GTG ACG      432
Val Arg Gln Thr Met Ala Glu Ala Leu Lys Val Trp Ser Asp Val Thr
                130                 135                 140

CCA CTC ACC TTT ACT GAG GTG CAC GAG GGC CGT GCT GAC ATC ATG ATC      480
Pro Leu Thr Phe Thr Glu Val His Glu Gly Arg Ala Asp Ile Met Ile
            145                 150                 155

GAC TTC GCC AGG TAC TGG CAT GGG GAC GAC CTG CCG TTT GAT GGG CCT      528
Asp Phe Ala Arg Tyr Trp His Gly Asp Asp Leu Pro Phe Asp Gly Pro
        160                 165                 170

GGG GGC ATC CTG GCC CAT GCC TTC TTC CCC AAG ACT CAC CGA GAA GGG      576
Gly Gly Ile Leu Ala His Ala Phe Phe Pro Lys Thr His Arg Glu Gly
    175                 180                 185

GAT GTC CAC TTC GAC TAT GAT GAG ACC TGG ACT ATC GGG GAT GAC CAG      624
Asp Val His Phe Asp Tyr Asp Glu Thr Trp Thr Ile Gly Asp Asp Gln
190                 195                 200                 205

GGC ACA GAC CTG CTG CAG GTG GCA GCC CAT GAA TTT GGC CAC GTG CTG      672
Gly Thr Asp Leu Leu Gln Val Ala Ala His Glu Phe Gly His Val Leu
                210                 215                 220

GGG CTG CAG CAC ACA ACA GCA GCC AAG GCC CTG ATG TCC GCC TTC TAC      720
Gly Leu Gln His Thr Thr Ala Ala Lys Ala Leu Met Ser Ala Phe Tyr
            225                 230                 235

ACC TTT CGC TAC CCA CTG AGT CTC AGC CCA GAT GAC TGC AGG GGC GTT      768
Thr Phe Arg Tyr Pro Leu Ser Leu Ser Pro Asp Asp Cys Arg Gly Val
        240                 245                 250

CAA CAC CTA TAT GGC CAG CCC TGG CCC ACT GTC ACC TCC AGG ACC CCA      816
Gln His Leu Tyr Gly Gln Pro Trp Pro Thr Val Thr Ser Arg Thr Pro
    255                 260                 265

GCC CTG GGC CCC CAG GCT GGG ATA GAC ACC AAT GAG ATT GCA CCG CTG      864
Ala Leu Gly Pro Gln Ala Gly Ile Asp Thr Asn Glu Ile Ala Pro Leu
270                 275                 280                 285

GAG CCA GAC GCC CCG CCA GAT GCC TGT GAG GCC TCC TTT GAC GCG GTC      912
Glu Pro Asp Ala Pro Pro Asp Ala Cys Glu Ala Ser Phe Asp Ala Val
                290                 295                 300

TCC ACC ATC CGA GGC GAG CTC TTT TTC TTC AAA GCG GGC TTT GTG TGG      960
Ser Thr Ile Arg Gly Glu Leu Phe Phe Phe Lys Ala Gly Phe Val Trp
            305                 310                 315

CGC CTC CGT GGG GGC CAG CTG CAG CCC GGC TAC CCA GCA TTG GCC TCT     1008
Arg Leu Arg Gly Gly Gln Leu Gln Pro Gly Tyr Pro Ala Leu Ala Ser
        320                 325                 330

CGC CAC TGG CAG GGA CTG CCC AGC CCT GTG GAC GCT GCC TTC GAG GAT     1056
Arg His Trp Gln Gly Leu Pro Ser Pro Val Asp Ala Ala Phe Glu Asp
    335                 340                 345

GCC CAG GGC CAC ATT TGG TTC TTC CAA GGT GCT CAG TAC TGG GTG TAC     1104
Ala Gln Gly His Ile Trp Phe Phe Gln Gly Ala Gln Tyr Trp Val Tyr
350                 355                 360                 365
```

GeneDX 1009, pg. 591

-continued

```
GAC GGT GAA AAG CCA GTC CTG GGC CCC GCA CCC CTC ACC GAG CTG GGC        1152
Asp Gly Glu Lys Pro Val Leu Gly Pro Ala Pro Leu Thr Glu Leu Gly
                370                 375                 380

CTG GTG AGG TTC CCG GTC CAT GCT GCC TTG GTC TGG GGT CCC GAG AAG        1200
Leu Val Arg Phe Pro Val His Ala Ala Leu Val Trp Gly Pro Glu Lys
            385                 390                 395

AAC AAG ATC TAC TTC TTC CGA GGC AGG GAC TAC TGG CGT TTC CAC CCC        1248
Asn Lys Ile Tyr Phe Phe Arg Gly Arg Asp Tyr Trp Arg Phe His Pro
        400                 405                 410

AGC ACC CGG CGT GTA GAC AGT CCC GTG CCC CGC AGG GCC ACT GAC TGG        1296
Ser Thr Arg Arg Val Asp Ser Pro Val Pro Arg Arg Ala Thr Asp Trp
    415                 420                 425

AGA GGG GTG CCC TCT GAG ATC GAC GCT GCC TTC CAG GAT GCT GAT GGC        1344
Arg Gly Val Pro Ser Glu Ile Asp Ala Ala Phe Gln Asp Ala Asp Gly
430                 435                 440                 445

TAT GCC TAC TTC CTG CGC GGC CGC CTC TAC TGG AAG TTT GAC CCT GTG        1392
Tyr Ala Tyr Phe Leu Arg Gly Arg Leu Tyr Trp Lys Phe Asp Pro Val
                450                 455                 460

AAG GTG AAG GCT CTG GAA GGC TTC CCC CGT CTC GTG GGT CCT GAC TTC        1440
Lys Val Lys Ala Leu Glu Gly Phe Pro Arg Leu Val Gly Pro Asp Phe
            465                 470                 475

TTT GGC TGT GCC GAG CCT GCC AAC ACT TTC CTC TGACCATGGC TTGGATGCCC      1493
Phe Gly Cys Ala Glu Pro Ala Asn Thr Phe Leu
        480                 485

TCAGGGGTGC TGACCCCTGC CAGGCCACGA ATATCAGGCT AGAGACCCAT GGCCATCTTT      1553

GTGGCTGTGG GCACCAGGCA TGGGACTGAG CCCATGTCTC CTGCAGGGGG ATGGGGTGGG      1613

GTACAACCAC CATGACAACT GCCGGGAGGG CCACGCAGGT CGTGGTCACC TGCCAGCGAC      1673

TGTCTCAGAC TGGGCAGGGA GGCTTTGGCA TGACTTAAGA GGAAGGGCAG TCTTGGGACC      1733

CGCTATGCAG GTCCTGGCAA ACCTGGCTGC CCTGTCTCAT CCCTGTCCCT CAGGGTAGCA      1793

CCATGGCAGG ACTGGGGGAA CTGGAGTGTC CTTGCTGTAT CCCTGTTGTG AGGTTCCTTC      1853

CAGGGGCTGG CACTGAAGCA AGGGTGCTGG GGCCCCATGG CCTTCAGCCC TGGCTGAGCA      1913

ACTGGGCTGT AGGGCAGGGC CACTTCCTGA GGTCAGGTCT TGGTAGGTGC CTGCATCTGT      1973

CTGCCTTCTG GCTGACAATC CTGGAAATCT GTTCTCCAGA ATCCAGGCCA AAAAGTTCAC      2033

AGTCAAATGG GGAGGGGTAT TCTTCATGCA GGAGACCCCA GGCCCTGGAG GCTGCAACAT      2093

ACCTCAATCC TGTCCCAGGC CGGATCCTCC TGAAGCCCTT TTCGCAGCAC TGCTATCCTC      2153

CAAAGCCATT GTAAATGTGT GTACAGTGTG TATAAACCTT CTTCTTCTTT TTTTTTTTTA      2213

AACTGAGGAT TGTCATTAAA CACAGTTGTT TTCTAAAAAA AAA                        2256
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 488 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ala Pro Ala Ala Trp Leu Arg Ser Ala Ala Ala Arg Ala Leu Leu
  1               5                  10                  15

Pro Pro Met Leu Leu Leu Leu Leu Gln Pro Pro Pro Leu Leu Ala Arg
            20                  25                  30

Ala Leu Pro Pro Asp Val His His Leu His Ala Glu Arg Arg Gly Pro
        35                  40                  45

Gln Pro Trp His Ala Ala Leu Pro Ser Ser Pro Ala Pro Ala Pro Ala
    50                  55                  60
```

-continued

```
Thr Gln Glu Ala Pro Arg Pro Ala Ser Ser Leu Arg Pro Pro Arg Cys
 65                  70                  75                  80

Gly Val Pro Asp Pro Ser Asp Gly Leu Ser Ala Arg Asn Arg Gln Lys
                 85                  90                  95

Arg Phe Val Leu Ser Gly Gly Arg Trp Glu Lys Thr Asp Leu Thr Tyr
            100                 105                 110

Arg Ile Leu Arg Phe Pro Trp Gln Leu Val Gln Glu Gln Val Arg Gln
        115                 120                 125

Thr Met Ala Glu Ala Leu Lys Val Trp Ser Asp Val Thr Pro Leu Thr
    130                 135                 140

Phe Thr Glu Val His Glu Gly Arg Ala Asp Ile Met Ile Asp Phe Ala
145                 150                 155                 160

Arg Tyr Trp His Gly Asp Asp Leu Pro Phe Asp Gly Pro Gly Gly Ile
                165                 170                 175

Leu Ala His Ala Phe Phe Pro Lys Thr His Arg Glu Gly Asp Val His
            180                 185                 190

Phe Asp Tyr Asp Glu Thr Trp Thr Ile Gly Asp Asp Gln Gly Thr Asp
        195                 200                 205

Leu Leu Gln Val Ala Ala His Glu Phe Gly His Val Leu Gly Leu Gln
    210                 215                 220

His Thr Thr Ala Ala Lys Ala Leu Met Ser Ala Phe Tyr Thr Phe Arg
225                 230                 235                 240

Tyr Pro Leu Ser Leu Ser Pro Asp Asp Cys Arg Gly Val Gln His Leu
                245                 250                 255

Tyr Gly Gln Pro Trp Pro Thr Val Thr Ser Arg Thr Pro Ala Leu Gly
            260                 265                 270

Pro Gln Ala Gly Ile Asp Thr Asn Glu Ile Ala Pro Leu Glu Pro Asp
        275                 280                 285

Ala Pro Pro Asp Ala Cys Glu Ala Ser Phe Asp Ala Val Ser Thr Ile
    290                 295                 300

Arg Gly Glu Leu Phe Phe Phe Lys Ala Gly Phe Val Trp Arg Leu Arg
305                 310                 315                 320

Gly Gly Gln Leu Gln Pro Gly Tyr Pro Ala Leu Ala Ser Arg His Trp
                325                 330                 335

Gln Gly Leu Pro Ser Pro Val Asp Ala Ala Phe Glu Asp Ala Gln Gly
            340                 345                 350

His Ile Trp Phe Phe Gln Gly Ala Gln Tyr Trp Val Tyr Asp Gly Glu
        355                 360                 365

Lys Pro Val Leu Gly Pro Ala Pro Leu Thr Glu Leu Gly Leu Val Arg
    370                 375                 380

Phe Pro Val His Ala Ala Leu Val Trp Gly Pro Glu Lys Asn Lys Ile
385                 390                 395                 400

Tyr Phe Phe Arg Gly Arg Asp Tyr Trp Arg Phe His Pro Ser Thr Arg
                405                 410                 415

Arg Val Asp Ser Pro Val Pro Arg Arg Ala Thr Asp Trp Arg Gly Val
            420                 425                 430

Pro Ser Glu Ile Asp Ala Ala Phe Gln Asp Ala Asp Gly Tyr Ala Tyr
        435                 440                 445

Phe Leu Arg Gly Arg Leu Tyr Trp Lys Phe Asp Pro Val Lys Val Lys
    450                 455                 460

Ala Leu Glu Gly Phe Pro Arg Leu Val Gly Pro Asp Phe Phe Gly Cys
465                 470                 475                 480

Ala Glu Pro Ala Asn Thr Phe Leu
                485
```

GeneDX 1009, pg. 593

-continued

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2260 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 11..1486

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
CCCGGGGCGG ATG GCA CGG GCC GCC TGT CTC CTC CGC GCG ATT TCG GGG          49
           Met Ala Arg Ala Ala Cys Leu Leu Arg Ala Ile Ser Gly
             1               5                  10

TGC CTC CTG CTC CCG CTG CCT CTG CTG CTC CTG TTG CTG CTG CTC CTG          97
Cys Leu Leu Leu Pro Leu Pro Leu Leu Leu Leu Leu Leu Leu Leu Leu
      15                  20                  25

CCG TCG CCG CTG ATG GCC CGG GCC AGG CCA CCG GAG AGT CAC CGT CAT         145
Pro Ser Pro Leu Met Ala Arg Ala Arg Pro Pro Glu Ser His Arg His
 30                  35                  40                  45

CAC CCT GTG AAG AAA GGG CCT CGG CTC CTG CAT GCA GCT CTG CCT AAT         193
His Pro Val Lys Lys Gly Pro Arg Leu Leu His Ala Ala Leu Pro Asn
                 50                  55                  60

ACC TTG ACA TCT GTC CCC GCG TCT CAT TGG GTC CCT AGT CCT GCC GGT         241
Thr Leu Thr Ser Val Pro Ala Ser His Trp Val Pro Ser Pro Ala Gly
             65                  70                  75

AGC TCC AGG CCT CTA CGA TGT GGT GTG CCC GAC CTG CCT GAT GTA CTG         289
Ser Ser Arg Pro Leu Arg Cys Gly Val Pro Asp Leu Pro Asp Val Leu
         80                  85                  90

AAT GCC CGG AAC CGA CAG AAG CGC TTC GTC CTG TCA GGA GGA CGC TGG         337
Asn Ala Arg Asn Arg Gln Lys Arg Phe Val Leu Ser Gly Gly Arg Trp
      95                 100                 105

GAG AAG ACA GAC CTC ACC TAT AGG ATC CTC CGG TTC CCA TGG CAG CTT         385
Glu Lys Thr Asp Leu Thr Tyr Arg Ile Leu Arg Phe Pro Trp Gln Leu
110                 115                 120                 125

GTA AGG GAG CAA GTC CGG CAG ACA GTG GCA GAG GCC CTC CAG GTA TGG         433
Val Arg Glu Gln Val Arg Gln Thr Val Ala Glu Ala Leu Gln Val Trp
                130                 135                 140

AGT GAA GTG ACC CCA CTC ACT TTC ACT GAG GTG CAC GAG GGA CGC GCT         481
Ser Glu Val Thr Pro Leu Thr Phe Thr Glu Val His Glu Gly Arg Ala
            145                 150                 155

GAC ATC ATG ATC GAC TTC GCA AGG TAC TGG GAT GGT GAC AAC TTG CCG         529
Asp Ile Met Ile Asp Phe Ala Arg Tyr Trp Asp Gly Asp Asn Leu Pro
        160                 165                 170

TTT GAC GGG CCT GGG GGC ATC CTG GCC CAT GGC TTC TTC CCT AAG ACC         577
Phe Asp Gly Pro Gly Gly Ile Leu Ala His Gly Phe Phe Pro Lys Thr
    175                 180                 185

CAC CGA GAA GGG GAT GTC CAC TTT GAC TAT GAT GAA ACT TGG ACT ATT         625
His Arg Glu Gly Asp Val His Phe Asp Tyr Asp Glu Thr Trp Thr Ile
190                 195                 200                 205

GGG GAC AAC CAG GGA ACT GAC CTG CTG CAA GTG GCG GCT CAT GAA TTT         673
Gly Asp Asn Gln Gly Thr Asp Leu Leu Gln Val Ala Ala His Glu Phe
                210                 215                 220

GGC CAT GTT CTG GGG CTA CAA CAC ACC ACA GCA GCT AAG GCC CTC ATG         721
Gly His Val Leu Gly Leu Gln His Thr Thr Ala Ala Lys Ala Leu Met
            225                 230                 235

TCC CCT TTC TAC ACC TTC CGC TAC CCT CTG AGC CTT AGC CCA GAT GAC         769
Ser Pro Phe Tyr Thr Phe Arg Tyr Pro Leu Ser Leu Ser Pro Asp Asp
        240                 245                 250

CGA AGG GGC ATC CAG CAC CTC TAT GGG CGG CCC CAG ATG ACC CCC ACC         817
```

GeneDX 1009, pg. 594

-continued

```
Arg Arg Gly Ile Gln His Leu Tyr Gly Arg Pro Gln Met Thr Pro Thr
    255                 260                 265

TCC CCC GCC CCA ACT TTG AGC TCC CAG GCT GGG ACA GAT ACC AAT GAG      865
Ser Pro Ala Pro Thr Leu Ser Ser Gln Ala Gly Thr Asp Thr Asn Glu
270                 275                 280                 285

ATT GCA CTG CTG GAG CCG GAA ACC CCG CCA GAT GTC TGT GAG ACT TCC      913
Ile Ala Leu Leu Glu Pro Glu Thr Pro Pro Asp Val Cys Glu Thr Ser
                290                 295                 300

TTC GAC GCG GTT TCC ACC ATC CGA GGA GAG CTC TTC TTC TTC AAG GCA      961
Phe Asp Ala Val Ser Thr Ile Arg Gly Glu Leu Phe Phe Phe Lys Ala
            305                 310                 315

GGC TTT GTG TGG AGG CTG CGC AGT GGG CGA CTG CAG CCC GGG TAT CCT     1009
Gly Phe Val Trp Arg Leu Arg Ser Gly Arg Leu Gln Pro Gly Tyr Pro
        320                 325                 330

GCT TTG GCC TCT CGG CAC TGG CAA GGA CTG CCC AGC CCT GTG GAT GCA     1057
Ala Leu Ala Ser Arg His Trp Gln Gly Leu Pro Ser Pro Val Asp Ala
    335                 340                 345

GCT TTT GAG GAT GCC CAG GGC CAG ATT TGG TTC TTC CAA GGT GCT CAG     1105
Ala Phe Glu Asp Ala Gln Gly Gln Ile Trp Phe Phe Gln Gly Ala Gln
350                 355                 360                 365

TAC TGG GTA TAT GAT GGT GAG AAG CCA GTC CTA GGC CCT GCA CCA CTC     1153
Tyr Trp Val Tyr Asp Gly Glu Lys Pro Val Leu Gly Pro Ala Pro Leu
                370                 375                 380

TCC AAG CTG GGC CTG CAA GGG TCC CCA GTT CAT GCC GCC TTG GTC TGG     1201
Ser Lys Leu Gly Leu Gln Gly Ser Pro Val His Ala Ala Leu Val Trp
            385                 390                 395

GGT CCT GAG AAG AAC AAG ATC TAC TTC TTC CGA GGT GGA GAC TAT TGG     1249
Gly Pro Glu Lys Asn Lys Ile Tyr Phe Phe Arg Gly Gly Asp Tyr Trp
        400                 405                 410

CGT TTC CAC CCC AGA ACC CAG CGA GTG GAC AAT CCC GTG CCC CGG CGC     1297
Arg Phe His Pro Arg Thr Gln Arg Val Asp Asn Pro Val Pro Arg Arg
    415                 420                 425

TCC ACT GAC TGG CGA GGG GTA CCT TCT GAG ATT GAT GCT GCC TTC CAG     1345
Ser Thr Asp Trp Arg Gly Val Pro Ser Glu Ile Asp Ala Ala Phe Gln
430                 435                 440                 445

GAT GCT GAG GGC TAT GCC TAC TTC CTT CGT GGC CAT CTC TAC TGG AAG     1393
Asp Ala Glu Gly Tyr Ala Tyr Phe Leu Arg Gly His Leu Tyr Trp Lys
                450                 455                 460

TTT GAT CCC GTG AAG GTG AAG GTC CTA GAA GGC TTT CCT CGC CCC GTA     1441
Phe Asp Pro Val Lys Val Lys Val Leu Glu Gly Phe Pro Arg Pro Val
            465                 470                 475

GGT CCT GAC TTC TTT GAC TGT GCT GAG CCT GCC AAT ACT TTC CGC         1486
Gly Pro Asp Phe Phe Asp Cys Ala Glu Pro Ala Asn Thr Phe Arg
        480                 485                 490

TGACAACACT TTGGATGCAT TCAGGGGTAC TGACTCCTGC CAGGGCACTT AGATCATGTA   1546

AGAGACCCAC AGCCATATCT GTGGCTCTGG CTTCAGGCAT GGGACAGACA GGGCCTATGT   1606

CTCCTCAGGG GAGTGGGTTG GGGTGCAGCC ACTGTTTGTA GGAACGACCA TGCTGTCATG   1666

TCACCTGCCA ACAATTGTCT CAGACTAGGC AAAGGCTTTG GTGTTACTTA AAAATAAGGG   1726

AGGTTTTGGG CTGGCAATAT TTCAGCTACC AATAATCCAC AGTCAGCCTG GTTGCCCAAG   1786

GTCTCCTATC TCTGTCCCTC AATGTAGAAC CCCCACACAA ACTCAGGAAT CACCTGCAAT   1846

GAGGTTCCTG TTGGGAGTGG TGTTGGTAAT GAGATGCCCA GGGTACCATG CTGCCCCTGC   1906

TAAGCAACTG GACCAGTATC TTTCCTGGTA AGTCAGCTCT GGAGAGATAG TGAACTGATC   1966

ATATTCTGGC AGGTGATTCA GACAAGTGCT TCCTGGAACT CAGGCCCCAA GGTACACAGC   2026

CAGCCAAGGA GGCAGCTGCT TCCTCCCAGA GACACGGAAC CTCAAAGGCC CCACATACCT   2086

CACAGCCTTG CCCCAGGCCA TTTCTTTCTG GGGCCCTCTT CCTAGCACAG GTACCCTCTA   2146

AGCCATGTAC ATGTGTATAC AGTGTATAAA GACTTTTTTA AAAAAACAAA AAACCAAACC   2206
```

-continued

CCAAAAAAGC CAAGACTGTC ATTAAACATG AGTGTTTTCT AAAAAAAAAA AAAA 2260

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 492 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ala Arg Ala Ala Cys Leu Leu Arg Ala Ile Ser Gly Cys Leu Leu
1 5 10 15

Leu Pro Leu Pro Leu Leu Leu Leu Leu Leu Leu Leu Leu Pro Ser Pro
20 25 30

Leu Met Ala Arg Ala Arg Pro Pro Glu Ser His Arg His His Pro Val
35 40 45

Lys Lys Gly Pro Arg Leu Leu His Ala Ala Leu Pro Asn Thr Leu Thr
50 55 60

Ser Val Pro Ala Ser His Trp Val Pro Ser Pro Ala Gly Ser Ser Arg
65 70 75 80

Pro Leu Arg Cys Gly Val Pro Asp Leu Pro Asp Val Leu Asn Ala Arg
85 90 95

Asn Arg Gln Lys Arg Phe Val Leu Ser Gly Gly Arg Trp Glu Lys Thr
100 105 110

Asp Leu Thr Tyr Arg Ile Leu Arg Phe Pro Trp Gln Leu Val Arg Glu
115 120 125

Gln Val Arg Gln Thr Val Ala Glu Ala Leu Gln Val Trp Ser Glu Val
130 135 140

Thr Pro Leu Thr Phe Thr Glu Val His Glu Gly Arg Ala Asp Ile Met
145 150 155 160

Ile Asp Phe Ala Arg Tyr Trp Asp Gly Asp Asn Leu Pro Phe Asp Gly
165 170 175

Pro Gly Gly Ile Leu Ala His Gly Phe Phe Pro Lys Thr His Arg Glu
180 185 190

Gly Asp Val His Phe Asp Tyr Asp Glu Thr Trp Thr Ile Gly Asp Asn
195 200 205

Gln Gly Thr Asp Leu Leu Gln Val Ala Ala His Glu Phe Gly His Val
210 215 220

Leu Gly Leu Gln His Thr Thr Ala Ala Lys Ala Leu Met Ser Pro Phe
225 230 235 240

Tyr Thr Phe Arg Tyr Pro Leu Ser Leu Ser Pro Asp Asp Arg Arg Gly
245 250 255

Ile Gln His Leu Tyr Gly Arg Pro Gln Met Thr Pro Thr Ser Pro Ala
260 265 270

Pro Thr Leu Ser Ser Gln Ala Gly Thr Asp Thr Asn Glu Ile Ala Leu
275 280 285

Leu Glu Pro Glu Thr Pro Pro Asp Val Cys Glu Thr Ser Phe Asp Ala
290 295 300

Val Ser Thr Ile Arg Gly Glu Leu Phe Phe Phe Lys Ala Gly Phe Val
305 310 315 320

Trp Arg Leu Arg Ser Gly Arg Leu Gln Pro Gly Tyr Pro Ala Leu Ala
325 330 335

Ser Arg His Trp Gln Gly Leu Pro Ser Pro Val Asp Ala Ala Phe Glu
340 345 350

Asp Ala Gln Gly Gln Ile Trp Phe Phe Gln Gly Ala Gln Tyr Trp Val

GeneDX 1009, pg. 596

-continued

```
          355                 360                 365

Tyr Asp Gly Glu Lys Pro Val Leu Gly Pro Ala Pro Leu Ser Lys Leu
    370                 375                 380

Gly Leu Gln Gly Ser Pro Val His Ala Ala Leu Val Trp Gly Pro Glu
385                 390                 395                 400

Lys Asn Lys Ile Tyr Phe Phe Arg Gly Gly Asp Tyr Trp Arg Phe His
                405                 410                 415

Pro Arg Thr Gln Arg Val Asp Asn Pro Val Pro Arg Arg Ser Thr Asp
            420                 425                 430

Trp Arg Gly Val Pro Ser Glu Ile Asp Ala Ala Phe Gln Asp Ala Glu
        435                 440                 445

Gly Tyr Ala Tyr Phe Leu Arg Gly His Leu Tyr Trp Lys Phe Asp Pro
    450                 455                 460

Val Lys Val Lys Val Leu Glu Gly Phe Pro Arg Pro Val Gly Pro Asp
465                 470                 475                 480

Phe Phe Asp Cys Ala Glu Pro Ala Asn Thr Phe Arg
                485                 490
```

We claim:

1. A cDNA encoding either mouse (SEQ. ID No. 3) or human (SEQ. ID No. 1) stromelysin-3, or fragment thereof.

2. The vector of claim 1 wherein said vector directs the expression of said cDNA.

3. The cDNA of claim 1 wherein said cDNA is contained in a vector.

4. The vector of claims 3 or 2 wherein said vector is selected from the group consisting of plasmid, phage, cosmid, retrovirus and baculovirus.

5. The cDNA of claim 1 wherein said cDNA encodes human stromelysin-3.

6. The cDNA of claim 5 wherein said cDNA is depicted in Sequence ID No. 1.

7. The cDNA of claim 1 wherein said cDNA encodes mouse stromelysin-3.

8. The cDNA of claim 7 wherein said cDNA is depicted in Sequence ID No. 3.

* * * * *

GeneDX 1009, pg. 597

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12P 19/34, C12Q 1/68 G01N 33/48, 33/00, 33/546 G01N 33/566, C07H 15/12, 17/00	A1	(11) International Publication Number: WO 91/09964 (43) International Publication Date: 11 July 1991 (11.07.91)

(21) International Application Number: PCT/US90/07314

(22) International Filing Date: 19 December 1990 (19.12.90)

(30) Priority data:
460,981 4 January 1990 (04.01.90) US

(71) Applicant: THE JOHNS HOPKINS UNIVERSITY [US/US]; 720 Rutland Avenue, Baltimore, MD 21205 (US).

(72) Inventor: VOGELSTEIN, Bert ; 3700 Breton Way, Baltimore, MD 21208 (US).

(74) Agents: HOSCHEIT, Dale, H. et al.; Banner, Birch, McKie & Beckett, 1001 G Street, NW, Eleventh Floor, Washington, DC 20001-4597 (US).

(81) Designated States: AT, AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH, CH (European patent), CM (OAPI patent), DE, DE (European patent), DK, DK (European patent), ES, ES (European patent), FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC, MG, ML (OAPI patent), MR (OAPI patent), MW, NL, NL (European patent), NO, RO, SD, SE, SE (European patent), SN (OAPI patent), SU, TD (OAPI patent), TG (OAPI patent).

Published
With international search report.

(54) Title: GENE DELETED IN COLORECTAL CANCER OF HUMANS

(57) Abstract

A new human gene termed DCC is disclosed. Methods and kits are provided for assessing mutations of the DCC gene in human tissues and body samples. Insertion, deletion, and point mutations in DCC are observed in human tumor cells. Normal tissues express DCC while most colorectal cancers do not. Loss of wild-type DCC genes is associated with neoplastic progression and a diminished life expectancy.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**ES**	Spain	**MG**	Madagascar
AU	Australia	**FI**	Finland	**ML**	Mali
BB	Barbados	**FR**	France	**MN**	Mongolia
BE	Belgium	**GA**	Gabon	**MR**	Mauritania
BF	Burkina Faso	**GB**	United Kingdom	**MW**	Malawi
BG	Bulgaria	**GN**	Guinea	**NL**	Netherlands
BJ	Benin	**GR**	Greece	**NO**	Norway
BR	Brazil	**HU**	Hungary	**PL**	Poland
CA	Canada	**IT**	Italy	**RO**	Romania
CF	Central African Republic	**JP**	Japan	**SD**	Sudan
CG	Congo	**KP**	Democratic People's Republic of Korea	**SE**	Sweden
CH	Switzerland			**SN**	Senegal
CI	Côte d'Ivoire	**KR**	Republic of Korea	**SU**	Soviet Union
CM	Cameroon	**LI**	Liechtenstein	**TD**	Chad
CS	Czechoslovakia	**LK**	Sri Lanka	**TG**	Togo
DE	Germany	**LU**	Luxembourg	**US**	United States of America
DK	Denmark	**MC**	Monaco		

GENE DELETED IN COLORECTAL CANCER OF HUMANS

The U.S. Government has a paid-up license in this invention and the right in limited circumstances to require the patent owner to license others on reasonable terms as provided for by the terms of grant numbers GM07309, GM07184, CA09243, and CA35494, awarded by the National Institutes of Health.

TECHNICAL AREA OF THE INVENTION

The invention relates to the area of cancer diagnostics. More particularly, the invention relates to detection of the loss and/or alteration of wild-type DCC genes in tumor tissues.

BACKGROUND OF THE INVENTION

Recent studies have elucidated several genetic alterations that occur during the development of colorectal tumors, the most common of which are deletions of the short arm of chromosome 17 (17p) and the long arm of chromosome 18 (18q). Vogelstein, et al., Science, vol. 244, p. 207 (1989); Fearon, et al., Science, vol. 238, p. 193 (1987); Muleris, et al., Ann. Genet. (Paris), vol. 28, p. 206 (1985); Monpezat, et al., Int. J. Cancer, vol. 41, p. 404 (1988). While some genetic alterations such as RAS gene mutations appear to occur relatively early during colorectal tumor development, chromosome 18q deletions are often late events associated with the transition from Class II to Class III adenomas or the transition from the benign (adenomatous) to the malignant (carcinomatous) state, (Vogelstein et al., New England Journal of Medicine, Vol. 319, p. 525, 1988) and appear to be related to metastasis and decreased survival time (Kern, et al., JAMA, vol. 261, pp. 13099-13103, 1989). Because carcinomas are often lethal, while the precursor adenomas are uniformly curable, the delineation of the molecular events mediating this transition are of considerable importance.

Allelic deletions have been reported to encompass a large area of chromosome 18q. (Vogelstein, et al., ibid.) This area is known to contain the BCL-2 gene (Tsujimoto, et al., Science, vol. 226, p. 1097 (1984); and Cleary, et al., Cell, vol. 47, p. 19 (1986),) the gastrin-releasing peptide gene (Spindel, et al., Proc. Natl. Acad. Sci., USA, vol. 81, p. 5699 (1984),) and the cellular homologue of the YES-1 oncogene (Semba, et al., Science, vol. 227, p. 1038 (1985) and Yoshida, et al., Cytogenet. Cell Genet., vol. 40, p. 786 (1985)). All of these genes are known to be associated with cancers. If a particular region of the chromosome is the target of the deletions, i.e., it is involved in the neoplastic process, precise delineation of the region is necessary to provide methods of diagnosis as well as therapy.

According to the model of Knudson for tumorigenesis (Cancer Research, vol. 45, p. 1482, 1985), there are tumor suppressor genes in all normal cells which, when they become non-functional due to mutation, cause neoplastic development. Evidence for this model has been found in the cases of retinoblastoma and colorectal tumors. The implicated suppressor genes in those tumors, RB and p53, were found to be deleted or altered in many cases of the tumors studied. There is a need in the art of cancer diagnosis and therapy to find other suppressor genes involved in tumorigenesis, so that defects in the suppressor genes or effected cells can be detected and the defects cured to abate or reverse the neoplastic processes.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide a method for diagnosing and prognosing a neoplastic tissue of a human.

It is another object of the invention to provide a method of supplying wild-type DCC gene function to a cell which has lost said gene function.

It is yet another object of the invention to provide a kit for determination of the nucleotide sequence of the DCC gene by the polymerase chain reaction.

It is still another object of the invention to provide nucleic acid probes for detection of mutations in the human DCC gene.

It is another object of the invention to provide a method of detecting genetic predisposition to cancer.

It is still another object of the invention to provide a cDNA molecule encoding the DCC gene product.

It is yet another object of the invention to provide a preparation of the human DCC protein.

These and other objects of the invention are provided by one or more of the embodiments which are described below. In one embodiment of the present invention a method of diagnosing or prognosing a neoplastic tissue of a human is provided comprising: isolating a tissue from a human; and detecting loss of wild-type DCC genes or their expression products from said tissue, said loss indicating neoplasia of the tissue and correlating with metastasis and early death.

In another embodiment of the present invention a method is provided for supplying wild-type DCC gene function to a cell which has lost said gene function by virtue of a mutation in the DCC gene, comprising: introducing a wild-type DCC gene into a cell which has lost said gene function such that said wild-type gene is expressed in the cell.

In another embodiment a method of supplying wild-type DCC gene function to a cell is provided comprising introducing a portion of a wild-type DCC gene into a cell which has lost said gene function such that said portion is expressed in the cell, said portion encoding a part of the DCC protein which is required for non-neoplastic growth of said cell.

In yet another embodiment a kit is provided for determination of the nucleotide sequence of the DCC gene by polymerase chain reaction. The kit comprises: a set of pairs of single stranded DNA primers, the sequence of said set derived from chromosome 18q, said set allowing synthesis of all nucleotides of the DCC gene coding sequences.

In still another embodiment of the invention a nucleic acid probe is provided which is complementary to human wild-type DCC gene coding sequences and which can form mismatches with mutant

DCC genes, thereby allowing their detection by enzymatic or chemical cleavage or by shifts in electrophoretic mobility.

In another embodiment a particular nucleic acid probe is provided which hybridizes to a DCC intron which is subject to insertional mutations in tumor cells.

In still another embodiment of the invention a method are provided for detecting the presence of a neoplastic tissue in a human. The methods comprise isolating a body sample from a human; detecting in said sample loss of a wild-type DCC gene sequence or wild-type DCC expression product, said loss indicating the presence of a neoplastic tissue in the human.

In yet another embodiment a method is provided of detecting genetic predisposition to cancer in a human, comprising: isolating a human sample selected from the group consisting of blood and fetal tissue; detecting loss of wild-type DCC gene coding sequences or their expression products from the sample, said loss indicating predisposition to cancer.

In still other embodiments a cDNA molecule is provided which comprises the coding sequence of the DCC gene.

In even another embodiment a preparation of the human DCC protein is provided which is substantially free of other human proteins. The amino acid sequence of the protein is shown in Figure 5.

The present invention provides the art with the information that the DCC gene, a heretofore unknown gene is, in fact, the target of deletional, insertional, and point mutational alterations on chromosome 18q and that these alterations are associated with the process of tumorigenesis. This information allows highly specific assays to be performed to assess the neoplastic status of a particular tissue or the potential neoplastic status of an individual.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 presents a map of the chromosomal walk and cross-hybridizing fragments in the DCC region. The DNA region of approximately 370 kb, cloned in 30 rounds of walking, is shown; only the maximal walk for each of the rounds is shown. The map position at "O" marks the location of p15-65. An EcoRI map for the region

was constructed and the EcoRI fragments that hybridized at reduced stringency (55°C) to rodent, chicken, or Xenopus DNA samples are indicated by solid boxes and alphabetical letters (A-X). Human fragments G, I, J, K, M, O, and P were used to isolate rat clones. The minimal region of cross-hybridization was identified and sequenced for both human and rat fragments. The locations of the EcoRI fragments that hybridized to human cDNA clones are indicated by arrows.

Figure 2 shows autoradiographs of Southern blots of DNA from mouse (M), rat (R), hamster (H), chicken (C), Xenopus laevis (F), and S. cerevisiae (Y) hybridized to human fragment O (in the panel on the left) and human fragment P (in the panel on the right). The sizes of the corresponding molecular weight markers in kilobases are indicated between the two panels.

Figure 3 shows the nucleotide sequence, predicted amino acid (AA) sequence (in single letter code) and splice acceptor and donor features of some human fragments and their corresponding rat homologues. Nucleotide sequences of the rat fragments were identical to those of the human except where indicated. When the predicted amino acid sequence of the rat differed from that of the human, the human sequence is given on the left of the slash and the rat sequence on the right. The regions in human fragment O and P where no corresponding sequence was available for the rat are indicated by the dashed lines. The predicted intron-exon boundaries are indicated by solid arrowheads, and the potential lariat signal preceding the splice acceptor sequence is overlined.

Figure 4 shows the nucleotide sequence derived from overlapping cDNA clones, prepared from mRNA of either the H82 cell line or from normal human brain. The methionine codon initiating the open reading frame (ORF) is designated as amino acid 202 and the last amino acid is numbered 1648.

Figure 5 demonstrates expression of DCC in human tumors and colorectal tumor cell lines. RNA was isolated and used as template to prepare a first strand of cDNA. The cDNA samples were used for polymerase chain reaction (PCR) analysis. The PCR products were separated by electrophoresis through an agarose gel, and after

Southern transfer, hybridized to a radioactively-labeled subclone of fragment P. RNA was isolated from normal human brain (lane 1), four different normal colonic mucosa specimens (lanes 2-5) or colorectal carcinoma cell lines (lane 6-16).

Figure 6A shows an autoradiograph of a Southern blot prepared from DNA of peripheral blood lymphocytes of two normal individuals hybridized to a radiolabeled cDNA probe (pDCC 1.65). pDCC 1.65 contained nucleotides 591 to 2250 of the cDNA shown in Figure 5. The sizes of the fragments in kb are indicated to the right of the blot. A 0.45 kb EcoRI fragment detected by pDCC 1.65 migrated off the gel and is thus not in the figure.

Figure 6B shows a Southern blot using probe pKC 430 (from the 3' region of the 1.65 kb cDNA) to normal (N) and cryostat sections of tumor (T) DNA from S115. This probe contained nucleotides 1760 to 2205 of the cDNA shown in Figure 5.

Figure 7A shows Southern blots of DNA from normal and tumor DNA samples digested with EcoRI and Eco0109. The DNA was then hybridized to a 0.4 kb genomic fragment which contained the exon from fragment P. Molecular weights in kb are indicated at the left; arrowheads indicate the tumor DNA fragments with insertions. DNA samples were derived from: lanes 1 and 2: peripheral blood lymphocytes from two normal individuals; lanes 3 and 4: non-neoplastic colon and colorectal carcinoma xenograft, respectively, from patient Cx7; lanes 5 and 6: non-neoplastic colon and cryostat sections of colorectal carcinoma, respectively, from patient S175; lane 7: carcinoma xenograft from patient Cx10; lane 8: colorectal carcinoma cell line RKO; lane 9: colorectal tumor cell line Vaco6; lane 10: colorectal carcinoma cell line SNU630; lane 11: colorectal carcinoma cell line SW48; lane 12: colorectal carcinoma cell line HCT16.

Figure 7B shows the sequence of the 170 bp XbaI-Eco0109 fragment to which the insertions shown in panel A were localized. The numbers above the sequence indicate the distance in bp from the 3' end of the exon contained in fragment P. The XbaI and Eco0109 restriction sites are indicated. The two regions of TA repeats are

overlined, and the 130 bp region of alternating purine-pyrimidine sequence is contained between the arrowheads.

Figure 8A shows a comparison of the sequence homology of the four immunoglobulin-like domains of DCC with one another and with chicken N-CAM [N-CAM(c)] and mouse N-CAM [N-CAM(m)]. The N-CAM(c) and N-CAM(m) sequences shown represent the consensus of the five Ig-like domains present in each protein; if no consensus was present at a particular position of N-CAM(c) or N-CAM(m) (i.e., if no two domains contained the identical residue), then the position is indicated by an X. Spaces inserted for alignment are indicated by a dash. The conserved cysteines thought to be involved in intra-domain disulfide pairing are indicated by solid triangles; other amino acid residues highly conserved in N-CAM and other similar Ig-like domains of the C2 class (Williams, et al., Ann. Rev. Immunol., vol. 6, p. 381, (1988)) are noted by open triangles. Sequences were aligned by inspection to give the greatest overall match. Residues in two or more of the DCC domains were boxed if they were identical. The N-CAM(c) and N-CAM(m) consensus sequences were boxed if they matched the DCC consensus.

Figure 8B shows a comparison of the sequence homology between DCC and chicken and mouse N-CAM in the fibronectin-type III-related regions.

DETAILED DESCRIPTION

It is a discovery of the present invention that mutational events associated with tumorigenesis occur in a previously unknown gene on chromosome 18q named here the DCC (Deleted in Colorectal Carcinomas) gene. Although it was previously known that deletion of alleles on chromosome 18q were common in certain types of cancers, it was not known that the target gene of these deletions was the DCC gene. Further it was not known that other types of mutational events in the DCC gene are also associated with cancers. The mutations of the DCC gene involve gross rearrangements such as insertions and deletions. However point mutations, which lead to loss of expression of wild-type DCC have also been observed.

According to the diagnostic and prognostic method of the present invention, loss of the wild-type gene is detected. The loss may be due to either insertional, deletional or point mutational events. If only a single allele is mutated, an early neoplastic state is indicated. However, if both alleles are mutated then a late neoplastic state is indicated. The finding of DCC mutations thus provides both diagnostic and prognostic information. A DCC allele which is not deleted (e.g., that on the sister chromosome to a chromosome carrying a DCC deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. It is believed that most mutations found in tumor tissues will be those leading to greatly decreased expression of the DCC gene product. However, mutations leading to non-functional gene products would also lead to a cancerous state. Point mutational events may occur in regulatory regions, such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the DCC gene product.

In order to detect the loss of the wild-type DCC gene in a tissue, it is helpful to isolate the tissue free from surrounding normal tissues. Means for enriching a tissue preparation for tumor cells are known in the art. For example, the tissue may be isolated from paraffin or cryostat sections. Cancer cells may also be separated from normal cells by flow cytometry. These as well as other techniques for separating tumor from normal cells are well known in the art. If the tumor tissue is highly contaminated with normal cells, detection of mutations is more difficult.

Detection of point mutations may be accomplished by molecular cloning of the allele (or alleles) present in the tumor tissue and sequencing that allele(s) using techniques well known in the art. Alternatively, the polymerase chain reaction can be used to amplify gene sequences directly from a genomic DNA preparation from the tumor tissue. The DNA sequence of the amplified sequences can then be determined. The polymerase chain reaction itself is well known in the art. See, e.g., Saiki et al., Science, Vol. 239, p. 487, 1988; U.S.

4,683,203; and U.S. 4,683,195. Specific primers which can be used in order to amplify the gene will be discussed in more detail below. Insertions and deletions of genes can also be detected by these techniques. In addition, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to score loss of an allele or an insertion in a polymorphic fragment. Other techniques for detecting insertions and deletions as are known in the art can be used.

Loss of wild-type genes can also be detected on the basis of the loss of a wild-type expression product of the gene. Such expression products include both the mRNA as well as the protein product itself. Point mutations may be detected by amplifying and sequencing the mRNA or via molecular cloning of cDNA made from the mRNA. The sequence of the cloned cDNA can be determined using DNA sequencing techniques which are well known in the art. The cDNA can also be sequenced via the polymerase chain reaction (PCR) which will be discussed in more detail below.

Mismatches, according to the present invention are hybridized nucleic acid duplexes which are not 100% homologous. The lack of total homology may be due to deletions, insertions, substitutions or frameshift mutations. Mismatch detection can be used to detect point mutations in the gene or its mRNA product. While these techniques are less sensitive than sequencing, they are simpler to perform on a large number of tumors. An example of a mismatch cleavage technique is the RNase protection method, which is described in detail in Winter et al., Proc. Natl. Acad. Sci. USA, Vol. 82, p. 7575, 1985 and Meyers et al., Science, Vol. 230, p. 1242, 1985. In the practice of the present invention the method involves the use of a labeled riboprobe which is complementary to the human wild-type gene coding sequence. The riboprobe and either mRNA or DNA isolated from the tumor tissue are annealed (hybridized) together and subsequently digested with the enzyme RNase A which is able to detect some mismatches in a duplex RNA structure. If a mismatch is detected by RNase A, it cleaves at the site of the mismatch. Thus, when the annealed RNA preparation is separated on an electrophoretic gel

matrix, if a mismatch has been detected and cleaved by RNase A, an RNA product will be seen which is smaller than the full-length duplex RNA for the riboprobe and the mRNA or DNA. The riboprobe need not be the full length of the DCC mRNA or gene but can be a segment of either. If the riboprobe comprises only a segment of the DCC mRNA or gene it will be desirable to use a number of these probes to screen the whole mRNA sequence for mismatches.

In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. See, e.g., Cotton et al., Proc. Natl. Acad. Sci. USA, vol. 85, 4397, 1988; and Shenk et al., Proc. Natl. Acad. Sci. USA, vol. 72, p. 989, 1975. Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. See, e.g., Cariello, Human Genetics, vol. 42, p. 726, 1988. With either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR (see below) before hybridization. Changes in DNA of the DCC gene can also be detected using Southern hybridization, especially if the changes are gross rearrangements, such as deletions and insertions.

DNA sequences of the DCC gene from the tumor tissue which have been amplified by use of polymerase chain reaction may also be screened using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the DCC gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length, corresponding to a portion of the DCC gene sequence. By use of a battery of such allele-specific probes, the PCR amplification products can be screened to identify the presence of a previously identified mutation in the DCC gene. Hybridization of allele-specific probes with amplified DCC sequences can be performed, for example, on a nylon filter. Hybridization to a particular probe under stringent hybridization conditions indicates the presence of the same mutation in the tumor tissue as in the allele-specific probe.

Loss of DCC mRNA expression can be detected by any technique known in the art. These include Northern analysis, PCR

amplification and RNase protection. Diminished mRNA expression indicates a loss of the wild-type DCC gene.

Loss of wild-type DCC genes can also be detected by screening for loss of wild-type DCC protein. For example, monoclonal antibodies immunoreactive with DCC can be used to screen a tissue. Lack of antigen would indicate a DCC mutation. Antibodies specific for mutant alleles could also be used to detect mutant DCC gene product. Such immunological assays could be done in any convenient format known in the art. These include Western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered DCC protein can be used to detect loss of wild-type DCC genes. Finding a mutant DCC gene product indicates loss of a wild-type DCC gene.

Mutant DCC genes or gene products can also be detected in other human body samples, such as, serum, stool, urine and sputum. The same techniques discussed above for detection of mutant DCC genes or gene products in tissues can be applied to other body samples. Cancer cells are sloughed off from tumors and appear in such body samples. In addition, the DCC gene product itself may be secreted into the extracellular space and found in these body samples even in the absence of cancer cells. By screening such body samples, a simple early diagnosis can be achieved for many types of cancers. In addition, the progress of chemotherapy or radiotherapy can be monitored more easily by testing such body samples for mutant DCC genes or gene products.

The methods of diagnosis of the present invention are applicable to any tumor in which DCC has a role in tumorigenesis. The diagnostic method of the present invention is useful for clinicians so that they can decide upon an appropriate course of treatment. For example, a tumor displaying loss of both DCC alleles might suggest a more aggressive therapeutic regimen than a tumor displaying loss of only one DCC allele.

The primer kit of the present invention is useful for determination of the nucleotide sequence of the DCC gene using the polymerase chain reaction. The kit comprises a set of pairs of single stranded DNA primers which can be annealed to sequences within or

surrounding the DCC gene on chromosome 18q in order to prime amplifying DNA synthesis of the DCC gene itself. The complete set allows synthesis of all of the nucleotides of the DCC gene coding sequences, i.e., the exons. The set of primers preferably allows synthesis of both intron and exon sequences, as a number of DCC mutations have been found in a DCC intron. The kit can also contain DNA polymerase, preferably Taq polymerase, and suitable reaction buffers. Such components are known in the art.

In order to facilitate subsequent cloning of amplified sequences, primers may have restriction enzyme sites appended to their 5' ends. Thus, all nucleotides of the primers are derived from DCC sequences or sequences adjacent to DCC except the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art. The primers themselves can be synthesized using techniques which are well known in the art. Generally, the primers can be made using synthesizing machines which are commercially available. Given the sequence of the DCC open reading frame shown in Figure 4, design of particular primers is well within the skill of the art.

The nucleic acid probes provided by the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA and in the RNase protection method for detecting point mutations already discussed above. The probes can be used to detect PCR amplification products. They may also be used to detect mismatches with the DCC gene or mRNA using other techniques. Mismatches can be detected using either enzymes (e.g., S1 nuclease), chemicals (e.g., hydroxylamine or osmium tetroxide and piperidine), or changes in electrophoretic mobility of mismatched hybrids as compared to totally matched hybrids. These techniques are known in the art. See, Cotton, supra, Shenk, supra, Myers, supra, Winter, supra, and Novack et al., Proc. Natl. Acad. Sci. USA, vol. 83, p. 586, 1986. Generally, the probes are complementary to DCC gene coding sequences, although probes to certain introns are also contemplated. One probe in particular hybridizes to the XbaI-Eco0109 fragment located 165 bp downstream from the DCC exon in fragment P

shown in Figure 3. Another probe hybridizes to fragment P itself, and is a probe for DCC coding sequences. An entire battery of nucleic acid probes is used to compose a kit for detecting loss of wild-type DCC genes. The kit allows for hybridization to the entire DCC gene. The probes may overlap with each other or be contiguous.

If a riboprobe is used to detect mismatches with mRNA, it is complementary to the mRNA of the human wild-type DCC gene. The riboprobe thus is an anti-sense probe in that it does not code for the DCC protein because it is of the opposite polarity to the sense strand. The riboprobe generally will be radioactively labeled which can be accomplished by any means known in the art. If the riboprobe is used to detect mismatches with DNA it can be of either polarity, sense or anti-sense. Similarly, DNA probes also may be used to detect mismatches.

Nucleic acid probes may also be complementary to mutant alleles of DCC gene. These are useful to detect similar mutations in other patients on the basis of hybridization rather than mismatches. These are discussed above and referred to as allele-specific probes. As mentioned above, the DCC probes can also be used in Southern hybridizations to genomic DNA to detect gross chromosomal changes such as deletions and insertions. The probes can also be used to select cDNA clones of DCC genes from tumor and normal tissues. In addition, the probes can be used to detect DCC mRNA in tissues to determine if expression is diminished as a result of loss of wild-type DCC genes. Provided with the DCC coding sequence shown in Figure 4, design of particular probes is well within the skill of the ordinary artisan.

According to the present invention a method is also provided of supplying wild-type DCC function to a cell which carries mutant DCC alleles. The wild-type DCC gene or a part of the gene may be introduced into the cell in a vector such that the gene remains extrachromosomal. In such a situation the gene will be expressed by the cell from the extrachromosomal location. If a gene portion is introduced and expressed in a cell carrying a mutant DCC allele, the gene portion should encode a part of the DCC protein which is required for

non-neoplastic growth of the cell. More preferred is the situation where the wild-type DCC gene or a part of it is introduced into the mutant cell in such a way that it recombines with the endogenous mutant DCC gene present in the cell. Such recombination requires a double recombination event which results in the correction of the DCC gene mutation. Vectors for introduction of genes both for recombination and for extrachromosomal maintenance are known in the art and any suitable vector may be used. Methods for introducing DNA into cells such as electroporation, calcium phosphate co-precipitation and viral transduction are known in the art and the choice of method is within the competence of the routineer. Cells transformed with the wild-type DCC-gene can be used as model systems to study cancer remission and drug treatments which promote such remission.

Polypeptides which have DCC activity can be supplied to cells which carry mutant or missing DCC alleles. The sequence of the DCC protein is disclosed in Figure 4. Protein can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors. Alternatively, DCC can be extracted from DCC-producing mammalian cells such as brain cells. In addition, the techniques of synthetic chemistry can be employed to synthesize DCC protein. Any of such techniques can provide the preparation of the present invention which comprises the DCC gene product having the sequence shown in Figure 4. The preparation is substantially free of other human proteins. This is most readily accomplished by synthesis in a microorganism or in vitro. Active DCC molecules can be introduced into cells by microinjection or by use of liposomes, for example. Alternatively, some such active molecules may be taken up by cells, actively or by diffusion. Extracellular application of DCC gene product may be sufficient to effect tumor growth, as DCC may act at the cell surface like its homologues, the neural cell adhesion molecules. Supply of molecules with DCC activity should lead to a partial reversal of the neoplastic state. Other molecules with DCC activity may also be used to effect such a reversal.

The present invention also provides a preparation of antibodies immunoreactive with human DCC protein. The antibodies may be polyclonal or monoclonal and may be raised against native DCC protein or DCC fusion proteins. The antibodies should be immunoreactive with DCC epitopes, preferably epitopes not present on other human proteins. In a preferred embodiment of the invention the antibodies will immunoprecipitate DCC proteins from solution as well as react with DCC protein on Western blots of polyacrylamide gels. Techniques for raising and purifying antibodies are well known in the art and any such techniques may be chosen to achieve the preparation of the invention.

Predisposition to cancers can be ascertained by testing normal tissues of humans for mutations of DCC gene. For example, a person who has inherited a germline DCC mutation would be prone to develop cancers. This can be determined by testing DNA from any tissue of the person's body. Most simply, blood can be drawn and DNA extracted from the cells of the blood. In addition, prenatal diagnosis can be accomplished by testing fetal cells or amniotic fluid for mutations of the DCC gene. Loss of a wild-type DCC allele, whether for example, by point mutation or by deletion, can be detected by any of the means discussed above.

Molecules of cDNA according to the present invention are intron-free, DCC gene coding molecules. They can be made by reverse transcriptase using the DCC mRNA as a template. These molecules can be propagated in vectors and cell lines as is known in the art. Such molecules have the sequence shown in Figure 4. The cDNA can also be made using the techniques of synthetic chemistry given the sequence disclosed herein.

The following are provided for exemplification purposes only and are not intended to limit the scope of the invention which has been described in broad terms above.

Example 1:

This example demonstrates that the locus on chromosome 18q which is the subject of frequent allelic deletions and of somatic mutations is expressed in cell lines of lung and colorectal carcinoma.

Tumor S123 was found to have a different hybridization pattern from normal colonic mucosa of the same patient when probed with probe p15-65. The tumor displayed a heterozygous pattern with two MspI alleles of 7.8 and 10.5 kb. The normal colonic mucosa was homozygous for the 7.8 kb allele.

The p15-65 probe contains a 2.7 kb SalI-EcoRI fragment derived from sequences adjacent to and including those contained within the anonymous DNA segment OLVII E10. Marlhens, et al., Nucl. Acids. Res., vol. 15, p. 1348, 1987. OLVII E10 is an 0.8 kb Hind III-EcoRI fragment that marks the D18S8 locus at 18q21.3 and detects MspI polymorphisms in normal individuals. Plasmid p15-65 was derived from a human genomic DNA library prepared from human DNA partially digested with MboI and cloned in lambda FIX (Stratagene). Two phage clones hybridizing to OLVII E10 were isolated by hybridization selection. Among several tested subfragments from the two phage clones identified, p15-65 produced the highest signal to noise ratio when used for Southern blot experiments. Plasmid p15-65 (and OLVII E10) detect polymorphic MspI sites, and thus give rise to four alleles of 10.5, 9.7, 7.8, and 7.0 kb at frequences of 17%, 4%, 49% and 30%, respectively (N=232).

In order to determine the molecular basis of the acquisition of heterozygosity in the carcinoma tissue of patient S123, genomic DNA clones containing the effected MspI site were isolated from the carcinoma and compared to normal DNA sequences. Except for MspI, the restriction maps to this region were identical in the tumor and normal tissue of patient S123 indicating no gross DNA additions or deletions in the tumor. The affected MspI site was found to lie within an Alu-type repeated element in a 1.8 kb EcoRI fragment which is 5 kb from p15-65. The sequence of the cloned DNA fragment from the S123 tumor DNA differed from the normal allele at a single base pair, resulting in the replacement of the internal G residue within the MspI recognition sequence 5'-CCGG-3' with an A residue. This mutation created a potential 3' splice acceptor site, identical to the consensus sequence for intron-exon junctions of primate genes (Shapiro, et al., Nucleic Acids Res., vol. 15, p. 7155, 1987; Ruskin, et al., Cell, vol. 38,

p. 317, 1984; and Zhuang, et al., Proc. Natl. Acad. Sci. USA, vol. 86, p. 2752, 1989.) It is noted that mutations creating splice acceptor sites associated with abnormal RNA processing have been previously found in inherited diseases, such as thalassemias (Ley, et al., Proc. Natl. Acad. Sci., USA, vol. 79, p. 4775, 1982; and Sharp, Annual Reviews of Biochem., vol. 55, p. 1119, 1986.)

Phage clones that encompassed a 35 kb region surrounding the mutated MspI site of tumor S123 were isolated. All EcoRI fragments from the phage clones were subcloned and used in hybridization experiments with Northern blots containing RNA of normal colonic mucosa and cell lines derived from tumors of the colon and several other organs. No expression was detected in these experiments, nor was expression detected in RNase protection studies of these RNA samples using selected subfragments from the phage clones. A more exhaustive strategy was therefore undertaken to identify expressed from this region of chromosome 18q.

A bidirectional chromosomal walk from this region was carried out using bacteriophage vectors. Over 140 unique clones, spanning approximately 370 kb, were isolated in 30 rounds of walking; the clones representing the maximal walk for each of the 30 rounds are shown in Figure 1. The clones were obtained as follows.

Human genomic DNA was partially-digested with MboI and fragments of 12-18 kb were cloned in the lambda FIX vector (Stratagene) using conditions recommended by the manufacturer. Clones were propagated in E. coli C600 or TAP 90 cells (Patterson, et al., Nucleic Acids Res., vol. 15, p. 6298, 1987). For each round of walking, EcoRI maps were constructed from comparison of digests of overlapping phage clones (starting with page clones containing p15-65). EcoRI fragments mapping furthest from previously obtained phage clones were used to re-screen the library. Approximately 1×10^6 phage clones were screened for each round of walking; three to seven new clones were generally obtained in each walk, and the clones purified through three rounds of hybridization selection.

In an effort to identify potential exons in the 370 kb region on the basis of their homology to other species, every EcoRI fragment

from the region was isolated and used as a hybridization probe at reduced stringency (conditions as described in Vogelstein, et al., Cancer Res., vol. 47, p. 4806, 1987 except that the hybridization buffer contained 0.5% non-fat dried milk and filter washing was performed in 44.5 mM sodium chloride, 1.8 mM sodium citrate, 0.3 mM Tris, pH7.5 at 55°C for 45 minutes) to screen DNA samples from various species (mouse, rat, hamster, chicken, Xenopus and yeast). Such a strategy for the identification of exons was employed for the Duchenne muscular dystrophy gene and cystic fibrosis genes. (Monaco, et al., Nature, vol. 323, p. 646, 1986 and Rommens, et al., Science, vol. 245, p. 1059, 1989). Twenty-four of the 117 EcoRI fragments hybridized to discrete DNA fragments of at least one of the species tested; these fragments are indicated by the solid boxes in Figure 1. The patterns observed with two of the fragments producing strong cross-species hybridization are illustrated in Figure 2. Fragment O hybridized strongly to mouse, rat, hamster, and Xenopus DNAs, and fragment P hybridized strongly to mouse, rat, and hamster DNAs.

Most of the cross-species hybridizing fragments were then used as probes to screen Northern blots prepared with RNA of various normal tissues or tumor cell lines, and also used to screen cDNA libraries from normal colonic mucosa specimens, a colorectal adenoma cell line, a brain tumor, a fibrosarcoma line, and an embryonal carcinoma line. No evidence for expression was obtained in these Northern blot studies, nor were any hybridizing clones identified in any of the cDNA libraries using the cross-hybridizing fragments as probes.

In order to determine if any of the cross-hybridizing fragments had exon-like structural features, we cloned several of the homologous fragments from the rat, and determined the regions of cross-hybridization for each fragment for both human and rat.

Rat genomic clones were obtained in two ways. In some cases, fragments identified on Southern blots by cross-species hybridization were eluted from agarose gels and cloned (as described for the human genomic clones above) using cross-hybridizing, radiolabeled human clones as probes. In other cases, MboI partial digests of rat genomic DNA were cloned in lambda DASH (Stratagene) using conditions

recommended by the manufacturer, and the library was screened with homologous human clones of interest.

The cross-hybridizing regions were then subcloned and sequenced from both species; the seven regions for which this analysis was performed are G, I, J, K, M, O and P. The sizes of the regions of homology studied ranged from 128 to over 534 bp, and their homology ranged from 75-89%. The sequences were examined for open reading frames (ORFs), conservation in the predicted amino acid sequence of the ORFs, consensus mammalian splice acceptor and lariat sequences at the 5' end of the ORFs, and consensus splice donor sequences at the 3' region of the ORFs (Shapiro, et al., Nucleic Acids Res., vol. 15, p. 7155, 1987; and Ruskin et al., Cell, vol. 38, p. 317, 1984). Several of these features were found in most of the fragments sequenced. Three of the human-rat fragment pairs (fragments G, O, and P in Figures 1 and 3) were found to have more exon-like features than the other fragments. The region of cross-species hybridization from human fragment G and its homologous rat fragment predicted ORFs that differed at two amino acid positions (Figure 3). Similarly, the ORFs predicted for the other two sets of fragments were highly conserved with a single amino acid substitution distinguishing human fragments O and P from their respective rat homologues (Figure 3). For all three sets of fragments, nucleotide substitutions were predominantly at the third position of codons in the exon-like regions, and sequence homology decreased to 75% or less outside these regions.

The striking homology between the human and rat sequences and their exon-like structural features suggested that the fragments might contain potential exons. However, as noted above, no expression was detected when these three fragments were used as probes of Northern blots of various RNAs or used to screen cDNA libraries. In addition, RNase protection experiments (Winter, Proc. Natl. Acad. Sci. USA, vol. 82, p. 7575, 1985), using these three fragments to generate anti-sense transcripts, failed to conclusively demonstrate evidence of expression in a variety of RNA samples, including normal colonic mucosa, colorectal cell lines, and other tumor cell lines of various types.

To increase the sensitivity of expression assays, we utilized the polymerase chain reaction (PCR, Saiki, et al., Science, vol. 239, p. 487, 1988) in an "exon-connection" strategy. cDNA was prepared from total RNA of various cell lines and tissues using reverse transcriptase to prime synthesis with random hexamers (Noonan, et al., Nucleic Acids Res., vol. 16, p. 10366, 1988). This single-stranded cDNA was then used in PCR experiments with oligonucleotide pairs derived from the sequence of two of the potential exons noted above. If the two potential exons were present in the same RNA transcript, then, using the appropriate oligonucleotides, it would be possible to amplify a cDNA product linking these two regions. Trace amounts of DNA, which often contaminate RNA preparations, would not give rise to the same sized PCR product in this assay because the exons were separated by an intron. Most pairs of oligonucleotides derived from the seven regions of human-rat homology described above did not generate discrete PCR fragments when tested by this strategy. However, an oligonucleotide pair derived from the sequence of fragments O and P of Figure 3 was found to generate a discrete 233 bp PCR product from cDNA of several of the mRNA samples studies, including that derived from a small cell carcinoma of the lung (H82) and a colorectal carcinoma (HCT116).

The oligonucleotide derived from fragment O was 5'-TTCCGCCATGGTTTTTAAATCA-3' and the oligonucleotide derived from fragment P was 5'-AGCCTCATTTTCAGCCACACA-3'. Cycle times were 1 minute at 95°C, 1 minute at 58°C, and 2 minutes at 70°C; twenty-five cycles were performed. The PCR products were phosphorylated using T4 DNA kinase (Bethesda Research Laboratories), ligated to EcoRI linkers (New England Biolabs), inserted into lambda gt10 phage arms (Stratagene), and cloned in E. coli C600 cells. Insert-containing clones were identified through hybridization with radiolabeled probes from fragments O and P (Figure 3). EcoRI inserts from the phage clones were subcloned in Bluescript SK (Stratagene) and sequenced as described by S. Tabor and C. C. Richardson, Proc. Natl. Acad. Sci. USA, vol. 84, p. 4767, 1987; and Kraft, et al., Biotech., vol. 6, p. 544, 1988. The PCR products from both cell lines

482 and HCT 116 were found to be the product of splicing the predicted exon of fragment O directly to the predicted exon of fragment P. Thus the region of chromosome 18q is expressed.

Example 2:

This example demonstrates the isolation and sequencing of the cDNA corresponding to the coding sequence of the gene which is deleted in colorectal carcinomas.

To confirm and extend these PCR experiments, we constructed a cDNA library from RNA of the H82 cell line (Gubler, et al., Gene, vol. 25, p. 263, 1983). Approximately 3.0 x 10^6 recombinant cDNA clones were screened with genomic DNA subclones containing regions of fragments G, O, and P of Figure 3. Four hybridizing clones were isolated and mapped with respect to one another and to the genomic clones shown in Figure 1. The longest clone was 1.65 kb in length and hybridized to at least eleven unique EcoRI fragments in human genomic DNA (Figure 6A), eight of which were present within the 370 kb region cloned in the chromosomal walk shown in Figure 1. The four cDNA clones isolated were sequenced and subsequently used as probes of cDNA libraries obtained from H82 cells or from normal human brain to obtain additional cDNA clones extending for a total of 2854 base pairs. Through sequencing, all clones were found to encode overlapping portions of a transcript in which there was a single long ORF of 2250 bp, which extended to the end of the sequenced region. The ORF began with a methionine codon in a favorable context for translation initiation (nucleotide 1 in Figure 4) according to the paradigms of Kozak, Nucl. Acids Res., vol. 15, p. 8125, 1987. The methionine was followed by a relatively hydrophobic sequence of 25 amino acids which resembled previously described signal sequences associated with membrane-bound proteins (Watson, Nucleic Acids Res., vol. 12, p. 5145, 1984). The signal sequence was immediately followed by 725 amino acids with significant homology to the neural cell adhesion molecules and other related cell surface glycoproteins. The gene encoding this transcript will be referred to as DCC (Deleted in Colorectal Carcinomas).

Example 3:

This example demonstrates that most normal tissues of rat and human produce the DCC transcript. However most colorectal tumor cell lines do not produce amounts as great as are produced in normal cells.

To identify the tissues in which the DCC gene was expressed, cDNA was prepared from several rat organs. These were: liver, kidney, adrenal, heart, lung, stomach, esophagus, spleen, small bowel, breast, bladder, uterus, aorta, psoas, brain, colon, tongue and skin. Because of the high degree of conservation of the DCC gene (see Figure 3), the same oligonucleotide primers used to demonstrate expression in human cells could also be used in the rat. To assess expression, oligonucleotide pairs from fragments O and P of Figure 3 were used in a PCR expression assay as described above.

Seventeen of the eighteen rat tissues tested appeared to produce the transcript at low levels, with greatest abundance observed in brain (Chomczynski, et al., Anal. Biochem., vol. 162, p. 156, 1987). Only the liver did not produce detectable transcript. Similar analysis of human tissues and cell lines revealed that the transcript was present in highest concentration in brain, and was also expressed in normal colonic mucosa and in several tumor cell lines, including those derived from tumors of the lung, brain and mesenchyme (Figure 5 and data not shown). In most colorectal carcinomas, however, expression was greatly reduced or absent; of seventeen colorectal tumor cell lines studied, only two expressed DCC mRNA at levels in excess of 5% of that produced in normal colonic mucosa (examples in Figure 5). The human colorectal carcinoma cell lines used were: lane 6, SW948; lane 7, SW1417; lane 8, SW1116; lane 9, SW403; lane 10, SW1463; lane 11, SW48; lane 12, HCT116; lane 13, RKO; lane 14, RCA; lane 15, "C"; lane 16, MOSER.

To determine the size of the transcript produced from this gene, Northern blots containing RNA from normal colonic mucosa or brain were hybridized with radioactively labelled cDNA clones. A major band of 10-12 kb was observed in normal brain RNA, but no bands were seen in the RNA from colonic mucosa (data not shown),

consistent with the higher level of expression observed in brain by PCR analysis.

Example 4:

This example demonstrates that the boundary of the deletion of one of the alleles in the S115 tumor is within the DCC gene and that rearrangements of the gene can be detected in human cells using cDNA probes.

In an attempt to establish the boundaries of the homozygous loss in the S115 tumor with respect to the DCC gene, the cDNA clones were used to probe Southern blots containing S115 DNA. A 430 bp subclone (pKC430, representing nucleotide 1760 to 2205 of the cDNA) detected three EcoRI fragments of 20 kb, 10 kb, and 1.8 kb in DNA from non-neoplastic colonic mucosa of patient S115 (Figure 6B). However, in DNA from the S115 tumor, the 20 kb fragment was not detected and the 10 kb and 1.8 kb fragments were present at approximately half the intensity observed in normal DNA. In addition, a new fragment of 5 kb was observed only in DNA from the tumor (marked with an arrowhead in Figure 6B). Probes more 3' than pKC430 also detected fragments in tumor DNA which were present at half the intensity of those seen in normal DNA, while probes 5' of pKC430 detected fragments which were homozygously deleted in the tumor. Thus, the DCC gene appeared to be broken by the deletion event on one of the two copies of chromosome 18, and the breakpoint established one boundary of the homozygous loss.

Example 5:

This example demonstrates that the cDNA probes of the present invention can be used to detect genetic alterations in DNA samples isolated from colorectal tumors.

To search for other genetic alterations, the cDNA probes were used in Southern blot analysis of colorectal tumor DNA samples. DNA from normal and tumor DNA samples were digested with EcoRI and Eco0109 and Southern blots were prepared as described above. The DNA was then hybridized to a 0.4 kb genomic fragment which contained the exon from fragment P.

Three of 51 primary tumors and two of twenty-one tumor xenografts were found to have new fragments not present in normal DNA of the same patient. In addition, five of twenty-two colorectal tumor cell lines were found to have altered fragments not present in forty-four DNA samples from normal individuals nor in any of the 45 DNA samples from tumor cell lines derived from tissues other than that of the colon or rectum.

In all cases, detailed mapping experiments showed that the new fragments detected by the cDNA probe resulted from insertions in an approximately 170 kb XbaI-Eco0109 fragment located 165 bp downstream of the exon in fragment P of Figure 3. The insertions were mapped by comparison of Southern blot patterns produced by digestion of tumor DNA samples with a combination of XbaI and Eco0109, and HindIII. The size of the insertion in the tumors varied from 120-300 bp (Figure 7A). Some variation in the size of the Xba1-Eco0109 fragment in alleles from normal individuals was seen; however, the maximal difference between the size of the smallest and largest of the 88 normal alleles studied was approximately 35 bp, and the largest of the normal alleles was found to be approximately 120 bp smaller than any of the altered alleles seen in the tumors.

The 1.4 kb EcoRI fragment (fragment P) containing the insertion site was isolated from one of the genomic clones of the chromosome walk (Figure 1) and sequenced; the sequence of the 170 bp XbaI-Eco0109 fragment from this fragment is shown in Figure 7B. There were two regions of TA repeats in the XbaI-Eco0109 fragment; one of the regions had eight repeats and the other had twenty-six. Both TA repeat regions were contained within a 130 bp region of alternating purine-pyrimidine base pairs which could potentially form Z-DNA (Rich, et al., Annu. Rev. Biochem., vol. 53, p. 791, 1984).

Example 6:

This example demonstrates the sequence similarity between four immunoglobulin-like domains of the DCC gene with the chicken N-CAM and mouse N-CAM genes.

The predicted amino acids sequence of DCC is highly homologous to the neural cell adhesion molecules (N-CAM) and other related

cell surface glycoproteins. (Edelman, Biochem., vol. 27, p. 3535, 1988). Two areas of high homology are noted. First the DCC gene contains four immunoglobulin-like (Ig-like) domains of the C2 class, defined by pairs of cysteines separated by 50 to 56 amino acids and other highly conserved residues surrounding the first and second cysteine of each pair. (See Figure 8A and Williams, Ann. Rev. Immunol., vol. 6, p. 381, 1988).

Sequences of the domains were aligned by inspection and spaces, indicated by dashes were inserted to give the greatest overall match. Residues in two or more of the DCC domains were boxed if they were identical. Each Ig-like domain is approximately 100 amino acids in length. DCC domain no. 1 includes amino acids 40 to 139, domain no. 2 includes amino acids 140-239, domain no. 3 includes amino acids 240-332, and domain no. 4 includes amino acids 333-422. The four Ig-like domains of DCC were found to be more homologous to one another than to N-CAM L1, or other members of the Ig super-family. A consensus sequence for the four Ig-like domains of DCC could be derived for 67% of the positions; the DCC consensus sequence matched the N-CAM consensus sequence at 42% of these positions.

Sequence homology between the DCC and chicken and mouse N-CAM genes was also found in the fibronectin-type III-related regions. DCC amino acid positions 423 to 605 were compared to amino acids 481 to 662 of the two N-CAM proteins. Potential sites of N-linked glycosylation are found at several positions within this region.

The fibronectin-type III-related domain is similar to the fibronectin-like domains present in N-CAM, L1, leucocyte common antigen related gene 1 (Lar1), fasciclin II, and other members of the cell adhesion molecule family. These fibronectin-related domains are carboxyl to the Ig-like domain in all of these proteins including DCC. Of 195 positions within the fibronectin-like region of DCC, 31% were identical in DCC and N-CAM and several conservative substitutions were also found (Figure 8B). The extensive homologies in both Ig-like and fibronectin-like domains between DCC and other members of this

family suggest that these proteins may all have been derived from a common precursor that included both these regions.

Example 7:

This example demonstrates the expression of the DCC protein in bacteria and the production of anti-DCC antibodies.

The DCC cDNA sequence shown in Figure 4 was inserted in a bacterial expression vector, pEX2. The vector will produce the protein of any open reading frame as a fusion product with beta-galactosidase. (See Stanley, et al., The EMBO Journal, vol. 3, pp. 1429-1434, 1984.) The bacterially expressed DCC protein was partially purified and injected into rabbits together with Freund's adjuvant. The rabbits were given three monthly booster injections. The rabbits produced antibodies which immunoprecipitate DCC proteins synthesized in in vitro translation reactions. The antibodies also bind to DCC on Western blots.

CLAIMS

1. A method of diagnosing or prognosing a neoplastic tissue of a human, comprising:

isolating a tissue from a human;

detecting loss of wild-type DCC gene coding sequences or their expression products from said tissue, said loss indicating neoplasia of the tissue and correlating with metastasis and early death.

2. The method of claim 1 wherein the expression products are mRNA molecules.

3. The method of claim 2 wherein the loss of wild-type DCC mRNA is detected by Northern hybridization of mRNA from said tissue to a DCC gene probe.

4. The method of claim 1 wherein loss of wild-type DCC gene coding sequences is detected by Southern hybridization of a DCC gene coding sequence probe to genomic DNA isolated from said tissue.

5. The method of claim 4 further comprising:

isolating a non-neoplastic tissue of the human;

isolating genomic DNA from said non-neoplastic tissue;

subjecting the genomic DNA to Southern hybridization with the DCC gene coding sequence probe; and

comparing the hybridizations of the DCC gene probe to said tissues.

6. The method of claim 4 wherein the DCC gene probe detects a restriction fragment length polymorphism.

7. The method of claim 1 wherein the loss of wild-type DCC gene coding sequences is detected by determining the sequence of all or part of a DCC gene in said tissue using a polymerase chain reaction, deviations in the DCC sequence determined from that of the sequence shown in Figure 4 indicating neoplasia.

8. The method of claim 1 wherein the loss of wild-type DCC gene coding sequences is detected by identifying a mismatch between molecules (1) a DCC gene or DCC mRNA isolated from said tissue and (2) a nucleic acid probe complementary to the human wild-

type DCC gene coding sequence, when molecules (1) and (2) are hybridized to each other to form a duplex.

9. The method of claim 4 wherein the DCC gene probe hybridizes to fragment P, a 1.4 kb EcoRI fragment shown in Figure 3.

10. The method of claim 1 wherein the loss of wild-type DCC gene coding sequences is detected by amplification of DCC gene sequences in said tissue and hybridization of the amplified DCC sequences to nucleic acid probes which comprise DCC sequences.

11. The method of claim 1 wherein the loss of wild-type DCC gene coding sequences is detected by molecular cloning of the DCC genes in said tissue and sequencing all or part of the cloned DCC gene.

12. The method of claim 1 wherein the detection of loss of wild-type DCC gene coding sequences comprises screening for a deletion mutation.

13. The method of claim 1 wherein the detection of loss of wild-type DCC gene coding sequences comprises screening for a point mutation.

14. The method of claim 1 wherein the detection of loss of wild-type DCC gene coding sequences comprises screening for an insertion mutation.

15. The method of claim 1 wherein the tissue isolated from a human is from colonic mucosa.

16. The method of claim 1 wherein the expression products are protein molecules.

17. The method of claim 16 wherein the loss of wild-type DCC protein is detected by Western blotting.

18. A method of supplying wild-type DCC gene function to a cell which has lost said gene function by virtue of mutation in a DCC gene, comprising:

introducing a wild-type DCC gene into a cell which has lost said gene function such that said gene is expressed in the cell.

19. The method of claim 18 wherein the wild-type DCC gene introduced recombines with the endogenous mutant DCC gene

present in the cell by a double recombination event to correct the DCC gene mutation.

20. A method of supplying wild-type DCC gene function to a cell which has lost said gene function by virtue of a mutation in a DCC gene, comprising:

introducing a portion of a wild-type DCC gene into a cell which has lost said gene function such that said portion is expressed in the cell, said portion encoding a part of the DCC protein which is required for non-neoplastic growth of said cell.

21. A kit for determination of the nucleotide sequence of a DCC gene by polymerase chain reaction, comprising:

a set of pairs of single stranded DNA primers, the sequence of said set derived from chromosome 18q, said set allowing synthesis of all nucleotides of the DCC gene coding sequences.

22. The kit of claim 21 wherein the primers have restriction enzyme sites at each 5' end.

23. A nucleic acid probe complementary to human wild-type DCC gene coding sequences.

24. A nucleic acid probe which hybridizes to the XbaI-Eco0109 fragment located 165 bp downstream from the DCC exon in fragment P shown in Figure 3.

25. A kit for detecting loss of wild-type DCC genes comprising a battery of nucleic acid probes which in the aggregate hybridize to all nucleotides of the DCC gene.

26. A method of detecting the presence of a neoplastic tissue in a human, comprising:

isolating a body sample from a human;

detecting in said sample loss of a wild-type DCC gene coding sequence or wild-type DCC expression product, said loss indicating the presence of a neoplastic tissue in the human.

27. The method of claim 26 wherein said body sample is selected from the group consisting of serum, stool, urine and sputum.

28. A method of detecting genetic predisposition to cancer in a human comprising:

isolating a human sample selected from the group consisting of blood and fetal tissue;

detecting loss of wild-type DCC gene coding sequences or their expression products from the sample, said loss indicating predisposition to cancer.

29. The method of claim 28 wherein the expression products are mRNA molecules.

30. The method of claim 29 wherein the loss of wild-type DCC mRNA is detected by Northern hybridization of mRNA from said tissue to a DCC gene probe.

31. The method of claim 28 wherein loss of wild-type DCC gene coding sequences is detected by Southern hybridization of a DCC gene coding sequence probe to genomic DNA isolated from said tissue.

32. The method of claim 31 wherein the DCC gene coding sequence probe detects a restriction fragment length polymorphism.

33. The method of claim 28 wherein the loss of wild-type DCC gene coding sequences is detected by determining the sequence of all or part of a DCC gene in said tissue using polymerase chain reaction, deviations in the DCC sequence determined from the sequence of Figure 4 indicating predisposition to cancer.

34. The method of claim 28 wherein the loss of wild-type DCC gene coding sequences is detected by identifying a mismatch between molecules (1) a DCC gene or DCC mRNA isolated from said tissue and (2) a nucleic acid probe complementary to the human wild-type DCC gene coding sequence, when molecules (1) and (2) are hybridized to each other to form a duplex.

35. The method of claim 31 wherein the DCC gene probe hybridizes to fragment P, a 1.4 kb EcoRI fragment shown in Figure 3.

36. The method of claim 28 wherein the loss of wild-type DCC gene coding sequences is detected by amplification of DCC gene sequences in said tissue and hybridization of the amplified DCC sequences to nucleic acid probes which comprise DCC gene coding sequences.

37. The method of claim 28 wherein the loss of wild-type DCC gene coding sequences is detected by molecular cloning of the

DCC genes in said tissue and sequencing all or part of the cloned DCC gene.

38. The method of claim 28 wherein the detection of loss of wild-type DCC gene coding sequences comprises screening for a deletion mutation.

39. The method of claim 28 wherein the detection of loss of wild-type DCC gene coding sequences comprises screening for a point mutation.

40. The method of claim 28 wherein the detection of loss of wild-type DCC gene coding sequences comprises screening for an insertion mutation.

41. The method of claim 28 wherein the expression products are protein molecules.

42. The method of claim 28 wherein the loss of wild-type DCC protein is detected by Western blotting.

43. A cDNA molecule comprising the coding sequence of the DCC gene.

44. A preparation of the human DCC protein substantially free of other human proteins, the amino acid sequence of said protein corresponding to that shown in Figure 4.

45. A preparation of antibodies immunoreactive with human DCC protein and not substantially immunoreactive with other human proteins.

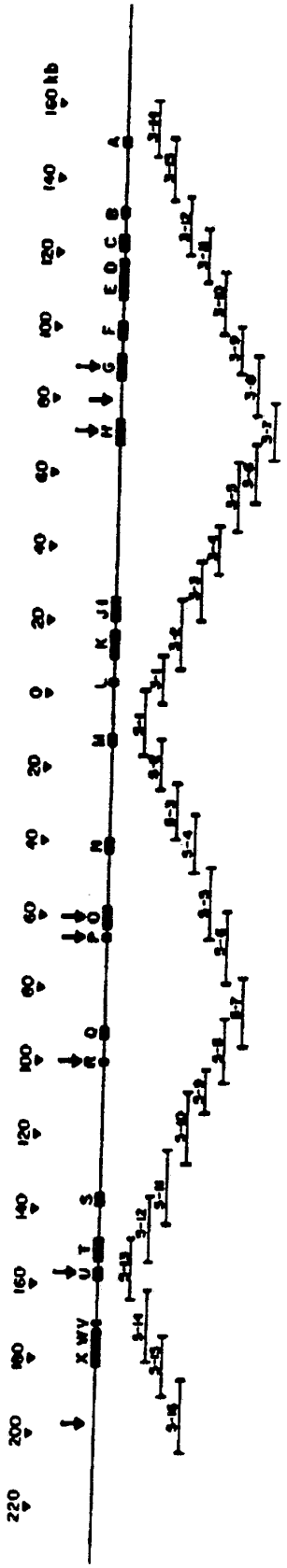

160 kb
140
120
100
80
60
40
20
0
20
40
60
80
100
120
140
160
180
200
220
A
B
C
D
E
F
G
H
J
K
L
M
N
O
P
Q
S
T
U
W
X

FIGURE 1

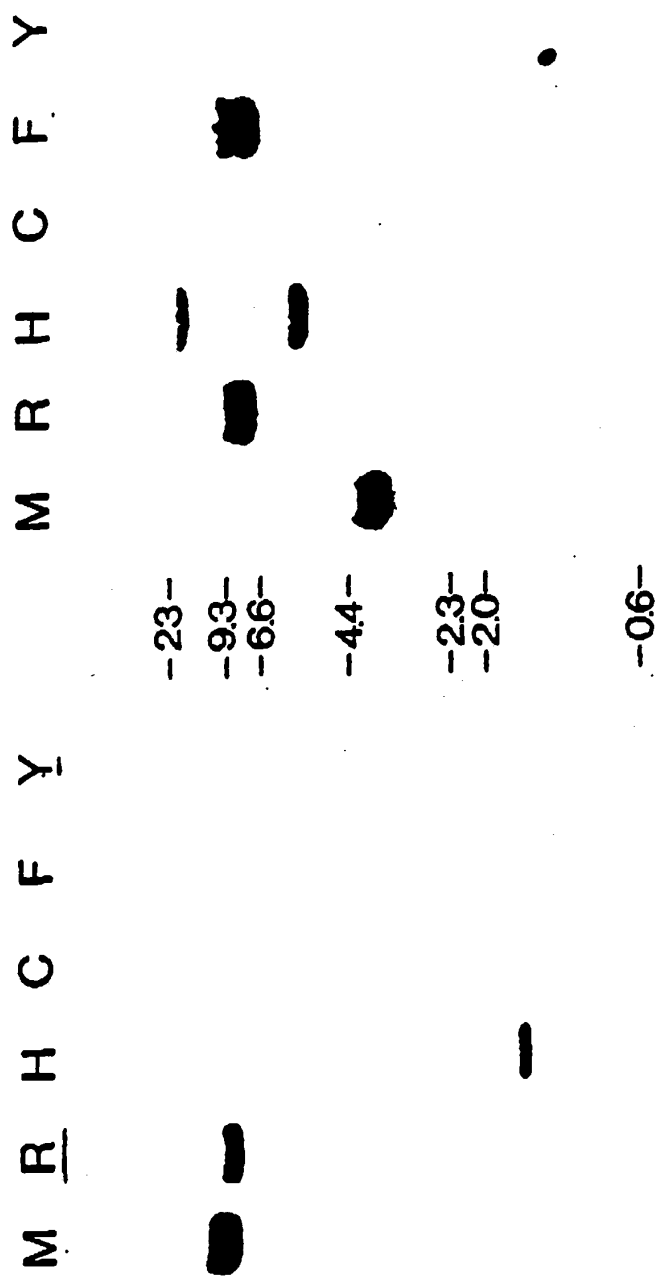


FIGURE 2

3/15

FRAGMENT G

HUMAN ATTTGGAAGACTTATTCTTCCTTCTTTGTTTTTCTCCTAG GA CCA CTG AGG TTC CTT TCA CAG ACA GAA TCT GTC ACA GCC
RAT T GAAAT CT A CC T CT G TG CT C C T C C A
AA P L R F L S Q T E S I T A

HUMAN TTC ATG GGA GAC ACA GTG CTA CTC AAG TGT GAA GTC ATT GGG GAG CCC ATG CCA ACA ATC CAC TGG CAG AAG
RAT T T C A
AA F M G D T V L L K C E V I G E/D P M P T I H W Q K

HUMAN AAC CAA CAA GAC CTG ACT CCA ATC CCA GGT GAC TCC CGA GTG GTG GTC
RAT T C C AC G A G T C
AA N Q Q D L T P I/M P G D S R V V V

FRAGMENT O

HUMAN CTCACTCACTTTTTTTTTTTTCTGTCTTTGCAG TT CCG CCA TGG TTT TTA AAT CAT CCT TCC AAC CTG TAT GCC TAT GAA AGC
RAT T C G C
AA P P W F L N H P S N L Y A Y E S

HUMAN ATG GAT ATT GAG TTT GAA TGT ACA GTC TCT GGA AAG CCT GTG CCC ACT GTG AAT TGG ATG AAG AAT GGA GAT GTG
RAT C G G C G
AA M D I E F E C T/A V S G K P V P T V N W H K N G D V

HUMAN GTC ATT CCT AGT GAT TAT TTT CAG ATA GTG GTAATT
RAT G --- --- --- --- --- --- --- ---
AA V I P S D Y F Q I V

FRAGMENT P

HUMAN TACTTTACCTGCTTTTTTCTTTTCCCTAG GGA GGA AGC AAC TTA CGG ATA CTT GGG GTG GTG AAG TCA GAT GAA GGC TTT
RAT ---------- --- --- --- --- -- C A
AA G G G S N L R I L G V V K S D E G F

HUMAN TAT CAA TGT GTG GCT GAA AAT GAG GCT GGA AAT GCC CAG ACC AGT GCA CAG CTC ATT GTC CCT AAG CCT GGTAAGACAAT
RAT C A GT C G C C C
AA Y Q C V A E N E A G N A Q T/S S A Q L I V P K P

FIGURE 3

```
                Gln Leu Pro Ala Arg Thr Pro Leu Ala Thr Pro Gln *** Gln Val
            C CAA CTC CCA GCT CGC ACA CCG CTG GCG ACA CCC CAG TAA CAA GTG

    Arg Ala Leu His Pro Ala Val Pro Pro Ala Ser Pro Pro Trp Val Pro
    AGA GCG CTC CAC CCC GCA GTC CCC CCC GCC TCT CCT CCC TGG GTC CCC

    Ser Ala Leu Gly Arg Lys Thr Asn Ser Ile Ser Ser Ser Arg Ala Glu
    TCG GCT CTC GGA AGA AAA ACC AAC AGC ATC TCC AGC TCT CGC GCG GAA

    Leu Ser Leu Gln Leu Tyr Pro Thr Asp Asp Lys Glu Pro Ala Ser Thr
    TTG TCT CTT CAA CTT TAC CCA ACC GAC GAC AAG GAA CCA GCC TCA ACC

    Phe *** Cys Thr Ala Arg Pro Gln Asp Cys Leu Pro Ser Pro Leu Gly
    TTT TAA TGC ACA GCC CGG CCA CAG GAT TGC CTT CCA TCT CCT CTT GGT

    Pro Ser Trp Met Trp Phe Ile Asp Asp Leu Arg Ala Pro Gln Arg Ala
    CCC TCC TGG ATG TGG TTT ATT GAT GAC TTG CGA GCC CCT CAG AGA GCT

    Val Phe Pro Pro Leu Ala Pro Ser Val Ser Leu Ser *** Phe Ser Lys
    GTC TTC CCT CCT CTG GCT CCC TCC GTT TCC TTG AGT TAG TTT TCT AAG

    Val Leu Pro Gly Leu Gly Ile Ser Trp Thr Glu Trp Asn Phe Leu Leu
    GTT TTA CCG GGG CTC GGG ATC TCT TGG ACC GAA TGG AAC TTT TTG CTG

    Pro Ala Phe Ala Ala Asp Ser Val Ser Gly Gln Gly Lys Arg Leu Arg
    CCT GCT TTT GCT GCT GAT TCT GTC AGT GGA CAA GGA AAA AGG CTT CGA

    Arg Gln Gln Arg Arg Arg Gly Gly Gly Glu Arg Gly Gly Gly Arg Gly
    AGG CAG CAG AGG CGC AGG GGA GGT GGA GAA AGA GGT GGA GGA AGA GGA

    Arg Gly Gly Gly Gly Ser Arg Arg Gly Ser Ala Arg Val Cys Ala Cys
    CGA GGA GGA GGA GGA AGC CGA AGG GGC TCG GCG CGT GTG TGT GCA TGT

    Val His Ala Cys Val Ser Ala Cys Val *** Val Leu Pro Leu Pro Ala
    GTG CAT GCG TGT GTG AGT GCA TGT GTG TGA GTG CTG CCG CTG CCC GCG
                                          START
    Thr Pro Gly Pro Glu Gly Val Gly *** Asn Met Glu Asn Ser Leu Arg
    ACC CCT GGC CCC GAA GGT GTT GGC TGA AAT ATG GAG AAT AGT CTT AGA

    Cys Val Trp Val Pro Lys Leu Ala Phe Val Leu Phe Gly Ala Ser Leu
    TGT GTT TGG GTA CCC AAG CTG GCT TTT GTA CTC TTC GGA GCT TCC TTG

    Leu Ser Ala His Leu Gln Val Thr Gly Phe Gln Ile Lys Ala Phe Thr
    CTC AGC GCG CAT CTT CAA GTA ACC GGT TTT CAA ATT AAA GCT TTC ACA
```

FIGURE 4
(Page 1 of 6)

```
[illegible]Leu Arg Phe Leu [illegible] Glu Pro Ser Asp Ala Val [illegible] Met Arg Gly   255
 [illegible] GCA [illegible] TTC [illegible] TCA [illegible] CCT TCT GAT [illegible] [illegible] [illegible] TTG [illegible] GGA

256  Gly Asn Val Leu Leu Asp Cys Ser Ala Glu Ser Asp Arg Gly Val Pro   271
767  GGA AAT GTC CTC CTC GAC TGC TCC GCG GAG TCC GAC CGA GGA GTT CCA   814

272  Val Ile Lys Trp Lys Lys Asp Gly Ile His Leu Ala Leu Gly Met Asp   287
815  GTG ATC AAG TGG AAG AAA GAT GGC ATT CAT CTG GCC TTG GGA ATG GAT   862

288  Glu Arg Lys Gln Gln Leu Ser Asn Gly Ser Leu Leu Ile Gln Asn Ile   303
863  GAA AGG AAG CAG CAA CTT TCA AAT GGG TCT CTG CTG ATA CAA AAC ATA   910

304  Leu His Ser Arg His His Lys Pro Asp Glu Gly Leu Tyr Gln Cys Glu   319
911  CTT CAT TCC AGA CAC CAC AAG CCA GAT GAG GGA CTT TAC CAA TGT GAG   958

320  Ala Ser Leu Gly Asp Ser Gly Ser Ile Ile Ser Arg Thr Ala Lys Val   335
959  GCA TCT TTA GGA GAT TCT GGC TCA ATT ATT AGT CGG ACA GCA AAA GTT  1006

336  Ala Val Ala Gly Pro Leu Arg Phe Leu Ser Gln Thr Glu Ser Val Thr   351
1007 GCA GTA GCA GGA CCA CTG AGG TTC CTT TCA CAG ACA GAA TCT GTC ACA  1054

352  Ala Phe Met Gly Asp Thr Val Leu Leu Lys Cys Glu Val Ile Gly Glu   367
1055 GCC TTC ATG GGA GAC ACA GTG CTA CTC AAG TGT GAA GTC ATT GGG GAG  1102

368  Pro Met Pro Thr Ile His Trp Gln Lys Asn Gln Gln Asp Leu Thr Pro   383
1103 CCC ATG CCA ACA ATC CAC TGG CAG AAG AAC CAA CAA GAC CTG ACT CCA  1150

384  Ile Pro Gly Asp Ser Arg Val Val Val Leu Pro Ser Gly Ala Leu Gln   399
1151 ATC CCA GGT GAC TCC CGA GTG GTG GTC TTG CCC TCT GGA GCA TTG CAG  1198

400  Ile Ser Arg Leu Gln Pro Gly Asp Ile Gly Ile Tyr Arg Cys Ser Ala   415
1199 ATC AGC CGA CTC CAA CCG GGG GAC ATT GGA ATT TAC CGA TGC TCA GCT  1246

416  Arg Asn Pro Ala Ser Ser Arg Thr Gly Asn Glu Ala Glu Val Arg Ile   431
1247 CGA AAT CCA GCC AGC TCA AGA ACA GGA AAT GAA GCA GAA GTC AGA ATT  1294

432  Leu Ser Asp Pro Gly Leu His Arg Gln Leu Tyr Phe Leu Gln Arg Pro   447
1295 TTA TCA GAT CCA GGA CTG CAT AGA CAG CTG TAT TTT CTG CAA AGA CCA  1342

448  Ser Asn Val Val Ala Ile Glu Gly Lys Asp Ala Val Leu Glu Cys Cys   463
1343 TCC AAT GTA GTA GCC ATT GAA GGA AAA GAT GCT GTC CTG GAA TGT TGT  1390

464  Val Ser Gly Tyr Pro Pro Pro Ser Phe Thr Trp Leu Arg Gly Glu Glu   479
1391 GTT TCT GGC TAT CCT CCA CCA AGT TTT ACC TGG TTA CGA GGC GAG GAA  1438

480  Val Ile Gln Leu Arg Ser Lys Lys Tyr Ser Leu Leu Gly Gly Ser Asn   495
1439 GTC ATC CAA CTC AGG TCT AAA AAG TAT TCT TTA TTG GGT GGA AGC AAC  1486

496  Leu Leu Ile Ser Asn Val Thr Asp Asp Asp Ser Gly Met Tyr Thr Cys   511
1487 TTG CTT ATC TCC AAT GTG ACA GAT GAT GAC AGT GGA ATG TAT ACC TGT  1534

512  Val Val Thr Tyr Lys Asn Glu Asn Ile Ser Ala Ser Ala Glu Leu Thr   527
1535 GTT GTC ACA TAT AAA AAT GAG AAT ATT AGT GCC TCT GCA GAG CTC ACA  1582

528  Val Leu Val Pro Pro Trp Phe Leu Asn His Pro Ser Asn Leu Tyr Ala   543
1583 GTC TTG GTT CCG CCA TGG TTT TTA AAT CAT CCT TCC AAC CTG TAT GCC  1630

544  Tyr Glu Ser Met Asp Ile Glu Phe Glu Cys Thr Val Ser Gly Lys Pro   559
1631 TAT GAA AGC ATG GAT ATT GAG TTT GAA TGT ACA GTC TCT GGA AAG CCT  1678

560  Val Pro Thr Val Asn Trp Met Lys Asn Gly Asp Val Val Ile Pro Ser   575
1679 GTG CCC ACT GTG AAT TGG ATG AAG AAT GGA GAT GTG GTC ATT CCT AGT  1726

576  Asp Tyr Phe Gln Ile Val Gly Gly Ser Asn Leu Arg Ile Leu Gly Val   591
1727 GAT TAT TT[illegible] [illegible] ATA GTG GGA GGA AGT AAC TTA CGG ATA CTT GGG GT[illegible]
```

FIGURE 4

```
   Val Lys Ser Asp Glu Gly Phe Tyr Gln Cys Val Ala Glu Asn Glu Ala     607
   GTG AAG TCA GAT GAA GGC TTT TAT CAA TGT GTG GCT GAA AAT GAG GCT    1822

   Gly Asn Ala Gln Thr Ser Ala Gln Leu Ile Val Pro Lys Pro Ala Ile     623
   GGA AAT GCC CAG ACC AGT GCA CAG CTC ATT GTC CCT AAG CCT GCA ATC    1870

   Pro Ser Ser Ser Val Leu Pro Ser Ala Pro Arg Asp Val Val Pro Val     639
   CCA AGC TCC AGT GTC CTC CCT TCG GCT CCC AGA GAT GTG GTC CCT GTC    1918

   Leu Val Ser Ser Arg Phe Val Arg Leu Ser Trp Arg Pro Pro Ala Glu     655
   TTG GTT TCC AGC CGA TTT GTC CGT CTC AGC TGG CGC CCA CCT GCA GAA    1966

   Ala Lys Gly Asn Ile Gln Thr Phe Thr Val Phe Phe Ser Arg Glu Gly     671
   GCG AAA GGG AAC ATT CAA ACT TTC ACG GTC TTT TTC TCC AGA GAA GGT    2014

   Asp Asn Arg Glu Arg Ala Leu Asn Thr Thr Gln Pro Gly Ser Leu Gln     687
   GAC AAC AGG GAA CGA GCA TTG AAT ACA ACA CAG CCT GGG TCC CTT CAG    2062

   Leu Thr Val Gly Asn Leu Lys Pro Glu Ala Met Tyr Thr Phe Arg Val     703
   CTC ACT GTG GGA AAC CTG AAG CCA GAA GCC ATG TAC ACC TTT CGA GTT    2110

   Val Ala Tyr Asn Glu Trp Gly Pro Gly Glu Ser Ser Gln Pro Ile Lys     719
   GTG GCT TAC AAT GAA TGG GGA CCG GGA GAG AGT TCT CAA CCC ATC AAG    2158

   Val Ala Thr Gln Pro Glu Leu Gln Val Pro Gly Pro Val Glu Asn Leu     735
   GTG GCC ACA CAG CCT GAG TTG CAA GTT CCA GGG CCA GTA GAA AAC CTG    2206

   Gln Ala Val Ser Thr Ser Pro Thr Ser Ile Leu Ile Thr Trp Glu Pro     751
   CAA GCT GTA TCT ACC TCA CCT ACC TCA ATT CTT ATT ACC TGG GAA CCC    2254

   Pro Ala Tyr Ala Asn Gly Pro Val Gln Gly Tyr Arg Leu Phe Cys Thr     767
   CCT GCC TAT GCA AAC GGT CCA GTC CAA GGT TAC AGA TTG TTC TGC ACT    2302

   Glu Val Ser Thr Gly Lys Glu Gln Asn Ile Glu Val Asp Gly Leu Ser     783
   GAG GTG TCC ACA GGA AAA GAA CAG AAT ATA GAG GTT GAT GGA CTA TCT    2350

   Tyr Lys Leu Glu Gly Leu Lys Lys Phe Thr Glu Tyr Ser Leu Arg Phe     799
   TAT AAA CTG GAA GGC CTG AAA AAA TTC ACC GAA TAT AGT CTT CGA TTC    2398

   Leu Ala Tyr Asn Arg Tyr Gly Pro Gly Val Ser Thr Asp Asp Ile Thr     815
   TTA GCT TAT AAT CGC TAT GGT CCG GGC GTC TCT ACT GAT GAT ATA ACA    2446

   Val Val Thr Leu Ser Asp Val Pro Ser Ala Pro Pro Gln Asn Val Ser     831
   GTG GTT ACA CTT TCT GAC GTG CCA AGT GCC CCG CCT CAG AAC GTC TCC    2494

   Leu Glu Val Val Asn Ser Arg Ser Ile Lys Val Ser Trp Leu Pro Pro     847
   CTG GAA GTG GTC AAT TCA AGA AGT ATC AAA GTT AGC TGG CTG CCT CCT    2542

   Pro Ser Gly Thr Gln Asn Gly Phe Ile Thr Gly Tyr Lys Ile Arg His     863
   CCA TCA GGA ACA CAA AAT GGA TTT ATT ACC GGC TAT AAA ATT CGA CAC    2590

   Arg Lys Thr Thr Arg Arg Gly Glu Met Glu Thr Leu Glu Pro Asn Asn     879
   AGA AAG ACG ACC CGC AGG GGT GAG ATG GAA ACA CTG GAG CCA AAC AAC    2638

   Leu Trp Tyr Leu Phe Thr Gly Leu Glu Lys Gly Ser Gln Tyr Ser Phe     895
   CTC TGG TAC CTA TTC ACA GGA CTG GAG AAA GGA AGT CAG TAC AGT TTC    2686

   Gln Val Ser Ala Met Thr Val Asn Gly Thr Gly Pro Pro Ser Asn Trp     911
   CAG GTG TCA GCC ATG ACA GTC AAT GGT ACT GGA CCA CCT TCC AAC TGG    2734

   Tyr Thr Ala Glu Thr Pro Glu Asn Asp Leu Asp Glu Ser Gln Val Pro     927
   TAT ACT GCA GAG ACT CCA GAG AAT GAT CTA GAT GAA TCT CAA GTT CCT    2782
```

Figure 4
(Page 3 of 6)

```
---  --- Gln Pro Se  Ser Leu His Val Arg Pro Gln Th. [redacted]
...  GAT CAA CCA AGC TCT CTT CAT GTG AGG CCC CAG ACT GAC TGC ... ATC   28?

 944 Met Ser Trp Thr Pro Pro Leu Asn Pro Asn Ile Val Val Arg Gly Tyr    ?.
2831 ATG AGT TGG ACT CCT CCC TTG AAC CCA AAC ATC GTG GTG CGA GGT TAT   28?

 960 Ile Ile Gly Tyr Gly Val Gly Ser Pro Tyr Ala Glu Thr Val Arg Val    97
2879 ATT ATC GGT TAT GGC GTT GGG AGC CCT TAC GCT GAG ACA GTG CGT GTG   29?

 976 Asp Ser Lys Gln Arg Tyr Tyr Ser Ile Glu Arg Leu Glu Ser Ser Ser    9?
2927 GAC AGC AAG CAG CGA TAT TAT TCC ATT GAG AGG TTA GAG TCA AGT TCC   29?

 992 His Tyr Val Ile Ser Leu Lys Ala Phe Asn Asn Ala Gly Glu Gly Val   10?
2975 CAT TAT GTA ATC TCC CTA AAA GCT TTT AAC AAT GCC GGA GAA GGA GTT   30?

1008 Pro Leu Tyr Glu Ser Ala Thr Thr Arg Ser Ile Thr Asp Pro Thr Asp   10?
3023 CCT CTT TAT GAA AGT GCC ACC ACC AGG TCT ATA ACC GAT CCC ACT GAC   30?

1024 Pro Val Asp Tyr Tyr Pro Leu Leu Asp Asp Phe Pro Thr Ser Val Pro   10?
3071 CCA GTT GAT TAT TAT CCT TTG CTT GAT GAT TTC CCC ACC TCG GTC CCA   31?

1040 Asp Leu Ser Thr Pro Met Leu Pro Pro Val Gly Val Gln Ala Val Ala   10?
3119 GAT CTC TCC ACC CCC ATG CTC CCA CCA GTA GGT GTA CAG GCT GTG GCT   31?

1056 Leu Thr His Asp Ala Val Arg Val Ser Trp Ala Asp Asn Ser Val Pro   10?
3167 CTT ACC CAT GAT GCT GTG AGG GTC AGC TGG GCA GAC AAC TCT GTC CCT   32?

1072 Lys Asn Gln Lys Thr Ser Glu Val Arg Leu Tyr Thr Val Arg Trp Arg   10?
3215 AAG AAC CAA AAG ACG TCT GAG GTG CGA CTT TAC ACC GTC CGG TGG AGA   32?

1088 Thr Ser Phe Ser Ala Ser Ala Lys Tyr Lys Ser Glu Asp Thr Thr Ser   11?
3263 ACC AGC TTT TCT GCA AGT GCA AAA TAC AAG TCA GAA GAC ACA ACA TCT   33?

1104 Leu Ser Tyr Thr Ala Thr Gly Leu Lys Pro Asn Thr Met Tyr Glu Phe   11?
3311 CTA AGT TAC ACA GCA ACA GGC CTC AAA CCA AAC ACA ATG TAT GAA TTC   33?

1120 Ser Val Met Val Thr Lys Asn Arg Arg Ser Ser Thr Trp Ser Met Thr   11?
3359 TCG GTC ATG GTA ACA AAA AAC AGA AGG TCC AGT ACT TGG AGC ATG ACT   34?

1136 Ala His Ala Thr Thr Tyr Glu Ala Ala Pro Thr Ser Ala Pro Lys Asp   11?
3407 GCA CAT GCC ACC ACG TAT GAA GCA GCC CCC ACC TCT GCT CCC AAG GAC   34?

1152 Phe Thr Val Ile Thr Arg Glu Gly Lys Pro Arg Ala Val Ile Val Ser   11?
3455 TTT ACA GTC ATT ACT AGG GAA GGG AAG CCT CGT GCC GTC ATT GTG AGT   35?

1168 Trp Gln Pro Pro Leu Glu Ala Asn Gly Lys Ile Thr Ala Tyr Ile Leu   11?
3503 TGG CAG CCT CCC TTG GAA GCC AAT GGG AAA ATT ACT GCT TAC ATC TTA   35?

1184 Phe Tyr Thr Leu Asp Lys Asn Ile Pro Ile Asp Asp Trp Ile Met Glu   1?
3551 TTT TAT ACC TTG GAC AAG AAC ATC CCA ATT GAT GAC TGG ATT ATG GAA   3?

1200 Thr Ile Ser Gly Asp Arg Leu Thr His Gln Ile Met Asp Leu Asn Leu   1?
3599 ACA ATC AGT GGT GAT AGG CTT ACT CAT CAA ATC ATG GAT CTC AAC CTT   3?

1216 Asp Thr Met Tyr Tyr Phe Arg Ile Gln Ala Arg Asn Ser Lys Gly Val   1?
3647 GAT ACT ATG TAT TAC TTT CGA ATT CAA GCA CGA AAT TCA AAA GGA GTG   3?

1232 Gly Pro Leu Ser Asp Pro Ile Leu Phe Arg Thr Leu Lys Val Glu His   1?
3695 GGG CCA CTC TCT GAT CCC ATC CTC TTC AGG ACT CTG AAA GTG GAA CAC   3?

1248 Pro Asp Lys Met Ala Asn Asp Gln Gly Arg His Gly Asp Gly Gly Tyr   1?
3743 CCT GAC AAA ATG GCT AAT GAC CAA GGT CGT CAT GGA GAT GGA GGT TAT   3?
```

FIGURE 4
(Page 4 of 6)

```
1264  Trp Pro Val Asp T[illegible] Asn Leu Ile Asp Arg Ser Thr L[illegible] Asn Glu Pro  1279
      TGG CCA GTT GAT ACT AAT TTG ATT GAT AGA AGC [illegible]

1280  Pro Ile Gly Gln Met His Pro Pro His Gly Ser Val Thr Pro Gln Lys  1295
3839  CCA ATT GGA CAA ATG CAC CCC CCG CAT GGC AGT GTC ACT CCT CAG AAG  3886

1296  Asn Ser Asn Leu Leu Val Ile Ile Val Val Thr Val Gly Val Ile Thr  1311
3887  AAC AGC AAC CTG CTT GTG ATC ATT GTG GTC ACC GTT GGT GTC ATC ACA  3934

1312  Val Leu Val Val Val Ile Val Ala Val Ile Cys Thr Arg Arg Ser Ser  1327
3935  GTG CTG GTA GTG GTC ATC GTG GCT GTG ATT TGC ACC CGA CGC TCT TCA  3982

1328  Ala Gln Gln Arg Lys Lys Arg Ala Thr His Ser Ala Gly Lys Arg Lys  1343
3983  GCC CAG CAG AGA AAG AAA CGG GCC ACC CAC AGT GCT GGC AAA AGG AAG  4030

1344  Gly Ser Gln Lys Asp Leu Arg Pro Pro Asp Leu Trp Ile His His Glu  1359
4031  GGC AGC CAG AAG GAC CTC CGA CCC CCT GAT CTT TGG ATC CAT CAT GAA  4078

1360  Glu Met Glu Met Lys Asn Ile Glu Lys Pro Ser Gly Thr Asp Pro Ala  1375
4079  GAA ATG GAG ATG AAA AAT ATT GAA AAG CCA TCT GGC ACT GAC CCT GCA  4126

1376  Gly Arg Asp Ser Pro Ile Gln Ser Cys Gln Asp Leu Thr Pro Val Ser  1391
4127  GGA AGG GAC TCT CCC ATC CAA AGT TGC CAA GAC CTC ACA CCA GTC AGC  4174

1[illegible]  His Ser Gln Ser Glu Thr Gln Leu Gly Ser Lys Ser Thr Ser His Ser  1407
4175  CAC AGC CAG TCA GAA ACC CAA CTG GGA AGC AAA AGC ACC TCT CAT TCA  4222

1408  Gly Gln Asp Thr Glu Glu Ala Gly Ser Ser Met Ser Thr Leu Glu Arg  1423
4223  GGT CAA GAC ACT GAG GAA GCA GGG AGC TCT ATG TCC ACT CTG GAG AGG  4270

1424  Ser Leu Ala Ala Arg Arg Ala Pro Arg Arg Lys Leu Met Ile Pro Met  1439
4271  TCG CTG GCT GCA CGC CGA GCC CCC CGG CGC AAG CTC ATG ATT CCC ATG  4318

1440  Asp Ala Gln Ser Asn Asn Pro Ala Val Val Ser Ala Ile Pro Val Pro  1455
4319  GAT GCC CAG TCC AAC AAT CCT GCT GTC GTG AGC GCC ATC CCG GTG CCA  4366

1456  Thr Leu Glu Ser Ala Gln Tyr Pro Gly Ile Leu Pro Ser Pro Thr Cys  1471
4367  ACG CTA GAA AGT GCC CAG TAC CCA GGA ATC CTC CCG TCT CCC ACC TGT  4414

1472  Gly Tyr Pro His Pro Gln Phe Thr Leu Arg Pro Val Pro Phe Pro Thr  1487
4415  GGA TAT CCC CAC CCG CAG TTC ACT CTC CGG CCT GTG CCA TTC CCA ACA  4462

1488  Leu Ser Val Asp Arg Gly Phe Gly Ala Gly Arg Ser Gln Ser Val Ser  1503
4 [illegible]  CTC TCA GTG GAC CGA GGT TTC GGA GCA GGA AGA AGT CAG TCA GTG AGT  4510

1504  Glu Gly Pro Thr Thr Gln Gln Pro Pro Met Leu Pro Pro Ser Gln Pro  1519
4511  GAA GGA CCA ACT ACC CAA CAA CCA CCT ATG CTG CCC CCA TCT CAG CCT  4558

1520  Glu His Ser Ser Ser Glu Glu Ala Pro Ser Arg Thr Ile Pro Thr Ala  1535
4559  GAG CAT TCT AGC AGC GAG GAG GCA CCA AGC AGA ACC ATC CCC ACA GCT  4606

1536  Cys Val Arg Pro Thr His Pro Leu Arg Ser Phe Ala Asn Pro Leu Leu  1551
4607  TGT GTT CGA CCA ACT CAC CCA CTC CGC AGC TTT GCT AAT CCT TTG CTA  4654

1552  Pro Pro Pro Met Ser Ala Ile Glu Pro Lys Val Pro Tyr Thr Pro Leu  1567
4655  CCT CCA CCA ATG AGT GCA ATA GAA CCG AAA GTC CCT TAC ACA CCA CTT  4702

1568  Leu Ser Gln Pro Gly Pro Thr Leu Pro Lys Thr His Val Lys Thr Ala  1583
4703  TTG TCT CAG CCA GGG CCC ACT CTT CCT AAG ACC CAT GTG AAA ACA GCC  4750

1584  Ser Leu Gly Leu Ala Gly Lys Ala Arg Ser Pro Leu Leu Pro Val Ser  1599
4751  TCC CTT GGG TTG GCT GGA AAA GCA AGA TCC CCT TTG CTT CCT GTG TCT  4798
```

FIGURE 4
(Page 5 of 6)

GeneDX 1009, pg. 638

9/15

```
      Val Pro Thr Al  Pro Glu Val Ser Glu Glu Ser H  [illegible]
4792  GTG CCA ACA GCC CCT GAA GTG TCT GAG GAG AGC C[illegible]

1616  Asp Ser Ala Asn Val Tyr Glu Gln Asp Asp Leu Ser Glu Gln Met Ala  1631
4847  GAT TCA GCC AAT GTG TAT GAA CAG GAT GAT CTG AGT GAA CAA ATG GCA  4894

1632  Ser Leu Glu Gly Leu Met Lys Gln Leu Asn Ala Ile Thr Gly Ser Ala  1647
4895  AGT TTG GAA GGA CTC ATG AAG CAG CTT AAT GCC ATC ACA GGC TCA GCC  4942

                 4 END
1648  [redacted]*** His Val Phe Leu Asn Gly *** Gly Glu Phe Ser Gly Asn Phe  1663
4943  TTT TAA CAT GTA TTT CTG AAT GGA TGA GGT GAA TTT TCC GGG AAC TTT  4990

1664  Ala Ala Tyr Gln Leu Pro Ile Asn Ser Thr Pro Val Ser Lys Asn Ser  1679
4991  GCA GCA TAC CAA TTA CCC ATA AAC AGC ACA CCT GTG TCC AAG AAC TCT  5038

1680  Asn Gln Cys Thr Gly His Pro Ser Gly Pro Leu Ser *** Gly Arg Ser  1695
5039  AAC CAG TGT ACA GGT CAC CCA TCA GGA CCA CTC AGT TAA GGA AGA TCC  5086

1696  *** Ser Ser Ser Glu Gly Ile Ser Ile Pro Ser Phe Thr Gly Ile Arg  1711
5087  TGA AGC AGT TCA GAA GGA ATA AGC ATT CCT TCT TTC ACA GGC ATC AGG  5134

1712  Asn Cys Gln Met Met Ile Met Ser Ser Leu Asn Lys Ser Lys Asp Ala  1727
5135  AAT TGT CAA ATG ATG ATT ATG AGT TCC CTA AAC AAA AGC AAA GAT GCA  5182

1728  Phe
5183  TTT T
```

FIGURE 4
(Page 6 of 6)

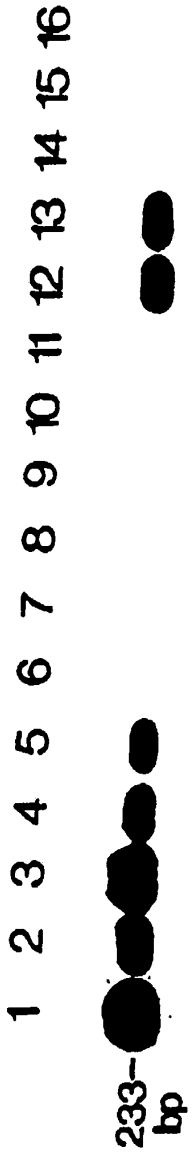


FIGURE 5

11/15

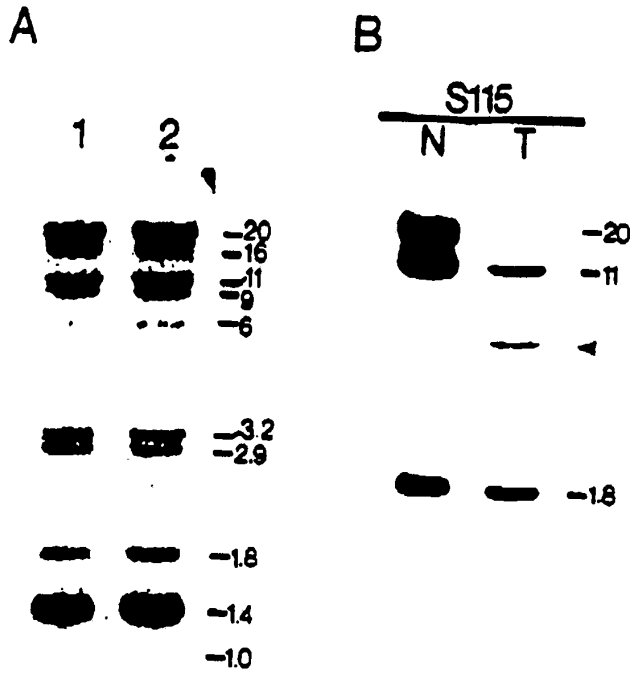


FIGURE 6

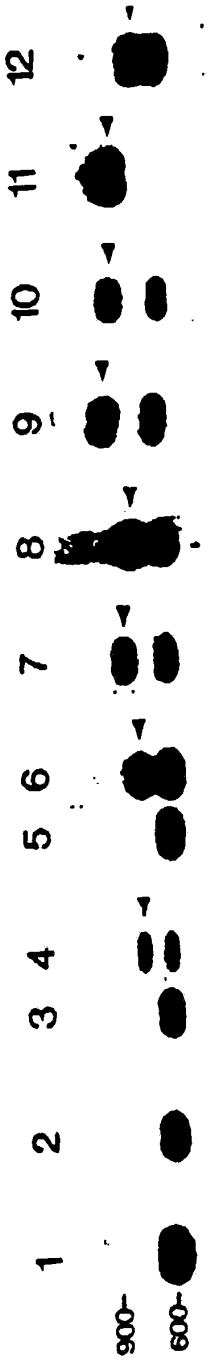


FIGURE 7A

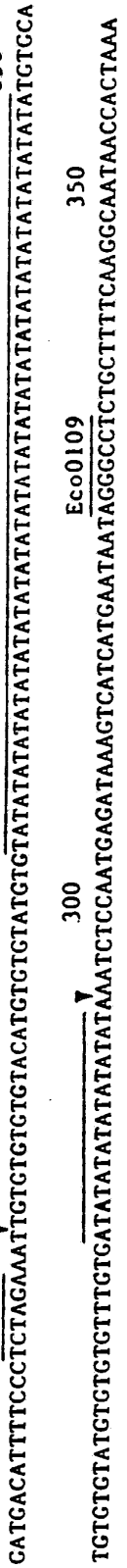

150
XbaI
200
250
GATGACATTTTCCCTCTAGAAATTGTGTGTGTACATGTGTATGTGTATATATATATATATATATATATATATATGTGCA
300
Eco0109
350
TGTGTGTATGTGTGTTTGTGATATATATATATAAATCTCCAATGAGATAAAGTCATCATGAATAATAGGGCCTCTGCTTTTCAAGGCAATAACCACTAAA

FIGURE 7B

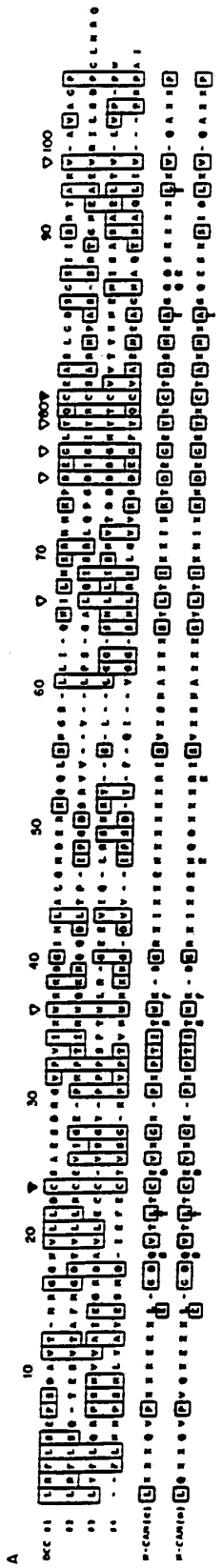


FIGURE 8A

15/15

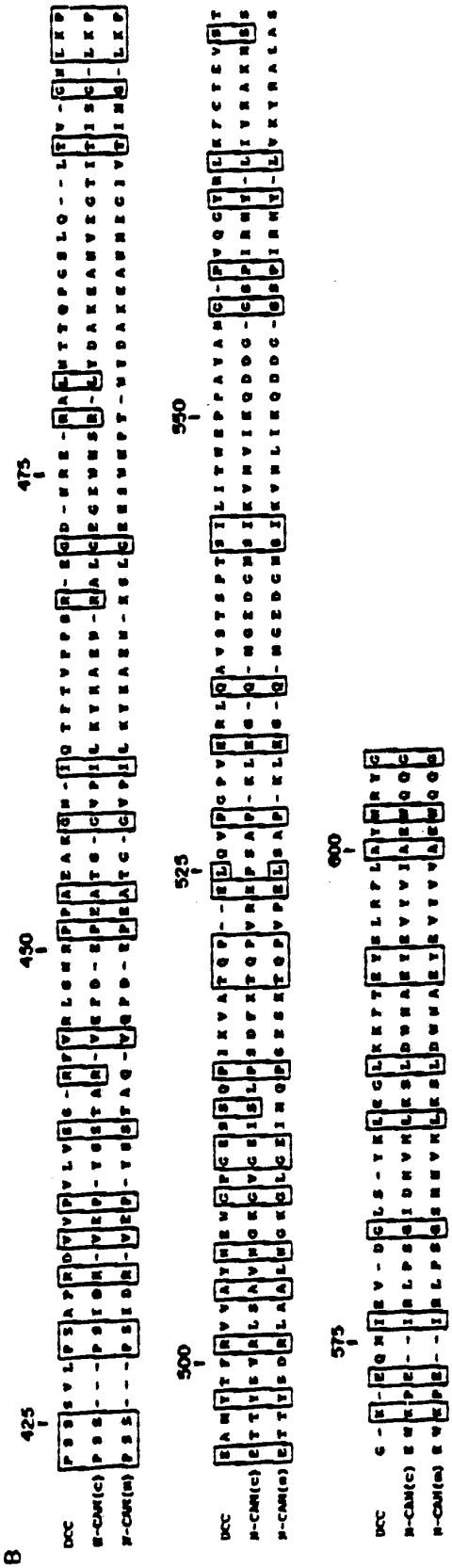


FIGURE 8B

GeneDX 1009, pg. 645

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 90/07314

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) *

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5): C12P 19/34; C12Q 1/68; G01 N 33/48 33/00, 33/564 33/566;C07 H 15/12, 17/00

US CL: 435/6; 435/91; 436/63; 436/94; 436/501; 436/508; 536/27; 536/28; 536/29

II. FIELDS SEARCHED

Minimum Documentation Searched 7

Classification System	Classification Symbols
U. S.	435/6,91; 436/63,94,501,508 536/27,28,29;

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched 8

III. DOCUMENTS CONSIDERED TO BE RELEVANT 9

Category *	Citation of Document, 11 with indication, where appropriate, of the relevant passages 12	Relevant to Claim No. 13
Y	Science, Volume 244, No. 4901, issued 14 April 1989,Baker et al., "Chromosome 17 Deletions and P. 53 Gene Mutations in Colorectal Carcinomas", pages 217-221, see abstract, p. 217, column 2, lines 3-10, p. 218, column 1, paragraph 4, lines 1-6, and p. 218, column 2, paragraph 2, lines 1-8.	1-17 & 25-42
Y	New England Journal of Medicine, Volume 319, No. 9, issued 01 September 1988, Vogelstein et al., "Genetic Alterations during Colorectal-Tumor Development", pages 525-532, see abstract and page 528, paragraphs 2 and 3.	1-17 & 25-42

* Special categories of cited documents: 10

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search

07 March 1991

Date of Mailing of this International Search Report

26 APR 1991

International Searching Authority

ISA/US

Signature of Authorized Officer

LORI YUAN

Form PCT/ISA/210 (second sheet) (1988)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, [16] with indication, where appropriate, of the relevant passages [17]	Relevant to Claim No [18]
Y	Journal of the American Medical Association. Volume 261, No. 21, issued 02 June 1989 Kern et al., "Clinical and Pathological Associations with Allelic Loss in Colorectal Carcinoma", pages 3099-3103, see abstract.	1-17 & 25-42
Y	Aaronson et al.,"Genetic and Phenotypic Markers of Tumors" published 1984 by Plenum Press (New York & London), pages 1-20 especially page 7, paragraphs 1-2 and page 1, paragraph 2.	1-17 & 25-42
Y	Chemical Abstracts, Volume 109, No. 19, issued 30 August 1988, Boman et al., "Chromosome 18 Allele Loss at the D1856 Locus in Human Colorectal Carcinomas," abstract No. 164792e, Biochemical and Biophysical Research Communications, 155(1) 463-9, See entire document.	1-17 & 25-42
Y	Old et al., "Principles of Gene Manipulation "published 1985 by Blackwell Scientific Publications , page 8-10 and 120-121, see entire document.	23 & 24
Y	Proceedings of the National Academy of Science U.S.A., Volume 74, No. 2, issued February 1977, Maxam et al., "A New Method for Sequencing DNA," pages 560-564, see entire document.	23 & 24
Y	Proceedings of the National Academy of Science U.S.A., Volume 74, No. 12, issued December 1977, Sanger et al. "DNA Sequencing with Chain-Terminating Inhibitors," pages 5463-5467, see entire document.	23 & 24

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers , because they relate to subject matter [1] not required to be searched by this Authority, namely:

2. ☐ Claim numbers , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [1], specifically:

3. ☐ Claim numbers_____, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

Claims 1-17, 23, 24, 25-42.

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

GeneDX 1009, pg. 648

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07H 15/12, 17/00, C07K 15/00
C07K 15/28

A1

(11) International Publication Number: **WO 92/00311**

(43) International Publication Date: 9 January 1992 (09.01.92)

(21) International Application Number: PCT/US91/04608

(22) International Filing Date: 27 June 1991 (27.06.91)

(30) Priority data:
543,963 27 June 1990 (27.06.90) US

(71) Applicant: PRINCETON UNIVERSITY [US/US]; Princeton, NJ 08540 (US).

(72) Inventors: LEVINE, Arnold, J. ; 138 Fritzrandolph Road, Princeton, NJ 08540 (US). SHENK, Thomas, E. ; 87 McCosh Circle, Princeton, NJ 08540 (US). FINLAY, Cathy, A. ; 660 Bayberry Lane, Yardley, PA 19067 (US).

(74) Agent: FEIT, Irving, N.; Imclone Systems Incorporated, 180 Varick Street, New York, NY 10014 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GB (European patent), GR (European patent), HU, IT (European patent), JP, KR, LU (European patent), NL (European patent), SE (European patent), SU.

Published
With international search report.

(54) Title: PROBES FOR DETECTING MUTANT P53

(57) Abstract

A panel of probes detects and distinguishes between sets of human p53 gene or protein mutations that frequently occur or are selected for in pre-cancer and cancer cells. Each set gives rise to a phenotype that is different from that of wild-type p53 and of at least one other set of p53 mutants.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**ES**	Spain	**MG**	Madagascar
AU	Australia	**FI**	Finland	**ML**	Mali
BB	Barbados	**FR**	France	**MN**	Mongolia
BE	Belgium	**GA**	Gabon	**MR**	Mauritania
BF	Burkina Faso	**GB**	United Kingdom	**MW**	Malawi
BG	Bulgaria	**GN**	Guinea	**NL**	Netherlands
BJ	Benin	**GR**	Greece	**NO**	Norway
BR	Brazil	**HU**	Hungary	**PL**	Poland
CA	Canada	**IT**	Italy	**RO**	Romania
CF	Central African Republic	**JP**	Japan	**SD**	Sudan
CG	Congo	**KP**	Democratic People's Republic of Korea	**SE**	Sweden
CH	Switzerland			**SN**	Senegal
CI	Côte d'Ivoire	**KR**	Republic of Korea	**SU**	Soviet Union
CM	Cameroon	**LI**	Liechtenstein	**TD**	Chad
CS	Czechoslovakia	**LK**	Sri Lanka	**TG**	Togo
DE	Germany	**LU**	Luxembourg	**US**	United States of America
DK	Denmark	**MC**	Monaco		

PROBES FOR DETECTING MUTANT P53

The present invention is directed to the use of molecular probes in the detection of cancer and pre-cancer states. More particularly, the invention is directed to different cancer and pre-cancer states by means of antibody and DNA probes. The cancer and pre-cancer states are those associated with the p53 protein.

Mutations of proto-oncogenes in somatic cells are increasingly being recognized as significant in the induction of human cancers. Some examples of oncogenes formed by such mutations include: neu, fes, fos, myc, myb, fms, Ha-ras, and Ki-ras. The mutations that convert proto-oncogenes to oncogenes are often point mutations. Much needs to be learned in order to understand how oncogenes and their expression products function to transform normal cells to cancer cells.

Oncogenes are generally believed to act in a dominant fashion. This is generally considered to mean that the conversion of a proto-oncogene to an oncogene results in the acquisition of a new function, i.e., enhancing transformation.

A different type of mutation associated with cancer occurs when a tumor suppressor gene is altered in a way that causes the product of the gene to lose its tumor suppressor function. An example of such a tumor suppressor gene is the retinoblastoma sensitivity gene, Rb. Tumor suppressor genes are sometimes called recessive oncogenes, although, strictly

speaking, the products of tumor suppressor genes do not contribute to tumor formation. The phenotype is recessive since, when both alleles are mutated, the absence of a tumor suppressor gene results in an enhancement of tumorigenesis.

The products of certain viral oncogenes are also able to transform cells. Examples of such products include the E6 and E7 proteins from human papilloma virus, the large T antigen from SV40, and E1a from adenovirus. Viral oncogene proteins are believed to bind to and, thereby, to inactivate tumor suppressor proteins, such as the retinoblastoma protein.

A gene product that exhibits some properties of both a dominant and a recessive oncogene is the 53kd phosphoprotein, p53. Evidence is growing that mutations in the p53 gene is associated with a large number of many types of cancers. For example, Iggo et al, Lancet 335, 675-679 (1990) has expressed the opinion that p53 is the most common proto-oncogene to undergo mutation in lung cancers.

Much of what is known about p53 has been derived from studying the effect of transfecting wild-type and mutant murine p53 in rat embryo fibroblast cells. This work has been reviewed by Levine et al, "The P53 Proto-Oncogene And Its Product," in Common Mechanisms Of Transformation By Small DNA Tumor Viruses, L. Villarreal, ed., American Society for Microbiology, Chapter 2 (1989); Hinds et al, ibid, Chapter 7; and Levine, BioEssays 12, 60-66 (1990).

Briefly, a number of point mutations between amino acids 130 and 240 of p53 (out of 390 amino acids) lead to significant, tumor-promoting changes in phenotype. Both wild-type and mutant p53 are often found at increased levels in transformed cells due to an increase in their metabolic

stabilities. The stabilization of mutant p53 is believed to occur through the formation of a complex with cellular proteins, such as the 70kd heat shock protein, hsc70. The stabilization of wild-type p53 is believed to be associated with its ability to form a complex with the mutant p53-hsc70 complex. These results are consistent with the proposition that alterations in p53 function are involved in the process of cellular transformation. The involvement of mutant murine p53 in the transformation of cells in culture is also apparent from the ability of mutant, but not wild-type, p53 to cooperate with activated Ha-ras to transform primary rat embryo cells.

The p53 gene resides on chromosome 17p. Many cancers, such as those discussed above, are associated with chromosome 17p deletions. Such allelic deletions often indicate the presence of a tumor suppressor gene. The mutation of one allele gives rise to a benign, pre-cancer state. The mutation of the second allele gives rise to the malignant cancer state.

Finlay et al, Cell 57, 1083-1093 (1989), has presented further evidence that wild-type murine p53 displays properties of a suppressor of transformation. Three observations are consistent with this theory.

First, the introduction of wild-type murine p53 into primary rodent cells, along with two cooperating transforming genes, ras and E1a, results in a decrease in the number of transformed foci. The transformed cell lines that were obtained were found to contain the murine p53 gene, but either failed to express it or produced high levels of an altered murine product. Thus, overexpression of the wild-type murine p53 protein appears to be detrimental to the process of transformation of cultured rat

cells by oncogenes.

Second, inactivation of the p53 gene is believed to be associated with the development of Friend virus-induced erythroleukemia in mice (Mowat et al, Nature 314, 633-636 (1985)). There are numerous examples in the literature of tumor cells derived from the spleens of mice infected with the Friend virus complex containing rearrangements or other mutations at the p53 gene locus; see, for example, Ben-David, Oncogene 3, 179-185 (1988).

The third line of evidence consistent with the possibility that the wild-type p53 protein is a member of a group of proteins involved in suppression of transformation is the ability, mentioned above, of p53 to form oligomeric protein complexes with viral oncogenes, such as the SV40 large T antigen, the adenovirus type 5 E1b-55 kd protein, and the human papilloma virus (HPV) type 16 or 18 E6 product; see, for example, Werness, Science 248, 76-79 (1990). Analogous complexes have also been observed between p105, the product of the retinoblastoma susceptibility gene, and the SV40 large T antigen (DeCaprio et al, Cell 54, 275-283 (1988)); the adenovirus E1a protein (Whyte et al, Nature 334, 124-129 (1988)); and the E7 protein of HPV-16 or -18 (Muenger et al, EMBO J. 8, 4099-4105 (1989)).

These interactions between viral proteins and p105 are thought to inactivate a growth-suppressive function of p105, thus mimicking deletions and mutations commonly found in the retinoblastoma gene in tumor cells. Similarly, oligomeric protein complex formation between these same viral proteins and p53 may eliminate or alter the growth-suppressive function of p53; see Finlay et al, id.

The clonal nature of p53-related tumors is consistent with a tumor progression model in which non-neoplastic pre-cancer cells bearing a wild-type p53 gene and a mutated p53 gene have a distinct proliferative advantage over normal cells, which contain two wild-type genes. The advantage is due to mutant p53-mediated interference with wild-type p53 function. The increased proliferative capacity of such non-neoplastic cells increases the probability of a second, inactivating mutation, i.e., gene conversion or deletion, at the p53 locus. The resulting cells, which now contain mutations in both p53 alleles, are able to express the fully neoplastic phenotype; see Finlay et al, id., and Baker et al, Science 244, 217-221 (1989).

The above model is supported by the discovery that human tumor cells from which a 17p chromosome allele been deleted contain mutations in the remaining allele. The mutations tended to be clustered in four "hot spots," which coincided with the four most highly conserved regions of the p53 gene; see Nigro et al, Nature 342, 705-708 (1989).

Gannon et al, EMBO J. 9, 1595-1602 (1990), propose that all human mutant p53 proteins are recognized by a monoclonal antibody, PAb240. These authors suggest that all p53 mutants exert a common conformational effect, which results in expression of the PAb240 epitope.

OBSERVATIONS OF THE INVENTORS THAT FORM THE BASIS OF THE PRESENT INVENTION

Previously unpublished experiments of the inventors relating to human p53 form the basis of the present invention. Different human p53 clones isolated from colorectal carcinomas possess mutations at amino acid residues 143, 175, 273 or 281 (out of a total of 393

residues). Such p53 mutants, when co-transfected into rat embryo fibroblasts (REFs) with activated wild-type ras oncogenes, cooperate with the oncogenes to transform the REFs in culture. All of the transformed cell lines derived from these experiments produce the human p53 protein in elevated levels.

The mutations are summarized in Table 1. The effect of the mutations on properties of human p53 mutant proteins is shown in Table 2.

Table 1.

Mutations of Human p53 Protein

Clone	DNA Nucleotide	Alteration	Protein Residue	Alteration
p53-c143A	430	T -> C	143	val -> ala
p53-175H	526	G -> A	175	arg -> his
p53-273H	820	G -> A	273	arg -> his
p53-281G	844	A -> G	281	asp -> gly

Table 2.

Properties of Human p53 Mutant Proteins

Clone	Tumors in Mice Frequency	Relative Tx Half-life	hsc Protein Bound	p90 Bound	Nude Bound
p53-cWT[1]	0	20 min.	-	+	+
p53-c143A[1]	1.6	1.5-2 hrs.	+	+	+
p53-WT[2]	0	ND[3]	ND[3]	ND[3]	ND[3]
p53-175H[2]	11.5	3.6-6.4 hrs.	+	+	+
p53-273H[2]	4.7	7 hrs.	-	+	+
p53-281G[2]	1.9	3.5 hrs.[4]	-	+	ND[3]

1 - cDNA. Does not contain introns.
2 - contains introns.
3 - Not Determined
4 - half-life estimated in cell lines expressing ras + E1a + mutant p53

In order to obtain the results shown in Table 2, cDNA clones or partial cDNA-genomic clones of p53, plus the activated ras oncogene, were co-transfected into primary rat embryo fibroblasts. In each set of transfections, transformed foci were scored two to three weeks later in duplicate cell cultures.

The results in Table 2 demonstrate that mutant human p53 cDNA or cDNA-genomic hybrid clones derived from colon carcinomas can behave as dominant oncogenes and cooperate with the ras oncogene in transforming rat embryo fibroblasts. Four different missense mutations, at amino acids residues 143, 175, 273 and 281, each contributed to the transformed phenotype. In all these cases, the human p53 mutant protein was produced in high levels in the transformed cell at least in part due to the extended half-life of these mutant proteins.

Unexpectedly, the phenotypes of cells containing the different mutants are not the same. For example, in one experiment, the three p53 DNA clones containing "hot spot" mutations at codons 175, 273, and 281, i.e., p53-175H, p53-273H, and p53-281G, had characteristic and reproducible transformation frequencies (number of foci produced) in a ratio of 6:2.4:1, indicating that these "hot spot" mutations are not equivalent in their phenotypes. In another experiment, the 175 mutant allele was 3-10 times more efficient than the 273 mutant allele in cooperating with the ras oncogene in transforming primary rat cells in culture.

More striking is the fact that p53-175H and p53-c143A mutant proteins bind to hsc70 in transformed cells while the p53-273H and p53-281G mutant proteins do not detectably interact with or bind to hsc70. Previous experiments have shown that some mutant murine p53 proteins have an altered

conformation (Hinds et al, Mol. Cell. Biol. 7, 2863-2869 (1987); Finlay et al, Ibid. 8, 531-539 (1988)). It is possible that such altered p53 molecules bind to hsc70 and sequester wild-type p53 in a complex that blocks proper folding, assembly or localization of p53. Thus, p53-c143A and p53-175H may represent mutant proteins which never fold correctly and thus retain their affinity for hsc70. If wild-type p53 is recruited into this complex, the p53-mutant-hsc70-p53-wild-type complex would poison the function of wild-type p53.

This cannot be the case for the p53-273H and p53-281G mutants., which, as mentioned above, do not bind to hsc70. The p53-273H human mutant protein can associate with the rat p53 protein and form an oligomeric complex as demonstrated by co-immunoprecipitation with a human-specific antibody, Ab-2. This complex is not mediated by hsc70 and may be less efficient in sequestering the rat cellular wild-type p53 protein. This is consistent with a poorer ability to transform cells in culture.

In contrast to co-transfections with activated ras plus mutant human p53, co-transfections with activated ras plus wild-type human p53 resulted in a very low frequency of focus formation, and only one focus could be cloned into an established cell line. Thus, it appears likely that mutation of human p53 activates a dominant transforming function that is not detectable in wild-type p53.

In summary, it now appears that all of the human p53 mutant proteins differ from the wild-type human p53 protein by having an extended half-life, by being expressed at higher levels, and by possessing the ability to transform cells in culture. These data support the suggestion that mutation of p53 on one allele could have a growth-promoting

phenotype in vivo, which expands the number of cells with such mutations and favors the selection of a second mutational event (deletion or gene conversion) in the cancer cells. The observation that many tumor cells retain only the mutant p53 allele suggests that these mutant proteins are not fully dominant over the wild-type allele, or that the mutants continue to confer a proliferative advantage on cells in the absence of wild-type p53. Such a positive effect of mutant p53 on cell proliferation in the absence of wild-type p53 could be an intrinsic function of the molecule, or could be mediated by titration of cellular proteins other than endogenous p53, for example, via the protein p90.

Most significantly, the results establish that there are classes of human p53 mutations that give rise to pre-cancer and cancer cells with different phenotypes. It is apparent that these different phenotypes can give rise to cancers that take different courses and have different prognoses.

A problem with present methods to detect cancer is that these different phenotypic changes have not been taken into account. This problem is addressed by the present invention.

A second problem addressed by the present invention is to determine how the different phenotypic changes affect the course and prognosis of cancer. Once this problem has been solved, a further problem is to be able to detect the various phenotypic changes in order to be able to predict the course a cancer will take and the best way to treat such cancer.

SUMMARY OF THE INVENTION

These and other problems as will be apparent to those having ordinary skill in the art have been solved by providing a panel of probes that detect and distinguish between sets of human p53 gene or protein mutations that frequently occur or are selected for in pre-cancer and cancer cells, each set giving rise to a phenotype that is different from that of wild-type p53 and of at least one other set of p53 mutants.

The invention further relates to a method of distinguishing between sets of human p53 gene or protein mutations that are frequently occurring or selected for in pre-cancer or cancer cells, each set giving rise to a phenotype that is different from the wild-type gene or protein and from at least one other set of mutations, the method comprising determining the mutations in a sample of p53 genes or proteins with a panel of probes that detect and distinguish between such sets.

The invention also provides a method of distinguishing between sets of human p53 protein mutations that are frequently occuring or selected for in pre-cancer or cancer cells, the method comprising the steps of determining whether such a mutation exists and, if so, whether the mutant protein binds to hsc70 less tightly than wild-type human p53 protein.

DETAILED DESCRIPTION OF THE INVENTION

DEFINITIONS

The present invention is directed to the detection of mutations in wild-type p53 genes and proteins. For the

purposes of the present specification, the term "wild-type" p53 means the nucleotide or amino acid sequence reported by Matlashewski et al, EMBO J. 13, 3257-3262 (1984); Zakut-Houri et al, EMBO J. 4, 1251-1255 (1985); and Lamb and Crawford, Mol. Cell. Biol. 5, 1379-1385 (1986). The sequences are available from GenBank. Wild-type p53 includes a proline/arginine polymorphism at amino acid 72 and the corresponding nucleotide polymorphism.

The detection of mutations in wild-type p53 genes and proteins is important, since such mutations indicate pre-cancer and cancer states. There are no apparent limitations in regard to the type of cancer that is associated with a p53 mutation. Such cancer include, generally, colorectal, lung, ovarian, cervical, adrenal cortex, bone, bladder, breast, brain, and mesenchyme cancers and, more specifically, chronic myelocytic leukemia, chronic myelogenous leukemia, and osteogenic sarcomas.

A pre-cancer cell is defined as a cell that has one normal p53 allele and one mutated p53 allele. The mutation is usually a point mutation.

In a cancer cell, both alleles are mutated. One mutation is usually a point mutation, as described above for a pre-cancer cell. The other mutation is usually a deletion of all or a significant part of the p53 gene.

MUTATIONS

The invention is based on the unexpected discovery that there are sets of mutations in p53 that correspond to sets of different conformations and different phenotypes. It is important to determine as many sets of mutations as possible. It is not, however, necessary to determine each individual mutation within a set. A set of mutations is

defined as having at least one mutation and giving rise to a phenotype that is different from the wild type and at least one other set of mutations.

Each set of phenotypes leads to a distinguishable course and severity of the same disease. Therefore, by determining sets of mutations, a physician can not only determine that a patient has a particular cancer, but can distinguish different subsets of prognoses and prescribe the best treatment.

For example, many colorectal cancers give rise to mutations at or near amino acid positions 143, 175, 273 and 281 of p53. These mutations are described in Table 1.

Families suffering with the Li-Fraumeni syndrome contain p53 mutations that cluster between codons 245 and 258. A mutation at codon 248 is particularly prevalent. These families have a high incidence of cancer.

Mutations in the same region of p53 implicated in the Li-Fraumeni syndrome are also frequently found in hepatocellular carcinomas. A mutation at codon 249 is frequently found in patients with liver cancer.

It should be emphasized that any single nucleotide or amino acid that is mutated or any region of the p53 gene or protein within which one or more mutation occurs constitutes a set of mutations in accordance with the present invention as long as the set of mutations satisfies the definition given above. In one embodiment of the invention, for example, sets of mutations comprise residues 117-142, 171-181, 234-258, and 270-286 of the p53 protein as well as the corresponding nucleotides of the p53 gene.

In another embodiment of the invention, one set of mutations comprises mutations at amino acids 143 and 175. A second set of mutations comprises mutations at amino acids 273 and 281.

PANEL OF PROBES

In general, a panel of probes in accordance with the invention includes at least two members. There may, for example, be as many as three, four, five or six sets of mutations and, therefore, the same number of probes. There may be as many as 8, 10, or 12 probes, and, in fact, even more.

The mutations in a set are commonly found in various types of human tumors. The ubiquitous nature of these mutations may result from their frequent occurrence or, as explained above, because they provide a proliferative advantage to cells containing them and, therefore, are selected for.

Alterations in either the nucleotide sequence of the gene or the amino acid sequence of the protein may be assayed in order to determine whether a mutation within one of the sets of mutations in accordance with the present invention exists. Alterations in the amino acid sequence may be probed by antibodies. Alterations in the nucleotide sequence may be probed by means of nucleotide probes or, in some cases, by restriction endonucleases. Nucleotides as used herein refer to RNA or DNA.

ANTIBODY PROBES

An "antibody" in accordance with the present specification is defined broadly as a polypeptide that binds

specifically to an epitope. The antibody may be polyclonal or monoclonal. Antibodies further include recombinant polyclonal or monoclonal Fab fragments prepared in accordance with the method of Huse et al, Science 246, 1275-1281 (1989).

Methods for preparing polyclonal and monoclonal antibodies that exhibit specificity toward single amino acid differences between oncogenes are described by McCormick et al in U.S. Patent 4,798,787. These methods are incorporated herein by reference.

Briefly, polyclonal antibodies may be produced by injecting a host mammal, such as a rabbit, mouse, rat, or goat, with the p53 protein or a fragment thereof capable of producing antibodies that distinguish between mutant p53 and wild-type p53. The peptide or peptide fragment injected may contain the wild-type sequence or the mutant sequence. Sera from the mammal are extracted and screened to obtain polyclonal antibodies that are specific to the peptide or peptide fragment.

In order to produce monoclonal antibodies, a host mammal is inoculated with a peptide or peptide fragment as described above, and then boosted. Spleens are collected from inoculated mammals a few days after the final boost. Cell suspensions from the spleens are fused with a tumor cell in accordance with the general method described by Kohler and Milstein in Nature 256, 495-497 (1975). In order to be useful, a peptide fragment must contain sufficient amino acid residues to define the epitope of the p53 molecule being detected.

If the fragment is too short to be immunogenic, it may be conjugated to a carrier molecule. Some suitable carrier

molecules include keyhold limpet hemocyanin and bovine serum albumen. Conjugation may be carried out by methods known in the art. One such method is to combine a cysteine residue of the fragment with a cysteine residue on the carrier molecule.

The peptide fragments may be synthesized by methods known in the art. Some suitable methods are described by Stuart and Young in "Solid Phase Peptide Synthesis," Second Edition, Pierce Chemical Company (1984).

A variety of assays are available for detecting proteins with labelled antibodies. Such methods may involve one step or two steps. In a one-step assay, the target p53 molecule, if it is present, is immobilized and incubated with a labelled antibody. The labelled antibody binds to the immobilized p53. After washing to remove unbound molecules, the sample is assayed for the presence of the label.

In a two-step assay, immobilized p53 is incubated with an unlabelled antibody. The p53-unlabelled antibody complex, if present, is then bound to a second, labelled antibody that is specific for the unlabelled antibody. The sample is washed and assayed for the presence of the label, as described above.

The label may be a radioactive atom, an enzyme, or a chromophoric moiety. Some examples of radioactive atoms include P^{32}, I^{125}, H^{3}, and C^{14}. Some examples of enzymes include horseradish peroxidase, alkaline phosphatase, beta-galactosidase, and glucose-6-phosphate dehydrogenase. Some examples of chromophoric moieties include fluorescein and rhodamine. The antibodies may be conjugated to these labels

by methods known in the art. For example, enzymes and chromophoric molecules may be conjugated to the antibodies by means of coupling agents, such as dialdehydes, carbodiimides, dimaleimides, and the like. Alternatively, conjugation may occur through a ligand-receptor pair. Some suitable ligand-receptor pairs include, for example, biotin-avadin or -streptavadin, and antibody-antigen.

OLIGONUCLEOTIDE PROBES

The probes of the present invention may also be oligonucleotides that distinguish wild-type from mutant DNA or RNA. The oligonucleotide probes may be prepared by methods known in the art. Suitable methods for synthesizing oligonucleotide probes are described by Caruthers in Science 230, 281-285 (1985).

The oligonucleotide probes may contain a sequence complementary to a sequence of wild-type or mutant p53 that comprises a nucleotide involved in a mutation. For example, the nucleotide involved in a mutation may be that at position 430, 526, 820, or 844 of wild-type or mutant p53.

The length of the oligonucleotide probe is not critical, as long as it is capable of hybridizing to a test sample containing wild-type or mutant p53 and distinguishing between the two. The oligonucleotide should contain at least 6 nucleotides, preferably at least 10 nucleotides, and, more preferably, at least 15 nucleotides.

The oligonucleotide probes are labelled for detection. The labels that can be conjugated to oligonucleotide probes for detection are the same as those that are conjugated to antibodies. Such labels are described above. Conjugating the labels to the oligonucleotides is achieved by methods

known in the art.

There is no upper limit to the length of the oligonucleotide probes. Longer probes are more difficult to prepare and require longer hybridization times. Therefore, the probe should not be longer than necessary. Normally, the oligonucleotide probe will not contain more than 50 nucleotides, preferably not more than 40 nucleotides, and, more preferably, not more than 30 nucleotides.

Methods for distinguishing wild-type oncogenes from mutants containing a single nucleotide change are described in PCT Application WO 87/07646. These methods are incorporated herein by reference.

Briefly, oligonucleotides containing either the wild-type or mutant sequence are hybridized under stringent conditions to dried agarose gels containing target p53 RNA or DNA digested with an appropriate restriction endonuclease. Suitable stringent conditions include two degrees below the calculated T_m of a perfect duplex. The oligonucleotide probe hybridizes to the target p53 detectably better when the probe and the target p53 are perfectly complementary.

The target p53 DNA is optionally amplified in order to improve the sensitivity of the assay. Amplification may be accomplished by methods known in the art. A suitable method is the polymerase chain reaction method, as described in Mullis et al, U.S. Patents 4,683,195 and 4,683,202.

A particularly convenient method for assaying a single point mutation by means of oligonucleotides is described in Segev, PCT Application WO 90/01069, licensed to ImClone

Systems Incorporated, New York City. This method is limited to cases wherein the nucleotide in wild type p53 that is mutated and the corresponding nucleotide in the mutant are not complementary.

Briefly, two oligonucleotide probes for each wild-type or mutated p53 strand being assayed are prepared. Each oligonucleotide probe is complementary to a sequence that straddles the nucleotide that either becomes or has been mutated. Thus, a gap is created between the two hybridized probes.

For example, in order to distinguish between wild type and mutant forms of p53, wherein the guanine at position 526 in the wild type form is mutated to adenine, probes that leave a gap at position 526 are prepared. The gap is filled with a mixture of a polymerase, a ligase, and the nucleotide complementary to that at position 526 to form a ligated oligonucleotide product. For example, if wild-type p53 is being detected, the gap is filled with cytosine. The mutant form will not be detected under these conditions.

On the other hand, if the mutant form is being detected, the gap will be filled with thymine. The wild-type p53 will not be detected under these conditions. Either of the oligonucleotides or the nucleotide filling the gap may be labelled by methods known in the art.

The ligated oligonucleotide product can be amplified by denaturing it from the p53, hybridizing it to additional oligonucleotide complement pairs, and filling the gap again, this time with the complement of the nucleotide that filled the gap in the first step.

To illustrate the method, structure (1) shows a pair of oligonucleotide probes hybridizing to wild-type p53 containing guanine at position 526. Structure (2) shows the gap between the two probes being filled with cytosine. Structure (3) shows the ligated oligonucleotide product from structure (2) hybridizing to two additional complementary oligonucleotides. Structure (4) shows the gap in Structure (3) being filled with guanine. This process can be repeated as often as is desired.

STRUCTURES

(1)
```
          ----------   ----------
---------------------- G ----------------------
                       526
```

(2)
```
          ---------- C ----------
---------------------- G ----------------------
```

(3)
```
          ---------- C ----------
          ----------   ----------
```

(4)
```
          ---------- C ----------
          ---------- G ----------
```

Following ligation of a ligated oligonucleotide product and, optionally, amplification, the oligonucleotide product is separated by size and the label is detected by methods known in the art. The description of the above procedure from PCT Application WO 90/01069 is incorporated herein by reference.

RESTRICTION ENDONUCLEASE PROBES

Mutations may also be detected if they create or abolish restriction sites. For example, the Hha I site is GCGC. The mutation in the human p53 gene at nucleotide 430

from thymine to cytosine creates an Hha I site. Such a mutation alters the amino acid sequence at residue 143 from valine to alanine; see Table 1.

A mutation of the human p53 gene at nucleotide 526 from guanine to adenine abolishes an Hha I site. Such a mutation causes an alteration at residue 175 from arginine to histidine; see Table 1.

Accordingly, the mutations indicated in Table 1 at residues 430 and 526 of the human p53 gene may be detected by restriction analysis; see Baker et al, Science 244, 217-221 (1989). Some additional examples of the use of restriction analysis to assay point mutations is given in Weinberg et al, U.S. Patent 4,786,718.

Some additional methods for distinguishing polynucleotide sequences differing by one nucleotide are described by De Ley et al, J. Bacteriol. 101, 738-754 (1970); Wood et al, Proc. Natl. Acad. USA 82, 1585-1588 (1985); Myers et al, Nature 313, 495-497 (1985); and Myers et al, Science 230, 1242-1246 (1985). These methods are incorporated herein by reference.

ASSAYS

The labeled probes described above are capable of distinguishing wild-type and sets of mutant forms of p53. Confirmation may be obtained by comparing the response of the probes to the different forms. The wild-type p53 gene and protein are known, and may be obtained by known methods, such as those described in Matlashewski et al, EMBO J. 13, 3257-3262 (1984); Zakut-Houri et al, EMBO J. 4, 1251-1255 (1985); and Lamb and Crawford, Mol. Cell. Biol. 5, 1379-1385 (1986). Mutants may be prepared from wild-type p53 by site-

directed mutagenesis; see, for example, Zoller and Smith, Nucl. Acids Res. 10, 6487-6500 (1982); Methods in Enzymology 100, 468-500 (1983); and DNA 3, 479-488 (1984).

Wild-type and mutant p53 structural genes may also be synthesized by known methods, such as by preparing overlapping double-stranded oligonucleotides, filling in the gaps, and ligating the ends together. The DNA may be cloned in a suitable host cell and expressed. The p53 DNA and protein may be recovered from the host cell. See, generally, Sambrook et al, "Molecular Cloning," Second Edition, Cold Spring Harbor Laboratory Press (1987).

Assays involving antibody and DNA probes are conducted in accordance with methods known in the art. The assay may be designed so that the probes test positive for wild-type p53 and negative for mutant p53. In such a case, it is preferable to use an oligonucleotide probe, which will not be affected by the specific mutation.

On the other hand, the probes may test positive for mutant p53 and negative for wild-type p53. In such a case, it is preferable to use antibodies, which detect protein. The preference for antibody probes is due to the presence of higher concentrations of mutant p53 protein than wild-type p53 protein in transformed cells.

It has unexpectedly been found that at least one set of human p53 mutants does not detectably bind to the heat shock protein hsc70 or, at least, binds significantly less tightly than wild-type p53. This set comprises mutations at amino acid positions 273 and 281. Therefore, this set may be distinguished from the other sets of mutants, such as the set that comprises mutations at amino acid positions 143 and 175, by determining whether any mutation exists and, if so,

whether the mutants bind to hsc70. Methods for determining relative binding affinities may be conducted by methods known in the art. For example, a method for determining whether a p53 protein binds to hsc70 is described by Finlay et al in Mol. and Cell. Biol. 8, 531-539 (1988) and by Hinds et al in Mol. and Cell. Biol., 7, 2863-2869 (1987). The method described in these papers, which is incorporated herein by reference, involves co-immunoprecipitation experiments with anti-p53 and anti-hsc70 antibodies.

A suitable antibody specific for hsc70, for example, may be prepared from the sera of rabbits immunized with the carboxy-terminal 21 amino acids of a 70kd heat shock protein family member, hsp70, as described in the Hinds et al article, Id. The method of preparing the antibody as described in the Hinds et al article is incorporated herein by reference.

p90

When the wild type or mutated p53 gene is transfected into rat embryo fibroblast (REF) cells, the REF cells express the p53 protein; see above. p53 may be isolated and purified by known methods. For example, the cells may be lysed and the p53 protein immunoprecipitated by adding p53-specific antibodies to solutions or suspensions of the protein. The immunoprecipitates are recovered by centrifugation. See Hinds et al., Cell Growth and Differentiation 1, 571-580 (1990); Finlay et al., Cell 57, 1083-1093 (1989) and Hinds et al., Molecular and Cell Biology 7, 2863-2869 (1987).

The co-immunoprecipitation may also be performed by means of an antibody affinity column using conditions suitable for purifying p53 with a PAb421 immunoaffinity

column. See Clarke et al. in Molecular and Cellular Biology 8, 1206-1215 (1988).

It has surprisingly been found that, when subjected to the above procedures, the p53 protein co-immunoprecipitates with a protein that has a molecular weight of 90kD, called p90. Suitable antibodies for the co-immunoprecipitation include PAb421 and Ab2. PAb421 recognizes the carboxy-terminus of p53 from various species, including human, mouse and rat p53, and is described by Harlow et al. in the Journal of Virology 39, 861-869 (1981). Ab2 is specific for the amino-terminus of human p53, and is available from Oncogene Science, Inc. of Manhassett, New York. The p90 protein does not immunoprecipitate when REF cells that do not express p53 are treated in the same way with the same antibodies.

Following centrifugation or elution from the column, p90 may be separated from the p53/p90 complex by means of SDS PAGE. The single band at 90kD is cut and sequenced. A partial sequence of p90 obtained from transformed rat embryo fibroblast cells is:

VAQMLLSQESDDYSQPST

Another method for purifying p90 is that generally described by Aebersold et al., Proc. Natl. Acad. Sci. USA 84, 6970-6974 (1987).

Mammals other than rats have proteins that are homologous to rat p90. A protein is homologous to p90 if it co-immunoprecipitates with p53 under the conditions described above and has substantially the same sequence as p90 - i.e., is at least about 30% identical, preferably at least about 50% identical, and more preferably at least

about 75% identical to the rat p90 gene.

The p90 protein or its homolog may also be prepared by well known recombinant DNA methods, such as those described by Sambrook, Fritsch and Maniatis (eds) in Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press (1989). Briefly, this method involves providing DNA that encodes the protein; amplifying or cloning the DNA in a suitable host; expressing the DNA in a suitable host; and harvesting the protein.

PROVIDING DNA

DNA encoding p90 is isolated by using the partial amino acid sequence provided above or a fragment thereof to prepare one or more oligonucleotide probes. The probe is labelled and used to screen a genomic or cDNA mammalian library in a suitable vector, such as in phage lambda. The homology between the DNA of the p53 from the species being screened and that of rat p53 is taken into account in determining the conditions for the hybridization.

The DNA isolated is sequenced, and the sequence used to prepare additional oligonucleotide probes. This procedure may be repeated to obtain overlapping fragments until a complete open reading frame is produced.

SCREENING GENOMIC DNA WITH OLIGONUCLEOTIDE PROBES

Methods for determining whether an oligonucleotide probe recognizes a specific nucleic acid molecule in a sample are known in the art. Preferably, the target nucleic acid molecule is immobilized. The presence of probe hybridized to the target nucleic acid molecule indicates the presence of the nucleic acid molecule in the sample.

Examples of some suitable screening methods are described by Dallas et al. in "The Characterization of an Escherichia Coli Plasmid Determinant that Encodes for the Production of a Heat-labile Enterotoxin." in K. N. Timmis and A. Puehler, eds, Plasmids of Medical, Environmental, and Commercial Importance, Elsevier/North-Holland Publishing Co., Amsterdam, pages 113-122 (1975); Grunstein and Hogness in Proc. Natl. Acad. Sci USA 72, 3961-3965 (1975); Palva et al. in U.S. Patent 4,731,325, which is assigned to Orion-yhtyma, Espoo, Finland; Mullis et al. in U.S. Patent 4,683,195, which is assigned to Cetus Corporation, Emeryville, California; Schneider et al in U.S. Patent 4,882,269, which is assigned to Princeton University, and Segev in PCT Application WO 90/01069. The Schneider et al patent and the Segev application are both licensed to ImClone Systems Inc., New York City.

AMPLIFYING DNA

The DNA obtained may be amplified by methods known in the art. One suitable method is the polymerase chain reaction (PCR) method described by Saiki et al. in Science 239, 487 (1988), by Mullis et al in U.S. Patent 4,683,195 and by Sambrook, Fritsch and Maniatis (eds) in Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press (1989). It is convenient to amplify the clones in lambda-gt10 or lambda-gt11 vectors using lambda-gt10 or lambda-gt11-specific oligomers as the Amplimers (available from Clontech, Palo Alto, California).

SEQUENCING DNA

Restriction fragments are cloned into a suitable vector, such as a plasmid or bacteriophage, and sequenced in accordance with methods known in the art. A suitable

sequencing method is the dideoxy chain terminating method described by Sanger et al in Proc. Natl. Acad. Sci. USA 74, 5463-5467 (1977). Suitable vectors and polymerases for sequencing are known. A suitable vector is the Bluescript vector of Stratagene. A suitable polymerase is Sequenase (United States Biochemical Corp., Cleveland, Ohio).

EXPRESSING DNA

The DNA encoding the protein of the invention may be replicated and used to express recombinant protein following insertion into a wide variety of host cells in a wide variety of cloning and expression vectors. The host may be prokaryotic or eukaryotic. The DNA may be obtained from natural sources and, optionally, modified. The genes may also be synthesized in whole or in part.

The vector into which the vector is spliced may comprise segments of chromosomal, non-chromosomal and synthetic DNA sequences. Some suitable prokaryotic cloning vectors include plasmids from E. coli, such as colE1, pCR1, pBR322, pMB9, pUC, pKSM, and RP4. Prokaryotic vectors also include derivatives of phage DNA such as fd, M13, and other filamentous single-stranded DNA phages.

Vectors for expressing proteins in bacteria, especially E.coli, are also known. Such vectors include the pKK233 plasmid family (or any of the tac family of plasmids), T7, and lambda P_L. Examples of vectors that express fusion proteins include the PATH vectors described by Dieckmann and Tzagoloff in J. Biol. Chem. 260, 1513-1520 (1985). The PATH vectors contain DNA sequences that encode anthranilate synthetase (TrpE) followed by a polylinker at the carboxy

terminus. Other expression vector systems are based on beta-galactosidase (pEX); maltose binding protein (pMAL); glutathione S-transferase (pGST) - see Gene 67, 31 (1988) and Peptide Research 3, 167 (1990).

Vectors useful in yeast are available. A suitable example is the 2u plasmid.

Suitable vectors for use in mammalian cells are also known. Such vectors include well-known derivatives of SV-40, adenovirus, retrovirus-derived DNA sequences and vectors derived from combination of plasmids and phage DNA.

Further eukaryotic expression vectors are known in the art (e.g., P.J. Southern and P. Berg, J. Mol. Appl. Genet. 1, 327-341 (1982); S. Subramani et al, Mol. Cell. Biol. 1, 854-864 (1981); R.J. Kaufmann and P.A. Sharp, "Amplification And Expression Of Sequences Cotransfected with A Modular Dihydrofolate Reductase Complementary DNA Gene," J. Mol. Biol. 159, 601-621 (1982); R.J. Kaufmann and P.A. Sharp, Mol. Cell. Biol. 159, 601-664 (1982); S.I. Scahill et al, "Expression And Characterization Of The Product Of A Human Immune Interferon DNA Gene In Chinese Hamster Ovary Cells," Proc. Natl. Acad. Sci. USA 80, 4654-4659 (1983); G. Urlaub and L.A. Chasin, Proc. Natl. Acad. Sci. USA 77, 4216-4220, (1980).

Useful expression hosts include well-known prokaryotic and eukaryotic cells. Some suitable prokaryotic hosts include, for example, E. coli, such as E. coli SG-936, E. coli HB 101, E. coli W3110, E. coli X1776, E. coli X2282, E. coli DHI, and E. coli MRCl, Pseudomonas, Bacillus, such as Bacillus subtilis, and Streptomyces. Suitable eukaryotic cells include yeasts and other fungi, insect, animal cells, such as COS cells and CHO cells, human cells and plant cells

in tissue culture.

The expression vectors useful in the present invention contain at least one expression control sequence that is operatively linked to the DNA sequence or fragment to be expressed. The control sequence is inserted in the vector in order to control and to regulate the expression of the cloned DNA sequence. Examples of useful expression control sequences are the lac system, the trp system, the tac system, the trc system, major operator and promoter regions of phage lambda, the control region of fd coat protein, the glycolytic promoters of yeast, e.g., the promoter for 3-phosphoglycerate kinase, the promoters of yeast acid phosphatase, e.g., Pho5, the promoters of the yeast alpha-mating factors, and promoters derived from polyoma, adenovirus, retrovirus, and simian virus, e.g., the early and late promoters or SV40, and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells and their viruses or combinations thereof.

FUSION PROTEINS

The proteins may be purified by methods known in the art. For example, the entire protein or portions thereof may be expressed in the form of a fusion protein with an appropriate fusion partner. The fusion partner preferably facilitates purification and identification. Some useful fusion partners include beta-galactosidase (Gray, et al., Proc. Natl. Acad. Sci. USA 79, 6598 (1982)); trpE (Itakura et al., Science 198, 1056 (1977)); protein A (Uhlen et al., Gene 23 369 (1983)); glutathione S-transferase (Johnson, Nature 338, 585 (1989)); Van Etten et al., Cell 58, 669 (1989)); and maltose binding protein (Guan et al., Gene 67, 21-30 (1987); Maina et al., Gene 74, 36-373 (1988); Riggs, P., in Ausebel, F.M. et al (eds) Current Protocols in

Molecular Biology, Greene Associates/Wiley Interscience, New York (1990)).

Such fusion proteins may be purified by affinity chromatography using reagents that bind to the fusion partner. The reagent may be a specific ligand of the fusion partner or an antibody, preferably a monoclonal antibody. For example, fusion proteins containing beta-galactosidase may be purified by affinity chromatography using an anti-beta-galactosidase antibody column (Ullman, Gene. 29, 27-31 (1984)). Similarly, fusion proteins containing maltose binding protein may be purified by affinity chromatography using a column containing cross-linked amylose; see Guan, European Patent Application 286,239.

Optionally, the DNA that encodes the fusion protein is engineered so that the fusion protein contains a cleavable site between the protein and the fusion partner. Both chemical and enzymatic cleavable sites are known in the art. Suitable examples of sites that are cleavable enzymatically include sites that are specifically recognized and cleaved by collagenase (Keil et al., FEBS Letters 56, 292-296 (1975)); enterokinase (Hopp et al., Biotechnology 6, 1204-1210 (1988)); factor Xa (Nagai et al., Methods Enzymol. 153, 461-481 (1987)); and thrombin (Eaton et al., Biochemistry 25, 505 (1986)). Collagenase cleaves between proline and X in the sequence Pro-X-Gly-Pro wherein X is a neutral amino acid. Enterkinase cleaves after lysine in the sequence Asp-Asp-Asp-Asp-Lys. Factor Xa cleaves after arginine in the sequence Ile-Glu-Gly-Arg. Thrombin cleaves between arginine and glycine in the sequence Arg-Gly-Ser-Pro.

Specific chemical cleavage agents are also known. For example, cyanogen bromide cleaves at methionine residues in proteins.

PURIFYING PROTEINS

The recombinant protein is purified by methods known in the art. Such methods include affinity chromatography using specific antibodies. Alternatively, the recombinant protein may be purified using a combination of ion-exchange, size-exclusion, and hydrophobic interaction chromatography using methods known in the art. These and other suitable methods are described by Marston, "The Purification of Eukaryotic Proteins Expressed in E. coli" in DNA Cloning, D. M. Glover, Ed., Volume III, IRL Press Ltd., England, 1987.

p90 may be used to prepare antibodies that are capable of co-immunoprecipitating p53. p53 may be isolated from the precipitate and purified. The anti-p90 antibodies may be polyclonal or monoclonal.

Polyclonal and monoclonal antibodies may be prepared by methods known in the art. See above and Campbell, "Monoclonal Antibody Technology, The Production and Characterization of Rodent and Human Hybridomas" in Burdon et al., Eds, Laboratoty Techniques in Biochemistry and Molecular Biology, Volume 13, Elsevier Science Publishers, Amsterdam (1985).

CLAIMS

What we claim is:

1. A panel of probes that detect and distinguish between sets of human p53 gene or protein mutations that frequently occur or are selected for in pre-cancer and cancer cells, each set giving rise to a phenotype that is different from that of wild-type p53 and of at least one other set of p53 mutants.

2. A panel of probes according to claim 1 wherein the mutations confer a proliferative advantage on cells containing such mutants.

3. A panel of probes according to claim 1 wherein the source of the p53 genes is colorectal carcinoma.

4. A panel of probes according to claim 1 wherein a first set of mutations comprises a mutation at position 143 or 175 of the p53 protein and a second set of mutations comprises a mutation at position 273 or 281 of the p53 protein or the corresponding positions of the p53 gene.

5. A panel of probes according to claim 1 wherein the sets of mutations comprise mutations between positions 117-142, 171-181, 234-258, or 270-286 of the p53 protein or the corresponding positions of the p53 gene.

6. A panel of probes according to claim 1 wherein the sets of mutations comprise mutations between positions 117-142, 171-181, 234-258, and 270-286 of the p53 protein or the corresponding positions of the p53 gene.

7. A panel of probes according to claim 1 wherein a first set of mutations comprises a mutation at position 430 or 526 of the p53 gene and a second set of mutations comprises a mutation at position 820 or 844 the p53 gene.

8. A panel of probes according to claim 1 wherein the sets of mutations comprise mutations between positions 352-427, 514-544, 703-775, or 811-859 of the p53 gene.

9. A panel of probes according to claim 1 wherein the sets of mutations comprise mutations between positions 352-427, 514-544, 703-775, and 811-859 of the p53 gene.

10. A panel of probes according to claim 1 wherein the probes are antibodies.

11. A panel of probes according to claim 1 wherein the probes are DNA probes.

12. A panel of probes according to claim 1 wherein the probes test positive for wild-type p53 and negative for mutant p53.

13. A probe according to claim 12 wherein the probe is a DNA probe.

14. A panel of probes according to claim 1 wherein the probes test positive for mutant p53 and negative for wild-type p53.

15. A panel of probes according to claim 14 wherein the probes are antibodies.

16. A panel of probes according to claim 15 wherein the antibodies are monoclonal.

17. A probe that detects a set of human p53 protein mutations in a pre-cancer or cancer cell wherein the protein does not bind to hsc70 or that detects a set of genetic mutations that produce such proteins.

18. A probe according to claim 17 wherein the set of mutations comprises a mutation at amino acid position 273 and/or 281 of human p53 protein or the corresponding nucleotide positions of the human p53 gene.

19. A probe according to claim 17 wherein the set of mutations comprises a mutation between amino acid positions 270-286 of human p53 protein or the corresponding nucleotide positions of the human p53 gene.

20. A probe according to claim 18 wherein the source of the p53 genes is colorectal carcinoma.

21. A probe according to claim 18 wherein the probe is an antibody.

22. A probe according to claim 18 wherein the probe is a DNA probe.

23. A probe according to claim 18 wherein the probe tests positive for wild-type p53 and negative for mutant p53.

24. A probe according to claim 23 wherein the probe is a DNA probe.

25. A probe according to claim 18 wherein the probe tests positive for mutant p53 and negative for wild-type p53.

26. A probe according to claim 25 wherein the probe is an antibody.

27. A probe according to claim 26 wherein the antibody is monoclonal.

28. A method of distinguishing between sets of human p53 gene or protein mutations that are frequently occurring or selected for in pre-cancer or cancer cells, each set giving rise to a phenotype that is different from the wild-type gene or protein and from at least one other set of mutations, the method comprising determining the mutations in a sample of p53 genes or proteins with a panel of probes that detect and distinguish between such sets.

29. A method of distinguishing between sets of human p53 protein mutations that are frequently occuring or selected for in pre-cancer or cancer cells, the method comprising the steps of determining whether such a mutation exists and, if so, whether the mutant protein binds to hsc70 less tightly than wild-type human p53 protein.

30. A protein having the following properties:

a. is capable of co-immunoprecipitating with p53 in the presence of p53 and p53-specific antibodies from rat embryo fibroblast cells that are transfected with, and express, a wild type or mutated p53 gene; and

b. is not capable of immunoprecipitating with p53-specific antibodies from rat embryo fibroblast cells that do not expresses a wild type or mutated p53 gene; and

c. is 30% - 100% homologous with rat p90.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/04608

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5): C07H 15/12, 17/00; C07K 15/00, 15/28
U.S. CL.: 530/387; 536/27,28,29

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
U.S. CL.	530/387,418 536/27,28,29

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
Y	US,A, 4,798,787 (McCormick et al.) 17 January 1989, see abstract and column 7, line 66- column 8, line 12.	1-27
Y	US,A, 4,871,838 (BOS et al.) 03 October 1989, see abstract.	1-27
Y	Science, Volume 244, issued 14 April 1989, Baker et al., "Chromosome 17 Deletions and p53 Gene Mutations in Colorectal Carcinomas," pages 217-221, see abstract.	1-27

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
06 SEPTEMBER 1991	02 OCT 1991
International Searching Authority	**Signature of Authorized Officer**
ISA/US	LORT YUAN

Form PCT/ISA/210 (second sheet) (Rev.11-87)

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

Y	US,A, 4,794,075 (FORD ET AL.) 27 December 1988, see column 5, line 21-column 6, line 2.	28,29

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers . , because they relate to subject matter [12] not required to be searched by this Authority, namely:

2. ☐ Claim numbers . . . . . , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [13], specifically:

3. ☐ Claim numbers_____, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☑ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

I. Claims 1-29 drawn to a first product of probes, classified in Class 536, subclass 27 and Class 530, subclass 387; and a first method of using first product, classified in Class 435, subclass 6 and Class 435, subclass 7.1.

II. Claim 30 drawn to a second product of a protein, classified in Class 530, subclass 350.

1. ☑ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number:
A61K 49/00, C12Q 1/68 C12N 15/00, 5/00	A1	WO 92/11874
		(43) International Publication Date: 23 July 1992 (23.07.92)

(21) International Application Number: PCT/US92/00295

(22) International Filing Date: 3 January 1992 (03.01.92)

(30) Priority data:
637,563 4 January 1991 (04.01.91) US

(71) Applicant: BAYLOR COLLEGE OF MEDICINE [US/US]; One Baylor Plaza, Houston, TX 77030 (US).

(72) Inventors: DONEHOWER, Lawrence, A. ; 4307 McDermed Drive, Houston, TX 11096 (US). BRADLEY, Allan ; 5619 Wigton, Houston, TX 77096 (US). BUTEL, Janet, S. ; 4051 Mischine, Houston, TX 77025 (US). SLAGLE, Betty ; 1605 Turnpike, Houston, TX 77008 (US).

(74) Agents: AUERBACH, Jeffrey, I. et al.; Weil, Gotshal & Manges, 1615 "L" Street, N.W., Washington, DC 20036 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), MC (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: TUMOR SUSCEPTIBLE NON-HUMAN ANIMALS

(57) Abstract

A desired non-human animal or an animal cell or human cell which contains a predefined, specific and desired alteration in at least one of its two p53 chromosomal alleles, such that at least one of these alleles contains a mutation which alters the expression of the allele, and the other of the alleles expresses either a normal p53 gene product, or comprises an identical or different p53 mutation.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**ES**	Spain	**MG**	Madagascar
AU	Australia	**FI**	Finland	**ML**	Mali
BB	Barbados	**FR**	France	**MN**	Mongolia
BE	Belgium	**GA**	Gabon	**MR**	Mauritania
BF	Burkina Faso	**GB**	United Kingdom	**MW**	Malawi
BG	Bulgaria	**GN**	Guinea	**NL**	Netherlands
BJ	Benin	**GR**	Greece	**NO**	Norway
BR	Brazil	**HU**	Hungary	**PL**	Poland
CA	Canada	**IT**	Italy	**RO**	Romania
CF	Central African Republic	**JP**	Japan	**RU**	Russian Federation
CG	Congo	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CH	Switzerland			**SE**	Sweden
CI	Côte d'Ivoire	**KR**	Republic of Korea	**SN**	Senegal
CM	Cameroon	**LI**	Liechtenstein	**SU**	Soviet Union
CS	Czechoslovakia	**LK**	Sri Lanka	**TD**	Chad
DE	Germany	**LU**	Luxembourg	**TG**	Togo
DK	Denmark	**MC**	Monaco	**US**	United States of America

TITLE OF THE INVENTION:

TUMOR SUSCEPTIBLE NON-HUMAN ANIMALS

FIELD OF THE INVENTION:

The invention is directed toward tumor-susceptible non-human animals. The invention further pertains to the use of such animals in the development of anti-cancer agents and therapies.

CROSS-REFERENCE TO RELATED APPLICATIONS:

This application is a continuation-in-part of U.S. Patent Application Serial No. 07/637,563 (filed January 4, 1991).

BACKGROUND OF THE INVENTION:

I. Chimeric and Transgenic Animals

Recent advances in recombinant DNA and genetic technologies have made it possible to introduce and express a desired gene sequence in a recipient animal. Through the use of such methods, animals have been engineered to contain gene sequences that are not normally or naturally present in an unaltered animal. The techniques have also been used to produce animals which exhibit altered expression of naturally present gene sequences.

GeneDX 1009, pg. 690

The animals produced through the use of these methods are known as either "chimeric" or "transgenic" animals. In a "chimeric" animal, only some of the animal's cells contain and express the introduced gene sequence, whereas other cells have been unaltered. The capacity of a chimeric animal to transmit the introduced gene sequence to its progeny depends upon whether the introduced gene sequences are present in the germ cells of the animal. Thus, only certain chimeric animals can pass along the desired gene sequence to their progeny.

In contrast, all of the cells of a "transgenic" animal contain the introduced gene sequence. Consequently, every transgenic animal is capable of transmitting the introduced gene sequence to its progeny.

II. Production of Transgenic Animals: Microinjection Methods

The most widely used method through which transgenic animals have been produced involves injecting a DNA molecule into the male pronucleus of a fertilized egg (Brinster, R.L. et al., Cell 27:223 (1981); Costantini, F. et al., Nature 294:92 (1981); Harbers, K. et al., Nature 293:540 (1981); Wagner, E.F. et al., Proc. Natl. Acad. Sci. (U.S.A.) 78:5016 (1981); Gordon, J.W. et al., Proc. Natl. Acad. Sci. (U.S.A.) 73:1260 (1976); Stewart, T.A. et al., Science 217:1046-1048 (1982); Palmiter, R.D. et al., Science 222:809 (1983); Evans, R.M et al. (U.S. Patent 4,870,009)).

The gene sequence being introduced need not be incorporated into any kind of self-replicating plasmid or virus

GeneDX 1009, pg. 691

(Jaenisch, R., Science, 240:1468-1474 (1988)). Indeed, the presence of vector DNA has been found, in many cases, to be undesirable (Hammer, R.E. et al., Science 235:53 (1987); Chada, K. et al., Nature 319:685 (1986); Kollias, G. et al., Cell 46:89 (1986); Shani, M., Molec. Cell. Biol. 6:2624 (1986); Chada, K. et al., Nature 314:377 (1985); Townes, T. et al., EMBO J. 4:1715 (1985)).

After being injected into the recipient fertilized egg, the DNA molecules are believed to recombine with one another to form extended head-to-tail concatemers. It has been proposed that such concatemers occur at sites of double-stranded DNA breaks at random sites in the egg's chromosomes, and that the concatemers are inserted and integrated into such sites (Brinster, R.L. et al., Proc. Natl. Acad. Sci. (U.S.A.) 82:4438 (1985)). Although it is, thus, possible for the injected DNA molecules to be incorporated at several sites within the chromosomes of the fertilized egg, in most instances, only a single site of insertion is observed (Jaenisch, R., Science, 240:1468-1474 (1988)).

Once the DNA molecule has been injected into the fertilized egg cell, the cell is implanted into the uterus of a recipient female, and allowed to develop into an animal. Since all of the animal's cells are derived from the implanted fertilized egg, all of the cells of the resulting animal (including the germ line cells) shall contain the introduced gene sequence. If, as occurs in about 30% of events, the first cellular division occurs before the introduced gene sequence has integrated into the

GeneDX 1009, pg. 692

cell's genome, the resulting animal will be a chimeric animal.

By breeding and inbreeding such animals, it has been possible to produce heterozygous and homozygous transgenic animals. Despite any unpredictability in the formation of such transgenic animals, the animals have generally been found to be stable, and to be capable of producing offspring which retain and express the introduced gene sequence.

Since microinjection causes the injected DNA to be incorporated into the genome of the fertilized egg through a process involving the disruption and alteration of the nucleotide sequence in the chromosome of the egg at the insertion site, it has been observed to result in the alteration, disruption, or loss of function of the endogenous egg gene in which the injected DNA is inserted. Moreover, substantial alterations (deletions, duplications, rearrangements, and translocations) of the endogenous egg sequences flanking the inserted DNA have been observed (Mahon, K.A. et al., Proc. Natl. Acad. Sci. (U.S.A.) 85:1165 (1988); Covarrubias, Y. et al., Proc. Natl. Acad. Sci. (U.S.A.) 83:6020 (1986); Mark, W. et al., Cold Spr. Harb. Symp. Quant. Biol. 50:453 (1985)). Indeed, lethal mutations or gross morphological abnormalities have been observed (Jaenisch, R., Science 240:1468-1474 (1988); First, N.L. et al., Amer. Meat Sci. Assn. 39th Reciprocal Meat Conf. 39:41 (1986))).

Significantly, it has been observed that even if the desired gene sequence of the microinjected DNA molecule is one that is naturally found in the recipient egg's genome, integration of the desired gene sequence rarely occurs at

GeneDX 1009, pg. 693

the site of the natural gene (Brinster, R.L. et al., Proc. Natl. Acad. Sci. (U.S.A.) 86:7087-7091 (1989)). Moreover, introduction of the desired gene sequence does not generally alter the sequence of the originally present egg gene.

Although the site in the fertilized egg's genome into which the injected DNA ultimately integrates cannot be predetermined, it is possible to control the expression of the desired gene sequence such that, in the animal, expression of the sequence will occur in an organ or tissue specific manner (reviewed by Westphal, H., FASEB J. 3:117 (1989); Jaenisch, R., Science 240:1468-1474 (1988); Meade, H. et al. (U.S. Patent 4,873,316)).

The success rate for producing transgenic animals is greatest in mice. Approximately 25% of fertilized mouse eggs into which DNA has been injected, and which have been implanted in a female, will become transgenic mice. A lower rate has been thus far achieved with rabbits, sheep, cattle, and pigs (Jaenisch, R., Science 240:1468-1474 (1988); Hammer, R.E. et al., J. Animal. Sci. 63:269 (1986); Hammer, R.E. et al., Nature 315:680 (1985); Wagner, T.E. et al., Theriogenology 21:29 (1984)). The lower rate may reflect greater familiarity with the mouse as a genetic system, or may reflect the difficulty of visualizing the male pronucleus of the fertilized eggs of many farm animals (Wagner, T.E. et al., Theriogenology 21:29 (1984)).

Thus, the production of transgenic animals by microinjection of DNA suffers from at least two major drawbacks. First, it can be accomplished only during the single-cell stage of an animal's life. Second, it requires the disruption of the natural sequence of the DNA, and thus

GeneDX 1009, pg. 694

is often mutagenic or teratogenic (Gridley, T. et al., Trends Genet. 3:162 (1987)).

III. Production of Chimeric and Transgenic Animals: Recombinant Viral and Retroviral Methods

Chimeric and transgenic animals may also be produced using recombinant viral or retroviral techniques in which the gene sequence is introduced into an animal at a multi-cell stage. In such methods, the desired gene sequence is introduced into a virus or retrovirus. Cells which are infected with the virus acquire the introduced gene sequence. If the virus or retrovirus infects every cell of the animal, then the method results in the production of a transgenic animal. If, however, the virus infects only some of the animal's cells, then a chimeric animal is produced.

The general advantage of viral or retroviral methods of producing transgenic animals over those methods which involve the microinjection of non-replicating DNA, is that it is not necessary to perform the genetic manipulations at a single cell stage. Moreover, infection is a highly efficient means for introducing the DNA into a desired cell.

Recombinant retroviral methods for producing chimeric or transgenic animals have the advantage that retroviruses integrate into a host's genome in a precise manner, resulting generally in the presence of only a single integrated retrovirus (although multiple insertions may occur). Rearrangements of the host chromosome at the site of integration are, in general, limited to minor duplications (4-6 base pairs) of host DNA at the host virus

GeneDX 1009, pg. 695

junctions (Jaenisch, R., Science 240:1468-1474 (1988); see also, Varmus, H., In: RNA Tumor Viruses (Weiss, R. et al., Eds.), Cold Spring Harbor Press, Cold Spring Harbor, NY, pp. 369-512 (1982)). The method is, however, as mutagenic as microinjection methods.

Chimeric animals have, for example, been produced by incorporating a desired gene sequence into a virus (such as bovine papilloma virus or polyoma) which is capable of infecting the cells of a host animal. Upon infection, the virus can be maintained in an infected cell as an extrachromosomal episome (Elbrecht, A. et al., Molec. Cell. Biol. 7:1276 (1987); Lacey, M. et al., Nature 322:609 (1986); Leopold, P. et al., Cell 51:885 (1987)). Although this method decreases the mutagenic nature of chimeric/transgenic animal formation, it does so by decreasing germ line stability, and increasing oncogenicity.

Pluripotent embryonic stem cells (referred to as "ES" cells) are cells which may be obtained from embryos until the early post-implantation stage of embryogenesis. The cells may be propagated in culture, and are able to differentiate either in vitro or in vivo upon implantation into a mouse as a tumor. ES cells have a normal karyotype (Evans, M.J. et al., Nature 292:154-156 (1981); Martin, G.R. et al., Proc. Natl. Acad. Sci. (U.S.A.) 78:7634-7638 (1981)).

Upon injection into a blastocyst of a developing embryo, ES cells will proliferate and differentiate, thus resulting in the production of a chimeric animal. ES cells are capable of colonizing both the somatic and germ-line lineages of such a chimeric animal (Robertson, E. et al.,

Cold Spring Harb. Conf. Cell Prolif. 10:647-663 (1983); Bradley A. et al., Nature 309:255-256 (1984); Bradley, A. et al., Curr. Top. Devel. Biol. 20:357-371 (1986); Wagner, E.F. et al., Cold Spring Harb. Symp. Quant. Biol. 50:691-700 (1985); (all of which references are incorporated herein by reference).

In this method, ES cells are cultured in vitro, and infected with a viral or retroviral vector containing the gene sequence of interest. Chimeric animals generated with retroviral vectors have been found to have germ cells which either lack the introduced gene sequence, or contain the introduced sequence but lack the capacity to produce progeny cells capable of expressing the introduced sequence (Evans, M.J. et al., Cold Spring Harb. Symp. Quant. Biol. 50:685-689 (1985); Stewart, C.L. et al., EMBO J. 4:3701-3709 (1985); Robertson, L. et al., Nature (1986); which references are incorporated herein by reference).

Because ES cells may be propagated in vitro, it is possible to manipulate such cells using the techniques of somatic cell genetics. Thus, it is possible to select ES cells which carry mutations (such as in the hprt gene (encoding hypoxanthine phosphoribosyl transferase) (Hooper, M. et al., Nature 326:292-295 (1987); Kuehn, M.R. et al., Nature 326:295-298 (1987)). Such selected cells can then be used to produce chimeric or transgenic mice which fail to express an active HPRT enzyme, and thus provide animal models for diseases (such as the Lesch-Nyhan syndrome which is characterized by an HPRT deficiency) (Doetschman, T. et al., Proc. Natl. Acad. Sci. (U.S.A.) 85:8583-8587 (1988)).

GeneDX 1009, pg. 697

As indicated above, it is possible to generate a transgenic animal from a chimeric animal (whose germ line cells contain the introduced gene sequence) by inbreeding.

The above-described methods permit one to screen for the desired genetic alteration prior to introducing the transfected ES cells into the blastocyst. One drawback of these methods, however, is the inability to control the site or nature of the integration of the vector.

IV. Production of Chimeric and Transgenic Animals: Plasmid Methods

The inherent drawbacks of the above-described methods for producing chimeric and transgenic animals have caused researchers to attempt to identify additional methods through which such animals could be produced.

Gossler, A. et al., for example, have described the use of a plasmid vector which had been modified to contain the gene for neomycin phosphotransferase (nptII gene) to transfect ES cells in culture. The presence of the nptII gene conferred resistance to the antibiotic G418 to ES cells that had been infected by the plasmid (Gossler, A. et al., Proc. Natl. Acad. Sci. (U.S.A.) 83:9065-9069 (1986), which reference is incorporated herein by reference). The chimeric animals which received the plasmid and which became resistant to G418, were found to have integrated the vector into their chromosomes.

Takahashi, Y. et al. have described the use of a plasmid to produce chimeric mice cells which expressed an avian crystallin gene (Development 102:258-269 (1988), incor-

GeneDX 1009, pg. 698

porated herein by reference). The avian gene was incorporated into a plasmid which contained the nptII gene. Resulting chimeric animals were found to express the avian gene.

V. Introduction of Gene Sequences into Somatic Cells

DNA has been introduced into somatic cells to produce variant (chimeric) cell lines. hprt-deficient Chinese hamster ovary (CHO) cells have been transformed with the CHO hprt gene in order to produce a prototrophic cell line (Graf, L.H. et al., Somat. Cell Genet. 5:1031-1044 (1979)). Folger et al. examined the fate of a thymidine kinase gene (tk gene) which had been microinjected into the nuclei of cultured mammalian cells. Recipient cells were found to contain from 1 to 100 copies of the introduced gene sequence integrated as concatemers at one or a few sites in the cellular genome (Folger, K.R. et al., Molec. Cell. Biol. 2:1372-1387 (1982)). DNA-mediated transformation of an RNA polymerase II gene into Syrian hamster cells has also been reported (Ingles, C. et al., Molec. Cell. Biol. 2:666-673 (1982)).

Plasmids conferring host neomycin resistance and guanosine phosphotransferase activity have been transfected into Chinese hamster ovary cells to generate novel cell lines (Robson, C.N. et al., Mutat. Res. 163:201-208 (1986)).

VI. Oncogenes and tumor suppressor genes

One mechanism through which cancer may arise is through a cell's exposure to a carcinogenic agent, either chemical or radiation. Such exposure may damage the DNA sequence of critical genes present in the genome of a cell of an animal. If this damage leads to either an impairment in the expression of the gene, or in the production of a mutant gene product, the cell may then proceed to proliferate, and ultimately result in the formation of a tumor.

One class of such critical genes has been referred to as "oncogenes." Oncogenes are genes which are naturally in an "inactivated" state, but which, through the effect of the DNA damage are converted to an "activated" state capable of inducing tumorigenesis (i.e. tumor formation). Oncogenes have been identified in 15-20% of human tumors. The products of oncogenes ("oncoproteins") can be divided into two broad classes according to their location in the cell.

Oncogene products which act in the cytoplasm of cells have readily identifiable biochemical or biological activities (Green, M.R., Cell 56:1-3 (1989)). Those that act in the nucleus of a cell have been more difficult to characterize. Some nuclear oncoproteins (such as E1A and myc) have transcriptional regulatory activity, and are believed to mediate their activities by the transcriptional activation of cellular genes (Kingston, R.E., Cell 41:3-5 (1985)). Other nuclear oncoproteins appear to have a complex array of activities (such as DNA binding activity, ability to initiate viral DNA synthesis, ATPase activity,

GeneDX 1009, pg. 700

helicase activity, and transcriptional regulatory activity) (Green, M.R., Cell 56:1-3 (1989)).

The creation of a mutant oncogene is only one of the requirements needed for tumor formation; tumorigenesis appears to also require the additional inactivation of a second class of critical genes: the "anti-oncogenes" or "tumor-suppressing genes." In their natural state these genes act to suppress cell proliferation. Damage to such genes leads to a loss of this suppression, and thereby results in tumorigenesis. Thus, the deregulation of cell growth may be mediated by either the activation of oncogenes or the inactivation of tumor-suppressing genes (Weinberg, R.A., Scientific Amer., Sept. 1988, pp 44-51).

Oncogenes and tumor-suppressing genes have a basic distinguishing feature. The oncogenes identified thus far have arisen only in somatic cells, and thus have been incapable of transmitting their effects to the germ line of the host animal. In contrast, mutations in tumor-suppressing genes can be identified in germ line cells, and are thus transmissible to an animal's progeny.

The classic example of a hereditary cancer is retinoblastomas in children. The incidence of retinoblastomas is determined by a tumor suppressor gene, the retinoblastoma (RB) gene (Weinberg, R.A., Scientific Amer., Sept. 1988, pp 44-51); Hansen M.F. et al., Trends Genet. 4:125-128 (1988)). Individuals born with a lesion in one of the RB alleles are predisposed to early childhood development of retinoblastomas (Weinberg, R.A., Scientific Amer., Sept. 1988, pp 44-51); Hansen M.F. et al., Trends Genet. 4:125-128 (1988)). Inactivation or mutation of (the

second RB allele in one of the somatic cells of these susceptible individuals appears to be the molecular event that leads to tumor formation (Cavenee, W.K. et al., Nature 305:779-784 (1983); Friend, S.H. et al., Proc. Nat'l. Acad. Sci. (U.S.A.) 84:9059-9063 (1987)).

The RB tumor-suppressing gene has been localized onto human chromosome 13. The mutation may be readily transmitted through the germ line of afflicted individuals (Cavenee, W.K. et al., New Engl. J. Med. 314:1201-1207 (1986)). Individuals who have mutations in only one of the two naturally present alleles of this tumor-suppressing gene are predisposed to retinoblastoma. Inactivation of the second of the two alleles is, however, required for tumorigenesis (Knudson, A.G., Proc. Nat'l. Acad. Sci. (U.S.A.) 68:820-823 (1971)).

A second tumor-suppressing gene is the p53 gene (Green, M.R., Cell 56:1-3 (1989); Mowat et al., Nature 314:633-636 (1985)). The protein encoded by the p53 gene is a nuclear protein that forms a stable complex with both the SV40 large T antigen and the adenovirus E1B 55 kd protein. The p53 gene product may be inactivated by binding to these proteins.

Initially, the p53 gene was thought to be an oncogene rather than a tumor-suppressing gene since it is capable of immortalizing primary rodent cells and can cooperate with the ras oncogene to cause transformation. Subsequent research revealed that the p53 genes used in those early experiments was a mutant allele of the normal p53 gene (Green, M.R., Cell 56:1-3 (1989)). Thus, the p53 gene is a tumor-suppressing gene rather than an oncogene.

GeneDX 1009, pg. 702

Mutations at any of a large number of positions in the p53 gene can result in the activation of the transforming potential of the p53 gene product (Eliyahu et al., Nature 312:646-649 (1984); Finlay et al., Molec. Cell. Biol. 8:531-539 (1988)). This has suggested that the activation of the p53 transforming activity is due to the inactivation of the normal p53 activity (Green, M.R., Cell 56:1-3 (1989)).

The p53 gene has been implicated as having a role in colorectal carcinoma (Baker, S.J. et al., Science 244:217-221 (1989)). Studies had shown that allelic deletions of the short arm of chromosome 17 occurred in over 75% of colorectal carcinomas. The region deleted was subsequently found to encompass the p53 gene locus (Baker, S.J. et al., Science 244:217-221 (1989)). The deletion of the region was found to mark a transition from a (benign) adenocarcinoma stage to a (malignant) carcinomatous stage (Vogelstein, B. et al., New Engl. J. Med. 319:525 (1988)).

Similar deletions in chromosome 17 have been identified in a wide variety of cancers including breast and lung cancers (Mackay, J. et al., Lancet ii:1384 (1988); James, C.D. et al., Canc. Res. 48:5546 (1988); Yakota, J. et al., Proc. Nat'l. Acad. Sci. (U.S.A.) 84:9252 (1987); Toguchida et al., Canc. Res. 48:3939 (1988)). In addition to p53 allele loss, Nigro et al. (Nature 342:705-708 (1989)) have demonstrated that the single remaining p53 allele in a variety of human tumors (brain, colon, breast, lung) undergo a point mutation which renders it tumorigenic. Fearon et al. (Cell 61:759-767 (1990)) have hypothesized that both point mutations and deletions in the p53 alleles may be required for a fully tumorigenic phenotype. These findings

GeneDX 1009, pg. 703

suggest that the p53 gene may have a role in many types of cancers.

VII. Conclusions

The application of the above-described technologies has the potential to produce animals which cannot be produced through classical genetics. For example, animals can be produced which suffer from human diseases (such as AIDS, diabetes, cancer, etc.), and may be valuable in elucidating therapies for such diseases. Chimeric and transgenic animals have substantial use as probes of natural gene expression.

Leder, P. et al. (U.S. Patent 4,736,866) disclose the production of transgenic non-human mammals which contain cells having an exogenously added activated oncogene sequence. Although the animals are disclosed as being useful for assaying for carcinogenic materials, the precise location and structure of the added oncogene sequence in the animals is unknown, and cannot be experimentally controlled. Thus, the value of the animals as a model for oncogenesis is significantly impaired.

Despite the successes of the above-described techniques, the methods have not yet led to the development of a model transgenic animal which can be used to study the conditions responsible for the initiation of neoplasia, and which can be used as a means for developing suitable antineoplastic agents and therapies. Indeed, prior to the present invention, research on transgenic or chimeric animals containing mutations in oncogene and "critical" genes had

GeneDX 1009, pg. 704

suggested that it would not be possible to produce viable animals containing mutations in the chromosomal alleles of their tumor-suppressing genes. It was believed that such animals would be non-viable, or would not survive to maturity. See, Soriano, P. et al., Cell 64:693-702 (1991) [c-src]; McMahon, A. and Bradley, B., Cell 62:1073-1085 (1990) [Wnt-1]; Koller, B. et al., Science 248:1227-1230 (1990) [MHC-I]; Tybulewicz, V.L.J., Cell 65:1153-1164 (1991) [c-abl]; Mucenski, M.L., Cell 65:677-690 (1991) [c-myb].

If however, animals predisposed to cancer could be obtained, they would facilitate a better understanding of cancer; they could be used to assay for the presence of cancer-causing agents in food, waste products, etc.; they could also be used to identify agents capable of suppressing or preventing cancer. Such animals would, therefore, be extremely desirable. The present invention provides such animals, and the methods to produce and use them.

BRIEF DESCRIPTION OF THE FIGURES:

Figure 1 shows the modifications made to a 3.7- kb fragment spanning exons 2 to 10 of the 11 exon p53 gene in order to facilitate the construction of the transgenic animals of the present invention.

Figure 2 shows the structure of the targeting construct used to successfully inactivate the p53 gene in mouse embryonic stem cells. The arrows labelled 5' and 3' represent the oligonucleotide primers used to screen the resistant stem cell colonies by PCR. H represents the hybridization probe for the PCR fragment.

GeneDX 1009, pg. 705

Figure 3 shows a comparison Southern blot of the p53 alleles present in two of the constructed embryonic stem cells (designated 1 and 2) and in wild type (wt) cells.

Figure 4 shows a Southern blot analysis of tail DNA of the progeny of the F_1[129xC57BL/6] chimeric mice. The analysis reveals that mouse pups p1 and p3 contain DNA which hybridizes to the neo gene, and thus contains the mutated p53 construct. "+" indicates positive control (i.e. embryonic stem cell with disrupted p53 gene); "-" indicates negative control (i.e. normal mouse DNA).

Figure 5 shows the procedures which are used to produce any desired mutation into a p53 allele of a cell.

Figure 6A shows the location and orientation of the primers used in the PCR amplification of the p53 mRNA. Figure 6B shows the results of the PCR amplification. The gel shows 19 lanes, of which the last (M) contains molecular size markers. The initial 18 lanes depict the amplification obtained in normal mice (++), p53 heterozygote mice (+-), and p53 homozygote mice (--), respectively, with each of six different pairs of PCR primers (1 and 4a; 7 and 10; B-actin control; 4b and 10; 1 and 10; and 4b and 6). The appearance of a band indicates that the primer pairs amplified a segment of an mRNA species present in the preparation.

SUMMARY OF THE INVENTION:

The present invention provides a desired non-human animal or an animal (including human) cell which contains a predefined, specific and desired alteration rendering the non-human animal or animal cell predisposed to cancer.

GeneDX 1009, pg. 706

Specifically, the invention pertains to a genetically altered non-human animal (most preferably, a mouse), or a cell (either non-human animal or human) in culture, that is defective in at least one of two alleles of a tumor-suppressor gene such as the p53 gene, the rb gene, etc., and most preferably, the p53 gene. The inactivation of at least one of these tumor suppressor alleles results in an animal with a higher susceptibility to tumor induction. A genetically altered mouse of this type is able to serve as a useful model for hereditary cancers and as a test animal for carcinogen studies. The invention additionally pertains to the use of such non-human animals or animal cells, and their progeny in research and medicine.

In detail, the invention provides a transgenic or chimeric animal cell whose genome comprises two chromosomal alleles of a tumor-suppressing gene (especially, the p53 gene), wherein at least one of the two alleles contains a mutation, or the progeny of this cell.

The invention includes the embodiment of the above animal cell, wherein one of the alleles expresses a normal tumor-suppressing gene product.

The invention includes the embodiments wherein the above animal cells are human cells, or the cells of a non-human animal. The invention includes the embodiment wherein the cell is an embryonic stem cell, and in particular, wherein the tumor-suppressing gene is a p53 gene, and the embryonic stem cell is ATCC CRL 10631.

The invention also includes a non-human transgenic or chimeric animal having an animal cell whose genome comprises two chromosomal alleles of a tumor-suppressing gene, wherein

GeneDX 1009, pg. 707

at least one of the two alleles contains a mutation, or the progeny of the animal, or an ancestor of the animal, at an embryonic stage (preferably the one-cell, or fertilized oocyte stage, and generally, not later than about the 8-cell stage).

The invention also includes the embodiment wherein the tumor suppressing gene of the non-human animal is a p53 gene.

The invention is also directed to the embodiments wherein the animal cell of the non-human animal is a human cell, or a cell of a non-human animal (of either the same species as the non-human animal or a different species).

The invention is also directed to the embodiments wherein the animal cell of the non-human animal is a germ-line cell, or a somatic cell.

The invention is also directed to the embodiment wherein the animal cell of the non-human animal is an embryonic stem cell (especially the embryonic stem cell, ATCC CRL 10631).

The invention additionally provides a method for identifying the presence of an agent suspected of being capable of affecting a characteristic of an animal cell that is attributable to the presence or expression of a tumor-suppressing gene, the method comprising:

A) administering an amount of the agent to an animal cell in cell culture, the cell having a genome that comprises two chromosomal alleles of the tumor-suppressing gene, wherein at least one of the two alleles contains a mutation;

B) maintaining the cell culture for a desired period of time after the administration;

GeneDX 1009, pg. 708

C) determining whether the administration of the agent has affected a characteristic of the animal cell that is attributable to the presence or expression of the alleles of the tumor-suppressing gene.

In particular, the invention includes the embodiment of the above method wherein the tumor-suppressing gene is a p53 gene.

The invention also includes the embodiments of this method wherein the agent is suspected of being able to increase a tumorigenic potential of the animal cell, wherein the agent is suspected of being able to decrease a tumorigenic potential of the animal cell, wherein the animal cell is a human cell, and wherein the animal cell is a non-human animal cell (such as an embryonic stem cell, and in particular, the embryonic stem cell, ATCC CRL 10631).

The invention also provides a method for identifying the presence of an agent suspected of being capable of affecting a characteristic of an animal cell that is attributable to the presence or expression of a tumor-suppressing gene, the method comprising:

A) administering an amount of the agent to an animal, the animal having a cell whose genome comprises two chromosomal alleles of the tumor-suppressing gene, wherein at least one of the two alleles contains a mutation;

B) maintaining the animal for a desired period of time after the administration;

C) determining whether the administration of the agent has affected a characteristic of the cell that is attributable to the presence or expression of the alleles of the tumor-suppressing gene.

GeneDX 1009, pg. 709

In particular, the invention includes the embodiments of the above method wherein the tumor-suppressing gene is a p53 gene.

The invention also includes the embodiments of this method wherein the agent is suspected of being able to increase a tumorigenic potential of the animal cell, wherein the agent is suspected of being able to decrease a tumorigenic potential of the animal cell, wherein the animal cell is a human cell, and wherein the animal cell is a non-human animal cell (such as an embryonic stem cell, and in particular, the embryonic stem cell, ATCC CRL 10631).

The invention also includes the embodiments of this method wherein the animal and the animal cell are of the same species, of different species.

The invention further provides a method of gene therapy comprising altering the genome of a cell of either a human or a non-human animal, wherein the cell has a genome that comprises two chromosomal alleles of a tumor-suppressing gene, and wherein at least one of the two alleles contains a mutation, to thereby form a cell wherein at least one of the alleles expresses a normal tumor-suppressing gene product.

In particular, the invention includes the embodiments of the above method of gene therapy wherein the tumor-suppressing gene is a p53 gene.

DESCRIPTION OF THE PREFERRED EMBODIMENTS:

As is well known, the cells of humans and animals (especially, rodents (i.e. mouse, rat, hamster, etc.),

GeneDX 1009, pg. 710

rabbits, sheep, goats, fish, pigs, cattle and non-human primates) are "diploid" cells, and thus naturally contain two copies ("alleles") of each and every gene of their genome. A cell's "genome" consists of all of its heritable DNA (either chromosomal or non-chromosomal (i.e. episomal, viral, etc.). One of the two alleles of a gene is provided by the animal's or cell's maternal parent; the other set is provided by its paternal parent. The diploid nature of human and animal cells is described by DeRobertis, E.D.P., et al. (Cell Biology, 6th Ed., W.B. Saunders Company, Philadelphia, (1975)), and in other similar treatises of cell biology.

Cancer in humans develops through a multi-step process, indicating that multiple changes must occur to convert a normal cell into one with a malignant phenotype. One class of involved genes includes cellular oncogenes. When activated by mutation or when expressed inappropriately, dominant-acting oncogenes override normal cellular control mechanisms and promote unbridled cell proliferation. A newly recognized class of cellular genes that appears to be equally important in cancer development includes the tumor suppressor genes, sometimes called "anti-oncogenes." These genes act to dampen cell growth; inactivation of their normal function appears to be a common denominator in the evolution of tumor cells. Both alleles of a tumor suppressor gene must be inactivated to result in loss of function in the cell. Inactivation of one allele (i.e., the gene copy on one of the two chromosomes) increases the probability that an event will damage the surviving allele,

GeneDX 1009, pg. 711

in effect making the host more susceptible to tumor induction.

The present invention relates to the production of non-human transgenic and chimeric animals and cells which contain at least one mutated chromosomal allele of a tumor suppressor gene (and, in a preferred embodiment, one normal allele of that gene). Where the cells and non-human animals of the present invention contain mutations in both of their chromosomal alleles, such mutations may be the same, or they may be different from one another.

As is well known, an allele may be capable of being expressed by the natural processes operating in a cell. The expression of an allele results in the production of a gene product. The term "allele" as used herein is intended to denote any nucleotide sequence that affects the expression of a particular gene. It thus is intended to refer to any enhancer, promoter, processing, intervening, coding or termination sequence or region of the gene, or any sequence that stabilizes the gene product, or its mRNA, etc.

An allele of a gene is said to be mutated if (1) it is not expressed in a cell or animal, (2) the expression of the allele is altered with respect to the expression of the normal allele of the gene, or (3) the allele expresses a gene product, but that gene product has altered structure, activity, or characteristics relative to the gene product of a normal allele of that gene.

Thus, the terms "mutation" or "mutated" as used herein are intended to denote an alteration in the "normal" or "wild-type" nucleotide sequence of any nucleotide sequence or region of the allele. As used herein, the terms "normal"

GeneDX 1009, pg. 712

and "wild-type" are intended to be synonymous, and to denote any nucleotide sequence typically found in nature. The terms "mutated" and "normal" are thus defined relative to one another; where a cell has two chromosomal alleles of a gene that differ in nucleotide sequence, at least one of these alleles is a "mutant" allele as that term is used herein. A "normal tumor-suppressing gene product" is the gene product that is expressed by a "normal" tumor-suppressing gene.

A mutation may be "cryptic." A cryptic mutation does not affect either the expression of the mutated gene, or the activity or function of the expressed gene product. Cryptic mutations may be detected through nucleotide sequence analysis. Examples of cryptic mutations include mutations that do not result in a change in the amino acid sequence of the expressed gene product, as well as mutations that result in the substitution of an equivalent amino acid at a particular position in the expressed gene product. Most preferably, the mutation will be "non-cryptic" and will therefore introduce a change in the nucleotide sequence of the allele that detectably alters either the expression or the activity or function of the allele. A "mutation that detectably alters the expression of an allele," as used herein denotes any change in nucleotide sequence affecting the extent to which the allele is transcribed, processed or translated. Such alterations may be, for example, in an enhancer, promoter, coding or termination region of the allele, mutations which stabilize the gene product, or its mRNA, etc. A "mutation that detectably alters the activity of an allele," as used herein denotes any change in nucleotide sequence that alters

GeneDX 1009, pg. 713

the capacity of the expressed gene product to mediate a function of the gene product. Such mutations include changes that diminish or inactivate one or more functions of the expressed product. Significantly, such mutations also include changes that result in an increase the capacity of the gene product to mediate any function (for example, a catalytic or binding activity) of that gene product. A "mutation that detectably alters the function of an allele," as used herein denotes any change in nucleotide sequence that alters the capacity of a binding molecule (such as a binding protein) to specifically bind to the allele.

Any of a wide variety of methods (treatment with mutagenic compounds, spontaneous isolation, insertional inactivation, site-specific insertions, deletions or substitutions, homologous recombination, etc.) may be used to produce mutations in accordance with the present invention. As indicated above, a large number of such mutations are known, and mutations can be readily identified by sequencing, tumorigenicity, resilience to tumorigenicity, binding activity, etc. (see, for example, Eliyahu et al., Nature 312:646-649 (1984); Finlay et al., Molec. Cell. Biol. 8:531-539 (1988); Nigro, J.M. et al., Nature 342:705-708 (1989), all herein incorporated by reference).

An allele is said to be "chromosomal" if it either is, or replaces, one of the two alleles of a gene which a cell inherits from its ancestors, or which an animal inherits from its parents. An allele is not "chromosomal," as that term is used herein, if the allele increases the copy number of the total number of alleles of a particular gene which are present in a cell.

GeneDX 1009, pg. 714

The cells that can be produced in accordance with the present invention include both "germ-line" and "somatic" cells. A germ-line cell is a sperm cell or egg cell, or a precursor or progenitor of either; such cells have the potential of transmitting their genome (including the altered tumor-suppressor allele) in the formation of progeny animals. A somatic cell is a cell that is not a germ-line cell. Such cells may be "substantially free of naturally occurring contaminants," or may be present in an animal of the same or of different species. A cell is "substantially free of naturally occurring contaminants" when it, or a precursor or ancestor cell, has been purified from tissue (normal, tumor, etc.) in which the cell is, or would be, naturally associated. Two species are said to be the same if they are capable of breeding with one another to produce fertile offspring. Two species are said to be different if they are either incapable of breeding to produce viable offspring, or are substantially incapable of producing fertile offspring.

The present invention encompasses the formation of such cells and non-human animals for any tumor suppressor gene. Examples of such genes include the rb gene (Weinberg, R.A., Scientific Amer., Sept. 1988, pp 44-51); Hansen M.F. et al., Trends Genet. 4:125-128 (1988)), the gene responsible for Wilms tumor (Maitland, N.J. et al., Anticanc. Res. 9:1417-1426 (1989); Shaw, A.P.W. et al., Oncogene 3:143-150 (1988); Kovacs, G. et al., Proc. Natl. Acad. Sci. (U.S.A.) 85:1571-1575 (1988); Kumar, S. et al., Int. J. Canc. 40:499-504 (1987); Allen, D.M. et al., Clin. Res. 35:416A (1987)) and the p53 gene (Lane, D.P. et al., Genes & Develop. 4:1-8

GeneDX 1009, pg. 715

(1990)). Other tumor suppressor genes may, however, be used in accordance with the present invention (Hiti, A.L., Molec. Cell. Biol. 9:4722-4730 (1989); Hastie, N., Anticanc. Res. 8:1074 (1988); Boada, M.B., Oncologia 10:9-30 (1987); Gallie, B.L., J. Cell. Biochem. 32:215-222 (1986); Alt, F.W. et al., Cold Spr. Harb. Symp. Quant. Biol. 51:931-942 (1986); Knudson, A.G., Symp. Molec. Cell. Biol. Biochem. Molec. Epidemiol. Canc. 40:6-13 (1985); Malcolm, S. Molec. Med. 1:79-94 (1984); Tricoli, J. et al. Amer. J. Hum. Genet. 36:121S (1984); all herein incorporated by reference).

The rb and p53 genes are the preferred tumor suppressor genes of the present invention. The cDNA and genomic forms of the rb gene have been cloned (Friend, S. et al., Nature 323:643 (1986); Lee, W. et al., Science 235:1394 (1987); Fung, Y. et al., Science 236:1657 (1987); Hong, F. et al., Proc. Nat'l. Acad. Sci. (U.S.A.) 86:5502 (1989), all herein incorporated by reference). Potential methods and animal (including transgenic) models for the study of retinoblastoma are discussed in Lee, W. et al., WO 90/05180, herein incorporated by reference. The invention is illustrated below with reference to the p53 tumor suppressor gene. The ability to manipulate this gene and to produce non-human transgenic animals which carry such mutated alleles is illustrated with respect to a particular mutated allele. It is to be understood, however, that the invention and the methods disclosed herein, can be used to produce any possible mutation in the p53 gene. In particular, the invention includes the production of animal cells and non-human transgenic or chimeric animals which carry the particular muta-

GeneDX 1009, pg. 716

tions of the p53 gene that are responsible for the Li-Fraumeni Syndrome discussed below.

I. The p53 Gene

The present invention concerns a non-human animal or an animal (including human) cell in which one of the two naturally present copies of the p53 gene of such non-human animal or animal cell has been rendered non-functional through a mutation (such as a deletion, insertion, or substitution in the naturally occurring p53 gene sequence).

The normal p53 gene product is a tumor-suppressing protein (Sager, R., Science 246:1406-1412 (1989); Finlay, C., Cell 57:1083 (1989), both of which references are herein incorporated by reference). The nature and characteristics of the p53 gene is reviewed by Lane, D. et al. (Genes Devel. 4:1-8 (1990), herein incorporated by reference). The p53 gene plays a protective role against the transforming effects of Friend erythroleukemia virus (Munroe, D. et al. Oncogene 2:621 (1988)). It is also believed to play a role in chromosome stability, differentiation and senescence, and cell proliferation (Sager, R., Science 246:1406-1412 (1989)).

As indicated above, allelic deletions of the p53 locus have been identified in colorectal carcinomas. Mutations in the p53 gene have also been identified in tumors of the brain, breast, and lung (including bronchioalveolar carcinoma, extrapulmonary small cell carcinoma, adenocarcinoma, small cell carcinoma, adenosquamous carcinoma, pulmonary carcinoid tumors, and in a

GeneDX 1009, pg. 717

neurofibrosarcoma (Nigro, J.M. et al., Nature 342:705-708 (1989); Takahashi, T. et al., Science 246:491-494 (1989)). Most tumors with such allelic deletions contain p53 point mutations resulting in amino acid substitutions. Such mutations are not confined to tumors with such allelic deletions, but also occur in some tumors that have retained both parental alleles. Mutations in the p53 gene which give rise to tumors are associated in four "hot-spots" which coincide with the most highly conserved sequences of the p53 gene (Nigro, J.M. et al., Nature 342:705-708 (1989), herein incorporated by reference).

In summary, there is now convincing evidence that the human p53 gene is a tumor suppressor gene (Weinberg, R.A., Scientific Amer., Sept. 1988, pp 44-51). Like the RB protein, discussed above, p53 is a nuclear protein that forms a complex with SV40 large T antigen (DeCaprio, J.A. et al., Cell 54:275-283 (1988); Crawford, L.V., Int. Rev. Exper. Pathol. 25:1-50 (1983)). The binding of these two proteins by viral tumor antigens presumably inactivates them and contributes to transformation. p53 gene deletions have been noted in several mouse erythroleukemic cell lines, reminiscent of the RB gene deletions in retinoblastomas (Mowat et al., Nature 314:633-636 (1985); Chow et al., J. Virol. 61:2777-2781 (1987); Hicks, G.G. et al., J. Virol. 62:4752-4755 (1988)). Cell lines and tumors derived from human osteogenic sarcomas often contain gross rearrangements of the p53 gene, including deletions (Masuda, H., Proc. Nat'l. Acad. Sci. (U.S.A.) 84:7716-7719 (1987)). Allelic deletions in chromosome 17p (which contains the p53 gene) occur in over 75% of colorectal carcinomas (Baker, S.J. et

GeneDX 1009, pg. 718

al., Science 244:217-221 (1989)). In two tumors, the remaining non-deleted p53 allele was shown to contain mutants in highly conserved regions previously found to be mutated in murine p53 genes (Baker, S.J. et al., Science 244:217-221 (1989)). Loss of heterozygosity in chromosome 17p has been noted in a high percentage of individuals with small cell lung carcinoma (Yokoto, J. et al., Proc. Nat'l. Acad. Sci. (U.S.A.) 84:9252-9526 (1987); Harbour, J.W. et al., Science 241:353-356 (1988)). In the HL-60 human leukemic cell line, major deletions in the p53 gene and absence of the p53 protein have been noted (Wolf, D. et al., Proc. Nat'l. Acad. Sci. (U.S.A.) 82:790-794 (1985)). These results indicate that the absence of a functional p53 allele is highly correlated with some forms of cancer in humans and mice, strongly suggesting a tumor-suppressor role for p53. Finally, Finlay et al. (Cell 57:1083-1093 (1989)) have recently demonstrated that the wild-type mouse p53 gene suppresses transformation in vitro after cotransfection of rat embryo cells with E1a and activated ras, indicating that the presence of the normal p53 gene acts negatively to block transformation.

The cDNA and genomic forms of the p53 gene have been cloned (Pennica, D. et al., Virol. 134:477-482 (1984); Jenkins, J. et al., Nature 312:651-654 (1984); Oren, M. et al., EMBO J. 2:1633-1639 (1983); Zahut-Houri, R. et al., Nature 306:594-597 (1983), all of which references are herein incorporated by reference).

Recent evidence has suggested that a mutation in the p53 gene may be responsible for the Li-Fraumeni Syndrome, a rare human genetic disorder (Malkin, D. et al., Science 250:1233-

1238 (1990); Marx, J., Science 250:1209 (1990), both references herein incorporated by reference). Individuals afflicted with this disease are highly susceptible to several malignant tumors - breast carcinomas, soft tissue sarcomas, brain tumors, osteosarcomas, leukemia, and adrenocortical carcinoma. The disease is also associated with a higher incidence of melanoma, gonadal germ cell tumors, and carcinomas of the lung, pancreas and prostate (Li, F.P. et al., Ann. Intern. Med. 71:747 (1969); Birch, J.M. et al., J. Clin. Oncol. 8:583 (1990); Birch, J.M. et al., Brit. J. Canc. 49:325 (1984); Li, F.P. et al., Canc. Res. 48:5358 (1988); Williams, W.R. et al., Familial Canc., 1st Int. Res. Conf. p. 151 (Karger, Basel, 1985); Strong, L.C. et al., J. Natl. Canc. Inst. 79:1213 (1987)).

The Li-Fraumeni Syndrome is associated with a particular set of mutations in exon 7 of the p53 gene (Malkin, D. et al., Science 250:1233-1238 (1990), herein incorporated by reference). Knowledge of the sequence and location of the mutation which causes the Li-Fraumeni syndrome permits one to use the methods of the present invention to produce a non-human transgenic animal which contains any of these mutations. As indicated above, such an animal, and its uses in diagnostics and gene therapy, is an embodiment of the present invention.

II. The Interaction of Mutant and Normal p53 Gene Products

A cascade of mutational events is believed to be required for tumorigenesis. It is believed that this

GeneDX 1009, pg. 720

cascade involves both the activation of one or more oncogene(s) and the inactivation of one or more tumor-suppressing gene(s) (Sager, R., Science 246:1406-1412 (1989); Fearson, E.R. et al., Cell 61:759-767 (1990), both herein incorporated by reference). Indeed, mutations in at least 4-5 genes are believed to be required for the formation of a malignant tumor.

Studies of the clonal nature of tumor formation have suggested that tumors have a monoclonal composition, and hence arise by the clonal propagation of a single progenitor cell (Fearson, E.R. et al., Science 247:193-197 (1987)). The loss (either by mutation or deletion) of functional p53 gene function has been found to be a characteristic of colorectal tumors (Fearson, E.R. et al., Cell 61:759-767 (1990)).

The simplest model to explain the mechanism of action of a tumor-suppressing gene is that malignancy requires two separate genetic events (i.e. loss by deletion or mutation of both functional p53 alleles in a cell). Inactivation of only one of the two natural p53 alleles causes a phenotypically silent, recessive mutation which does not result in malignancy in the absence of the second mutational event.

It has been observed that, by introducing DNA containing a mutated p53 gene into rodent cells, such cells can be transformed. This transformation is due to a cooperative action of the mutant p53 gene product with the gene product of the ras oncogene. Since such cells also express the normal p53 gene, it has been proposed that the inactivation of only one p53 allele is not phenotypically silent

(Fearson, E.R. et al., Cell 61:759-767 (1990)). The ability of the cells to undergo transformation in the presence of both the normal and the mutant p53 gene products suggests that the two gene products may interact to exert their effects. One explanation of the mixed recessive-dominant characteristic of the p53 mutant/normal allele interaction is that oligomerization is needed for activity, and that the oligomer formed in cells having both a mutant and a normal p53 allele contains a mixture of the gene products of these alleles (Eliyahu et al., Oncogene 3:313-321 (1988); Kraiss, et al., J. Virol. 62:4737-4744 (1988)). Due to the presence of the mutant p53 gene product in such mixed oligomers, the oligomers would be expected to have a diminished ability to suppress tumor formation (Fearson, E.R. et al., Cell 61:759-767 (1990), herein incorporated by reference). Consistent with this explanation has been the finding that mutant p53 gene products form heterodimers with the wild-type p53 gene product (Finlay, C.A., Cell 57:1083 (1989)).

Transgenic animals may be used to investigate the biological implications of tumor-suppressing genes (Capecchi, M.R., Science 244:1288-1292 (1989)). Lavigueur, A. et al. constructed a transgenic mouse which had a single added mutant p53 gene in addition to the endogenous two wild-type p53 alleles. The mouse, and its progeny overexpressed the added p53 gene. The mice were found to have a high incidence of lung, bone, and lymphoid tumors (Lavigueur, A. et al., Molec. Cell. Biol. 9:3982-3991 (1989). The high levels of mutant p53 in these mice led to tumor formation in 22 of 112 mice examined. Due to the overexpression of the mutant p53 gene product, and to the

GeneDX 1009, pg. 722

ability of this mutant protein to impair the biological activity of the normal p53 gene product, the transgenic mice of Lavigueur et al. exhibit the phenotype of an animal in which all normal, endogenous p53 gene product activity has been functionally inactivated (Lavigueur, A. et al., Molec. Cell. Biol. 9:3982-3991 (1989); Sager, R., Science 246:1406-1412 (1989)).

Thus, the mice described by Lavigueur, A. et al. (Molec. Cell. Biol. 9:3982-3991 (1989)) contained an added oncogenic p53 allele in addition to its two normal p53 genes. Importantly, the effect of gene dosage in such animals is unknown. Moreover, it is not known whether the added p53 gene in such animals contains all of the normal upstream and downstream regulatory elements. Even if all of such elements were present, the context of the integration of the construct may render the regulation of the gene abnormal. For these reasons, the utility of the mice in assessing the role of the p53 gene on oncogenesis is unclear.

It would be desirable to produce a transgenic animal whose genome possesses one normal and functional p53 allele and one non-functional (mutant) p53 allele. Such animals could be used to study the consequences resulting from the loss of one p53 allele, and thus would more clearly aid in elucidating the processes of oncogenesis and tumorigenesis. Such animals would also be useful in screening potential carcinogens, in developing novel antineoplastic therapeutics, and in gene therapy. The present invention provides such an animal.

GeneDX 1009, pg. 723

III. Homologous Recombination

The present invention uses the process of homologous recombination to introduce a specific mutation into the naturally present p53 sequence of an animal cell, most preferably an embryonic stem (ES) cell. The mutated ES cells of non-human animals can then be either cultured in suitable cell culture medium, or introduced into the uterus of a suitable recipient and permitted to develop into a non-human animal.

Alternatively, the methods of the present invention may be used to alter the somatic cells of a non-human animal to produce a chimeric non-human animal.

An understanding of the process of homologous recombination (Watson, J.D., In: Molecular Biology of the Gene, 3rd Ed., W.A. Benjamin, Inc., Menlo Park, CA (1977), which reference is incorporated herein by reference) is thus desirable in order to fully appreciate the present invention.

In brief, homologous recombination is a well-studied natural cellular process which results in the scission of two nucleic acid molecules having identical or substantially similar sequences (i.e. "homologous"), and the ligation of the two molecules such that one region of each initially present molecule is now ligated to a region of the other initially present molecule (Sedivy, J.M., Bio-Technol. 6:1192-1196 (1988), which reference is incorporated herein by reference).

Homologous recombination is, thus, a sequence specific process by which cells can transfer a "region" of DNA from

GeneDX 1009, pg. 724

one DNA molecule to another. As used herein, a "region" of DNA is intended to generally refer to any nucleic acid molecule. The region may be of any length from a single base to a substantial fragment of a chromosome.

For homologous recombination to occur between two DNA molecules, the molecules must possess a "region of homology" with respect to one another. Such a region of homology must be at least two base pairs long. Two DNA molecules possess such a "region of homology" when one contains a region whose sequence is so similar to a region in the second molecule that homologous recombination can occur.

Recombination is catalyzed by enzymes which are naturally present in both prokaryotic and eukaryotic cells. The transfer of a region of DNA may be envisioned as occurring through a multi-step process.

If either of the two participant molecules is a circular molecule, then the recombination event results in the integration of the circular molecule into the other participant.

Importantly, if a particular region is flanked by regions of homology (which may be the same, but are preferably different), then two recombinational events may occur, and result in the exchange of a region of DNA between two DNA molecules. Recombination may be "reciprocal," and thus results in an exchange of DNA regions between two recombining DNA molecules. Alternatively, it may be "non-reciprocal," (also referred to as "gene conversion") and result in both recombining nucleic acid molecules having the same nucleotide sequence. There are no constraints

GeneDX 1009, pg. 725

regarding the size or sequence of the region which is exchanged in a two-event recombinational exchange.

The frequency of recombination between two DNA molecules may be enhanced by treating the introduced DNA with agents which stimulate recombination. Examples of such agents include trimethylpsoralen, UV light, etc.

IV. Production of Chimeric and Transgenic Animals: Gene Targeting Methods

One approach to producing animals having defined and specific genetic alterations has used homologous recombination to control the site of integration of an introduced marker gene sequence in tumor cells and in fusions between diploid human fibroblast and tetraploid mouse erythroleukemia cells (Smithies, O. et al., Nature 317:230-234 (1985)).

This approach was further exploited by Thomas, K. R., and co-workers, who described a general method, known as "gene targeting," for targeting mutations to a preselected, desired gene sequence of an ES cell in order to produce a transgenic animal (Mansour, S.L. et al., Nature 336:348-352 (1988); Capecchi, M.R., Trends Genet. 5:70-76 (1989); Capecchi, M.R., Science 244:1288-1292 (1989); Capecchi, M.R. et al., In: Current Communications in Molecular Biology, Capecchi, M.R. (ed.), Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), pp. 45-52; Frohman, M. A. et al., Cell 56:145-147 (1989); all of which references are incorporated herein by reference).

GeneDX 1009, pg. 726

It may now be feasible to deliberately alter any gene in a mouse (Capecchi, M.R., Trends Genet. 5:70-76 (1989); Frohman, M. A. et al., Cell 56:145-147 (1989)). Gene targeting involves the use of standard recombinant DNA techniques to introduce a desired mutation into a cloned DNA sequence of a chosen locus. That mutation is then transferred through homologous recombination to the genome of a pluripotent, embryo-derived stem (ES) cell. The altered stem cells are microinjected into mouse blastocysts and are incorporated into the developing mouse embryo to ultimately develop into chimeric animals. In some cases, germ line cells of the chimeric animals will be derived from the genetically altered ES cells, and the mutant genotypes can be transmitted through breeding.

Gene targeting has been used to produce chimeric and transgenic mice in which an nptII gene has been inserted into the β_2-microglobulin locus (Koller, B.H. et al., Proc. Natl. Acad. Sci. (U.S.A.) 86:8932-8935 (1989); Zijlstra, M. et al., Nature 342:435-438 (1989); Zijlstra, M. et al., Nature 344:742-746 (1989); DeChiaba et al., Nature 345:78-80 (1990)). Similar experiments have enabled the production of chimeric and transgenic animals having a c-abl gene which has been disrupted by the insertion of an nptII gene (Schwartzberg, P.L. et al., Science 246:799-803 (1989)). The technique has been used to produce chimeric mice in which the en-2 gene has been disrupted by the insertion of an nptII gene (Joyner, A.L. et al., Nature 338:153-155 (1989)).

Gene targeting has also been used to correct an hprt deficiency in an hprt$^-$ ES cell line. Cells corrected of the

GeneDX 1009, pg. 727

deficiency were used to produce chimeric animals. Significantly, all of the corrected cells exhibited gross disruption of the regions flanking the hprt locus; all of the cells tested were found to contain at least one copy of the vector used to correct the deficiency, integrated at the hprt locus (Thompson, S. et al., Cell 56:313-321 (1989); Koller, B.H. et al., Proc. Natl. Acad. Sci. (U.S.A.) 86:8927-8931 (1989)).

In order to utilize the "gene targeting" method, the gene of interest must have been previously cloned, and the intron-exon boundaries determined. The method results in the insertion of a marker gene (i.e. the nptII gene) into a translated region of a particular gene of interest. Thus, use of the gene targeting method results in the gross destruction of the gene of interest.

Recently, chimeric mice carrying the homeobox hox 1.1 allele have been produced using a modification of the gene targeting method (Zimmer, A. et al., Nature 338:150-154 (1989). In this modification, the integration of vector sequences was avoided by microinjecting ES cells with linear DNA containing only a portion of the hox 1.1 allele, without any accompanying vector sequences. The DNA was found to cause the gene conversion of the cellular hox allele. Selection was not used to facilitate the recovery of the "converted" ES cells, which were identified using the polymerase chain reaction ("PCR"). Approximately 50% of cells which had been clonally purified from "converted" cells were found to contain the introduced hox 1.1 allele, suggesting to Zimmer, A. et al. either chromosomal instability or contamination of sample. None of the

GeneDX 1009, pg. 728

chimeric mice were found to be able to transmit the "converted" gene to their progeny (Zimmer, A. et al., In: Current Communications in Molecular Biology, Capecchi, M.R. (ed.), Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), pp. 53-58).

Significantly, the use of gene targeting to alter a gene of a cell results in the formation of a gross alteration in the sequence of that gene. The efficiency of gene targeting depends upon a number of variables, and is different from construct to construct. For the p53 gene constructs used herein, such efficiency is approximately 1/300.

V. Production of Chimeric and Transgenic Animals: Use of Insertion Vectors

In contrast to the above-described methods, the present invention uses methods capable of producing subtle, precise, and predetermined mutations in the sequence of one of the two alleles of the p53 gene of a human or animal cell. Although the methods discussed below are capable of mutating both alleles of the cell's p53 gene, it is possible to readily identify (for example through the use of PCR (discussed below), or other methods) such dual mutational events. Since the frequency of such dual mutational events is the square of the frequency of a single mutational event, cells having mutations in both of their p53 alleles will be only a very small proportion of the total population of mutated cells.

The present invention has several embodiments. In the simplest embodiment, an insertion vector is used to mutate

GeneDX 1009, pg. 729

the nucleotide sequence of one of the two alleles of the p53 gene of a cell. The use of this vector type in combination with a second selectable reversion event prevents the disruption of the chromosome by the selectable marker gene (an nptII gene, for example) of the vector or by other vector sequences. Thus, gross distortions of the recipient chromosome are avoided. The efficiency of the gene targeting is substantially greater than in the gene targeting methods discussed above.

Most preferably, the DNA molecule(s) which are to be introduced into the recipient cell contains a region of homology with a region of the p53 gene. In a preferred embodiment, the DNA molecule will contain two regions of homology with the cell's p53 gene. These regions of homology will preferably flank the precise sequence whose incorporation into the p53 gene is desired. As stated above, the regions of homology may be of any size greater than two bases long. Most preferably, the regions of homology will be greater than 10 bases long.

The DNA molecule(s) may be single stranded, but are preferably double stranded. The DNA molecule(s) may be introduced to the cell as one or more RNA molecules which may be converted to DNA by reverse transcriptase or by other means. Preferably, the DNA molecule will be double stranded linear molecule. In the best mode for conducting the invention, such a molecule is obtained by cleaving a closed covalent circular molecule to form a linear molecule. Preferably, a restriction endonuclease capable of cleaving the molecule at a single site to produce either a blunt end or staggered end linear molecule is employed. Most

GeneDX 1009, pg. 730

preferably, the nucleotides on each side of this restriction site will comprise at least a portion of the preferred two regions of homology between the DNA molecule being introduced and the p53 gene.

The invention thus provides a method for altering the natural p53 gene sequence through the introduction of a "desired gene sequence" into that gene. The "desired gene sequence" may be of any length, and have any nucleotide sequence. It may comprise one or more gene sequences which encode complete proteins, fragments of such gene sequences, regulatory sequences, etc. Significantly, the desired gene sequence may differ only slightly from a native gene of the recipient cell (for example, it may contain single, or multiple base alterations, insertions or deletions relative to the native gene). The use of such desired gene sequences permits one to create subtle and precise changes in the p53 gene of the recipient cell. Thus, the present invention provides a means for manipulating and modulating the expression and regulation of the p53 gene.

In particular, the invention provides a mean for manipulating and modulating p53 gene expression and protein structure through the replacement of a naturally present p53 gene sequence with a "non-selectable" "desired gene sequence." A gene sequence is non-selectable if its presence or expression in a recipient cell provides no survival advantage to the cell under the culturing conditions employed. Thus, by definition, one cannot select for cells which have received a "non-selectable" gene sequence in their p53 gene. In contrast, a "dominant" gene sequence is one which can under certain circumstances

GeneDX 1009, pg. 731

provide a survival advantage to a recipient cell. The neomycin resistance conferred by the nptII gene is a survival advantage to a cell cultured in the presence of neomycin or G418. The nptII gene is thus a dominant, rather than a non-selectable gene sequence.

In particular, the invention permits the replacement of the naturally present p53 gene sequence of a recipient cell with an "analog" sequence. A sequence is said to be an analog of another sequence if the two sequences are substantially similar in sequence, but have minor changes in sequence corresponding to single base substitutions, deletions, or insertions with respect to one another, or if they possess "minor" multiple base alterations. Such alterations are intended to exclude insertions of dominant selectable marker genes.

When the desired gene sequence, flanked by regions of homology with the p53 gene sequence of the recipient cell, is introduced into the recipient cell as a linear double stranded molecule, whose termini correspond to the regions of homology, a single recombination event with the p53 gene of the cell will occur in approximately 5% of the transfected cells. Such a single recombinational event will lead to the integration of the entire linear molecule into the genome of the recipient cell.

The structure generated by the integration of the linear molecule will undergo a subsequent, second recombinational event (approximately 10^{-5} - 10^{-7} per cell generation). This second recombinational event will result in the elimination of all DNA except for the flanking regions of homology, and the desired DNA sequence from the integrated structure.

GeneDX 1009, pg. 732

Thus, the consequence of the second recombinational event is to replace the DNA sequence which is normally present between the flanking regions of homology in the cell's p53 gene, with the desired DNA sequence, and to eliminate the instability of gene replacement.

The DNA molecule containing the desired gene sequence may be introduced into the pluripotent cell by any method which will permit the introduced molecule to undergo recombination at its regions of homology. Some methods, such as direct microinjection, or calcium phosphate transformation, may cause the introduced molecule to form concatemers upon integration. These concatemers may resolve themselves to form non-concatemeric integration structures. Since the presence of concatemers is not desired, methods which produce them are not preferred. In a preferred embodiment, the DNA is introduced by electroporation (Toneguzzo, F. et al., Nucleic Acids Res. 16:5515-5532 (1988); Quillet, A. et al., J. Immunol. 141:17-20 (1988); Machy, P. et al., Proc. Natl. Acad. Sci. (U.S.A.) 85:8027-8031 (1988); all of which references are incorporated herein by reference).

After permitting the introduction of the DNA molecule(s), the cells are cultured under conventional conditions, as are known in the art.

In order to facilitate the recovery of those cells which have received the DNA molecule containing the desired gene sequence, it is preferable to introduce the DNA containing the desired gene sequence in combination with a second gene sequence which would contain a detectable marker gene sequence. For the purposes of the present invention, any

GeneDX 1009, pg. 733

gene sequence whose presence in a cell permits one to recognize and clonally isolate the cell may be employed as a detectable marker gene sequence.

In one embodiment, the presence of the detectable marker sequence in a recipient cell is recognized by hybridization, by detection of radiolabelled nucleotides, or by other assays of detection which do not require the expression of the detectable marker sequence. Preferably, such sequences are detected using PCR (Mullis, K. et al., Cold Spring Harbor Symp. Quant. Biol. 51:263-273 (1986); Erlich H. et al., EP 50,424; EP 84,796, EP 258,017, EP 237,362; Mullis, K., EP 201,184; Mullis K. et al., US 4,683,202; Erlich, H., US 4,582,788; and Saiki, R. et al., US 4,683,194), which references are incorporated herein by reference).

PCR achieves the amplification of a specific nucleic acid sequence using two oligonucleotide primers complementary to regions of the sequence to be amplified. Extension products incorporating the primers then become templates for subsequent replication steps. PCR provides a method for selectively increasing the concentration of a nucleic acid molecule having a particular sequence even when that molecule has not been previously purified and is present only in a single copy in a particular sample. The method can be used to amplify either single or double stranded DNA.

Most preferably, however, the detectable marker gene sequence will be expressed in the recipient cell, and will result in a selectable phenotype. Examples of such preferred detectable gene sequences include the hprt gene (Littlefield, J.W., Science 145:709-710 (1964), herein

GeneDX 1009, pg. 734

incorporated by reference), the tk gene (i.e. thymidine kinase gene) and especially the tk gene of herpes simplex virus (Giphart-Gassler, M. et al., Mutat. Res. 214:223-232 (1989) herein incorporated by reference), the nptII gene (Thomas, K.R. et al., Cell 51:503-512 (1987); Mansour, S.L. et al., Nature 336:348-352 (1988), both references herein incorporated by reference), or other genes which confer resistance to amino acid or nucleoside analogues, or antibiotics, etc.

Cells which express an active HPRT enzyme are unable to grow in the presence of certain nucleoside analogues (such as 6-thioguanine, 8-azapurine, etc.), but are able to grow in media supplemented with HAT (hypoxanthine, aminopterin, and thymidine). Conversely, cells which fail to express an active HPRT enzyme are unable to grow in media containing HATG, but are resistant to analogues such as 6-thioguanine, etc. (Littlefield, J.W., Science 145:709-710 (1964)). Cells expressing active thymidine kinase are able to grow in media containing HATG, but are unable to grow in media containing nucleoside analogues such as 5-azacytidine (Giphart-Gassler, M. et al., Mutat. Res. 214:223-232 (1989)). Cells containing an active HSV-tk gene are incapable of growing in the presence of gangcylovir or similar agents.

The detectable marker gene may be any gene which can complement for a recognizable cellular deficiency. Thus, for example, the gene for HPRT could be used as the detectable marker gene sequence when employing cells lacking HPRT activity. Thus, this agent is an example of agents may be used to select mutant cells, or to "negatively select" for cells which have regained normal function.

GeneDX 1009, pg. 735

The nptII gene (Southern, P.J., et al., J. Molec. Appl. Genet. 1:327-341 (1982); Smithies, O. et al., Nature 317:230-234 (1985), which references are incorporated herein by reference) is the most preferred detectable marker gene sequence. Constructs which contain both an nptII gene and either a tk gene or an hprt gene are especially preferred.

A. Use of a Single DNA Molecule Containing Both the Detectable Marker Sequence and the Desired Gene Sequence

In a first preferred embodiment, the detectable marker gene sequence, flanked by the regions of homology to the p53 gene, is provided to the recipient cells on the same DNA molecule which contains the desired gene sequence. As discussed previously, it is preferred that this DNA molecule be a linear molecule.

After selection for cells which have incorporated the desired DNA molecule (for example by selection for G418 resistant cells when the detectable marker gene sequence is an expressible nptII gene sequence), the cells are cultured, and the presence of the introduced DNA molecule is confirmed as described above. Approximately 10^7 cells are cultured and screened for cells which have undergone the second recombinational event (discussed above) resulting in the replacement of a native sequence (i.e. a gene sequence which is normally and naturally present in the recipient cell) with the desired gene sequence.

Any of a variety of methods may be used to identify cells which have undergone the second recombinational event. Direct screening of clones, use of PCR, use of hybridization

GeneDX 1009, pg. 736

probes, etc., may all be employed for this purpose. In a preferred embodiment, the DNA molecule will, in addition to the desired gene sequence, the flanking regions of homology and the detectable marker gene sequence, contain an additional gene sequence which will permit the selection or recognition of cells which have undergone the second recombinational event. This additional gene sequence will be excised from the cell's p53 gene as a direct consequence of the second recombinational event. Thus, gene sequences which are suitable for this purpose include any gene sequence whose loss from a cell can be detected or selected for. Examples of such "negative selection" gene sequences include the hprt gene, and the tk gene (especially the tk gene of herpes simplex virus).

In the first preferred embodiment, the frequency of the second recombinational event is approximately 10^{-5}. However, the use of a "negative selection" gene sequence permits one to identify such recombinant cells at a frequency of approximately 100%.

The DNA molecule may have a region of heterology located at the proposed insertion site. Insertion of such a vector permits one to select for recombinants which have recombined at the insertion site (and not at other potential sites). If recombination occurs at the desired insertion site, it will lead to the loss of the sequence of heterology located at the proposed insertion site of the DNA molecule (HSVtk, for example). Insertions which result from other recombinational events will retain the sequence of heterology. Thus, by employing a region of heterology which encodes an assayable gene product, or which can be used as

GeneDX 1009, pg. 737

a "negative selectable" marker, one can readily determine that the locus of insertion of the recipient cell contains the precise sequence desired. The efficiency of such a vector is approximately 0.5%.

The region of heterology which may be introduced at the insertion site of the DNA molecule may be either short or of substantial size (for example, 2 kb). The site of linearization may be 5', 3', or within the region of heterology. When the site of linearization is within the region of heterology, the efficiency of gene targeting is approximately 2%.

The region of heterology may be located at a site internal to the region of homology where the desired recombination shall occur. Such a construct can be used when one desires to introduce a subtle mutation into a locus of the cellular gene at a site other than that of the site of desired recombination.

B. Use of a Different DNA Molecules to Provide the Detectable Marker Sequence and the Desired Gene Sequence

In a second preferred embodiment, the detectable marker gene sequence, flanked by the regions of homology, will be provided to the recipient cell on a different DNA molecule from that which contains the desired gene sequence. It is preferred that these molecules be linear molecules.

When provided on separate DNA molecules, the detectable marker gene sequence and the desired gene sequence will most preferably be provided to the recipient cell by co-electroporation, or by other equivalent techniques.

GeneDX 1009, pg. 738

After selection of such recipients (preferably through the use of a detectable marker sequence which expresses the nptII gene and thus confers cellular resistance to the antibiotic G418), the cells are grown up and screened to confirm the insertion event (preferably using PCR).

In the absence of any selection, only one cell in 10^7 would be expected to have the predicted recombinant structures. If, however, one selects for recipient cells which contain and express a detectable marker sequence (such as the nptII gene), it is possible to obtain a 10^3 to 10^5 fold enrichment for cells which have taken up both DNA molecules. Typically, such enrichment enables one to identify the desired recipient cell (in which the introduced DNA has integrated into the cell's genome) by screening only 800 -1,500 cells. Such screening is preferably done using PCR, or other equivalent methods. Using such negative selection techniques, one may manipulate the vector copy number.

The two introduced DNA molecules will generally not have integrated into the same site in the genome of the recipient cell. Thus, in some cases, the desired gene sequence will have integrated in a manner so as to replace the native cellular gene sequence between the flanking regions of homology. The locus of integration of the detectable marker gene is unimportant for the purposes of the present invention, provided it is not genetically linked to the locus of the p53 gene. If desired, however, it is possible to incorporate a gene sequence capable of negative selection along with the DNA containing the detectable marker sequence. Thus, one can ultimately select for cells which

GeneDX 1009, pg. 739

have lost the introduced selectable marker gene sequence DNA.

C. Use of Direct Selection to Identify Homologous Recombination Events

Although all of the above-described preferred embodiments enable the isolation of cells in which one of the cell's p53 alleles has been mutated to contain a desired gene sequence, each embodiment requires the screening of a significant number of candidate cells in order to identify the desired recombinant cell. It is, however, possible to directly select for the desired recombinant cell by employing a variation of the above embodiments.

The method for direct selection of the desired cells relies upon the phenotypic difference in targeted and non-targeted cells and the use of a single gene which can be used for both positive and negative selection.

Typically, in any homologous recombination experiment performed with an insertion vector, three populations of cells will be created. The first class of cells will be those which have failed to receive the desired DNA molecule. This class will comprise virtually all of the candidate cells isolated on completion of the experiment. The second class of cells will be those cells in which the desired gene sequence has been incorporated at a random insertion site (i.e. a site other than in the p53 gene). Approximately one cell in 10^3-10^4 total cells will be in this class. The third class of cells will be those cells in which the desired gene sequence has been incorporated by homologous recombination

GeneDX 1009, pg. 740

into a site in the p53 gene. Approximately one cell in 10^5-10^6 total cells will be in this class.

In the above-described embodiments, the cells of the first class (non-transfected cells) can be eliminated by positive selection, thus necessitating the screening of only about 1,000 cells in order to identify the desired recombinant cell. In the present embodiment, cells of the third class (homologous recombinants) may be selected from the cells of the second class (random insertions) if a phenotypic difference exists between the cells of the two classes.

Since random integration sites are likely to be concatemeric with few single copy clones (depending upon the DNA concentration with which the cells were transfected), such integration events are inherently unstable. Thus, such concatemeric constructs will typically undergo intrachromosomal recombination. Such recombination will always leave one intact copy of the vector in the genome. Thus, all random insertion events may be negatively selected from the population if a negatively selectable marker is included on the vector.

In contrast, cells in which the desired gene sequence has been incorporated into the p53 gene by homologous recombination will revert with a relatively high frequency (approximately 1 in 10^4-10^5 per cell division (depending upon the size of the duplicated structure) to produce a mutated p53 gene that does not contain vector sequences. Therefore, even if the vector contained a negatively selectable gene sequence, such cells will survive negative selection, and can be recovered. The small percentage of homologous

GeneDX 1009, pg. 741

recombinant cells which have not undergone reversion will also be eliminated by the negative selection.

A preferred negative selectable marker is the hprt gene (cells expressing an active HPRT enzyme are unable to grow in the presence of certain nucleoside analogues such as 6-thioguanine etc.). When using 6-thioguanidine as a negative selection agent, a density of 10^7 cells is preferably used since the efficiency of 6-thioguanidine selection is cell density dependent. A typical experiment with 10^7 transfected cells would yield approximately 10 revertant cells after successive selection. The relative yield of revertant clones can be substantially increased by using "Poly A Selection" for the first round of selection.

In such a "Poly A Selection" one exploits the fact that, if an introduced DNA molecule were to integrate at random into the host chromosome, it would generally not integrate at a site adjacent to a necessary 3' polyadenylation site. Thus, the mRNA produced by the transcription of such randomly inserted constructs would generally lack polyadenylation. This fact can be exploited by using vectors which permit one to select for a recombinational event that results in integration adjacent to the natural polyadenylation site of the introduced gene sequence (i.e. by homologous recombination rather than by random insertion). As stated above, the frequency of obtaining a desired recombinant cell is approximately 10^{-3}. By using Poly A Selection, desired cells can be recovered at a frequency of approximately 10^{-2}. Thus, the poly A selection results in an approximate increase of overall efficiency of nearly 10 fold. Poly A selection may therefore be advantageously used

GeneDX 1009, pg. 742

in situations where one desires to minimize or avoid the screening of colonies to identify random versus homologous recombinants.

D. Production of altered p53 alleles containing heterologous sequences

As stated above, the desired gene sequence may be of any length, and have any nucleotide sequence. In particular, it is possible to design the sequence of the desired gene sequence in order to create single, or multiple base alterations, insertions or deletions in any preselected gene of a cell.

If such changes are within a translated region of the p53 gene sequence, then a new protein variant of the p53 gene product can be obtained.

The present invention may be used to produce cells in which the natural p53 gene has been replaced with an altered gene sequence, or a heterologous p53 gene. A p53 gene is said to be heterologous to a transgenic cell if it is derivable from a species other than that of the transgenic cell.

In one embodiment, this replacement may be accomplished in a single step. To accomplish such replacement, a DNA molecule containing a desired gene sequence and a region of homology with the p53 gene is introduced into a recipient cell. A selectable marker gene is also introduced into the cell, and used to select for cells which have underwent recombination. The method results in the replacement of the normal sequences adjacent to the region of homology with the heterologous sequences of the desired DNA sequence.

GeneDX 1009, pg. 743

In a second embodiment, this replacement may be accomplished in two steps. As in the embodiment described above, a cell is provided with a DNA molecule containing a desired gene sequence and a region of homology with the p53 gene. The DNA molecule also contains a selectable marker gene used to select for cells which have undergone a recombinational event that has resulted in the insertion of the introduced DNA molecule into their chromosomes at the site of homology.

Significantly, in this embodiment, the introduced DNA molecule will also contain a "negative selectable" marker gene which can be used to select for cells which undergo a second recombinational event that results in the loss of the inserted DNA.

A second DNA molecule is employed to complete the gene replacement. This second DNA molecule need not contain any selectable marker gene. Upon receipt of the second DNA molecule, a second recombinational event occurs which exchanges the "second" DNA molecule for the integrated "first" DNA molecule (including the desired DNA sequence, the selectable marker sequence, and the "negative selectable" marker sequence contained on that molecule).

In another embodiment of the invention, subtle mutations may be introduced into a desired locus using a "cassette" construct containing both a positive selection marker (such as the nptII gene or the gpt gene) and a negative selection marker (such as the tk gene). In this embodiment, one first uses the positive selection capacity of the construct to introduce the two selection markers into a desired locus. One then introduces the desired subtle mutations (substi-

GeneDX 1009, pg. 744

tutions, insertions, deletions, etc.) by providing a cell with a DNA molecule that contains the desired mutation. By selecting for the loss of the "cassette" (using the negative selection marker), one can select for recombinational events which result in the replacement of the "cassette" sequence with the DNA sequence containing the desired mutation.

The present invention may also be used to replace contiguous regions of a chromosome with any desired gene sequence. Thus, the present invention is not limited in the size of the DNA regions which may be altered or replaced. This aspect of the present invention may be considered as a series of 5 steps. The first step in replacing a large region of a chromosome with a desired sequence involves setting up an initial target. In this step, a recipient cell is provided with a DNA molecule which contains a "first fragment" of the total desired replacement sequence. This "first fragment" of the desired replacement sequence contains a selectable marker sequence (most preferably the nptII gene) at its end.

The DNA molecule also contains a "dual selection" gene sequence which encodes a non-functional fragment of a gene sequence for which both a positive and a negative selection exists. An example of such a gene is the gpt gene when used in the context of an hprt$^-$ cell. Cells which express a functional gpt gene can be selected for by their ability to grow in HAT medium; Cells which lack a functional gpt gene can be selected for by their ability to grow in the presence of 6-thioguanine.

Homologous recombination results in the insertion of the DNA molecule into the cell's genome at the region of

homology. Importantly, since this step results in the creation of a cell whose genome contains the selectable marker gene, it is possible to select for the desired recombinational event.

In the second step of the method, a second DNA molecule is provided to the cell. This second DNA molecule contains a "second fragment" of the desired replacement sequence as well as a sequence of the dual selection gene that, due to an internal deletion, is incapable of encoding a functional gene product. Homologous recombination results in the insertion of the second DNA molecule into the cell's genome in a manner so as to create a functional dual selection gene. Recombination also results in the integration of a non-functional fragment of the dual selection gene. Importantly, since this step results in the creation of a cell whose genome contains a functional dual selection gene, it is possible to select for the desired recombinational event.

In the third step of the method, a third DNA molecule is provided to the cell. This third DNA molecule contains both the "first" and "second" fragments of the desired replacement sequence. Homologous recombination results in the insertion of the third DNA molecule into the cell's genome in a manner so as to delete the functional dual selection gene. The non-functional fragment of the dual selection gene (formed in step 2) is not affected by the recombination, and is retained. Importantly, since this step results in the creation of a cell whose genome lacks the dual selection gene, it is possible to select for the desired recombinational event.

GeneDX 1009, pg. 746

In the fourth step of the method, a fourth DNA molecule is provided to the cell. This fourth DNA molecule contains a "third fragment" of the desired replacement sequence as well as a sequence of the dual selection gene that, as in step 2, is incapable of encoding a functional gene product due to an internal deletion. Homologous recombination results in the insertion of the fourth DNA molecule into the cell's genome in a manner so as to create a functional dual selection gene. Recombination also results in the integration of a non-functional fragment of the dual selection gene. Importantly, since this step results in the creation of a cell whose genome contains a functional dual selection gene, it is possible to select for the desired recombinational event.

In the fifth step of the method, a fifth DNA molecule is provided to the cell. This fifth DNA molecule contains both the "second" and "third" fragments of the desired replacement sequence. Homologous recombination results in the insertion of the fifth DNA molecule into the cell's genome in a manner so as to delete the functional dual selection gene. The non-functional fragment of the dual selection gene (formed in step 4) is not affected by the recombination, and is retained. Importantly, since this step results in the creation of a cell whose genome lacks the dual selection gene, it is possible to select for the desired recombinational event.

As will be appreciated, the net effect of the above-described steps is to produce a cell whose genome has been engineered to contain a "first," "second," and "third" "fragment" of a particular desired gene in a contiguous

manner. The steps may be repeated as desired in order to introduce additional "fragments" into the cell's genome. In this manner, cells can be constructed which contain heterologous genes, chromosome fragments, or chromosomes, that could not be introduced using a single vector. As indicated above, it is possible to select for each step of the method.

VI. The Production of Chimeric and Transgenic Animals

The chimeric or transgenic animal cells of the present invention are prepared by introducing one or more DNA molecules into a precursor pluripotent cell, most preferably an ES cell, or equivalent (Robertson, E.J., In: Current Communications in Molecular Biology, Capecchi, M.R. (ed.), Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), pp. 39-44, which reference is incorporated herein by reference). The term "precursor" is intended to denote only that the pluripotent cell is a precursor to the desired ("transfected") pluripotent cell which is prepared in accordance with the teachings of the present invention. The pluripotent (precursor or transfected) cell may be cultured in vivo, in a manner known in the art (Evans, M.J. et al., Nature 292:154-156 (1981)) to form a chimeric or transgenic animal.

Any ES cell may be used in accordance with the present invention. It is, however, preferred to use primary isolates of ES cells. Such isolates may be obtained directly from embryos such as the CCE cell line disclosed by Robertson, E.J., In: Current Communications in Molecular

GeneDX 1009, pg. 748

Biology, Capecchi, M.R. (ed.), Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), pp. 39-44), or from the clonal isolation of ES cells from the CCE cell line (Schwartzberg, P.A. et al., Science 246:799-803 (1989), which reference is incorporated herein by reference). Such clonal isolation may be accomplished according to the method of E.J. Robertson (In: Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, (E.J. Robertson, Ed.), IRL Press, Oxford, 1987) which reference and method are incorporated herein by reference. The purpose of such clonal propagation is to obtain ES cells which have a greater efficiency for differentiating into an animal. Clonally selected ES cells are approximately 10-fold more effective in producing transgenic animals than the progenitor cell line CCE. For the purposes of the recombination methods of the present invention, clonal selection provides no advantage. An example of ES cell lines which have been clonally derived from embryos are the ES cell lines, AB1 (hprt$^+$) or AB2.1 (hprt$^-$).

The ES cells are preferably cultured on stromal cells (such as STO cells (especially SNC4 STO cells) and/or primary embryonic fibroblast cells) as described by E.J. Robertson (In: Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, (E.J. Robertson, Ed.), IRL Press, Oxford, 1987, pp 71-112), which reference is incorporated herein by reference. Methods for the production and analysis of chimeric mice are disclosed by Bradley, A. (In: Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, (E.J. Robertson, Ed.), IRL Press, Oxford, 1987, pp 113-151), which reference is incorporated herein by

GeneDX 1009, pg. 749

reference. The stromal (and/or fibroblast) cells serve to eliminate the clonal overgrowth of abnormal ES cells. Most preferably, the cells are cultured in the presence of leukocyte inhibitory factor ("lif") (Gough, N.M. et al., Reprod. Fertil. Dev. 1:281-288 (1989); Yamamori, Y. et al., Science 246:1412-1416 (1989), both of which references are incorporated herein by reference). Since the gene encoding lif has been cloned (Gough, N.M. et al., Reprod. Fertil. Dev. 1:281-288 (1989)), it is especially preferred to transform stromal cells with this gene, by means known in the art, and to then culture the ES cells on transformed stromal cells that secrete lif into the culture medium.

ES cell lines may be derived or isolated from any species (for example, chicken, etc.), although cells derived or isolated from mammals such as rodents (i.e. mouse, rat, hamster, etc.), rabbits, sheep, goats, fish, pigs, cattle, primates and humans are preferred.

VII. Uses of the Present Invention

The present invention provides human or animal cells which contain a desired gene sequence in one of the two p53 alleles of the cell's genome.

In a first embodiment, the invention also provides a means for producing non-human chimeric or transgenic animals whose cells contain such a sequence. The animals which may be produced through application of the described method include chicken, non-human mammals (especially, rodents (i.e. mouse, rat, hamster, etc.), rabbits, sheep, goats, fish, pigs, cattle and non-human primates).

GeneDX 1009, pg. 750

The cells and non-human animals of the present invention have both diagnostic and therapeutic utility.

A. Diagnostic Utility

Since the invention provides a cell, or a transgenic or chimeric non-human animal that contains a single functional allele of the p53 gene, and since such cells will become tumor cells upon the mutation of the functional allele to a non-functional form, the present invention can be used to identify an agent that is capable of affecting a characteristic of an animal cell that is attributable to the presence or expression of a tumor-suppressing gene. A characteristic of an animal cell is said to be "attributable to the presence or expression of a tumor-suppressing gene," if the characteristic is altered by the absence or lack of expression of the tumor-suppressing gene. Examples of such characteristics include tumorigenesis, resilience to tumorigenesis, the extent, distribution, incidence, location, grade, etc. of tumors, etc.

In one embodiment, such agents can decrease the tumorigenic (or neoplastic) potential of the cells or animals. Such agents are discussed below with regard to the therapeutic potential of the invention.

In a second embodiment, such agent can increase the tumorigenic (or neoplastic) potential of the cells or animals. Thus, the cells and non-human animals of the present invention have utility in testing potential or suspected carcinogens for tumorigenic activity. They may be used to identify and assess the tumorigenic effects of

GeneDX 1009, pg. 751

agents that may be present, for example, in the environment (such as environmental pollutants in air, water or soil), or resulting from environmental exposures to chemicals, radioisotopes, etc. They may also be used to facilitate studies of the effects of diet on oncogenesis. They may be used to determine whether potential or present food additives, chemical waste products, chemical process by-products, water sources, proposed or presently used pharmaceuticals, cosmetics, etc., have tumorigenic activity. They may also be used to determine the tumorigenic potential of various energy forms (such as UV rays, X-rays, ionizing radiation, gamma rays of elemental isotopes, etc.).

The frequency at which a mutational event occurs is dependent upon the concentration of a mutagenic chemical agent, or the intensity of a mutagenic radiation. Thus, since the frequency of a single cell receiving two mutational events is the square of the frequency at which a single mutational event will occur, the cells and non-human animals of the present invention shall be able to identify neoplastic (mutagenic) agents at concentrations far below those needed to induce neoplastic changes in natural cells or animals.

One especially preferred cell is a non-human cell in which one of the natural p53 alleles has been replaced with a functional human p53 allele and the other of the natural p53 alleles has been mutated to a non-functional form. Alternatively, one may employ a non-human cell in which the two natural p53 alleles have been replaced with a functional and a non-functional allele of the human p53 gene.

GeneDX 1009, pg. 752

Such cells may be used, in accordance with the methods described above to assess the neoplastic potential of agents in cells containing the human p53 allele. More preferably, such cells are used to produce non-human animals which do not contain any natural functional p53 alleles, but which contain only one functional human p53 allele. Such non-human animals can be used to assess the tumorigenicity of an agent in a non-human animal expressing the human p53 gene product.

1. In Vitro Assays

In one embodiment, one may employ the cells of the present invention, in in vitro cell culture, and incubate such cells in the presence of an amount of the agent whose tumorigenic potential is to be measured. This embodiment therefore comprises an in vitro assay of tumorigenic activity.

Although many carcinogenic agents may directly mediate their tumorigenic effects, some agents will not exhibit tumorigenic potential until metabolized, or until presented to a susceptible cell along with one or more "co-carcinogenic" factors. The present invention permits the identification of such "latent" carcinogenic and "co-carcinogenic" agents. In accordance with this embodiment of the invention, the presence of a "latent" carcinogen can be identified by merely maintaining cell or animal exposure to a candidate agent. Alternatively, the cells of the present invention can be incubated in "spent" culture medium (i.e. medium containing the candidate agent that was used to

GeneDX 1009, pg. 753

culture other cells before being used to culture the cells of the present invention).

The present invention permits the identification of co-carcinogenic factors capable of inducing neoplastic effects in the presence of a second agent. Such factors can be identified by culturing the cells of the present invention in the presence of two or more candidate agents simultaneously, and then assaying for neoplasia.

The transformation of the cells to a neoplastic state would be indicative of tumorigenic (or neoplastic) activity of the assayed agent. Such a neoplastic state may be evidenced by a change in cellular morphology, by a loss of contact inhibition, by the acquisition of the capacity to grow in soft agar, or most preferably, by the initiation of expression of tumor antigens.

The use of tumor antigens as a means of detecting neoplastic activity is preferred since such antigens may be readily detected.

As is well known in the art, antibodies, or fragments of antibodies, may be used to quantitatively or qualitatively detect the presence tumor of antigens on cell surfaces. Since any cell type (i.e. lung, kidney, colon, etc.) may be employed to form the p53- mutated cells of the present invention, it is possible to determine whether an agent has a tissue specific tumorigenic potential. To accomplish this goal, one would incubate a candidate agent in the presence of p53-mutated cells derived from any of a variety of tissue types. Since tumors have tumor-specific antigens, and since antibodies capable of binding to such antigens have been isolated, it is possible to use such antibodies to

GeneDX 1009, pg. 754

characterize any tumor antigens which may be expressed by the p53-mutated cells.

Such detection may be accomplished using any of a variety of immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect the antigen through the use of radioimmune assays. A good description of a radioimmune assay (RIA) may be found in Laboratory Techniques and Biochemistry in Molecular Biology, by Work, T.S., et al., North Holland Publishing Company, NY (1978), with particular reference to the chapter entitled "An Introduction to Radioimmune Assay and Related Techniques" by Chard, T., incorporated by reference herein. Examples of suitable radioisotopic labels include ^{3}H, ^{111}In, ^{125}I, ^{131}I, ^{32}P, ^{35}S, ^{14}C, ^{51}Cr, ^{57}To, ^{58}Co, ^{59}Fe, ^{75}Se, ^{152}Eu, ^{90}Y, ^{67}Cu, ^{217}Ci, ^{211}At, ^{212}Pb, ^{47}Sc, ^{109}Pd, etc.

Alternatively, enzyme labels, non-radioactive isotopic labels, fluorescent labels, chemiluminescent labels or other suitable labels can be employed.

Examples of suitable enzyme labels include malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast-alcohol dehydrogenase, alpha-glycerol phosphate dehydrogenase, triose phosphate isomerase, peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase, acetylcholine esterase, etc.

Examples of suitable non-radioactive isotopic labels include ^{157}Gd, ^{55}Mn, ^{162}Dy, ^{52}Tr, ^{56}Fe, etc.

GeneDX 1009, pg. 755

Examples of suitable fluorescent labels include an ^{152}Eu label, a fluorescein label, an isothiocyanate label, a rhodamine label, a phycoerythrin label, a phycocyanin label, an allophycocyanin label, an o-phthaldehyde label, a fluorescamine label, etc.

Examples of chemiluminescent labels include a luminal label, an isoluminal label, an aromatic acridinium ester label, an imidazole label, an acridinium salt label, an oxalate ester label, a luciferin label, a luciferase label, an aequorin label, etc.

Those of ordinary skill in the art will know of other suitable labels which may be employed in accordance with the present invention. The binding of these labels to antibodies or fragments thereof can be accomplished using standard techniques commonly known to those of ordinary skill in the art. Typical techniques are described by Kennedy, J.H., et al. (Clin. Chim. Acta 70:1-31 (1976)), and Schurs, A.H.W.M., et al. (Clin. Chim. Acta 81:1-40 (1977)). Coupling techniques mentioned in the latter are the glutaraldehyde method, the periodate method, the dimaleimide method, the m-maleimidobenzyl-N-hydroxy-succinimide ester method, all of which methods are incorporated by reference herein.

The above-described in vitro assay has the advantageous features of potentially lower cost than presently used assays, and the capacity to readily screen large numbers of agents. Use of this embodiment facilitates comparisons of test results obtained at different times and conditions. Moreover, because it is possible to use very large numbers of cells in such assays, it is possible to detect the

GeneDX 1009, pg. 756

tumorigenic activity of tumorigenic agents even at very low concentrations. Lastly, since this embodiment can be performed using human cells, it provides a means for determining the tumorigenic (or neoplastic) potential of a test compound on human cells.

2. In Vivo Assays

In a second embodiment, one may employ the non-human animals of the present invention, and provide to such animals (by, for example, inhalation, ingestion, injection, implantation, etc.) an amount of the agent whose tumorigenic potential is to be measured. The formation of tumors in such animals (as evidenced by direct visualization by eye, or by biopsy, imaging, detection of tumor antigens, etc.) would be indicative of tumorigenic activity of the assayed agent.

The use of the non-human animals of the present invention is preferred over naturally occurring non-human animals since such natural animals contain two functional p53 alleles, and thus require two mutational events in order to lead to loss of functional p53 activity. In contrast, since the non-human animals of the present invention have only one functional p53 allele, only one mutational event is needed to cause total loss of p53 function.

The detection of tumors in such animals can be accomplished by biopsy, imaging, or by assaying the animals for the presence of cells which express tumor antigens.

For example, such detection may be accomplished by removing a sample of tissue from a subject and then treating

GeneDX 1009, pg. 757

the isolated sample with any suitably labeled antibodies (or antibody fragments) as discussed above. Preferably, such in situ detection is accomplished by removing a histological specimen from the subject, and providing the labeled antibody to such specimen. The antibody (or fragment) is preferably provided by applying or by overlaying the labeled antibody (or fragment) to a sample of tissue. Through the use of such a procedure, it is possible to determine not only the presence of antigen, but also the distribution of the antigen on the examined tissue. Using the present invention, those of ordinary skill will readily perceive that any of a wide variety of histological methods (such as staining procedures) can be modified in order to achieve such in situ detection.

Alternatively, the detection of tumor cells may be accomplished by in vivo imaging techniques, in which the labeled antibodies (or fragments thereof) are provided to the subject, and the presence of the tumor is detected without the prior removal of any tissue sample. Such in vivo detection procedures have the advantage of being less invasive than other detection methods, and are, moreover, capable of detecting the presence of antigen-expressing cells in tissue which cannot be easily removed from the patient.

Additionally, it is possible to assay for the presence of tumor antigens in body fluids (such as blood, lymph, etc.), stools, or cellular extracts. In such immunoassays, the antibodies (or antibody fragments) may be utilized in liquid phase or bound to a solid-phase carrier, as described below.

GeneDX 1009, pg. 758

The use of an in vivo assay has several advantageous features. The in vivo assay permits one not only to identify tumorigenic agents, but also to assess the kind(s) of tumors induced by the agent, the number and location (i.e. whether organ or tissue specific) of any elicited tumors, and the grade (clinical significance) of such elicited tumors. It further permits an assessment of tumorigenicity which inherently considers the possible natural metabolism of the introduced agent, the possibility that the introduced agent (or its metabolic by-products) might selectively accumulate in specific tissues or organs of the recipient animal, the possibility that the recipient animal might recognize and repair or prevent tumor formation. In short, such an assay provides a true biological model for studying and evaluating the tumorigenic potential of an agent in a living non-human animal.

3. Immunoassays of Tumor Antigens

The in vitro, in situ, or in vivo detection of tumor antigens using antibodies (or fragments of antibodies) can be improved through the use of carriers. Well-known carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, agaroses, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen. Thus, the support configuration may

GeneDX 1009, pg. 759

be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Those skilled in the art will note many other suitable carriers for binding monoclonal antibody, or will be able to ascertain the same by use of routine experimentation.

The binding molecules of the present invention may also be adapted for utilization in an immunometric assay, also known as a "two-site" or "sandwich" assay. In a typical immunometric assay, a quantity of unlabeled antibody (or fragment of antibody) is bound to a solid support that is insoluble in the fluid being tested (i.e., blood, lymph, liquified stools, tissue homogenate, etc.) and a quantity of detectably labeled soluble antibody is added to permit detection and/or quantitation of the ternary complex formed between solid-phase antibody, antigen, and labeled antibody.

Typical immunometric assays include "forward" assays in which the antibody bound to the solid phase is first contacted with the sample being tested to extract the antigen from the sample by formation of a binary solid phase antibody-antigen complex. After a suitable incubation period, the solid support is washed to remove the residue of the fluid sample, including unreacted antigen, if any, and then contacted with the solution containing an unknown quantity of labeled antibody (which functions as a "reporter molecule"). After a second incubation period to permit the labeled antibody to complex with the antigen bound to the solid support through the unlabeled antibody, the solid support is washed a second time to remove the unreacted

GeneDX 1009, pg. 760

labeled antibody. This type of forward sandwich assay may be a simple "yes/no" assay to determine whether antigen is present or may be made quantitative by comparing the measure of labeled antibody with that obtained for a standard sample containing known quantities of antigen. Such "two-site" or "sandwich" assays are described by Wide at pages 199-206 of Radioimmune Assay Method, edited by Kirkham and Hunter, E. & S. Livingstone, Edinburgh, 1970.

In another type of "sandwich" assay, which may also be useful to identify tumor antigens, the so-called "simultaneous" and "reverse" assays are used. A simultaneous assay involves a single incubation step as the antibody bound to the solid support and labeled antibody are both added to the sample being tested at the same time. After the incubation is completed, the solid support is washed to remove the residue of fluid sample and uncomplexed labeled antibody. The presence of labeled antibody associated with the solid support is then determined as it would be in a conventional "forward" sandwich assay.

In the "reverse" assay, stepwise addition first of a solution of labeled antibody to the fluid sample followed by the addition of unlabeled antibody bound to a solid support after a suitable incubation period is utilized. After a second incubation, the solid phase is washed in conventional fashion to free it of the residue of the sample being tested and the solution of unreacted labeled antibody. The determination of labeled antibody associated with a solid support is then determined as in the "simultaneous" and "forward" assays.

GeneDX 1009, pg. 761

As explained above, the immunometric assays for antigen require that the particular binding molecule be labeled with a "reporter molecule." These reporter molecules or labels, as identified above, are conventional and well-known to the art. In the practice of the present invention, enzyme labels are a preferred embodiment. No single enzyme is ideal for use as a label in every conceivable immunometric assay. Instead, one must determine which enzyme is suitable for a particular assay system. Criteria important for the choice of enzymes are turnover number of the pure enzyme (the number of substrate molecules converted to product per enzyme site per unit of time), purity of the enzyme preparation, sensitivity of detection of its product, ease and speed of detection of the enzyme reaction, absence of interfering factors or of enzyme-like activity in the test fluid, stability of the enzyme and its conjugate, availability and cost of the enzyme and its conjugate, and the like. Included among the enzymes used as preferred labels in the immunometric assays of the present invention are peroxidase, alkaline phosphatase, beta-galactosidase, urease, glucose oxidase, glycoamylase, malate dehydrogenase, and glucose-6-phosphate dehydrogenase. Urease is among the more preferred enzyme labels, particularly because of chromogenic pH indicators which make its activity readily visible to the naked eye.

B. Therapeutic Utility

Significantly, the cells and animals of the present invention can be used to identify agents that decrease the

GeneDX 1009, pg. 762

tumorigenic (or neoplastic) potential of the cells or animals. Such agents can be "anti-tumor agents" and/or "chemopreventative agents." "Anti-tumor agents" act to decrease the proliferation of the cells (or the growth, dissemination, or metastasis of tumors in the chimeric or transgenic animals). "Chemopreventative agents" act to inhibit the formation of new tumors. Such agents may have general activity (inhibiting all new tumor formation), or may have a specific activity inhibiting the distribution, frequency, grade, etc.) of specific types of tumors in specific organs and tissue. Thus, the present invention permits the identification of novel antineoplastic therapeutics. Any of the above assays of tumor-suppressing activity may be used for this purpose.

The transgenic cells and non-human animals of the present invention can be used to study human gene regulation of the p53 gene. For example, such cells and animals can be used to investigate the interactions of the p53 gene with oncogenes or other tumor suppressor genes. Thus, they may be used to identify therapeutic agents which have the ability to impair or prevent neoplastic or tumorigenic development. Such agents have utility in the treatment and cure of cancer in humans and animals.

Significantly, potential therapeutic agents are frequently found to induce toxic effects in one animal model but not in another animal model. To resolve the potential of such agents, it is often necessary to determine the metabolic patterns in various species, and to then determine the toxicities of the metabolites. The present invention

GeneDX 1009, pg. 763

permits one to produce transgenic cells or animals which could facilitate such determinations.

When providing the therapeutic agents of the present invention to the cells of an animal, pharmaceutically acceptable carriers (i.e. liposomes, etc.) are preferably employed. Such agents can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, are described, for example, in Nicolau, C. et al. (Crit. Rev. Ther. Drug Carrier Syst. 6:239-271 (1989)), which reference is incorporated herein by reference.

In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of the desired gene sequence together with a suitable amount of carrier vehicle.

Additional pharmaceutical methods may be employed to control the duration of action. Control release preparations may be achieved through the use of polymers to complex or absorb the desired gene sequence (either with or without any associated carrier). The controlled delivery may be exercised by selecting appropriate macromolecules (for example polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine, sulfate) and the concentration of macromolecules as well as the methods of incorporation in order to control release. Another possible method to control the duration of action by controlled

GeneDX 1009, pg. 764

release preparations is to incorporate the agent into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatine-microcapsules and poly(methylmethacylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions.

C. Use in Research and in Gene Therapy

The cells and non-human animals of the present invention, quite apart from their uses in veterinary and human medicine, may be used to investigate gene regulation, expression and organization in animals. The methods of the present invention may be used to produce alterations in a regulatory region for the native p53 gene sequence. Thus, the invention provides a means for altering the nature or control of transcription or translation of the p53 gene, and of altering the p53 gene itself. For example, the invention enables one to introduce mutations which result in increased or decreased gene expression. Similarly, it enables one to impair or enhance the transcriptional capacity of the natural p53 gene in order to decrease or increase its

GeneDX 1009, pg. 765

expression. Thus, the present invention permits the manipulation and dissection of the p53 gene.

Such abilities are especially valuable in gene therapy protocols, and in the development of improved animal models of cancer.

In one embodiment of the present invention, DNA encoding either a functional p53 gene, variants of that gene, or other genes which influence the activity of the p53 gene, may be introduced into the somatic cells of an animal (particularly mammals including humans) in order to provide a treatment for cancer (i.e. "gene therapy"). Most preferably, viral or retroviral vectors are employed for this purpose.

The principles of gene therapy are disclosed by Oldham, R.K. (In: Principles of Biotherapy, Raven Press, NY, 1987), and similar texts. Disclosures of the methods and uses for gene therapy are provided by Boggs, S.S. (Int. J. Cell Clon. 8:80-96 (1990)); Karson, E.M. (Biol. Reprod. 42:39-49 (1990)); Ledley, F.D., In: Biotechnology, A Comprehensive Treatise, volume 7B, Gene Technology, VCH Publishers, Inc. NY, pp 399-458 (1989)); all of which references are incorporated herein by reference.

Although, as indicated above, such gene therapy can be provided to a recipient in order to treat (i.e. suppress, attenuate, or cause regression) an existing neoplastic state, the principles of the present invention can be used to provide a prophylactic gene therapy to individuals who, due to inherited genetic mutations, or somatic cell mutation, contain cells having impaired p53 gene expression (for example, only a single functional allele of the p53

GeneDX 1009, pg. 766

gene). Such therapy would be administered in advance of the detection of cancer in order to lessen the individual's predisposition to the disease.

Having now generally described the invention, the same will be more readily understood through reference to the following examples which are provided by way of illustration, and are not intended to be limiting of the present invention, unless specified.

EXAMPLE 1

INACTIVATION OF THE p53 TUMOR SUPPRESSOR GENE IN THE MOUSE: OVERVIEW OF THE EXPERIMENT

Structural alterations of the p53 tumor suppressor gene have been associated with a wide array of human cancers. To examine the role of p53 in tumorigenesis and mammalian development, embryonic stems (ES) cell lines were generated in which one of the two endogenous p53 alleles has been inactivated by the insertion of a neo^R gene (eg. the nptII gene of tn5) following homologous recombination.

The gene targeting strategy utilized a murine p53 3.7 kb genomic construct flanked by an HSV TK gene and interrupted in exon 5 by a polyA - neo^R gene driven by a pol II promoter. The 3.7 kb fragment is derived from a genomic plasmid clone of murine p53 (Oren, M. et al., EMBO J. 2:1633-1639 (1983); Pinhasi et al. Molec. Cell. Biol. 4:2180-2186 (1984)). The DNA was cloned from a liver cell of a normal Balb/c mouse. Following gene transfer into ES cells and G418/FIAU selection (as described above), two clones were identified by PCR and Southern analysis which have the expected altered

GeneDX 1009, pg. 767

p53 gene structure (Figure 3). Three probes were used for this purpose. The first probe (neo probe) indicates the presence of the neomycin resistance determinant in the stem cells. The second and third probes (exon 1-2 and exon 5-10 probes) reveal that the wild type and mutant alleles are present in the two embryonic stem cells, but that the mutant allele is not present in the wild type (wt) cell. The fact that the three probes identify the same bands in the mutant embryonic stem cells indicates that, as expected, the neo determinant in these cells is linked to the p53 gene.

One of the heterozygous ES clones was injected into C57BL/6 recipient blastocysts and implanted into pseudopregnant C57BL/6 female mice. This ES clone was designated ESΔp53.

Embryonic stem cell line ESΔp53, (containing the p53 targeting construct of Figure 2, as verified in Figure 3) was deposited with the American Type Culture Collection, Rockville, MD. on December 27, 1990 and was accorded the ATCC accession number: CRL 10631.

The offspring resulting from the injection of the above-described ES cell into a blastocyst, and the implantation of the blastocyst into a mouse uterus, were highly chimeric. As discussed below, the chimeric males have been bred, and germ line heterozygotes have been generated. These heterozygotes are then examined for increased tumor susceptibility and bred to determine the effects of a p53 nullizygous state on embryonic development.

GeneDX 1009, pg. 768

EXAMPLE 2
THE GENERATION OF THE RECOMBINOGENIC p53 CONSTRUCTS

To maximize an ability to obtain and select ES cells containing homologous recombination events, the strategy of Mansour et al. was utilized (Mansour, S.L. et al., Nature 336:348-352 (1988), herein incorporated by reference). The procedure used a positive selection method for isolating cells that had stably integrated the introduced mutating DNA and a negative selection against cells that did not contain a homologous recombination event, allowing an enrichment for cells containing the desired homologous recombination event. The adaptation of these procedures for obtaining homologous recombination into the p53 gene is outlined in Figure 1.

As indicated above, a segment of the genomic sequences from the mouse p53 gene (a 3.7-kb fragment spanning exons 2 to 10 of the 11 exon gene) was obtained from plasmid pSVpcG3 (Pinhasi et al. Molec. Cell. Biol. 4:2180-2186 (1984)). As shown in Figure 1, this sequence of DNA was modified in two ways. First, a neo marker gene driven by the MC1 promoter enhancer was designed to obtain high levels of expression in ES cells (Thomas, K.R. et al., Cell 51:503-512 (1988)). This sequence was inserted into the unique Bal 1 site in exon 5. This insertion provided a positive selectable marker (neo) for stable gene transfer into ES cells and disrupted the coding sequence of p53, producing an inactive allele following successful homologous recombination. The second alteration entailed the attachment of a herpes simplex virus thymidine kinase gene (HSV TK) to the 3' end of the gene-targeting construct (Figure 2). This attachment

GeneDX 1009, pg. 769

provided the negative selection (using the HSV-TK-specific thymidine analogue FIAU (1-(2 deoxy, 2 fluoro, β-D arabinofuranosyl)-5-iodouracil) against cells that have random integrations of the targeting construct (Figure 2).

EXAMPLE 3

TRANSFER OF CONSTRUCTS INTO ES CELLS AND IDENTIFICATION OF CELLS CONTAINING HOMOLOGOUS RECOMBINATION EVENTS

After generating the above-described p53 targeting construct (figure 2), the construct was introduced into cultured ES cells by electroporation. ES cells were cultured as described by E.J. Robertson (In: Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, (E.J. Robertson, Ed.), IRL Press, Oxford, 1987, pp 71-112). Cells for electroporation were collected by trypsinization at 60-80% confluency, sedimented and resuspended in buffered saline with DNA at 25 U/ml. 10^7 cells/ml are treated with a single pulse from a Bio-Rad Gene Pulser (.240 volts, 500 uF, 0.4 cm cuvette) in order to achieve electroporation. Under these conditions, efficiencies of stable gene transfer of approximately 10^{-3} were obtained. After pulsing, the ES cells were plated onto feeder cells (as described above) at 5 X 106 cells/6-cm plate for G418/gancyclovir selection. Selection was allowed to proceed for 10-12 days (until control plate showed no colonies).

G418/FIAU-resistant companies were isolated, amplified, and genomic DNA purified from about 50-100 individual isolates. These DNAs were restricted with Hind III, Eco RI, and Pvu II. Colonies were separated into two halves; one

GeneDX 1009, pg. 770

half of each colony was lysed with detergent and subjected to PCR with oligonucleotide probes derived from the neo gene and p53 gene sequences outside the borders of the targeting construct. If the insertion of the construct had occurred through homologous recombination, then a 1.2 kb fragment would be generated. Two of 100 colonies tested gave the expected 1.2 kb PCR band. The DNA from these positive colonies were then subjected to agarose gel electrophoresis, Southern blotting, and hybridization to three probes. The probes used were a p53 exon 4 probe (130 bp), an exon 11 probe (200 bp), and the neo probe (1 kb). Each of these probes hybridized to Southern fragments diagnostic for a homologous recombination event. Taken together, the Southern hybridization conclusively identified colonies with p53 gene disruptions generated by homologous recombination.

Initial constructs utilized an MC1-neo gene; these constructs did not work well. The successful constructs were a - Pol II - neo - polyA$^-$ - gene construct.

EXAMPLE 4

CONSTRUCTION OF MOUSE CHIMERAS FROM SPECIFIC ES CELL CLONES

Chimeras were constructed from suitable clones as described by Bradley, A. (In: Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, (E.J. Robertson, Ed.), IRL Press, Oxford, 1987, pp 113-151). Briefly, 3.5-day-old blastocyst-stage embryos were collected from the dissected uterine tracts of C57BL females 3 days after plugging. 12-

GeneDX 1009, pg. 771

15 individual cells were microinjected into the cavity of the blastocyst-stage embryos and, after a brief culture period, transferred back into the uterine horns of pseudopregnant F_1(CBAxC57BL) foster mothers 2 days after mating with a vasectomized male.

The methods for introducing the ES cell into the blastocyst, and for producing offspring have been described above, and comprise techniques which are well-known to those of ordinary skill (Mansour, S.L. et al., Nature 336:348-352 (1988); Capecchi, M.R., Trends Genet. 5:70-76 (1989); Capecchi, M.R., Science 244:1288-1292 (1989); Capecchi, M.R. et al., In: Current Communications in Molecular Biology, Capecchi, M.R. (ed.), Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), pp. 45-52; Frohman, M. A. et al., Cell 56:145-147 (1989); all of which references have been incorporated herein by reference).

In order to generate an adequate number of chimeras of high quality (i.e., high contribution of donor cells) the blastocyst injection experiments were repeated over a period of a few weeks. Approximately 53 blastocyst injections were performed.

Following embryo transfer, the young are born 17 days later. The levels of chimerism could not be assessed at this time because the cell line used carries the Black Agouti coat color markers and the embryos are Black non-Agouti. The Agouti marker becomes visible in the coat at day 8. At this stage the chimeras were scored and non-chimeric animals discarded. The presence of the Y chromosome in the donor cell line ensured that the majority of any germ line contribution was through the germ line of

GeneDX 1009, pg. 772

male chimeras. In addition, dominant effects of the Y chromosome distort the sex ratio among the chimeric offspring in favor of males by the conversion of some female embryos to phenotypic males. The sex ratio consequently is an effective early measure of the levels of chimerism in the experimental population, and this tends to reflect germ line chimerism.

The mice ranged from total black (i.e. no chimerism) to greater than 90% agouti (i.e. very high chimerism). When the above-described stromal feeder cell procedures were employed, greater than 50% of the injected pups showed high degrees of chimerism (as determined by the ratio of agouti coat color to black coat color in the offspring).

EXAMPLE 5

TEST BREEDING OF CHIMERAS TO ASSAY FOR GERM LINE CONTRIBUTIONS AND IDENTIFICATION OF HETEROZYGOTES

Of the 53 injected blastocysts, approximately 12 male mice were obtained which exhibited high to very high chimerism (i.e. levels of contributions of agouti hair in their coat that exceeded 50%). These mice were subjected to inbreeding and further analysis as described below.

The 12 male chimeric mice were test bred when they reached sexual maturity (about 8 weeks of age). Test breeding proceeded by caging a single germ line chimera with two virgin C57BL females, and permitting the animals to naturally breed. Successive litters were scored after 8 days of age. The presence of the dominant Agouti coat color among the litter demonstrated the successful colonization of

GeneDX 1009, pg. 773

the germ line. This approach assumed, that, as in the case of the rb gene, a single normal germ line allele would be able to supply normal p53 function and that gene dosage effects would not be important.

Unexpectedly, this assumption was not confirmed. Of the 12 chimeric mice, only four mice (#96, #101, #102 and #103) were found to produce progeny mice. The great majority of these mouse pups (approximately 95-97%) died at about 24 hours post partum. Nevertheless, after several litters, five viable pups, designated p1-p5 were obtained from parents #101 and #102. These parental mice (#101 and #102) were found to be equivalent for the purposes of the invention. The five mice pups were grown to adolescence, and are fully viable.

The 5 mice pups were tested by Southern blot hybridization of tail DNA to determine whether they contained the mutated p53 allele. Thus, when the pups reached six to eight weeks of age, approximately one inch of tail was snipped off and used to prepare high molecular weight DNA according to the following protocol for the preparation of high molecular weight DNA from mouse embryo and/or yolk sacs:

In the procedure, it is desirable to reserve materials and solutions (such as Eppendorf tubes, proteinase K, distilled water (dH_2O), phenol, phenol/chloroform, chloroform, isopropanol, TE, etc.) for genomic use, and to then use such materials only when preparing samples of genomic DNA.

GeneDX 1009, pg. 774

Steps of the Protocol:

1. The embryos of the required gestation are carefully isolated from maternal tissue. The embryos and/or yolk sacs are then placed into Eppendorf tubes containing 250 µl of Cutting Buffer (50 mM Tris pH 7.5, 50 mM EDTA pH 8.0, 100 mM NaCl, 5 mM DTT, 0.5 mM Spermidine). The embryos can either be processed singly or fixed for staining, using their respective yolk sacs for the isolation of DNA.

2. The embryonic tissue is finely minced with a scissors. It may, alternatively, be broken up by passing it twice through a 26G 5/8" Sub-Q needle attached to a 1cc syringe. Yolk sac is very soft tissue that is easily broken up during lysis; therefore, no further manipulation is necessary.

3. 250 µl of Lysis Buffer (50 mM Tris pH 7.5, 50 mM EDTA pH 8.0, 100 mM NaCl, 5 mM DTT, 0.5 mM Spermidine, 2% SDS) and 10 µl of freshly prepared 10 mg/ml Proteinase K is added to each tube.

4. The tubes are gently rocked overnight at 55°C.

5. 500 µl of phenol is added to each tube, and the tubes are rotated at 5-10 rpm for 15-20 minutes. The tubes are then centrifuged in a microcentrifuge at maximum speed for 10 minutes. The DNA is then transferred to clean Eppendorf tubes using pipette tips from which the

tips have been snipped off. [When pipetting genomic DNA samples, it is preferable to use pipette tips from which the tips have been snipped off. This will prevent shearing the DNA and keep the average size in excess of 80-100 kb.]

6. 500 µl of phenol/chloroform is added to each tube, and step 5 is repeated.

7. 500 µl of chloroform is added to each tube, and step 5 is repeated. When transferring the DNA to clean tubes, the volume transferred is to be recorded.

8. A volume of isopropanol equal to the volume of DNA is added to each tube. The tubes are inverted a few times to precipitate the DNA.

9. The tubes are centrifuged in a microcentrifuge at maximum speed for 10 minutes.

10. The supernatant is carefully aspirated off. The pelleted DNA is rinsed in 1 ml of 70% of ethanol (EtOH). The tubes are centrifuged again in the microcentrifuge at maximum speed for 10 minutes.

11. The EtOH is poured off and allow the pellets are allowed to air dry until there is no visible sign of remaining EtOH. The DNA is resuspended in 15 µl of TE buffer (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0).

GeneDX 1009, pg. 776

12. Enough DNA can be prepared for at least one lane of a 20 cm x 15 cm agarose gel with 30 sample wells; usually, a couple of hundred micrograms of DNA is obtained. The DNA is digested by adding 2 μl of the appropriate 10X buffer, 2 μl of dH_2O, and 1 μl of the appropriate restriction enzyme (i.e., such that the reaction is in 20 μl total volume).

13. Since genomic DNA is difficult to load in this size of well (in which it has to displace the electrophoresis buffer), the gel is loaded at the bench before it is submerged in the running buffer.

Once the DNAs were obtained, their concentrations were determined by A_{260} determination on a UV spectrophotometer. 5-10 μg of the DNAs were then cleaved with either BamHI or EcoR1 restriction endonucleases for several hours at 37°C. The cut DNAs were loaded on a 0.8% agarose gel and electrophoresed at 100V for approximately 3 hours on a horizontal submarine gel apparatus. The DNAs were then blotted to Zetaprobe (BioRad) nylon membranes and hybridized in 20% dextran sulfate-3XSSPE, both steps according to the protocol of Reed and Mann (Reed, K.C. et al., Nucleic Acids Res. 13:7207-7221 (1985)). The probe for hybridization was ^{32}P-labelled according to the protocol of Feinberg and Vogelstein (Feinberg, A. et al., Anal. Biochem. 132:6-13 (1983)). Hybridization was permitted to continue overnight at 68°C and the filters were washed according to the protocol of Reed and Mann (Reed, K.C. et al., Nucleic Acids Res. 13:7207-7221 (1985))). The probe for Figure 4 was a

GeneDX 1009, pg. 777

pol II-neo probe. The Pol II part of the probe hybridizes to a 4.0 kb fragment of the Pol II gene which is present in all mice. The Neo part of the probe hybridizes to a larger 6.7 kb fragment present only in the p53 targeted germ line heterozygote mice.

Two of the five mice tested, pups p1 and p3, were found to contain the mutated p53 allele in its germ line (Figure 4). Figure 4 shows a Southern blot analysis of tail DNA of the progeny of the F_1[129xC57BL/6] chimeric mice. The DNA was restricted with BamHI and hybridized to the above-described ^{32}P-labelled Pol II-Neo probe. The band between the offspring and the controls is a marker which happened to hybridize to the probe. The analysis thus revealed that mouse pups p1 and p3 contain DNA which hybridizes to the neo gene, and thus contain the mutated p53 construct.

Following the demonstration of germ line chimericism, the initial F_1(C57BLx129) males were crossed with strain 129 females to generate a larger population of heterozygote mice. Germ line transmission of the mutated p53 allele to offspring may be confirmed by restriction mapping analysis.

Thus, in summary, 4 of the 12 mice exhibiting very high chimerism (#96, #101, #102 and #103) were found to be good breeders, when permitted to breed naturally. Of these four mice, mice #101 and #102 were found to produce viable offspring having the agouti phenotype. Thus, these chimeric mice have been demonstrated to be germ line chimeric animals (i.e. they contain germ cells which possess the desired p53 construct, and are capable to transmitting this construct to progeny animals). Indeed, progeny animals of mice #101 and #102 were obtained, through routine breeding. Two of these

GeneDX 1009, pg. 778

progeny animals (p1 and p3) contained the mutated p53 allele in its germ line, and were thus capable of transferring it indefinitely to future generations by further routine breeding.

EXAMPLE 6

PRODUCTION OF TRANSGENIC ANIMALS HAVING LI-FRAUMENI AND OTHER p53 MUTANT ALLELES

The above-described non-human transgenic animal contains a mutation in the p53 gene. The same methods can be used to produce animals having other mutations in this gene (Figure 5). Thus, a targeting construct is prepared which contains a desired mutation (deletion, insertion or point mutation). The construct is introduced into embryonic stem cells, and permitted to undergo homologous recombination with a p53 allele of that cell. This process leads to the integration of the construct into the p53 locus of the stem cell, and results in a duplication of the p53 gene at that locus. The duplication is resolved, via intrachromosomal recombination to yield an embryonic stem cell having the desired mutated p53 allele (Figure 5).

EXAMPLE 7

ASSAY OF TUMOR FREQUENCY IN THE HETEROZYGOUS ANIMALS

A sizable group (n=50-100) of F_1 mice with a single disrupted p53 allele are observed over a period of 12-24 months for signs of spontaneous tumors, abnormal symptoms, or early deaths. Careful records of deaths and tumor

GeneDX 1009, pg. 779

incidence are kept and compared to a control population of F_1 mice that are homozygous for the normal p53 allele. Mice are sacrificed at 3, 6, 12, and 18 months of age and autopsied for any unusual histopathology. Mice exhibiting distress, disease symptoms, or evidence of tumors are sacrificed and examined to establish the exact nature of disease.

Groups of mice are treated with a known carcinogen, such as DMBA, following standard procedures (Butel, J.S. et al., J. Virol. 38:571-580 (1981); Knepper, J.E. et al., Intl. J. Canc. 40:414-422 (1987), both of which references are herein incorporated by reference) and observed for tumor development.

The transgenic mice are infected with Friend murine leukemia virus (F-MuLv) and assayed for time of onset and severity of erythroleukemia. F-MuLv integration into one or both alleles of the p53 gene is associated with erythroleukemia in virus-infected mice (Mowat et al., Nature 314:633-636 (1985); Chow et al., J. Virol. 61:2777-2781 (1987); Hicks, G.G. et al., J. Virol. 62:4752-4755 (1988)).

When differences in tumor incidence between normal mice and heterozygotes are observed, the data is analyzed by standard statistical procedures to determine the significance of the results. Significant increases in tumor incidence in the p53 heterozygote mice is preferably established using larger scale confirmatory studies and carcinogen sensitivity studies documenting the utility of the mice as research and carcinogen screening tools.

The transgenic mice are inter-bred, by conventional, non-recombinant methods, in order to determine whether p53

GeneDX 1009, pg. 780

is a recessive lethal gene and, if so, to determine what stage of embryogenesis is affected. If development proceeds significantly, the organs and issues most affected are identified. Tumors derived from p53 heterozygotes are screened to determine the range of mutations in the remaining "normal" p53 allele. Embryo fibroblast cultures are derived from the altered animals to determine which parameters of transformation are affected under normal culture conditions and following transfection with oncogenes or treatment with carcinogens.

The p53 heterozygotes are interbred to obtain transgenic animals in which both of the animal's p53 alleles contain the mutated construct.

The p53 heterozygotes are bred to other transgenic mice containing activated dominant oncogenes (e.g., the transgenic mouse of Leder, P. et al. (U.S. Patent 4,736,866)) to determine whether offspring containing both mutant genes have an accelerated rate of tumor formation.

GeneDX 1009, pg. 781

EXAMPLE 8
CHARACTERIZATION OF THE p53-DEFICIENT MICE

The above-described p53 heterozygous mice were obtained and interbred to produce progeny. Among the progeny identified were p53 homozygous animals. Thus, unexpectedly, the present invention permitted the isolation and propagation of both p53 heterozygous mice and p53 homozygous mice.

A. p53-Homozygotes Have No Intact p53 RNA or Protein

To conclusively show that the gene targeting of the p53 allele in the mouse germ line generated a null mutation, p53 RNA and protein were analyzed from homozygote tissues and cells. The rationale was that no intact p53 RNA or protein should be present if the disrupted p53 gene was completely inactivated. To assess p53 RNA levels in the homozygotes, RNA-PCR procedures were used because tissue p53 RNA levels are often so low that Northern analyses are inadequate for detection. Five sets of primers (Figure 6A) were used to amplify cDNA synthesized from total RNA purified from spleen, kidney, and testes of wild type, heterozygote, and homozygote mice. The results of these RNA-PCR assays using total spleen RNA (Figure 6B) clearly indicated that an intact p53 mRNA was synthesized only in the wild type and heterozygote mice. Note that for the primers which amplify a fragment spanning the exon 5 neo insertion site there is no apparent PCR fragment in the homozygotes, indicating the failure of the homozygotes to produce an intact message.

GeneDX 1009, pg. 782

However, exon 7 and 10 primers generate a strong fragment from the homozygotes cDNA. The data supports the conclusion that the RNA generated 3' of the neo insertion site is part of a hybrid neo p53 message initiated by the pol II promoter. Exon 1 and 4 primers amplified a very faint PCR band from the homozygote cDNAs relative to the wild type cDNAs. The low amounts of this band indicates that transcription from the endogenous p53 promoter in the homozygotes appears to result in an unstable truncated transcript as a result of the neo cassette insertion.

To demonstrate that a normal p53 protein is not expressed in the homozygotes, an immunoprecipitation assay was performed on proteins derived from tertiary embryo fibroblasts obtained from normal, heterozygote, and homozygote embryos. Using a polyclonal p53 antiserum, the wild type and heterozygote cells were found to contain a precipitable p53, while the homozygote cells did not.

In order to further investigate whether any intact or truncated p53 molecule was expressed in the homozygotes, an immunoblot analysis was performed using a control antibody (E1a-specific antibody M73 (Harlow, E. et al., J. Virol. 55:533-546 (1985)) and three different p53-specific monoclonal antibodies selected from a set of six such antibodies: PAb 242 (Yewdell, J.W. et al., J. Virol. 59:444-452 (1986)); PAb 246 (Yewdell, J.W. et al., J. Virol. 59:444-452 (1986)); PAb 248 (Yewdell, J.W. et al., J. Virol. 59:444-452 (1986)); PAb 421 (Harlow, E. et al., J. Virol. 39:861-869 (1981)); RA3.2C2 (Coffman, R.L. et al., J. Exper. Med. 153:269-279 (1981)); and 200.47 (Dippold et al., Proc. Natl. Acad. Sci. (U.S.A.) 78:1695-1699 (1981)). The

GeneDX 1009, pg. 783

epitopes identified by these antibodies have been mapped (Gannon, J.V. et al., EMBO J. 9:1595-1602 (1990); Wade-Evans, A. et al., EMBO J. 4:699-706 (1985); Mole, S.E. et al., Nucl. Acids Res. 17:3319 (1989); Yewdell, J.W. et al., J. Virol. 59:444-452 (1986)). Cell lysates were initially immunoprecipated with a collection of either (1) PAb 242, PAb 246, and PAb 248, mapping to the p53 epitope located at amino acids 14-25, 88-109, and 14-69, respectively, or (2) PAb 421, RA3.2C2, and 200.47, mapping to the p53 epitope located at amino acids 370-378, 41-69, and 41-111, respectively. The immunoprecipitates were then subjected to SDS polyacrylamide gel electrophoresis, followed by transfer of the proteins to nylon. These blots were then probed with all six of the above-mentioned anti-p53 monoclonal antibodies. The results clearly indicated that homozygotes did not contain any intact or truncated p53 detectable by this assay.

In conclusion, the RNA and protein data indicate that neither intact or truncated p53 is produced in the homozygotes and that the targeted p53 mutation indeed represents a null mutation.

B. p53-Homozygote Mice Are Susceptible to Tumors

Although the p53-negative homozygous mice appeared developmentally normal, they were clearly susceptible to spontaneous tumor formation (naturally occurring tumors are rare in mice less than 6 months of age (Madison, R.M et al., J. Nat'l. Canc. Inst. 40:683-685 (1968); Maita, K. et al., Toxicol. Pathol. 16:340-349 (1988); Squire, R.A. et al., In:

GeneDX 1009, pg. 784

Pathology of Laboratory Animals, Volume II, (Eds., Benirshke, K. et al.), 1054-1055 (Springer-Verlag, NY, 1978)).

The chimeric mice and the germ line p53 heterozygotes and homozygotes were monitored daily for tumors or other abnormalities. The incidence of tumors was noted when the chimeric mice were 14 months old, the oldest heterozygotes were over nine months old and the oldest homozygotes were about six months old (Table 1, age is in weeks (wks) or months (mons)). At that time, no tumors were observed in the normal control mice, a variety of tumors were detected in the homozygote mice (Table 1; cases # 1-20), two tumors were detected in the heterozygote mice (Table 1, cases # 21 and # 22), and three tumors were detected in the chimeric mice (Table 1, cases # 23 and # 24; in case # 23, the phenotypically male mouse had a choriocarcinoma surrounded by recognizable ovarian tissue).

GeneDX 1009, pg. 785

TABLE 1: TUMORS OBSERVED IN p53-DEFICIENT MICE

#	Sex	Age	Histologic type	Anatomic site
1	M	18 wks	undifferentiated sarcoma	subcutis
2	M	8 wks	gonadoblastoma malignant lymphoma	testes thymus
3	M	18 wks	hemangiosarcoma (2)	subcutis/liver
4	F	9 wks	malignant lymphoma	generalized
5	F	15 wks	malignant lymphoma	generalized
6	M	24 wks	undifferentiated sarcoma hemangiosarcoma	vertebra subcutis
7	M	24 wks	hemangiosarcoma malignant lymphoma	subcutis/muscle generalized
8	M	15 wks	hemangiosarcoma	subcutis
9	M	24 wks	hemangiosarcoma malignant lymphoma	subcutis generalized
10	M	16 wks	osteosarcoma	pelvis
11	F	15 wks	malignant lymphoma hemangioma	generalized heart
12	F	18 wks	malignant lymphoma	thymus
13	M	14 wks	malignant lymphoma	generalized
14	F	14 wks	malignant lymphoma	generalized
15	F	15 wks	mammary adenocarcinoma	mammary gland
16	M	16 wks	malignant lymphoma hemangiosarcoma hemangioma	generalized subcutis muscle
17	M	17 wks	malignant lymphoma	generalized
18	M	17 wks	malignant lymphoma hemangiosarcoma (2)	generalized heart/perirenal fat
19	M	21 wks	malignant lymphoma	thymus
20	M	22 wks	malignant lymphoma	generalized
21	M	24 wks	embryonal carcinoma	testis
22	M	37 wks	malignant lymphoma	generalized

GeneDX 1009, pg. 786

TABLE 1: TUMORS OBSERVED IN p53-DEFICIENT MICE

#	Sex	Age	Histologic type	Anatomic site
23	M	14 mons	osteosarcoma choriocarcinoma	sacrum ovary*
24	M	14 mons	Leydig cell tumor	testis

More than 70 wild type mice, 100 heterozygotes, and 60 homozygotes have been analyzed. The data collected on the mouse tumors indicate that homozygote mice are susceptible to tumors at an early age. By six months, greater than 50% of homozygotes develop tumors. Thus, the tumor incidence in such mice is considerably higher than the 20% incidence observed by 18 months of age in mice carrying a mutant p53 transgene (Lavigueur, A. et al., Molec. Cell. Biol. 9:3982-3991 (1989)). Four of the homozygote mice were found to have had multiple primary tumors of different cell type origin. The tumors in the older chimeric and heterozygote mice suggest that mice with a single mutant p53 allele are also susceptible to tumors, but at a decreased rate compared to homozygotes. No wild type control mice have developed tumors currently under observation. These results thus support the conclusion that loss of one or both p53 alleles predispose the mice to cancer. The variety of tumors identified in the homozygote mice demonstrate that p53 is involved in tumorigenesis in many tissues and cell types.

Despite having a developmentally normal appearance, a significant number of the homozygous animals succumbed to premature death in the absence of obvious tumors (tissue autolysis prevented a complete pathological analysis in some cases). Several of the deaths appeared associated with

GeneDX 1009, pg. 787

unresolved infections, suggesting the presence of a subtle defect of the immune system in the homozygotes. Such immune defects, thus appear able to contribute to the accelerated rate of tumor development in these animals.

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth and as follows in the scope of the appended claims.

GeneDX 1009, pg. 788

WHAT IS CLAIMED IS:

1. A transgenic or chimeric animal cell whose genome comprises two chromosomal alleles of a tumor-suppressing gene, wherein at least one of said two alleles contains a mutation.

2. The animal cell of claim 1, wherein one of said alleles expresses a normal tumor-suppressing gene product.

3. The animal cell of any of claims 1 or 2, wherein said tumor suppressing gene is a p53 gene.

4. The animal cell of claim 1 which is a human cell.

5. The animal cell of claim 1 which is a non-human animal cell.

6. The animal cell of claim 5 which is an embryonic stem cell.

7. The embryonic stem cell of claim 6 which is ATCC CRL 10631, and wherein said tumor-suppressing gene is a p53 gene.

8. A non-human transgenic or chimeric animal having an animal cell whose genome comprises two chromosomal alleles of a tumor-suppressing gene, wherein at least one of said two alleles contains a mutation, or a progeny of said animal, or an ancestor of said animal, at an embryonic stage.

GeneDX 1009, pg. 789

9. The non-human animal of claim 8, wherein said tumor suppressing gene is a p53 gene.

10. The non-human animal of claim 8, wherein said animal cell is a human cell.

11. The non-human animal of claim 8, wherein said animal cell is a non-human cell.

12. The non-human animal of claim 11, wherein said animal cell is a germ-line cell.

13. The non-human animal of claim 11, wherein said animal cell is a somatic cell.

14. The non-human animal of claim 11, wherein said animal and said animal cell are of the same species.

15. The non-human animal of claim 11, wherein said animal and said animal cell are of different species.

16. A non-human animal containing the embryonic stem cell of claim 6, or a progeny of said animal, or an ancestor of said animal, at an embryonic stage.

17. A non-human animal containing the embryonic stem cell of claim 7, or a progeny of said animal, or an ancestor of said animal, at an embryonic stage.

GeneDX 1009, pg. 790

18. A method for identifying the presence of an agent suspected of being capable of affecting a characteristic of an animal cell that is attributable to the presence or expression of a tumor-suppressing gene, said method comprising:

A) administering an amount of said agent to an animal cell in cell culture, said cell having a genome that comprises two chromosomal alleles of said tumor-suppressing gene, wherein at least one of said two alleles contains a mutation;

B) maintaining said cell culture for a desired period of time after said administration;

C) determining whether the administration of said agent has affected a characteristic of said animal cell that is attributable to the presence or expression of said alleles of said tumor-suppressing gene.

19. The method of claim 18 wherein said tumor-suppressing gene is a p53 gene.

20. The method of claim 18 wherein said agent is suspected of being able to increase a tumorigenic potential of said animal cell.

21. The method of claim 18 wherein said agent is suspected of being able to decrease a tumorigenic potential of said animal cell.

22. The method of claim 18 wherein said animal cell is a human cell.

GeneDX 1009, pg. 791

23. The method of claim 18 wherein said animal cell is a non-human animal cell.

24. The method of claim 23 wherein said non-human animal cell is an embryonic stem cell.

25. The method of claim 24 wherein said embryonic stem cell is ATCC CRL 10631.

26. A method for identifying the presence of an agent suspected of being capable of affecting a characteristic of an animal cell that is attributable to the presence or expression of a tumor-suppressing gene, said method comprising:

A) administering an amount of said agent to an animal, said animal having a cell whose genome comprises two chromosomal alleles of said tumor-suppressing gene, wherein at least one of said two alleles contains a mutation;

B) maintaining said animal for a desired period of time after said administration;

C) determining whether the administration of said agent has affected a characteristic of said cell that is attributable to the presence or expression of said alleles of said tumor-suppressing gene.

27. The method of claim 26 wherein said tumor-suppressing gene is a p53 gene.

GeneDX 1009, pg. 792

28. The method of claim 26 wherein said agent is suspected of being able to increase a tumorigenic potential of said animal cell.

29. The method of claim 26 wherein said agent is suspected of being able to decrease a tumorigenic potential of said animal cell.

30. The method of claim 26 wherein said animal cell is a human cell.

31. The method of claim 26 wherein said animal cell is a non-human animal cell.

32. The method of claim 31 wherein said non-human animal cell is an embryonic stem cell.

33. The method of claim 32 wherein said embryonic stem cell is ATCC CRL 10631.

34. The method of claim 26, wherein said animal and said animal cell are of the same species.

35. The method of claim 26, wherein said animal and said animal cell are of different species.

36. A method of gene therapy comprising altering the genome of a cell of an animal, wherein said cell has a genome that comprises two chromosomal alleles of a tumor-suppressing gene, wherein at least one of said two alleles

GeneDX 1009, pg. 793

contains a mutation, to thereby form a cell wherein at least one of said alleles expresses a normal tumor-suppressing gene product.

37. The method of claim 36, wherein said animal is a non-human animal.

38. The method of claim 36, wherein said animal is a human.

39. The method of claim 36, wherein said tumor-suppressing gene is a p53 gene.

GeneDX 1009, pg. 794

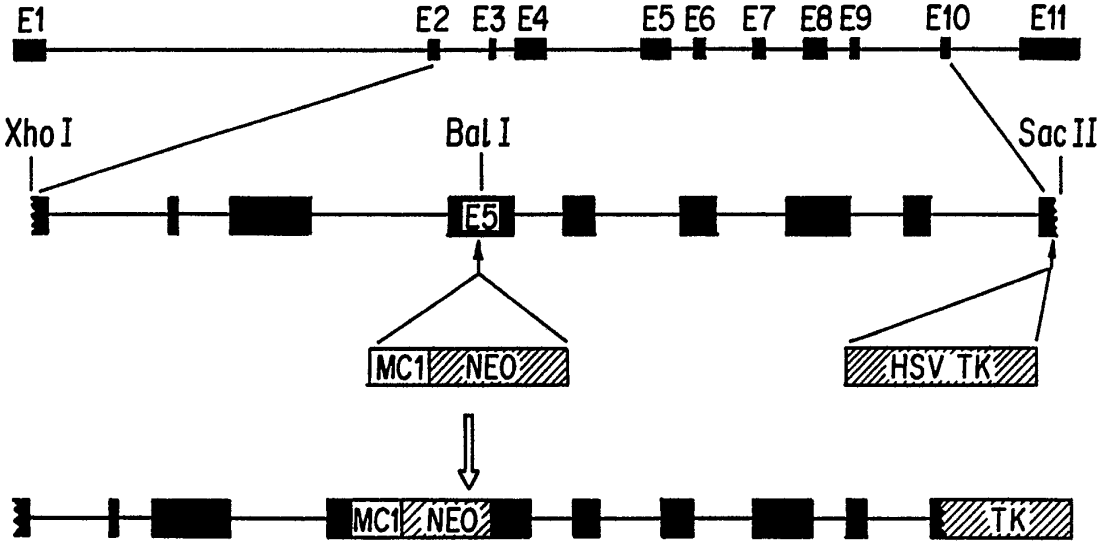
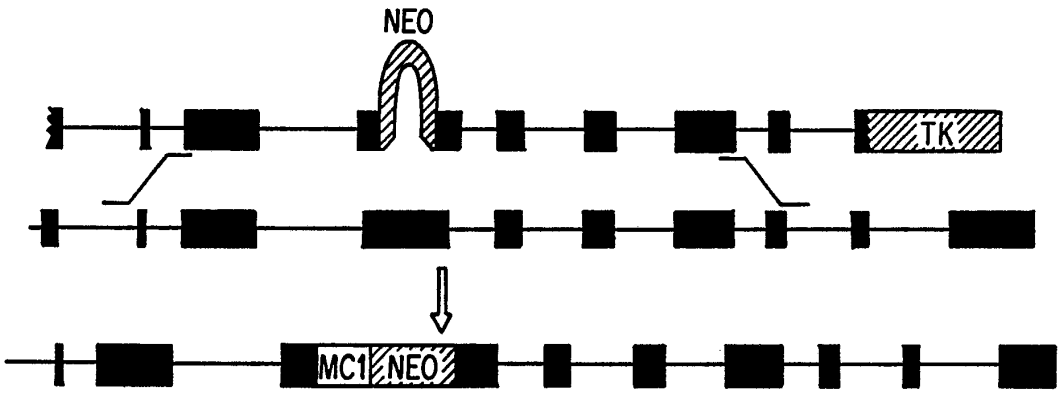
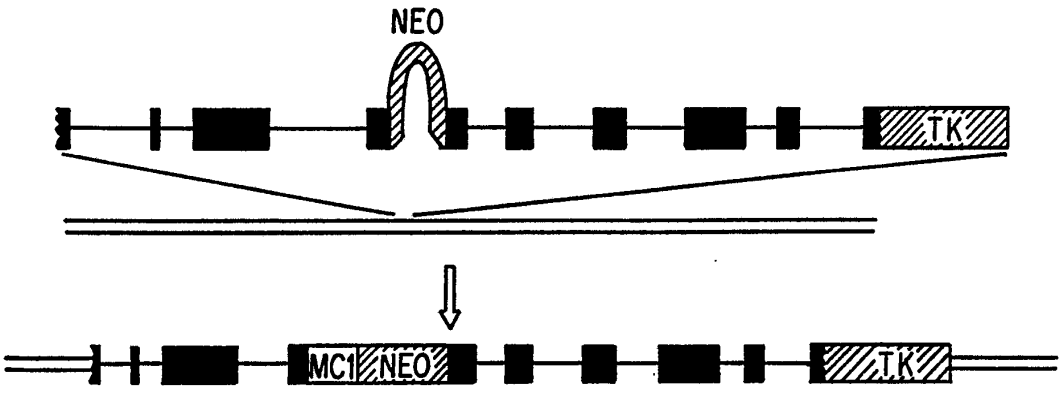


FIG. 1



$p53^{-}$ NEO^{R} HSV-TK^{-} ($G418^{R}$, $GANC^{R}$)

FIG. 1A



$p53^{+}$ NEO^{R} HSV-TK^{+} ($G418^{R}$, $GANC^{S}$)

FIG. 1B

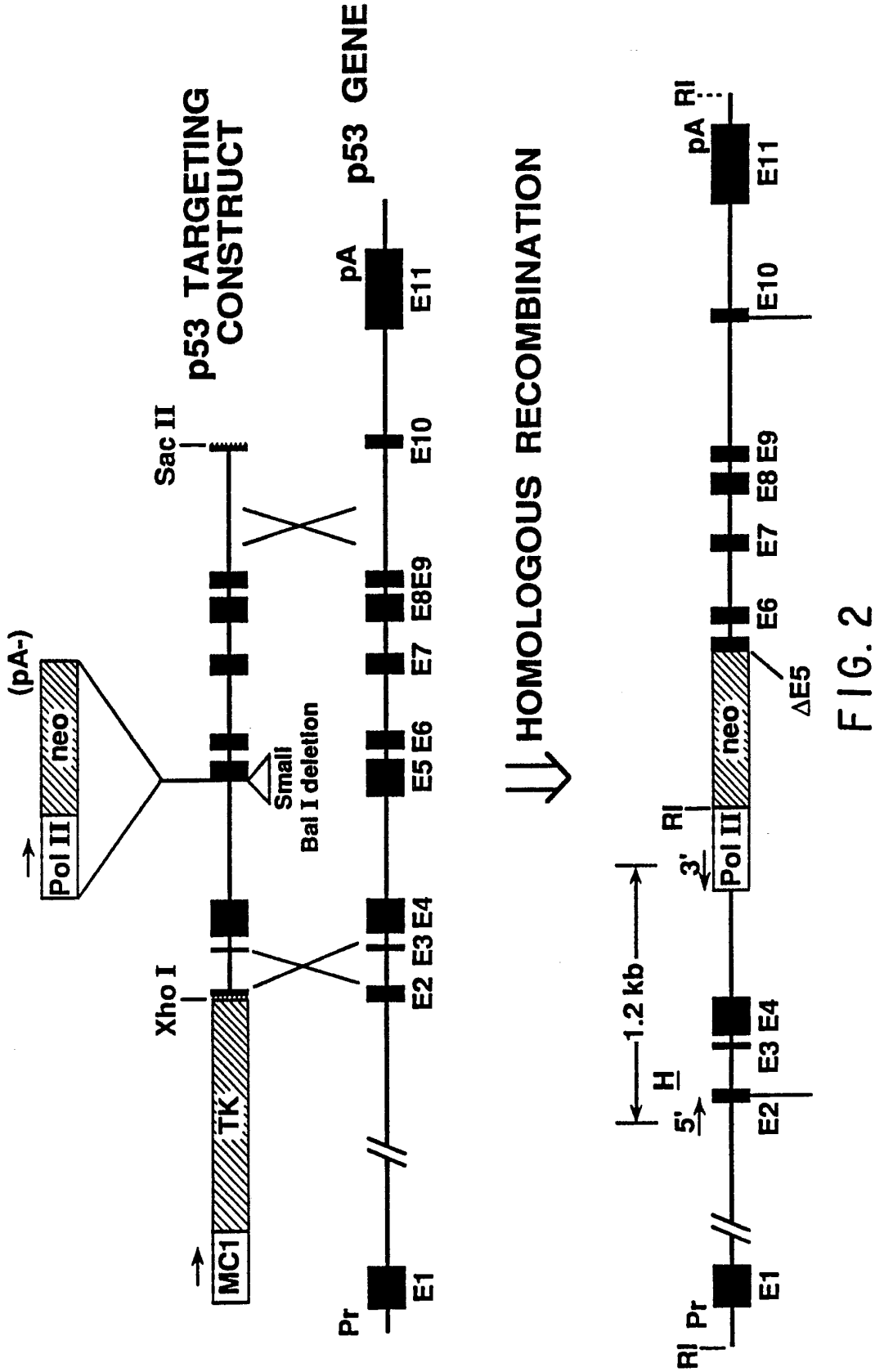

p53 TARGETING CONSTRUCT
p53 GENE
Sac II
Xho I
(pA-)
neo
Pol II
TK
MC1
SmaII
Bal I deletion
pA
E11
E10
E8E9
E7
E5 E6
E4
E3
E2
E1
Pr
HOMOLOGOUS RECOMBINATION
RI
pA
E11
E10
E9
E8
E7
E6
ΔE5
neo
Pol II
3'
5'
H
1.2 kb
E4
E3
E2
E1
Pr
RI

FIG. 2

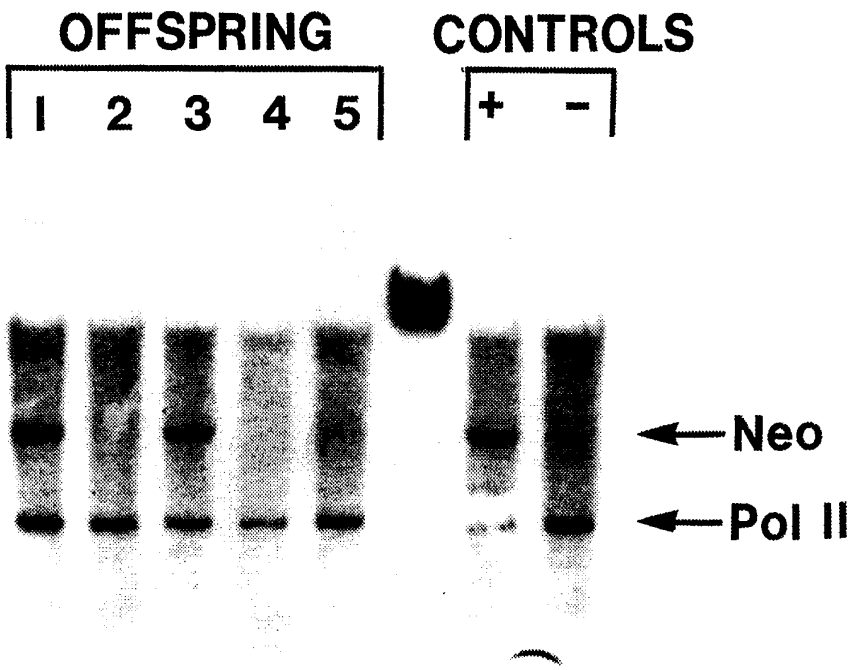
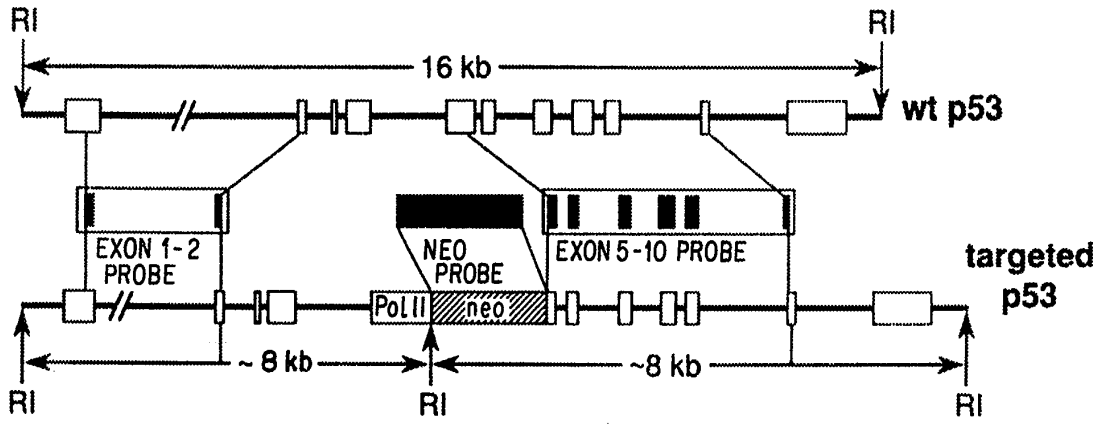


FIG. 4


RI
RI
16 kb
wt p53
EXON 1-2 PROBE
NEO PROBE
EXON 5-10 PROBE
targeted p53
PolII
neo
~ 8 kb
~8 kb
RI
RI
RI

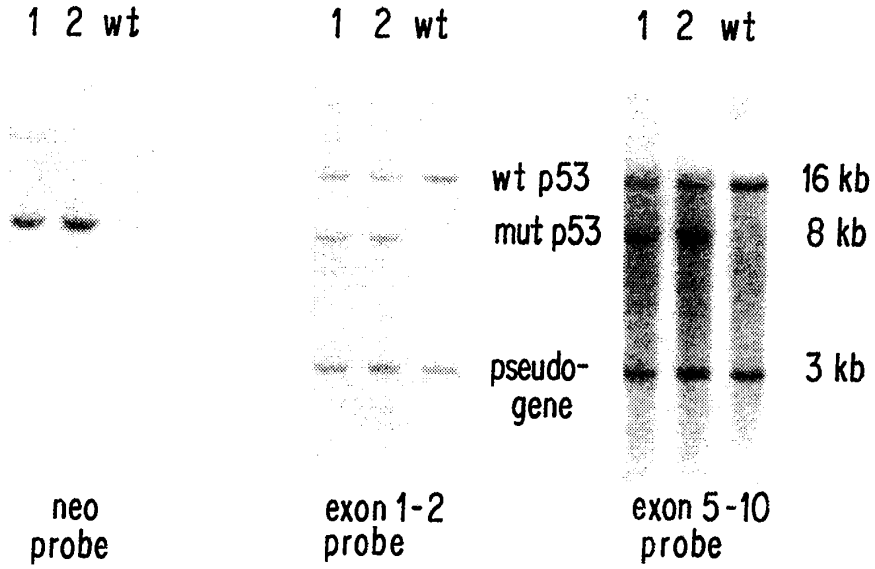

1 2 wt
1 2 wt
1 2 wt
wt p53
mut p53
pseudo-gene
16 kb
8 kb
3 kb
neo probe
exon 1-2 probe
exon 5-10 probe

FIG. 3

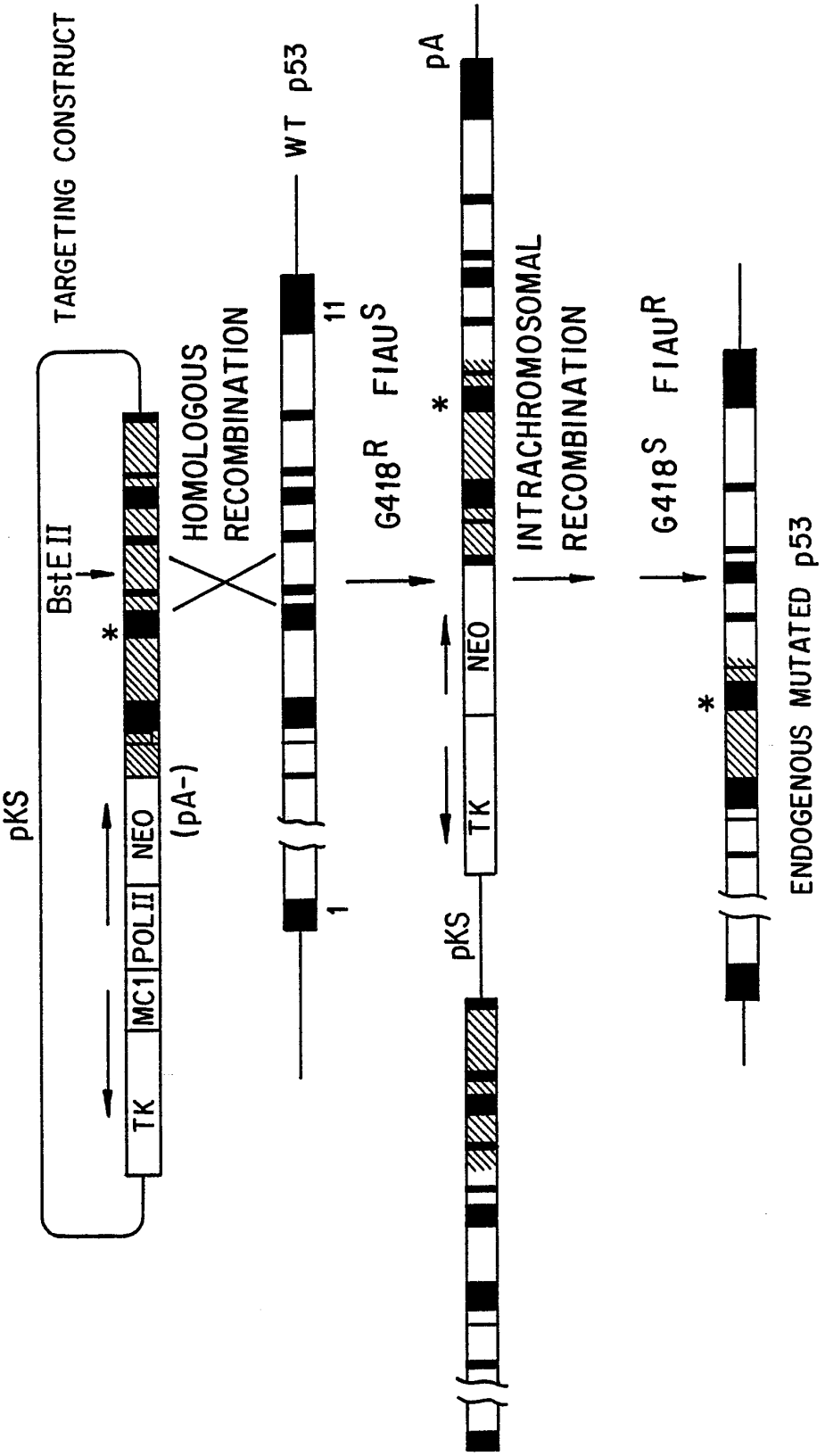
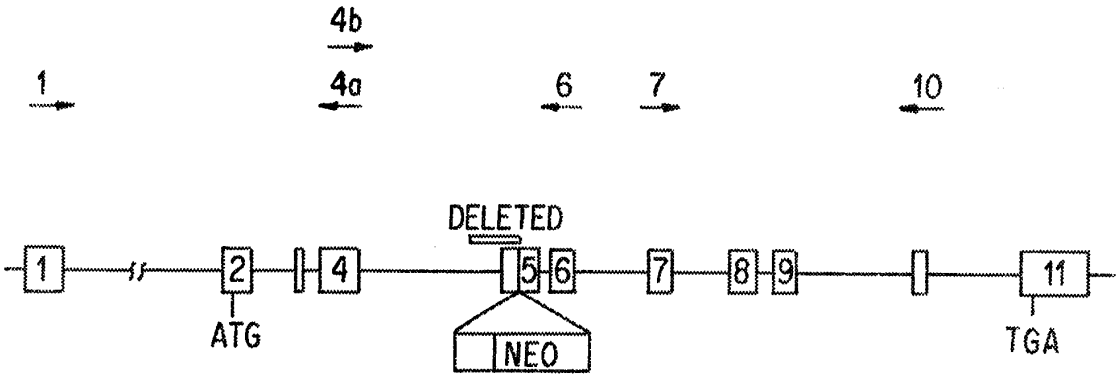

TARGETING CONSTRUCT
BstE II
pKS
TK
MC1
POLII
NEO
(pA-)
*
HOMOLOGOUS RECOMBINATION
WT p53
1
11
G418R FIAUS
pA
INTRACHROMOSOMAL RECOMBINATION
G418S FIAUR
ENDOGENOUS MUTATED p53

FIG. 5


4b
1
4a
6
7
10
DELETED
1
2
4
5
6
7
8
9
11
ATG
NEO
TGA

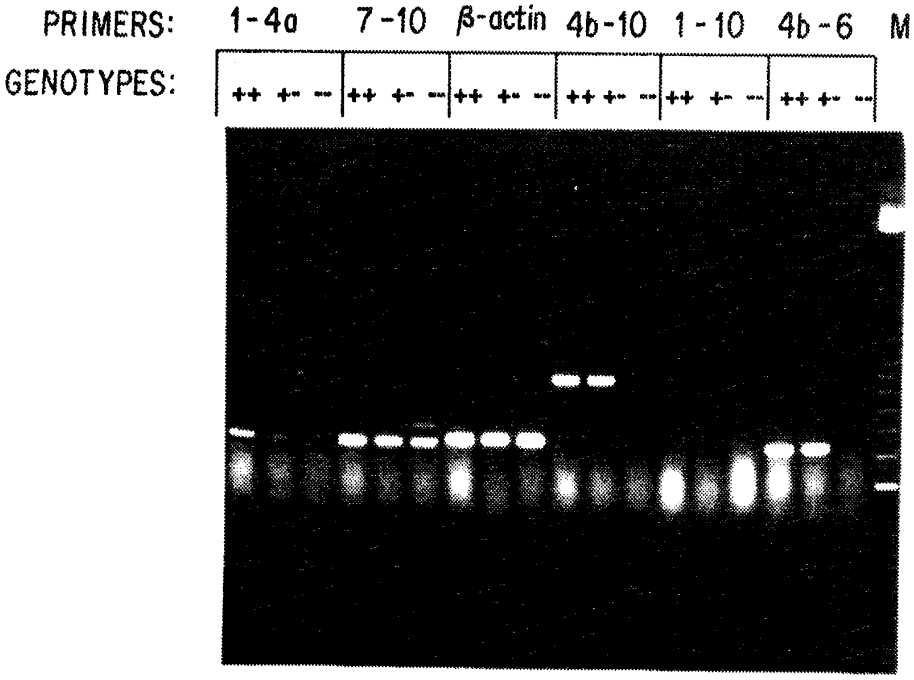

PRIMERS:
1-4a
7-10
β-actin
4b-10
1-10
4b-6
M
GENOTYPES:
++ +- -- ++ +- -- ++ +- -- ++ +- -- ++ +- -- ++ +- --

FIG. 6

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US92/00295

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[3]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC (5): A61K 49/00; C12Q 1/68; C12N 15/00, 5/00
US CL : 424/9; 435/6, 172.3, 240.2; 935/53, 57, 62, 70, 71, 76

II. FIELDS SEARCHED

Minimum Documentation Searched[4]

Classification System	Classification Symbols
U.S.	424/9; 435/6, 172.3, 240.2; 935/53, 57, 62, 70, 71, 76

Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched[5]

APS, BIOSIS

III. DOCUMENTS CONSIDERED TO BE RELEVANT[14]

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages[17]	Relevant to Claim No.[18]
X/Y	Molecular and Cellular Biology, Volume 9, No. 9, issued September 1989, Lavigueur et al, "High Incidence of Lung, Bone, and Lymphoid Tumors in Transgenic Mice Overexpressing Mutant Alleles of the p53 Oncogene", pages 3982-3991, see the entire document.	1-3, 6, 8, 9, 11-14, 16-19/4, 7, 10, 15, 20-22
X/Y	Science, Volume 250, issued 14 December 1990, Chen et al, "Genetic Mechanisms of Tumor Suppression by the Human p53 Gene", pages 1576-1580, see the entire document.	1-4/5-22
Y	US, A, 4,736,866, (Leder et al), 12 April 1988, see the entire document.	1-22
Y	US, A, 4,735,895 (Kopelovich), 5 April 1988, see column 6, lines 26-53 and column 10, line 60 to column 11, line 48.	18-22
Y	WO, A, 89/05864 (Shenk et al) 29 June 1989, see the entire document.	18-22
Y	Cell, volume 51, issued 06 November 1987, Thomas et al, "Site-Directed Mutagenesis by Gene Targeting in Mouse Embryo-Derived Stem Cells", pages 503-512, see the entire document.	1-22

* Special categories of cited documents:[16]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search[2]	Date of Mailing of this International Search Report[2]
30 MARCH 1992	23 APR 1992
International Searching Authority[1]	Signature of Authorized Officer[20]
ISA/US	Jacqueline Stone

Form PCT/ISA/210 (second sheet)(May 1986) ß

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

Y	Proc. Natl. Acad. Sci. USA, Vol. 86, issued November 1989, Koller et al, "Germ-line transmission of a planned alteration made in a hypoxanthine phosphoribosyltransferase gene by homologous recombination in embryonic stem cells", pages 8927-8931, see the entire document.	1-17

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE[1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers _, because they relate to subject matter (1) not required to be searched by this Authority, namely:

2. ☐ Claim numbers _, because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out (1), specifically:

3. ☐ Claim numbers _, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☐ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING[2]

This International Searching Authority found multiple inventions in this international application as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Search Authority did not invite payment of any additional fee.

Remark on protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

GeneDX 1009, pg. 802

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number:
G01N 33/574, C12Q 1/68 C07K 15/00	A1	WO 94/00764
		(43) International Publication Date: 6 January 1994 (06.01.94)

(21) International Application Number: PCT/US93/05952

(22) International Filing Date: 25 June 1993 (25.06.93)

(30) Priority data:
07/903,588 26 June 1992 (26.06.92) US

(71) Applicant: THE UNITED STATES OF AMERICA, represented by THE SECRETARY OF HEALTH AND HUMAN SERVICES [US/US]; Office of Technology Transfer, National Institutes of Health, Box OTT, Bethesda, MD 20892-9902 (US).

(72) Inventors: TSARFATY, Ilan ; 1812 Hammond Court, Frederick, MD 21702 (US). RESAU, James, H. ; 4145 Henhawk Court, Ellicot City, MD 21042 (US). KEYDAR, Iafa ; 16 Mania and Israel Shouet, Ramat Aviv, Tel Aviv (IL). FALETTO, Donna ; 5551 Talbot Court, Mt. Airy, MD 21771 (US). VANDE WOUDE, George, F. ; Rt. 1, Box 2905, Berryville, VA 22611 (US).

(74) Agents: BENT, Stephen, A. et al.; Foley & Lardner, 3000 K Street, N.W., Suite 500, P.O. Box 25696, Washington, DC 20007-8696 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: THE MET PROTO-ONCOGENE AND A METHOD FOR PREDICTING BREAST CANCER PROGRESSION

(57) Abstract

A method for predicting breast tumor metastasis by determining the amount of met protein in tumor tissue relative to normal breast duct tissue.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**MR**	Mauritania
AU	Australia	**GA**	Gabon	**MW**	Malawi
BB	Barbados	**GB**	United Kingdom	**NE**	Niger
BE	Belgium	**GN**	Guinea	**NL**	Netherlands
BF	Burkina Faso	**GR**	Greece	**NO**	Norway
BG	Bulgaria	**HU**	Hungary	**NZ**	New Zealand
BJ	Benin	**IE**	Ireland	**PL**	Poland
BR	Brazil	**IT**	Italy	**PT**	Portugal
BY	Belarus	**JP**	Japan	**RO**	Romania
CA	Canada	**KP**	Democratic People's Republic of Korea	**RU**	Russian Federation
CF	Central African Republic			**SD**	Sudan
CG	Congo	**KR**	Republic of Korea	**SE**	Sweden
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovak Republic
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LU**	Luxembourg	**TD**	Chad
CS	Czechoslovakia	**LV**	Latvia	**TG**	Togo
CZ	Czech Republic	**MC**	Monaco	**UA**	Ukraine
DE	Germany	**MG**	Madagascar	**US**	United States of America
DK	Denmark	**ML**	Mali	**UZ**	Uzbekistan
ES	Spain	**MN**	Mongolia	**VN**	Viet Nam
FI	Finland				

THE *MET* PROTO-ONCOGENE AND A METHOD FOR PREDICTING BREAST CANCER PROGRESSION

This is a continuation-in-part of application serial number 07/642,971, filed January 18, 1991, which is a continuation-in-part of application serial number 07/457,566 filed December 27, 1989 and application serial number 07/582,063 filed September 14, 1990, all applications pending. The entire contents of these applications are hereby incorporated by reference.

Background of the Invention

The human *met* protooncogene product (Met or Met protein) a member of the family of tyrosine kinase growth factor receptors, was identified via the activated oncogene *tpr-met* (C.S. Cooper *et al.*, *Nature* 311: 29 (1984); M. Park *et al.*, *Cell* 45: 895 (1986)). Met is synthesized as a glycosylated 170-kD precursor and cleaved in the external (ligand binding) domain to yield a mature disulfide-linked α- (50-kD), β- (140-kD) heterodimer (C.S. Cooper *et al.*, *Nature* 311: 29 (1984); S. Girodano *et al.*, *Oncogene* 4: 1383 (1989); D.L. Faletto *et al.*, *Oncogene*, in press (1992)). The Met receptor is expressed in a wide variety of tissue and cell types, but the highest levels are found in epithelial cells (M.F. Di Renzo *et al.*, *Oncogene* 6: 1997-2003 (1991); A. Iyer *et al.*, *Cell Growth & Diff.* 1: 87 (1990)).

Hepatocyte growth factor (HGF) was first purified from human and rabbit plasma and rat platelets on the basis of its ability to stimulate mitogenesis of rat hepatocytes (E. Gohoda et al., *J. Clin. Invest.* 81: 414 (1988); R. Zarnegar and G. Michalopoulos, *Cancer Res.* 49: 3314 (1989); T. Nakamura et al. *FEBS Lett.* 224: 311 (1987)). Thus, HGF may act as a humoral factor promoting liver regeneration after partial hepatectomy or liver injury (G.K. Michalopoulos, *FASEB J.* 4: 176 (1990)).

The same factor was purified from human fibroblast culture medium and shown to act on melanocytes and a variety of epithelial and endothelial cells (J.S. Rubin et al., *Proc. Natl. Acad. Sci. U.S.A.* 88: 415 (1990)). Together with evidence of HGF expression in several organs (J.S. Rubin et al., *Proc. Natl. Acad. Sci. U.S.A.* 88: 415 (1990); K. Tashiro et al. *Proc. Natl. Acad. Sci. U.S.A.* 87: 3200 (1990); R. Zarnegar et al., *Proc. Natl. Acad. Sci. U.S.A.* 87: 1252 (1990); T. Kinoshita et al. *Biochem. Biophys. Res. Comm.* 165: 1229 (1989)), these findings indicate that HGF may also act as a paracrine mediator of proliferation for a broad spectrum of cell types. Molecular cloning of HGF revealed a remarkable structural homology to plasminogen and related serine proteases (J.S. Rubin et al., *Proc. Natl. Acad. Sci. U.S.A.* 415 (1990); T. Nakamura et al., *Nature* 342: 440 (1989); K. Miyazawa et al., *Biophys. Res. Comm.* 163: 967 (1989)).

HGF is structurally related to the family of serine proteases that includes plasminogen, prothrombin, urokinase, and tissue plasminogen activator (J.S. Rubin et al., *Proc. Natl. Acad. Sci. U.S.A.* 88: 415 (1990)); T. Nakamura et al., *Nature* 342: 440 (1989)). As defined in the present invention, HGF includes a variant of HGF previously characterized as a broad-spectrum mitogen called plasminogen like growth factor (PLGF). Several proteases, including members of the serine protease family, stimulate DNA synthesis presumably through a proteolytic mechanism similar to tryptic activation of the insulin receptor (S.E. Shoelson et al. *J. Biol. Chem.* 263: 4852 (1988)). Only urokinase has been found to associate with a specific cell-surface receptor, which itself bears no homology to any known tyrosine kinase receptors (A.L. Roldan et al., *EMBO J.* 9: 467 (1990)).

U.S. Patent Application Serial Number 07/642,971, incorporated by reference above, describes the complex comprising HGF and met protooncogene protein and identifies the *met* protooncogene as the receptor for HGF.

Scatter factor (SF) originally had been considered to be related to but different from HGF, SF being associated with cell motogenicity (motility), and HGF being associated with cell mitogenicity (growth). However, recently it has been shown that HGF is identical to SF, and this factor is now referred to as "HGF/SF" (E. Gerardi *et al.*, *Cancer Cells* 3: 227 (1991); E.M. Rosen *et al.*, *Cell Growth & Diff.* 2: 603 (1991); L. Naldini *et al.*, *EMBO J.* 10: 2876 (1991b); K.M. Weidner *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7001 (1991)) and was independently shown both to promote epithelial cell motility (scattering) and to cause certain epithelial cell lines to become invasive in *in vitro* assays (E.M. Rosen *et al.*, *Cell Growth & Diff.* 2: 603 (1991); M. Stoker *et al.*, *Nature* 327: 239 (1987)).

The scattering response of HGF/SF establishes that this factor is responsible for cell motility and differentiation (G.F. Vandewoude *Japan. J. Can. Res.* 83: cover (1992)). For instance, MDCK cells grown in collagen gels in the presence of HGF/SF form bracing tubules, suggesting that the three-dimensional geometry of cell-substrate interactions directs MDCK cells to organize into tubules in response to HGF/SF (R. Montesano *et al.*, *Cell* 67: 901 (1991)).

It is known that mammary gland epithelia undergo developmental changes during pregnancy to become secretory. In culture, mammary epithelial cells can regain their differentiated phenotype only when appropriate hormonal and substratum conditions are provided (M.J. Bissell *et al.*, *The Mammary Gland*, M. Neville and C. Daniel, Eds. (Plenum Press Publishing Corp., New York, pp. 97-146 (1987); C.H. Streuli *et al.*, *J. Cell Biol.* 115: 1383 (1991)). Thus, a need existed to determine whether there was *met* expression in normal breast duct cells and to determine the role that HGF/SF and *met* play in the differentiation of ductal epithelium in the mammary gland.

In this connection, a need also existed to determine whether there was a general phenotype in epithelial cancers arising from organs that normally involve *met* expression for differentiation.

The work of Bieche *et al.*, *The Lancet*, 339: 139 (1992) has shown that the loss of heterozygosity on chromosome 7q is associated with aggressive primary breast cancer. Specifically, Bieche *et al.* used the *c-met* proto-oncogene probe, which detects sequences on chromosome 7q31, to analyse tumor and blood leucocyte DNA samples from 245 patients with primary breast cancers. The p*met* H polymorphic probe detected a high frequency of allele loss (40-50%) among the 121 informative (heterozygous) patients. This genetic alteration was not significantly associated with standard prognostic features including tumor size, histopathologic grade, and lymph-node or steroid receptor status. However, patients with loss of heterozygosity on chromosome 7q31 in primary tumor DNA had significantly shorter metastasis-free survival and overall survival after surgery than patients without this alteration. Based upon this observation, Bieche et al. hypothesized that the region on chromosome 7 detected by the probe might be the site of a breast cancer or metastasis suppressor gene. Bieche *et al.* did not identify a specific genetic locus responsible for poor prognosis in breast cancer patients.

Thus, one object of the present invention is to provide a method for predicting the progression of breast cancer based upon the presence of Met protein in normal and breast cancer tissue. Such a method would be advantageous in the management of breast cancer therapy, used either alone or in conjunction with other prognostic features such as tumor size, histo- and cytopathological grade and lymph node or steroid receptor status.

Summary of the Invention

The present invention relates to a method for predicting breast tumor progression. This method involves contacting a section from a breast tumor with an antibody reagent specific for Met protein under antibody binding conditions, wherein the section contains normal breast duct tissue and tumor tissue; determining the binding of the reagent to Met protein in the normal tissue and the tumor tissue; and comparing the binding of the reagent to met protein in the normal tissue with the binding in the tumor tissue; wherein, greater binding of the reagent to normal than to tumor tissue indicates a high likelihood of tumor metastasis.

Brief Description of the Drawings

Figure 1 shows the expression of Met in cells bordering lumen-like structures of the breast cancer cell line T47D. T47D cells were treated with Bouin's fixative, embedded in paraffin, serial sectioned, and stained by indirect immunofluorescence. Cells were analyzed using confocal laser scanning microscopy (CLSM).

Figure 1A shows a Nomarski image of the lumen-like structures in T47D cells in culture.

Figure 1B shows immunofluorescent staining of the same region as in Figure 1A with Met C28 antibody. Figures 1A and 1B are X2.45 magnification of the lower right hand portion of Figures 1A and 1B. Overall magnification is X490.

Figure 1C through 1E show quantitative determination of the Met-specific immunofluorescence. The histograms show the distribution of fluorescence of the bracketed areas.

Figure 1C specifically shows the fluorescence in the apical region of the lumen for the right lumen.

Figure 1D specifically shows the fluorescence in the apical region of the lumen for the upper left lumen.

Figure 1E specifically shows the fluorescence in the apical region of the lumen for an area adjacent to the upper left lumen.

Figure 2 shows subcellular localization of Met in T47D cells. Immunoelectron microscopy was performed using secondary antibody labeled with 10-nm colloidal gold particles (magnification 30,000X) (For postembedding immunoelectron microscopy, fixed cells were embedded in L.R. gold resin (Electron Microscopy Science) at -25°C, sectioned with an LKB Nova Ultratome, and picked up with Formvar-coated 200 mesh gold grids. The grids were washed three times in PBS 10 min and incubated in 1% BSA (in PBS) for 2 hours and in rabbin anti-human C28 Met antibody (diluted 1:50 in 1% BSA) at room temperature for 1 hour. Controls were incubated either in the presence of C28 competing peptide or in the absence of the primary antibody. The grids were washed again in PBS, incubated in RPM11640 medium for 20 min in 1% BSA as above, and reacted with goat anti-rabbit IgG gold (10-nm diameter; 1:10 diluted in 1% BSA; Amersham) at room temperature for 1 hour. The grids were finally washed in PBS and distilled water and stained with uranyl acetate and lead citrate. The sections were observed and photographed using a Philips EM 410 electron microscope.). Met is localized in microvilli that protrude into lumen (L) and could account for the apical staining observed in cells bordering the lumen (A and B).

Figure 3 shows expression of Met in breast tissue biopsies. A paraffin-embedded section of biopsy from breast adenocarcinoma was subjected to indirect immunofluorescence staining with C28 antibody or monoclonal antibody 4G10. The cells and immunofluorescence were analyzed by CLSM.

Figure 3A shows immunofluorescence of normal mammary duct, performed with C28. Magnification X250.

Figure 3B shows immunofluorescence of normal mammary duct, performed with anti-P-Tyr. Magnification X250.

Figure 3C shows immunofluorescence of normal mammary duct, performed with C28. Magnification X250.

(4G10); (F) analysis of the overlap between the Met and the anti-P-Tyr staining, using the confocal image processing program (red indicates increased overlap).

Figures 4A-F relate to expression of Met in the mouse embryonal digestive tract. Eleven-day-old BALB/c mouse embryos were fixed, embedded in paraffin, and subjected to immunofluorescence staining with SP260 and 4G10 antibodies. Magnification is X220.

Figure 4A shows apical staining of the lumen of the digestive tract with SP260.

Figure 4B shows staining in the absence of the primary SP260 antibody.

Figure 4C shows staining in the presence of SP260 competing peptide.

Figure 4D shows apical staining of the lumen of the digestive tract with SP260.

Figure 4E shows immunofluorescence with 4G10.

Figure 4F shows an analysis of the overlap between the Met and the anti-P-Tyr staining by a confocal image processing program.

Figures 5A-F relates to HGF/SF induction of lumen formation in SW480 and HT29 human colon carcinoma cell lines. The cells were treated with the specified concentration of HGF/SF when they were -70% confluent. Magnification is X164.

Figure 5A shows SW480 cells grown for 18 hours in complete medium without HGF/SF.

Figure 5B shows SW480 cells grown for 18 hours in complete medium with HGF/SF at 1 ng/ml.

Figure 5C shows SW480 cells grown for 18 hours in complete medium with HGF/SF at 5 ng/ml.

Figure 5D shows SW480 cells grown for 18 hours in complete medium with HGF/SF at 10 ng/ml.

Figure 5E shows SW480 cells grown for 18 hours in complete medium with HGF/SF at 100 ng/ml.

Figure 6 shows the influence of HGF/SF concentration on lumen formation. The number of cells forming lumen structures was calculated for each HGF/SF concentration.

Detailed Description of the Invention

Applicants have discovered that Met protein and RNA products are prognostic in evaluating tumor progression towards metastasis in human breast cancer. The expression of *met* in certain organs is associated with epithelial cell differentiation in tumor cells. More specifically, applicants have discovered that expression of the *met* protooncogene product in normal tissues is associated with the formation of highly ordered, fully differentiated lumen structures and that the ligand for Met, HGF, plays a key role in inducing the formation of these structures. It also has been discovered that while *met* expression is very high in cells that form normal breast ducts, expression of *met* in adjacent, poorly differentiated, cancer tissue is markedly reduced. This information has led to the further discovery that reduced *met* expression correlates with poor prognosis in human breast cancer.

The method of the present invention involves comparing the amount of Met in normal breast tissue with the amount of Met in tumor tissue. This is accomplished by determining one or more of *met* DNA abundance, *met* mRNA abundance, or Met protein abundance in normal breast tissue and breast tumor tissue from the same patient. The detection of a higher abundance of *met* DNA, *met* mRNA or Met protein in normal tissue than in tumor tissue is indicative of a poor prognosis.

In one embodiment of the present invention, the method involves contacting a sample of breast tissue containing both normal and tumor cells with an antibody reagent specific for Met protein, under antibody binding conditions. Greater binding of the reagent to normal

than to tumor tissue indicates a high likelihood of tumor metastasis and a poor prognosis.

Preferably, the breast tissue sample is a "section," i.e., a histological section obtained from a surgical or biopsy procedure, according to techniques well known in the art. The term "contacting" is intended to encompass any technique which permits the antibody reagent to bind with Met protein.

By antibody reagents is meant whole antibodies and parts thereof, either alone or conjugated with other moieties. Antibodies include polyclonal antibodies, monoclonal antibodies, and single chain antibodies. Antibody fragments are those the bind the Met protein, including Fab and $F(ab)_2$ fragments, *inter alia*. The antibody reagents may comprise antibodies made in animals or by recombinant DNA techniques. Also the antibody reagents include antibody and antibody fragments conjugated to, among other moieties, detectable labels, such as enzymatic labels and fluorescent labels. Other useful labels which may be comprised by the antibody reagents include radionuclides.

Antibody binding conditions generally are well known in the art and, for the most part will include neutral pH, moderate salt, temperatures between 2-3 C and 37° C, incubation times between several minutes and overnight or longer. Preferred conditions include those described by M. Gonzatti-Haces *et al., Proc. Natl. Acad. Sci. USA* 85: 21 (1988) for a rabbit antibody prepared against the extracellular domain of the Met peptide, those of M. Park *et al., Proc. Natl. Acad. Sci. USA* 84: 6379 (1987) for a rabbit anti-mouse Met and the conditions described by D.K. Morrison *et al. Cell* 58: 649 (1989)) for a mouse monoclonal antibody to phosphorylated tyrosine (P-Tyr). It will be readily appreciated by those of ordinary skill in the art that the details of an antibody binding procedure may be adjusted to favor improved signal to noise ratios or sensitivity, rapidity or completeness, and the like. Moreover, conditions may be adjusted to

accommodate different histological procedures for fixation and staining. Finally, it will be appreciated that titers and appropriate dilutions will be different for different antibody reagent preparations.

Binding of the antibody reagent with Met protein may be determined by any method known to the skilled artisan and may include visualization by conventional or confocal laser scanning microscopy.

The term "greater binding" is readily understood to be a relative term which derives meaning by comparing the amount of bound Met in normal with the amount of bound Met in tumor cells. An 80- to 150-fold or greater reduction in Met in tumor cells is considered predictive of a poor prognosis. Whether a prognosis is good or poor depends upon the likelihood of metastasis. A "high likelihood" of tumor metastasis means that there is about an 80-90% chance of tumor metastasis within about two years.

In the preferred embodiment of the present invention, a histological tumor breast biopsy section containing both normal duct tissue and adjacent tumor tissue is embedded in paraffin and blocked for about 10 minutes using a commercially available immunostaining blocking reagent. A primary antibody is then incubated with the tissue. The preferred primary antibody is selected from the group consisting of C28 (rabbit anti-human Met), C200 (rabbit antibody against the extracellular domain of the Met peptide), SP260 (rabbit anti-mouse Met), 4G10 (mouse monoclonal antibody to phosphorylated tyrosine (P-Tyr) and the 19S monoclonal (previously described in U.S. patent application serial no. 07/457,556). The tissue is then further incubated with a secondary antibody coupled to a fluorochrome. The secondary antibody may be donkey anti-rabbit coupled to phycoerythrin or donkey anti-mouse coupled to fluorescein isothiocyanate. Antibody binding is determined by examining the fluorochrome labeled cells with a conventional light or confocal laser scanning microscope, such as the Zeiss laser scan microscope. The

fluorescent intensity is then quantitated using methods known in the art. In the preferred embodiment, the relative fluorescence intensity (the intensity of the normal compared with tumor tissue) is quantitated with an Indec Laser Scan Microscope Image Processor Option. Optionally, photographs may be taken to document visualized tissue. A poor prognosis is indicated when an 80- 150-fold reduction in Met in the tumor tissue as compare with normal tissue is observed.

In yet another embodiment of the present invention, the abundance of *met* mRNA or *met* DNA in a breast tissue sample may be detected by *in situ* hybridization using *met* sequence specific probes, or by hybridization of *met* sequence specific probes to mRNA or DNA from normal and tumor tissue. Probe complementary to *met* are prepared by methods known in the art and allowed to hybridize to mRNA or DNA within a section of a tissue sample (either embedded by standard techniques, e.g. paraffin, or otherwise preserved). Unhybridized probe are removed by nuclease. Hybridization can be detected by autoradiography or other methods. The intensity of hybridization reflects the amount of *met* within the cells of the tissue. The observation of tumor tissue containing reduced levels of *met* relative to levels in normal tissue, indicates a poor prognosis.

In yet another embodiment of the present invention, the polymerase chain reaction ("PCR") is used to detect *met* DNA or mRNA in a breast tissue sample. To carry out the PCR, a pair of *met* sequence specific primers is employed, which hybridize to opposite strands of the *met* gene at offset positions on the double helix. The primers provide initiation points for DNA synthesis.

In the presence of DNA polymerase, the four nucleotide triphosphates ("NTPs"), and other necessary co-factors, which are all well known to the art, new DNA strands are synthesized complementary to the templates which hybridized with the primers. Several rounds of synthesis are carried out, the double stranded products

being denatured between rounds. Preferably, a thermal stable DNA polymerase is used so that it is not necessary to add enzyme anew for each round of synthesis.

The PCR produces a double stranded DNA amplification product which has the same sequence as the original stretch of the *met* DNA defined by the ends of the primer pair sequences. The amount of PCR product indicates the amount of *met* DNA or *met* mRNA in the sample. The product can be detected by a variety of methods well-known in the art.

Where the products are produced in a test tube, or the like, they can be resolved by agarose or polyacrylamide electrophoresis and detected by fluorescence staining, such as ethidium bromide. Alternatively, one of the NTPs may be labelled and the PCR products may be determined by measuring incorporation of the labelled NTP. A variety of other methods for resolving, detecting and measuring the amount of PCR product are well-known to the art that are suitable for use in the present invention.

In situ techniques may employ the use of fluorescent and radiolabels which can be easily quantitated by fluorescence microscopy or autoradiography, respectively. Generally, fluorescent labels will be preferred. Another labelling technique may employ enzymatic tags which generate readily quantifiable colorimetric or chemiluminescent signals.

PCR may be rendered specific for *met* DNA or *met* mRNA in *in situ* and in liquid PCRs. For instance, RNAse or DNAse may be used to remove one template or the other from the sample, and the use of primers that distinguish between the gene and the message (e.g. a primer that hybridizes to a sequence in the untranscribed region of the promoter will be gene specific).

In accordance with the methods of the present invention, PCR is used to detect the presence of *met* DNA or *met* mRNA in both normal and tumor tissue. Therefore, when liquid PCR is employed, it is essential

to distinguish samples containing normal and tumor tissue. Reduced amounts of *met* DNA or *met* mRNA in tumor, relative to normal tissue, is predictive of tumor metastasis.

In another embodiment, RNA ("Northern") blotting may be used in the methods of the invention. According to this method, RNA is isolated from tumor and normal breast tissue by any of a number of standard procedures. (Lehrach, H., *Biochemistry*, 16: 4743 (1975)) Again, it is important that separate assays be run on the tumor and normal breast tissue so that respective hybridization results can be compared. RNA is subjected to denaturing gel electrophoresis and transferred to nitrocellulose or other support matrix. The *met* mRNA can be detected by hybridization of radioactively or non-radioactively labelled *met*. The presence of *met* mRNA is detected by the intensity of hybridization. The intensity observed in normal tissue is compared with that in tumor tissue; a reduced amount of intensity in tumor tissue is indicative of poor prognosis.

Other techniques suitable to the claimed methods are readily apparent to the skilled artisan and include Nuclease Protection Assays, ELISA and Western blotting, as described in Example 1.

Although the present invention is exemplified with a method of predicting the progression of breast cancer, the claimed methods are likewise applicable to predicting the progression of any tumor associated with tubule forming epithelial tissues, such as those found in the digestive tract.

Although the method of the present invention involves the detection of *met* DNA, *met* mRNA or Met, the detection and comparison of HGF/SF DNA, mRNA or protein in normal and tumor tissue is likewise predictive of tumor progression, in accordance with the teachings of the present application.

It is noted that unless defined otherwise, all technical and scientific terms used herein have the same

meaning as commonly understood by one of ordinary skill in the art of the invention. Although any methods and materials similar or equivalent to those described herein can be used in the practice of the invention, the preferred methods and materials have been described. Unless mentioned otherwise, the techniques employed or contemplated herein are standard methodologies well known to one of ordinary skill in the art. The materials, methods and examples are illustrative only and not limiting.

Example 1

Expression of Met in cells and tissues of both human and mouse origin. Immunoprecipitation and Western analyses showed that *met* is expressed in human breast carcinoma cell line T47D (I. Keydar *et al., Eur. J. Cancer* 15: 659 (1979)) Applicants then examined the distribution of Met in T47D cells by immunofluorescence and confocal laser scanning microscopy (CLSM), using antibody C28 (M. Gonzatti-Haces *et al., Proc. Natl. Acad. Sci. USA* 85: 21 (1988)) or C200 rabbit antibodies directed against a C-terminal or an N-terminal peptide of Met, respectively, as shown in Figure 1. Controls were prepared with the C28 antiserum in the presence of competing peptide or without the C28 antiserum.

It is known that T47D cells in suspension form lumen-like structures resembling mammary ducts (I. Keydar *et al., Eur. J. Cancer* 15: 659 (1979)). Analyses of T47D cells in paraffin sections stained with C28 antibody showed intense fluorescent staining in cells bordering lumen-like structures, as shown in Figure 1A. A marked decrease in intensity was observed when competing C28 peptide was added to the primary antibody or when the primary antibody was omitted. Moreover, T47D cells stained with the C200 antibody, directed against the met extracellular domain, also showed intense fluorescent staining in cells bordering the lumen-like structures, as

shown in Figure 1A. By CLSM, applicants observed an 80-fold greater fluorescence in cells forming the lumen border than in adjacent cells, as shown in Figure 1B.

Applicants further investigated the subcellular localization of Met by electron microscopy using indirect immunogold labeling and the C28 Met antibody. Through this analysis, applicants discovered that Met was localized to microvilli that protrude into lumen, as shown in Figure 1C.

The intense met-specific staining of cells lining the lumen borders in T47D breast carcinoma cells led the applicants to examine *met* expression in normal and abnormal human breast tissue. Specifically, applicants examined fifty human breast carcinoma biopsies by CLSM using the C28 Met antibody. The normal and tumor tissue from a representative breast biopsy is shown in Figure 2. Applicants observed intense Met fluorescence in cells that form a normal mammary duct, as shown in Figure 2A. The intensity of staining was 80-fold greater in the duct-forming cells than in the adjacent nonductal cells. The pattern of Met staining in the mammary duct was similar to, but higher than, that observed in the T47D lumen-like structures. Applicants further discovered that, in contrast, Met staining in adjacent breast tumor tissue was always reduced, but even in the disrupted architecture of the tumor, met fluorescence was evident in lumen- or duct-like structures, as shown in Figure 2C.

It is known that Met is rapidly phosphorylated in tyrosine residues after HGF/SF activation (D.P. Bottaro *et al.*, *Science* 251: 802 (1991); L. Naldini *et al.*, *Oncogene* 6: 501 (1991)). Applicants have further observed that anti-phosphotyrosine (anti-P-Tyr) monoclonal antibody labeling colocalizes with Met staining in cells activated with the met ligand, HGF/SF. In light of this, applicants costained breast tissue samples with the anti-P-Tyr antibody and found intense anti-P-Tyr fluorescence that colocalized with Met staining in the cells forming

normal breast ducts (Figure 2B). Applicants conclude from these results that Met and its substrates are activated in the cells bordering the normal breast duct. As with the lower Met staining of tumor tissue (Fig. 2C), applicants also observed reduced levels of anti-P-Tyr staining in the tumor cells (Fig. 2D); however, even the lower levels of anti-P-Tyr fluorescence colocalized with Met staining.

Thus, applicants have also observed that human *met* expression both *in vitro* and *in vivo* is concentrated in cells bordering lumen-like structures. Since extensive lumen and duct formation occurs in the early development of the digestive tract, applicants further investigated *met* expression in this tissue in mouse embryos. These analyses, performed using a C-terminal rabbit peptide antibody, SP260 (A. Iyer *et al., Cell Growth & Diff.* 1: 87 (1990)), showed that in 11-day-old mouse embryos *met* is expressed in the lumen-bordering cells, as shown in Figure 3. Applicants noted that, as in human samples, intense Met staining was evident on the apical side of the cells forming the lumen of the digestive tract, as shown in Figures 3A and 3D, and was 50-fold higher in these cells than in the surrounding tissue. Serial sections stained in the absence of the primary C-terminal mouse Met antibody or in the presence of competing peptide showed much lower levels of fluorescence, as shown in Figures 3B and 3C. Moreover, as in the cells lining the human breast duct, mouse Met staining colocalized with anti-P-Tyr staining in the lining of the embryonic digestive tract, as shown in Figure 3E. Thus, applicants have discovered that the Met receptor appears to be preferentially expressed in the border cells of the breast duct and the gastrointestinal tract lumen. Similarly, Met staining in lumen-like structures has been observed in human gastrointestinal lumen and lumen of biliary ducts and esophagus (M. Pratt *et al. Int. J. Cancer* 49: 323 (1991)). The colocalization of Met

staining with anti-P-Tyr immunofluorescence implies that the Met receptor is activated in the lumen-forming cells of the breast duct and gastrointestinal tract, a shown in Figures 2 and 3.

Having established a correlation between the expression of Met and lumen formation, applicants tested whether Met-HGF/SF could induce lumen formation in human epithelial carcinoma cells *in vitro*. Specifically, cells expressing Met were exposed to varying concentrations of HGF/SF (1-100 ng/ml). More specifically, cells were grown on 16-chamber Labtek slides (Nunc) and treated with specific concentrations of HGF/SF (95% pure, Collaborative Research) for 24 hours. After two washes in PBS, cells were fixed for 10 minutes in cold methanol (-20°C) and washed extensively with PBS. Cells were also stained for 10 minutes with 0.1% methylene blue in PBS and were visualized and photographed using a Zeiss microscope. The results with two human colon carcinoma cell lines, SW480 and HT29, are shown in Figure 4. Indeed, applicants observed dramatic formation of lumen structures in these cells (as well as in breast carcinoma cell lines) when the cells were grown on glass (70% confluent) and treated with HGF/SF. Lumen formation was cell density dependent and not obvious when cells were grown in plastic tissue culture flasks or treated with epidermal growth factor instead of HGF/SF under the same conditions. Furthermore, lumen formation was dependent on the dose of HGF/SF, as shown in Figures 4A and 4B. Thus, when cells were treated with 1-10 ng of HGF/SF per milliliter of medium, applicants observed lumen-like structures with a uniform size. The regularity in the size indicates that the number of cells forming the lumen border is controlled and that cell-cell interactions must play an important role in this formation.

Example 2

Immunofluorescence Assays

For immunofluorescence assays, fixed cells or paraffin-embedded tissues, cells, and embryo sections (I. Keydar *et al., Eur. J. Cancer* 15: 659 (1979)) were blocked for 10 minutes using the Biomeda immunostaining kit blocking reagents. Primary antibody (C28, rabbit anti-human Met at 1:100 dilution in PBS according to the methods of M. Gonzatti-Haces *et al., Proc. Natl. Acad. Sci. USA* 85: 21 (1988); C200, rabbit antibody prepared against the extracellular domain of the Met peptide, amino acids 643-663, according to the methods of M. Park *et al., Proc. Natl. Acad. Sci. USA* 84: 6379 (1987); SP260, rabbit anti-mouse Met at a 1:100 dilution in PBS; 4G10, mouse monoclonal antibody to phosphorylated tyrosine (P-Tyr), according to D.K. Morrison *et al. Cell* 58: 649 (1989)) was added and incubated for 2 hours at room temperature. Secondary antibody incubation (donkey anti-rabbit coupled to phycoerythrin at 1:50 and donkey anti-mouse coupled to fluorescein isothiocyanate at 1:100 (Jackson ImmunoResearch Laboratories, Inc.) were performed for 1 hour at room temperature. After an extensive washing, cells were fixed using gel mount (Biomedia).

Fluorochrome-labeled cells were examined using a Zeiss laser scan microscope (LSM) having the following configuration: 25 nW argon and HeNe lasers, 488, 514, and 543 maximum lines; control Indec (Sungate) software for image acquisition of X-Y scan, Z-series scan three-dimensional visualization. Photographs were taken using a Sony color video printer mavigraph and Sony UPC-5010a color print paper. When comparing the fluorescence intensity, applicants used identical parameters for each image (e.g., scanning line, laser light, contrast brightness) and assessed quantitation of the relative fluorescence by using an Indec LSM Image Processor Option, "Histogram."

Example 3

Immunoelectron microscopy

For postembedding immunoelectron microscopy, fixed cells were embedded in L.R. gold resin (Electron Microscopy Science) at -25°C, sectioned with an LKB Nova Ultratome, and picked up with Formvar-coated 200 mesh gold grids. The grids were washed three times in PBS 10 min and incubated in 1% BSA (in PBS) for 2 hours and in rabbin anti-human C28 Met antibody (diluted 1:50 in 1% BSA) at room temperature for 1 hour. Controls were incubated either in the presence of C28 competing peptide or in the absence of the primary antibody. The grids were washed again in PBS, incubated in RPM11640 medium for 20 min in 1% BSA as above, and reacted with goat anti-rabbit IgG gold (10-nm diameter; 1:10 diluted in 1% BSA; Amersham) at room temperature for 1 hour. The grids were finally washed in PBS and distilled water and stained with uranyl acetate and lead citrate. The sections were observed and photographed using a Philips EM 410 electron microscope.

What Is Claimed Is:

1. A method for predicting the progression of a breast cancer by (a) determining one or more of Met DNA abundance, Met mRNA abundance, or Met protein abundance in normal breast tissue and in breast tumor tissue, and (b) comparing the abundance of said met DNA, met RNA or met protein in normal breast tissue with said met DNA, met RNA or Met protein in tumor breast tissue, wherein said normal and tumor tissue is from the same breast, and, wherein a higher abundance of Met DNA, Met RNA or Met protein in said normal tissue than in said tumor tissue indicates a poor prognosis.

2. A method according to claim 1, comprising the steps of:

(a) contacting a section from a breast tumor with an antibody reagent specific for Met protein under antibody binding conditions, wherein said section contains normal breast duct tissue and tumor tissue;

(b) determining the binding of the reagent to Met protein in said normal tissue and said tumor tissue; and

(c) comparing said binding of said reagent to Met protein in said normal tissue with said binding in said tumor tissue; wherein, greater binding of said reagent to said normal than to said tumor tissue indicates a high likelihood of tumor metastasis.

3. The method according to claim 2, wherein said antibody reagent is an immunofluorescent stain comprising and antibody selected from the group consisting the C28, C200 and 19S, SP260 and 4G10 monoclonal antibodies.

FIG. 1B

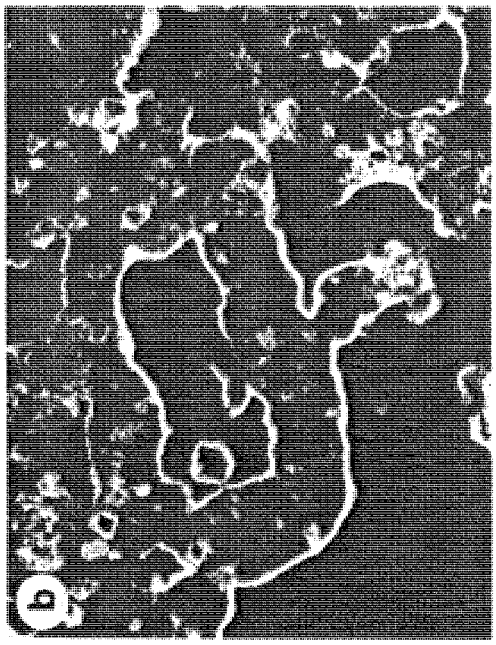


FIG. 1A



GeneDX 1009, pg. 825

FIG. 1D

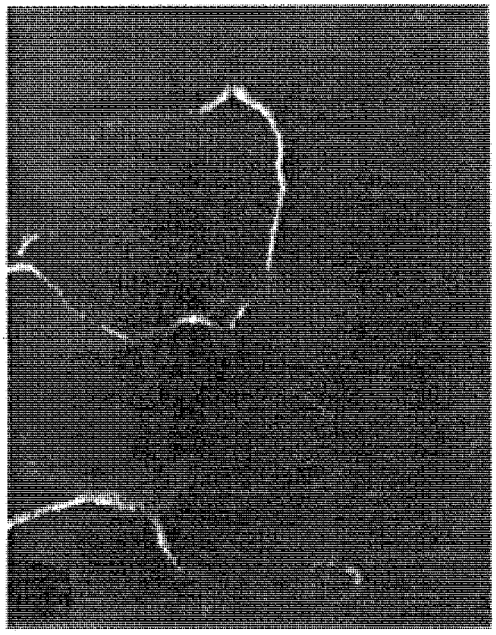


FIG. 1G

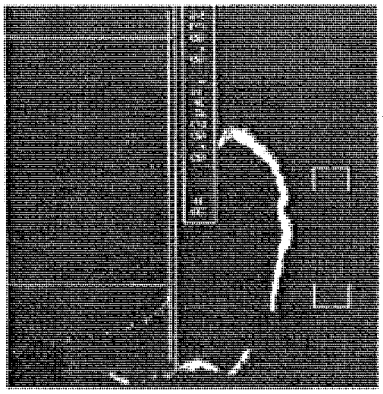


FIG. 1C



FIG. 1F

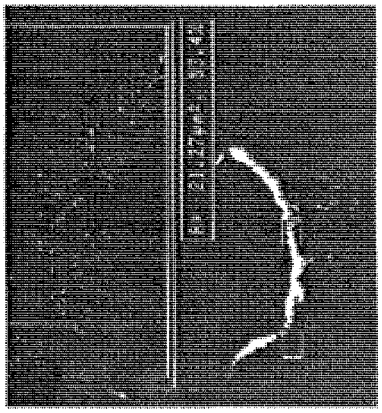
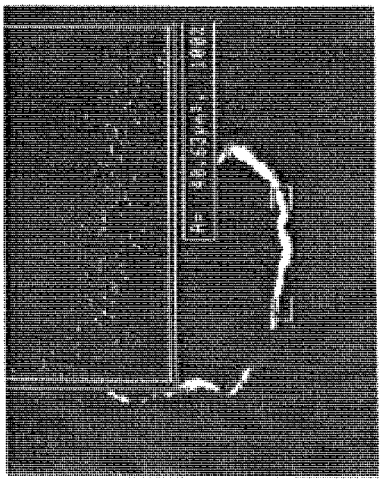
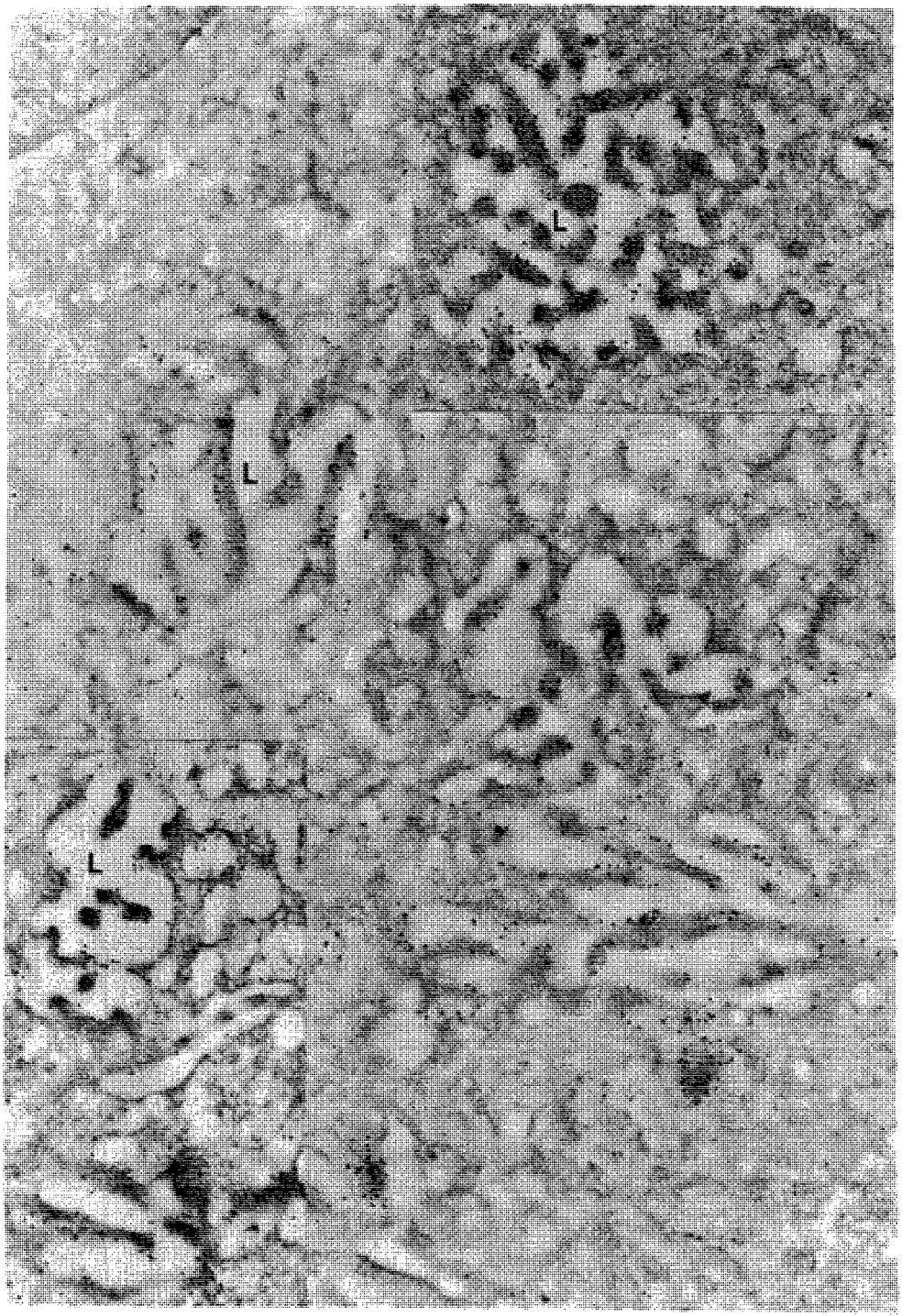


FIG. 1E



GeneDX 1009, pg. 826

FIG. 2



GeneDX 1009, pg. 827

FIG. 3B



FIG. 3A



GeneDX 1009, pg. 828

FIG. 3D

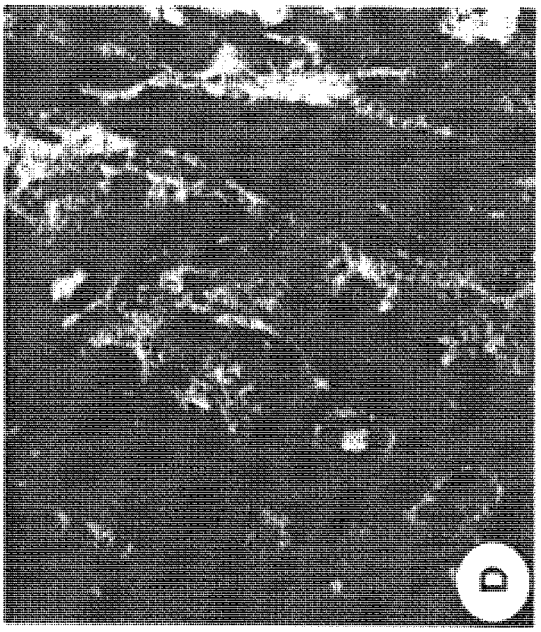
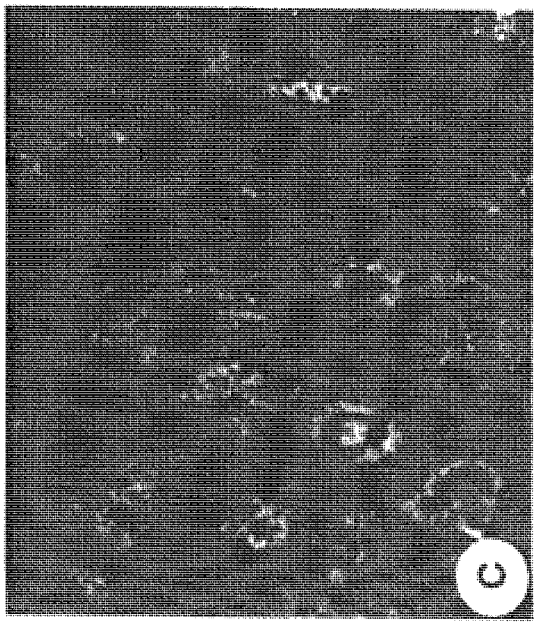


FIG. 3C



GeneDX 1009, pg. 829

FIG. 4C



FIG. 4B

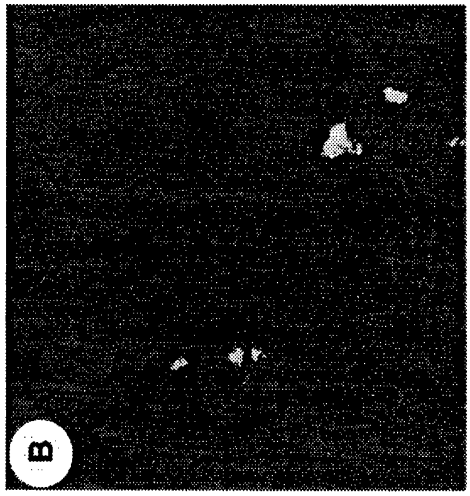


FIG. 4A



GeneDX 1009, pg. 830

FIG. 4F

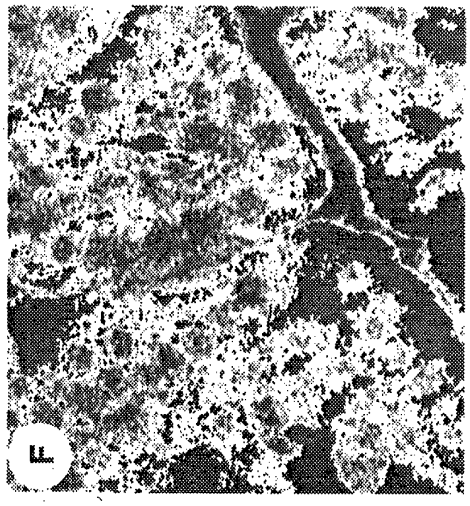
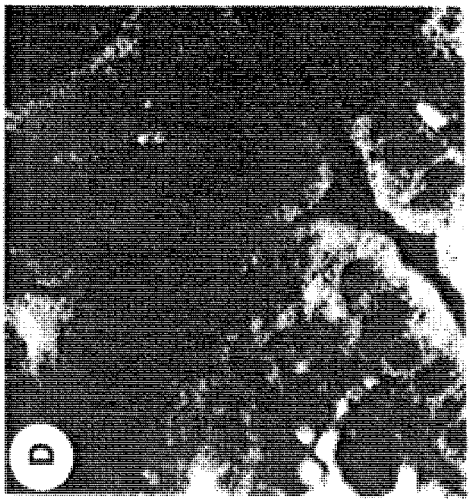


FIG. 4E



FIG. 4D



GeneDX 1009, pg. 831

FIG. 5A

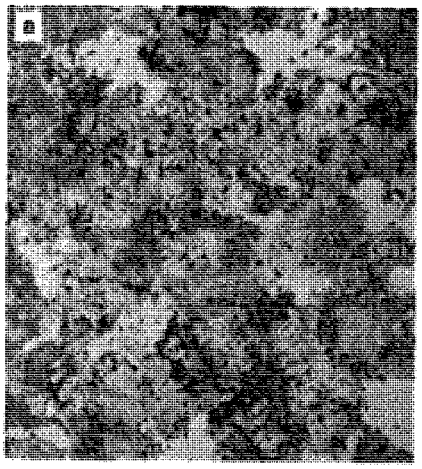


FIG. 5D

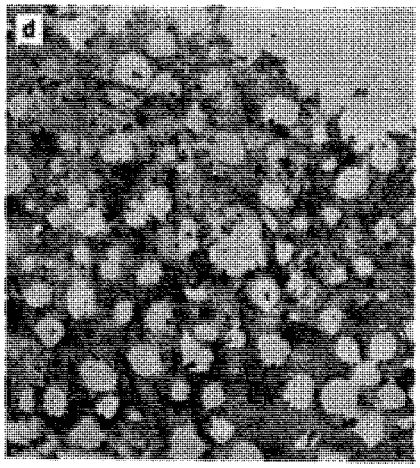


FIG. 5B

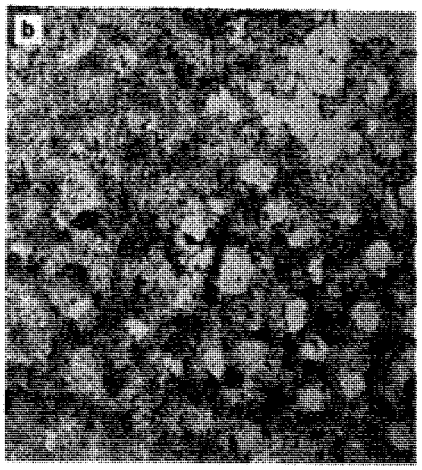


FIG. 5E



FIG. 5C

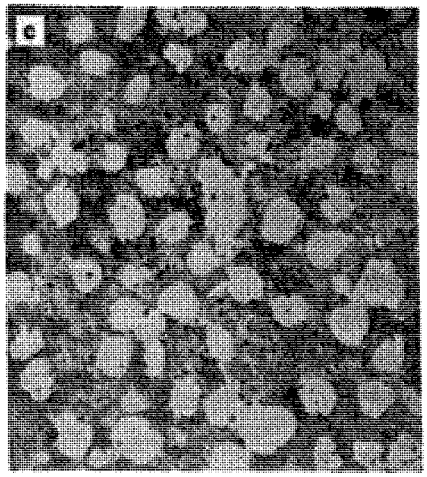
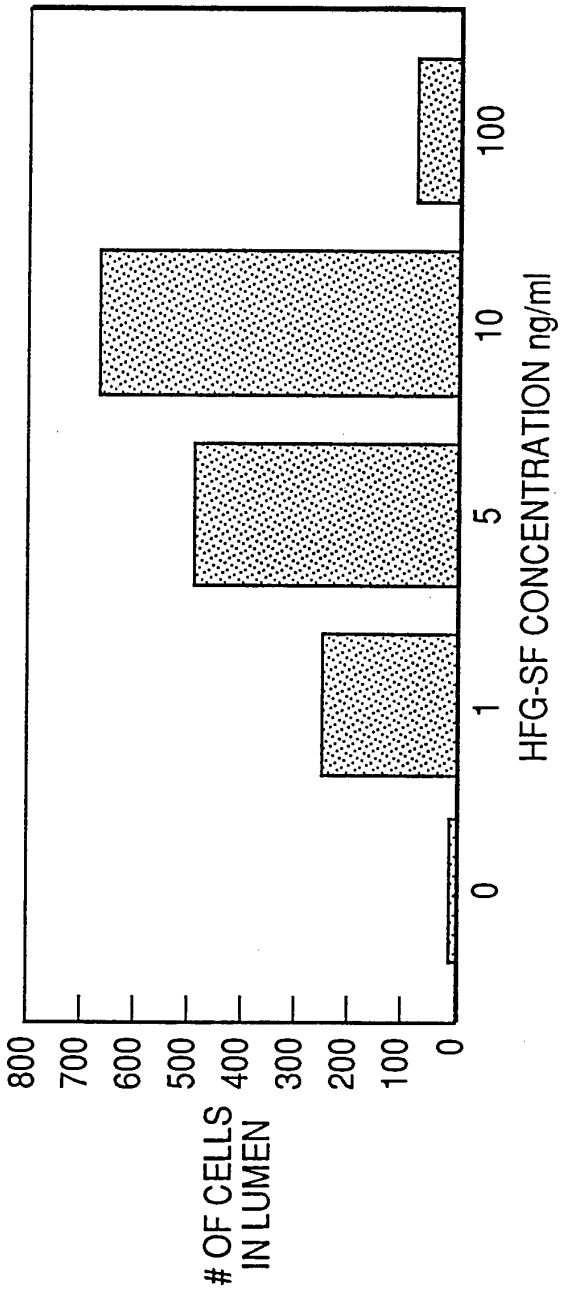


FIG. 5F



FIG. 6



GeneDX 1009, pg. 833

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 93/05952

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 G01N33/574; C12Q1/68; C07K15/00

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	G01N ; C12Q ; C07K

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
Y	WO,A,9 109 974 (THE UNITED STATES OF AMERICA) 11 July 1991 see the whole document ---	1-3
Y,P	SCIENCE vol. 257, 28 August 1992, pages 1258 - 1261 I.TSARFATY ET AL. 'The met Proto-Oncogene Receptor and Lumen Formation.' see the whole document ---	1-3

-/--

° Special categories of cited documents : [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
30 SEPTEMBER 1993	06. 10. 93
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	HITCHEN C.

Form PCT/ISA/210 (second sheet) (January 1985)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	INTERNATIONAL JOURNAL OF CANCER vol. 49, no. 3, 30 September 1991, pages 323 - 328 M.PRAT ET AL. 'The Receptor Encoded by the Human c-met Oncogene is Expressed in Hepatocytes, Epithelial Cells and Solid Tumours' cited in the application see the whole document	1-3
A	THE LANCET vol. 339, 18 January 1992, pages 139 - 143 I.BIECHE ET AL. 'Loss of heterozygosity on chromosome 7q and aggressive primary breast cancer' cited in the application see page 139 - page 140	1-3
A	ONCOGENE vol. 7, no. 1, January 1992, pages 3 - 7 C.S.COOPER 'The met oncogene : from detection by transfection to transmembrane receptor for hepatocate growth factor' see abstract	1-3
A	WO,A,8 910 412 (APPLIED BIOTECHNOLOGY, INC.) 2 November 1989 see abstract	1-3

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO.

US 9305952
SA 76404

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information. 30/09/93

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9109974	11-07-91	AU-B-	639532	29-07-93
		AU-A-	7140691	24-07-91
		EP-A-	0507868	14-10-92
		JP-T-	5504476	15-07-93
WO-A-8910412	02-11-89	AU-A-	3568289	24-11-89
		EP-A-	0412116	13-02-91
		JP-T-	3505964	26-12-91

EPO FORM P0479

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) **International Patent Classification** 6 :
C07H 21/02, 21/04, C12Q 1/68, G01N 33/53, C12P 21/00, C12N 15/63, 15/85, A61K 48/00

A1

(11) **International Publication Number:** **WO 95/19369**

(43) **International Publication Date:** 20 July 1995 (20.07.95)

(21) **International Application Number:** PCT/US95/00608

(22) **International Filing Date:** 17 January 1995 (17.01.95)

(30) **Priority Data:**

08/182,961	14 January 1994 (14.01.94)	US
08/373,799	17 January 1995 (17.01.95)	US

(71) **Applicant:** VANDERBILT UNIVERSITY [US/US]; 405 Kirkland Hall, Nashville, TN 37240 (US).

(72) **Inventors:** HOLT, Jeffrey, T.; 1708 Leaton Court, Franklin, TN 37064 (US). JENSEN, Roy, A.; 2701 Longwood Lane, Franklin, TN 37064 (US). PAGE, David, L.; 5905 Robert E. Lee Court, Nashville, TN 37215 (US). OBERMILLER, Patrice, S.; 5305 Frankliln Pike, Nashville, TN 37220 (US). ROBINSON-BENION, Cheryl, L.; 2105 Summit Avenue, Nashville, TN 37218 (US). THOMPSON, Marilyn, E.; 105 Southwood Park Place, Nashville, TN 37217 (US).

(74) **Agent:** LANQUIST, Edward, D., Jr.; Waddey & Patterson, 27th floor, L & C Tower, 401 Church Street, Nashville, TN 37219 (US).

(81) **Designated States:** AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
With amended claims and statement.

(54) **Title:** METHOD FOR DETECTION AND TREATMENT OF BREAST CANCER

(57) **Abstract**

The present invention provides a method of detecting and diagnosing pre-invasive breast cancer by identifying differentially expressed genes in early, pre-invasive breast cancer tissue. Differentially expressed genes can be used as genetic markers to indicate the presence of pre-invasive cancerous tissues. Microscopically directed tissue sampling techniques combined with differential display or differential screening of cDNA libraries are used to determine differential expression of genes in the early stages of breast cancer. Differential expression of genes in pre-invasive breast cancer tissue is confirmed by RT-PCR, nuclease protection assays and in-situ hybridization of ductal carcinoma in situ tissue RNA and control tissue RNA. The present invention also provides a method of screening for compounds that induce expression of the BRCA1 gene, whose product negatively regulates cell growth in both normal and malignant mammary epithlial cells. The present invention also relates to gene therapy method using this gene.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

DESCRIPTION

"METHOD FOR DETECTION AND TREATMENT OF BREAST CANCER"

TECHNICAL FIELD

The present invention relates generally to methods of detection and diagnosis of breast cancer and more particularly to a diagnostic method which relies on the identification of marker genes expressed in pre-invasive cancers by microscopically-directed cloning. Furthermore, this invention concerns the prevention, detection, and diagnosis of breast cancer by addressing the molecular events which occur during the earliest alterations in breast tissue.

The present invention also relates generally to methods of treatment of breast cancer, and more particularly to gene therapy methods and methods for screening compounds that induce expression of the BRCA1 gene product.

BACKGROUND ART

It will be appreciated by those skilled in the art that there exists a need for a more sensitive and less invasive method of early detection and diagnosis of breast cancer than those methods currently in use. Breast cancer presents inherent difficulties in regard to the ease with which it is detected and diagnosed. This is in contrast to detection of some other common cancers, including skin and cervical cancers, the latter of which is based on cytomorphologic screening techniques.

There have been several attempts to develop improved methods of breast cancer detection and diagnosis. In the attempts to improve methods of detection and diagnosis of breast cancer, numerous studies have searched for oncogene mutations, gene amplification, and loss of heterozygosity in invasive breast cancer (Callahan, et al., 1992; Cheickh, et al., 1992; Chen,et al, 1992; and, Lippman, et al, 1990). However, few studies of breast cancer have analyzed gene mutations and/or altered gene expression in ductal carcinoma in situ (DCIS). Investigators have demonstrated high levels of p53 protein in 13-40% of DCIS lesions employing a monoclonal antibody to p53, and subsequent sequencing demonstrated mutations in several cases (Poller et al, 1992). The neu/erbB2 gene appears to be amplified in a subset of DCIS lesions (Allred et al, 1992; Maguire et al, 1992). Histologic analysis of DCIS cases suggests that mutations and altered gene expression events, as well as changes in chromatin and

DNA content, occur predominantly in comedo DCIS (Böcker et al, 1992; Killeen et al, 1991; and, Komitowski et al, 1990), which has a rapid rate of local invasion and progression to metastasis. Thus, there are presently no reliable marker genes for non-comedo DCIS (NCDCIS, hereafter).

Cancer in humans appears to be a multi-step process which involves progression from pre-malignant to malignant to metastatic disease which ultimately kills the patient. Epidemiologic studies in humans have established that certain pathologic conditions are "pre-malignant" because they are associated with increased risk of malignancy. There is precedent for detecting and eliminating pre-invasive lesions as a cancer prevention strategy: dysplasia and carcinoma in-situ of the uterine cervix are examples of pre-malignancies which have been successfully employed in the prevention of cervical cancer by cytologic screening methods. Unfortunately, because the breast cannot be sampled as readily as cervix, the development of screening methods for breast pre-malignancy involves more complex approaches than cytomorphologic screening now currently employed to detect cervical cancer.

Pre-malignant breast disease is also characterized by an apparent morphological progression from atypical hyperplasias, to carcinoma in-situ (pre-invasive cancer) to invasive cancer which ultimately spreads and metastasizes resulting in the death of the patient. Careful histologic examination of breast biopsies has demonstrated intermediate stages which have acquired some of these characteristics but not others. Detailed epidemiological studies have established that different morphologic lesions progress at different rates, varying from atypical hyperplasia (with a low risk) to comedo ductal carcinoma-in-situ which progresses to invasive cancer in a high percentage of patients (London et al, 1991; Page et al, 1982; Page et al, 1985; Page et al, 1991; and Page et al, 1978). Family history is also an important risk factor in the development of breast cancer and increases the relative risk of these pre-malignant lesions (Dupont et al, 1985; Dupont et al, 1993; and, London et al, 1991). Of particular interest is non-comedo carcinoma-in-situ which is associated with a greater than ten-fold increased relative risk of breast cancer compared to control groups (Ottesen et al, 1992; Page et al, 1982). Two other reasons besides an increased relative risk support the concept that DCIS is pre-malignant: 1) When breast cancer occurs in

these patients it regularly occurs in the same region of the same breast where the DCIS was found; and 2) DCIS is frequently present in tissue adjacent to invasive breast cancer (Ottesen et al, 1992; Schwartz et al, 1992). For these reasons DCIS very likely represents a rate-limiting step in the development of invasive breast cancer in women.

DCIS (sometimes called intraductal carcinoma) is a group of lesions in which the cells have grown to completely fill the duct with patterns similar to invasive cancer, but do not invade outside the duct or show metastases at presentation. DCIS occurs in two forms: comedo DCIS and non-comedo DCIS. Comedo DCIS is often a grossly palpable lesion which was probably considered "cancer" in the 19th and early 20th century and progresses to cancer (without definitive therapy) in at least 50% of patients within three years (Ottesen et al, 1992; Page et al, 1982). Most of the molecular alterations which have been reported in pre-malignant breast disease have been observed in cases of comedo DCIS (Poller et al, 1993; Radford et al, 1993; and, Tsuda et al, 1993). Non-comedo DCIS is detected by microscopic analysis of breast aspirates or biopsies and is associated with a 10 fold increased risk of breast cancer, which corresponds to a 25-30% absolute risk of breast cancer within 15 years (Ottesen et al, 1992; Page et al, 1982; and, Ward et al, 1992).

Widespread application of mammography has changed the relative incidence of comedo and non-comedo DCIS such that NCDCIS now represents the predominant form of DCIS diagnosed in the United States (Ottesen et al, 1992; Page et al, 1982; and Pierce et al, 1992). Both forms of DCIS generally recur as invasive cancer at the same site as the pre-malignant lesion (without definitive therapy). The precursor lesions to DCIS are probably atypical ductal hyperplasia and proliferative disease without atypia which are associated with lower rates of breast cancer development, but show further increased risk when associated with a family history of breast cancer (Dupont et al, 1985; Dupont et al, 1989; Dupont et al, 1993; Lawrence, 1990; London et al, 1991; Page et al, 1982; Page et al, 1985; Page et al, 1991; Page et al, 1978; Simpson et al, 1992; Solin et al, 1991; Swain, 1992; Weed et al, 1990).

What is needed, then, is a sensitive method of detection and diagnosis of breast cancer when the cancerous cells are still in the pre-invasive stage. To illustrate the usefulness in early breast cancer detection of a marker gene and its encoded protein,

consider the dramatic impact that prostate specific antigen has had on early stage prostate cancer. This method of early detection and diagnosis of breast cancer is presently lacking in the prior art.

Breast cancer occurs in hereditary and sporadic forms. Recently the BRCA 1 gene has been cloned and shown to be mutated in kindreds with hereditary breast and ovarian cancer (Hall et al. 1990, Miki, Y. et al. 1994, Friedman et al. 1994, Castilla et al. 1994, Simard et al. 1994). Although 92% of families with two or more cases of early-onset breast cancer and two cases of ovarian cancer have germ-line mutations in BRCA 1 (Narod et al. in press), the gene has not been shown to be mutated in any truly sporadic case to date (Futreal et al. 1994). Despite the surprising paucity of somatically acquired mutations in sporadic breast cancer, it is still a likely tumor suppressor gene with a key role in breast epithelial cell biology. The BRCA 1 gene encodes a protein of 1863 amino acids with a predicted zinc finger domain observed in proteins which regulate gene transcription. Until the discovery of the function of the BRCA1 gene in conjucntion with the delopment of the present invention, the function was unknown.

DISCLOSURE OF THE INVENTION

Epidemiologic studies have established that NCDCIS of the breast is associated with a ten-fold increased risk of breast cancer (absolute risk of 25-30%). It seems likely that this pre-invasive lesion is a determinate precursor of breast cancer because the subsequent development of breast cancer is regularly in the same region of the same breast in which the NCDCIS lesion was found. Important aspects of the present invention concern isolated DNA segments and those isolated DNA segments inserted into recombinant vectors encoding differentially expressed marker genes in abnormal tissue, specifically in NCDCIS, as compared with those expressed in normal tissue, and the creation and use of recombinant host cells through the application of DNA technology, which express these differentially expressed marker genes (Sambrook et al, 1989).

Because there are no cell lines or animal models which clearly display known characteristics of pre-invasive breast disease, human breast tissue samples are essential

for studying pre-invasive breast disease. Using human tissue samples, we subsequently have developed a method for cDNA cloning from histologically identified lesions in human breast biopsies. We have used this method to clone genes which are differentially expressed in pre-invasive breast lesions such as NCDCIS lesions as compared to genes expressed in normal tissue. The differentially expressed genes detected in pre-invasive breast cancer are called marker genes. Identification of marker genes for pre-invasive breast disease provides improved methods for detection and diagnosis of pre-invasive breast cancer tissue, and further provides marker genes for studies of the molecular events involved in progression from pre-invasive to malignant breast disease.

Analysis of marker gene expression in NCDCIS presents the advantage that cancerous breast tissue at that stage is non-invasive. Detection and diagnosis of NCDCIS by means of differentially expressed marker genes compared to the same marker genes in normal breast tissue, would allow a greater ability to detect, prevent and treat the disease before it becomes invasive and metastasizes. The stage or intermediate condition of NCDCIS is a particularly good candidate for early intervention because it is 1) prior to any invasion and thus prior to any threat to life; 2) it is followed by invasive carcinoma in over 30% of cases if only treated by biopsy; and, 3) there is a long "window" of opportunity (4-8 years) approximately before invasive neoplasia occurs. Thus, NCDCIS is an ideal target for early diagnosis. While these morphologically defined intermediate endpoints have been widely accepted, progress in defining the molecular correlates of these lesions has been hampered by an inability to identify and sample them in a manner which would allow the application of molecular techniques.

Frozen tissue blocks from breast biopsies were used to construct and screen cDNA libraries prepared from NCDCIS tissue, normal breast tissue, breast cancer tissue, and normal human breast epithelial cells. Several cDNAs which were differentially expressed in human DCIS epithelial cells compared to normal breast epithelial cells were cloned and sequenced. One gene which is differentially expressed is the M2 subunit of RibRed which is expressed at low levels in human breast epithelial cells but at higher levels in 4 out of 5 DCIS tissue samples. It is presumed that the

altered morphologic appearance and determinant biologic behavior of DCIS results from altered expression of genes (such as RibRed) which is important in the induction of breast cancer in humans.

This invention, therefore, provides a method of detecting and diagnosing pre-invasive breast cancer by analyzing marker genes which are differentially expressed in non-comedo DCIS cells. Histopathologic studies have demonstrated that these morphologic patterns in breast tissue lead to invasive breast cancer in at least 20-30% of patients. The present method analyzes gene expression in normal, pre-malignant and malignant breast biopsies; and, it allows simultaneous comparison and cloning of marker genes which are differentially expressed in pre-invasive breast cancer. These marker genes can then be used as probes to develop other diagnostic tests for the early detection of pre-invasive breast cancer.

The present invention concerns DNA segments, isolatable from both normal and abnormal human breast tissue, which are free from total genomic DNA. The isolated DCIS-1 protein product is the regulatory element of the RibRed enzyme. This and all other isolatable DNA segments which are differentially expressed in preinvasive breast cancer can be used in the detection, diagnosis and treatment of breast cancer in its earliest and most easily treatable stages. As used herein, the term "abnormal tissue" refers to pre-invasive and invasive breast cancer tissue, as exemplified by collected samples of non-comedo or comedo DCIS tissues.

As used herein, the term "DNA segment" refers to a DNA molecule which has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding a differentially expressed protein (as measured by the expression of mRNA) in abnormal tissue refers to a DNA segment which contains differentially expressed-coding sequences in abnormal tissue as compared to those expressed in normal tissue, yet is isolated away from, or purified free from, total genomic DNA of Homo sapiens sapiens. Furthermore, a DNA segment encoding a BRCA1 protein refers to a DNA segment which contains BRCA1 coding sequences, yet is isolated away from, or purified free from, total genomic DNA of Homo sapiens sapiens. Included within the term "DNA segment", are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids,

phage, viruses, and the like.

Similarly, a DNA segment comprising an isolated or purified differentially expressed gene or comprising an isolated or purified BRCA1 gene refers to a DNA segment including differentially expressed coding sequences or BRCA1 coding sequences isolated substantially away from other naturally occurring genes or protein encoding sequences. In this respect, the term "gene" is used for simplicity to refer to a functional protein, polypeptide or peptide encoding unit. As will be understood by those in the art, this functional term includes both genomic sequences and cDNA sequences. "Isolated substantially away from other coding sequences" means that the gene of interest, in this case, any differentially expressed marker gene or the BRCA1 gene, forms the significant part of the coding region of the DNA segment, and that the DNA segment does not contain large portions of naturally-occurring coding DNA, such as large chromosomal fragments or other functional genes or cDNA coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

In particular embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating DNA sequences which encode differentially expressed genes in pre-invasive breast cancer, each which includes within its amino acid sequence an amino acid sequence in accordance with SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7, all seq id no:s 1-7 are derived from non-comedo DCIS samples from Homo sapiens sapiens. In other particular embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating DNA sequences which encode the M2 subunit of human RibRed that includes within its amino acid sequence the similar amino acid sequence of hamster RibRed corresponding to the M2 subunit of hamster RibRed.

In certain embodiments, the invention concerns isolated DNA segments and recombinant vectors which partially or wholly encode a protein or peptide that includes within its amino acid sequence an amino acid sequence essentially as partially or wholly encoded, respectively, by SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7. Naturally, where the DNA segment or vector encodes a full length differentially expressed protein, or is intended

for use in expressing the differentially expressed protein, the most preferred sequences are those which are essentially as set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7 and which encode a protein that exhibits differential expression, e.g., as may be determined by the differential display or differential sequencing assay, as disclosed herein.

The term "a sequence essentially as set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7" means that the sequence substantially corresponds to a portion of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7, respectively, and has relatively few nucleotides which are not identical to, or a biologically functional equivalent of, the nucleotides of the respective SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7. The term "biologically functional equivalent" is well understood in the art and is further defined in detail herein, for example see pages 24 through 25. Accordingly, sequences which have between about 70% and about 80%; or more preferably, between about 81% and about 90%; or even more preferably, between about 91% and about 99%; of amino acids which are identical or functionally equivalent to the amino acids of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7 will be sequences which are "essentially as set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7", respectively.

In particular embodiments, the invention concerns a drug screening method and a gene therapy method that use isolated DNA segments and recombinant vectors incorporating DNA sequences which encode a protein that includes within its amino acid sequence an amino acid sequence in accordance with SEQ ID NO:49, SEQ ID NO:49 derived from breast tissue from Homo sapiens. In other particular embodiments, the invention concerns isolated DNA sequences and recombinant DNA vectors incorporating DNA sequences wich encode a protein taht includes with its amino acid sequence the amino acid sequence of the BRCA1 gene product from human breast tissue.

In certain embodiments, the invention concerns methods using isolated DNA segments and recombinant vectors which partially or wholly encode a protein or peptide that includes within its amino acid sequence an amino acid sequence essentially as set forth in SEQ ID NO:49. Naturally, where the DNA segment or vector encodes a full length BRCA1 protein, or is intended for use in expressing the BRCA1 protein, the most preferred sequences are those which are essentially as set forth in SEQ ID NO:47 and which encode a protein that retains activity as a negative growth regulator in human breast cells, as may be determined by antisense assay, as disclosed herein.

The term "a sequence essentially as set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7" means that the sequence substantially corresponds to a portion of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7, respectively, and has relatively few nucleotides which are not identical to, or a biologically functional equivalent of, the nucleotides of the respective SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7. The term "biologically functional equivalent" is well understood in the art and is further defined in detail herein, for example see pages 24 through 25. Accordingly, sequences which have between about 70% and about 80%; or more preferably, between about 81% and about 90%; or even more preferably, between about 91% and about 99%; of amino acids which are identical or functionally equivalent to the amino acids of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7 will be sequences which are "essentially as set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7", respectively.

The term "a sequence essentially as set forth in SEQ ID NO:49" means that the sequence substantially corresponds to a portion of SEQ ID NO:49 and has relatively few amino acids which are not identical to, or a biologically functional equivalent of, the nucleotides of SEQ ID NO:49. The term "biologically functional equivalent" is well understood in the art and is further defined in detail herein, for example see pages 24 through 25. Accordingly, sequences which have between about 70% and about 80%; or more preferably, between about 81% and about 90%; or even more

preferably, between about 91% and about 99%; of amino acids which are identical or functionally equivalent to the amino acids of SEQ ID NO:49 will be sequences which are "essentially as set forth in SEQ ID NO:49".

In certain other embodiments, the invention concerns isolated DNA segments and recombinant vectors that include within their sequence a nucleic acid sequence essentially as set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7. The term "essentially as set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7" is used in the same sense as described above and means that the nucleic acid sequence substantially corresponds to a portion of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7, respectively, and has relatively few codons which are not identical, or functionally equivalent, to the codons of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7, respectively. Again, DNA segments which encode proteins exhibiting differential expression will be most preferred. The term "functionally equivalent codon" is used herein to refer to codons that encode the same amino acid, such as the six codons for arginine or serine, and also refers to codons that encode biologically equivalent amino acids (see Figure 8).

In certain other embodiments, the invention concerns a method for screening drugs and a gene therapy method which involve the use of isolated DNA segments and recombinant vectors that include within their sequence a nucleic acid sequence essentially as set forth in SEQ ID NO:47 and SEQ ID NO:48. The term "essentially as set forth in SEQ ID NO:47 and SEQ ID NO:48" is used in the same sense as described above and means that the nucleic acid sequence substantially corresponds to a portion of SEQ ID NO:47 and SEQ ID NO:48 respectively, and has relatively few codons which are not identical, or functionally equivalent, to the codons of SEQ ID NO:47 and SEQ ID NO:48, respectively. Again, DNA segments which encode proteins exhibiting the negative regulatory activity of the BRCA1 will be most preferred. The term "functionally equivalent codon" is used herein to refer to codons

that encode the same amino acid, such as the six codons for arginine or serine, and also refers to codons that encode biologically equivalent amino acids (see Figure 8).

It will also be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5' or 3' sequences, and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the sequence meets the criteria set forth above, including the maintenance of biological protein activity where protein expression is concerned. The addition of terminal sequences particularly applies to nucleic acid sequences which may, for example, include various non-coding sequences flanking either of the 5' or 3' portions of the coding region or may include various internal sequences, i.e., introns, which are known to occur within genes.

Excepting intronic or flanking regions, and allowing for the degeneracy of the genetic code, sequences which have between about 20% and about 50%; or more preferably, between about 50% and about 70%; or even more preferably, between about 70% and about 99%; of nucleotides which are identical to the nucleotides of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7 will be sequences which are "essentially as set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7", respectively. Sequences which are essentially the same as those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7 may also be functionally defined as sequences which are capable of hybridizing to a nucleic acid segment containing the complement of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7, respectively, under relatively stringent conditions. Suitable relatively stringent hybridization conditions will be well known to those of skill in the art (Sambrook et al, 1989).

Excepting intronic or flanking regions, and allowing for the degeneracy of the genetic code, sequences which have between about 20% and about 50%; or more preferably, between about 50% and about 70%; or even more preferably, between about 70% and about 99%; of nucleotides which are identical to the nucleotides of SEQ ID NO:47 and SEQ ID NO:48 will be sequences which are "essentially as set forth in

SEQ ID NO:47 and SEQ ID NO:48", respectively. Sequences which are essentially the same as those set forth in SEQ ID NO:47 and SEQ ID NO:48 may also be functionally defined as sequences which are capable of hybridizing to a nucleic acid segment containing the complement of SEQ ID NO:47 and SEQ ID NO:48, respectively, under relatively stringent conditions. Suitable relatively stringent hybridization conditions will be well known to those of skill in the art (Sambrook et al, 1989).

It is also important to understand the molecular events which lead to progression from pre-invasive to invasive breast cancer. Breast cancer is a disease that is presumed to involve a series of genetic alterations that confer increasing growth independence and metastatic capability on somatic cells. Identifying the molecular events that lead to the initial development of a neoplasm is therefore critical to understanding the fundamental mechanisms by which tumors arise and to the selection of optimal targets for gene therapy and chemopreventive agents. As intermediate endpoints in neoplastic development, some pre-malignant breast lesions represent important, and possibly rate-limiting steps in the progression of human breast cancer, and careful epidemiological studies have established the relative risk for breast cancer development for specific histologic lesions. In particular, invasive breast cancer develops in the region of the previous biopsy site in at least 25-30% of patients following diagnosis of non-comedo DCIS providing strong evidence that this pre-malignant lesion is a determinant event in breast cancer progression. While these morphologically defined intermediate endpoints have been widely accepted, progress in defining the molecular correlates of these lesions has been hampered by an inability to identify and sample them in a manner which would allow the application of molecular techniques.

The present invention includes a comparison of gene expression between multiple breast tissue biopsy samples as a means to identify differentially expressed genes in pre-malignant breast disease compared with normal breast tissue. These genetic markers should be extremely useful reagents for early diagnosis of breast cancer, and for the delineation of molecular events in progression of breast cancer.

Identification of gene markers which are expressed in the majority of pre-invasive breast cancer tissue samples involves cDNA library preparation from both

normal and abnormal tissue. This is followed by either a modified differential display method or a differential screening method to identify differential expression of genes which is subsequently confirmed by RT-PCR, nuclease protection assays and in situ hybridization of DCIS tissue RNA and control tissue RNAs (Sambrook et al, 1989). Use of genetic engineering methods can bias the screening to specifically identify genes whose encoded proteins are secreted or are present at the cell surface, in order to find proteins which will be useful markers for diagnostic blood tests (secreted proteins) or for diagnostic imaging studies (cell surface proteins).

Thus, the method of the present invention begins with the collection of at least one tissue sample by a microscopically-directed collection step in which a punch biopsy is obtained exclusively from abnormal tissue which exhibits histological or cytological characteristics of pre-invasive breast cancer. Preferably, the sample site will be an isolatable tissue structure, such as ductal epithelial cells from pre-invasive breast cancer tissue. The mRNA is purified from the sample. Then, a cDNA library is prepared from the mRNA purified from the abnormal tissue sample (Sambrook et al, 1989).

A normal tissue sample is then obtained from the patient, using a sample site from an area of tissue which does not exhibit histological or cytological characteristics of pre-invasive cancer. A cDNA library is also prepared from this normal tissue sample.

The abnormal tissue cDNA library can then be compared with the normal tissue cDNA library by differential display or differential screening to determine whether the expression of at least one marker gene in the abnormal tissue sample is different from the expression of the same marker gene in the normal tissue sample.

Further diagnostic steps can be added to the method by cloning the marker gene using sequence-based amplification to create a cloned marker gene which can then be DNA-sequenced in order to derive the protein sequence. The protein sequence is then used to generate antibodies which will recognize these proteins by antibody recognition of the antigen. The presence of the antibody-recognized antigen can then be detected by means of conventional medical diagnostic tests.

This invention also includes methods of screening for compounds and gene therapy methods using the BRCA1 gene. BRCA1 mRNA is expressed at 5-10 fold higher levels in normal mammary tissue than in invasive breast cancer samples. Having demonstrated that mRNA expression levels of BRCA1 are higher in normal mammary cells than in cancer cells, antisense methods were used to test the hypothesis that BRCA1 expression inhibits cell growth. These tests showed that diminished expression of BRCA1 increased the proliferative rate of breast cells.

An object of the present invention, then, is to provide a method of early detection of pre-invasive breast cancer in human tissue.

It is a further object of this invention to identify early marker genes for pre-invasive breast disease which can be used in screening methods for early pre-invasive breast cancer.

It is also an object of this invention to produce a cDNA library from pre-invasive breast cancer tissue resulting in a permanent genetic sample of that pre-invasive breast cancer tissue.

It is also an object of this invention to provide a drug or biological screening method using the BRCA 1 promoter region and gene therapy method using the BRCA 1 gene.

List of Abbreviations

TPA	Phorbol 12-myristate 13-acetate
MCF-7	An immortalized cell line derived from a metastasis of human breast cancer
HMEC	A primary (non-immortalized) cell line derived from breast epithelial cells obtained during reduction mammoplasty
DCIS	Ductal Carcinoma-in-situ
NCDC	Non-Comedo Ductal Carcinoma in situ
cDNA	Complementary DNA obtained from an RNA template
DNA	Deoxyribonucleic Acid
RT-PCR	Reverse Transcriptase-Polymerase Chain Reaction
RibRed	Ribonucleotide Reductase

Fig. 1 shows Table I which describes anatomic lesion types in the human breast with pre-malignant implication.

Fig. 2 shows a model for pre-malignant conditions, highlighting magnitude of risk for progression to clinical malignancy.

Fig. 3 contains color photos of DCIS tissue, before (upper left panel) and after microscopically-directed excisional punch biopsy (upper right panel). The lower panels show tissue samples of normal breast tissue (lower left panel), and invasive breast cancer (lower right panel).

Fig. 4 shows expression of collagen III mRNA in tissue mRNA samples, analyzed by RNase protection assay methods.

Fig. 5 shows differential display of cDNAs obtained from patient tissue samples and controls.

Fig. 6 shows a comparison of the sequence between DCIS-1 and the human and hamster genes.

Fig. 7 shows expression of DCIS-1 mRNA in tissue mRNA samples analyzed by RNase protection assay as described in the legend to Figure 4.

Fig. 8 is Table II which displays the genetic code.

Fig. 9 is a Table which lists differentially expressed marker genes.

Figs. 10A and 10B shows expression of BRCA1 mRNA during breast cancer progression by PCR detection and nuclease protection assay, respectively.

Figs. 11A and 11B is a comparison of BRCA1 expression in normal breast and invasive breast cancer using nuclease protection assay of RNA, respectively.

Figs. 12A, 12B, and 12C show that antisense inhibition of BRCA1 accelerates mammary cell proliferation.

Figs. 13A and 13B includes a Northern blot of mRNA and nuclear runon studies that show that ribonucleotide reductase M2 mRNA is cell cycle regulated in MCF-7 cells.

Fig. 14 includes a nuclease protection assay that shows that antisense inhibition of BRCA1 in human mammary cells decreases BRCA1 mRNA and increases ribonucleotide reductase mRNA.

UTILITY STATEMENT

The detection of differentially expressed genes in pre-invasive breast tissue, specifically in non-comedo ductal carcinoma in situ as compared to genes expressed in normal tissue, is useful in the diagnosis, prognosis and treatment of human breast cancer. Such differentially expressed genes are effective marker genes indicating the significantly increased risk of breast cancer in a patient expressing these differentially expressed marker genes. These marker genes are useful in the detection, early diagnosis, and treatment of breast cancer in humans.

The discovery of the function of the BRCA 1 gene has broad utility including, in the present invention, development of methods to treat familial and sporadic breast cancers as well as screen for therapeutic drugs through production of important indicator compounds.

ACTIVITY STATEMENT

Of the differentially expressed genes described in this invention, DCIS-1 encodes a gene similar to the M2 subunit of hamster ribonucleotide reductase. The M2 subunit of ribonucleotide reductase (RibRed, hereafter) is responsible for regulation of RibRed. The differential levels of expression of the marker genes described in this invention (Seq ID No.s 1-7), indicate genetic changes which have been linked to the presence of pre-invasive breast cancer.

The BRCA1 gene (Seq. ID No. 47) is differentially expressed in invasive breast cancer cells. The BRCA1 gene product is a negative regulator of mammary cell proliferation which is expressed at diminished levels in sporadic breast cancer.

BEST MODE FOR CARRYING OUT THE INVENTION

For the purposes of the subsequent description, the following definitions will be used:

Nucleic acid sequences which are "complementary" are those which are capable of base-pairing according to the standard Watson-Crick complementarity rules. That is, that the larger purines will always base pair with the smaller pyrimidines to form only combinations of Guanine paired with Cytosine (G:C) and Adenine paired with

either Thymine (A:T) in the case of DNA or Adenine paired with Uracil (A:U) in the case of RNA.

"Hybridization techniques" refer to molecular biological techniques which involve the binding or hybridization of a probe to complementary sequences in a polynucleotide. Included among these techniques are northern blot analysis, southern blot analysis, nuclease protection assay, etc.

"Hybridization" and "binding" in the context of probes and denatured DNA are used interchangeably. Probes which are hybridized or bound to denatured DNA are aggregated to complementary sequences in the polynucleotide. Whether or not a particular probe remains aggregated with the polynucleotide depends on the degree of complementarity, the length of the probe, and the stringency of the binding conditions. The higher the stringency, the higher must be the degree of complementarity and/or the longer the probe.

"Probe" refers to an oligonucleotide or short fragment of DNA designed to be sufficiently complementary to a sequence in a denatured nucleic acid to be probed and to be bound under selected stringency conditions.

"Label" refers to a modification to the probe nucleic acid that enables the experimenter to identify the labeled nucleic acid in the presence of unlabeled nucleic acid. Most commonly, this is the replacement of one or more atoms with radioactive isotopes. However, other labels include covalently attached chromophores, fluorescent moeities, enzymes, antigens, groups with specific reactivity, chemiluminescent moeities, and electrochemically detectable moeities, etc.

"Marker gene" refers to any gene selected for detection which displays differential expression in abnormal tissue as opposed to normal tissue. It is also referred to as a differentially expressed gene.

"Marker protein" refers to any protein encoded by a "marker gene" which protein displays differential expression in abnormal tissue as opposed to normal tissue.

"Tissuemizer" describes a tissue homogenization probe.

"Abnormal tissue" refers to pathologic tissue which displays cytologic, histologic and other defining and derivative features which differ from that of normal

tissue. This includes in the case of abnormal breast tissue, among others, pre-invasive and invasive neoplasms.

"Normal tissue" refers to tissue which does not display any pathologic traits.

"PCR technique" describes a method of gene amplification which involves sequenced-based hybridization of primers to specific genes within a DNA sample (or library) and subsequent amplification involving multiple rounds of annealing, elongation and denaturation using a heat-stable DNA polymerase.

"RT-PCR" is an abbreviation for reverse transcriptase-polymerase chain reaction. Subjecting mRNA to the reverse transcriptase enzyme results in the production of cDNA which is complementary to the base sequences of the mRNA. Large amounts of selected cDNA can then be produced by means of the polymerase chain reaction which relies on the action of heat-stable DNA polymerase produced by Thermus aquaticus for its amplification action.

"Microscopically-directed" refers to the method of tissue sampling by which the tissue sampled is viewed under a microscope during the sampling of that tissue such that the sampling is precisely limited to a given tissue type, as the investigator requires. Specifically, it is a collection step which involves the use of a punch biopsy instrument. This surgical instrument is stereotactically manually-directed to harvest exclusively from abnormal tissue which exhibits histologic or cytologic characteristics of pre-invasive cancer. The harvest is correlated with a companion slide, stained to recognize the target tissue.

"Differential display" describes a method in which expressed genes are compared between samples using low stringency PCR with random oligonucleotide primers.

"Differential screening" describes a method in which genes within cDNA libraries are compared between two samples by differential hybridization of cDNAs to probes prepared from each library.

"Nuclease protection assay" refers to a method of RNA quantitation which employs strand specific nucleases to identify specific RNAs by detection of duplexes.

"Differential expression" describes the phenomenon of differential genetic expression seen in abnormal tissue in comparison to that seen in normal tissue.

"Isolatable tissue structure" refers to a tissue structure which when visualized microscopically or otherwise is able to be isolated from other different surrounding tissue types.

"In situ hybridization of RNA" refers to the use of labeled DNA probes employed in conjunction with histological sections on which RNA is present and with which the labeled probe can hybridize allowing an investigator to visualize the location of the specific RNA within the cell.

"Comedo DCIS cells" refers to cells comprising an in situ lesion with the combined features of highest grade DCIS.

"Non-comedo DCIS cells" refers to cells of DCIS lesions without comedo features.

"Cloning" describes separation and isolation of single genes.

"Sequencing" describes the determination of the specific order of nucleic acids in a gene or polynucleotide.

The present invention provides a method for detecting and diagnosing cancer by analyzing marker genes which are differentially expressed in early, pre-invasive breast cancer, specifically in non-comedo DCIS cells. Our histopathologic studies have demonstrated that certain morphologic patterns in breast tissue are pre-malignant, leading to invasive breast cancer in at least 20-30% of patients. We have developed a new method for analyzing gene expression in normal, pre-malignant and malignant breast biopsies which allows simultaneous comparison and cloning of marker genes which are differentially expressed in pre-invasive breast cancer. These marker genes (which appear as differentially expressed genes in pre-invasive breast cancer) can be used as probes to develop diagnostic tests for the early detection of pre-invasive breast cancer (Sambrook, 1989).

The present invention thus comprises a method of identification of marker genes which are expressed in the majority of pre-invasive breast cancer tissue samples. It involves cDNA library preparation followed by a modified differential display method. Use of genetic engineering methods (Sambrook, 1989) can bias the screening to specifically identify genes whose encoded proteins are secreted or are present at the cell

surface, in order to find proteins which will be useful markers for diagnostic blood tests (secreted proteins) or for diagnostic imaging studies (cell surface proteins).

Naturally, the present invention also encompasses DNA segments which are complementary, or essentially complementary, to the sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:47 and SEQ ID NO:48. Nucleic acid sequences which are "complementary" are those which are capable of base-pairing according to the standard Watson-Crick complementarity rules. As used herein, the term "complementary sequences" means nucleic acid sequences which are substantially complementary, as may be assessed by the same nucleotide comparison set forth above, or as defined as being capable of hybridizing to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:47 and SEQ ID NO:48 under relatively stringent conditions such as those described herein.

The nucleic acid segments of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, nucleic acid fragments may be prepared which include a short stretch complementary to SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:47 and SEQ ID NO:48, such as about 10 nucleotides, and which are up to 10,000 or 5,000 base pairs in length, with segments of 500 being preferred in most cases. DNA segments with total lengths of about 1,000, 500, 200, 100 and about 50 base pairs in length are also contemplated to be useful.

It will also be understood that this invention is not limited to the particular nucleic acid and amino acid sequences of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:47, SEQ ID NO:48, and SEQ ID NO:49. Recombinant vectors and isolated DNA

segments may therefore variously include the differentially expressed coding regions or the BRCAI coding regions themselves, coding regions bearing selected alterations or modifications in the basic coding region, or they may encode larger polypeptides which nevertheless include differentially expressed-coding regions and the BRCA1 coding regions or may encode biologically functional equivalent proteins or peptides which have variant amino acids sequences.

The DNA segments of the present invention encompass biologically functional equivalent differentially expressed proteins and peptides biologically functional equivalent proteins of BRCA1. Such sequences may arise as a consequence of codon redundancy and functional equivalency which are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally equivalent proteins or peptides may be created via the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques, e.g., to introduce improvements to the antigenicity of the protein or to test site-directed mutants or others in order to examine carcinogenic activity of the differentially expressed marker genes at the molecular level.

If desired, one may also prepare fusion proteins and peptides, e.g., where the differentially expressed marker gene coding regions are aligned within the same expression unit with other proteins or peptides having desired functions, such as for purification or immunodetection purposes (e.g., proteins which may be purified by affinity chromatography and enzyme label coding regions, respectively).

Recombinant vectors form important further aspects of the present invention. Particularly useful vectors are contemplated to be those vectors in which the coding portion of the DNA segment is positioned under the control of a promoter. The promoter may be in the form of the promoter which is naturally associated with a RIBRED gene, e.g., in human cells, as may be obtained by isolating the 5' non-coding sequences located upstream of the coding segment or exon, for example, using recombinant cloning and/or PCR technology, in connection with the compositions disclosed herein.

In other embodiments, it is contemplated that certain advantages will be gained by positioning the coding DNA segment under the control of a recombinant, or heterologous, promoter. As used herein, a recombinant or heterologous promoter is intended to refer to a promoter that is not normally associated with a differentially expressed marker gene or the BRCA1 gene in its natural environment. Such promoters may include MMTV promoters normally associated with other genes, and/or promoters isolated from any other bacterial, viral, eukaryotic, or mammalian cell. Naturally, it will be important to employ a promoter that effectively directs the expression of the DNA segment in the cell type chosen for expression. The use of promoter and cell type combinations for protein expression is generally known to those of skill in the art of molecular biology, for example, see Sambrook *et al.* (1989). The promoters employed may be constitutive, or inducible, and can be used under the appropriate conditions to direct high level expression of the introduced DNA segment, such as is advantageous in the large-scale production of recombinant proteins or peptides. Appropriate promoter systems contemplated for use in high-level expression include, but are not limited to appropriate bacterial promoters.

As mentioned above, in connection with expression embodiments to prepare recombinant differentially expressed marker gene encoded proteins and peptides, it is contemplated that longer DNA segments will most often be used, with DNA segments encoding the entire differentially expressed protein or subunit being most preferred. However, it will be appreciated that the use of shorter DNA segments to direct the expression of differentially expressed peptides or epitopic core regions, such as may be used to generate anti-marker protein antibodies, also falls within the scope of the invention (Harlow et al, 1988).

DNA segments which encode peptide antigens from about 15 to about 50 amino acids in length, or more preferably, from about 15 to about 30 amino acids in length are contemplated to be particularly useful. The C terminus of proteins provide an excellent region for peptide antigen recogition (Harlow et al, 1988). DNA segments encoding peptides will generally have a minimum coding length in the order of about 45 to about 147, or to about 90 nucleotides. DNA segments encoding partial length peptides may have a minimum coding length in the order of about 50 nucleotides for

a polypeptide in accordance with seq id no:3, or about 264 nucleotides for a polypeptide in accordance with SEQ ID NO: 1.

In addition to their use in directing the expression of the differentially expressed marker proteins, the nucleic acid sequences disclosed herein also have a variety of other uses. For example, they also have utility as probes or primers in nucleic acid hybridization embodiments. As such, it is contemplated that oligonucleotide fragments corresponding to the sequences of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7 for stretches of between about 10 to 15 nucleotides and about 20 to 30 nucleotides will find particular utility. Longer complementary sequences, e.g., those of about 40, 50, 100, 200, 500, 1000, and even up to full length sequences of about 2,000 nucleotides in length, will also be of use in certain embodiments.

The ability of such nucleic acid probes to specifically hybridize to differentially expressed marker gene sequences will enable them to be of use in detecting the presence of complementary sequences in a given sample. However, other uses are envisioned, including the use of the sequence information for the preparation of mutant species primers, or primers for use in preparing other genetic constructions.

Nucleic acid molecules having stretches of 20, 30, 50, or even of 500 nucleotides or so, complementary to SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7 are particularly contemplated as hybridization probes for use in, e.g., Southern and Northern blotting. This would allow differentially expressed structural or regulatory genes to be analyzed, both in patients and sample tissue from pre-invasive and invasive breast tissue. The total size of fragment, as well as the size of the complementary stretch(es), will ultimately depend on the intended use or application of the particular nucleic acid segment. Smaller fragments will generally find use in hybridization embodiments, wherein the length of the complementary region may be varied, such as between about 10 and about 100 nucleotides, but larger complementary stretches of up to about 300 nucleotides may be used, according to the length complementary sequences one wishes to detect.

Nucleic Acid Hybridization

The use of a hybridization probe of about 10 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having complementary sequences over stretches greater than 10 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of 15 to 20 nucleotides, or even longer where desired.

Hybridization probes may be selected from any portion of any of the sequences disclosed herein. All that is required is to review the sequences set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7 and to select any continuous portion of one of the sequences, from about 10 nucleotides in length up to and including the full length sequence, that one wishes to utilise as a probe or primer. The choice of probe and primer sequences may be governed by various factors, such as, by way of example only, one may wish to employ primers from towards the termini of the total sequence, or from the ends of the functional domain-encoding sequences, in order to amplify further DNA; one may employ probes corresponding to the entire DNA, or to the 5' region, to clone marker-type genes from other species or to clone further marker-like or homologous genes from any species including human; and one may employ randomly selected, wild-type and mutant probes or primers with sequences centered around the RibRed M2 subunit encoding sequence to screen DNA samples for differentially expressed levels of RibRed, such as to identify human subjects which may be expressing differential levels of RibRed and thus may be susceptible to breast cancer.

The process of selecting and preparing a nucleic acid segment which includes a sequence from within SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7 may alternatively be described as "preparing a nucleic acid fragment". Of course, fragments may also be obtained by other techniques such as, e.g., by mechanical shearing or by restriction enzyme digestion. Small nucleic acid segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, as is commonly

practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such as the PCR technology of U.S. Patent 4,603,102 (incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular biology.

Accordingly, the nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of differentially expressed marker genes or cDNAs. Depending on the application envisioned, one will desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, e.g., one will select relatively low salt and\or high temperature conditions, such as provided by 0.02M-0.15M NaCl at temperatures of 50°C to 70°C. Such selective conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating specific differentially expressed marker genes.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template or where one seeks to isolate marker gene sequences from related species, functional equivalents, or the like, less stringent hybridization conditions will typically be needed in order to allow formation of the heteroduplex. In these circumstances, one may desire to employ conditions such as 0.15M-0.9M salt, at temperatures ranging from 20°C to 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as increased temperature. Thus, hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

In certain embodiments, it will be advantageous to employ nucleic acid sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of giving a detectable signal. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmental undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known which can be employed to provide a means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

In general, it is envisioned that the hybridization probes described herein will be useful both as reagents in solution hybridization as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to specific hybridization with selected probes under desired conditions. The selected conditions will depend on the particular circumstances based on the particular criteria required (depending, for example, on the G+C contents, type of target nucleic acid, source of nucleic acid, size of hybridization probe, etc.). Following washing of the hybridized surface so as to remove nonspecifically bound probe molecules, specific hybridization is detected, or even quantified, by means of the label. (Sambrook et al, 1989).

In a preferred embodiment of the method, certain preliminary procedures are necessary to prepare the sample tissue and the probes before the detection of differential expression of marker genes in abnormal tissue as compared to that in normal tissue can be accomplished.

SAMPLE PREPARATION

RNA purification

RNA was isolated from frozen tissue samples by mincing of microdisected frozen tissue fragments with a razor blade and then adding 800 microliter of 5.6M

guanidinium to increase mixing, followed by a 30 second microcentrifuge centrifugation at 14,000 rpm to remove particulate matter. The supernatant was then removed and the viscosity was reduced by multiple aspirations through a 22 gauge needle and then 200 ul of chloroform was added and the sample was incubated on ice for 15 minutes (during this time the sample was vortexed multiple times). Following incubation with chloroform, the sample was centrifuged for 15 minutes at 14,000 rpm and the aqueous layer was removed and ethanol precipitated. This extraction method produces RNA which is primarily derived from cells of epithelial origin. In order to obtain RNA samples which presumably includes RNA derived from these stromal cells; the particulate material (remaining in the pellet from the 30 second centrifugation) was homogenized with a tissuemizer, washed with PBS, treated with collagenase at 37°C for 30 minutes, sonicated, extracted with phenol/chloroform and ethanol precipitated.

cDNA libraries were constructed in lambda phage using polyA-selected mRNA from the following samples; cultured human breast epithelial cells, tissue from three reduction mammoplasty patients, tissue from three DCIS patients, and tissue from one DCIS patient (patient #10) that showed a focus of microinvasion adjacent to an area of DCIS. Multiple punches were needed to obtain sufficient RNA for polyA selection and library construction. 200 ug of total RNA was obtained by pooling 20 punches from normal breast tissue (reduction mammoplasty samples) and 5-8 punches from DCIS lesions, presumably reflecting the greater cellularity of the DCIS samples. cDNA libraries were constructed by first and second strand cDNA synthesis followed by the addition of directional synthetic linkers (ZAP-cDNA Synthesis Kit, Stratagene, La Jolla, California). The Xho I-Eco Rl linkered cDNA was then ligated into lambda arms, packaged with packaging extracts, and then used to infect XL1-blue bacteria resulting in cDNA libraries.

PROBE PREPARATION

The collagen III probe employed for nuclease protection assays was constructed by subcloning the 208 bp Hinc II-Pst I fragment from the 3' untranslated region of the human type III procollagen gene into pGem4Z. This region of the human procollagen III gene was obtained by PCR amplification of published sequence (Ala-Kokko et al, 1989) followed by restriction with Hinc II and Pst I. For a control probe to assure

equal loading and recovery of RNA, we used a T7 polymerase-generated probe for human glyceraldehyde phosphate dehydrogenase (GADP) which protects a 140 bp Sac I-Xba I fragment; (a generous gift from Janice Nigro, Vanderbilt University). Probe DCIS-1 was generated by linearizing the rescued plasmid with Pvu II, which should generate a 200 bp protected fragment. RNase protection assays were performed with 1 ug of unselected RNA and the above-cited probes using the methods we have reported previously (Holt, 1993).

Differential Display-based cloning of cDNAs:

Rescued cDNA library samples were used as templates for low stringency PCR with the either a pair of 25 bp primers or an anchored 14 bp primer paired with a random 25 bp primer. Random 25 bp primers were generated by a computer-based algorithm (Jotte and Holt, unpublished). Samples were denatured for two minutes at 95°C followed by 40 cycles, each cycle consisting of denaturation for 1 minute at 94°C., annealing for 2 minutes at 25°C., and extension for 1 minute at 72°C. The samples were then run on an 6% non-denaturing polyacrylamide gel, which was dried and autoradiographed. Specific bands were excised then reamplified with the same primers used for their generation. Specificity was confirmed on 6% polyacrylamide gel, and samples were purified by ethanol precipitation of the remainder of the PCR reaction. Fragments were then individually cloned into SrfI cut vectors by standard methods using PCR-Script™SK(+) Cloning Kit (Stratagene, LaJolla, California) and then sequenced.

EXAMPLE 1

Studies showing Increased Risk of Breast Cancer in Patients with DCIS

Since the 1970's, studies of pre-invasive lesions associated with the development of breast cancer have been undertaken in an attempt to refine histologic and cytologic criteria for the hyperplastic lesions analogous to those of the uterine cervix and colon. Because of the availability of tissue from breast biopsies done many years previously, cohorts of women who underwent breast biopsies 15 to 20 years ago, can be studied to determine the risk for development of breast cancer attributable to specific lesions.

Many concurrent studies evaluating lesions associated with cancer at time of cancer diagnosis led the way in pointing out lesions of potential interest (Wellings et al, 1975). Hopefully, these intermediate stages in cancer development will serve to provide indicators of breast cancer development sufficiently precise to guide prevention and intervention strategies (Weed et al, 1990; Lippman et al, 1990). Such intermediate elements prior to the development of metastatic capable cancers also provide the opportunity to define the molecular biology of these elements. Studies of the development of pre-invasive breast disease have provided insight into different types of lesions with different implications for breast cancer risk and the process of carcinogenesis (See Figure 1). Pre-invasive breast disease is herewith defined to be any reproducibly defined condition which confers an elevated risk of breast cancer approaching double that of the general population (Komitowski et al, 1990). The specifically-defined atypical hyperplasias and lobular carcinoma in situ confer relative risks of four to ten times that of the general population. This risk is for carcinoma to develop anywhere in either breast (Page et al, 1985; Page et al, 1991). The statistical significance of these observations have regularly been $<.0001$. Thus, absolute risk figures of 10-20% likelihood of developing into invasive carcinoma in 10 to 15 years arise. DCIS is a very special element in this story because the magnitude of risk is as high as any other condition noted ($P< .00005$), but remarkably, the developing invasive cancer is in the same site in the same breast. This local recurrence and evolution to invasiveness marks these lesions as determinate precursors of invasive breast cancer (Betsill et al, 1978; Page et al, 1982). These figures are for the type of DCIS which has become detected very commonly since the advent of mammography, the small and NCDCIS variety. It is likely that the comedo DCIS variety indicates a much greater risk, often presenting as larger lesions, and treated regularly by mastectomy in the past 50 years making follow-up studies impossible (Figure 1).

The precision of histopathologic diagnosis in this area as noted in Table I (shown in Figure 1) was most convincingly confirmed in a large, prospective study (London et al, 1991). There has also been a recent review of the reproducibility of the assignment of diagnosis by a panel of pathologists (Schnitt et al, 1992). The precision has been fostered by combining histologic pattern criteria with cytologic and extent of

lesion criteria. Classic surgical pathology criteria were predominantly derived from histologic pattern only. A further point of relevance to the importance of these histopathologically defined lesions of pre-malignancy in the breast is the relationship to familiality. A family history of breast cancer in a first degree relatives confers about a doubling of breast cancer risk. However, women with the atypical hyperplasias at biopsy and a family history of breast cancer are at 9-10 times the risk of developing invasive breast cancer as the general population (Dupont et al, 1985; Dupont et al, 1989).

Careful consideration of all of the above-mentioned epidemiologic data has led to the following model for progression from generalized pre-malignant lesions to determinant lesions to invasive cancer. Figure 2 shows this model for the induction and progression of pre-invasive breast disease based on study of the Vanderbilt cohort (Dupont et al, 1985) of more than 10,000 breast biopsies (follow-up rate 85%; median time of 17 years; 135 women developed breast cancer).

EXAMPLE 2

Identification of genes which are differentially expressed in DCIS

Construction of cDNA libraries from DCIS lesions

In order to study differential gene expression in DCIS, we collected cases of NCDCIS. The diagnosis of DCIS is made on histomorphologic grounds based on architectural, cytologic, and occasionally extent criteria. NCDCIS lacks comedo features and consists of microscopic intraductal lesions which fill and extend the duct, contain rigid internal architecture, and often have hyperchromatic and monomorphic nuclei.

Study of non-comedo DCIS for differential marker gene expression indicates the diagnostic utility of comparison of marker gene expression in these tissues. Although the morbidity and mortality of breast cancer clearly results from invasion and metastasis, the development of breast cancer is clearly significant in its early stages for two basic reasons:

1) The molecular changes will presumably be simpler in early lesions than in later lesions which may have acquired numerous mutations or "hits";

and

2) Successful prevention strategies may require attacking cancer before it develops the capacity to invade or metastasize.

Non-comedo DCIS is the earliest determinant lesion which recurs locally as invasive cancer. Although comedo DCIS may be technically easier to study because the tumors are larger, its aggressiveness and the presence of numerous genetic alterations (such as p53 and erbB2) suggest that it may have advanced beyond the earliest stages of carcinogenesis.

The commercial utility of a method for prevention of cancer is clear. In order to study differential gene expression in DCIS, breast tissue with extensive microscopic non-comedo DCIS was identified and banked in a frozen state. cDNA libraries were constructed from mRNA isolated from frozen sections of DCIS lesions. Tissue samples from patients with mammographic results consistent with DCIS were cryostat frozen and a definitive diagnosis was made by the histopathologic criteria which we have described (Jensen et al, Submitted for publication; Holt et al, In press).

Control mRNA was obtained from frozen tissue samples obtained from reduction mammoplasties and from cultured human breast epithelial cells. Because non-comedo DCIS is a microscopic lesion, we had to microlocalize regions of DCIS in biopsy samples. To accomplish this we prepared frozen sections in which we located regions of DCIS and then employed a 2 mm punch to obtain an abnormal tissue sample only from those regions that contained DCIS. This selective harvesting was accomplished by carefully aligning the frozen section slide with the frozen tissue block and identifying areas of interest. The harvest of the appropriate area was then confirmed with a repeat frozen section. A similar approach was used to isolate mRNA from lobules of normal breast in samples collected from a reduction mammoplasty. Prior studies have shown that breast lobules are approximately 2.5 mm in diameter, thus the 2 mm punch provided a well-tailored excision. This microlocation and collection step, in which abnormal tissue samples are collected from an isolatable tissue structure, was performed with extreme care and was absolutely crucial to the success of these studies. Contamination by normal breast epithelial cells or by breast stromal cells would clearly negatively skew the differential screening approach. If the punch biopsy did not cleanly

excise DCIS without contamination by other cell types or tissues then the sample was not used for mRNA isolation (Jensen et al, Submitted for publication). Figure 3 contains color photos of DCIS (abnormal) tissue, before (upper left panel) and after excisional punch biopsy (upper right panel). The lower panels show tissue samples of normal breast tissue (lower left panel), and invasive breast cancer (lower right panel).

Following microlocation punch harvesting of the frozen tissue, RNA was isolated, purified, and employed to construct cDNA libraries. RNA was isolated following mincing of tissue in 5.6M guanidinium isothiocyanate and 40% phenol, centrifugation to remove particulate matter, viscosity reduction by repeated aspiration through a 22 gauge needle, chloroform extraction and ethanol precipitation. In most samples there was particulate matter resistant to guanidinium-phenol extraction that was white in color and fibrous in appearance and was presumed to represent breast stroma. This stromal material was sparse in DCIS samples but abundant in samples obtained from normal breast tissue derived from reduction mammoplasties. The stromal material was minced with a tissuemizer, washed with PBS, treated with collagenase at 37°C for 30 minutes, sonicated, extracted with phenol/chloroform and ethanol precipitated. 200 ug of total RNA was obtained by pooling 20 punches from normal breast tissue (reduction mammoplasty samples) and 5-8 punches from DCIS lesions, presumably reflecting the greater cellularity of the DCIS samples. All libraries had greater than 50% inserts and contained between 2 X 10^6 and 7 X 10^7 phage recombinants with an average insert size varying between 500 and 1000 base pairs.

EXAMPLE 3

Development of an extraction method which produces breast epithelial RNA

It was necessary that tissue samples not be contaminated by non-epithelial stromal cells. Such contamination would complicate efforts to compare gene expression between samples. In order to test the extent of stromal contamination of the mRNA samples, we determined the level of expression of collagen III mRNA by an RNase protection assay. RNase protection assays were employed in these and subsequent studies because it is a quantitative method and can be performed on small amounts of unselected RNA. Collagen III mRNA was identified in the presumed stromal fraction

of the normal breast tissue and to a lesser extent in the microinvasive breast cancer sample, but no expression of collagen III was detected in the DCIS samples which were subsequently employed for cDNA library construction. Figure 4 compares expression in NL 2 and #10CA with other patient samples and NL1 to determine collagen III expression.

Expression of Collagen III mRNA in tissue mRNA samples was analyzed by RNase protection assay by methods we have reported previously (Holt, 1993). One μg of mRNA was hybridized with two labeled RNA probes: a T7 polymerase-generated probe for human glyceraldehyde phosphate dehydrogenase (GADP) which protects a 140 bp Sac I-Xba I fragment; and a T7 polymerase-generated probe which protects a 208 bp Hinc II-Pst I fragment from the 3' untranslated region of the human type III procollagen gene (Coll III) obtained by PCR subcloning of the published sequence (Ala-Kokko et al, 1991). RNA samples were labeled as follows: NL1 is RNA from cultured human breast epithelial cells (Hammond et al, 1984), NL2 is RNA from normal breast tissue, NL3 is RNA derived from the fibrous stromal fraction of breast tissue as described (Jensen et al, Submitted for publication), NL4 is another sample from normal breast tissue. This is described in greater detail on page 30 of this patent application. #12,#8,#4,#6, and #10 are from patient samples with DCIS. Sample #10CA is RNA obtained from the small focus of microinvasion shown in Figure 3. Con is a control sample using tRNA.

EXAMPLE 4

Screening of cDNA libraries

Following successful testing which demonstrated that stromal contamination was not a problem, cDNA libraries were constructed in lambda phage using polyA-selected mRNA from the following samples: cultured human breast epithelial cells, tissue from three reduction mammoplasty patients, tissue from three DCIS patients, and tissue from one DCIS patient (patient #10) that showed a small focus of invasion adjacent to an area of DCIS. Multiple punches were needed to obtain sufficient RNA for polyA selection and library construction. Selective handling of tissue was accomplished.

Comparison of gene expression between samples was performed by either differential screening or a modification of differential display (Liang et al, 1992a; Liang et al, 1992b; Saiki et al, 1988; Melton et al, 1984). Plasmid DNA was prepared from the cDNA libraries following helper phage rescue and screened by two independent methods. Figure 5 below shows the results of differential display comparing cDNAs of several patient DCIS samples with cDNA obtained from normal breast epithelial cells and an early invasive cancer. Although few genes shown in this Figure are differentially expressed in the majority of samples with DCIS, the heterogeneity of gene expression in patient samples is seen.

The differential display method (Liang et al, 1992a and 1992b) allows simultaneous comparison of multiple tissue samples. Initial studies using this method (reverse transcriptase followed by PCR) were unsatisfactory because of unwanted amplification of contaminating DNA in tissue samples and the small size of many of the fragments identified by display. To circumvent some of these problems, we have attempted to combine the advantages of cDNA library screening with the advantages of differential display by:

1) Constructing cDNA libraries from the tissue mRNA samples;
2) Performing differential display on the plasmid DNA prepared from the cDNA libraries;
3) Subcloning the fragments identified by differential display;
4) Using the subcloned fragment as a probe to clone the cDNA from the appropriate library.

Example 5

Identification of a gene (RibRed) which is differentially expressed in multiple NCDCIS cases

Employing these methods, 10 differentially expressed clones were identified and the seven that showed the greatest difference in expression between multiple samples were further characterized by DNA sequencing. Comparison of the sequenced clones with GenBank demonstrated that six of the clones are apparently unique sequences (although further DNA sequencing is necessary); but that one of the clones (here termed DCIS-1 and described in Sequence Listing No. 1) showed 90% homology to the

previously cloned hamster gene encoding the M2 subunit of ribonucleotide reductase (Pavloff et al, 1992; Hurta et al, 1991; Hurta et al, 1991). Although human M2 ribonucleotide reductase has been cloned previously, comparison of the hamster cDNA sequence with our clone and with the prior human clone indicates that DCIS-1 is homologous to an alternatively poly-adenylated form of the human ribonucleotide reductase which has not been cloned previously. Figure 6 shows a comparison of the sequence between DCIS-1 and the human and hamster genes.

Because of our concern that different patients may have differential gene expression which is idiosyncratic (or related to morphological differences in biopsy appearance) and not necessarily related to the induction or progression of DCIS, we simultaneously analyzed gene expression in multiple DCIS samples compared to multiple control samples. We constructed cDNA libraries from the following samples:

1) Cultured HMEC epithelial cells;
2) Reduction mammoplasty: 11 year old with virginal hyperplasia;
3) Reduction mammoplasty: 28 year old patient;
4) Reduction mammoplasty: 35 year old patient;
5) DCIS patient #12;
6) DCIS patient #8;
7) DCIS patient #10;
8) DCIS patient #10 from an area of invasive cancer adjacent to DCIS;

In addition to the samples we employed to construct cDNA libraries shown above, we also obtained frozen tissue samples from 7 more DCIS patients, 2 cellular fibroadenoma samples, and samples of "usual hyperplasia" and atypical hyperplasia.

Because the DCIS clones were identified by cloning methods which include selection and amplification, it was important to confirm by nuclease protection assays that the genes were differentially expressed in the original unselected, unamplified RNA samples (Figure 7).

This approach allowed identification of a human gene similar to the hamster RibRed gene (coding for the M2 subunit) and 7 other human genes as genes which are differentially expressed in a majority of cases of DCIS in human breast tissue. The table of differentially expressed genes lists the genes which have been identified as

differentially expressed genes in DCIS tissue samples as compared to that in normal tissue (Figure 9).

EXAMPLE 6

Methods for studying potential use of differentially expressed genes for diagnostic screening

One advantage of the differential display method is that it allows comparison of multiple tissue samples of pre-invasive or invasive breast cancer. For example, use of this method has successfully demonstrated that the M2 subunit ribonucleotide reductase gene is differentially expressed in 4 out of 5 pre-invasive breast cancer tissue samples. It is significant that the M2 subunit is involved in the regulation of the ribonucleotide reductase gene and is found to be over-expressed in abnormal tissue samples.

Identification of differentially expressed genes may lead to discovery of genes which are potentially useful for breast cancer screening. Of particular interest are genes whose expression is restricted to breast epithelial cells and whose gene products are secreted. Screening for secreted proteins is possible by using the known hydrophobic sequences which encode leader sequences as one primer for differential display. The identification of secreted proteins which are specific for early breast pre-malignancy (or even early invasive cancer) would provide an important tool for early breast cancer screening programs. If a differentially expressed gene has not been cloned previously (or if details of its expression are unknown or uncertain) then nuclease protection assays or Northern blots can be performed on RNA prepared from tissue samples from a variety of tissues to determine if expression of this gene is restricted to breast. If necessary cDNA libraries prepared from other tissues can be added to the differential display screen as a way to identify only those genes which are expressed in early breast cancer and, in addition, are only expressed in breast tissue.

Once differentially expressed genes have been initially characterized for expression in pre-malignant and malignant breast disease, antibodies to the protein products of potentially useful genes can be developed and employed for immunohistochemistry (Harlow et al, 1988). This will provide an additional test to determine whether the expression of this gene is restricted to the breast. Subsequently, these antibodies will

be used to detect the presence of this protein present in the blood of patients with pre-invasive and/or invasive cancer. By assaying for serum protein levels in the same patients who exhibited elevated expression of the gene in their tissue samples it will be possible to determine whether a gene product is being secreted into the blood.

EXAMPLE 7

Decreased expression of BRCA1 accelerates growth and is observed during breast cancer progression

Breast cancer occurs in hereditary and sporadic forms. Recently the BRCA 1 gene has been cloned and shown to be mutated in kindreds with hereditary breast and ovarian cancer (Hall et al. 1990, Miki, Y. et al. 1994, Friedman et al. 1994, Castilla et al. 1994, Simard et al. 1994). Although 92% of families with two or more cases of early-onset breast cancer and two cases of ovarian cancer have germ-line mutations in BRCA 1 (Narod et al. in press), the gene has not been shown to be mutated in any truly sporadic case to date (Futreal et al. 1994). Despite the surprising paucity of somatically acquired mutations in sporadic breast cancer, it is still a likely tumor suppressor gene with a key role in breast epithelial cell biology. The BRCA 1 gene encodes a protein of 1863 amino acids with a predicted zinc finger domain observed in proteins which regulate gene transcription.

As an initial characterization of the regulation and function of the BRCA 1 gene, we analyzed and manipulated expression of BRCA 1 mRNA levels. The results taken together indicate that the BRCA 1 gene product is a negative regulator of mammary cell proliferation which is expressed at diminished levels in sporadic breast cancer.

Expression of BRCA1 mRNA during breast cancer progression

As described above, microscopy-directed cloning has been employed to compare gene expression in normal mammary epithelium, carcinoma in-situ, and invasive breast cancer. This method produces predominantly epithelial mRNA with minimal contamination from stromal elements and we used this approach to obtain mRNA from normal neoplastic tissues from patients without a family history of breast cancer. Expression of BRCA1 exon 24 in human breast tissue samples is shown in Fig. 1. The legend of Fig. 1 is as follows.

The following tissue samples were used for mRNA isolation: Normal tissue samples: NL1-cultured human breast epithelial cells, NL2- Histologically normal breast tissue from an 11 year old undergoing a reduction mammoplasty, NL4- histologically normal breast tissue from an 14 year old undergoing a reduction mammoplasty. Carcinoma-in-situ samples are #6, #8, #10, #12, #23 (comedo type), #41, #55; and invasive cancer samples #10CA (invasive cancer from the same patient with carcinoma-in-situ), 36CA, 1CA. All of these tissue samples were obtained from patients who had no family history of hereditary breast cancer and RNA preparation was performed as described above.

PCR detection of BRCA1 exon 24 in cDNA libraries from the following tissue samples is described in Figure 10A. Lane 1: human genomic DNA, lane 2: NL1, lane 3: NL4, lane 4: $8, lane 5: #12, lane 6: #10, lane 7: #10CA, lane 8: #41, lane 9: #23, lane 10: 36CA, lane 11: lambda DNA. The arrow points to the expected 113 bp band.

Nuclease protection assays of microdissected mRNA from tissue samples are described in Fig. 10B. One ug of mRNA from each tissue sample was hybridized with 32P-labelled, T7 polymerase-generated RNA probes for BRCA1 and human glyceraldehyde-3-phosphate dehydrogenase (GAPD) which produce expected protected fragments of 113 and 140 respectively as indicated by the lines on the right. Data were quantitated by phosphorimaging. The hybridizing intensity of each BRCA1 band was normalized to its respective GAPD band. The normalized values of NL1, NL2, and NL4 were intensity in each sample relative to 1. Sample 1 employs human leukocyte mRNA; Samples 2-4 are NL1, NL2, and NL4; Samples 5-9 are #6(2.8), 8(3.7), 10(2.8), 12 (5.9), and 55 (1.4); and 10-12 are #10CA (0.07), 36CA (0.13), and 1CA (0.2).

Fig. 10 shows that BRCA1 exon 24 mRNA is expressed at 5-10 fold higher levels in normal mammary tissue than in invasive breast cancer samples. Initial studies showed detectable levels of BRCA1 cDNA in a cDNA library prepared from a tissue sample with preinvasive carcinoma-in-situ but not in normal breast cancer invasive breast cancer cDNA libraries (Figure 10A). Because this method is relatively insensitive we directly quantitated BRCA1 mRNA by nuclease protection assays in RNA samples obtained by our microdissection method described above. These assays

indicate that expression of BRCA1 mRNA in micro-dissected normal mammary epithelial tissue (lanes 2-4, Figure 10B) is 5-15 fold higher than that in breast cancer (lanes 10-12, Figure 10B). The highest levels of BRCA1 are observed in samples from non-comedo ductal carcinoma-in-situ (lanes 5-9, Figure 10B), a premalignant breast lesion with a finite, but relatively low rate of progression to invasion (Betsill et at., 1978, Page, D.L. et al., 1982, Page and Dupont, 1990).

Because these studies suggested that invasive breast cancer exhibited lower mRNA levels than normal breast epithelial cells, we compared expression of paired samples of normal breast and invasive cancer from the same patient (Figure 11A; compare lanes 2 and 3, 4 and 5, 6 and 7). The legend of Fig. 11 is as follows.

Nuclease protection assays of RNA obtained from paired samples of invasive breast cancer and histologically normal breast tissue are shown in Fig. 11A. Samples in lanes 2 and 3 (first patient), 4 and 5 (second patient), 6 and 7 (third patient) are from invasive cancer and normal breast tissue respectively. Lane 1 is NL1 mRNA as described in legend to Fig. 10 and lane 8 is human leukocyte mRNA. Ratios of BRCA1/GAPD for each sample: lane 1: 25.9, lane 2: 1.8, lane 3: 7.6, lane 4: 2.0, lane 5: 12.4, lane 6: 0.7, lane 7: 6.0. The probes and methods are as described in Fig. 10 except the GAPD probe was of lower specific activity to improve quantitation.

Nuclease protection assays of RNA from a series of invasive breast cancer tissue samples (lanes 2-9 compared with NL1 (lane 1) and leukocyte mRNA (lane 10) are shown in Fig. 11B. Ratios of BRCA1/GAPD for each sample: lane 1: 19.1, lane 2: 0.3, lane 3: 1.8, lane 4: 1.6, lane 5: 0.2, lane 6: 0.3, lane 7: 1.9, lane 8: 0, lane 9: 0.6.

Although the samples were paired in Fig. 11A, they were not microdissected so this approach overestimates the relative expression level of invasive samples because they have a greater percentage of epithelial cells. RNA levels were four to eight fold higher in samples derived from normal breast than in samples derived from invasive breast cancer. We next analyzed expression levels in 8 non-hereditary invasive cancer samples (Figure 11B: lanes 2-7). Although these samples showed some variability in expression level, all had lower levels of BRCA1 mRNA (determined by ratio of

BRCA1 to GAPD) than the primary breast epithelial cell line or the normal breast samples shown in Figure 11A.

Effects of BRCA1 gene inhibition on proliferative rate and gene expression

Having demonstrated that mRNA expression levels of BRCA1 are higher in normal mammary cells than in cancer cells, we used antisense methods to test the hypothesis that BRCA1 expression inhibits cell growth. Unmodified 18 base deoxyribonucleotide complementary to the BRCA1 translation initiation site were synthesized and added to cultures of primary mammary epithelial cells (Stampfer et al. 1980) or MCF-7 breast cancer cells (Soule and McGrath, 1980). Figure 12 is graph showing growth rate of human primary mammary epithelial cells (A), MCF-7 cells (B), retinal pigmented epithelial cells (C), cultured as described below. Points and bars represent the mean and the 95% confidence interval of triplicate counts of cells incubated with a single bolus of the indicated concentration of antisense or control sense deoxyribonucleotide.

The morphologic appearance of the cell lines was not noticeably changed by addition of antisense oligonucleotide, but the proliferative rate was faster. Incubation of cells with 40 uM anti-BRCA1 oligonucleotide produced accelerated growth of both normal (Figure 12A) and malignant mammary cells (Figure 12B), but did not affect the growth of human retinal pigmented epithelial cells (Figure 12C). An intermediate dose of anti-BRCA1 oligonucleotide produced a less pronounced but significant increase in cell growth rate. This was not a toxic effect of the oligonucleotide since a control "sense" oligomer with the same GC content did not increase the proliferation rate, and because an addition of a 10 fold excess of sense oligomer to the anti-BRCA1 oligomer reversed the growth activation.

In order to critically evaluate the function of BRCA1 gene inhibition on growth stimulation and cell cycle progression it was necessary to identify a gene whose expression is cell cycle regulated in human mammary cells. The gene encoding the M2 subunit of ribonucleotide reductase is amplified in conditions of nucleotide starvation (Hurta and Wright 1992) and as shown above, exhibits elevated levels of expression in premalignant breast disease. Because ribonucleotide reductase constitutes the rate limiting step in DNA synthesis, we reasoned that it might be cell cycle regulated in a

synchronous growth model such as MCF-7 cells which can be growth arrested by tamoxifen and then restimulated by estrogen (Aitken et al. 1985, Arteaga et al. 1989). MCF-7 cells were growth arrested by tamoxifen for 48 hours and then stimulated at time zero (0) with 1uM estradiol (+E) or control vehicle (-E). Inhibition of DNA synthesis by tamoxifen and induction of synthesis by estrogen were confirmed by nuclear labelling studies with tritiated thymidine.

Fig. 13 panels A and B show that transcription of the ribonucleotide reductase M2 gene is cell cycle regulated, inhibited by tamoxifen, and induced by estrogen. Fig. 13A is a Northern blot of mRNA from synchronized MCF-7 cells. At the indicated time in hours, total cellular RNA was isolated and Northern blotting performed using the 1.6 Kb Eco RI fragment from our cloned human ribonucleotide reductase cDNA described above. Two mRNA species of 1.6 and 3.4 Kb are observed in these studies.

Fig. 13B shows nuclear runon studies of synchronized MCF-7 cells were performed by our published methods (Holt et al 1988) employing the 1.6 Kb fragment of ribonucleotide reductase described above (RR); the 1.8 Kb fragment of Topoisomerase II (Topo) described in the Olsen et al. 1993); the 1.0 Kb cyclophilin gene (Thompson et al. 1994) used as a constitutive control; and 18S ribosomal RNA (Thompson et al. 1994). Con represents cells which were grown for 48 hours but not treated with tamoxifen.

Antisense inhibition is a useful strategy for studying gene expression which is dependent on expression of the antisense target gene (Robinson-Benion and Holt, in press, 1995), e.g. genes whose expression is directly or indirectly dependent on BRCA1 levels. Fig. 14 demonstrates that antisense inhibition of BRCA1 results in a corresponding increased expression of M2 ribonucleotide reductase mRNA. A nuclease protection assay of mRNA derived from primary mammary epithelial cells (lanes 1-4, 9-10) or MCF-7 cells (lanes 5-8, 11-12) cultured for 4 days with antisense or control oligonucleotide was performed under the following conditions: no oligonucleotide (lanes 1 and 5); 40uM antiBRCA1 (lanes 2,6,10,12); 4uM antiBRCA1 (lanes 3 and 7); 40uM sense control (lanes 4,8,9,11). Probes for BRCA1 and GAPD are as described for

Figure 10, and the ribonucleotide reductase M2 probe (RR) detects the 200 bp probe is described above.

Ribonucleotide reductase mRNA levels are highest in samples treated with 40 uM anti-BRCA1 oligonucleotide for both primary mammary epithelial cells and for MCF-7 cells (Fig. 14). Antisense inhibition of BRCA1 results in a 70-90% inhibition of mRNA levels in anti-BRCA1 treated cells compared with cells treated with the "sense" control oligonucleotide (compare lanes 9 and 10, Fig. 14). Note that MCF-7 cells have lower levels of BRCA1 than the normal mammary epithelial cells (compare lanes 1 and 5, Fig. 14) anti-BRCA 1 since the antisense inhibition may drop BRCA1 levels below a critical threshold which normally functions to inhibit growth.

Methodology

Tissue samples. Freshly obtained breast biopsy or reduction mammoplasty specimens were frozen and then RNA was obtained following the microdissection method described above. Lesions were selected which were microlocalized and homogenous so that pure lesions could be obtained by 2 mm punches. Samples which had admixed normal epithelial, carcinoma-in-situ, or invasive cancer were not used for this study. Family history was obtained by chart review and/or interview to exclude familial breast cancer cases.

Nuclease Protection Assays. PCR primers were derived from BRCA1 sequence in GenBank (Accession number U14680); forward 5' CAATTGGGCAGATGTGT 3' and reverse 5' CTGGGGGATCTGGGGTATCA 3' which amplify a 113 bp region from exon 24, corresponding to bases 5587 to 5699 of the human BRCA1. This region was selected because this exon has not been reported to be differentially spliced unlike more 5' exons. The BRCA1 probe was cloned by subcloning this 113 bp band from normal human genomic DNA into PCRscriptSK and screening for correct orientation. One ug of mRNA from each tissue sample was hybridized with 32P-labelled, T7 polymerase-generated RNA probes for BRCA1 and human glyceraldehyde-3-phosphate dehydrogenase (GADP) which would produce expected protected fragments of 113 and 140 respectively. The construction and use of the GADP probe for RNA standardization has been described above. The probe for

ribonucleotide reductase M2 mRNA is the same as above and detects a 200 bp protected fragment.

Antisense oligonucleotide studies. Unmodified deoxyribonucleotide were analyzed by gel electrophoresis and UV shadowing and shown to be homogenous and of appropriate size. These oligonucleotide were purified by multiple lyophilization and solubilized in buffered media as described (Holt et al. 1988). Sequence of the unmodified antiBRCA1 oligonucleotide 5' AAGAGCAGATAAATCCAT 3' and the complementary sense oligonucleotide 5' ATGGATTTATCTGCTCTT 3' correspond to the presumed translation initiation site at bases 12-137 of the GenBank sequence. The antisense oligonucleotide sequence was searched against Genbank and no significant homologies were identified to genes except BRCA1. Oligonucleotides were used according to our published methods (Holt et al. 1988). Primary mammary epithelial cells were cultured in serum-free medium supplemented with epidermal growth factor, insulin, hydrocortisone, ethanolamine, phosphorylethanolamine, and bovine pituitary extract. MCF-7 cells were cultured in Minimum Essential Medium Eagle (Modified) with Earle's salts and 2g/L sodium bicarbonate m supplemented with 2mM L-glutamine, GMS-A (Gibco Cat. #680-1300AD), nonessential amino acids, and 2.5% fetal calf serum. Retinal pigmented perithelial cells were cultured in DMEM and 10% calf serum.

Our results indicate that the BRCA1 gene is expressed at higher levels in normal mammary cells than in breast cancer cells and that diminished expression of BRCA1 increased the proliferative rate of breast cells. This correlates well with the recent finding that patients with BRCA1 gene-linked hereditary breast cancer have tumors that grow more rapidly than comparable sporadic tumors (Marcus, J. et al. 1994). The decreased mRNA levels which were observed in sporadic breast cancers are not a consequence of differential splicing of the gene since the RNAs were quantitated with probes from the 3' end of the mRNA which is not a region where differential splicing is reported to occur (Miki, Y. et al 1994). Invasive sporadic cancers have BRCA1 mRNA levels which vary from 0 (in one case) to 20% of the levels observed in normal human mammary epithelium.

Examples 8 and 9 describe applications of the discovery of the function of the BRCA1 gene. Example 8 describes a gene therapy method and example 9 describes a drug screening method. The discovery of the diminished expression of the BRCA1 mRNA in breast cancer using the microdissection techniques of this invention provides an important scientific basis for these examples.

Example 8

Gene Therapy method based on determination of the function of the BRCA1 Gene

Viral vectors containing a DNA sequence that codes for a protein having an amino acid sequence as essentially set forth in SEQ ID NO:49 can be constructed using techniques that are well known in the art. This sequence includes the BRCA1 gene product. Viral vectors containing a DNA sequence essentially as set forth in SEQ ID NO:47 (the BRCA1 gene) can be also constructed using techniques that are well known in the art. Retroviral vectors, adenoviral vectors, or adeno-associated viral vectors are all useful methods for delivering genes into breast cancer cells. An excellent candidate for use in breast cancer gene therapy is a Moloney-based retroviral vector with a breast selective MMTV promoter which we have reported previously (Wong et al). The viral vector is constructed by cloning the DNA sequence essentially as set forth in SEQ ID:47 into a retroviral vector such as a breast selective vector. Most preferably, the full-length (coding region) cDNA for BRCA1 is cloned into the retroviral vector. The retroviral vector would then be transfected into virus producing cells in the following manner: Viruses are prepared by transfecting PA317 cells with retroviral vector DNAs which were purified as described in Wong et al. Following transfection, the PA317 cells are split and then treated with G418 until individual clones can be identified and expanded. Each clone is then screened for its titer by analyzing its ability to transfer G418 resistance (since the retroviral vector contains a Neomycin resistance gene). The clones which have the highest titer are then frozen in numerous aliquots and tested for sterility, presence of replication-competent retrovirus, and presence of mycoplasm. The methods generally employed for construction and production of retroviral vectors have been described in Muller, 1990.

Once high titer viral vector producing clones have been identified, then patients with breast cancer can be treated by the following protocol: Viral vector expressing

BRCA1 is infused into either solid tumors or infused into malignant effusions as a means for altering the growth of the tumor (since it is shown above that the BRCA1 gene product decreases the growth rate of breast cancer cells). Because viral vectors can efficiently transduce a high percentage of cancer cells, the tumors would be growth inhibited.

Example 9

Method of Screening Compounds Capable of Activating Promoter Region of the BRCA1 Gene

The discovery of the function of the BRCA1 gene provides a clear utility in that induction of expression of the gene and the resulting increase in level of protein encoded by the gene in the breast cancer cell should slow the proliferation of the breast cancer cells. Induction of expression of the gene can be caused by administering a compound to a patient that stimulates the regulatory regions of this gene, such as the promoter.

A method for screening compounds that activate the promoter of the BRCA1 gene is designed in the following way. A promoter sequence is a DNA segment that upregulates the expression of a gene. A sequence essentially as set forth in SEQ ID NO:48 can be ligated into a suitable vector, such as a plasmid, that contains a reporter gene using standard recombinant DNA techniques of restriction enzyme digests, ligation of fragment into vector, and transformation of bacteria. SEQ ID NO:48 includes the promoter sequence of the BRCA1 gene. A reporter gene is a gene that produces a readily detectable product. Examples of appropriate reporter genes which could be employed for this purpose include Beta-galactosidase or the chloramphenicol acetyltransferase gene.

The BRCA1 promoter/reporter gene combination can then be cloned into an expression vector or viral vector by standard recombinant DNA methods. Breast cancer cells can then be transfected with the expression vector containing the BRCA1 promoter/reporter gene using standard transfection methods which we have reported previously (Holt et al. PNAS 1986). A stable transformant with appropriate low level expression (breast cancer cells have low level BRCA1 expression as shown above) will be identified and then characterized to demonstrate proper DNA integration and

expression. Methods of establishing and characterizing stable transformants have been described (Holt. MCB, 1994). Once an appropriate stable transformant cell line is identified, then we can plate the cell line in a manner than permits screening of hundreds or thousands of drugs or biological agents (for example in multiple 96 well microtiter plates). Level of expression of the reporter gene can be quantitated and agents which activate expression are thus identified. A positive result (i.e. induction of the promoter region) results in increased levels of the reporter gene resulting in either an increase in color (Beta-galactosidase assay) or specific radioactivity (Chloramphenicol acetyltransferase activity) through a reaction between the protein encoded by the reporter gene and a compound in the reaction medium. The compound produced by the reaction between the reporter gene protein and the compound in the reaction medium is the cause of the increase in color or specific radioactivity. These compounds can be called indicator compounds in that their presence indicates that the drug or biologial agent activitated the BRCA1 promoter. Methods for standardizing and performing Beta-galactosidase or chloramphenicol acetyltransferase assays have been reported (Holt et. al. MCB 1994). This method would be useful for initial screening of agents which increase BRCA1 expression. These agents could then be tested in more rigorous assays of breast cancer growth such as nude mouse tumor assays (Arteaga et al). This approach allows mass screening of large numbers of agents, sparing more rigorous animal tests for only promising compounds which score in the reporter gene assay described herein.

Thus, although there have been described particular embodiments of the present invention of a new and useful "Method for Detection and Treatment of Breast Cancer", it is not intended that such embodiments be construed as limitations upon the scope of this invention except as set forth in the following claims. It will be apparent to those skilled in the art that many changes and modifications may be made without departing from the invention in its broader aspects. For example, the above described techniques may be used in the diagnosis of other diseases and detection of differential genetic expression from microscopically-directed tissue samples of pathologic tissue. The production of a cDNA library produced as a result of the differential expression of genes in pathologic tissue in comparison to normal tissue provides the opportunity for

further adiagnostic capabilities. Further, although there have been described certain experimental conditions used in the preferred embodiment, it is not intended that such conditions be construed as limitations upon the scope of this invention except as set forth in the claims.

The following references are included to provide details of scientific technology herein incorporated by reference to the extent that they provide additional information for the purposes of indicating the background of the invention or illustrating the state of the art.

REFERENCES

Abendroth, C. S., H. H. Wang, et al. 1991. Comparative features of carcinoma in situ and atypical ductal hyperplasia of the breast on fine-needle aspiration biopsy specimens. Am J Clin Pathol 96(5): 654-659.

Aitken, S.C. & Lippman, M.E. Effect of estrogens and antiestrogens on growth-regulatory enzymes in human breast cancer cells in tissue culture. *Cancer Res.* **45**, 1611-1620 (1985).

Ala-Kokko,L., S. Kontusaari, C. T. Baldwin, H. Kuivaniemi, D. J. Prockop. 1991. Structure of cDNA clones coding for the entire preproalpha I (III) chain of human type III procollagen. Biochemical Journal 260, 509-516.

Allred, D. C., G. M. Clark, et al. 1992. Overexpression of HER-2/neu and its relationship with other prognostic factors change during the progression of in situ to invasive breast cancer. Hum Pathol 23(9): 974-979.

Arteaga, C.L. et al. Blockade of the type I somatomedin receptor inhibits growth of human breast cancer cells in athymic mice. *J. Clin. Invest.* **84**, 1418-1423 (1989).

Betsill, WL, Jr., PP Rosen, PH Lieberman, and GF Robbins.1978. Intraductal carcinoma. Long term follow-up after biospy only. J.A.M.A. 239:1863-1867.

Böcker, W., B. Bier, et al. 1992. An immunohistochemical study of the breast using antibodies to basal and luminal keratins, alpha-smooth muscle actin, vimentin, collagen IV and laminin. Part II: Epitheliosis and ductal carcinoma in situ. Virchows Arch A Pathol Anat Hispathol 421: 323-330.

Callahan R, CS Cropp, GR Merlo, DS Liscia, APM Cappa, and R. Lidereau. 1992. Somatic mutations and human breast cancer. Cancer 69:1582-86.

Carter D, RRL Smith. 1977. Carcinoma in situ of the breast. Cancer 40:1189-1193.

Castilla, L.C. et al. Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. *Nature Genet* **8**, 387-391 (1994).

Cheickh MB, P Rouanet, G Louason, P Jeanteur, and C Theillet. 1992. An attempt to define sets of cooperating genetic alterations in human breast cancer. Int. J. Cancer 51:542-547.

Chen L-C, W Kurisu, B-M Ljung, ES Goldman, Dmorre II, and HS Smith. 1992. Heterogeneity for allelic loss in human breast cancer. JNCI 84:506-510.

Dupont, WD, Parl FF, Hartmann WH, Brinton LA, Winfield AC, Worrell JA, Schuyler PA, Plummer WD: Breast cancer risk associated with proliferative breast disease and atypical hyperplasia. Cancer 71:1258-1265, 1993.

Dupont WD, and DL Page. 1985. Risk factors for breast cancer in women with proliferative breast disease. N.Engl.J.Med. 312:146-151.

Dupont WD, DL Page, LW Rogers, and FF Parl. 1989. Influence of exogenous estrogens, proliferative breast disease, and other variables on breast cancer risk. Cancer 63:948-957.

Fentiman, I. S. 1992. Ductal carcinoma in situ. BMJ 304: 1261-1262.

Fisher, E. R. and R. Siderit .1992. Value of cytometric analysis for distinction of intraductal carcinoma of the breast. Breast Cancer Res Treat 21: 165-172.

Friedman, L.S. et al. Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nature Genet* **8**, 399-404 (1994).

Futreal, A. et al. BRCA1 mutations in primary breast and ovarian carcinomas. *Science* **266**, 120-122 (1994).

Hall, J.M. et al. Linkage of early-onset breast cancer to chromosome 17q21.*Science* **250**, 1684-1689 (1990).

Hammond SL, RG Ham, and MR Stampfer.1984. Serum-free growth of human mammary epithelial cells: rapid clonal growth in defined medium and extended serial passage with pituitary extract. Proc. Natl. Acad. Sci. USA 81:5435-5439.

Harlow E, D Lane. 1988. Antibodies. A Laboratory Manual. Cold Spring Harbor Laboratory Press.

Holt JT, RA Jensen, and DL Page. Histopatholoy: Old Principles and new methods. 1994. Cancer Surveys. In Press.

Holt, JT. Antisense rescue defines specialized and generalized functional domains for c-fos protein. 1993. Molecular and Cellular Biology 13, 3821-3830.

Holt, J.T., Redner, R.L., and Nienhuis, A.W. An oligomer complementary to c-myc mRNA inhibits proliferation of H-60 promyelocytic cells and induces differentiation. *Mol. Cell. Biol* **8**, 963-973 (1988).

Hurta, RAR, J. A. Wright. 1992. Alteration in the activity and regulation of mammalian ribonucleotide reductase by a DNA damaging agent. Journal of Biological Chemistry 267, 7066-
7071.

Hurta, RAR, S. K. Samuel, A. H. Greenberg, J. A. Wright. 1991. Early induction of ribonucleotide reductase gene expression by TGF-beta1 in malignant Ras transformed cell lines. Journal of Biological Chemistry 266, 24097-24100.

Jensen, RA, DL Page, and JT Holt. Identification of genes expressed in pre-malignant breast disease by microscopically-directed cloning. Proc. Natn. Acad. Sci. U.S.A. 91, 9257-9261 (1994).

Killeen, J. L. and H. Namiki .1991. DNA analysis of ductal carcinoma in situ of the breast: A comparison with histologic features. Cancer 68: 2602-2607.

Komitowski D, and C. Janson. 1990. Quantitative features of chromatin structure in the prognosis of breast cancer. Cancer 65:2725-2730.

Lawrence, G. 1991. Evaluation of treatment options for ductal carcinoma in situ of the breast. Arch Surg 126: 1541-1541.

Lennington, WJ, RA Jensen, LW Dalton, and DL Page. Ductal carcinoma in situ of the breast. Cancer 73: 118-124, 1994.

Liang P, and AB Pardee. 1992. Differential display of eukaryotic mRNA by means of the polymerase chain reaction. Science 257:967-971.

Liang P, L Averboukh, K Keyomarsi, R Sager, and AB Pardee. 1992. Differential display and cloning of messenger RNAs from human breast cancers versus mammary epithelial cells. Cancer 52:6966-6968.

Lippman SM, JS Lee, R Lotan, W Hittelman, MJ Wargovich, and WK Hong. 1990. Biomarkers as intermediate endpoints in chemoprevention trials. JNCI 82:555-560.

London SJ, JL Connolly, SJ Schnitt, and GA Colditz. 1991. A prospective study of benign breast disease and risk of breast cancer. JAMA 267:941-944.

Maguire, H. J., M. E. Hellman, et al. 1992. Expression of c-erbB-2 in in situ and in adjacent invasive ductal adenocarcinomas of the female breast. Pathobiology 60: 117-121.

Marcus, J.N., Watson, P., Page, D.L. & Lynch, H.T. Pathology and heredity of breast cancer in younger women. *J. Nat. Cancer Inst. Monograph No.* 16, 23-34 (1994).

Melton, D.A., D.A. Krieg, M.R. Rebagliati, T. Flaniatis, K. Zinn, and M.R. Green. 1984. SP6 RNA polymerase. Nucleic.Acid.Res. 12:7035-7056.

Miki, Y, et al. A strong candidate gene for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66-71 (1994).

Narod, S.A. et al. A heterogeneity analysis of 145 breast-ovarian cancer families. *Am. J. Human Genet.* (In Press)

Olsen, J.E., Winston, J.T., Whitlock, J.A., and Pledger, W.J. Cell cycle-dependent gene expression in V point-arrested Balb c 3T3 cells. *J. Cell. Phys.* **154**, 333-342 (1993).

Ottesen, G. L., H. P. Graversen, et al. 1992. Ductal carcinoma in situ of the female breast: Short-term results of a prospective nationwide study. Am J Surg Pathol 16: 1183-1196.

Page DL, WD Dupont, LW Rogers, and M Landenberger. 1982. Intraductal carcinoma. Follow-up after biopsy only. Cancer 49:751-758.

Page DL, WD Dupont, LW Rogers, and MS Rados. 1985. Atypical hyperplastic lesions of the female breast. A long term follow-up study. cancer 55:2698-2708.

Page, D.L. & Dupont, W.D. A typical hyperplastic lesions of the female breast. A long term follow up study. *Cancer* **66**, 1326-1335 (1990).

Page DL, TE Kidd, WD Dupont, JF Simpson, and LW Rogers. 1991. Lobular neoplasia of the breast: higher risk for subsequent invasion predicted by more extensive disease. Hum Pathol. 22:1232-1239.

Page DL, R Vander Zwaag, LW Rogers, LT Williams, WE Alker, and WH Hartman. 1978. Relation between component parts of fibrocystic disease complex and breast cancer. J.Natl. Cancer Inst. 61:1055-1063.

Pavloff, N., D. Rivard, S. Masson, S. H. Shen, A. M. Mes-Masson. 1992.Sequence analysis of the large and small subunits of human ribonucleotide reductase. DNA Sequence 2, 227-234 .

Pierce, S. M., S. J. Schnitt, et al. 1992. What to do about mammographically detected ductal carcinoma in situ? Cancer 70: 2576-2578.

Poller, D. N., C. E. Hutchings, et al. 1992. p53 Protein expression in human breast carcinoma: Relationship to expression of epidermal growth factor receptor, c-erbB-2 protein overexpression, and oestrogen receptor. Br J Cancer 66: 583-588.

Poller, DN, EC Roberts, JA Bell, CW Elston, RW Blamey, and IO Ellis. p53 expression in mammary ductal carcinoma in situ: relationship to immunohistochemical expression of estrogen receptor and cerbB2 protein. Human Pathology 1993, 24:463-468.

Posner, M. C. and N. Wolmark 1992. Non-invasive breast carcinoma. Breast Cancer Res Treat 21: 155-164.

Radford DM, K Fair, AM Thompson, JH Ritter, M Holt, T Steinbrueck, W Wallace, Wells SA Jr, and HR Donis-Keller. Allelic loss on chromosome 17 in ductal carcinoma in situ of the breast. Cancer Res 1993, 53:2947-2950.

Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Sambrook J, EF Fritsch, T Maniatis. 1989. Molecular Cloning Laboratory Manual, 2d Edition. Cold Spring Harbor Laboratory Press.

Sanchez LM, F Vizoso, I Diez-Itza, C Lopez-Otin. 1992. Cancer Res 32:95-100.

Schnitt, SJ, JL Connolly, FA Tavassoli, et al. 1992. Interobserver reproducibility in the diagnosis of ductal proliferative breast lesions using standardized criteria. Am.J. Surg. Pathol. 16:1133-1143.

Schwartz, G. F., G. C. Finkel, et al. 1992. Subclinical ductal carcinoma in situ of the breast: Treatment by local excision and surveillance alone. Cancer 70: 2468-2474.

Simard, J. et al. Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families. *Nature Genet* **8**, 392-398 (1994).

Simpson, T., R. C. Thirlby, et al. 1992. Surgical treatment of ductal carcinoma in situ of the breast: 10- to 20-year follow-up. Arch Surg 127: 468-472.

Solin, L. J., A. Recht, et al. 1991. Ten-year results of breast-conserving surgery and definitive irradiation for intraductal carcinoma (ductal carcinoma in situ) of the breast. Cancer 68: 2337-2344.

Soule, H.D. & McGrath, C.M. Estrogen responsive proliferation of clonal human breast carcinoma cells in athymic mice. *Cancer Letters* **10**, 177-189 (1980).

Stampfer, M., Hallowes, R.C., & Hackett, A.J. Growth of normal human mammary cells in culture. *In Vitro* **16**, 415-425 (1980).

Swain, S. M. 1992. Ductal carcinoma in situ. Cancer Invest 10: 443-454.

Tsuda H, K Iwaya, T Fukutomi, S Hirohashi. p53 mutations and c-erbB2 amplification in intraductal and invasive breast carcinomas of high histologic grade. Jpn J Cancer Res 1993, 84:394-401.

Thompson, M.E., Valentine, D.L., Strada, S.J., Wagner, J.A., and Scammell, J.G. Transcriptional regulation of secretogranin II and chromogranin B by cyclic AMP in a rat pheochromocytoma cell line. *Mol. Pharmacol.* 880-889 (1994).

Ward, B. A., C. F. McKhann, et al. 1992.Ten-year follow-up of breast carcinoma in situ in Connecticut. Arch Surg 12: 1392-1395.

Weed DL, P Greenwald, and JW Cullem. 1990. The future of cancer prevention and control. Semin.Oncol. 17:504-509.

Wellings SR, HM Jensen, and RG Marcum. 1975. An atlas of subgross pathology of the human breast with special reference to possible precancerous lesions. J.Natl. Cancer Inst. 55:231-273.

ADDITIONAL DESCRIPTION OF THE FIGURES

Figure 2: Model for premalignant conditions, highlighting magnitude of risk for progression to clinical malignancy. Terms from human breast neoplasia are used: no proliferative disease (No Pro), proliferative disease without alypia (PDWA), typical hyperplasia (AH), carcinoma in situ (CIS). As is proposal of tumor progression each stage is more likely to proceed to the next (dotted lines), but could also remain stable (horizontal lines, probably fairly frequent), or directly proceed to develop a clone of cells with malignant behavior (vertical lines, becoming more likely further to right.)

Figure 5: Differential display of cDNAs obtained from patient tissue samples and controls. Rescued cDNA library samples were used as templates for low stringency PCR with the primers 5'GATGAGTTCGTGTCCGTACAACTGG3' and 5' GGTTATCGAAATCAGCCACAGCGCC3'; 40 cycles were performed at conditions described above. Samples (See legend to Figure 4): Lane 1 - #12; Lanes 2 and 3: separate phage rescues of NL1 to show reproducibility of the assay; Lane 4 - #8; Lane 5 - #10; Lane 6 - #10CA; Lane 7 - control from the rescued phage vector without cDNA inserts. Arrows mark cDNAs which are overexpressed in DCIS versus normal. Arrowheads mark cDNAs which are differentially expressed in the invasive cancer (note this may reflect contamination from stromal cells). The bar marks a cDNA which is expressed in normal breast cells at higher levels than in DCIS or invasive cancer.

Figure 7: Expression of DCIS-1 mRNA in tissue mRNA samples analyzed by RNase protection assay. Probes: GADH probe and DCIS-1 clone probe which was generated by linearizing the rescued plasmid with Pvu II and should generate a 200 bp protected fragment. RNA samples were labeled as in the legend to Figure 4.

GeneDX 1009, pg. 891

SEQUENCE LISTINGS

(1) GENERAL INFORMATION:

(i) APPLICANT: HOLT, JEFFREY T.
JENSEN, ROY A.
PAGE, DAVID L.
OBERMILLER, PATRICE S.
ROBINSON-BENION, CHERYL L.
THOMPSON, MARILYN E.

(ii) TITLE OF INVENTION: METHOD FOR DETECTION AND TREATMENTS OF BREAST CANCER

(iii) NUMBER OF SEQUENCES: 49

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: I.C. WADDEY, JR.
(B) STREET: 27TH FLOOR, L & C TOWER, 401 CHURCH
(C) CITY: NASHVILLE
(D) STATE: TENNESSEE
(E) COUNTRY: USA
(F) ZIP: 37219

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette, 3.50 inch, 800 kB storage
(B) COMPUTER: IBM PC/XT/AT compatible
(C) OPERATING SYSTEM: MS-DOS (version 5.0)
(D) SOFTWARE: WordPerfect 5.1/WordPerfect Editor

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: U.S. 08/182,961
(B) FILING DATE: 14 JAN 1994

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: I.C. WADDEY, JR.

(B) REGISTRATION NUMBER: 25,180

(C) REFERENCE/DOCKET NUMBER: 0216-9409

(ix) TELECOMMUNICATION INFORMATION (O):

(A) TELEPHONE: (615) 242-2400

(B) TELEFAX: (615) 242-2221

(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 264

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(v) ORIGINAL SOURCE

(A) ORGANISM: Homo sapiens sapiens

(C) INDIVIDUAL/ISOLATE: sample of non-comedo DCIS

(D) DEVELOPMENTAL STAGE: adult

(F) TISSUE TYPE: female breast

(G) CELL TYPE: ductal carcinoma in situ

(H) CELL LINE: not derived from a cell line

(I) ORGANELLE: no

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: cDNA library derived from human

(B) CLONE: obtained from identification of differential gene expression

(viii) POSITION IN GENOME:

(A) CHROMOSOME/SEGMENT: unknown

(B) MAP POSITION: unknown

(C) UNITS: unknown

(ix) FEATURE:

(A) NAME/KEY: DCIS-1

(B) LOCATION: GenBank accession no. L2736

(C) IDENTIFICATION METHOD: microscopically-directed sampling and differential display

(D) OTHER INFORMATION: gene encoding M2 subunit of humanribonucleotide reductase

(x) PUBLICATION INFORMATION: unpublished

(K) RELEVANT RESIDUES IN SEQ ID NO: 1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
TTGGGAATTG GGTACGCGGG CCCCCCACTG TGCCGAATTC CTGCATGCGG GGGATCCACT  60
AGTTCAGAGC AGGCCGCCAC CCGTAGGACT CCAGCTTTTG TTCGTTCCCT TTAGTGAGGG 120
TTAATTTTCG AGCTTGGCGT AATCATGGTC ATAGCTGTTT CCTGTGTGAA ATTGTTATCC 180
GCTCACAATT CCACACAACA TACGAGCCGG AAGCATAAAA GTGTAAAGCC TGGGGTGCCT 240
AATGAGTGAG CTAACTCACA TTAA                                        264
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 73

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(v) ORIGINAL SOURCE

ORGANISM: Homo sapiens sapiens

(C) INDIVIDUAL/ISOLATE: sample of non-comedo DCIS

(D) DEVELOPMENTAL STAGE: adult

(F) TISSUE TYPE: female breast

(G) CELL TYPE: ductal carcinoma in situ

(H) CELL LINE: not derived from a cell line

(I) ORGANELLE: no

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: cDNA library derived from human

(B) CLONE: obtained from identification of differential gene expression

(viii) POSITION IN GENOME:

(A) CHROMOSOME/SEGMENT: unknown

(B) MAP POSITION: unknown

(C) UNITS: unknown

(ix) FEATURE:

(A) NAME/KEY: DCIS-2

(B) LOCATION: GenBank accession no. L27637

(C) IDENTIFICATION METHOD: microscopically-directed sampling and differential display

(x) PUBLICATION INFORMATION: unpublished

(K) RELEVANT RESIDUES IN SEQ ID NO: 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

TAGCCCGGTT ATCGAAATAG CCACAGCGCC TCTTCACTAT CAGCAGTACG CCGCCCAGTT 60
GTACGGACAC GGA 73

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(v) ORIGINAL SOURCE

(A) ORGANISM: Homo sapiens sapiens

(C) INDIVIDUAL/ISOLATE: sample of non-comedo DCIS

(D) DEVELOPMENTAL STAGE: adult

(F) TISSUE TYPE: female breast

(G) CELL TYPE: ductal carcinoma in situ

(H) CELL LINE: not derived from a cell line

(I) ORGANELLE: no

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: cDNA library derived from human

(B) CLONE: obtained from identification of differential gene expression

(viii) POSITION IN GENOME:

(A) CHROMOSOME/SEGMENT: unknown

(B) MAP POSITION: unknown

(C) UNITS: unknown

(ix) FEATURE:

(A) NAME/KEY: DCIS-3

(B) LOCATION: L27638

(C) IDENTIFICATION METHOD: microscopically-directed sampling and differential display

(x) PUBLICATION INFORMATION: unpublished

(K) RELEVANT RESIDUES IN SEQ ID NO: 3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TGCCCGATGT GTGTCGTACA ACTGGCGCTG TGGCTGATTT CGATAA 46

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 72

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(v) ORIGINAL SOURCE

(A) ORGANISM: Homo sapiens sapiens

(C) INDIVIDUAL/ISOLATE: sample of non-comedo DCIS

(D) DEVELOPMENTAL STAGE: adult

(F) TISSUE TYPE: female breast

(G) CELL TYPE: ductal carcinoma in situ

(H) CELL LINE: not derived from a cell line

(I) ORGANELLE: no

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: cDNA library derived from human

(B) CLONE: obtained from identification of differential gene expression

(viii) POSITION IN GENOME:

(A) CHROMOSOME/SEGMENT: unknown

(B) MAP POSITION: unknown

(C) UNITS: unknown

(ix) FEATURE:

(A) NAME/KEY: DCIS-4

(B) LOCATION: L27640

(C) IDENTIFICATION METHOD: microscopically-directed sampling and differential display

(x) PUBLICATION INFORMATION: unpublished

(K) RELEVANT RESIDUES IN SEQ ID NO: 4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TAGCCCATGA GTTCGTGTCC GTACAACTGG GGCGCTGTGG CTGATTTCGA TANNNNNAGC 60
ATCAGCCCGA CG 72

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 84

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(v) ORIGINAL SOURCE

(A) ORGANISM: Homo sapiens sapiens

(C) INDIVIDUAL/ISOLATE: sample of non-comedo DCIS

(D) DEVELOPMENTAL STAGE: adult

(F) TISSUE TYPE: female breast

(G) CELL TYPE: ductal carcinoma in situ

(H) CELL LINE: not derived from a cell line

(I) ORGANELLE: no

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: cDNA library derived from human

(B) CLONE: obtained from identification of differential gene expression

(viii) POSITION IN GENOME:

(A) CHROMOSOME/SEGMENT: unknown

(B) MAP POSITION: unknown

(C) UNITS: unknown

(ix) FEATURE:

(A) NAME/KEY: DCIS-5

(B) LOCATION: L27641

(C) IDENTIFICATION METHOD: microscopically-directed sampling and differential display

(x) PUBLICATION INFORMATION: unpublished

(K) RELEVANT RESIDUES IN SEQ ID NO: 5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
TAGCCCGGTT ATCGAAATCA GCCACAGCGC CTAACTTCTG CAGAAGCCTT TGACCATCAC  60
CAGTTGTACG GACACGAACT CATC                                         84
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 99

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(v) ORIGINAL SOURCE

(A) ORGANISM: Homo sapiens sapiens

(C) INDIVIDUAL/ISOLATE: sample of non-comedo DCIS

(D) DEVELOPMENTAL STAGE: adult

(F) TISSUE TYPE: female breast

(G) CELL TYPE: ductal carcinoma in situ

(H) CELL LINE: not derived from a cell line

(I) ORGANELLE: no

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: cDNA library derived from human

(B) CLONE: obtained from identification of differential gene expression

(viii) POSITION IN GENOME:

(A) CHROMOSOME/SEGMENT: unknown

(B) MAP POSITION: unknown

(C) UNITS: unknown

(ix) FEATURE:

(A) NAME/KEY: DCIS-6

(B) LOCATION: L27642

(C) IDENTIFICATION METHOD: microscopically-directed sampling and differential display

(x) PUBLICATION INFORMATION: unpublished

(K) RELEVANT RESIDUES IN SEQ ID NO: 6

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GTGGTTTCCG AAATTCCTGG GAAGGGGGGT GCTGGCGTGT GGAATTGTCG CGGCCCCTGG 60
TCTGCCGCGG CGTTTTTTGT CTACATTCGT CGTAGCTCG 99

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 88

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(v) ORIGINAL SOURCE

(A) ORGANISM: Homo sapiens sapiens

(C) INDIVIDUAL/ISOLATE: sample of non-comedo DCIS

(D) DEVELOPMENTAL STAGE: adult

(F) TISSUE TYPE: female breast

(G) CELL TYPE: ductal carcinoma in situ

(H) CELL LINE: not derived from a cell line

(I) ORGANELLE: no

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: cDNA library derived from human

(B) CLONE: obtained rom identification of differential gene expression

(viii) POSITION IN GENOME:

(A) CHROMOSOME/SEGMENT: unknown

(B) MAP POSITION: unknown

(C) UNITS: unknown

(ix) FEATURE:

(A) NAME/KEY: DCIS-7

(B) LOCATION: L27643

(C) IDENTIFICATION METHOD: microscopically-directed sampling and differential display

(x) PUBLICATION INFORMATION: unpublished

(K) RELEVANT RESIDUES IN SEQ ID NO: 7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ATCAGCGCGC GACATTCGGG TACCCGCGCC CCCCCCTCCG TCGGAATTCC TCGAGCCGGG 60
ATCCATAGGA TGTGGAGTTA GTTTTGTT 88

(2) INFORMATION FOR SEQ ID NO:8

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CGCGACGGCC GCGCGTCTGC CAGGG 25

(2) INFORMATION FOR SEQ ID NO:9

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CGCCCCTGCG TTACCCTCCC CGCCG 25

(2) INFORMATION FOR SEQ ID NO:10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GGATGGCGTC CTGTAACCCG ACGCT 25

(2) INFORMATION FOR SEQ ID NO:11

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

ACTGGGCTGT CCTGCGGTGG CGGGG 25

(2) INFORMATION FOR SEQ ID NO:12

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

CTGAGAGGTA GCCGCGCGGA GGCTG 25

(2) INFORMATION FOR SEQ ID NO:13

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GCCTGGCCGC GACACGGATT ACCGC 25

(2) INFORMATION FOR SEQ ID NO:14

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TTAGCGCATG GTGGACCTGG AGACG 25

(2) INFORMATION FOR SEQ ID NO:15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TGTGGTTACG TCAGCGAAGG TAATA 25

(2) INFORMATION FOR SEQ ID NO:16

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AGTCGCACGC ATGTCACGCT CCGCC 25

(2) INFORMATION FOR SEQ ID NO:17

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

TATCCAAGCG GCAGGCTACG AGGCC 25

(2) INFORMATION FOR SEQ ID NO:18

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GGCGCGCCCG ACGGTCTGGT ATCTA 25

(2) INFORMATION FOR SEQ ID NO:19

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

CTCCCTCCCC GGACTCGGGG TTAGT 25

(2) INFORMATION FOR SEQ ID NO:20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

ATGCGGGCGG CTCGGGCCTG GTCGC 25

(2) INFORMATION FOR SEQ ID NO:21

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CGTGAAGCCT ATGCCCTCCC TCAAC 25

(2) INFORMATION FOR SEQ ID NO:22

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GTGCCGTCGT AGCCCTTCAG CGATC 25

(2) INFORMATION FOR SEQ ID NO:23

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GCGACACTAG GCTCCCGGAG GAGGG 25

(2) INFORMATION FOR SEQ ID NO:24

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

TGGGCCAGGC CTCCGGGCCC GGTAT 25

(2) INFORMATION FOR SEQ ID NO:25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

CCGGAACTGC GATAGCGTCC GTCCC 25

(2) INFORMATION FOR SEQ ID NO:26

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

AGCGGACACC TGTTTCCCGA GAGCC 25

(2) INFORMATION FOR SEQ ID NO:27

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

AACGGGTGGA CATCCGCCTG CCGCC 25

(2) INFORMATION FOR SEQ ID NO:28

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

TGAACCACGA TGTCAATCGT CCCGA 25

(2) INFORMATION FOR SEQ ID NO:29

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

TCATCCCCGC CGAAAGACGC TCGCC 25

(2) INFORMATION FOR SEQ ID NO:30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

ATAGGCTGCG GCACGCGCTG GGACT 25

(2) INFORMATION FOR SEQ ID NO:31

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

GACCAGGTGC GCACGAGCAT GTACA 25

(2) INFORMATION FOR SEQ ID NO:32

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

AGCGTAGTCA TCGGCCTTCG CGCCC 25

(2) INFORMATION FOR SEQ ID NO:33

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

GGCCCCTAGC CCAGGGTGAA GCCCA 25

(2) INFORMATION FOR SEQ ID NO:34

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CCCAGTGCTA CGGGCCGCCC CAAGC 25

(2) INFORMATION FOR SEQ ID NO:35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CCTTCCTGGG TTACCTGCCC TCGGG 25

(2) INFORMATION FOR SEQ ID NO:36

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

TCCGGACAGC AGCCACGCCA AGGGC 25

(2) INFORMATION FOR SEQ ID NO:37

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

ACGCGCTGGT CCACCGAGGC CTGAT 25

(2) INFORMATION FOR SEQ ID NO:38

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

CGATGCAAGG CCAGCAGCAC TCGAC 25

(2) INFORMATION FOR SEQ ID NO:39

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CCCCCGGAGC GGACCACCGG ACGTG 25

(2) INFORMATION FOR SEQ ID NO:40

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

AGCGGGGAGG GATCGGGGGC CAAGC 25

(2) INFORMATION FOR SEQ ID NO:41

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GCCTGGTGTA GGCAGGCAGC TCTTA 25

(2) INFORMATION FOR SEQ ID NO:42

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

CCACCCCTGT AGTGCGGGCT GCGAG 25

(2) INFORMATION FOR SEQ ID NO:43

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

GGAACCCGAC GCCCGTCCAG GGTTC 25

(2) INFORMATION FOR SEQ ID NO:44

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

TCGGGCAGCA AGGCCGGGAC GCTCC 25

(2) INFORMATION FOR SEQ ID NO:45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GACGGGGGAC GGGCTAGGTG GCTTA 25

(2) INFORMATION FOR SEQ ID NO:46

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(A) DESCRIPTION: PCR primer

(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no

(v) FRAGMENT TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

CTTGTTGCCG GCGGAGAGGG CTGCC 25

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5712

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(v) ORIGINAL SOURCE

(A) ORGANISM: Homo sapiens sapiens

(C) INDIVIDUAL/ISOLATE:

(D) DEVELOPMENTAL STAGE: adult

(F) TISSUE TYPE: female breast

(G) CELL TYPE: ductal carcinoma in situ, invasive breast cancer and normal breast tissue

(H) CELL LINE: not derived from a cell line

(I) ORGANELLE: no

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: cDNA library derived from human

(B) CLONE: obtained using published sequence

(viii) POSITION IN GENOME:

(A) CHROMOSOME/SEGMENT: unknown

(B) MAP POSITION: unknown

(C) UNITS: unknown

(ix) FEATURE:

(A) NAME/KEY: BRCA1

(B) LOCATION: GenBank accession no. U14680

(C) IDENTIFICATION METHOD: microscopically-directed sampling and nuclease protection assay

(D) OTHER INFORMATION: gene encoding BRCA1 protein

(x) PUBLICATION INFORMATION:

(A) AUTHORS: Miki, Y., et. al.

(B) TITLE: A strong candidate gene for the breast and ovarian cancer susceptibility gene BRCA1.

(C) JOURNAL: Science

(D) VOLUME: **266**

(E) PAGES: 66-71

(F) DATE: 1994

(K) RELEVANT RESIDUES IN SEQ ID NO: 47

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

```
agctcgctga gacttcctgg accccgcacc aggctgtggg gtttctcaga taactgggcc  60
cctgcgctca ggaggccttc accctctgct ctgggtaaag ttcattggaa cagaaagaa   119
atg gat tta tct gct ctt cgc gtt gaa gaa gta caa aat gtc att aat     167
Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
1               5                  10                  15
gct atg cag aaa atc tta gag tgt ccc atc tgt ctg gag ttg atc aag    215
Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
            20                  25                  30
gaa cct gtc tcc aca aag tgt gac cac ata ttt tgc aaa ttt tgc atg    263
Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
        35                  40                  45
ctg aaa ctt ctc aac cag aag aaa ggg cct tca cag tgt cct tta tgt    311
Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
    50                  55                  60
aag aat gat ata acc aaa agg agc cta caa gaa agt acg aga ttt agt    359
Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
65                  70                  75                  80
caa ctt gtt gaa gag cta ttg aaa atc att tgt gct ttt cag ctt gac    407
Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
                85                  90                  95
aca ggt ttg gag tat gca aac agc tat aat ttt gca aaa aag gaa aat    455
Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
            100                 105                 110
aac tct cct gaa cat cta aaa gat gaa gtt tct atc atc caa agt atg    503
Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
        115                 120                 125
ggc tac aga aac cgt gcc aaa aga ctt cta cag agt gaa ccc gaa aat    551
Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
    130                 135                 140
cct tcc ttg cag gaa acc agt ctc agt gtc caa ctc tct aac ctt gga    599
Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
145                 150                 155                 160
```

```
cact gtg aga act ctg agg aca aag cag cgg ata caa cct caa aag acg    647
Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
                165                 170                 175

tct gtc tac att gaa ttg gga tct gat tct tct gaa gat acc gtt aat    695
Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
            180                 185                 190
aag gca act tat tgc agt gtg gga gat caa gaa ttg tta caa atc acc    743
Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
        195                 200                 205
cct caa gga acc agg gat gaa atc agt ttg gat tct gca aaa aag gct    791
Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
    210                 215                 220
gct tgt gaa ttt tct gag acg gat gta aca aat act gaa cat cat caa    839
Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
225                 230                 235                 240
ccc agt aat aat gat ttg aac acc act gag aag cgt gca gct gag agg    887
Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg
                245                 250                 255
cat cca gaa aag tat cag ggt agt tct gtt tca aac ttg cat gtg gag    935
His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu
            260                 265                 270
cca tgt ggc aca aat act cat gcc agc tca tta cag cat gag aac agc    983
Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser
        275                 280                 285
agt tta tta ctc act aaa gac aga atg aat gta gaa aag gct gaa ttc   1031
Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe
    290                 295                 300
tgt aat aaa agc aaa cag cct ggc tta gca agg agc caa cat aac aga   1079
Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg
305                 310                 315                 320
tgg gct gga agt aag gaa aca tgt aat gat agg cgg act ccc agc aca   1127
Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr
                325                 330                 335
gaa aaa aag gta gat ctg aat gct gat ccc ctg tgt gag aga aaa gaa   1175
Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu
            340                 345                 350
tgg aat aag cag aaa ctg cca tgc tca gag aat cct aga gat act gaa   1223
Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu
        355                 360                 365
gat gtt cct tgg ata aca cta aat agc agc att cag aaa gtt aat gag   1271
Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu
    370                 375                 380
tgg ttt tcc aga agt gat gaa ctg tta ggt tct gat gac tca cat gat   1319
Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp
385                 390                 395                 400
ggg gag tct gaa tca aat gcc aaa gta gct gat gta ttg gac gtt cta   1367
Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu
                405                 410                 415
```

```
aat gag gta gat gaa tat tct ggt tct tca gag aaa ata gac tta ctg  1415
Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu
          420                 425                 430
gcc agt gat cct cat gag gct tta ata tgt aaa agt gaa aga gtt cac  1463
Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Asp Arg Val His
       435                 440                 445
tcc aaa tca gta gag agt aat att gaa gac aaa ata ttt ggg aaa acc  1511
Ser Lys Ser Val Glu Ser Asp Ile Glu Asp Lys Ile Phe Gly Lys Thr
    450                 455                 460
tat cgg aag aag gca agc ctc ccc aac tta agc cat gta act gaa aat  1559
Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn
465                 470                 475                 480
cta att ata gga gca ttt gtt act gag cca cag ata ata caa gag cgt  1607
Leu Ile Ile Gly Ala Phe Val Ser Glu Pro Gln Ile Ile Gln Glu Arg
                485                 490                 495
ccc ctc aca aat aaa tta aag cgt aaa agg aga cct aca tca ggc ctt  1655
Pro Leu Thr Asn Lys Leu Lys Aeg Lys Arg Arg Pro Thr Ser Gly Leu
                500                 505                 510
cat cct gag gat ttt atc aag aaa gca gat ttg gca gtt caa aag act  1703
His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr
          515                 520                 525
cct gaa atg ata aat cag gga act aac caa acg gag cag aat ggt caa  1751
Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln
       530                 535                 540
gtg atg aat att act aat agt ggt cat gag aat aaa aca aaa ggt gat  1799
Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp
    545                 550                 555
tct att cag aat gag aaa aat cct aac cca ata gaa tca ctc gaa aaa  1847
Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys
560                 565                 570                 575
gaa tct gct ttc aaa acg aaa gct gaa cct ata agc agc agt ata agc  1895
Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser
                580                 585                 590
aat atg gaa ctc gaa tta aat atc cac aat tca aaa gca cct aaa aag  1943
Asn Glu Leu Glu Leu Asn Ile Met His Asn Ser Lys Ala Pro Lys Lys
          595                 600                 605
aat agg ctg agg agg aag tct tct acc agg cat att cat gcg ctt gaa  1991
Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Als Leu Glu
       610                 615                 620
cta gta gtc agt aga aat cta agc cca cct aat tgt act gaa ttg caa  2039
Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln
    625                 630                 635
att gat agt tgt tct agc agt gaa gag ata aag aaa aaa aag tac aac  2087
Ile Asp Ser Cys Ser Ser Ser Glu Glu Ile Lys Lys Lys Lys Tyr Asn
640                 645                 650                 655
caa atg cca gtc agg cac agc aga aac cta caa ctc atg gaa ggt aaa  2135
Gln Met Pro Val Arg His Ser Arg Asn Leu Gln Leu Met Glu Gly Lys
                660                 665                 670
```

```
gaa cct gca act gga gcc aag aag agt aac aag cca aat gaa cag aca   2183
Glu Pro Ala Thr Gly Ala Lys Lys Ser Asn Lys Pro Asn Glu Gln Thr
          675                 680                 685
agt aaa aga cat gac agc gat act ttc cca gag ctg aag tta aca aat   2231
Ser Lys Arg His Asp Ser Asp Thr Phe Pro Glu Leu Lys Leu Thr Asn
       690                 695                 700
gca cct ggt tct ttt act aag tgt tca aat acc agt gaa ctt aaa gaa   2279
Ala Pro Gly Ser Phe Thr Lys Cys Ser Asn Thr Ser Glu Leu Lys Glu
    705                 710                 715
ttt gtc aat cct agc ctt cca aga gaa gaa aaa gaa gag aaa cta gaa   2327
Phe Val Asn Pro Ser Leu Pro Arg Glu Glu Lys Glu Glu Lys Leu Glu
720                 725                 730                 735
aca gtt aaa gtg tct aat aat gct gaa gac ccc aaa gat ctc atg tta   2375
Thr Val Lys Val Ser Asn Asn Ala Glu Asp Pro Lys Asp Leu Met Leu
                740                 745                 750
agt gga gaa agg gtt ttg caa act gaa aga tct gta gag agt agc agt   2423
Ser Gly Glu Arg Val Leu Gln Thr Glu Arg Ser Val Glu Ser Ser Ser
          755                 760                 765
att tca ttg gta cct ggt act gat tat ggc act cag gaa agt atc tcg   2471
Ile Ser Leu Val Pro Gly Thr Asp Tyr Gly Thr Gln Glu Ser Ile Ser
       770                 775                 780
tta ctg gaa gtt agc act cta ggg aag gca aaa aca gaa cca aat aaa   2519
Leu Leu Glu Val Ser Thr Leu Gly Lys Ala Lys Thr Glu Pro Asn Lys
    785                 790                 795
tgt gtg agt cag tgt gca gca ttt gaa aac ccc aag gga cta att cat   2567
Cys Val Ser Gln Cys Ala Ala Phe Glu Asn Pro Lys Gly Leu Ile His
800                 805                 810                 815
ggt tgt tcc aaa gat aat aga aat gac aca gaa ggc ttt aag tat cca   2615
Gly Cys Ser Lys Asp Asn Arg Asn Asp Thr Glu Gly Phe Lys Tyr Pro
                820                 825                 830
ttg gga cat gaa gtt aac cac agt cgg gaa aca agc ata gaa atg gaa   2663
Leu Gly His Glu Val Asn His Ser Arg Glu Thr Ser Ile Glu Met Glu
          835                 840                 845
gaa agt gaa ctt gat gct cag tat ttg cag aat aca ttc aag gtt tca   2711
Glu Ser Glu Leu Asp Ala Gln Tyr Leu Gln Asn Thr Phe Lys Val Ser
       850                 855                 860
aag cgc cag tca ttt gct ccg ttt tca aat cca gga aat gca gaa gag   2759
Lys Arg Gln Ser Phe Ala Pro Phe Ser Asn Pro Gly Asn Ala Glu Glu
    865                 870                 875
gaa tgt gca aca ttc tct gcc cac tct ggg tcc tta aag aaa caa agt   2807
Glu Cys Ala Thr Phe Ser Ala His Ser Gly Ser Leu Lys Lys Gln Ser
880                 885                 890                 895
cca aaa gtc act ttt gaa tgt gaa caa aag gaa gaa aat caa gga aag   2855
Pro Lys Val Thr Phe Glu Cys Glu Gln Lys Glu Glu Asn Gln Gly Lys
                900                 905                 910
aat gag tct aat atc aag cct gta cag aca gtt aat atc act gca ggc   2903
Asn Glu Ser Asn Ile Lys Pro Val Gln Thr Val Asn Ile Thr Ala Gly
          915                 920                 925
```

ttt cct gtg gtt ggt cag aaa gat aag cca gtt gat aat gcc aaa tgt 2951
Phe Pro Val Val Gly Gln Lys Asp Lys Pro Val Asp Asn Ala Lys Cys
930 935 940

agt atc aaa gga ggc tct agg ttt tgt cta tca tct cag ttc aga ggc 2999
Ser Ile Lys Gly Gly Ser Arg Phe Cys Leu Ser Ser Gln Phe Arg Gly
945 950 955

aac gaa act gga ctc att act cca aat aaa cat gga ctt tta caa aac 3047
Asn Glu Thr Gly Leu Ile Thr Pro Asn Lys His Gly Leu Leu Gln Asn
960 965 970 975

cca tat cgt ata cca cca ctt ttt ccc atc aag tca ttt gtt aaa act 3095
Pro Tyr Arg Ile Pro Pro Leu Phe Pro Ile Lys Ser Phe Val Lys Thr
980 985 990

aaa tgt aag aaa aat ctg cta gag gaa aac ttt gag gaa cat tca atg 3143
Lys Cys Lys Lys Asn Leu Leu Glu Glu Asn Phe Glu Glu His Ser Met
995 1000 1005

tca cct gaa aga gaa atg gga aat gag aac att cca agt aca gtg agc 3191
Ser Pro Glu Arg Glu Met Gly Asn Glu Asn Ile Pro Ser Thr Val Ser
1010 1015 1020

aca att agc cgt aat aac att aga gaa aat gtt ttt aaa gaa gcc agc 3239
Thr Ile Ser Arg Asn Asn Ile Arg Glu Asn Val Phe Lys Glu Ala Ser
1025 1030 1035

tca agc aat att aat gaa gta ggt tcc agt act aat gaa gtg ggc tcc 3287
Ser Ser Asn Ile Asn Glu Val Gly Ser Ser Thr Asn Glu Val Gly Ser
1040 1045 1050 1055

agt att aat gaa ata ggt tcc agt gat gaa aac att caa gca gaa cta 3335
Ser Ile Asn Glu Ile Gly Ser Ser Asp Glu Asn Ile Gln Ala Glu Leu
1060 1065 1070

ggt aga aac aga ggg cca aaa ttg aat gct atg ctt aga tta ggg gtt 3383
Gly Arg Asn Arg Gly Pro Lys Leu Asn Ala Met Leu Arg Leu Gly Val
1075 *1080* *1085*

ttg caa cct gag gtc tat aaa caa agt ctt cct gga agt aat tgt aag *3431*
Leu Gln Pro Glu Val Tyr Lys Gln Ser Leu Pro Gly Ser Asn Cys Lys
1090 1095 1100

cat cct gaa ata aaa aag caa gaa tat gaa gaa gta gtt cag act gtt 3479
His Pro Glu Ile Lys Lys Gln Glu Tyr Glu Glu Val Val Gln Thr Val
1105 1110 1115

aat aca gat ttc tct cca tat ctg att tca gat aac tta gaa cag cct 3527
Asn Thr Asp Phe Ser Pro Tyr Leu Ile Ser Asp Asn Leu Glu Gln Pro
1120 1125 1130 1135

atg gga agt agt cat gca tct cag gtt tgt tct gag aca cct gat gac 3575
Met Gly Ser Ser His Ala Ser Gln Val Cys Ser Glu Thr Pro Asp Asp
1140 1145 1150

ctg tta gat gat ggt gaa ata aag gaa gat act agt ttt gct gaa aat 3623
Leu Leu Asp Asp Gly Glu Ile Lys Glu Asp Thr Ser Phe Ala Glu Asn
1155 1160 1165

gac att aag gaa agt tct gct gtt ttt agc aaa agc gtc cag aaa gga 3671
Asp Ile Lys Glu Ser Ser Ala Val Phe Ser Lys Ser Val Gln Lys Gly
1170 1175 1180

```
gag ctt agc agg agt cct agc cct ttc acc cat aca cat ttg gct cag   3719
Glu Leu Ser Arg Ser Pro Ser Pro Phe Thr His Thr His Leu Ala Gln
    1185                1190                1195
ggt tac cga aga ggg gcc aag aaa tta gag tcc tca gaa gag aac tta   3767
Gly Tyr Arg Arg Gly Ala Lys Lys Leu Glu Ser Ser Glu Glu Asn Leu
1200                1205                1210                1215
tct agt gag gat gaa gag ctt ccc tgc ttc caa cac ttg tta ttt ggt   3815
Ser Ser Glu Asp Glu Glu Leu Pro Cys Phe Gln His Leu Leu Phe Gly
                1220                1225                1230
aaa gta aac aat ata cct tct cag tct act agg cat agc acc gtt gct   3863
Lys Val Asn Asn Ile Pro Ser Gln Ser Thr Arg His Ser Thr Val Ala
            1235                1240                1245
acc gag tgt ctg tct aag aac aca gag gag aat tta tta tca ttg aag   3911
Thr Glu Cys Leu Ser Lys Asn Thr Glu Glu Asn Leu Leu Ser Leu Lys
        1250                1255                1260
aat agc tta aat gac tgc agt aac cag gta ata ttg gca aag gca tct   3959
Asn Ser Leu Asn Asp Cys Ser Asn Gln Val Ile Leu Ala Lys Als Ser
    1265                1270                1275
cag gaa cat cac ctt agt gag gaa aca aaa tgt tct gct agc ttg ttt   4007
Gln Glu His His Leu Ser Glu Glu Thr Lys Cys Ser Ala Ser Leu Phe
1280                1285                1290                1295
tct tca cag tgc agt gaa ttg gaa gac ttg act gca aat aca aac acc   4055
Ser Ser Gln Cys Ser Glu Leu Glu Asp Leu Thr Ala Asn Thr Asn Thr
                1300                1305                1310
cag gat cct ttc ttg att ggt tct tcc aaa caa atg agg cat cag tct   4103
Gln Asp Pro Phe Leu Ile Gly Ser Ser Lys Gln Met Arg His Gln Ser
            1315                1320                1325
gaa agc cag gga gtt ggt ctg agt gac aag gaa ttg gtt tca gat gat   4151
Glu Ser Gln Gly Val Gly Leu Ser Asp Lys Glu Leu Val Ser Asp Asp
        1330                1335                1340
gaa gaa aga gga acg ggc ttg gaa gaa aat aat caa gaa gag caa agc   4199
Glu Glu Arg Gly Thr Gly Leu Glu Glu Asn Asn Gln Glu Glu Gln Ser
    1345                1350                1355
atg gat tca aac tta ggt gaa gca gca tct ggg tgt gag agt gaa aca   4247
Met Asp Ser Asn Leu Gly Glu Ala Ala Ser Gly Cys Glu Ser Glu Thr
1360                1365                1370                1375
agc gtc tct gaa gac tgc tca ggg cta tcc tct cag agt gac att tta   4295
Ser Val Ser Glu Asp Cys Ser Gly Leu Ser Ser Gln Ser Asp Ile Leu
                1380                1385                1390
acc act cag cag agg gat acc atg caa cat aac ctg ata aag ctc cag   4343
Thr Thr Gln Gln Arg Asp Thr Met Gln His Asn Leu Ile Lys Leu Gln
            1395                1400                1405
cag gaa atg gct gaa cta gaa gct gtg tta gaa cag cat ggg agc cag   4391
Gln Glu Met Ala Glu Leu Glu Ala Val Leu Glu Gln His Gly Ser Gln
        1410                1415                1420
cct tct aac agc tac cct tcc atc ata agt gac tct tct gcc ctt gag   4439
Pro Ser Asn Ser Tyr Pro Ser Ile Ile Ser Asp Ser Ser Ala Leu Glu
    1425                1430                1435
```

```
gac ctg cga aat cca gaa caa agc aca tca gaa aaa gca gta tta act   4487
Asp Leu Arg Asn Pro Glu Gln Ser Thr Ser Glu Lys Val Leu Gln Thr
1440            1445                1450                1455
tca cag aaa agt agt gaa tac cct ata agc cag aat cca gaa ggc ctt   4535
Ser Gln Lys Ser Ser Glu Tyr Pro Ile Ser Gln Asn Pro Glu Gly Xaa
            1460                1465                1470
tct gct gac aag ttt gag gtg tct gca gat agt tct acc agt aaa aat   4583
Ser Ala Asp Lys Phe Glu Val Ser Ala Asp Ser Ser Thr Ser Lys Asn
        1475                1480                1485
aaa gaa cca gga gtg gaa agg tca tcc cct tct aaa tgc cca tca tta   4631
Lys Glu Pro Gly Val Glu Arg Ser Ser Pro Ser Lys Cys Pro Ser Leu
    1490                1495                1500
gat gat agg tgg tac atg cac agt tgc tct ggg agt ctt cag aat aga   4679
Asp Asp Arg Trp Tyr Met His Ser Cys Ser Gly Ser Leu Gln Asn Arg
1505                1510                1515                1520
aac tac cca tct caa gag gag ctc att aag gtt gtt gat gtg gag gag   4727
Asn Tyr Pro Pro Gln Glu Glu Leu Ile Lys Val Val Asp Val Glu Glu
                1525                1530                1535
caa cag ctg gaa gag tct ggg cca cac gat ttg acg gaa aca tct tac   4775
Gln Gln Leu Glu Glu Ser Gly Pro His Asp Leu Thr Glu Thr Ser Tyr
            1540                1545                1550
ttg cca agg caa gat cta gag gga acc cct tac ctg gaa tct gga atc   4823
Leu Pro Arg Gln Asp Leu Glu Gly Thr Pro Tyr Leu Glu Ser Gly Ile
        1555                1560                1565
agc ctc ttc tct gat gac cct gaa tct gat cct tct gaa gac aga gcc   4871
Ser Leu Phe Ser Asp Asp Pro Glu Ser Asp Pro Ser Glu Asp Arg Ala
    1570                1575                1580
cca gag tca gct cgt gtt ggc aac ata cca tct tca acc tct gca ttg   4919
Pro Glu Ser Ala Arg Val Gly Asn Ile Pro Ser Ser Thr Ser Ala Leu
1585                1590                1595                1600
aaa gtt ccc caa ttg aaa gtt gca gaa tct gcc cag agt cca gct gct   4967
Lys Val Pro Gln Leu Lys Val Ala Glu Ser Ala Gln Ser Pro Ala Ala
                1605                1610                1615
gct cat act act gat act gct ggg tat aat gca atg gaa gaa agt gtg   5015
Ala His Thr Thr Asp Thr Ala Gly Tyr Asn Ala Met Glu Glu Ser Val
            1620                1625                1630
agc agg gag aag cca gaa ttg aca gct tca aca gaa agg gtc aac aaa   5063
Ser Arg Glu Lys Pro Glu Leu Thr Ala Ser Thr Glu Arg Val Asn Lys
        1635                1640                1645
aga atg tcc atg gtg gtg tct ggc ctg acc cca gaa gaa ttt atg ctc   5111
Arg Met Ser Met Val Val Ser Gly Leu Thr Pro Glu Glu Phe Met Leu
    1650                1655                1660
gtg tac aag ttt gcc aga aaa cac cac atc act tta act aat cta att   5159
Val Tyr Lys Phe Ala Arg Lys His His Ile Thr Leu Thr Asn Leu Ile
1665                1670                1675                1680
act gaa gag act act cat gtt gtt atg aaa aca gat gct gag ttt gtg   5207
Thr Glu Glu Thr Thr His Val Val Met Lys Thr Asp Ala Glu Phe Val
                1685                1690                1695
```

```
tgt gaa cgg aca ctg aaa tat ttt cta gga att gcg gga gga aaa tgg   5255
Cys Glu Arg Thr Leu Lys Tyr Phe Leu Gly Ile Ala Gly Gly Lys Trp
          1700                1705                1710
gta gtt agc tat ttc tgg gtg acc cag tct att aaa gaa aga aaa atg   5303
Val Val Ser Tyr Phe Trp Val Thr Gln Ser Ile Lys Glu Arg Lys Met
       1715                1720                1725
ctg aat gag cat gat ttt gaa gtc aga gga gat gtg gtc aat gga aga   5351
Leu Asn Glu His Asp Phe Glu Val Arg Gly Asp Val Val Asn Gly Arg
   1730                1735                1740
aac cac caa ggt cca aag cga gca aga gaa tcc cag gac aga aag atc   5399
Asn His Gln Gly Pro Lys Arg Ala Arg Glu Ser Gln Asp Arg Lys Ile
1745                1750                1755                1760
ttc agg ggg cta gaa atc tgt tgc tat ggg ccc ttc acc aac atg ccc   5447
Phe Arg Gly Leu Glu Ile Cys Cys Tyr Gly Pro Phe Thr Asn Met Pro
               1765                1770                1775
aca gat caa ctg gaa tgg atg gta cag ctg tgt ggt gct tct gtg gtg   5495
Thr Asp Gln Leu Glu Trp Met Val Gln Leu Cys Gly Ala Ser Val Val
          1780                1785                1790
aag gag ctt tca tca ttc acc ctt ggc aca ggt gtc cac cca att gtg   5543
Lys Glu Leu Ser Ser Phe Thr Leu Gly Thr Gly Val His Pro Ile Val
       1795                1800                1805
gtt gtg cag cca gat gcc tgg aca gag gac aat ggc ttc cat gca att   5591
Val Val Gln Pro Asp Ala Trp Tht Glu Asp Asn Gly Phe His Ala Ile
   1810                1815                1820
ggg cag atg tgt gag gca cct gtg gtg acc cga gag tgg gtg ttg gac   5639
Gly Gln Met Cys Glu Ala Pro Val Val Thr Arg Glu Trp Val Leu Asp
1825                1830                1835                1840
agt gta gca ctc tac cag tgc cag gag ctg gac acc tac ctg ata ccc   5687
Ser Val Ala Leu Tyr Gln Cys Gln Glu Leu Asp Thr Tyr Leu Ile Pro
               1845                1850                1855
cag atc ccc cac agc cac tac tgat                                 5712
Gln Ile Pro His Ser His Tyr
          1860
```

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1237

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA regulatory sequence

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(v) ORIGINAL SOURCE

(A) ORGANISM: Homo sapiens sapiens

(C) INDIVIDUAL/ISOLATE:

(D) DEVELOPMENTAL STAGE: adult

(F) TISSUE TYPE: female breast

(G) CELL TYPE: normal breast

(H) CELL LINE: not derived from a cell line

(I) ORGANELLE: no

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: cDNA library derived from human

(B) CLONE: obtained using published sequence

(viii) POSITION IN GENOME:

(A) CHROMOSOME/SEGMENT: unknown

(B) MAP POSITION: unknown

(C) UNITS: unknown

(ix) FEATURE:

(A) NAME/KEY: BRCA1 promoter

(B) LOCATION:

(C) IDENTIFICATION METHOD: restriction enzyme digest

(D) OTHER INFORMATION: DNA sequence regulating gene encoding BRCA1 protein

(x) PUBLICATION INFORMATION:

(A) AUTHORS: Brown et al.

(B) TITLE: Scientific Correspondence

(C) JOURNAL: Nature

(D) VOLUME: 372

(E) PAGES: 733

(F) DATE: 22/29 DECEMBER 1994

(K) RELEVANT RESIDUES IN SEQ ID NO: 48

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

TTCCGGGACT CTACTACCTT TACCCAGACG AGAGGGTGAA GGCCTCCTGA TCGCAGGGGC 60

CCAGTTATCT GAGAAACCCC ACAGCCTGGT GCGGGGTCCA GGAAGTCTCA GCGAGCTCAC 120

GCCGCGCAGT CGCAGTTTTA ATTTATCTGT AATTCCCGCG CTTTTCCGTT GCCACGGAAA 180
CCAAGGGGCT ACCGCTAAGC AGCAGCCTCT CAGAATACGA AATCAAGGTA CAATCAGAGG 240
AAGGGAGGGA CAGAAAGAGC CAAGCGTCTC TCGGGGCTCT GGATTGGCCA CCCAGTCTGC 300
CCCCGGATGA CGTAAAAGGA AAGAGACGGA AGAGGAAGAA TTCTACCTGA GTTCGCCGTA 360
AAGCGCCCGC CCTCTCGCCT CTACGCTTCC AGTTGCGGCT TATTACGTCA CAGTAATTGC 420
TGTACCAAGG TCAGAATCGC CACCTGAGGC CTGAATATCA GCGTAAGATA GTGTCCAAAG 480
CAGTCTTAAG AAGAGGTCCC ATTACCCCAC TCTTTCCGCC CTAATGGAGT CCTCCAGTTT 540
AGGTAAATAA AAGGATTGTT GGGAGGTGGA GGGAAAGAAC TACTATTTCC AACATGCATT 600
GCGGAACGAA AGGCCTTGGC CACACTGTTC CTTGGAAACT GTAGTCTTAT GGAGAGGAAC 660
ATCCAATACC AAAGCGGGCA CAATTCTCAC GGAAATCCAG TGGATAGATT GGAGACCTCC 720
GCGGGCTTAT ACATGTCAAC AGTAATATTG GGTTGTTATG TTCTCCTATC TTGAGAGCAG 780
AGACTAGGCC AAAAAAAGAT ATAGGAAGAC TACGATTCCC ATCCAGCCCC ACGAGTCTCG 840
GGCAAGTAGT CCTCTAAGGT CAGTGGCCTG CGGGGACGCA GTGGGCGCCG AATTTGCCTG 900
GGGAAGGGGA AATCCCTCTC TGGTCACATC TGCGCACTCC TAGTTCCGCC CCTCAGCATC 960
AATGTTTGTT ATTGTTGTTC GGGTTCAGGT TGCTTCTGCC CCGCCCCATC GACGCAATCT 1020
CCACCAATCA ATGGCGTGGT CGTTTTGAGG GACAAGTGGT GAGAGCCAAT CATCTTGGCG 1080
AACACTCGGA GAAACAGGGG ACTAGTTACT GTCTTTATCC GCCATGTTAG ATTCACCCCA 1140
CAGGGATAGC GGCAGAGCCG GTAGCGGACG GTCCTTGCAT TGGCCTCCGG CAGGCGCCCC 1200
CCGGGGGCGG GAAGCTGGTA AGGAAGCAGC TGCGGTT 1237

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1863

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(v) ORIGINAL SOURCE

(A) ORGANISM: Homo sapiens sapiens

(C) INDIVIDUAL/ISOLATE:

(D) DEVELOPMENTAL STAGE: adult

(F) TISSUE TYPE: female breast

(G) CELL TYPE: normal breast tissue

(H) CELL LINE: not derived from a cell line

(I) ORGANELLE: no

(ix) FEATURE:

(A) NAME/KEY: BRCA1 protein

2 (B) LOCATION: 1 to 1863

(C) IDENTIFICATION METHOD: observation of mRNA and antisense inhibition of BRCA1 gene

(D) OTHER INFORMATION: BRCA1 protein has a negative regulatory effect on growth of human mammary cells.

(x) PUBLICATION INFORMATION:

(A) AUTHORS: Miki, Y., et. al.

(B) TITLE: A strong candidate gene for the breast and ovarian cancer susceptibility gene BRCA1.

(C) JOURNAL: Science

(D) VOLUME: 266

(E) PAGES: 66-71

(F) DATE: 1994

(K) RELEVANT RESIDUES IN SEQ ID NO: 49

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

```
Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
1               5                  10                  15
Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
            20                  25                  30
Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
        35                  40                  45
Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
    50                  55                  60
Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
65                  70                  75                  80
Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
                85                  90                  95
Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
            100                 105                 110
Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
        115                 120                 125
Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
    130                 135                 140
Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
145                 150                 155                 160
Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
                165                 170                 175
Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
            180                 185                 190
```

Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
195 200 205
Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
210 215 220
Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
225 230 235 240
Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg
245 250 255
His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu
260 265 270
Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser
275 280 285
Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe
290 295 300
Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg
305 310 315 320
Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr
325 330 335
Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu
340 345 350
Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu
355 360 365
Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu
370 375 380
Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp
385 390 395 400
Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu
405 410 415
Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu
420 425 430
Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Asp Arg Val His
435 440 445
Ser Lys Ser Val Glu Ser Asp Ile Glu Asp Lys Ile Phe Gly Lys Thr
450 455 460
Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn
465 470 475 480
Leu Ile Ile Gly Ala Phe Val Ser Glu Pro Gln Ile Ile Gln Glu Arg
485 490 495
Pro Leu Thr Asn Lys Leu Lys Aeg Lys Arg Arg Pro Thr Ser Gly Leu
500 505 510
His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr
515 520 525
Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln
530 535 540
Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp
545 550 555
Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys
560 565 570 575

Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser
580 585 590
Asn Glu Leu Glu Leu Asn Ile Met His Asn Ser Lys Ala Pro Lys Lys
595 600 605
Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Als Leu Glu
610 615 620
Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln
625 630 635
Ile Asp Ser Cys Ser Ser Ser Glu Glu Ile Lys Lys Lys Lys Tyr Asn
640 645 650 655
Gln Met Pro Val Arg His Ser Arg Asn Leu Gln Leu Met Glu Gly Lys
660 665 670
Glu Pro Ala Thr Gly Ala Lys Lys Ser Asn Lys Pro Asn Glu Gln Thr
675 680 685
Ser Lys Arg His Asp Ser Asp Thr Phe Pro Glu Leu Lys Leu Thr Asn
690 695 700
Ala Pro Gly Ser Phe Thr Lys Cys Ser Asn Thr Ser Glu Leu Lys Glu
705 710 715
Phe Val Asn Pro Ser Leu Pro Arg Glu Glu Lys Glu Glu Lys Leu Glu
720 725 730 735
Thr Val Lys Val Ser Asn Asn Ala Glu Asp Pro Lys Asp Leu Met Leu
740 745 750
Ser Gly Glu Arg Val Leu Gln Thr Glu Arg Ser Val Glu Ser Ser Ser
755 760 765
Ile Ser Leu Val Pro Gly Thr Asp Tyr Gly Thr Gln Glu Ser Ile Ser
770 775 780
Leu Leu Glu Val Ser Thr Leu Gly Lys Ala Lys Thr Glu Pro Asn Lys
785 790 795
Cys Val Ser Gln Cys Ala Ala Phe Glu Asn Pro Lys Gly Leu Ile His
800 805 810 815
Gly Cys Ser Lys Asp Asn Arg Asn Asp Thr Glu Gly Phe Lys Tyr Pro
820 825 830
Leu Gly His Glu Val Asn His Ser Arg Glu Thr Ser Ile Glu Met Glu
835 840 845
Glu Ser Glu Leu Asp Ala Gln Tyr Leu Gln Asn Thr Phe Lys Val Ser
850 855 860
Lys Arg Gln Ser Phe Ala Pro Phe Ser Asn Pro Gly Asn Ala Glu Glu
865 870 875
Glu Cys Ala Thr Phe Ser Ala His Ser Gly Ser Leu Lys Lys Gln Ser
880 885 890 895
Pro Lys Val Thr Phe Glu Cys Glu Gln Lys Glu Glu Asn Gln Gly Lys
900 905 910
Asn Glu Ser Asn Ile Lys Pro Val Gln Thr Val Asn Ile Thr Ala Gly
915 920 925
Phe Pro Val Val Gly Gln Lys Asp Lys Pro Val Asp Asn Ala Lys Cys
930 935 940
Ser Ile Lys Gly Gly Ser Arg Phe Cys Leu Ser Ser Gln Phe Arg Gly
945 950 955

Asn Glu Thr Gly Leu Ile Thr Pro Asn Lys His Gly Leu Leu Gln Asn
960 965 970 975
Pro Tyr Arg Ile Pro Pro Leu Phe Pro Ile Lys Ser Phe Val Lys Thr
980 985 990
Lys Cys Lys Lys Asn Leu Leu Glu Glu Asn Phe Glu Glu His Ser Met
995 1000 1005
Ser Pro Glu Arg Glu Met Gly Asn Glu Asn Ile Pro Ser Thr Val Ser
1010 1015 1020
Thr Ile Ser Arg Asn Asn Ile Arg Glu Asn Val Phe Lys Glu Ala Ser
1025 1030 1035
Ser Ser Asn Ile Asn Glu Val Gly Ser Ser Thr Asn Glu Val Gly Ser
1040 1045 1050 1055
Ser Ile Asn Glu Ile Gly Ser Ser Asp Glu Asn Ile Gln Ala Glu Leu
1060 1065 1070
Gly Arg Asn Arg Gly Pro Lys Leu Asn Ala Met Leu Arg Leu Gly Val
1075 1080 1085
Leu Gln Pro Glu Val Tyr Lys Gln Ser Leu Pro Gly Ser Asn Cys Lys
1090 1095 1100
His Pro Glu Ile Lys Lys Gln Glu Tyr Glu Glu Val Val Gln Thr Val
1105 1110 1115
Asn Thr Asp Phe Ser Pro Tyr Leu Ile Ser Asp Asn Leu Glu Gln Pro
1120 1125 1130 1135
Met Gly Ser Ser His Ala Ser Gln Val Cys Ser Glu Thr Pro Asp Asp
1140 1145 1150
Leu Leu Asp Asp Gly Glu Ile Lys Glu Asp Thr Ser Phe Ala Glu Asn
1155 1160 1165
Asp Ile Lys Glu Ser Ser Ala Val Phe Ser Lys Ser Val Gln Lys Gly
1170 1175 1180
Glu Leu Ser Arg Ser Pro Ser Pro Phe Thr His Thr His Leu Ala Gln
1185 1190 1195
Gly Tyr Arg Arg Gly Ala Lys Lys Leu Glu Ser Ser Glu Glu Asn Leu
1200 1205 1210 1215
Ser Ser Glu Asp Glu Glu Leu Pro Cys Phe Gln His Leu Leu Phe Gly
1220 1225 1230
Lys Val Asn Asn Ile Pro Ser Gln Ser Thr Arg His Ser Thr Val Ala
1235 1240 1245
Thr Glu Cys Leu Ser Lys Asn Thr Glu Glu Asn Leu Leu Ser Leu Lys
1250 1255 1260
Asn Ser Leu Asn Asp Cys Ser Asn Gln Val Ile Leu Ala Lys Als Ser
1265 1270 1275
Gln Glu His His Leu Ser Glu Glu Thr Lys Cys Ser Ala Ser Leu Phe
1280 1285 1290 1295
Ser Ser Gln Cys Ser Glu Leu Glu Asp Leu Thr Ala Asn Thr Asn Thr
1300 1305 1310
Gln Asp Pro Phe Leu Ile Gly Ser Ser Lys Gln Met Arg His Gln Ser
1315 1320 1325
Glu Ser Gln Gly Val Gly Leu Ser Asp Lys Glu Leu Val Ser Asp Asp
1330 1335 1340

Glu Glu Arg Gly Thr Gly Leu Glu Glu Asn Asn Gln Glu Glu Gln Ser
1345 1350 1355
Met Asp Ser Asn Leu Gly Glu Ala Ala Ser Gly Cys Glu Ser Glu Thr
1360 1365 1370 1375
Ser Val Ser Glu Asp Cys Ser Gly Leu Ser Ser Gln Ser Asp Ile Leu
1380 1385 1390
Thr Thr Gln Gln Arg Asp Thr Met Gln His Asn Leu Ile Lys Leu Gln
1395 1400 1405
Gln Glu Met Ala Glu Leu Glu Ala Val Leu Glu Gln His Gly Ser Gln
1410 1415 1420
Pro Ser Asn Ser Tyr Pro Ser Ile Ile Ser Asp Ser Ser Ala Leu Glu
1425 1430 1435
Asp Leu Arg Asn Pro Glu Gln Ser Thr Ser Glu Lys Val Leu Gln Thr
1440 1445 1450 1455
Ser Gln Lys Ser Ser Glu Tyr Pro Ile Ser Gln Asn Pro Glu Gly Xaa
1460 1465 1470
Ser Ala Asp Lys Phe Glu Val Ser Ala Asp Ser Ser Thr Ser Lys Asn
1475 1480 1485
Lys Glu Pro Gly Val Glu Arg Ser Ser Pro Ser Lys Cys Pro Ser Leu
1490 1495 1500
Asp Asp Arg Trp Tyr Met His Ser Cys Ser Gly Ser Leu Gln Asn Arg
1505 1510 1515 1520
Asn Tyr Pro Pro Gln Glu Glu Leu Ile Lys Val Val Asp Val Glu Glu
1525 1530 1535
Gln Gln Leu Glu Glu Ser Gly Pro His Asp Leu Thr Glu Thr Ser Tyr
1540 1545 1550
Leu Pro Arg Gln Asp Leu Glu Gly Thr Pro Tyr Leu Glu Ser Gly Ile
1555 1560 1565
Ser Leu Phe Ser Asp Asp Pro Glu Ser Asp Pro Ser Glu Asp Arg Ala
1570 1575 1580
Pro Glu Ser Ala Arg Val Gly Asn Ile Pro Ser Ser Thr Ser Ala Leu
1585 1590 1595 1600
Lys Val Pro Gln Leu Lys Val Ala Glu Ser Ala Gln Ser Pro Ala Ala
1605 1610 1615
Ala His Thr Thr Asp Thr Ala Gly Tyr Asn Ala Met Glu Glu Ser Val
1620 1625 1630
Ser Arg Glu Lys Pro Glu Leu Thr Ala Ser Thr Glu Arg Val Asn Lys
1635 1640 1645
Arg Met Ser Met Val Val Ser Gly Leu Thr Pro Glu Glu Phe Met Leu
1650 1655 1660
Val Tyr Lys Phe Ala Arg Lys His His Ile Thr Leu Thr Asn Leu Ile
1665 1670 1675 1680
Thr Glu Glu Thr Thr His Val Val Met Lys Thr Asp Ala Glu Phe Val
1685 1690 1695
Cys Glu Arg Thr Leu Lys Tyr Phe Leu Gly Ile Ala Gly Gly Lys Trp
1700 1705 1710
Val Val Ser Tyr Phe Trp Val Thr Gln Ser Ile Lys Glu Arg Lys Met
1715 1720 1725

Leu Asn Glu His Asp Phe Glu Val Arg Gly Asp Val Val Asn Gly Arg
1730 1735 1740
Asn His Gln Gly Pro Lys Arg Ala Arg Glu Ser Gln Asp Arg Lys Ile
1745 1750 1755 1760
Phe Arg Gly Leu Glu Ile Cys Cys Tyr Gly Pro Phe Thr Asn Met Pro
1765 1770 1775
Thr Asp Gln Leu Glu Trp Met Val Gln Leu Cys Gly Ala Ser Val Val
1780 1785 1790
Lys Glu Leu Ser Ser Phe Thr Leu Gly Thr Gly Val His Pro Ile Val
1795 1800 1805
Val Val Gln Pro Asp Ala Trp Tht Glu Asp Asn Gly Phe His Ala Ile
1810 1815 1820
Gly Gln Met Cys Glu Ala Pro Val Val Thr Arg Glu Trp Val Leu Asp
1825 1830 1835 1840
Ser Val Ala Leu Tyr Gln Cys Gln Glu Leu Asp Thr Tyr Leu Ile Pro
1845 1850 1855
Gln Ile Pro His Ser His Tyr
1860

CLAIMS

What I claim is:

1. A method for detecting differential expression of at least one marker gene in pre-invasive cancerous breast tissue, said method comprising the steps of:

(a) obtaining an abnormal breast tissue sample by a collection step wherein said abnormal breast tissue sample comprises substantially exclusively abnormal breast tissue which exhibits histological or cytological characteristics of pre-invasive breast cancer;

(b) isolating mRNA from said abnormal breast tissue sample;

(c) preparing at least one abnormal breast tissue cDNA library from said mRNA isolated from said abnormal breast tissue sample;

(d) obtaining a normal breast tissue sample from humans either with or without disease, said normal breast tissue sample comprising substantially exclusively normal breast tissue which does not exhibit histological or cytological characteristics of pre-invasive breast cancer;

(e) preparing at least one normal breast tissue cDNA library from said normal breast tissue sample; and

(f) comparing said abnormal breast tissue cDNA library with said normal tissue cDNA library to determine whether the expression of at least one marker gene in said abnormal breast tissue sample is different from the expression of said marker gene in said normal breast tissue sample.

2. The method according to Claim 1 wherein said collection step is microscopically-directed.

3. The method according to Claim 2 wherein the size of said abnormal tissue sample substantially conforms to an isolatable tissue structure such that only cells exhibiting abnormal cytological or histological characteristics are collected.

4. The method according to Claim 3 wherein said isolatable tissue structure comprises ductal epithelial cells in pre-invasive breast cancer tissue.

5. The method according to Claim 1 further comprising confirming said differential expression of said marker gene in said normal tissue sample and in said abnormal tissue sample by using a hybridization or PCR technique.

6. The method according to Claim 5 wherein said hybridization technique comprises RT-PCR.

7. The method according to Claim 5 wherein said hybridization technique comprises nuclease protection assays.

8. The method according to Claim 5 wherein said hybridization technique comprises in-situ hybridization of RNA in said abnormal tissue sample and in said normal tissue sample.

9. The method according to Claim 1 wherein said abnormal cDNA library and said normal cDNA library are compared by means of differential display.

10. The method according to Claim 1 wherein said abnormal cDNA library and said normal cDNA library are compared by means of differential screening.

11. The method according to claim 1, wherein said normal tissue comprises normal breast tissue cells.

12. The method according to claim 1, wherein said abnormal breast tissue cells are non-comedo ductal carcinoma in situ cells.

13. The method according to claim 1, wherein the primer used in the PCR amplification technique is selected from the group consisting of randomly selected primers having the sequences

5′-CGCGACGGCCGCGCGTCTGCCAGGG-3′, 5′-CTTGCGCGCATACGCACAAC-3′,
5′-AACCCTCACCCTAACCCCAA-3′, 5′-CGCCCCTGCGTTACCCTCCCCGCCG-3′,
5′-GGATGGCGTCCTGTAACCCGACGCT-3′, 5′-ACTGGGCTGTCCTGCGGTGGCGGGG-3′,
5′-CTGAGAGGTAGCCGCGCGGAGGCTG-3′, 5′-GCCTGGCCGCGACACGGATTACCGC-3′,
5′-TTAGCGCATGGTGGACCTGGAGACG-3′, 5′-TGTGGTTACGTCAGCGAAGGTAATA-3′,

15. A method of diagnosing the presence of pre-invasive breast cancer in human pathologic tissues, said method comprising the steps of:

(a) obtaining an abnormal breast tissue sample by a collection step wherein said abnormal breast tissue sample comprises substantially exclusively abnormal breast tissue which exhibits histological or cytological characteristics of pre-invasive breast cancer;

(b) isolating mRNA from said abnormal breast tissue sample;

(c) preparing at least one abnormal breast tissue cDNA library from said mRNA isolated from said abnormal breast tissue sample;

(d) obtaining a normal breast tissue sample from humans either with or without disease, said normal breast tissue sample comprising substantially exclusively normal breast tissue which does not exhibit histological or cytological characteristics of pre-invasive breast cancer;

(e) preparing at least one normal breast tissue cDNA library from said normal breast tissue sample; and

(f) comparing said abnormal breast tissue cDNA library with said normal tissue cDNA library to determine whether the expression of at least one marker gene in said abnormal breast tissue sample is different from the expression of said marker gene in said normal breast tissue sample.

(g) cloning said differentially expressed marker gene using sequence-based amplification to create a cloned marker gene;

(h) sequencing said cloned marker gene;

(i) producing proteins encoded by said cloned marker gene;

18. The method according to claim 15, wherein said medical diagnostic tests comprise blood tests.

19. The method according to claim 15, wherein said sequence-based amplification employs DNA sequences designed to clone marker genes encoding secreted proteins.

20. The method according to claim 15, wherein said sequence-based amplification employs DNA sequences designed to clone marker genes encoding transcription factors.

21. The method according to claim 15, wherein said sequence-based amplification employs DNA sequences designed to clone marker genes encoding RibRed.

22. The method according to claim 15, wherein said cloned marker genes encoding secreted proteins are employed in the diagnosis of specific diseases by using a blood test.

23. The method according to claim 15, wherein said sequence-based amplification employs DNA sequences adapted to clone marker genes which encode cell surface proteins.

24. The method according to claim 15, wherein said proteins encoded by said cloned marker comprise cell surface proteins and wherein the presence of said proteins as a diagnostic indicator is detected by using a diagnostic imaging test.

25. A diagnostic method to determine the presence of pre-invasive breast cancer using detection of a differentially expressed marker gene, according to claim 15, wherein said diagnostic method comprises:

a) obtaining a substantially purified marker gene which is expressed to a greater degree in cells collected by a microscopically-directed cloning method from abnormal tissue than in cells collected from normal tissue;

b) probing tissues using a hybridization technique to determine whether said substantially purified marker gene is differentially expressed; and,

c) probing nucleic acids of tissues using a standard hybidization technique to determine the presence of said substantiallly purified marker gene in a tissue, the

presence of the marker gene indicating the presence of non-comedo DCIS which is pre-invasive breast cancer.

26. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed
according to SEQ ID NO:1, which comprises

```
TTGGGAATTG GGTACGCGGG CCCCCCACTG TGCCGAATTC CTGCATGCGG GGGATCCACT  60
AGTTCAGAGC AGGCCGCCAC CCGTAGGACT CCAGCTTTTG TTCGTTCCCT TTAGTGAGGG 120
TTAATTTTCG AGCTTGGCGT AATCATGGTC ATAGCTGTTT CCTGTGTGAA ATTGTTATCC 180
GCTCACAATT CCACACAACA TACGAGCCGG AAGCATAAAA GTGTAAAGCC TGGGGTGCCT 240
AATGAGTGAG CTAACTCACA TTAA                                        264
```

27. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed according to SEQ ID NO:2, which comprises

```
TAGCCCGGTT ATCGAAATAG CCACAGCGCC TCTTCACTAT CAGCAGTACG CCGCCCAGTT  60
GTACGGACAC GGA                                                     73
```

28. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed according to SEQ ID NO:3, which comprises

```
TGCCCGATGT GTGTCGTACA ACTGGCGCTG TGGCTGATTT CGATAA                 46
```

29. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed according to SEQ ID NO:4, which comprises

```
TAGCCCATGA GTTCGTGTCC GTACAACTGG GGCGCTGTGG CTGATTTCGA TANNNNNAGC  60
ATCAGCCCGA CG                                                      72
```

30. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed according to SEQ ID NO:5, which comprises

```
TAGCCCGGTT ATCGAAATCA GCCACAGCGC CTAACTTCTG CAGAAGCCTT TGACCATCAC  60
CAGTTGTACG GACACGAACT CATC                                          84
```

31. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed according to SEQ ID NO:6, which comprises

```
GTGGTTTCCG AAATTCCTGG GAAGGGGGGT GCTGGCGTGT GGAATTGTCG CGGCCCCTGG  60
TCTGCCGCGG CGTTTTTTGT CTACATTCGT CGTAGCTCG                          99
```

32. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed according to SEQ ID NO:7, which comprises

```
ATCAGCGCGC GACATTCGGG TACCCGCGCC CCCCCCTCCG TCGGAATTCC TCGAGCCGGG  60
ATCCATAGGA TGTGGAGTTA GTTTTGTT                                     88
```

33. A method for detecting differential expression of at least one marker gene in pre-invasive cancerous breast tissue, said method comprising the steps of:

(a) obtaining an abnormal tissue sample by a collection step wherein said abnormal tissue sample comprises substantially exclusively abnormal tissue which exhibits histological or cytological characteristics of pre-invasive cancer;

(b) isolating mRNA from said abnormal tissue sample;

(c) preparing at least one abnormal tissue cDNA library from said mRNA isolated from said abnormal tissue sample;

(d) obtaining a normal tissue sample from humans either with or without disease, said normal tissue sample comprising substantially exclusively normal tissue which does not exhibit histological or cytological characteristics of pre-invasive cancer;

(e) preparing at least one normal tissue cDNA library from said normal tissue sample; and

(f) comparing said abnormal tissue cDNA library with said normal tissue cDNA library to determine whether the expression of at least one marker gene in said abnormal tissue sample is different from the expression of said marker gene in said normal tissue sample.

34. The method according to Claim 33 wherein said collection step is microscopically-directed.

a) obtaining a substantially purified marker gene which is expressed to a greater degree in cells collected by a microscopically-directed cloning method from abnormal tissue than in cells collected from normal tissue;

b) probing tissues using a hybridization technique to determine whether the marker gene is differentially expressed; and,

c) probing nucleic acids of tissues using a standard PCR technique to determine the presence of the marker gene in a tissue, the presence of the marker gene indicating the presence of pre-invasive cancer.

35. Substantially purified DNA having the nucleotide sequences selected from the group of seqences consisting of: SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7.

36. An expression vector for the differentially expressed polypeptides encoded by said substantially purified DNA comprising one of the group of DNA sequences of claim 28 operatively linked to at least one control sequence compatible with a suitable bacterial host cell.

37. The vector of claim 36 wherein the DNA encoding the differentially expressed polypeptides encoded by said substantially purified DNA comprising one of the group of DNA sequences of claim 28 is linked to at least one sequence from bacteriophage.

38. Substantially purified polypeptides encoded by substantially purified DNA comprising one of the group of DNA sequences of claim 35 free of proteins other than proteins encoded by said substantially purified DNA.

39. An antibody specifically binding one of the group of polypeptides encoded by one of the nucleotide sequences selected from the group of seqences consisting of: SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID ID NO:7.

40. A method of detecting and/or determining said antibody in a test sample, comprising the steps:

(a) providing a test sample suspected of containing said marker protein;

(b) adding a quantity of said marker protein of claim 38 to the antibody of claim 39; and

(c) determining a level of said marker protein in said test sample.

41. A method of screening compounds for activity in the treatment of breast cancer, comprising the steps of:

(a) ligating a DNA sequence that regulates expression of the BRCA1 gene into a vector, the vector having a reporter gene, so that the DNA sequence is located such that the DNA sequence regulates expression of the reporter gene;

(b) introducing the ligated DNA sequence/reporter gene into a breast cancer cell;

(c) administering a compound to the breast cancer cell; and

(d) detecting levels of a protein produced by the reporter cell.

42. The method according to claim 41 wherein the DNA sequence is as essentially set forth in SEQ ID NO:48.

43. The method according to claim 42 wherein the DNA sequence is selected from among:

a. a DNA sequence which hybridizes to SEQ ID NO:48 or fragments thereof; and

b. DNA sequences which but for the degeneracy of the genetic code would hybridize to the DNA sequences defined in (a) and (b).

44. The method according to claim 41 wherein the ligated DNA sequence/reporter gene is introduced into the breast cancer cell by cloning the ligated DNA sequence/reporter gene into an expression vector and transfecting the breast cancer cells with the expression vector.

45. The method according to claim 44 wherein the DNA sequence is essentially set forth in SEQ ID NO:48 or its complementary strands.

46. A method of producing an indicator compound, comprising the steps of:

(a) ligating a DNA sequence that regulates expression of the BRCA1 gene into a vector, the vector having a reporter gene, so that the DNA sequence is located such that the DNA sequence regulates expression of the reporter gene;

(b) introducing the ligated DNA sequence/reporter gene into a breast cancer cell;

(c) administering a biological agent to the breast cancer cell; and

(d) producing a protein encoded by the reporter gene; and

(e) reacting the protein encoded by the reporter gene with a compound in the reaction media to produce the indicator compound.

47. The method according to claim 46 wherein the ligated DNA sequence/reporter gene is introduced into the breast cancer cell by cloning the ligated

DNA sequence/reporter gene into an expression vector and transfecting the breast cancer cells with the expression vector.

48. The method according to claim 46 wherein the DNA sequence is as essentially set forth in SEQ ID NO:48 or its complementary strands.

49. The method according to claim 46 wherein the DNA sequence is selected from among:

a. a DNA sequence which hybridizes to SEQ ID NO:48 or fragments thereof; and

b. DNA sequences which but for the degeneracy of the genetic code would hybridize to the DNA sequences defined in (a) and (b).

50. A method of treating breast cancer in a patient comprising the steps of ligating a gene that encodes a protein having an amino acid sequence as essentiallly set forth in SEQ ID NO:49 with a promoter capable of inducing expression of the gene in a breast cancer cell and introducing the ligated gene into a breast cancer cell.

51. The method of treating breast cancer described in claim 50 wherein the gene has a DNA sequence selected from among:

a. the DNA sequence as essentially set forth in SEQ ID NO:47 or its complementary strands;

b. a DNA sequence which hybridizes to SEQ ID NO:47 or fragments thereof; and

c. DNA sequences which but for the degeneracy of the genetic code would hybridize to the DNA sequences defined in (a) and (b).

52. The method of treating breast cancer described in claim 50 wherein the gene has a DNA sequence having 20-99% homology with SEQ ID NO:47.

53. The method according to claim 50 wherein the ligated gene is introduced into the cell in a viral expression vector.

54. The method according to claim 50 wherein the breast cancer is gene-linked hereditary breast cancer.

55. The method described in claim 50 wherein the breast cancer is sporadic breast cancer.

AMENDED CLAIMS

[received by the International Bureau on 14 June 1995 (14.06.95); original claims 13 and 15 amended; new claims 14,16 and 17 added; remaining claims unchanged (8 pages)]

6. The method according to Claim 5 wherein said hybridization technique comprises RT-PCR.

7. The method according to Claim 5 wherein said hybridization technique comprises nuclease protection assays.

8. The method according to Claim 5 wherein said hybridization technique comprises in-situ hybridization of RNA in said abnormal tissue sample and in said normal tissue sample.

9. The method according to Claim 1 wherein said abnormal cDNA library and said normal cDNA library are compared by means of differential display.

10. The method according to Claim 1 wherein said abnormal cDNA library and said normal cDNA library are compared by means of differential screening.

11. The method according to claim 1, wherein said normal tissue comprises normal breast tissue cells.

12. The method according to claim 1, wherein said abnormal breast tissue cells are non-comedo ductal carcinoma in situ cells.

13. The method according to claim 1, wherein the primer used in the PCR amplification technique is selected from the group consisting of randomly selected primers having the sequences

5'-CGCGACGGCCGCGCGTCTGCCAGGG-3', 5'-CTTGCGCGCATACGCACAAC-3',
5'-AACCCTCACCCTAACCCCAA-3', 5'-CGCCCCTGCGTTACCCTCCCCGCCG-3',
5'-GGATGGCGTCCTGTAACCCGACGCT-3', 5'-ACTGGGCTGTCCTGCGGTGGCGGGG-3',
5'-CTGAGAGGTAGCCGCGCGGAGGCTG-3', 5'-GCCTGGCCGCGACACGGATTACCGC-3',
5'-TTAGCGCATGGTGGACCTGGAGACG-3', 5'-TGTGGTTACGTCAGCGAAGGTAATA-3',
5'-AGTCGCACGCATGTCACGCTCCGCC-3', 5'-TATCCAAGCGGCAGGCTACGAGGCC-3',
5'-GGCGCGCCCGACGGTCTGGTATCTA-3', 5'-CTCCCTCCCCGGACTCGGGGTTAGT-3',
5'-ATGCGGGCGGCTCGGGCCTGGTCGC-3', 5'-CGTGAAGCCTATGCCCTCCCTCAAC-3',
5'-GTGCCGTCGTAGCCCTTCAGCGATC-3', 5'-GCGACACTAGGCTCCCGGAGGAGGG-3',
5'-TGGGCCAGGCCTCCGGGCCCGGTAT-3', 5'-CCGGAACTGCGATAGCGTCCGTCCC-3',
5'-AGCGGACACCTGTTTCCCGAGAGCC-3', 5'-AACGGGTGGACATCCGCCTGCCGCC-3',
5'-TGAACCACGATGTCAATCGTCCCGA-3', 5'-TCATCCCCGCCGAAAGACGCTCGCC-3',
5'-ATAGGCTGCGGCACGCGCTGGGACT-3', 5'-GACCAGGTGCGCACGAGCATGTACA-3',
5'-AGCGTAGTCATCGGCCTTCGCGCCC-3', 5'-GGCCCCTAGCCCAGGGTGAAGCCCA-3',
5'-CCCAGTGCTACGGGCCGCCCCAAGC-3', 5'-CCTTCCTGGGTTACCTGCCCTCGGG-3',
5'-TCCGGACAGCAGCCACGCCAAGGGC-3', 5'-ACGCGCTGGTCCACCGAGGCCTGAT-3',
5'-CGATGCAAGGCCAGCAGCACTCGAC-3', 5'-CCCCCGGAGCGGACCACCGGACGTG-3',
5'-AGCGGGGAGGGATCGGGGGCCAAGC-3', 5'-GCCTGGTGTAGGCAGGCAGCTCTTA-3',
5'-CCACCCCTGTAGTGCGGGCTGCGAG-3', 5'-GGAACCCGACGCCCGTCCAGGGTTC-3',

5'-TCGGGCAGCAAGGCCGGGACGCTCC-3', 5'-GACGGGGGACGGGCTAGGTGGCTTA-3',
and 5'-CTTGTTGCCGGCGGAGAGGGCTGCC-3'.

14. The method according to claim 2, wherein said abnormal tissue sample is approximately 2 mm in diameter.

15. A method of diagnosing the presence of pre-invasive breast cancer in human pathologic tissues, said method comprising the steps of:

(a) obtaining an abnormal breast tissue sample by a collection step wherein said abnormal breast tissue sample comprises substantially exclusively abnormal breast tissue which exhibits histological or cytological characteristics of pre-invasive breast cancer;

(b) isolating mRNA from said abnormal breast tissue sample;

(c) preparing at least one abnormal breast tissue cDNA library from said mRNA isolated from said abnormal breast tissue sample;

(d) obtaining a normal breast tissue sample from humans either with or without disease, said normal breast tissue sample comprising substantially exclusively normal breast tissue which does not exhibit histological or cytological characteristics of pre-invasive breast cancer;

(e) preparing at least one normal breast tissue cDNA library from said normal breast tissue sample; and

(f) comparing said abnormal breast tissue cDNA library with said normal tissue cDNA library to determine whether the expression of at least one marker gene in said abnormal breast tissue sample is different from the expression of said marker gene in said normal breast tissue sample.

(g) cloning said differentially expressed marker gene using sequence-based amplification to create a cloned marker gene;

(h) sequencing said cloned marker gene;

(i) producing proteins encoded by said cloned marker gene;

(j) generating antibodies which will recognize said proteins encoded by said cloned marker gene by antigen recognition; and

(k) detecting said recognized antigen by means of medical diagnostic tests.

16. The method according to claim 15, wherein said medical diagnostic tests comprise diagnostic tissue tests.

17. The method according to claim 15, wherein said medical diagnostic tests comprise X-ray tests.

18. The method according to claim 15, wherein said medical diagnostic tests comprise blood tests.

19. The method according to claim 15, wherein said sequence-based amplification employs DNA sequences designed to clone marker genes encoding secreted proteins.

20. The method according to claim 15, wherein said sequence-based amplification employs DNA sequences designed to clone marker genes encoding transcription factors.

21. The method according to claim 15, wherein said sequence-based amplification employs DNA sequences designed to clone marker genes encoding RibRed.

22. The method according to claim 15, wherein said cloned marker genes encoding secreted proteins are employed in the diagnosis of specific diseases by using a blood test.

23. The method according to claim 15, wherein said sequence-based amplification employs DNA sequences adapted to clone marker genes which encode cell surface proteins.

24. The method according to claim 15, wherein said proteins encoded by said cloned marker comprise cell surface proteins and wherein the presence of said proteins as a diagnostic indicator is detected by using a diagnostic imaging test.

25. A diagnostic method to determine the presence of pre-invasive breast cancer using detection of a differentially expressed marker gene, according to claim 15, wherein said diagnostic method comprises:

a) obtaining a substantially purified marker gene which is expressed to a greater degree in cells collected by a microscopically-directed cloning method from abnormal tissue than in cells collected from normal tissue;

b) probing tissues using a hybridization technique to determine whether said substantially purified marker gene is differentially expressed; and,

c) probing nucleic acids of tissues using a standard hybidization technique

to determine the presence of said substantiallly purified marker gene in a tissue, the presence of the marker gene indicating the presence of non-comedo DCIS which is pre-invasive breast cancer.

26. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed according to SEQ ID NO:1, which comprises

```
TTGGGAATTG GGTACGCGGG CCCCCCACTG TGCCGAATTC CTGCATGCGG GGGATCCACT  60
AGTTCAGAGC AGGCCGCCAC CCGTAGGACT CCAGCTTTTG TTCGTTCCCT TTAGTGAGGG  120
TTAATTTTCG AGCTTGGCGT AATCATGGTC ATAGCTGTTT CCTGTGTGAA ATTGTTATCC  180
GCTCACAATT CCACACAACA TACGAGCCGG AAGCATAAAA GTGTAAAGCC TGGGGTGCCT  240
AATGAGTGAG CTAACTCACA TTAA                                          264
```

27. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed according to SEQ ID NO:2, which comprises

```
TAGCCCGGTT ATCGAAATAG CCACAGCGCC TCTTCACTAT CAGCAGTACG CCGCCCAGTT  60
GTACGGACAC GGA                                                     73
```

28. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed according to SEQ ID NO:3, which comprises

```
TGCCCGATGT GTGTCGTACA ACTGGCGCTG TGGCTGATTT CGATAA                  46
```

29. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed according to SEQ ID NO:4, which comprises

```
TAGCCCATGA GTTCGTGTCC GTACAACTGG GGCGCTGTGG CTGATTTCGA TANNNNNAGC  60
ATCAGCCCGA CG                                                      72
```

30. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed according to SEQ ID NO:5, which comprises

```
TAGCCCGGTT ATCGAAATCA GCCACAGCGC CTAACTTCTG CAGAAGCCTT TGACCATCAC  60
CAGTTGTACG GACACGAACT CATC                                          84
```

31. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed according to SEQ ID NO:6, which comprises

```
GTGGTTTCCG AAATTCCTGG GAAGGGGGGT GCTGGCGTGT GGAATTGTCG CGGCCCCTGG  60
TCTGCCGCGG CGTTTTTTGT CTACATTCGT CGTAGCTCG                          99
```

32. The method according to claim 25, wherein said substantially purified marker gene has the sequence listed according to SEQ ID NO:7, which comprises

```
ATCAGCGCGC GACATTCGGG TACCCGCGCC CCCCCCTCCG TCGGAATTCC TCGAGCCGGG 60
ATCCATAGGA TGTGGAGTTA GTTTTGTT                                    88
```

33. A method for detecting differential expression of at least one marker gene in pre-invasive cancerous breast tissue, said method comprising the steps of:

(a) obtaining an abnormal tissue sample by a collection step wherein said abnormal tissue sample comprises substantially exclusively abnormal tissue which exhibits histological or cytological characteristics of pre-invasive cancer;

(b) isolating mRNA from said abnormal tissue sample;

(c) preparing at least one abnormal tissue cDNA library from said mRNA isolated from said abnormal tissue sample;

(d) obtaining a normal tissue sample from humans either with or without disease, said normal tissue sample comprising substantially exclusively normal tissue which does not exhibit histological or cytological characteristics of pre-invasive cancer;

(e) preparing at least one normal tissue cDNA library from said normal tissue sample; and

(f) comparing said abnormal tissue cDNA library with said normal tissue cDNA library to determine whether the expression of at least one marker gene in said abnormal tissue sample is different from the expression of said marker gene in said normal tissue sample.

34. The method according to Claim 33 wherein said collection step is microscopically-directed.

a) obtaining a substantially purified marker gene which is expressed to a greater degree in cells collected by a microscopically-directed cloning method from abnormal tissue than in cells collected from normal tissue;

b) probing tissues using a hybridization technique to determine whether the marker gene is differentially expressed; and,

c) probing nucleic acids of tissues using a standard PCR technique to determine the presence of the marker gene in a tissue, the presence of the marker gene indicating the presence of pre-invasive cancer.

35. Substantially purified DNA having the nucleotide sequences selected from the group of seqences consisting of: SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7.

36. An expression vector for the differentially expressed polypeptides encoded by said substantially purified DNA comprising one of the group of DNA sequences of claim 28 operatively linked to at least one control sequence compatible with a suitable bacterial host cell.

37. The vector of claim 36 wherein the DNA encoding the differentially expressed polypeptides encoded by said substantially purified DNA comprising one of the group of DNA sequences of claim 28 is linked to at least one sequence from bacteriophage.

38. Substantially purified polypeptides encoded by substantially purified DNA comprising one of the group of DNA sequences of claim 35 free of proteins other than proteins encoded by said substantially purified DNA.

39. An antibody specifically binding one of the group of polypeptides encoded by one of the nucleotide sequences selected from the group of seqences consisting of: SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID ID NO:7.

40. A method of detecting and/or determining said antibody in a test sample, comprising the steps:

(a) providing a test sample suspected of containing said marker protein;

(b) adding a quantity of said marker protein of claim 38 to the antibody of claim 39; and

(c) determining a level of said marker protein in said test sample.

41. A method of screening compounds for activity in the treatment of breast cancer, comprising the steps of:

(a) ligating a DNA sequence that regulates expression of the BRCA1 gene into a vector, the vector having a reporter gene, so that the DNA sequence is located such that the DNA sequence regulates expression of the reporter gene;

(b) introducing the ligated DNA sequence/reporter gene into a breast cancer

cell;

(c) administering a compound to the breast cancer cell; and

(d) detecting levels of a protein produced by the reporter cell.

42. The method according to claim 41 wherein the DNA sequence is as essentially set forth in SEQ ID NO:48.

43. The method according to claim 42 wherein the DNA sequence is selected from among:

a. a DNA sequence which hybridizes to SEQ ID NO:48 or fragments thereof; and

b. DNA sequences which but for the degeneracy of the genetic code would hybridize to the DNA sequences defined in (a) and (b).

44. The method according to claim 41 wherein the ligated DNA sequence/reporter gene is introduced into the breast cancer cell by cloning the ligated DNA sequence/reporter gene into an expression vector and transfecting the breast cancer cells with the expression vector.

45. The method according to claim 44 wherein the DNA sequence is essentially set forth in SEQ ID NO:48 or its complementary strands.

46. A method of producing an indicator compound, comprising the steps of:

(a) ligating a DNA sequence that regulates expression of the BRCA1 gene into a vector, the vector having a reporter gene, so that the DNA sequence is located such that the DNA sequence regulates expression of the reporter gene;

(b) introducing the ligated DNA sequence/reporter gene into a breast cancer cell;

(c) administering a biological agent to the breast cancer cell; and

(d) producing a protein encoded by the reporter gene; and

(e) reacting the protein encoded by the reporter gene with a compound in the reaction media to produce the indicator compound.

47. The method according to claim 46 wherein the ligated DNA sequence/reporter gene is introduced into the breast cancer cell by cloning the ligated DNA sequence/reporter gene into an expression vector and transfecting the breast

cancer cells with the expression vector.

48. The method according to claim 46 wherein the DNA sequence is as essentially set forth in SEQ ID NO:48 or its complementary strands.

49. The method according to claim 46 wherein the DNA sequence is selected from among:

a. a DNA sequence which hybridizes to SEQ ID NO:48 or fragments thereof; and

b. DNA sequences which but for the degeneracy of the genetic code would hybridize to the DNA sequences defined in (a) and (b).

50. A method of treating breast cancer in a patient comprising the steps of ligating a gene that encodes a protein having an amino acid sequence as essentiallly set forth in SEQ ID NO:49 with a promoter capable of inducing expression of the gene in a breast cancer cell and introducing the ligated gene into a breast cancer cell.

51. The method of treating breast cancer described in claim 50 wherein the gene has a DNA sequence selected from among:

a. the DNA sequence as essentially set forth in SEQ ID NO:47 or its complementary strands;

b. a DNA sequence which hybridizes to SEQ ID NO:47 or fragments thereof; and

c. DNA sequences which but for the degeneracy of the genetic code would hybridize to the DNA sequences defined in (a) and (b).

52. The method of treating breast cancer described in claim 50 wherein the gene has a DNA sequence having 20-99% homology with SEQ ID NO:47.

53. The method according to claim 50 wherein the ligated gene is introduced into the cell in a viral expression vector.

54. The method according to claim 50 wherein the breast cancer is gene-linked hereditary breast cancer.

55. The method described in claim 50 wherein the breast cancer is sporadic breast cancer.

STATEMENT UNDER ARTICLE 19

Pursuant to Article 19 of the Patent Cooperation Treaty and Rule 46, Applicant respectfully submits the attached sheets of amended claims. The sheets are replacement sheets for pages 98-105 of the above referenced International application. These sheets contain Claims 6-55 of the above referenced international application. New Claims 14, 16 and 17 have been added to replacement pages 99-100. Additional primers have been listed in Claim 13 on replacement pages 98-99. These primers are described in the Sequence Listing. Claim 15 has been amended to include steps (j) and (k) on replacement page 99. The new claims and the amended claims do not go beyond the scope of the application as filed. The remaining replacement sheets include no amendments, but are filed to maintain the correct numbering of the claim pages.

Figure 1:

TABLE I: Anatomic Lesion Types in the Human Breast with Pre-malignant Implication

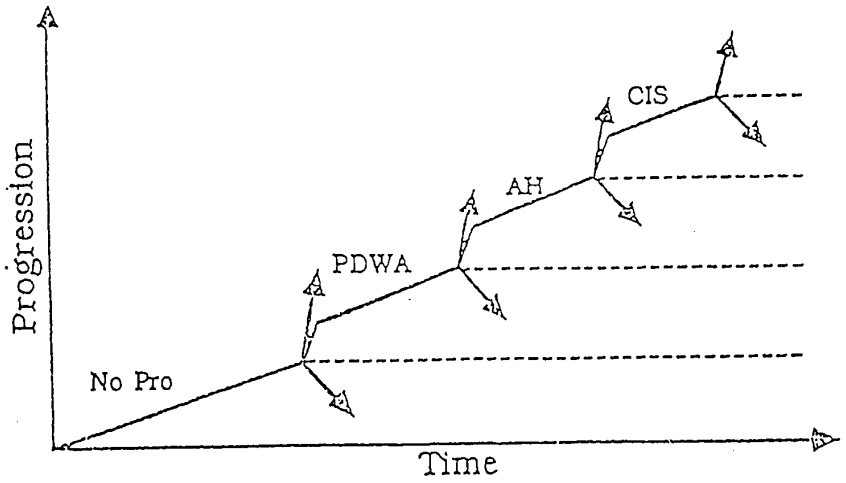
Pre-malignant Lesions	Relative Risk*	P value	Reference
Indicators of generalized increased risk			
Atypical ductal hyperplasia	4-5 fold	< .00001	(Dupont, et al, 1985 and 1993.)
Lobular CIS	9-10 fold	< .00001	(Page, et al, 1991.)
Determinant Lesions with Regional Risk			
Non-comedo DCIS	10-11 fold	< .00005	(Page, et al, 1982.)

* represents the 95% confidence interval for relative risk.

Figure 1: Table I describes anatomic lesion types in the human breast with pre-malignant implication.

GeneDX 1009, pg. 953

Fig. 2



GeneDX 1009, pg. 954

Fig. 3



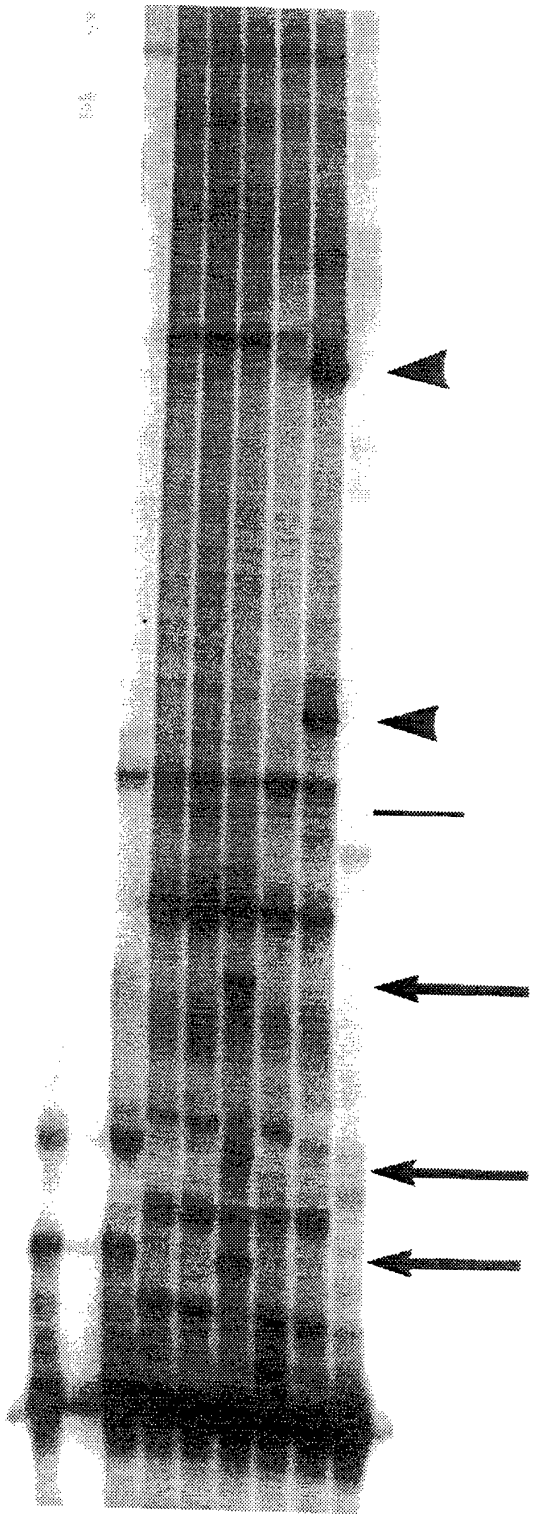
GeneDX 1009, pg. 955

Fig. 3



GeneDX 1009, pg. 956

Fig. 5



GeneDX 1009, pg. 957

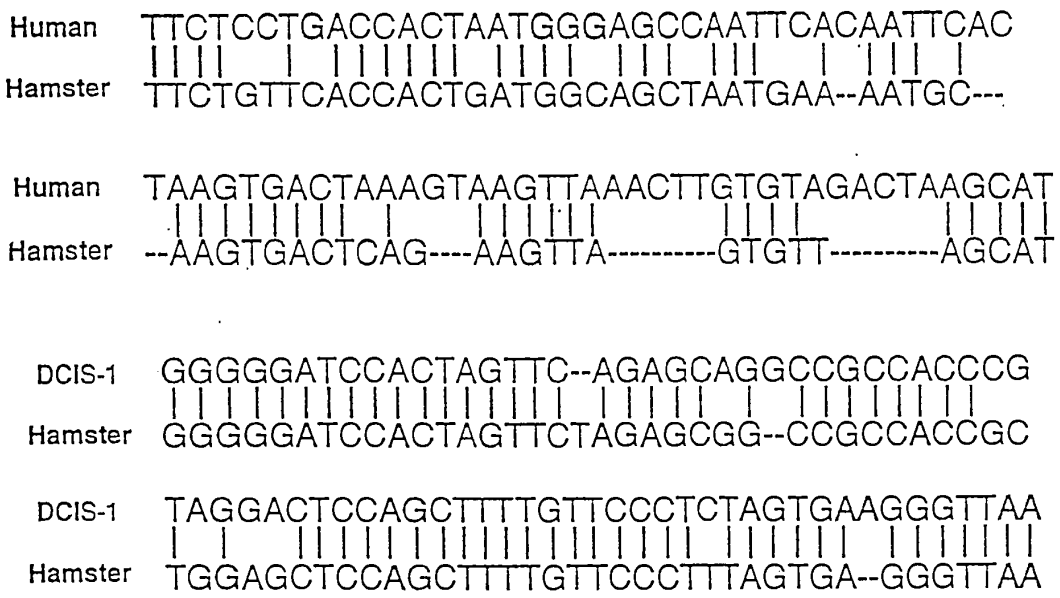
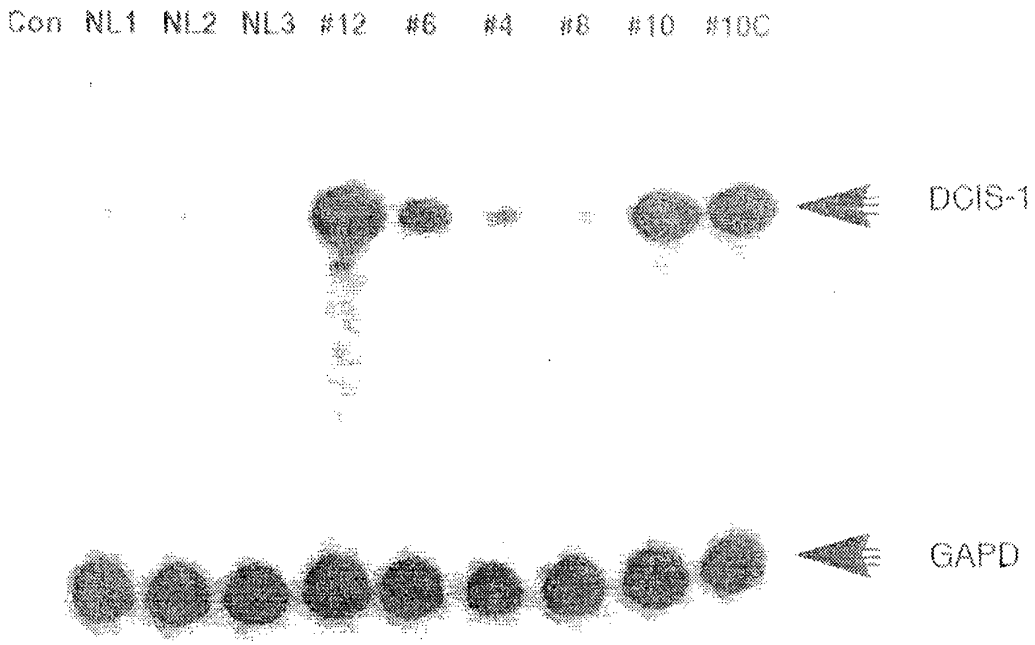


Figure 6: Comparison of the sequence between DCIS-1 and the human and hamster genes.

Fig. 7



GeneDX 1009, pg. 959

Fig. 8 - Table of the Genetic Code

Amino Acids			Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic acid	Asp	D	GAC	GAU				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				

Figure 8: Table of the Genetic Code.

SEQ ID NO: 1: (DCIS-1)

```
TTGGGAATTG GGTACGCGGG CCCCCCACTG TGCCGAATTC CTGCATGCGG GGGATCCACT  60
AGTTCAGAGC CCGTAGGACT CCAGCTTTTG TTCGTTCCCT TTAGTGAGGG TTAATTTTCG 120
AGCTTGGCGT AATCATGGTC ATCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAA 180
CATACGAGCC GGAAGCATAA AAGTGTAAGC AATGAGTGAG CTAACTCACA TTAA        234

SEQ ID NO: 2: (DCIS-2)
TAGCCCGGTT ATCGAAATAG CCACAGCGCC TCTTCACTAT CAGCAGTACG CCGCCCAGTT  60
GTA CGG ACA CGG                                                    72

SEQ ID NO: 3: (DCIS-3)
TGCCCGATGA GTTGTGTCGT ACAACTGGCG CTGTGGCTGA TTTCGATAA              49

SEQ ID NO: 4: (DCIS-4)
TAGCCCATGA GTTCGTGTCC GTACAACTGG GGCGCTGTGG CTGATTTCGA TANNNNNAGC  60
ATCAGCCCGA CG                                                      72

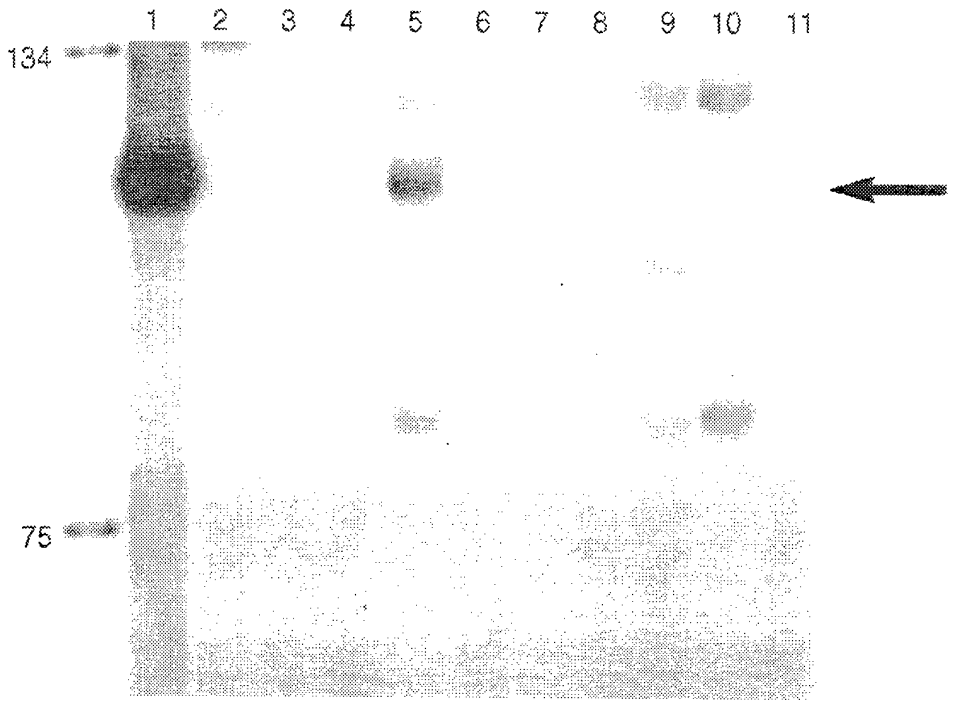
SEQ ID NO: 5: (DCIS-5)
TAGCCCGGTT ATCGAAATCA GCCACAGCGC CTAACTTCTG CAGAAGCCTT TGACCATCAC  60
CAGTTGTACG GAAACGAACT CATC                                         84

SEQ ID NO: 6: (DCIS-6)
GTGGTTTCCG AAATTCCTG GGAAGGGGGG TGCTGGCGTG TGGAATTGTC GCGGCCCCTG   60
GTCTGCCGCG GCGTTTTTT GTCTACATTC GTCGTAGCTC G                       101

SEQ ID NO: 7: (DCIS-7)
ATCAGCGCGC GACATTCGGG TACCCGCGCC C*****TCCG TCGGAATTCC TCGAGCCGGG  60
AT**ATAGGA TGTGGAGTTA GTTTTGTT                                     88
```

Figure 9: Table of Differentially Expressed Marker Genes From Pre-Invasive Human Breast Tissue

Fig. 10A



GeneDX 1009, pg. 962

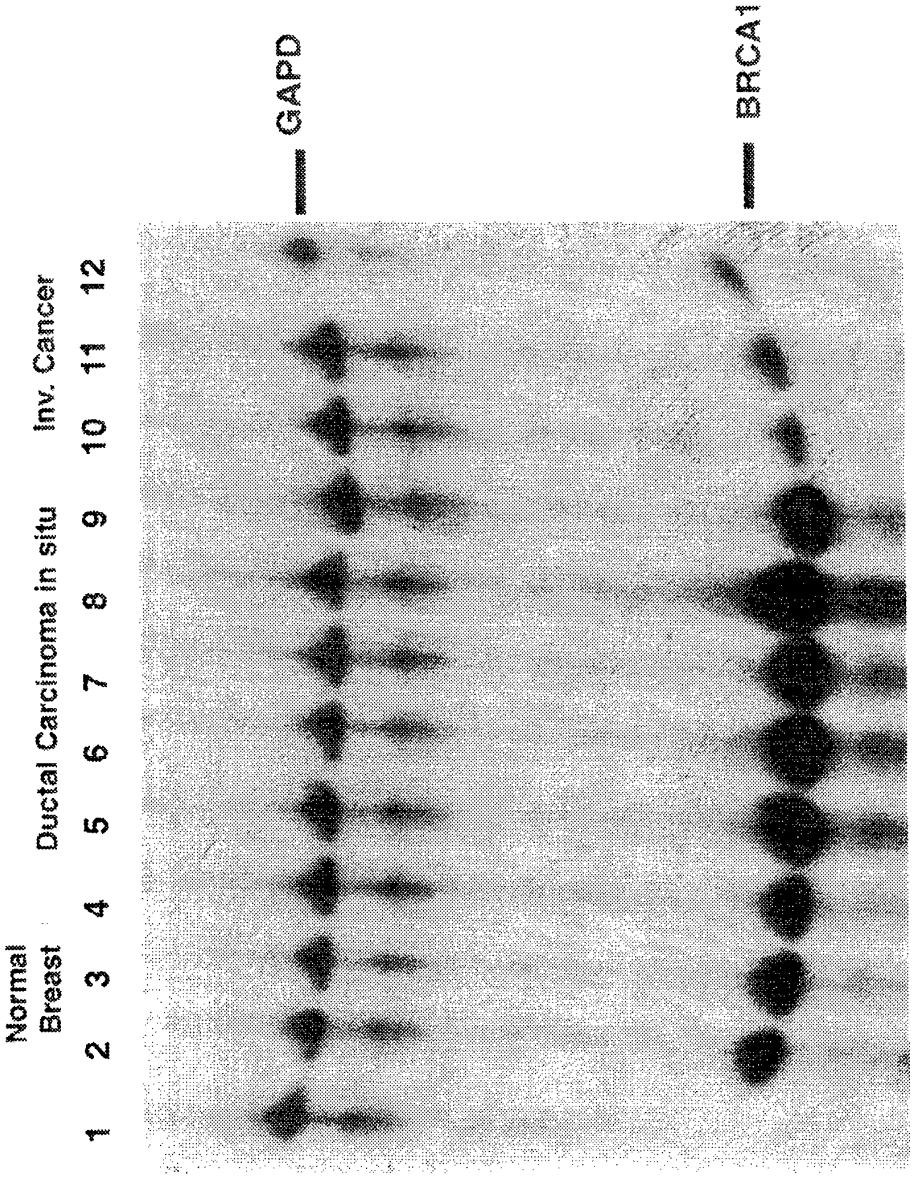
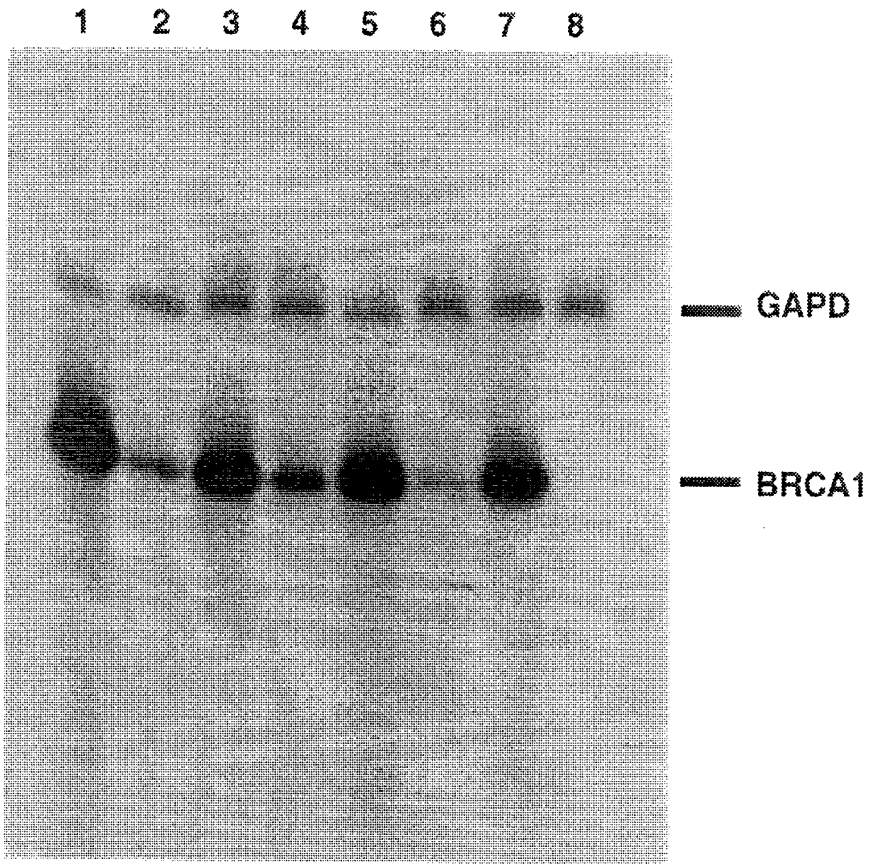


Fig. 10B

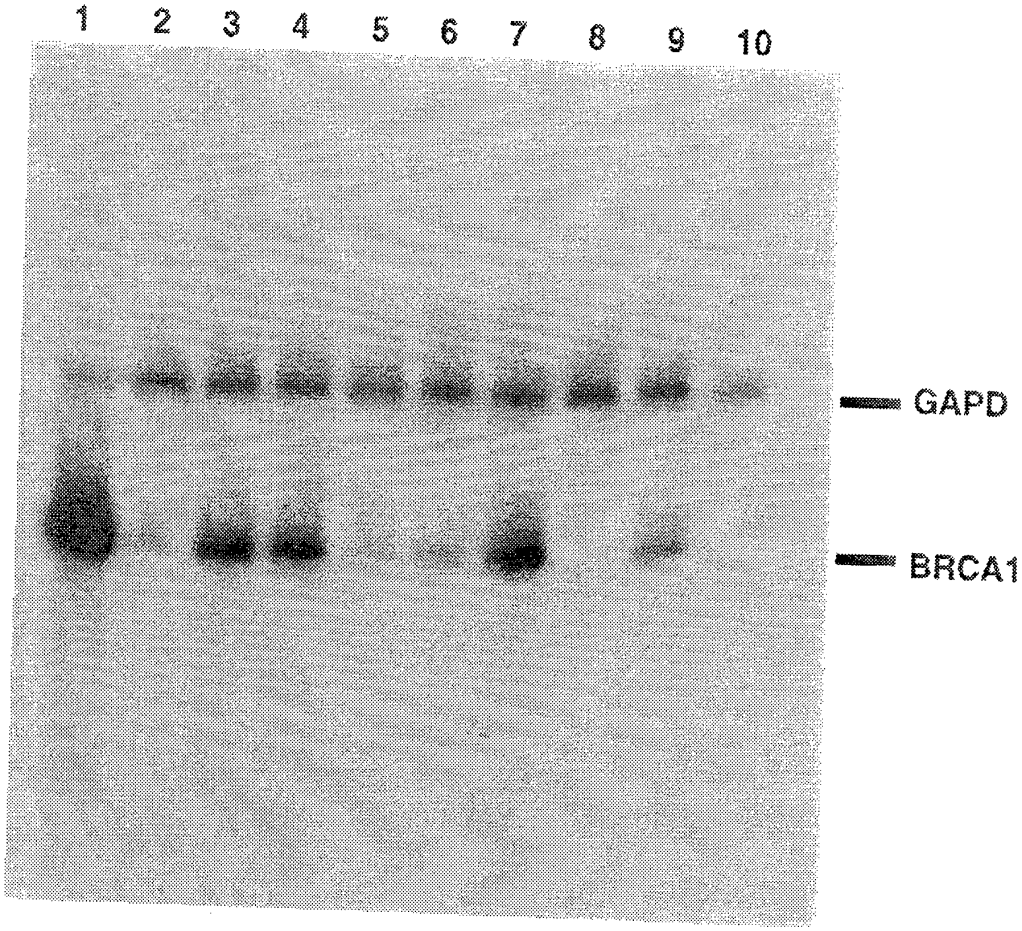
GeneDX 1009, pg. 963

Fig. 11A



GeneDX 1009, pg. 964

Fig. 11B



GeneDX 1009, pg. 965

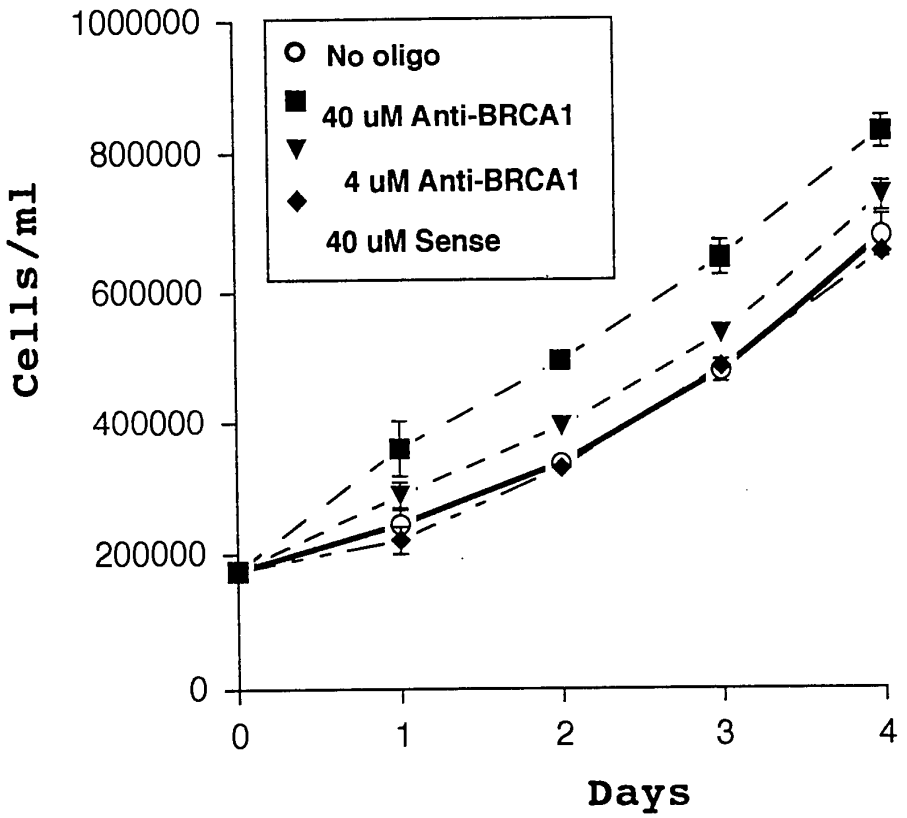

1000000
800000
600000
400000
200000
0
Cells/ml
No oligo
40 uM Anti-BRCA1
4 uM Anti-BRCA1
40 uM Sense
0
1
2
3
4
Days

Fig. 12A

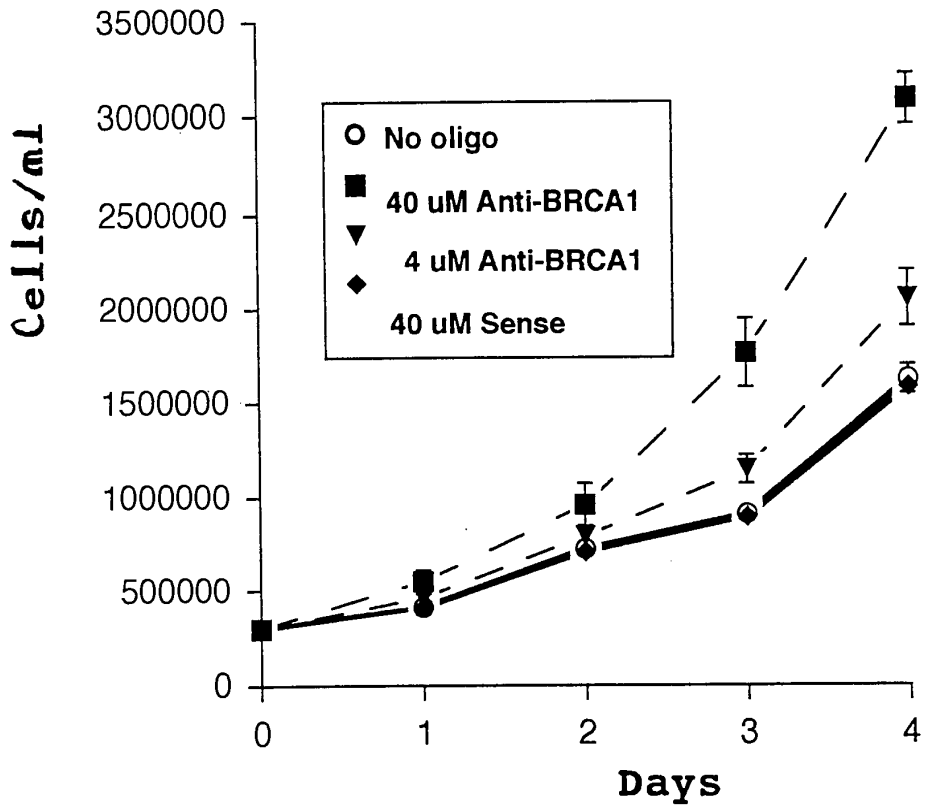

3500000
3000000
2500000
2000000
1500000
1000000
500000
0
Cells/ml
No oligo
40 uM Anti-BRCA1
4 uM Anti-BRCA1
40 uM Sense
0
1
2
3
4
Days

Fig. 12B

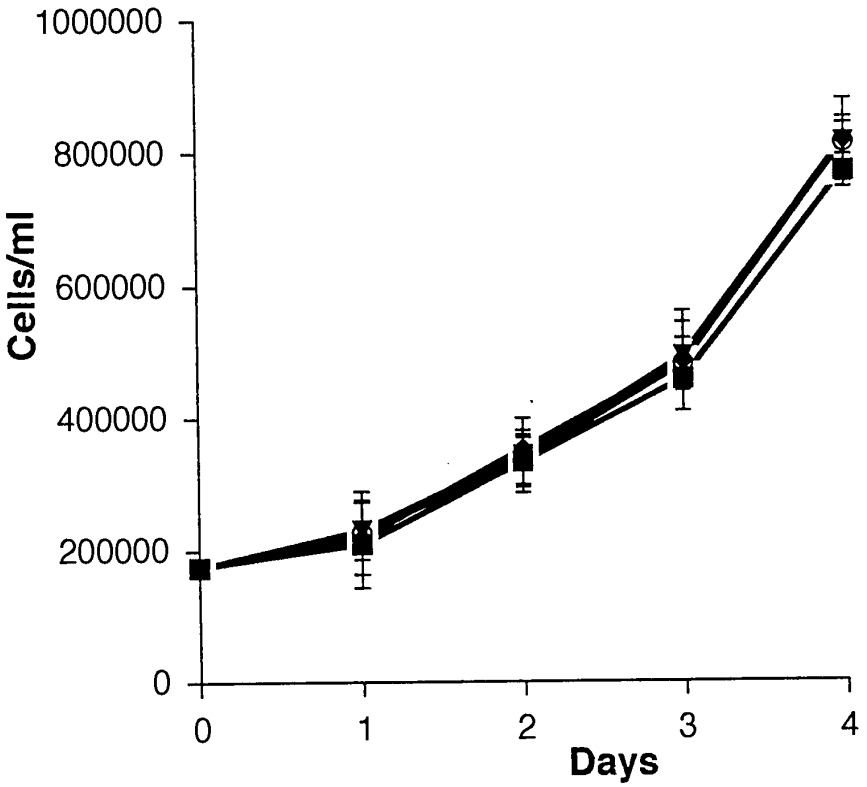
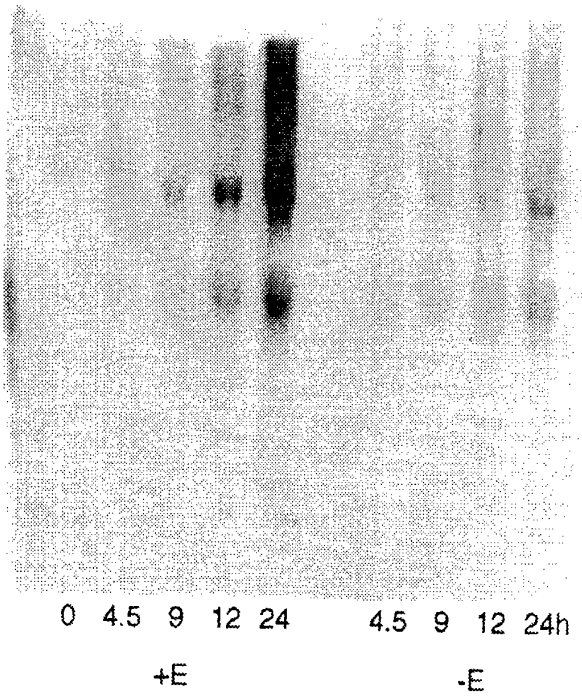


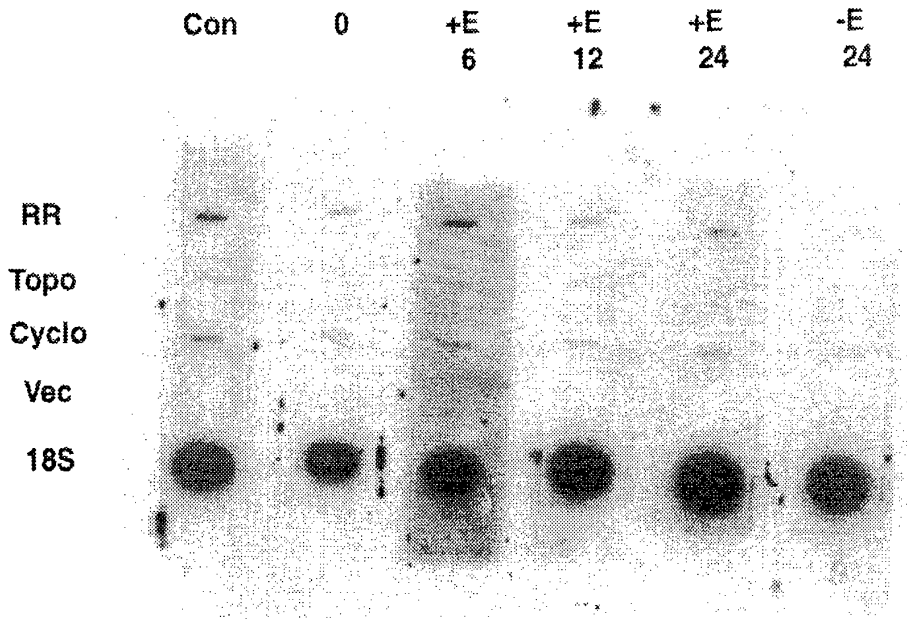
Fig. 12C

Fig. 13A



GeneDX 1009, pg. 969

Fig. 13B



GeneDX 1009, pg. 970

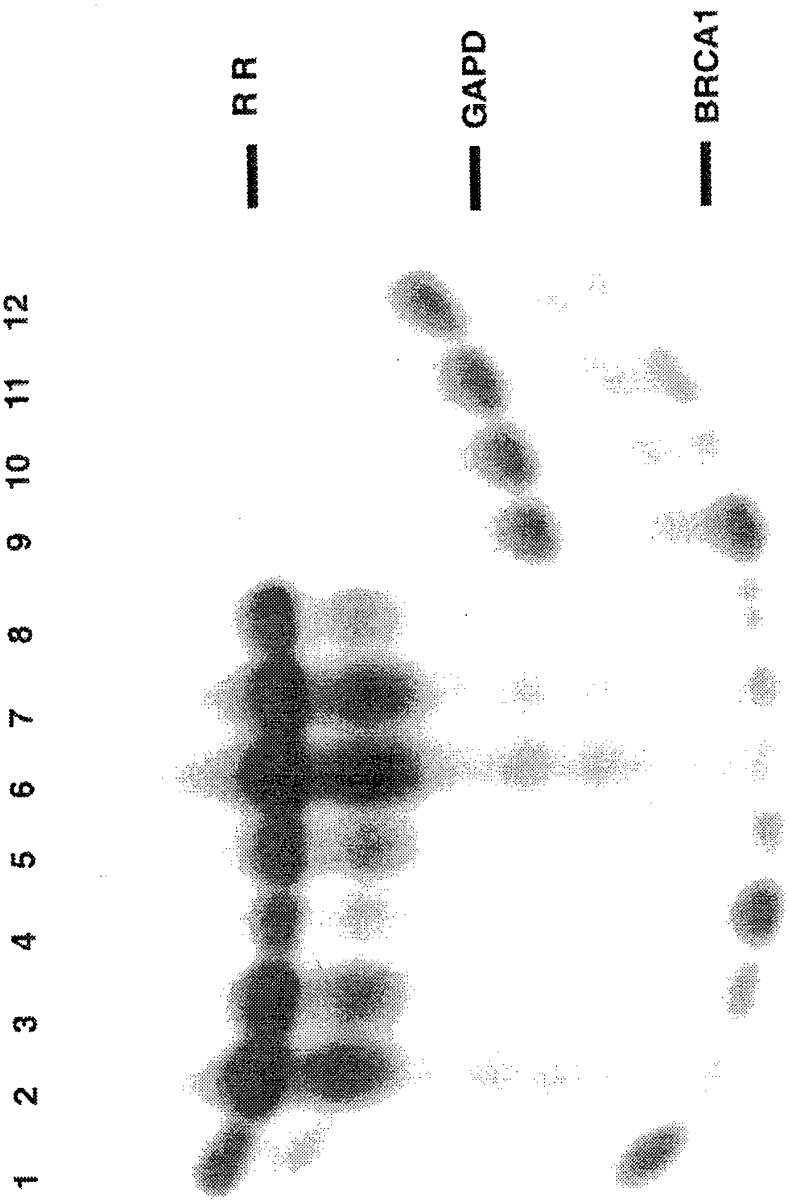


Fig. 14

GeneDX 1009, pg. 971

INTERNATIONAL SEARCH REPORT

Int... ...al application No.
PCT/US95/00608

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 7.1, 69.2, 172.3, 320.1; 514/44; 536/23.2, 23.5; 935/3, 6, 9, 11, 14, 23, 77, 78

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, Y	Science, Vol. 266, issued 07 October 1994, Y. Miki et al., "A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene *BRCA1*", pages 66-71, see entire document.	1-13, 15, 18-37, 40
Y	Cancer Research, Vol. 52, issued 15 December 1992, P. Liang et al., "Differential Display and Cloning of Messenger RNAs from Breast Cancer *versus* Mammary Epithelial Cells", pages 6966-6968, see entire document.	1-13, 15, 18-37
Y	Cancer Surveys, Vol. 18, issued 1993, J. T. Holt et al., "Histopathology: Old Principles and New Methods", pages 1-16, Tables 1 and 2 and Figures 1-5b, see pages 7-12.	1-13

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 21 APRIL 1995	Date of mailing of the international search report 04 MAY 1995
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Stephanie W. Zitomer, Ph.D. Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/00608

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	DNA (N.Y.), Vol. 5, No. 5, issued 1986, Neuhold et al., "Dioxin-Inducible Enhancer Region Upstream from the Mouse P-1450 Gene and Interaction with a Heterologous SV-40 Promoter", abstract, see entire document.	41-49
E, Y	US, A, 5,399,346 (ANDERSON ET AL.) 21 March 1995.	50-55

Form PCT/ISA/210 (continuation of second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

in[illegible]nal application No.	PCT/US95/00608

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
1-13, 15, 18-37, 40-55

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

GeneDX 1009, pg. 974

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/00608

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C07H 21/02, 21/04; C12Q 1/68; G01N 33/53; C12P 21/00; C12N 15/63, 15/85; A61K 48/00

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/6, 7.1, 69.2, 172.3, 320.1; 514/44; 536/23.2, 23.5; 935/3, 6, 9, 11, 14, 23, 77, 78

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, DIALOG: Biosis, Derwent Biotech. Abstracts, WPI, Chem. Abstr., Diss. Abstr., Embase, Medline, Current Biotech. Abstr. (Royal Soc.); search strat: (cancer or carcinoma)(p)(breast or ovar?)(p)(gene or nucleic)(p)gene(..) sequence, nucleic(..) sequence?

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1-13 and 25-37, drawn to a first process of using and a first product used: a diagnostic process using the product nucleic acids.

Group II, claims 15, 18-24 and 40 , drawn to a second process of using, an immunoassay.

Group III, claims 38 and 39 , drawn to a second product used, proteins including polypeptides and antibodies.

Group IV, claims 41-45, drawn to a third process of using, a process of screening compounds for activity in breast cancer treatment.

Group V, claims 46-49, drawn to a fourth process of using, a process for producing an indicator compound.

Group VI, claims 50-55, drawn to a fifth process of using, a process for treating breast cancer.

The inventions listed as Groups I-VI do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The inventions of Groups III-VI lack the diagnostic technical feature and the inventions of Groups II-VI lack the nucleic acid special feature of Group I while the inventions of Groups II and IV-VI each have a different result such that they individually lack the special features of the others that are responsible for that result: The Group II process has a diagnostic result; the Group IV invention identifies a compound that affects the expression of the BRCA1 gene; the Group V process produces an indicator compound; the invention of Group VI treats cancer.

GeneDX 1009, pg. 975