ISSUE CLASSIFICATION	Class	Subclass

08/

[redacted]

UTILITY SERIAL NUMBER	PATENT D.	PATENT NUMBER
08/348824		

SERIAL NUMBER	FILING DATE	CLASS	SUBCLASS	GROUP ART UNIT	EXAMINER
08/348824	11/29/94	435	6	1807	Rees

APPLICANTS

RONDA N. BATTULIS-ELDENS, SALT LAKE CITY, UT; JACQUES SIMARD, QUEBEC, CANADA.

CONTINUING DATA*******************
VERIFIED THIS APPLN 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 08/269,221 08/12/94
WHICH IS A CIP OF 08/248,602 05/20/94 ABN

DR 7/A/96

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

NONE DR 7/19/96

FILED WITH 7 DRWGS, 1 CLAIMS, 0 TOTAL CLAIMS
POWER OF ATTORNEY FILED 06/07/95

Foreign priority claimed 35 USC 119 conditions met	AS FILED	STATE OR COUNTRY	SHEETS DRWGS.	TOTAL CLAIMS	INDEP. CLAIMS	FILING FEE RECEIVED	ATTORNEY'S DOCKET NO.
☐ yes ☑ no ☐ yes ☐ no Verified and Acknowledged DR Examiner's initials	→	UT	10	18	2	$860.00	2488A1093474

ADDRESS

BONNIE A. STRAUCH, MORRIS & MANNING
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

TITLE

[illegible] BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE MUTATIONS

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

PARTS OF APPLICATION FILED SEPARATELY	Applications Examiner

NOTICE OF ALLOWANCE MAILED	Assistant Examiner	CLAIMS ALLOWED: Total Claims	CLAIMS ALLOWED: Print Claim
1			

ISSUE FEE: Amount Due	ISSUE FEE: Date Paid	DRAWING: Sheets Drwg.	DRAWING: Figs. Drwg.	DRAWING: Print Fig.

Label Area	Primary Examiner	ISSUE BATCH NUMBER
	PREPARED FOR ISSUE	

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.

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

(FACE)

ABANDONED

08/348824

briefed in group 1800

PATENT APPLICATION

08348824

APPROVED FOR LICENSE

INITIALS LDEL 14426

ABANDONED

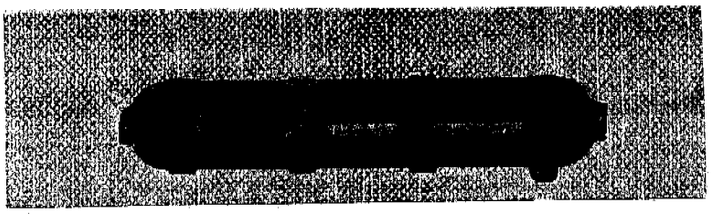
Date Entered or Counted

CONTENTS

Received or Mailed

	1. *Application* _____ *papers.*		
	2. Ltr re: Raw Sequence List w/ERRORS	12/20/94	1-4-95
	3. Ltr re: Rec. & Ext.	3-6-96	
	4. Raw Seq Listing (OK)	4-7-96	
	5. Rej 3 Mos.	7-22-96	
	6. Extr. Abandt.	9-10-96	10/11
	7. Notice of Aband	10-15-96	
	8. Request for Access	1-8-98	
	9. Copy of Cover	1-12-98	10/11
	10. Request for Access	1-28-98	
	11. Request for CLAIMS	4/30/02	
	12.		
	13.		
	14.		
	15.		
	16.		
	17.		
	18.		
	19.		
	20.		
	21.		
	22.		
	23.		
	24.		
	25.		
	26.		
	27.		
	28.		
	29.		
	30.		
	31.		
	32.		

(FRONT)



SEARCHED

Class	Sub.	Date	Exmr.
536	23.1	7/19/96	DR
536	24.3		
435	6		
435	91.2		
435	7.1 or 8.9		

SEARCH NOTES

	Date	Exmr.
APS, STN: BIOSIS, BIOTECHABS, BIOTECHDS, CANCERLIT, CABA, DRUGU, EMBASE, BIOBUSINESS, DGENE, MEDLINE, USPATFULL, CABA, WPIDS, IFIPAT, JAPIO, Searched related cases more extensive Search notes — related cases see 08/409,305 RSP.	7/19/96	DR

INTERFERENCE SEARCHED

Class	Sub.	Date	Exmr.

(RIGHT OUTSIDE)

Staple Issue Slip Here

POSITION		ID NO.	DATE
CLASSIFIER		8	12-14-94
EXAMINER		230	12-30-94
TYPIST		18	6-7
VERIFIER	429	614	
CORPS CORR.			
SPEC. HAND		401	5/31/96
FILE MAINT.		309	1-3-95
DRAFTING			

INDEX OF CLAIMS

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

PATENT APPLICATION SERIAL NO. 08/348824

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

PTO-1556
(5/87)

BAR CODE LABEL

U.S. PATENT APPLICATION

SERIAL NUMBER	FILING DATE	CLASS	GROUP ART UNIT
08/348,824	11/29/94	435	1807

APPLICANT

DONNA M. SHATTUCK-EIDENS, SALT LAKE CITY, UT; JACQUES SIMARD, QUEBEC, CANADA.

CONTINUING DATA********
VERIFIED THIS APPLN 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
WHICH IS A CIP OF 08/246,602 05/20/94 ABN

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

FOREIGN FILING LICENSE GRANTED 06/07/96

STATE OR COUNTRY	SHEETS DRAWING	TOTAL CLAIMS	INDEPENDENT CLAIMS	FILING FEE RECEIVED	ATTORNEY DOCKET NO.
UT	10	18	2	$860.00	248841093474

ADDRESS

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

TITLE

17Q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE MUTATIONS

This is to certify that annexed hereto is a true copy from the records of the United States Patent and Trademark Office of the application which is identified above.

By authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

Date Certifying Officer

05.12.95

PATENT APPLICATION SERIAL NO. 08/348824

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

PTO-1556
(5/87)

PATENT APPLICATION SERIAL NO. 08/348824

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

PTO-1556
(5/87)

BAR CODE LABEL

U.S. PATENT APPLICATION

SERIAL NUMBER	FILING DATE	CLASS	GROUP ART UNIT
08/348,824	11/29/94	435	1807

APPLICANT

DONNA M. SHATTUCK-EIDENS, SALT LAKE CITY, UT; JACQUES SIMARD, QUEBEC, CANADA.

CONTINUING DATA********
VERIFIED THIS APPLN 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
WHICH IS A CIP OF 08/246,602 05/20/94 ABN

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

FOREIGN FILING LICENSE GRANTED 06/07/96

STATE OR COUNTRY	SHEETS DRAWING	TOTAL CLAIMS	INDEPENDENT CLAIMS	FILING FEE RECEIVED	ATTORNEY DOCKET NO.
UT	10	18	2	$860.00	248841093474

ADDRESS

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

TITLE

17Q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE MUTATIONS

This is to certify that annexed hereto is a true copy from the records of the United States Patent and Trademark Office of the application which is identified above.

By authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

Date Certifying Officer

08/348824

Commissioner of Patents and Trademarks
Washington, D.C. 20231

MAIL ROOM
NOV 29 1994
PAT & TRADE MARK OFF

Sir:

Transmitted herewith for filing is the patent application of:

Inventor(s): Donna M. SHATTUCK-EIDENS and Jacques SIMARD

ETAC

For: **17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE MUTATIONS**

Enclosed are:

[X] ten (10) sheet(s) of drawings (8 figures), in triplicate.

[ ] Assignment of the invention (with cover sheet) to _____

[X] Statement Pursuant to 37 CFR §1.821(f) and Computer Disk.

[X] APPLICATION FILED WITH NO DECLARATION AND NO FEE PURSUANT TO 37 CFR §1.37(d).

	(COL. 1)	(COL. 2)	SMALL ENTITY			OTHER THAN SMALL ENTITY	
FOR:	NO. FILED	EXTRA	RATE	FEE		RATE	FEE
BASIC FEE:				$365.00	OR		$730.00
TOTAL CLAIMS:	18-20		X $ 11	$	OR	X $ 22	$
INDEP CLAIMS:	2- 3		X $ 38	$	OR	X $ 76	$
MULTIPLE DEPENDENT CLAIM PRESENTED:			+ $120	$	OR	+ $240	$
			TOTAL	$	OR	TOTAL	$730.00

[ ] A check in the amount of $_____ to cover the filing fee and assignment recordation fee is attached.

[ ] The Commissioner is authorized to charge the following fees associated with this communication or credit any overpayment to Deposit Account No. 22-0261. Duplicate copies are attached.

[ ] Additional filing fees required under 37 CFR 1.16.
[ ] Patent application processing fee under 37 CFR 1.17.

Please direct all correspondence to: Jeffrey L. Ihnen, Reg. No. 28,957

Respectfully submitted,

Jeffrey L. Ihnen
Registration No. 28,957
Attorney for Applicant

VENABLE, BAETJER, HOWARD & CIVILETTI
1201 New York Avenue, N.W., Suite 1000
Washington, D.C. 20005
202-962-4810

Atty. Docket No. 24884-109347-4
Dated: 29 November 1994

MAIL ROOM
NOV
29
1994
PAT. & TRADE MARK OFF.

TITLE OF THE INVENTION

17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE MUTATIONS

This work has been supported in part by National Cancer Institute grant Numbers CA-48711 and CA-55914. The United States government has certain rights in the invention.

CROSS REFERENCE TO RELATED APPLICATIONS

This application 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, which is a continuation-in-part of application Serial No. 08/246,602, filed on 20 May 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 bibliography.

08/348824 BOX SEQUENCE
05.12.95



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. Less 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).

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. (Hill *et al.*, 1978; 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 (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 basepairs) 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

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 STS's, P1s and BAC's 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 cystines 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 3575 to 3874 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 8A shows a mutation and cosegregation analysis in BRCA1 kindreds. Carrier individuals are represented as filled circles and squares in the pedigree diagrams. Nonsense mutation in Kindred 2082. P indicates the person originally screened, b and c are haplotype carriers, a, d, 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 8B 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 8C 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.

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 basepairs, 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.

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 M.H., and Wallace B.R., 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.*, 1993).

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 Mfdl91 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 Mfdl91 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.

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 Cot-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 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 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.*, 1988) 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 adaptors 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 adaptors. These adaptors serve as amplification sites for PCR. The target and probe sequences are denatured and mixed with human Cot-1 DNA to block repetitive sequences. Solution hybridization is carried out to high Cot-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 suceptibility 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 delations 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 make 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.*), 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 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 determintions, 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 fuction, directly correlates to a 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.*, 1989 (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 sequence 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"** 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 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.), pendant 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 Ausebel *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 labelling probes see, e.g., Sambrook *et al.*, 1989 or Ausebel *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 labelled 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, labelling, 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 labelling polypeptides are well known in the art. See, e.g., Sambrook *et al.*, 1989 or Ausebel *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 homogenous"** 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 or 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 for purification utilized.

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 tecniques. The modified nucleic acid is prepared by conventionl 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, substitutions, 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, 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; Ausebel *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 & 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 are described, e.g., in Sambrook *et al.*, 1989 or Ausebel *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 *et al.*, 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 strand 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 Ausebel *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 Ausebel *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 molony 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, Kruse & Patterson, 1973. 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). A preferred PCR based strategy contemplated within the scope of this invention is provided in Example 8. 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 3575 to 3874 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 labelling 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 labelling nucleic acid probes and their use in biotin-avidin based assays see Rigby, et al., 1977 and Nguyen, et al. Example 11 details a method using a biotin-avidin type assay employing BRCA1.

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 differenct 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 claim 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 immunoenzymetric assays (IEMA), including sandwich assays using monoclonal and/or polyclonal antibodies. Exemplary sandwich assays are described by David et al, in US 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, European Patent Application 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.

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 modelling 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-idiotypic 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. 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; Zenke *et al.*, 1990; Wu *et al.*, 1989; Wolff *et al.*, 1991; Wagner *et al.*, 1990; Wagner *et al.*, 1991; Cotten *et al.*, 1990; Curiel *et al.*, 1991; Curiel *et al.*, 1991). 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 exmple, 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 mutgenesis 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

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.

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 (KIOO1, 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.*, 1993), 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 *et al.*, 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		42D6(D17S588) Recombination		Mfd188 (D17S579) 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.

GeneDX 1035, pg. 60

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 (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, Mfd191 (provided by James Weber), THRA1 (Futreal et al. 1992), 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 the 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

GeneDX 1035, pg. 64

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 1035, pg. 65

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 BAC. 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 were 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.

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

GeneDX 1035, pg. 71

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 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

In kindred K1901, the new markers showed no observable recombination with breast cancer susceptability, 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 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 an Region Flanked by tdj11474 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 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 centomere - 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 consegregating 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 Narod *et al.* (1993).

Kindred 1813

Kindred 1813 is a small family with four cases of breast cancer diagnosed under the age of 40 whose mother had breast cancer diagnosed at age 45 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 Smard *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 Figure 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, DNA 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 $NaPO_4$ pH 6.8, 100 μg/ml Salmon Sperm DNA, lx Denhardts, 50% formamide, 0.1% SDS, and 2 μg/ml cot-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.*, 1989). The eluted Not I digested DNA (~15 kb to 25 kb) was then digested with EcoRI restriction enzyme to give smaller fragments (~O.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 O.l 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 Cot-1 DNA for 2-4 hrs. as per the manufacturer's instructions (Gibco/BRL, Cat. #5279SA). The Cot-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.

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 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 later 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 rain (Stratagene, Cat. 936206), human testing (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.*(1989). 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 EcoRI 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 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 PIs 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 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 eletrophoresis to determine positions in the sequence that contain either mutations such as detections 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 gene 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 (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 cystine 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 five potential alternate splice donors, the longest of which would add 1,050 bp to the sequence in SEQ ID NO:2. The predominate 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 sit in testis. The A in position 148 represents the potential start site in somatic tissue. The G at position 363 represents the cannonical first splice donor site. The G at position 452 represents the first splice donor site in two clones (testis 1 and testis 2). The G at position 897 represents the first splice donor site in thymus 3. The T at position 1519 represents the first splice acceptor site. The A at position 1538 is the cannonical start site, which appears at position 64 of SEQ ID NO:1. Partial genomic DNA sequences determined for BRCA1 are set forth in SEQ ID NO:14. The lower case letters denote intron sequence while the upper case letters denote exon sequence. Uncertainty about sequence at an intron/exon boundary is indicated by vvvvv. Indefinite intervals (between introns) are designated with vvvvvvvvvvvvv. The intron/exon junctions are shown in Table 9.

TABLE 9

Exon No.	Base Position 5′	Base Position 3′	Length	Intron Borders 5′	Intron Borders 3′
e1	1	100	99	GATAAATTAAAACTGCGACTGCGCGGCGTG[15*]	...AAAAATAAAGGACGCTGTCGAYTAGTTCTT[16]
e2	101	204	103	AAGATATATATATATGTTTTTCTAATGTGT[17]	GTAAGTCAGCACAAGAGTGTATTAATTTGG[18]
e3	205	258	53	TTATTGAGCCTCATTTATTTTCTTTTTCTC...[19]	...ATGTGGCTCCATTATTAGCTTTTGTTTTTG[20]
e4	**	**	110	AGCTACTTTTTTTTTTTTTTTTTGAGACAG[21]	GTAAGTGCACACCACCATATCCAGCTAAAT[22]
e5	259	336	77	TTGTTCTTTCTTTCTTTATAATTTATAGAT...[23]	GATATAANNNGAANTGATGCTAGGTTGGAA[24]
e6	337	425	88	GATAATCACTTGCTGAGTGTGTTTCTCAAA...[25]	GTAAGTGTTGAATATCCCAAGAATGAGAGT[26]
e7	426	568	142	AAACATAATGTTTTCCCTTGTATTTTACAG[27]	AAACCATTTGTTTTCTTCTTCTTCTTCTTC[28]
e8	569	670	101	TTGACTGTTCTTTACCAWACTGTTWARCAG[29]	GGTAAGGGTCTCAGGTTTTTTAAGTATTTA[30]
e9	671	716	45	TTGATTTATTTTTTGGGGGGAAATTTTTTA[31]	GGTGAGTCAAAGAGAACCTTTGTCTATGAA[32]
e10	717	793	76	TTCTTATTAGGACTCTGTCTTTTCCCTATA[33]	GGTAATGGCAAAGTTTGCCAACTTAACAGG[34]
e11	794	4220	3426	TGAGTACCTTGTTATTTTTGTATATTTTCA[35]	GTATTGGAACCAGGTTTGTGTGTGTGGCCC[36]
e12	4221	4313	92	ACATCTGAACCTCTGTTTTTGTTATTTAAG[37]	NGNGAATGTAATCCTAATATTTCNCNCCNA[38]
e13	4314	4480	166	TCTTGGTACCATNNATCGTTTTTGAATNAG[39]	TGTGTATTGTTGGCCAAACACTGATATCTT[40]
e14	4481	4605	124	AGATTTGTTTTCTCATTCCATTTAAAGCAG[41]	GTAAGAAACATCAATGTAAAGATGCTGTGG[42]
e15	4606	4798	192	CGATGGTTTTCTCCTTCCATTTATCTTTCT[43]	GTAATATTTCATCTGCTGTATTGGARCAMA[44]
e16	4799	5109	310	TGGAAGGGAAGCCCACTCNAAAAAANNTTN...[45]	GTGAGTGTATCCATATGTATCTCCCTAATG[46]

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

GeneDX 1035, pg. 86

TABLE 9 (cont'd)

Exon No.	Base Position 5'	3'	Length	Intron Borders 5'	3'
e17	5110	5198	88	ATGATAATGGAATATTTGATTTAATTTCAG[47*]	TATACCCAAGAACCTTTACAGAATACCTTK[48]
e18	5199	5277	78	TAATCCTTTGAGTGTKTCTCATTCTGCAGA[49]	KAAGKATAATACTATWTCTCCCCYCCTCCC[50]
e19	5278	5316	38	TAACCTGTCTTTTCTATGATCTCTTTAGGG[51]	GTAAGTACTTGATGTTACAAACTAACCAGA[52]
e20	5317	5400	83	TTATCCTGATGGGTGTGNTGTTTCTTTCAG[53]	GTAAAGCTCCCTCCCTCAAGTTGACNNAAA[54]
e21	5401	5455	54	TAATTTGGATTCCCGTCTCGGGTTCAGATC[55]	GTAAGAGCCTCGGGAGAANCCAGAGTTCCA[56]
e22	5456	5530	74	GTGATTTTAGATCCTAAATGTCCATTTTAG[57]	...CAGAGCAAGACCCTGTCTCAAAAACAAACA[58]
e23	5531	5590	59	TTGAATGCTCTTTCCTTCCTGGGGANCCAG[59]	TTTATCACTCATTACCTGGTGCTTGAGTAG[60]
e24	5591	5918	327	CTAATCTCTGCTTGTGTTCTCTGTCTCCAG[61]	

* 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).

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.

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, 3 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	4000
1910	extra C	frameshift	5329
2099	T→G	Met→Arg	5387
2035	?	loss of transcript	
1901	11 bp deletion	frameshift	133

* In Sequence ID NO:1

All three sequence variants are heterozygous and each appears in only one of the kindreds. Kindred 2082 contains a nonsense mutation in coding exon 10 (Fig. 8A), Kindred 1910 contains a single nucleotide insertion in coding exon 19 (Fig. 8B), and Kindred 2099 contains a missense mutation in coding exon 20, resulting in a Met->Arg substitution (Fig. 8C). 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 107

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 548 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 3921) appeared homozygous, whereas her genomic DNA revealed heterozygosity at this position (Fig. 8C). 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. 8B). 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. 8A). 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 haploytpe, 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. 8C).

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 caucasions 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 3 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. In particular, no mutation has been found in Kindred 1901, which has a phenotypic pattern similar to Kindred 2035 and is highly unlikely to be due to BRCA1. It may segregate a similar regulatory mutation, but is uninformative for all the polymorphisms examined to date.

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 mutations identified thus far are all located in the C-terminal half of the peptide. 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.*, 1992; 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 were found, one in an ovarian tumor, four in breast tumors and one in a male unaffected haplotype carrier (Table 12). One mutation, Glu1548Ter, introduced a stop codon that would create a truncated protein missing 273 amino acids at the carboxy terminus. In addition, three missense mutations were identified. Two cases, Ala1715Glu and Met1782Arg,

involve substitutions of small, hydrophobic residues by charged residues, and a third, Pro1644Leu, found in an ovarian tumor, involves a more conservative substitution. The missense mutation in the male carrier involve a substitution of a charged residue by a hydrophobic residue.

TABLE 12

Predisposing Mutations

Patient	Codon	Nucleotide Change	Amino Acid Change	Age of Onset	Family History
BT098	1548	GAG → TAG	Glu → Stop	39	-
OV24	1644	CCA → CTA	Pro → Leu	44	-
BT106	1715	GCG → GAG	Ala → Glu	24	+
MC44	1782	ATG → AGG	Met → Arg	42	+
17764	954/5	2bp deletion	frameshift	31	+
19964	841	CGG → TGG	Arg → Trp		+*

* Unaffected haplotype carrier, male

Several lines of evidence suggest that all four 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 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. This patient had a sister who died of breast cancer at age 34, another sister who died of lymphoma, and a brother who died of lung cancer. Her mutation (Met1782Arg) 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 Met1782Arg 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., 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 3575-3874 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 $N/aP0_4$. 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, 40mM $NaPO_4$ at 65°C, followed by two 30-min washes in 1% SDS, 40mM $NaPO_4$ 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, 0.3 M NaCl, and 5X Denhardt's solution (see Sambrook, et al. *supra*). 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 6M urea, 1x standard saline citrate (SSC), 0.1%SDS and one 10 min. wash in 1x SSC, O.1% Triton®X-100. The blots are rinsed for 10 min. at room temp. 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.

Spleens are removed from immune mice and a single cell suspension is prepared (see Harlow, et al. 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, et al. 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 BRCA-1 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, 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 mutants) 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.

EXAMPLE 15

Analysis of BRCA1 Mutations

The DNA samples which were screened for BRCA1 mutations were extracted from blood or tumor samples from patients with breast or ovarian cancer (or known carriers by haplotype analysis) who were participating in research studies on the genetics of breast cancer. All subjects signed appropriate informed consent. Table 13 details the number of samples, ascertainment criteria, and screening method for each set of samples screened.

TABLE 13

Sets of DNA Samples Screened for Mutations in BRCA1

Source of Samples	Description of Samples	Screening Method	No. Samples Screened	No. Mutations Found to Date
UTAH-2	Br/Ov Families	SEQ	10	2
MONTREAL	Br/Ov Families	SEQ	30	13
MSKCC	Br and Br/Ov Families	SEQ	14	2
MSK/UT-1	Early Onset Br Cases	SEQ	24	1
STRANG	Br and Br/Ov Families	SEQ	12	4
STOCKHOLM	Br and Br/Ov Families	SEQ	15	3
USC-1	Bilat Br Proband, High-Risk	SEQ	7	3
TUMOR-3	Early Onset Br Tumors	SEQ	14	1
USC-2	Bilat Br <50+1° rel Br	ASO	59	5
MSK/UT-2	Early Onset Br Cases	ASO	109	3

[1] Most sample groups contained heterogenous mixture of samples. The most representative description of each set is given.

[2] SEQ-Direct sequencing of PCR products; SSCA-Single Strand Confirmation Assay; CMC-Chemical Mismatch Cleavage; HA-Heteroduplex analysis; ASO-Allele-Specific Oligonucleotide.

Although the original mutations described in Miki *et al.* were detected through screening of cDNA, 25 pairs of intronic PCR primers were used to amplify the complete coding sequence and splice junctions from genomic DNA for the majority of the remaining samples. Updated primer information is publicly available via anonymous ftp from morgan.med.utah.edu in the directory pub/BRCA1. Where possible, DNA sequence variations were tested for cosegregation with breast or ovarian cancer in the family. Further evidence of a casual role of a sequence variant in cancer was provided by proving the absence of the putative mutation in a set of control individuals. Screening for specific, previously-identified mutations in large sets of selected samples was performed using ASO hybridization.

Table 14 describes several of the 67 mutations found screening the entire BRCA1 coding sequence, the 21 intron/exon boundaries and by finding polymorphic sites in genomic DNA reduce to monomorphic sites in cDNA. Two common mutations were found and their frequencies in other samples were examined by ASO analysis (Table 15). Tables 16 and 17 describe the distribution of mutations by type and by location within the BRCA1 coding sequence, respectively. By far, the majority of mutations identified were frameshifts which accounted for 56% of the 39 different mutations found to date in BRCA1 and in 65% of all samples shown to have BRCA1 mutations through complete screening. Globally, no statistically significant departure from a random distribution across the coding sequence of BRCA1 was found ($\chi^2 = 2.00$, 2 df, $p = 0.37$) among the 37 distinct mutations found in the coding sequence of BRCA1 to date.

TABLE 14

Mutations Identified by Complete Screening of the BRCA1 Gene

Sample Set	Family	# Cases		Mutation Description			
		BR	OV	Type[1]	Exon	Codon	Mutation[2]
MONTREAL	180	2	2	FS	2	23	185 del AG -> ter 39
MONTREAL	235	4	2	FS	2	23	185 del AG -> ter 39
MONTREAL	253	1	3	FS	2	23	185 del AG -> ter 39
MONTREAL	255	0	7	FS	2	23	185 del AG -> ter 39
MSKCC	210311	3	0	FS	2	23	185 del AG -> ter 39
USC-1	008	2	1	FS	2	23	185 del AG -> ter 39
TUMOR-3[4]	-	1	0	FS	11	270	926 ins 11 -> ter 301
MSK/UT-1	-	1	0	FS	11	270	926 ins 11 -> ter 301
MONTREAL	270	4	3	FS	11	339	1128 ins A -> ter 345
STRANG	2903	1	2	FS	11	339	1128 ins A -> ter 345
MONTREAL	185	1	3	FS	11	392	1294 del 40 -> ter 398
USC-1	052	5	1	FS	11	655	2080 ins A -> ter 672
USC-1	068	2	1	FS	11	655	2080 ins A -> ter 672
STRANG	2802	2	2	FS	11	725	2294 del G -> ter 735
MONTREAL	179	2	3	MS	11	826	Thr 826 Lys
STOCKHOLM	AL48	3	1	FS	11	826	2477 del C -> ter 845
STOCKHOLM	BR33	5	1	FS	11	826	2477 del C -> ter 845
UTAH-2	2305	2	7	FS	11	956	2982 del 5 -> ter 968
MONTREAL	218	5	1	FS	11	1002	3121 del A -> ter 1023
STOCKHOLM	BR24	2	1	FS	11	1016	3166 ins 5 -> ter 1025

TABLE 14 (Cont'd)

Mutations Identified by Complete Screening of the BRCA1 Gene

Sample Set	Family	# Cases BR	# Cases OV	Mutation Description Type[1]	Exon	Codon	Mutation[2]
UTAH-2	2039	3	2	MS	11	1347	Arg 1347 Gly
MONTREAL	183	4	1	FS	11	1355	4184 del 4 -> ter 1364
STRANG	1900[3]	3	1	NS	13	1443	Arg 1443 ter
STRANG	8622[3]	4	1	FS	16	1656	5085 del 19 -> ter 1671
MONTREAL	101	2	2	FS	20	1756	5382 ins C -> ter 1829
MONTREAL	162	3	1	FS	20	1756	5382 ins C -> ter 1829
MONTREAL	166	5	2	FS	20	1756	5382 ins C -> ter 1829
MONTREAL	279	4	0	FS	20	1756	5382 ins C -> ter 1829
MSKCC	193549	0	3	FS	20	1756	5382 ins C -> ter 1829

[1] FS-Frameshift; NS-Nonsense; MS-Missense; SP-Splice Site; IR-Inferred Regulatory.

[2] For Missense and Nonsense mutations, the mutation description contains: wild type amino acid, affected codon, altered amino acid (or ter). For frameshift mutations, the format is: nucleotide, insertion or deletion, specific nucleotides changed (if <3) or number inserted or deleted (if >2) and the amino acid in which the frameshift results in a termination signal. Nucleotides refer to the BRCA1 cDNA sequence in GENBANK under Accession No. U-14680.

[3] The mutation in this family was independently identified in both the Myriad and University of Pennsylvania Labs.

[4] The mutation identified in this tumor was also found in the germline of the individual.

TABLE 15

Frequency of Two Common BRCA1 Mutations

Set	Number Studied	Number of Mutations Found *185 del AG*	*5382 ins C*
USC-1	59	4	1
MSK/UT-2	109	3	0
GLASGOW-2	100	Not tested	3
GLASGOW-3	100	Not tested	2
CRC-OV	250	Not tested	1

TABLE 16

Observed Frequency of Different Types of Mutations

Mutation Type	N (Percent) Distinct Mutations[1]	All Mutations[2]
Frameshift	22 (56)	44 (65)
Nonsense	7 (18)	10 (15)
Missense	8 (21)	12 (17)
Other	2 (5)	2 (3)

[1] Identical mutatiions are counted only once in this column.
[2] Each sample in which a mutation has been identified is counted in this column.

TABLE 17

Distribution of Identified Mutations in BRCA1 Coding Sequence

Mutations	Amino Acids 1-621	622-1242	1243-1863
Distinct	9	12	16
All	21	15	30

Mutations have been found in many different regions of the gene - phenotypically severe mutations have been found both in the extreme 5' end of BRCA1 as well as in the extreme 3' portion

of the gene. One such mutation found in a family with seven early-onset breast cancer cases produces a protein that is only missing the terminal 10 amino acids, indicating that this region of BRCA1 plays a role in normal gene function. It is noteworthy the overwhelming majority of alterations in BRCA1 have been either frameshift or nonsense mutations. Of the total of 68 independent individuals with mutations identified by complete screening of the coding sequence, 54 (79%) are frameshift or nonsense mutations, resulting in an unstable or truncated protein product.

In BRCA1, to date, two mutations appear to be relatively common. The *5382 ins C* BRCA1 mutation in codon 1756 and the *185 del AG* mutation in codon 23 were identified by direct sequencing in seven (10%) and eight (12%) of the 68 probands in which mutations have been identified, respectively. In addition to these common mutations, two mutations were identified three times and there are eleven additional mutations which have been found twice by a complete screen of the cDNA. The total fraction of mutations which are recurrent is 38% and these 15 recurrent mutations comprise 63% of the 68 mutations detected in probands. Many of the probands screened to date for BRCA1 mutations were selected for having a high prior probability of having such mutations. Thus the mutations found in this set may not be representative of those which would be identified in other sets of patients. However, the two most frequent BRCA1 mutations (*5382 ins C* and *185 del AG*) have been found multiple times in targeted screening in sets of probands who were either unselected for family history or ascertained with minimal family history.

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. Nat. Acad. Sci. USA* 77:5399-5403.

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

Anderson, D.E. (1992). *J. Natl. Cancer Inst.* **48**:1029-1034.

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

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

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

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.

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. Nat. 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. Nat. Acad. Sci. USA.* ___ _____

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

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

Cotton, *et al.* (1988). *Proc. Nat. 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.* (1991). *Proc. Nat. Acad. Sci. USA* **88**:8850-8854.

Curiel, *et al.* (1991). *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.

Erickson, J. *et al.*, (1990). *Science* **249**:%27-533.

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

Felgner, *et al.* (1987). *Proc. Nat. 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.* (1992). *Hum. Molec. Genet.* **1**:66.

Futreal, P.A., *et al.* (1992). *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. Nat. 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.* (1992). *Am. J. Hum. Genet.* **50**:1235-1241.

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

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.*

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

Johnson, J.L., *et al.* (1994). *Mol. Cell Biol.* 14:1956-63.

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.

Lovett, *et al.* (1991). *Proc. Nat. 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.* (1989). *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.

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.

Miki, Y., *et al.*, 1994. *Science* **266**:66-71.

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. Microbiol. Immunol.* **158**:25-38.

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

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

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

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. Nat. Acad. Sci. USA* **85**:3044-3048.

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

Novack, *et al.* (1986). *Proc. Nat. 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. Nat. Acad. Sci. USA* **86**:276-2770.

Orita, *et al.* (1989). *Genomics* **5**:874-879.

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

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

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

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

Pierce, *et al.* (1992). *Proc. Nat. 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.

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. Nat. Acad. Sci. USA* **86**: 232-236.

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

Smith, T.F., *et al.* (1981). *J. Mol. Biol.* **147**:195-197.

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

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.

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

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

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

Skolnick, M.H. (1980). Banbury Report No 4: Cancer Incidence in Defined Populations (J. Cairns, J. Lyon, M. Skolnick, eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 285-297.

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

Skolnick, M.H. (1977). *Genet. Hum. Cancer*, (J.J. Mulvill, R.W. Miller and J.F. Freeman, Jr., eds.), Raven Press, N.Y.

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

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

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

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

Sternberg (1990). *Proc. Nat. 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.

Thomas, A., *et al.*, *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.* (1991). *Proc. Nat. Acad. Sci. USA* **88**:4255-4259.

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

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

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

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

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

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 & 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*, (In press).

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

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

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

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

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Shattuck-Eidens, Donna M.
Simard, Jacques

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

(iii) NUMBER OF SEQUENCES: 14

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Venable, Baetjer, Howard & Civiletti
(B) STREET: 1201 New York Avenue, N.W., Suite 1000
(C) CITY: Washington
(D) STATE: DC
(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 #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US
(B) FILING DATE: 02-SEP-1994
(C) CLASSIFICATION:

(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

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/246,602
(B) FILING DATE: 20-MAY-1994

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Ihnen, Jeffrey L.
(B) REGISTRATION NUMBER: 28,957
(C) REFERENCE/DOCKET NUMBER: 24884-109347-2

(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: 5857 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(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: 64..5769

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

GGCCCCTGCG CTCAGGAGGC CTTCACCCTC TGCTCTGGGT AAAGTTCATT GGAACAGAAA 60

GAAATGGATT TATCTGCTCT TCGCGTTGAA GAAGTACAAA ATGTCATTAA TGCTATGCAG 120

AAAATCTTAG AGTGTCCCAT CTGTCTGGAG TTGATCAAGG AACCTGTCTC CACAAAGTGT 180

GACCACATAT TTTGCAAATT TTGCATGCTG AAACTTCTCA ACCAGAAGAA AGGGCCTTCA 240

CAGTGTCCTT TATGTAAGAA TGATATAACC AAAAGGAGCC TACAAGAAAG TACGAGATTT 300

AGTCAACTTG TTGAAGAGCT ATTGAAAATC ATTTGTGCTT TTCAGCTTGA CACAGGTTTG 360

GAGTATGCAA ACAGCTATAA TTTTGCAAAA AAGGAAAATA ACTCTCCTGA ACATCTAAAA 420

GATGAAGTTT CTATCATCCA AAGTATGGGC TACAGAAACC GTGCCAAAAG ACTTCTACAG 480

AGTGAACCCG AAAATCCTTC CTTGCAGGAA ACCAGTCTCA GTGTCCAACT CTCTAACCTT 540

GGAACTGTGA GAACTCTGAG GACAAAGCAG CGGATACAAC CTCAAAAGAC GTCTGTCTAC 600

ATTGAATTGG GATCTGATTC TTCTGAAGAT ACCGTTAATA AGGCAACTTA TTGCAGTGTG 660

GGAGATCAAG AATTGTTACA AATCACCCCT CAAGGAACCA GGGATGAAAT CAGTTTGGAT 720

TCTGCAAAAA AGGCTGCTTG TGAATTTTCT GAGACGGATG TAACAAATAC TGAACATCAT 780

CAACCCAGTA ATAATGATTT GAACACCACT GAGAAGCGTG CAGCTGAGAG GCATCCAGAA 840

AAGTATCAGG GTAGTTCTGT TTCAAACTTG CATGTGGAGC CATGTGGCAC AAATACTCAT 900

GCCAGCTCAT TACAGCATGA GAACAGCAGT TTATTACTCA CTAAAGACAG AATGAATGTA 960

GAAAAGGCTG AATTCTGTAA TAAAAGCAAA CAGCCTGGCT TAGCAAGGAG CCAACATAAC 1020

AGATGGGCTG GAAGTAAGGA AACATGTAAT GATAGGCGGA CTCCCAGCAC AGAAAAAAAG 1080

GTAGATCTGA ATGCTGATCC CCTGTGTGAG AGAAAAGAAT GGAATAAGCA GAAACTGCCA 1140

TGCTCAGAGA ATCCTAGAGA TACTGAAGAT GTTCCTTGGA TAACACTAAA TAGCAGCATT 1200

CAGAAAGTTA ATGAGTGGTT TTCCAGAAGT GATGAACTGT TAGGTTCTGA TGACTCACAT 1260

GATGGGGAGT CTGAATCAAA TGCCAAAGTA GCTGATGTAT TGGACGTCCT AATTGAGGTA 1320

GATGAATATT CTGGTTCTTC AGAGAAAATA GACTTACTGG CCAGTGATCC TCATGAGGCT 1380

TTAATATGTA AAAGTGACAG AGTTCACTCC AAATCAGTAG AGAGTAATAT TGAAGGCCAA 1440

ATATTTGGGA AAACCTATCG GAAGAAGGCA AGCCTCCCCA ACTTAAGCCA TGTAACTGAA 1500

AATCTAATTA TAGGNGCATT TGTTTCTGAG CCACAGATAA TACAAGAGCG TCCCCTCACA 1560

AATAAATTAA AGCGTAAAAG GAGACCTACA TCAGGCCTTC ATCCTGAGGA TTTTATCAAG 1620

AAAGCAGATT TGGCAGTTCA AAAGACTCCT GAAATGATAA ATCAGGGAAC TAACCAAACG 1680

GAGCAGAATG GTCAAGTGAT GAATATTACT AATAGTGGTC ATGAGAATAA AACAAAAGGT 1740

GATTCTATTC AGAATGAGAA AAATCCTAAC CCAATAGAAT CACTCGAAAA AGAATCTGCT 1800

TTCAAAACGA AAGCTGAACC TATAAGCAGC AGTATAAGCA ATATGGAACT CGAATTAAAT 1860

ATCCACAATT CAAAAGCACC TAAAAAGAAT AGGCTGAGGA GGAAGTCTTC TACCAGGCAT 1920

ATTCATGCGC TTGAACTAGT AGTCAGTAGA AATCTAAGCC CACCTAATTG TACTGAATTG 1980

CAAATTGATA GTTGTTCTAG CAGTGAAGAG ATAAAGAAAA AAAAGTACAA CCAAATGCCA 2040

GTCAGGCACA GCAGAAACCT ACAACTCATG GAAGGTAAAG AACCTGCAAC TGGAGCCAAG 2100

AAGAGTAACA AGCCAAATGA ACAGACAAGT AAAAGACATG ACAGTGATAC TTTCCCAGAG 2160

CTGAAGTTAA CAAATGCACC TGGTTCTTTT ACTAAGTGTT CAAATACCAG TGAACTTAAA 2220

GAATTTGTCA ATCCTAGCCT TCCAAGAGAA GAAAAAGAAG AGAAACTAGA AACAGTTAAA 2280

GTGTCTAATA ATGCTGAAGA CCCCAAAGAT CTCATGTTAA GTGGAGAAAG GGTTTTGCAA 2340

ACTGAAAGAT CTGTAGAGAG TAGCAGTATT TCACTGGTAC CTGGTACTGA TTATGGCACT 2400

CAGGAAAGTA TCTCGTTACT GGAAGTTAGC ACTCTAGGGA AGGCAAAAAC AGAACCAAAT 2460

AAATGTGTGA GTCAGTGTGC AGCATTTGAA AACCCCAAGG GACTAATTCA TGGTTGTTCC 2520

AAAGATAATA GAAATGACAC AGAAGGCTTT AAGTATCCAT TGGGACATGA AGTTAACCAC 2580

AGTCGGGAAA CAAGCATAGA AATGGAAGAA AGTGAACTTG ATGCTCAGTA TTTGCAGAAT 2640

ACATTCAAGG TTTCAAAGCG CCAGTCATTT GCTCTGTTTT CAAATCCAGG AAATGCAGAA 2700

GAGGAATGTG CAACATTCTC TGCCCACTCT GGGTCCTTAA AGAAACAAAG TCCAAAAGTC 2760

ACTTTTGAAT GTGAACAAAA GGAAGAAAAT CAAGGAAAGA ATGAGTCTAA TATCAAGCCT 2820

GTACAGACAG TTAATATCAC TGCAGGCTTT CCTGTGGTTG GTCAGAAAGA TAAGCCAGTT 2880

GATAATGCCA AATGTAGTAT CAAAGGAGGC TCTAGGTTTT GTCTATCATC TCAGTTCAGA 2940

GGCAACGAAA CTGGACTCAT TACTCCAAAT AAACATGGAC TTTTACAAAA CCCATATCGT 3000

ATACCACCAC TTTTTCCCAT CAAGTCATTT GTTAAAACTA AATGTAAGAA AAATCTGCTA 3060

GAGGAAAACT TTGAGGAACA TTCAATGTCA CCTGAAAGAG AAATGGGAAA TGAGAACATT 3120

CCAAGTACAG TGAGCACAAT TAGCCGTAAT AACATTAGAG AAAATGTTTT TAAAGAAGCC 3180

AGCTCAAGCA ATATTAATGA AGTAGGTTCC AGTACTAATG AAGTGGGCTC CAGTATTAAT 3240

GAAATAGGTT CCAGTGATGA AAACATTCAA GCAGAACTAG GTAGAAACAG AGGGCCAAAA 3300

TTGAATGCTA TGCTTAGATT AGGGGTTTTG CAACCTGAGG TCTATAAACA AAGTCTTCCT 3360

GGAAGTAATT GTAAGCATCC TGAAATAAAA AAGCAAGAAT ATGAAGAAGT AGTTCAGACT 3420

GTTAATACAG ATTTCTCTCC ATCTCTGATT TCAGATAACT TAGAACAGCC TATGGGAAGT 3480

AGTCATGCAT CTCAGGTTTG TTCTGAGACA CCTGATGACC TGTTAGATGA TGGTGAAATA 3540

AAGGAAGATA CTAGTTTTGC TGAAAATGAC ATTAAGGAAA GTTCTGCTGT TTTTAGCAAA 3600

AGCGTCCAGA AAGGAGAGCT TAGCAGGAGT CCTAGCCCTT TCACCCATAC ACATTTGGCT 3660

CAGGGTTACC GAAGAGGGGC CAAGAAATTA GAGTCCTCAG AAGAGAACTT ATCTAGTGAG 3720

GATGAAGAGC TTCCCTGCTT CCAACACTTG TTATTTGGTA AAGTAAACAA TATACCTTCT 3780

CAGTCTACTA GGCATAGCAC CGTTGCTACC GAGTGTCTGT CTAAGAACAC AGAGGAGAAT 3840

TTATTATCAT TGAAGAATAG CTTAAATGAC TGCAGTAACC AGGTAATATT GGCAAAGGCA 3900

TCTCAGGAAC ATCACCTTAG TGAGGAAACA AAATGTTCTG CTAGCTTGTT TTCTTCACAG 3960

TGCAGTGAAT TGGAAGACTT GACTGCAAAT ACAAACACCC AGGCTCCTTT CTTGATTGGT 4020

TCTTCCAAAC AAATGAGGCA TCAGTCTGAA AGCCAGGGAG TTGGTCTGAG TGACAAGGAA 4080

TTGGTTTCAG ATGATGAAGA AAGAGGAACG GGCTTGGAAG AAAATAATCA AGAAGAGCAA 4140

AGCATGGATT CAAACTTAGG TGAAGCAGCA TCTGGGTGTG AGAGTGAAAC AAGCGTCTCT 4200

GAAGACTGCT CAGGGCTATC CTCTCAGAGT GACATTTTAA CCACTCAGCA GAGGGATACC 4260

ATGCAACATA ACCTGATAAA GCTCCAGCAG GAAATGGCTG AACTAGAAGC TGTGTTAGAA 4320

CAGCATGGGA GCCAGCCTTC TAACAGCTAC CCTTCCATCA TAAGTGACTC TTCTGCCCTT 4380

GAGGACCTGC GAAATCCAGA ACAAAGCACA TCAGAAAAAG CAGTATTAAC TTCACAGAAA 4440

AGTAGTGAAT ACCCTATAAG CCAGAATCCA GAAGGCCTNT CTGCTGACAA GTTTGAGGTG 4500

TCTGCAGATA GTTCTACCAG TAAAAATAAA GAACCAGGAG TGGAAAGGTC ATCCCCTTCT 4560

AAATGCCCAT CATTAGATGA TAGGTGGTAC ATGCACAGTT GCTCTGGGAG TCTTCAGAAT 4620

AGAAACTACC CACCTCAAGA GGAGCTCATT AAGGTTGTTG ATGTGGAGGA GCAACAGCTG 4680

GAAGAGTCTG GGCCACACGA TTTGACGGAA ACATCTTACT TGCCAAGGCA AGATCTAGAG 4740

GGAACCCCTT ACCTGGAATC TGGAATCAGC CTCTTCTCTG ATGACCCTGA ATCTGATCCT 4800

TCTGAAGACA GAGCCCCAGA GTCAGCTCGT GTTGGCAACA TACCATCTTC AACCTCTGCA 4860

TTGAAAGTTC CCCAATTGAA AGTTGCAGAA TCTGCCCAGA GTCCAGCTGC TGCTCATACT 4920

ACTGATACTG CTGGGTATAA TGCAATGGAA GAAAGTGTGA GCAGGGAGAA GCCAGAATTG 4980

ACAGCTTCAA CAGAAAGGGT CAACAAAAGA ATGTCCATGG TGGTGTCTGG CCTGACCCCA 5040

GAAGAATTTA TGCTCGTGTA CAAGTTTGCC AGAAAACACC ACATCACTTT AACTAATCTA 5100

ATTACTGAAG AGACTACTCA TGTTGTTATG AAAACAGATG CTGAGTTTGT GTGTGAACGG 5160

ACACTGAAAT ATTTTCTAGG AATTGCGGGA GGAAAATGGG TAGTTAGCTA TTTCTGGGTG 5220

ACCCAGTCTA TTAAAGAAAG AAAAATGCTG AATGAGCATG ATTTTGAAGT CAGAGGAGAT 5280

GTGGTCAATG GAAGAAACCA CCAAGGTCCA AAGCGAGCAA GAGAATCCCA GGACAGAAAG 5340

ATCTTCAGGG GGCTAGAAAT CTGTTGCTAT GGGCCCTTCA CCAACATGCC CACAGATCAA 5400

CTGGAATGGA TGGTACAGCT GTGTGGTGCT TCTGTGGTGA AGGAGCTTTC ATCATTCACC 5460

CTTGGCACAG GTGTCCACCC AATTGTGGTT GTGCAGCCAG ATGCCTGGAC AGAGGACAAT 5520

GGCTTCCATG CAATTGGGCA GATGTGTGAG GCGCCTGTGG TGACCCGAGA GTGGGTGTTG 5580

GACAGTGTAG CACTCTACCA GTGCCAGGAG CTGGACACCT ACCTGATACC CCAGATCCCC 5640

CACAGCCACT ACTGACTGCA GCCAGCCACA GGTACAGAGC CACAGGACCC CAAGAATGAG 5700

CTTACAAAGT GGCCTTTCCA GGCCCTGGGA GCTCCTCTCA CTCTTCAGTC CTTCTACTGT 5760

CCTGGCTACT AAATATTTTA TGTACATCAG CCTGAAAAGG ACTTCTGGCT ATGCAAGGGT 5820

CCCTTAAAGA TTTTCTGCTT GAAGTCTCCC TTGGAAA 5857

(2) INFORMATION FOR SEQ ID NO:2:

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

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

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

(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

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

Ile 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 Asn Ile Glu Gly Gln 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 Xaa Ala Phe Val Ser 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

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 Leu 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
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 Ser 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 Ala 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 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 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

(iv) ANTI-SENSE: NO

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

(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

(iv) ANTI-SENSE: NO

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

(vii) IMMEDIATE SOURCE:
(B) CLONE: s754 B

(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

(iv) ANTI-SENSE: NO

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

(vii) IMMEDIATE SOURCE:
(B) CLONE: s975 A

(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

(iv) ANTI-SENSE: 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

(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

(iv) ANTI-SENSE: 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

(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

(iv) ANTI-SENSE: 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

(iv) ANTI-SENSE: 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

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: 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: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

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

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGKCTTACTC TGTTGTCCCA GCTGGAGTAC AGWGTGCGAT CATGAGGCTT ACTGTTGCTT 60

GACTCCTAGG CTCAAGCGAT CCTATCACCT CAGTCTCCAA GTAGCTGGAC T 111

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

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

(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 Leu
35

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1540 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(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:13

1 GAGGCTAGAG GGCAGGCACT TTATGGCAAA CTCAGGTAGA ATTCTTCCTC

51 TTCCGTCTCT TTCCTTTTAC GTCATCCGGG GGCAGACTGG GTGGCCAATC

101 CAGAGCCCCG AGAGACGCTT GGCTCTTTCT GTCCCTCCCA TCCTCTGATT

151 GTACCTTGAT TTCGTATTCT GAGAGGCTGC TGCTTAGCGG TAGCCCCTTG

```
 201  GTTTCCGTGG CAACGGAAAA GCGCGGGAAT TACAGATAAA TTAAAACTGC
 251  GACTGCGCGG CGTGAGCTCG CTGAGACTTC CTGGACGGGG GACAGGCTGT
 301  GGGGTTTCTC AGATAACTGG GCCCCTGCGC TCAGGAGGCC TTCACCCTCT
 351  GCTCTGGGTA AAGGTAGTAG AGTCCCGGGA AAGGGACAGG GGGCCCAAGT
 401  GATGCTCTGG GGTACTGGCG TGGGAGAGTG GATTTCCGAA GCTGACAGAT
 451  GGGTATTCTT TGACGGGGGG TAGGGGCGGA ACCTGAGAGG CGTAAGGCGT
 501  TGTGAACCCT GGGGAGGGGG GCAGTTTGTA GGTCGCGAGG GAAGCGCTGA
 551  GGATCAGGAA GGGGGCACTG AGTGTCCGTG GGGGAATCCT CGTGATAGGA
 601  ACTGGAATAT GCCTTGAGGG GGACACTATG TCTTTAAAAA CGTCGGCTGG
 651  TCATGAGGTC AGGAGTTCCA GACCAGCCTG ACCAACGTGG TGAAACTCCG
 701  TCTCTACTAA AAATACnAAA ATTAGCCGGG CGTGGTGCCG CTCCAGCTAC
 751  TCAGGAGGCT GAGGCAGGAG AATCGCTAGA ACCCGGGAGG CGGAGGTTGC
 801  AGTGAGCCGA GATCGCGCCA TTGCACTCCA GCCTGGGCGA CAGAGCGAGA
 851  CTGTCTCAAA ACAAAACAAA ACAAAACAAA ACAAAAAACA CCGGCTGGTA
 901  TGTATGAGAG GATGGGACCT TGTGGAAGAA GAGGTGCCAG GAATATGTCT
 951  GGGAAGGGGA GGAGACAGGA TTTTGTGGGA GGGAGAACTT AAGAACTGGA
1001  TCCATTTGCG CCATTGAGAA AGCGCAAGAG GGAAGTAGAG GAGCGTCAGT
1051  AGTAACAGAT GCTGCCGGCA GGGATGTGCT TGAGGAGGAT CCAGAGATGA
1101  GAGCAGGTCA CTGGGAAAGG TTAGGGGCGG GGAGGCCTTG ATTGGTGTTG
1151  GTTTGGTCGT TGTTGATTTT GGTTTTATGC AAGAAAAAGA AAACAACCAG
1201  AAACATTGGR GAAAGCTAAG GCTACCACCA CCTACCCGGT CAGTCACTCC
1251  TCTGTAGCTT TCTCTTYCTT GGAGAAAGGA AAAGACCCAA GGGGTTGGCA
1301  GCRATATGTG AAAAAATTCA GAATTTATGT TGTCTAATTA CAAAAAGCAA
1351  CTTCTAGAAT CTTTAAAAAT AAAGGACGTT GTCATTAGTT CTTTGGTTTG
1401  TATYATTCTA AAACCTTCCA AATCTTAAAT TTACTTTATT TTAAAATGAT
1451  AAAATGAAGT TGTCATTTTA TAAACCTTTT AAAAAGATAT ATATATATGT
1501  TTTTCTAATG TGTTAnAGTT CATTGGAACA GAAAGAAATG
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21781 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(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:14

```
   1 ccttttacgt catcggggag actgggtggc aatcgcagcc cgagagacgc atggctcttt
  61 ctgccctcca tcctctgatg taccttgatt tcgtattctg agaggctgct gcttagcggt
 121 agccccttgg tttccgtggc aacggaaaag cgcgggaatt acagataaat taaaactgcg
 181 actgcgcggc gtgAGCTCGC TGAGACTTCC TGGACCCCGC ACCAGGCTGT GGGGTTTCTC
 241 AGATAACTGG GCCCCTGCGC TCAGGAGGCC TTCACCCTCT GCTCTGGGTA AAGvvvvvaa
 301 aaataaagga cgctgtcgay tagttcttct ggtttgtatt attctmaaac cttccaaatc
 361 ttcaaattta cttyatttta aaatgatgaa aatgaagttg tcattttata aacctttaa
 421 aaagatatat atatatgttt ttctaatgtg tTAAAGTTCA TTGGAACAGA AAGAAATGGA
 481 TTTATCTGCT CTTCGCGTTG AAGAAGTACA AAATGTCATT AATGCTATGC AGAAAATCTT
 541 AGAGTGTCCC ATCTGgtaag tcagcacaag agtgtattaa tttgggattc ctatgattat
 601 ctcctatgca aatgaacaga attgaccttа catactaggg aagaaaagac atgtctagta
 661 agattaggct attgtaattg ctgattttct taactgaaga actttaaaaa tatagaaaat
 721 gattccttgt tctccatcca ctctgcctct cccactcctc tccttttcaa cacatcctgt
 781 ggtccgggaa agcagggctc tgtcttgatt ggttctgcac tgggcaggat ctgttagata
 841 ctgcatttgc tttctccagc tctaaavvvv vvvvvvvvaa atgctgatga tagtatagag
 901 tattgaaggg atcaatataa ttctgttttg atatctgaaa gctcactgaa ggtaaggatc
 961 gtattctctg ctgtattctc agttcctgac acagcagaca tttaataaat attgaacgaa
1021 cttgaggcct tatgttgact cagtcataac agctcamagt tgaacttatt cactaagaat
1081 agctttattt ytaaataaat tattgagcct catttatttt ctttttctcv vvvvTCTGGA
1141 GTTGATCAAG GAACCTGTCT CCACAAAGTG TGACCACATA TTTTGCAAvv vvvatgtggc
1201 tccattatta gcttttgttt ttgtccttca taacccagga aacacctaac tttatagaag
1261 ctttactttc ttcaattaag tgagaacgaa aatccaactc catttcattc tttctcagag
1321 agtatatagt tatcaaaagt tggttgtaat catagttcct ggtaaagttt tgacatatat
1381 tatctttttt ttttttgag acaagtctcg ctctgtcgcc caggctggag tgcagtggca
1441 tgaggcttgc tcactgcacc tccgcccccg agttcagcga ctctvvvvvv vvvvvvvtga
1501 gatctagacc acatggtcaa agagatagaa tgtgagcaat aaatgaacct taaatttttc
1561 aacagctact ttttttttt tttttgaga cagGGkCTTA CTCTGTTGTC CCAGCTGGAG
1621 TACAGw-G-T GCGATCATGA GGCTTACTGT TGC-TTGAC- TCCTA-GGCT CAAGCGATCC
1681 TATCACCTCA GTCTC-CAAG -TAGCT-GGA CTgtaagtgc acaccaccat atccagctaa
1741 attttgtgtt ttctgtagag acggggtttc gccatgtttc ccaggctggt cttgaacttt
```

1801 gggcttaacc cgtctgccca cctaggcatc ccaaagtgct aggattacag gtgtgagtca
1861 tcatgcctgg ccagtatttt agttagctct gtcttttcaa gtcatataca agttcatttt
1921 ctttaagtta gttaacaacc ttatatcatg tattcttttc tagcavvvvv vvvvvvvvtg
1981 tgatcataac agtaagccat atgcatgtaa gttcagtttc atagatcatt gcttatgtag
2041 tttaggtttt tgcttatgca gcatccaaaa acaattagga aactattgct tgtaattcac
2101 ctgccattac tttttaaatg gctcttaagg gcagttgtga gattatcttt tcatggctat
2161 ttgccttttg agtattcttt ctacaaaagg aagtaaatta aattgttctt tctttcttta
2221 taatttatag atvvvvvATT TTGCATGCTG AAACTTCTCA ACCAGAAGAA AGGGCCTTCA
2281 CAGTGTCCTT TATGTAAGAA TGATATAACC AAAAGgatat aannngaant gatgctaggt
2341 tggaagcaac cacagtagga aaagtagaaa ttatttaata acatagcgtt cctataaaac
2401 cattcatcag aaaaatttat aaaagagttt ttagcacaca gtaaattatt tccaaagtta
2461 ttttcctgaa agttttatgg gcatctgcct tatacaggtn ttgvvvvvvv vvvvvggtaa
2521 ggcttaaatg aatgacaaaa agttactaaa tcactgccat cacacggtta tatacagatg
2581 tcaatgatgt attgatttag aggttttcta ctgttgctgc atcttatttt tatttgttta
2641 catgtctttt cttattttag tgtccttaaa aggttgataa tcacttgctg agtgtgtttc
2701 tcaaavvvvv GAGCCTACAA GAAAGTACGA GATTTAGTCA ACTTGTTGAA GAGCTATTGA
2761 AAATCATTTG TGCTTTTCAG CTTGACACAG GTTTGGAGTA TGCAAACAGC TATAATTTTG
2821 CAAAAAAGGA AAATAACTCT CCTGAACATC TAAAAGATGA AGTTTCTATC ATCCAAAGTA
2881 TGGGCTACAG AAACCGTGCC AAAAGACTTC TACAGAGTGA ACCCGAAAAT CCTTCCTTGC
2941 AGaaaccatt tgttttcttc ttcttcttct tcttcttttc ttttttttt ctttttttt
3001 tcwtgagagt ggagtcyttg gctcgtssyg gcgccgaggc taggaagcag ctcctcctgc
3061 gctcgtacgc cacgctagta ggtagcgtgg gcakacacgg cacgcgcacc atggccaggc
3121 taatttttgt attgttvvvv vvvvvvvvva tgatggagat cttaaaaagt aatcattctg
3181 gggctgggcg tagtagcttg cacctgtaat cccagcactt cgggaggctg aggcaggcag
3241 ataatttgag gtcaggagtt tgagaccagc ctggccaaca tggtgaaacc catctctact
3301 aaaaatacaa aaattagctg ggtgtggtgg cacgtacctg taatcccagc tactcgggag
3361 gcggaggcac aagaattgct tgaacctagg acgcggaggt tgcagcgagc caagatcgcg
3421 ccactgcact ccagcctggg ccgtagagtg agactctgtc tcaaaaaaga aaaaaaagta
3481 attgttctag ctgggcgcag tggctcttgc ctgtaatccc agcactttgg gaggccaagg
3541 cgggtggatc tcgagtccta gagttcaaga ccagcctagg caatgtggtg aaaccccatc
3601 gctacaaaaa atacaaaaat tagccaggca tggtggcgtg cgcatgtagt cccagctcct
3661 tgggaggctg aggtgggagg atcacttgaa cccaggagac agaggttgca gtgaaccgag
3721 atcacgccac cacgctccag cctgggcaac agaacaagac tctgtctaaa aaaatacaaa
3781 taaaataaaa gtagttctca cagtaccagc attcattttt caaaagatat agagctaaaa

3841 aggaaggaaa aaaaaagtaa tgttgggctt ttaaatactc gttcctatac taaatgttct
3901 taggagtgct ggggttttat tgtcatcatt tatccttttt aaaaatgtta ttggccaggc
3961 acggtggctc atggctgtaa tcccagcact ttgggaggcc gaggcaggca gatcacctga
4021 ggtcaggagt gtgagaccag cctggccaac atggcgaaac ctgtctctac taaaaataca
4081 aaaattaact aggcgtggtg gtgtacgcct gtagtcccaG ctactcggga ggctgaggca
4141 ggagaatcaa ctgaaccagg gaggtggagg ttgcagtgtg ccgagatcac gccactgcac
4201 tctagcctgg caacagagca agattctgtc tcaaaaaaaa aaaacatata tacacatata
4261 tcccaaagtg ctgggattac atatatatat atatatatat atTatatata tatatatata
4321 tatatgtgat atatatgtga tatatatata acatatatat atgtAatata tatgtgatat
4381 atatataata tatatatgtA atatatatgt gatatatata tatacacaca cacacacata
4441 tatatgtAtg tgtgtgtAca Cacacacaca caaattagcc aggcatagtt gcacacgctt
4501 ggtagaccCa gctactcagg aggCtgaggG aggagaATct cttgaActta ggaggcggag
4561 gttgcagtga gctgagattg cgccactgca ctccagcctg ggtgacagag caggactctg
4621 tacacccccc aaaacaaaaa aaaaagttat cagatgtgat tggaatgtat atcaagtatc
4681 agcttcaaaa tatgctatat taatacttca aaaattacac aaataataca taatcaggtt
4741 tgaaaaattt aagacaacms aaraaaaaaw ycmaatcaca matatcccac acattttatt
4801 attmctmctm cwattatttt gwagagmctg ggtctcacyc ykttgctwat gctggtcttt
4861 gaacyccykg ccycaarcar tcctsctcca bcctcccaar gtgctgggga twataggcat
4921 garctacccg ctcccagccc magacaattt tagtgtgwgc aaaattcctg ggcatttttt
4981 cmaggchdnm wayayvtwds mtgcwhrtga hbggctgatg wgtcaattta ttttgyccat
5041 ggtgtcaakt ytctcttcag gaggaaaagc acamgaactg gccaayaatt gcttgactgt
5101 tctttaccaw actgttwarc agGAAACCAr TCTCAGTGTC CCAACTCTCT AACCTTGGAA
5161 CTTGTGAGAA CTCTGAGGAC AAAGCAGCGG ATACAACCTC AAAAGACGTC TGTCTACATT
5221 GAATTGggta agggtctcag gtttttttaag tatttaataa taattgctgg attccttatc
5281 ttatagtttt gccamaaatc ttggtcataa tttgtatttg tggtaggcag cwttgggaag
5341 tgaattttat gagccctatg gtgagttata aaaaatgtaa aagacgcagt tcccaccttg
5401 aagaatctta ctttaaaaag ggagcaaaag aggccaggca tggtggctca cacctgtaat
5461 cccagcactt tgggaggcca aagtgggtgg atcacctgag gtcgggagtt cgagaccagc
5521 ctagccaaca tggagaaact ctgtctgtac caaaaaataa aaaattagcc aggtgtggtg
5581 gcacataact gtaatcccag ctactcggga ggctgaggca ggagaatcac ttgaaCccgg
5641 gaggtggagg ttgcggtgaa ccgagatcgc accattgcac tccagcctgg gcaaaaataG
5701 cgaaactcca tctaaaaaaa aaaaagagag caaaagaaag amtmtctggt tttaamtmtg
5761 tgtaaatatg ttttTggaaa gatggagagt agcaataaga aaaaacatga tggattgcta
5821 cagtatttag ttccaagata aattgtacta gatgaggaag ccttttaaga agagctgaat

5881 tgccaggcgc agtggctcac gcctgtaatc ccagcacttt gggaggccga ggtgggcgga
5941 tcacctgagg tcgggagttc aagaccagcc tgaccaacat ggagaaaccc catctctact
6001 aaaaaaaaaa aaaaaaaaat tagccggggt ggtggcttat gcctgtaatc ccagctactc
6061 aggaggctga ggcaggagaa tcgcttgaac ccaggaagca gaggttgcag tgagccaaga
6121 tcgcaccatt gcactccagc ctaggcaaca agagtgaaac tccatctcaa aaaaaaaaaa
6181 aAAgagctga Atcttggctg ggcaggatgg ctcgtgcctg taatcctaac gctttggaag
6241 accgaggcag aaggattggt tgagtccacg agtttaagac cagcctggcc aacatagggg
6301 aaccctgtct ctatttttaa aataataata catttttggc cggtgcggtg gctcatgcct
6361 gtaatcccaa tactttggga ggctgaggca ggtagatcac ctgaggtcag agttcgagac
6421 cagcctggat aacctggtga aacccctctt tactaaaaat acaaaaaaaa aaaaaaatta
6481 gctgggtgtg gtagcacatg cttgtaatcc cagctacttg ggaggctgag gcaggagaat
6541 cgcttgaacc agggaggcgg aggttacaat gagccaacac tacaccactg cactccagcc
6601 tgggcaatag agtgagactg catctcaaaa aaataataat ttttaaaaat aataaatttt
6661 tttaagctta taaaaagaaa agttgaggcc agcatagtag ctcacatctg taatctcagc
6721 agtggcagag gattgcttga agccaggagt ttgagaccag cctgggcaac atagcaagac
6781 ctcatctcta caaaaaaatt tctttttaa attagctggg tgtggtggtg tgcatctgta
6841 gtcccagcta ctcaggaggc agaggtgagt ggatacattg aacccaggag tttgaggctg
6901 tagtgagcta tgatcatgcc actgcactcc aacctgggtg acagagcaag acctccaaaa
6961 aaaaaaaaaa aagagctgct gagctcagaa ttcaaactgg gctctcaaat tggattttct
7021 tttagaatat atttataatt aaaaaggata gccatctttt gagctcccag gcaccaccat
7081 ctatttatca taacacttac tgttttcccc ccttatgatc ataaattcct agacaacagg
7141 cattgtaaaa atagttatag tagttgatat ttaggagcac ttaactatat tccaggcact
7201 attgtgcttt tcttgtataa ctcattagat gcttgtcaga cctctgagat tgttcctatt
7261 atacttattt tacagatgag aaaattaagg cacagagaag ttatgaaatt tttccaaggt
7321 attaaaccta gtaagtggct gagccatgat tcaaacctag gaagttagat gtcagagcct
7381 gtgctttttt tttgtttttg tttttgtttt cagTagaaac gggggtctca ctttgttggc
7441 caggctggtc ttgaactcct aacctcaaat aatccaccca tctcggcctc ctcaagtgct
7501 gggattacag gtgagagcca ctgtgcctgg cgaagcccat gcctttaacc acttctctgt
7561 attacatact agcttaacta gcattgtacc tgccacagta gatgctcagt aaatatttct
7621 agttgaatat ctgtttttca acaagtacat ttttttaacc cttttaatta agaaaacttt
7681 tattgattta ttttttgggg ggaaattttt taGGATCTGA TTCTTCTGAA GATACCGTTA
7741 ATAAGGCAAC TTATTGCAgg tgagtcaaag agaacctttg tctatgaagc tggtattttc
7801 ctatttagtt aatattaagg attgatgttt ctctcttttt aaaaatattt taacttttat
7861 tttaggttca gggatgtatg tgcagtttgt tatataggta aacacacgac ttgggatttg

7921 gtgtatagat ttttttcatc atccgggtac taagcatacc ccacagtttt ttgtttgctt
7981 tctttctgaa tttctccctc ttcccacctt cctccctcaa gtaggctggt gtttctccag
8041 actagaatca tggtattgga agaaaccttа gagatcatct agtttagttc tctcatttta
8101 tagtggagga aatacccttt ttgtttgttg gatttagtta ttagcactgt ccaaaggaat
8161 ttaggataac agtagaactc tgcacatgct tgcttctagc agattgttct ctaagttcct
8221 catatacagt aatattgaca cagcagtaat tgtgactgat gaaaatgttc aaggacttca
8281 ttttcaactc tttctttcct ctgttcctta tttccacata tctctcaagc tttgtctgta
8341 tgttatataa taaactacaa gcaaccccaa ctatgttacc taccttcctt aggaattatt
8401 gcttgaccca ggtttttttt ttttttttтt tggagacggg gtcttgccct gttgccagga
8461 tggagtgtag tggcgccatc tcggctcact gcaatctcca actccctggt tcaagcgatt
8521 ctcctgtctc aatctcacga gtagctggga ctacaggtat acaccaccac gcccggttaa
8581 ttgaccattc catttctttc tttctctctt ttttttttтt ttttttgaga cagagtcttg
8641 ctctgttgcc caggctggag tacagaggtg tgatctcacc tctccgcaac gtctgcctcc
8701 caggttgaag ccatactcct gcctcagcct ctctagtagc tgggactaca ggcgcgcgcc
8761 accacacccg gctaattttt gtatttttag tagagatggg gtttcaccat gttggccagg
8821 ctggtcttga actcatgacc tcaagtggtc cacccgcctc agcctcccaa agtgctggaa
8881 ttacaggctt gagccaccgt gcccagcaac catttcattt caactagaag tttctaaagg
8941 agagagcagc tttcactaac taaataagat tggtcagctt tctgtaatcg aaagagctaa
9001 aatgtttgat cttggtcatt tgacagttct gcatacatgt aactagtgtt tcttattagg
9061 actctgtctt ttccctataG TGTGGGAGAT CAAGAATTGT TACAAATCAC CCCTCAAGGA
9121 ACCAGGGATG AAATCAGTTT GGATTCTGCA AAAAAGggta atggcaaagt ttgccaactt
9181 aacaggcact gaaaagagag tgggtagata cagtactgta attagattat tctgaagacc
9241 atttgggacc tttacaaccc acaaaatctc ttggcagagt tagagtatca ttctctgtca
9301 aatgtcgtgg tatggtctga tagatttaaa tggtactaga ctaatgtacc tataataaga
9361 ccttctTgta actgattgtt gccctttcgc ttttttttтt gtttgtttgt ttgttttttt
9421 ttgagatggg gtctcactct gttgcccagg ctggagtgca gtgatgcaat cttggctcac
9481 tgcaacctcc acctccaaAg gctcaagcta tcctcccact tcagcctcct gagtagctgg
9541 gactacaggc gcatgccacc acaCccggTt aatttttтgt ggttttatag agatggggtt
9601 tcaccatgtt accgaggctg gtctcaaact cctggactca agcagtctgc ccacttcagc
9661 ctcccaaagt gctgcagtta caggcttgag ccactgtgcc tggcctgccc tttacttttа
9721 attggtgtat ttgtgtttca tcttttacct actggttttt aaatatаggg agtggtaagt
9781 ctgtagatag aacagagtat taagtagact taatggccag taatctttag agtacatcag
9841 aaccagtttt ctgatggcca atctgctttt aattcactct tagacgttag agaaataggt
9901 gtggtttctg catagggaaa attctgaaat taavvvvvvv vvvvvvgatc ctaagtggaa

```
 9961 ataatctagg taaataggaa ttaaatgaaa gagtatgagc tacatcttca gtatacttgg
10021 tagtttatga ggttagtttc tctaatatag ccagttggTt gatttccacc tccaaggtgt
10081 atgaagtatg tatttttta atgacaattc agtttktgag taccttgtta tttttgtata
10141 ttttcaGCTG CTTGTGAATT TTCTGAGACG GATGTAACAA ATACTGAACA TCATCAACCC
10201 AGTAATAA-T GATTTGAACA CCACTGAGAA GCGTGCAGCT GAGAGGCATC CAGAAAAGTA
10261 TCAGGGTAGT TCTGTTTCAA ACTTGCATGT GGAGCCATGT GGCACAAATT ACTCATGCCA
10321 GCTCATTACA GCATGAGAAC AGCAGTTTAT TACTCACTAA AGACAGAATG AATGTAGAAA
10381 AGGCTGAATT CTGTAATAAA AGCAAACAGC CTGGCTTAGC AAGGAGCCAA CATAACAGAT
10441 GGGCTGGAAG TAAGGAAACA TGTAATGATA GGCGGACTCC CAGCACAGAA AAAAAGGTAG
10501 ATCTGAATGC TGATCCCCTG TGTGAGAGAA AAGAATGGAA TAAGCAGAAA CTGCCATGCT
10561 CAGAGAATCC TAGAGATACT GAAGATGTTC CTTGGATAAC ACTAAATAGC AGCATTCAGA
10621 AAGTTAATGA GTGGTTTTCC AGAAGTGATG AACTGTTAGG TTCTGATGAC TCACATGATG
10681 GGGAGTCTGA ATCAAATGCC AAAGTAGCTG ATGTATTGGA CGTCCTAATT GAGGTAGATG
10741 AATATTCTGG TTCTTCAGAG AAAATAGACT TACTGGCCAG TGATCCTCAT GAGGCTTTAA
10801 TATGTAAAAG TGACAGAGTT CACTCCAAAT CAGTAGAGAG TAATATTGAA GGCCAAA-TA
10861 TTTGGGAAAA CCTATCGGAA GAAGGCAAGC CTCCCCAACT TAAGCCATGT AACTGAAAAT
10921 CTAATTATAG GnGCATTTGT TTCTGAGCCA CAGATAATAC AAGAGCGTCC CCTCACAAAT
10981 AAATTAAAGC GTAAAAGGAG ACCTACATCA GGCCTTCATC CTGAGGATTT TATCAAGAAA
11041 GCAGATTTGG CAGTTCAAAA GACTCCTGAA ATGATAAATC AGGGAACTAA CCAAACGGAG
11101 CAGAAtggtc aagtgaTGAA TATTACTAAT AGTGG-TCA- TGAGAATAAA ACAAAAGGTG
11161 ATTCTATTCA GAATGAGAAA AATCCTAACC CAATAGAATC ACTCGAAAAA GAATCTGCTT
11221 TCAAAACGAA AGCTGAACCT ATAAGCAGCA GTATAAGCAA TATGGAACTC GAATTAAATA
11281 TCCACAATTC AAAAGCACCT AAAAAGAATA GGCTGAGGAG GAAGTCTTCT ACCAGGCATA
11341 TTCATGCGCT TGAACTAGTA GTCAGTAGAA ATCTAAGCCC ACCTAATTGT ACTGAATTGC
11401 AAATTGATAG TTGTTCTAGC AGTGAAGAGA TAAAGAAAAA AAAGTACAAC CAAATGCCAG
11461 TCAGGCACAG CAGAAACCTA CAACTCATGG AAGGTAAAGA ACCTGCAACT GGAGCCAAGA
11521 AGAGTAACAA GCCAAATGAA CAGACAAGTA AAAGACATGA CAGTGATACT TTCCCAGAGC
11581 TGAAGTTAAC AAATGCACCT GGTTCTTTTA CTAAGTGTTC AAATACCAGT GAACTTAAAG
11641 AATTTGTCAA TCCTAGCCTT CCAAGAGAAG AAAAAGAAGA GAAACTAGAA ACAGTTAAAG
11701 TGTCTAATAA TGCTGAAGAC CCCAAAGATC TCATGTTAAG TGGAGAAAGG GTTTTGCAAA
11761 CTGAAAGATC TGTAGAGAGT AGCAGTATTT CACTGGTACC TGGTACTGAT TATGGCACTC
11821 AGGAAAGTAT CTCGTTACTG GAAGTTAGCA CTCTAGGGAA GGCAAAAACA GAACCAAATA
11881 AATGTGTGAG TCAGTGTGCA GCATTTGAAA ACCCCAAGGG ACTAATTCAT GGTTGTTCCA
11941 AAGATAATAG AAATGACACA GAAGGCTTTA AGTATCCATT GGGACATGAA GTTAACCACA
```

12001 GTCGGGAAAC AAGCATAGAA ATGGAAGAAA GTGAACTTGA TGCTCAGTAT TTGCAGAATA
12061 CATTCAAGGT TTCAAAGCGC CAGTCATTTG CTCTGTTTTC AAATCCAGGA AATGCAGAAG
12121 AGGAATGTGC AACATTCTCT GCCCACTCTG GGTCCTTAAA GAAACAAAGT CCAAAAGTCA
12181 CTTTTGAATG TGAACAAAAG GAAGAAAATC AAGGAAAGAA TGAGTCTAAT ATCAAGCCTG
12241 TACAGACAGT TAATATCACT GCAGGCTTTC CTGTGGTTGG TCAGAAAGAT AAGCCAGTTG
12301 ATAATGCCAA ATGTAGTATC AAAGGAGGCT CTAGGTTTTG TCTATCATCT CAGTTCAGAG
12361 GCAACGAAAC TGGACTCATT ACTCCAAATA AACATGGACT TTTACAAAAC CCATATCGTA
12421 TACCACCACT TTTTCCCATC AAGTCATTTG TTAAAACTAA ATGTAAGAAA AATCTGCTAG
12481 AGGAAAACTT TGAGGAACAT TCAATGTCAC CTGAAAGAGA AATGGGAAAT GAGAACATTC
12541 CAAGTACAGT GAGCACAATT AGCCGTAATA ACATTAGAGA AAATGTTTTT AAAGAAGCCA
12601 GCTCAAGCAA TATTAATGAA GTAGGTTCCA GTACTAATGA AGTGGGCTCC AGTATTAATG
12661 AAATAGGTTC CAGTGATGAA AACATTCAAG CAGAACTAGG TAGAAACAGA GGGCCAAAAT
12721 TGAATGCTAT GCTTAGATTA GGGGTTTTGC AACCTGAGGT CTATAAACAA AGTCTTCCTG
12781 GAAGTAATTG TAAGCATCCT GAAATAAAAA AGCAAGAATA TGAAGAAGTA GTTCAGACTG
12841 TTAATACAGA TTTCTCTCCA TCTCTGATTT CAGATAACTT AGAACAGCCT ATGGGAAGTA
12901 GTCATGCATC TCAGGTTTGT TCTGAGACAC CTGATGACCT GTTAGATGAT GGTGAAATAA
12961 AGGAAGATAC TAGTTTTGCT GAnAATGACA TTAAGGAAAG TTCTGCTGTT TTTAGCAAAA
13021 GCGTCCAGAG GGGAGAGCTT AGCAGGnGTC CTAGCCCTTT CACCCATACA CATTTGGCTC
13081 AGGGTTACCG AAGAGGGGCC AAGAAATTAG AGTCCTCAGA AGAGAACTTA TCTAGTGAGG
13141 ATGAAGAGCT TCCCTGCTTC CAACACTTGT TATTTGGTAA AGTAAACAAT ATACCTTCTC
13201 AGTCTACTAG GCATAGCACC GTTGCTACCG AGTGTCTGTC TAAGAACACA GAGGAGAATT
13261 TATTATCATT GAAGAATAGC TTAAATGACT GCAGTAACCA GGTAATATTG GCAAAGGCAT
13321 CTCAGGAACA TCACCTTAGT GAGGAAACAA AATGTTCTGC TAGCTTGTTT TCTTCACAGT
13381 GCAGTGAATT GGAAGACTTG ACTGCAAATA CAAACACCCA GGCTCCTTTC TTGATTGGTT
13441 CTTCCAAACA AATGAGGCAT CAGTCTGAAA GCCAGGGAGT TGGTCTGAGT GACAAGGAAT
13501 TGGTTTCAGA TGATGAAGAA AGAGGAACGG GCTTGGAAGA AAATAATCAA GAAGAGCAAA
13561 GCATGGATTC AAACTTAGgt attggaacca ggttt--gtg tgtgtggccc agtctattta
13621 tagaagtgag ctaaatgttt atgcttttgg ggagcacatt ttacaaattt ccaagtatag
13681 ttaaaggaac tgcttcttaa acttganaca tgttcctcct aaggtggttn ncatagaaaa
13741 aagtccttca cacagctagg acgtcatctt tggcttgaat gagctttaac atcctantta
13801 ctggtggact tacttctggt ttcattttat taaagaaatc caggtgtccc aagcaaggat
13861 ttatcactgg gtgaaagaaa gtaatccagc cccgcantga gaggaaagca nagaagtgga
13921 gngnttggct tgttncactg acccnctttg gnnttaggga gagactgtcc vvvvvvvvvv
13981 vvvggtgtcc caaagcaagg aatttaatca ttttgtgtga catgaaagta natccagtcc

14041 tgccaatgag aagaaaaagc acagcaagnn ngcagcgttt atagtctgct tttacatctg
14101 aacctctgtt tttgttattt aagGT-AAGG TGAAGCAGCA TCTGGGTGTG AGAGTGAAAC
14161 AAGCGTCTCT -GAAGACTGC TCAGGGCTAT CCTCTCAGAG TGACATTTTA -ACCACTCAG
14221 CAGAGvvvvv vvvvvvvvng ngaatgtaat cctaatattt cncnccna-c tta-aaagaa
14281 -tacc-actc caanggcatc ncaatacatc aatcaattgg gg-aattggg atttt--ccc
14341 tcnctaacat ca-ntggaat aatttcatgg c-attaattg catgaatgtg gttagattaa
14401 aaggtgttca tgcta-gaac ttgtagttcc atacta-ggt -atttcaat- cnttgttcct
14461 caaan-aatt tgtatgatat attntcattt aatggaaa-g cttctcaaag tatttc-att
14521 tt-cttggta ccatnnatcg tttttgaatn ag-GG-AT-A CCATGCAACA TAACCTGATA
14581 AAGCTCCAGC AGG-AAATGG CTGAACTAGA AGCTGTGTTA GAACAGCATG GGAGCCAGCC
14641 TTCTAACAGC TACCCTTCCA TCATAAGTGA CTCTTCTGCC CTTGAGGACC TGCGAAATCC
14701 AGAACA-A-A GCACATCAGA -AAAAGGtgt gtattgttgg ccaaacactg atatcttaan
14761 caaaattctt tccttcccct ttatctcctt ctgaagagta aggacctagc tccaacattt
14821 tatgatcctt gctcagcaca tgggtaatta tggagccttg gttcttgtcc ctgctcacaa
14881 ctaatatacc agtcagaggg acccaaggca gtcattcatg ttgtcatctg agavvvvvvv
14941 vvvvvvaaat tctgcctgat atacttgttn caaaaccaat ttgtgtatca tagattgatg
15001 cttttgaana aaatcagtat tctaacctga attatcacta tcagaacaaa gcagtaaagt
15061 agatttgttt tctcattcca tttaaagcag TATTAACTTC ACAGA-AAAG TAGTGAATAC
15121 CCTATAAGCC AGAATCCAGA AGGCCTnTCT GCTGACAAGT TTGAGGTGTC TGCAGATAGT
15181 TCTACCAGTA AAAATAAAGA ACCAGGAGTG GAAgtaagaa acatcaatgt aaagatgctg
15241 tggtatctga catctttatt tatattgaac tctgattgtt aatttttttc accatacttt
15301 ctccagtttt tttgcataca ggcatttata cacttttatt gctctaggat acttcttttg
15361 tttaatccta tataggvvvv vvvvvvvvvg gataagntca agagatattt tgataggtga
15421 tgcagtgatn aattgngaaa atttnctgcc tgcttttaat cttcccccgt tctttcttcc
15481 tncctcccte ccttcctncc tcccgtcctt ncctttcctt tccctccctt ccnccttctt
15541 tccntctntc tttcctttct ttcctgtcta cctttctttc cttcctccct tcctttttctt
15601 ttctttcttt cctttccttt tctttccttt ctttccttte ctttctttct tgacagagtc
15661 ttgctctgtc actcaggctg gagtgcagtg gcgtgatctc gnctc-actg caacctct--
15721 gtc-tc-cca ggt-tcaagc aattttcctg cctcagcctc ccgagtagct gagatta-ca
15781 ggcgccagcc -ac-cac-ac cc-ag-ctac tgacctgctt ttvvvaaaca gctgggagat
15841 atggtgcctc agaccatccc catgttatat gtcaaccctg acatattggc aggcaacatg
15901 aatccagact tctaggctgt catgcgggct ctttctttgc cagtcattnc tgatctctct
15961 gacatgaact gtntcaggta tgctttggct gcccagcaag tatgaaatgt cctntcacaa
16021 ggggtggcga tggttttctc ctt-ccattt atctttctAG GTCATCCCCT TCTAAATGCC

16081 CATCA-TTA- GATGATAGGT GGTACATG-C ACAGTTGCTC TGGGAGTCTT CAGAA-TAGA
16141 AACTACCCA- CCTCAAGAGG AGCT-CATTA AGGTTGTTGA TGTGGAGGAG CAACAGCTGG
16201 AAGAGTCTGG GCCACACGAT TTGACGGAAA CATCTTACTT GCCAAGGCAA GATCTAGgta
16261 atatttcatc tgctgtattg garcamacac tytgatttta ctctgaatcc tacataaaga
16321 tatt-ctggt tarccaactt -ttagat-gt rctagtctat -catggacac ttttgt-tat
16381 acttaattaa gcccacttta gaaaaatagc tcaagtgtta atcwaggttt acttgwaaat
16441 tattgaaact gttaatccat ctatatttta attaatggtt taactaatga ttttgaggat
16501 gwgggagtck tggtgtactc tamatgtatt atttcaggcc aggcatagtg gctcacgcct
16561 ggtaatccca gtayycmrga gcccgaggca ggtggagcca gctgaggtca ggagttcaag
16621 acctgtcttg gccaacatgg gngaaaccct gtcttcttct taaaaaanac aaaaaaaatt
16681 aactgggttg tgcttaggtg natgccccgn atcctagttn ttcttgnggg ttgagggagg
16741 agatcacntt ggaccccgga ggggngggtg ggggngagca ggncaaaaca cngacccagc
16801 tggggtggaa gggaagccca ctcnaaaaaa nnttnvvvvv vvvvvvvvAG GGAACCCCTT
16861 ACCTGGAATC TGGAATCAGC CTCTT-CTCT GATGACCC-T GAATCTGATC C-TTCTGAAG
16921 ACA-GAGCCC C-A-GAGTCA -GCTCGTGTT GGC-AACA-T ACCATCTTC- AACCTCTGCA
16981 TTGAAAGTT- CCCCAATTGA AA-GTTGCAG AATCTGCCCA GAGTCCAGCT GCTGCTCA-T
17041 ACTACTGATA CTGCTGGGTA TAATGCAATG GAAGAAAGTG TGAGCAGGGA GAAGCCAGAA
17101 TTGACAGCTT CAACAGAAAG GGTCAACAAA AGAATGTCCA TGGTGG-TGT CTGGCCTGAC
17161 CCCAGAAGAA TTTgtgagtg tatccatatg tatctcccta atgactaaga cttaacaaca
17221 ttctggaaag agttttatgt aggtattgtc aattaataac ctagaggaag aaatctagaa
17281 ancaatcaca gttctgtgta atttaatttc gattactaat ttctgaaaat ttagaayvvv
17341 vvvvvvvvvv ncccnncccc cnaatctgaa atgggggtaa ccccccccca accganacnt
17401 gggtn-gcnt agaga-nttt -aa-tggccc nt-tctgagg -nacanaagc ttaagc-cag
17461 gng-acgtgg ancnat-gng ttgtttnttg tt-tggttac ctccagcctg gtgacaganc
17521 aagctctgtc tngnaaaant aggngagggn naagattcac nntaaatagt tccaggacac
17581 gtgtagaacg tgcaggattg ctacgtaggt aaacatatgc catggtggga taactagtat
17641 tctgagctgt gtgctagagg taactcatga taatggaata tttgatttaa tttcagATGC
17701 TCGTGTACAA GTTTGCCAGA AAACACCACA TCACTTTAAC TAATCTAATT ACTGAAGAGA
17761 CTACTCATGT TGTTATGAAA ACAGAtatac ccaagaacct ttacagaata ccttkcatct
17821 sctgcataaa accacatgag gcgaggcaca gtggcgcatg cctgtawtcg caRcactttg
17881 ggaggccgag gcgggcagat accccgagRt tMSgagtcga gccatcctgg ccagcatggt
17941 gaaaccccgt ctctactaNN aaatggNaaa attaNctggg tgtggtcgcg tgcNcctgta
18001 gtcccagcta ctcgtgaggc tgaggcagga gaatcacttg aaccggggaa atggaggttt
18061 cagtgagcag agatcatNcc cctNcattcc agcctggcga cagagcaagg ctccgtcNcc

18121 Naaaaaataa aaaaaaacgt gaacaaataa gaatatttgt tgagcatagc atggatgata
18181 gtcttctaat agtcaatcaa ttactttatg aaagacaaat aatagttttg ctgcttcctt
18241 acctcctttt gttttgggtt aagatttgga gtgtgggcca ggcacvvvvv vvvvvvvvga
18301 tctat-agct agccttggCg tctagaagat gggtgttgag aagagggagt ggaaagatat
18361 ttcctctggt cttaacttca tatcagcctc ccctagactt ccaaatatcc atacctgctg
18421 gttataatta gtggtgtttt cagcctctga ttctgtcacc aggggtttta gaatcataaa
18481 tccagattga tcttgggagt gtaaaaaact gaggctcttt agcttcttag gacAgcactt
18541 cctgattttg ttttcaactt ctaatccttt gagtgTktct cattctgcag aTGCTGAGTT
18601 TGTGTGTGAA CGGACACTGA AATATTTTCT AGGAATTGCG GGAGGAAAAT GGGTAGTTAG
18661 CTATTTCTGG kaagkataat actatwtctc cccycctccc CtAacamcTc agaattgsat
18721 tttacaccta acatttacam ctaAgnyyyt kgctgatgct tgagtctgag tcmcmaaagk
18781 cctttaatTg Taatactaac tactTttatc ttTaatatca ctttgTtcag AtaagCtggt
18841 gaTGCtggga aaatggtctc ttttataact aataggacct aAtctgctcc tagcaaTgtt
18901 agca-tatga gctagggatt tatttaatag tcggcaggaa tccatgtgca rcagncaaac
18961 ttataatgtt taaattaaac atcaactctg tctccagaag gaaactgctg ctacaagcct
19021 tattaaaggg ctgtggcttt agagggaagg acctctcctc tgtcattctt cctgtgctct
19081 tttgtgaatc gctgacctct ctatctccgt gaaaagagca cgttcttctg ctgtatgtaa
19141 cctgtctttt ctatgatctc tvvvvvvvvv vvvvnaaaaa cggggnnggg antgggcctt
19201 aaanccaaag ggcnaactcc ccaaccattn aaaaantgac nggggattat taaaancggc
19261 gggaaacatt tcacngccca actaatattg ttaaattaaa accaccaccn ctgcnccaag
19321 gagggacctt ctccnncaaa ncttattaaa gggcctgngc cttaaagagg gaaggaccct
19381 ctcctctgtc aatcttcctg tgctctnttt gtgaatcgct gacctctcta tctccgtgaa
19441 aaagagcacg ttcttctgct gtatgtaacc tgtctttttct atgatctctt tagggGTGAC
19501 CCAGTCTATT AAAGAAAGAA AAATGCTGAA TGAGgtaagt acttgatgtt acaaactaac
19561 cagagatatt cattcagtca tatagttaaa aatgtatttg cttccttcca tcaatgcacc
19621 actttcctta acaatgcaca aattttccat gataatgagg atcatcaaga attatgcagg
19681 cctgcactgt ggctcatacc tataatccca gcgctttggg aggctgaggc gcttggatcv
19741 vvvvvvvvvv vvtgcacccc aagacggatg ggnnctcccc cnccnctatt ttgatttaga
19801 aantgggttc ccatgngtaa gcngtgcgac ccctgcctca atgactnctc ctaatctccc
19861 aaattctagg atagngggtg anccc-tccn c-tgnc-tga -t-ccntaaa atg-gacgtg
19921 tctgctncnc cact-ccatt gaagga-gct tctctttctc ttatcctgat gggt-gtgnt
19981 -gtttctttc agCATGATTT TGAAGTCAGA GGAGATGTGG TCAATGGAAG AAACCACCAA
20041 GGTCCAAAGC GAGCAAGAGA ATCCCAGGAC AGAAAGgtaa agctccctcc ctcaagttga
20101 cnnaaatctc accccaccac tctgtattcc actccccttt gcagagatgg gccgcttcat

20161 tttgtaagac ttattacata catacacagt gctagatact ttcacacagg ttctttttc
20221 actcttccat cccaaccaca taaataagta ttgtctctac tttatgaatg ataaaactaa
20281 gagatttaga gaggctgtgt aatttggatt cccgtctcgg gttcagatcv vvvvvvvvvv
20341 vvATCTTCAG GGGGCTAGAA ATCTGTTGCT ATGGGCCCTT CACCAACATG CCCACAGgta
20401 agagcctcgg gagaanccag agttccagca ccagcctttg tcttacatag tggagtatta
20461 taagcaaggt cccacgatgg gggttcctca gattgctgaa atgttctaga ggctattcta
20521 tttctctacc actctccaaa caaaacagca cctaaatgtt atcctatggc aaaaaaaaac
20581 tataccttgt ccccttctc aagagcatga aggtggttaa tagttaggat tcagtatgtt
20641 atgtgttcag atggcgttga gctgctgtta gtgccvvvvv vvvvvvvvtt tgagagacta
20701 tcaaaccttа taccaagtgg ccttatggag actgataacc agagtacatg gcatatcagt
20761 ggcaaattga cttaaaatcc atacccctac tattttaaga ccattgtcct ttggagcaga
20821 gagacagact ctcccattga gaggtcttgc tataagcctt catccggaga gtgtagggta
20881 gagggcctgg gttaagtatg cagattactg cagtgatttt aGatcctaaa tgtccatttt
20941 agATCAACTG GAATGGATGG TACAGCTGTG TGGTGCTTCT GTGGTGAAGG AGCTTTCATC
21001 ATTCACCCTT GGCACAGvvv vvvvvvvvvv cagagcaaga ccctgtctca aaaacaaaca
21061 aaaaaaatga tgaagtgaca gttccagtag tcctactttg acactttgaa tgctctttcc
21121 ttcctgggga nccagGTGTC CACCCAATTG TGGTTGTGCA GCCAGATGCC TGGACAGAGG
21181 ACAATGGCTT CCATGtttat cactcattac ctggtgcttg agtagcacag ttcttggcac
21241 atttttaaat atttgttgaa tgaatggcta aaatgtcttt ttgatgtttt tattgttatt
21301 tgttttatat tgtaaaagta atacatgaac tgtttccatg gggtgggagt aagatatgaa
21361 tgttcatcac vvvvvvvvvv vvvcagtaat cctnagaact catacgaccg ggcccctgga
21421 gtcgnt-gnt tngagcctag tccngg-ag- aatgaattga cactaatctc tgcttgtgtt
21481 ctctgtctcc agCAATTGGG CAGATGTGTG AGGCGCCTGT GGTGACCCGA GAGTGGGTGT
21541 TGGACAGTGT AGCACTCTAC CAGTGCCAGG AGCTGGACAC CTACCTGATA CCCCAGATCC
21601 CCCACAGCCA CTACTGACTG CAGCCAGCCA CAGGTACAGA GCCACAGGAC CCCAAGAATG
21661 AGCTTACAAA GTGGCCTTTC CAGGCCCTGG GAGCTCCTCT CACTCTTCAG TCCTTCTACT
21721 GTCCTGGCTA CTAAATATTT TATGTACATC AGCCTGAAAA GGACTTCTGG CTATGCAAGG
21781 GTCCCTTAAA GATTTTCTGC TTGAAGTCTC CCTTGGAAAT

WHAT IS CLAIMED IS:

1. An isolated DNA sequence consisting essentially of DNA coding for a mutant BRCA1 polypeptide, said DNA selected from the group consisting of the mutations set forth in Table 14.

2. A nucleic acid probe complementary to human altered BRCA1 gene sequences, said altered BRCA1 gene sequences selected from the group consisting of the mutations set forth in Table 14.

3. A method for diagnosing a polymorphism associated with predisposition for breast and ovarian cancer in a human comprising: detecting a germline alteration of the wild-type BRCA1 gene or its expression products in a human sample, said alteration indicating a predisposition to said cancer and selected from the group consisting of the mutations of claim 1.

4. The method of claim 3 wherein said expression product is mRNA of BRCA1.

5. The method of claim 4 wherein the alteration of the wild-type BRCA1 gene mRNA is detected by hybridization of the mRNA from said human sample to a BRCA1 gene probe.

6. The method of claim 3 wherein the alteration of the wild-type BRCA1 gene is detected by observing shifts in electrophoretic mobility of single-stranded DNA on non-denaturing polyacrylamide gels.

7. The method of claim 3 wherein the alteration of the wild-type BRCA1 gene is detected by hybridization of a BRCA1 gene probe to genomic DNA isolated from said sample.

8. The method of claim 3 wherein the alteration of the wild-type BRCA1 gene is detected by hybridization with an allele-specific probe.

9. The method of claim 3 wherein the alteration of the wild-type BRCA1 gene is detected by amplifying all or part of the BRCA1 gene in said sample to produce an amplified sequence and sequencing the amplified sequence.

10. The method of claim 3 wherein the alteration of the wild-type BRCA1 gene is detected by amplifying all or part of the BRCA1 gene for a specific BRCA1 allele.

11. The method of claim 3 wherein the alteration of the wild-type BRCA1 gene is detected by molecularly cloning all or part of the BRCA1 gene in said sample to produce a cloned sequence and sequencing the cloned sequence.

12. The method of claim 3 wherein the alteration of the wild-type BRCA1 gene is detected by identifying a mismatch between molecules (1) BRCA1 gene genomic DNA or BRCA1 mRNA isolated from said tissue, 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.

13. The method of claim 3 wherein the alteration of the wild-type BRCA1 gene 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.

14. The method of claim 3 wherein the alteration of the wild-type BRCA1 gene is detected by amplification of BRCA1 gene sequnces in said sample and hybridization of the amplified sequences to nucleic acid probes which comprise non-wild-type BRCA1 gene sequences.

15. The method of claim 3 wherein the alteration of the wild-type BRCA1 gene is detected by screening for a deletion mutation.

16. The method of claim 3 wherein the alteration of the wild-type BRCA1 gene is detected by screening for a point mutation.

17. The method of claim 3 wherein the alteration of the wild-type BRCA1 gene is detected by screening for an insertion.

18. The method of claim 3 wherein the alteration of the wild-type BRCA1 gene is detected by *in situ* hybridization of the BRCA1 gene with nucleic acid probes which comprise the BRCA1 gene.

TITLE OF THE INVENTION

17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE MUTATIONS

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.

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 the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) 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 MUTATIONS

for which an application for Letters Patent of the United States was filed on 29 November 1994 and assigned Serial No. 08/348,824.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment(s) referred to above.

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 application(s) or PCT international application(s) designating the United States of America that is/are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in that/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 application(s) and the national or PCT international filing date of this application:

Ser. No. 08/308,104, filed 16 September 1994

Ser. No. 08/300,266, filed 2 September 1994

Ser. No. 08/289,221, filed 12 August 1994

Ser. No. 08/246,002, filed 20 May 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; Fred W. Hathaway, Registration No. 32,236; Royal W. Craig, Registration No. 34,145; E. Brendan Magrab, Registration No. 36,205; John J. Bettino, Registration No. 36,747; 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: Donna m. SHATTUCK-EIDENS

Inventor's Signature: Donna M. Shattuck-Eidens

Date of Signature: Feb 17, 1995

Residence Address: 1874 Texas Street
Salt Lake city, Utah 84108

Post Office Address: SAME

Country of Citizenship: USA

Full Name of Second Inventor: Jacques SIMARD

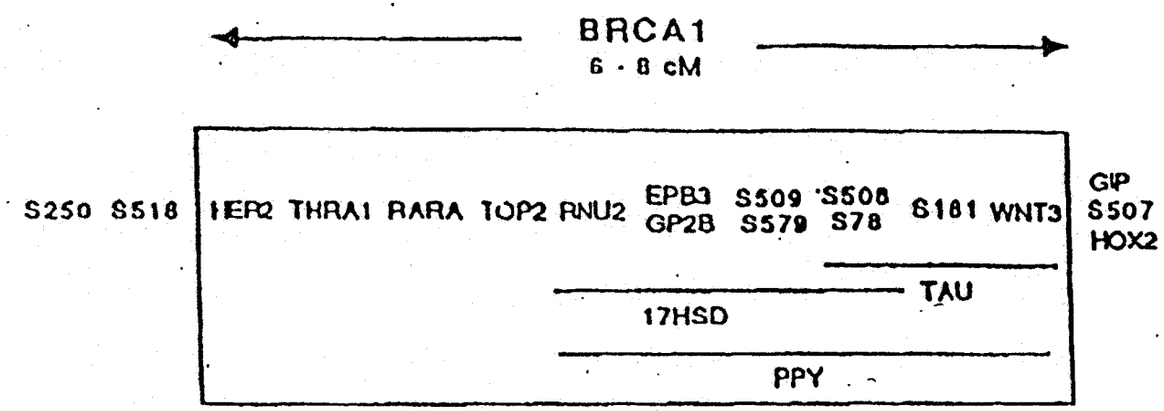
Inventor's Signature: Jacques Simard

Date of Signature: February 16th 1995

Residence Address: 4808 Pilet
St-Augustin de Desmaures
Québec, Canada, G3A 1Z9

Post Office Address: SAME

Country of Citizenship: CANADA



Map of the early onset breast and ovarian cancer region (BRCA1)

FIGURE 1

08/348824

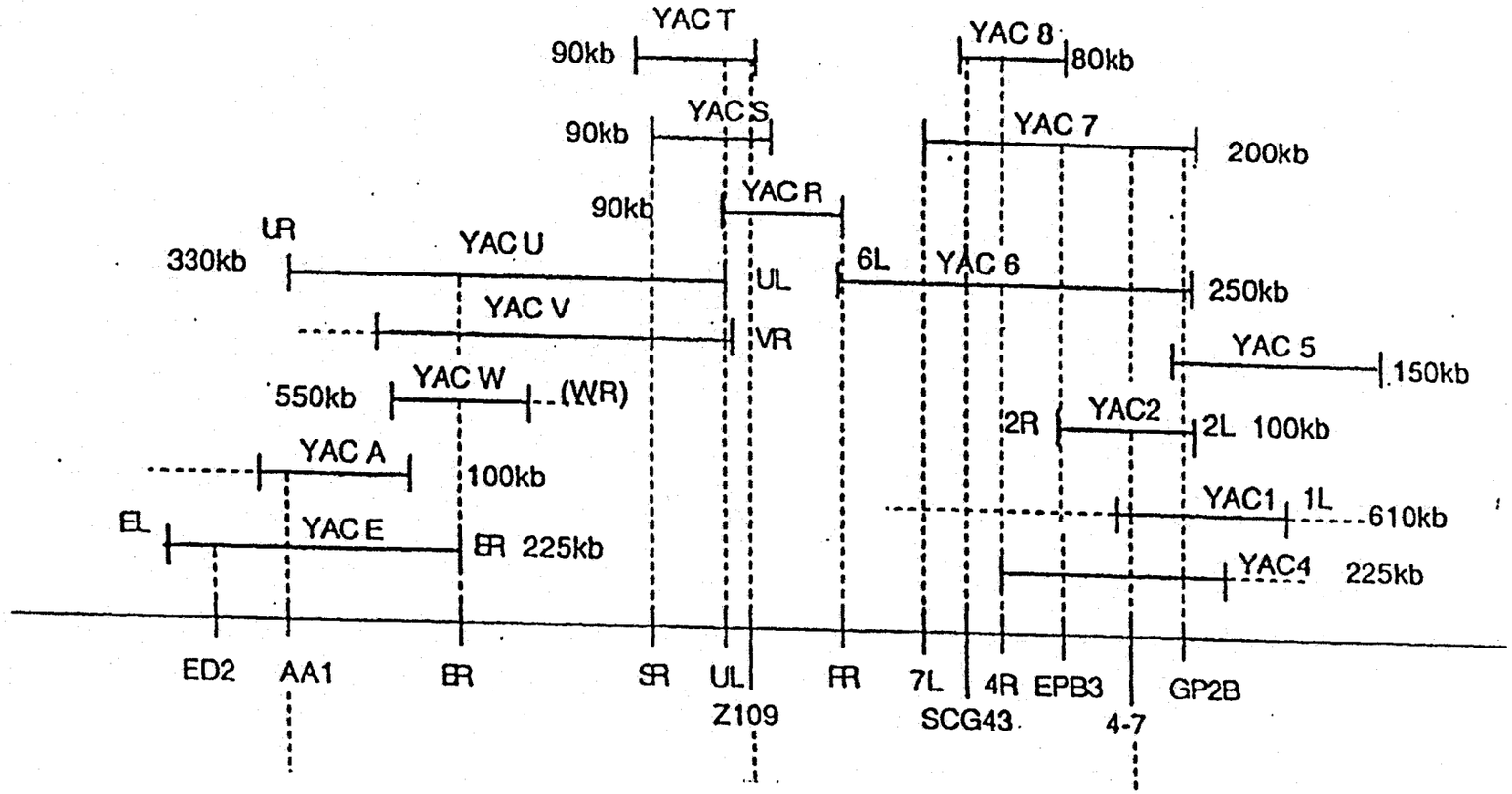


FIGURE 2

09/348824

GeneDX 1035, pg. 148

Schematic Map of STS', P1s, and Bac's in the Region

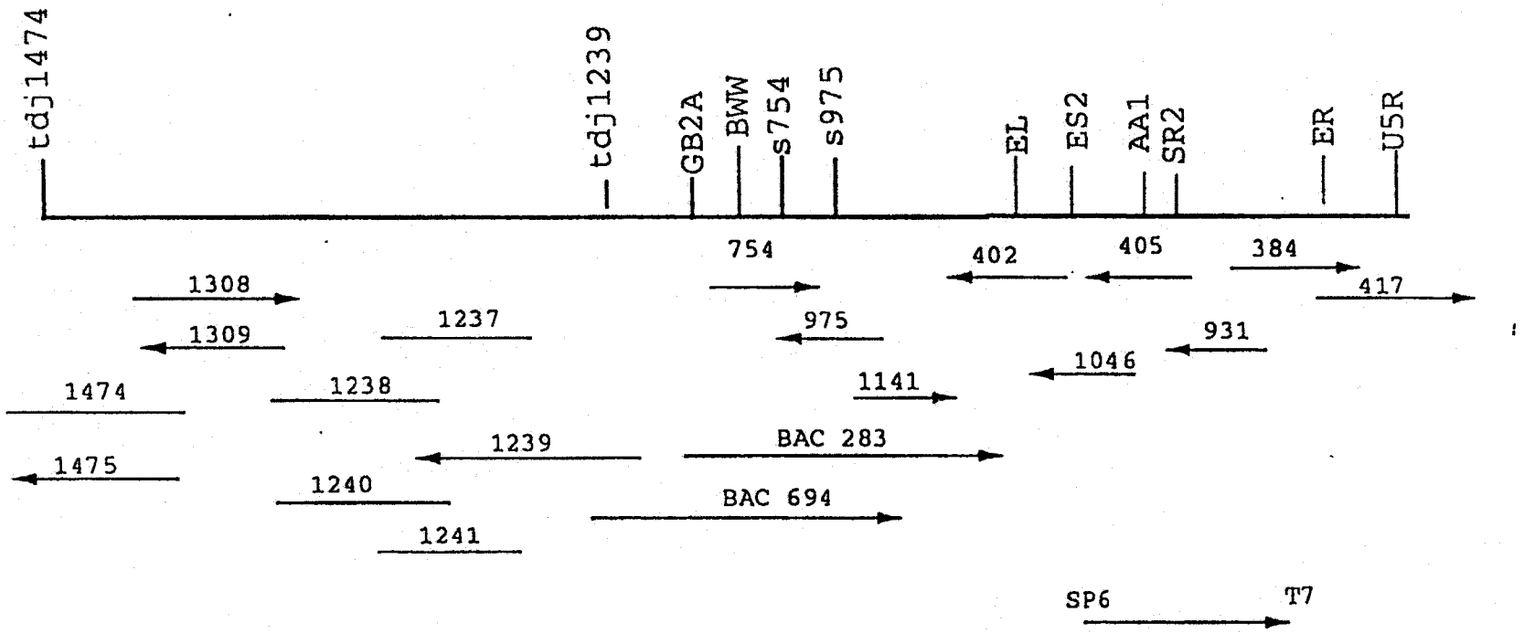


FIGURE 3

08/398,821

08/348824

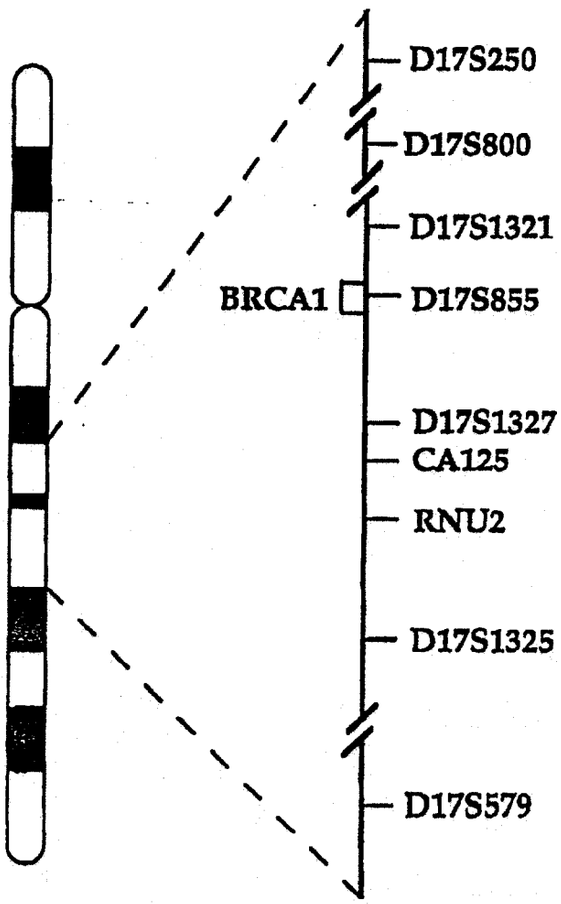


FIGURE 4

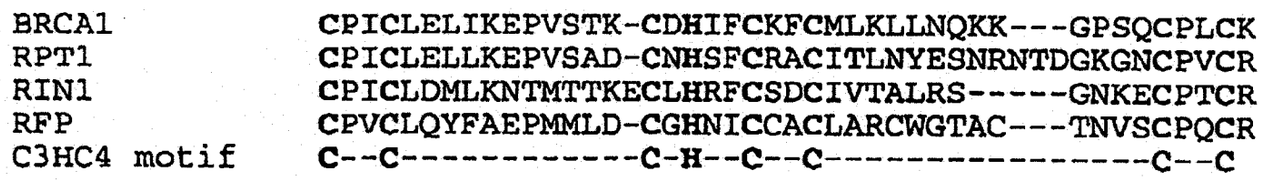


FIGURE 5

08/345824

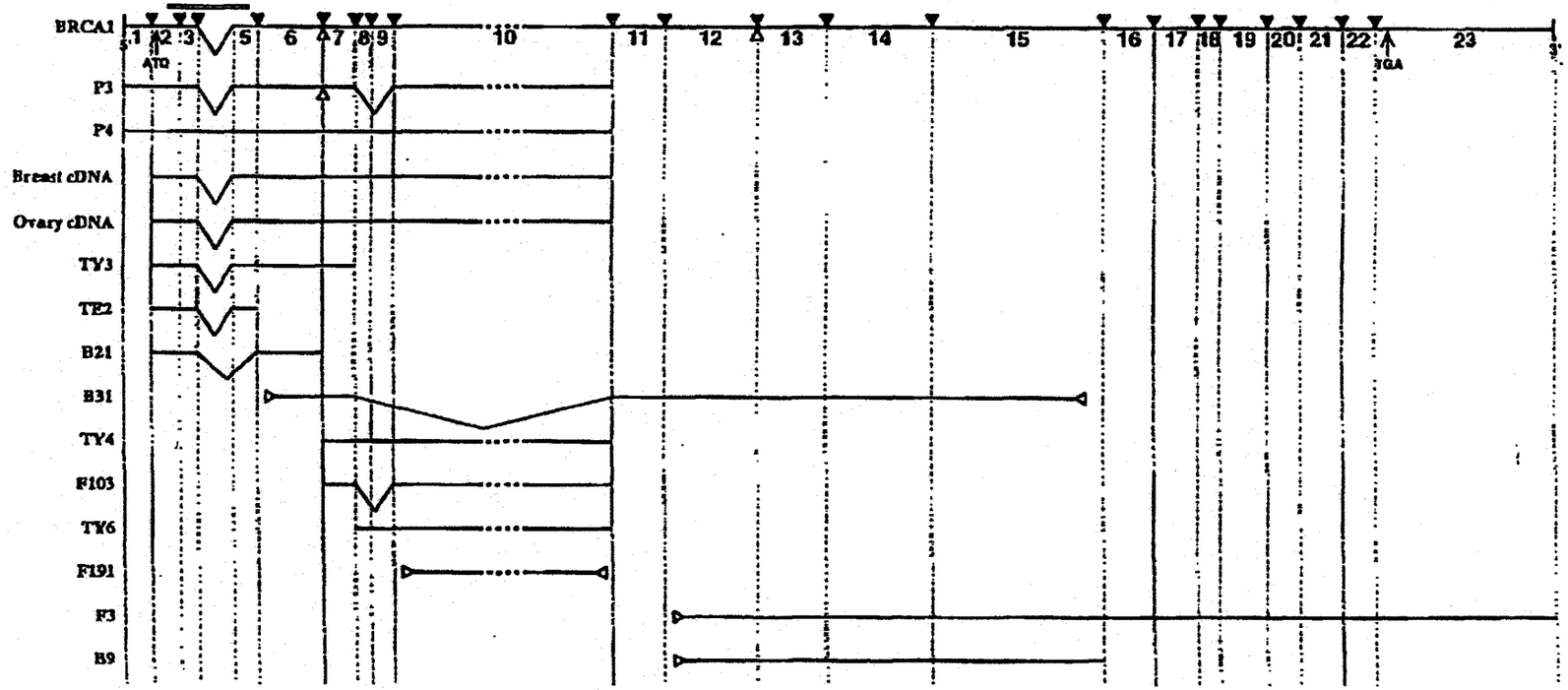


FIGURE 6

09/324,824

US 10/348824

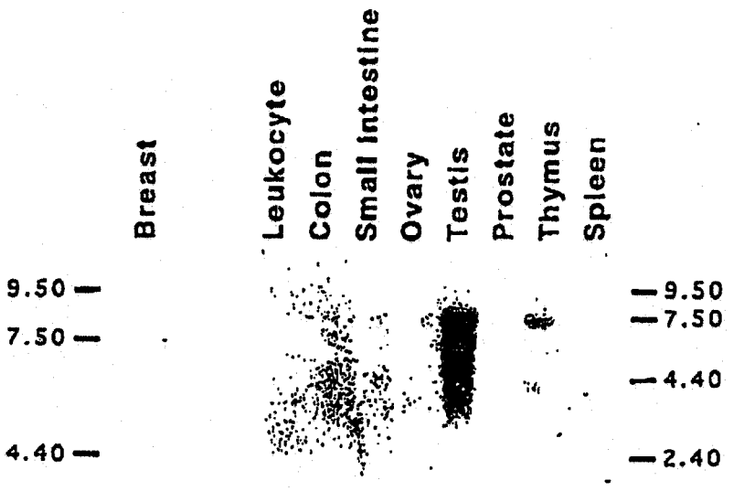


FIGURE 7

08/343824

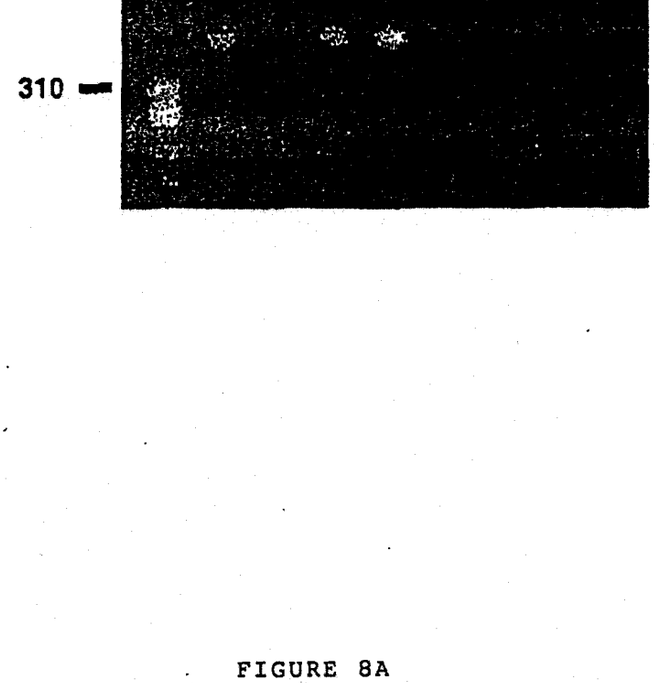


FIGURE 8A

US 10/348824

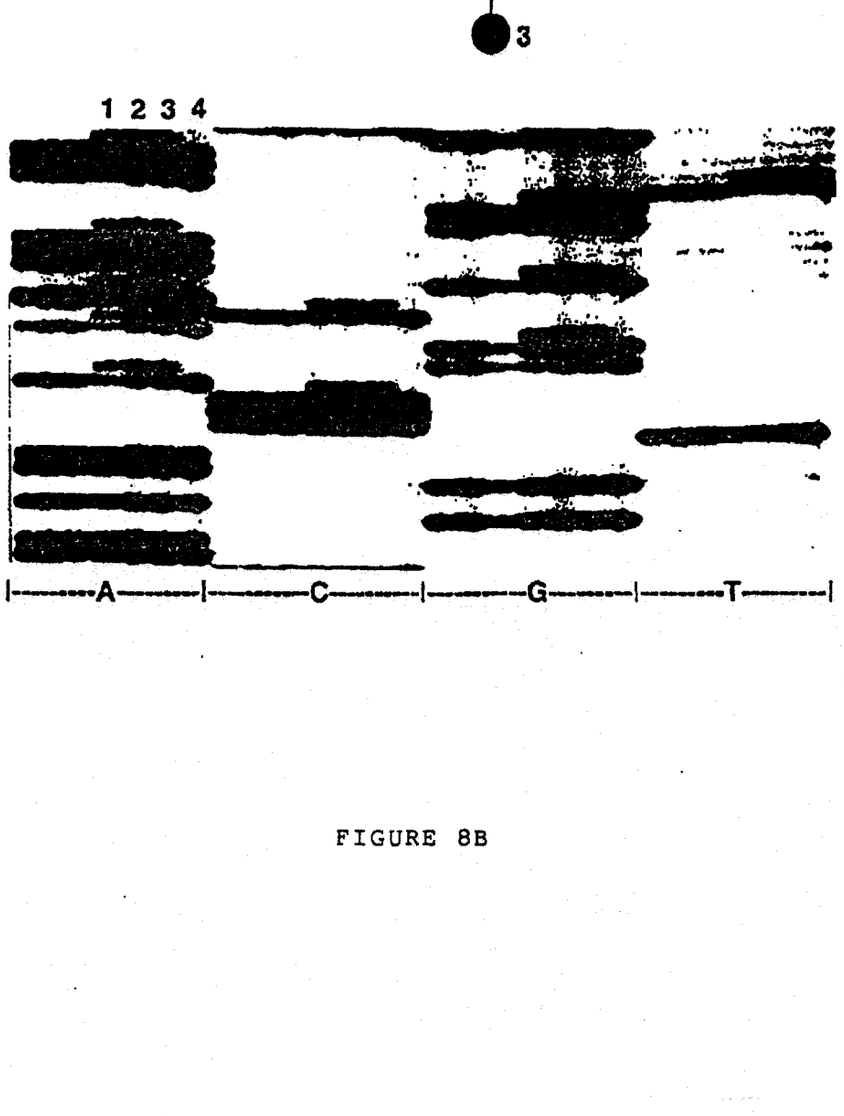


FIGURE 8B

08/348824

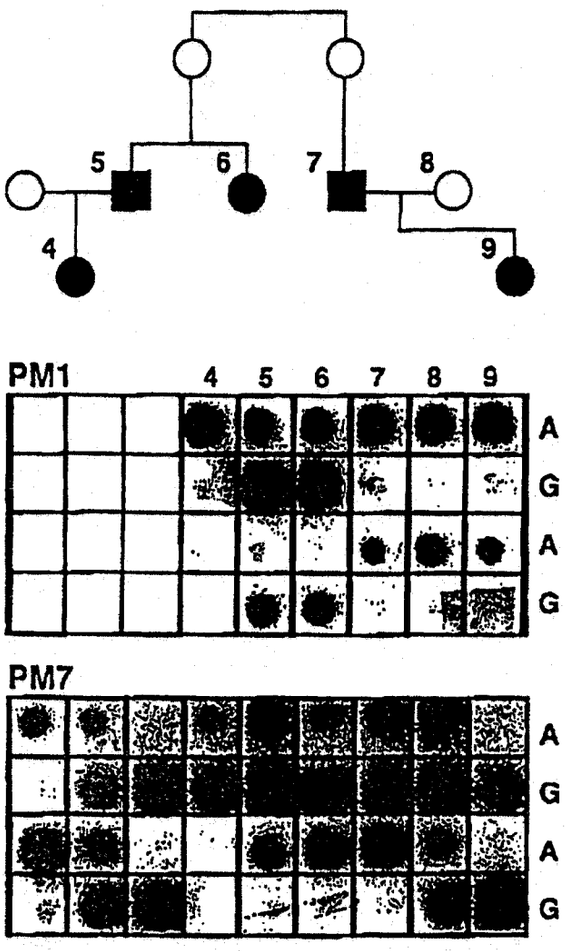


FIGURE 8C

05.12.95

08/348824



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In ~~re:~~ application of:

Donna M. SHATTUCK-EIDENS et al.

Serial No. (to be assigned)

Filed: herewith

For: 17q-LINKED BREAST AND OVARIAN
CANCER SUSCEPTIBILITY GENE MUTATIONS

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, filed concurrently herewith, Applicants present 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. 24884-109347-4
Dated: 29 November 1994
1935.01

Does not comply
- Send a notice -

YELLOW

This Raw Listing contains the General Information Section and those Sequences containing ERRORS.

SEQUENCE LISTING

(1) General Information:

(i) APPLICANT: Shattuck-Eidens, Donna M.
Simard, Jacques

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

(iii) NUMBER OF SEQUENCES: 14

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Venable, Baetjer, Howard & Civiletti
(B) STREET: 1201 New York Avenue, N.W., Suite 1000
(C) CITY: Washington
(D) STATE: DC
(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 #1.25

(vi) CURRENT APPLICATION DATA:
--> OK
(A) APPLICATION NUMBER: US
(B) FILING DATE: 02-SEP-1994
(C) CLASSIFICATION:

(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

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/246,602
(B) FILING DATE: 20-MAY-1994

YELLOW

RAW SEQUENCE LISTING
PATENT APPLICATION *US/08/348,824*

DATE: 12/20/94
TIME: 11:28:28

INPUT SET: S1614.raw

This Raw Listing contains the General Information Section and those Sequences containing ERRORS.

SEQUENCE LISTING

(1) General Information:

(i) APPLICANT: Shattuck-Eidens, Donna M.
Simard, Jacques

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

(iii) NUMBER OF SEQUENCES: 14

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Venable, Baetjer, Howard & Civiletti
(B) STREET: 1201 New York Avenue, N.W., Suite 1000
(C) CITY: Washington
(D) STATE: DC
(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 #1.25

--> OK

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US
(B) FILING DATE: 02-SEP-1994
(C) CLASSIFICATION:

(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

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/246,602
(B) FILING DATE: 20-MAY-1994

```
        46
        47     (viii) ATTORNEY/AGENT INFORMATION:
        48              (A) NAME: Ihnen, Jeffrey L.
        49              (B) REGISTRATION NUMBER: 28,957
        50              (C) REFERENCE/DOCKET NUMBER: 24884-109347-2
        51
        52       (ix) TELECOMMUNICATION INFORMATION:
        53              (A) TELEPHONE: 202-962-4810
        54              (B) TELEFAX: 202-962-8300
        55
        56
```

ERRORED SEQUENCES FOLLOW:

```
        57   (2) INFORMATION FOR SEQ ID NO:1:
        58
        59        (i) SEQUENCE CHARACTERISTICS:
-->     60             (A) LENGTH: 5857 base pairs
        61             (B) TYPE: nucleic acid
        62             (C) STRANDEDNESS: single
        63             (D) TOPOLOGY: linear
        64
        65       (ii) MOLECULE TYPE: cDNA
        66
        67      (iii) HYPOTHETICAL: NO
        68
        69       (iv) ANTI-SENSE: NO
        70
        71       (vi) ORIGINAL SOURCE:
        72             (A) ORGANISM: Homo sapiens
        73
        74       (ix) FEATURE:
        75             (A) NAME/KEY: CDS
        76             (B) LOCATION: 64..5769
        77
        78
        79       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
        80
        81   GGCCCCTGCG CTCAGGAGGC CTTCACCCTC TGCTCTGGGT AAAGTTCATT GGAACAGAAA      60
        82
        83   GAAATGGATT TATCTGCTCT TCGCGTTGAA GAAGTACAAA ATGTCATTAA TGCTATGCAG     120
        84
        85   AAAATCTTAG AGTGTCCCAT CTGTCTGGAG TTGATCAAGG AACCTGTCTC CACAAAGTGT     180
        86
        87   GACCACATAT TTTGCAAATT TTGCATGCTG AAACTTCTCA ACCAGAAGAA AGGGCCTTCA     240
        88
        89   CAGTGTCCTT TATGTAAGAA TGATATAACC AAAAGGAGCC TACAAGAAAG TACGAGATTT     300
        90
        91   AGTCAACTTG TTGAAGAGCT ATTGAAAATC ATTTGTGCTT TTCAGCTTGA CACAGGTTTG     360
        92
```

```
 93  GAGTATGCAA ACAGCTATAA TTTTGCAAAA AAGGAAAATA ACTCTCCTGA ACATCTAAAA     420
 94
 95  GATGAAGTTT CTATCATCCA AAGTATGGGC TACAGAAACC GTGCCAAAAG ACTTCTACAG     480
 96
 97  AGTGAACCCG AAAATCCTTC CTTGCAGGAA ACCAGTCTCA GTGTCCAACT CTCTAACCTT     540
 98
 99  GGAACTGTGA GAACTCTGAG GACAAAGCAG CGGATACAAC CTCAAAAGAC GTCTGTCTAC     600
100
101  ATTGAATTGG GATCTGATTC TTCTGAAGAT ACCGTTAATA AGGCAACTTA TTGCAGTGTG     660
102
103  GGAGATCAAG AATTGTTACA AATCACCCCT CAAGGAACCA GGGATGAAAT CAGTTTGGAT     720
104
105  TCTGCAAAAA AGGCTGCTTG TGAATTTTCT GAGACGGATG TAACAAATAC TGAACATCAT     780
106
107  CAACCCAGTA ATAATGATTT GAACACCACT GAGAAGCGTG CAGCTGAGAG GCATCCAGAA     840
108
109  AAGTATCAGG GTAGTTCTGT TTCAAACTTG CATGTGGAGC CATGTGGCAC AAATACTCAT     900
110
111  GCCAGCTCAT TACAGCATGA GAACAGCAGT TTATTACTCA CTAAAGACAG AATGAATGTA     960
112
113  GAAAAGGCTG AATTCTGTAA TAAAAGCAAA CAGCCTGGCT TAGCAAGGAG CCAACATAAC    1020
114
115  AGATGGGCTG GAAGTAAGGA AACATGTAAT GATAGGCGGA CTCCCAGCAC AGAAAAAAAG    1080
116  1080
117
--> 118  GTAGATCTGA ATGCTGATCC CCTGTGTGAG AGAAAAGAAT GGAATAAGCA GAAACTGCCA    1140
119
120  TGCTCAGAGA ATCCTAGAGA TACTGAAGAT GTTCCTTGGA TAACACTAAA TAGCAGCATT    1200
121
122  CAGAAAGTTA ATGAGTGGTT TTCCAGAAGT GATGAACTGT TAGGTTCTGA TGACTCACAT    1260
123
124  GATGGGGAGT CTGAATCAAA TGCCAAAGTA GCTGATGTAT TGGACGTCCT AATTGAGGTA    1320
125
126  GATGAATATT CTGGTTCTTC AGAGAAAATA GACTTACTGG CCAGTGATCC TCATGAGGCT    1380
127
128  TTAATATGTA AAAGTGACAG AGTTCACTCC AAATCAGTAG AGAGTAATAT TGAAGGCCAA    1440
129
130  ATATTTGGGA AAACCTATCG GAAGAAGGCA AGCCTCCCCA ACTTAAGCCA TGTAACTGAA    1500
131
132  AATCTAATTA TAGGNGCATT TGTTTCTGAG CCACAGATAA TACAAGAGCG TCCCCTCACA    1560
133
134  AATAAATTAA AGCGTAAAAG GAGACCTACA TCAGGCCTTC ATCCTGAGGA TTTTATCAAG    1620
135
136  AAAGCAGATT TGGCAGTTCA AAAGACTCCT GAAATGATAA ATCAGGGAAC TAACCAAACG    1680
137
138  GAGCAGAATG GTCAAGTGAT GAATATTACT AATAGTGGTC ATGAGAATAA AACAAAAGGT    1740
139
140  GATTCTATTC AGAATGAGAA AAATCCTAAC CCAATAGAAT CACTCGAAAA AGAATCTGCT    1800
141
142  TTCAAAACGA AAGCTGAACC TATAAGCAGC AGTATAAGCA ATATGGAACT CGAATTAAAT    1860
143
144  ATCCACAATT CAAAAGCACC TAAAAAGAAT AGGCTGAGGA GGAAGTCTTC TACCAGGCAT    1920
145
```

```
146   ATTCATGCGC TTGAACTAGT AGTCAGTAGA AATCTAAGCC CACCTAATTG TACTGAATTG    1980
147
148   CAAATTGATA GTTGTTCTAG CAGTGAAGAG ATAAAGAAAA AAAAGTACAA CCAAATGCCA    2040
149
150   GTCAGGCACA GCAGAAACCT ACAACTCATG GAAGGTAAAG AACCTGCAAC TGGAGCCAAG    2100
151
152   AAGAGTAACA AGCCAAATGA ACAGACAAGT AAAAGACATG ACAGTGATAC TTTCCCAGAG    2160
153
154   CTGAAGTTAA CAAATGCACC TGGTTCTTTT ACTAAGTGTT CAAATACCAG TGAACTTAAA    2220
155
156   GAATTTGTCA ATCCTAGCCT TCCAAGAGAA GAAAAAGAAG AGAAACTAGA AACAGTTAAA    2280
157
158   GTGTCTAATA ATGCTGAAGA CCCCAAAGAT CTCATGTTAA GTGGAGAAAG GGTTTTGCAA    2340
159
160   ACTGAAAGAT CTGTAGAGAG TAGCAGTATT TCACTGGTAC CTGGTACTGA TTATGGCACT    2400
161
162   CAGGAAAGTA TCTCGTTACT GGAAGTTAGC ACTCTAGGGA AGGCAAAAAC AGAACCAAAT    2460
163
164   AAATGTGTGA GTCAGTGTGC AGCATTTGAA AACCCCAAGG GACTAATTCA TGGTTGTTCC    2520
165
166   AAAGATAATA GAAATGACAC AGAAGGCTTT AAGTATCCAT TGGGACATGA AGTTAACCAC    2580
167
168   AGTCGGGAAA CAAGCATAGA AATGGAAGAA AGTGAACTTG ATGCTCAGTA TTTGCAGAAT    2640
169
170   ACATTCAAGG TTTCAAAGCG CCAGTCATTT GCTCTGTTTT CAAATCCAGG AAATGCAGAA    2700
171
172   GAGGAATGTG CAACATTCTC TGCCCACTCT GGGTCCTTAA AGAAACAAAG TCCAAAAGTC    2760
173
174   ACTTTTGAAT GTGAACAAAA GGAAGAAAAT CAAGGAAAGA ATGAGTCTAA TATCAAGCCT    2820
175
176   GTACAGACAG TTAATATCAC TGCAGGCTTT CCTGTGGTTG GTCAGAAAGA TAAGCCAGTT    2880
177
178   GATAATGCCA AATGTAGTAT CAAAGGAGGC TCTAGGTTTT GTCTATCATC TCAGTTCAGA    2940
179
180   GGCAACGAAA CTGGACTCAT TACTCCAAAT AAACATGGAC TTTTACAAAA CCCATATCGT    3000
181
182   ATACCACCAC TTTTTCCCAT CAAGTCATTT GTTAAAACTA AATGTAAGAA AAATCTGCTA    3060
183
184   GAGGAAAACT TTGAGGAACA TTCAATGTCA CCTGAAAGAG AAATGGGAAA TGAGAACATT    3120
185
186   CCAAGTACAG TGAGCACAAT TAGCCGTAAT AACATTAGAG AAAATGTTTT TAAAGAAGCC    3180
187
188   AGCTCAAGCA ATATTAATGA AGTAGGTTCC AGTACTAATG AAGTGGGCTC CAGTATTAAT    3240
189
190   GAAATAGGTT CCAGTGATGA AAACATTCAA GCAGAACTAG GTAGAAACAG AGGGCCAAAA    3300
191   3300
192
--> 193   TTGAATGCTA TGCTTAGATT AGGGGTTTTG CAACCTGAGG TCTATAAACA AAGTCTTCCT    3360
194
195   GGAAGTAATT GTAAGCATCC TGAAATAAAA AAGCAAGAAT ATGAAGAAGT AGTTCAGACT    3420
196
197   GTTAATACAG ATTTCTCTCC ATCTCTGATT TCAGATAACT TAGAACAGCC TATGGGAAGT    3480
198
```

```
      199  AGTCATGCAT CTCAGGTTTG TTCTGAGACA CCTGATGACC TGTTAGATGA TGGTGAAATA    3540
      200
      201  AAGGAAGATA CTAGTTTTGC TGAAAATGAC ATTAAGGAAA GTTCTGCTGT TTTTAGCAAA    3600
      202
      203  AGCGTCCAGA AAGGAGAGCT TAGCAGGAGT CCTAGCCCTT TCACCCATAC ACATTTGGCT    3660
      204
      205  CAGGGTTACC GAAGAGGGGC CAAGAAATTA GAGTCCTCAG AAGAGAACTT ATCTAGTGAG    3720
      206
      207  GATGAAGAGC TTCCCTGCTT CCAACACTTG TTATTTGGTA AAGTAAACAA TATACCTTCT    3780
      208
      209  CAGTCTACTA GGCATAGCAC CGTTGCTACC GAGTGTCTGT CTAAGAACAC AGAGGAGAAT    3840
      210
      211  TTATTATCAT TGAAGAATAG CTTAAATGAC TGCAGTAACC AGGTAATATT GGCAAAGGCA    3900
      212
      213  TCTCAGGAAC ATCACCTTAG TGAGGAAACA AAATGTTCTG CTAGCTTGTT TTCTTCACAG    3960
      214
      215  TGCAGTGAAT TGGAAGACTT GACTGCAAAT ACAAACACCC AGGCTCCTTT CTTGATTGGT    4020
      216
      217  TCTTCCAAAC AAATGAGGCA TCAGTCTGAA AGCCAGGGAG TTGGTCTGAG TGACAAGGAA    4080
      218
      219  TTGGTTTCAG ATGATGAAGA AAGAGGAACG GGCTTGGAAG AAAATAATCA AGAAGAGCAA
      220  4140
      221
-->   222  AGCATGGATT CAAACTTAGG TGAAGCAGCA TCTGGGTGTG AGAGTGAAAC AAGCGTCTCT    4200
      223
      224  GAAGACTGCT CAGGGCTATC CTCTCAGAGT GACATTTTAA CCACTCAGCA GAGGGATACC    4260
      225
      226  ATGCAACATA ACCTGATAAA GCTCCAGCAG GAAATGGCTG AACTAGAAGC TGTGTTAGAA    4320
      227
      228  CAGCATGGGA GCCAGCCTTC TAACAGCTAC CCTTCCATCA TAAGTGACTC TTCTGCCCTT    4380
      229
      230  GAGGACCTGC GAAATCCAGA ACAAAGCACA TCAGAAAAAG CAGTATTAAC TTCACAGAAA    4440
      231
      232  AGTAGTGAAT ACCCTATAAG CCAGAATCCA GAAGGCCTNT CTGCTGACAA GTTTGAGGTG    4500
      233
      234  TCTGCAGATA GTTCTACCAG TAAAAATAAA GAACCAGGAG TGGAAAGGTC ATCCCCTTCT    4560
      235
      236  AAATGCCCAT CATTAGATGA TAGGTGGTAC ATGCACAGTT GCTCTGGGAG TCTTCAGAAT    4620
      237
      238  AGAAACTACC CACCTCAAGA GGAGCTCATT AAGGTTGTTG ATGTGGAGGA GCAACAGCTG    4680
      239
      240  GAAGAGTCTG GGCCACACGA TTTGACGGAA ACATCTTACT TGCCAAGGCA AGATCTAGAG    4740
      241
      242  GGAACCCCTT ACCTGGAATC TGGAATCAGC CTCTTCTCTG ATGACCCTGA ATCTGATCCT    4800
      243
      244  TCTGAAGACA GAGCCCCAGA GTCAGCTCGT GTTGGCAACA TACCATCTTC AACCTCTGCA    4860
      245
      246  TTGAAAGTTC CCCAATTGAA AGTTGCAGAA TCTGCCCAGA GTCCAGCTGC TGCTCATACT    4920
      247
      248  ACTGATACTG CTGGGTATAA TGCAATGGAA GAAAGTGTGA GCAGGGAGAA GCCAGAATTG    4980
      249
      250  ACAGCTTCAA CAGAAAGGGT CAACAAAAGA ATGTCCATGG TGGTGTCTGG CCTGACCCCA    5040
      251
```

4140

```
     252  GAAGAATTTA TGCTCGTGTA CAAGTTTGCC AGAAAACACC ACATCACTTT AACTAATCTA    5100
     253
     254  ATTACTGAAG AGACTACTCA TGTTGTTATG AAAACAGATG CTGAGTTTGT GTGTGAACGG    5160
     255
     256  ACACTGAAAT ATTTTCTAGG AATTGCGGGA GGAAAATGGG TAGTTAGCTA TTTCTGGGTG    5220
     257
     258  ACCCAGTCTA TTAAAGAAAG AAAAATGCTG AATGAGCATG ATTTTGAAGT CAGAGGAGAT    5280
     259
     260  GTGGTCAATG GAAGAAACCA CCAAGGTCCA AAGCGAGCAA GAGAATCCCA GGACAGAAAG
     261  5340
     262
-->  263  ATCTTCAGGG GGCTAGAAAT CTGTTGCTAT GGGCCCTTCA CCAACATGCC CACAGATCAA    5400
     264
     265  CTGGAATGGA TGGTACAGCT GTGTGGTGCT TCTGTGGTGA AGGAGCTTTC ATCATTCACC    5460
     266
     267  CTTGGCACAG GTGTCCACCC AATTGTGGTT GTGCAGCCAG ATGCCTGGAC AGAGGACAAT    5520
     268
     269  GGCTTCCATG CAATTGGGCA GATGTGTGAG GCGCCTGTGG TGACCCGAGA GTGGGTGTTG    5580
     270
     271  GACAGTGTAG CACTCTACCA GTGCCAGGAG CTGGACACCT ACCTGATACC CCAGATCCCC    5640
     272
     273  CACAGCCACT ACTGACTGCA GCCAGCCACA GGTACAGAGC CACAGGACCC CAAGAATGAG    5700
     274
     275  CTTACAAAGT GGCCTTTCCA GGCCCTGGGA GCTCCTCTCA CTCTTCAGTC CTTCTACTGT    5760
     276
     277  CCTGGCTACT AAATATTTTA TGTACATCAG CCTGAAAAGG ACTTCTGGCT ATGCAAGGGT    5820
     278
     279  CCCTTAAAGA TTTTCTGCTT GAAGTCTCCC TTGGAAA                             5857
     280
     281
```

```
     921  (2) INFORMATION FOR SEQ ID NO:13:
     922
     923       (i) SEQUENCE CHARACTERISTICS:
-->  924            (A) LENGTH: 1540 base pairs
     925            (B) TYPE: nucleic acid
     926            (C) STRANDEDNESS: single
     927            (D) TOPOLOGY: linear
     928
     929      (ii) MOLECULE TYPE: DNA (genomic)
     930
     931     (iii) HYPOTHETICAL: NO
     932
     933      (iv) ANTI-SENSE: NO
     934
     935      (vi) ORIGINAL SOURCE:
     936            (A) ORGANISM: Homo sapiens
     937
     938      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13
     939
     940       1  GAGGCTAGAG GGCAGGCACT TTATGGCAAA CTCAGGTAGA ATTCTTCCTC
     941
     942      51  TTCCGTCTCT TTCCTTTTAC GTCATCCGGG GGCAGACTGG GTGGCCAATC
```

delete
wrong format.

delete these #'s

RAW SEQUENCE LISTING
PATENT APPLICATION *US/08/348,824*

DATE: 12/20/94
TIME: 11:28:58

INPUT SET: S1614.raw

150 etc...

```
943
944    101  CAGAGCCCCG AGAGACGCTT GGCTCTTTCT GTCCCTCCCA TCCTCTGATT
945
946    151  GTACCTTGAT TTCGTATTCT GAGAGGCTGC TGCTTAGCGG TAGCCCCTTG
947
948    201  GTTTCCGTGG CAACGGAAAA GCGCGGGAAT TACAGATAAA TTAAAACTGC
949
950    251  GACTGCGCGG CGTGAGCTCG CTGAGACTTC CTGGACGGGG GACAGGCTGT
951
952    301  GGGGTTTCTC AGATAACTGG GCCCCTGCGC TCAGGAGGCC TTCACCCTCT
953
954    351  GCTCTGGGTA AAGGTAGTAG AGTCCCGGGA AAGGGACAGG GGGCCCAAGT
955
956    401  GATGCTCTGG GGTACTGGCG TGGGAGAGTG GATTTCCGAA GCTGACAGAT
957
958    451  GGGTATTCTT TGACGGGGGG TAGGGGCGGA ACCTGAGAGG CGTAAGGCGT
959
960    501  TGTGAACCCT GGGGAGGGGG GCAGTTTGTA GGTCGCGAGG GAAGCGCTGA
961
962    551  GGATCAGGAA GGGGGCACTG AGTGTCCGTG GGGGAATCCT CGTGATAGGA
963
964    601  ACTGGAATAT GCCTTGAGGG GGACACTATG TCTTTAAAAA CGTCGGCTGG
965
966    651  TCATGAGGTC AGGAGTTCCA GACCAGCCTG ACCAACGTGG TGAAACTCCG
967
968    701  TCTCTACTAA AAATACnAAA ATTAGCCGGG CGTGGTGCCG CTCCAGCTAC
969
970    751  TCAGGAGGCT GAGGCAGGAG AATCGCTAGA ACCCGGGAGG CGGAGGTTGC
971
972    801  AGTGAGCCGA GATCGCGCCA TTGCACTCCA GCCTGGGCGA CAGAGCGAGA
973
974    851  CTGTCTCAAA ACAAAACAAA ACAAAACAAA ACAAAAAACA CCGGCTGGTA
975
976    901  TGTATGAGAG GATGGGACCT TGTGGAAGAA GAGGTGCCAG GAATATGTCT
977
978    951  GGGAAGGGGA GGAGACAGGA TTTTGTGGGA GGGAGAACTT AAGAACTGGA
979
980   1001  TCCATTTGCG CCATTGAGAA AGCGCAAGAG GGAAGTAGAG GAGCGTCAGT
981
982   1051  AGTAACAGAT GCTGCCGGCA GGGATGTGCT TGAGGAGGAT CCAGAGATGA
983
984   1101  GAGCAGGTCA CTGGGAAAGG TTAGGGGCGG GGAGGCCTTG ATTGGTGTTG
985
986   1151  GTTTGGTCGT TGTTGATTTT GGTTTTATGC AAGAAAAAGA AAACAACCAG
987
988   1201  AAACATTGGR GAAAGCTAAG GCTACCACCA CCTACCCGGT CAGTCACTCC
989
990   1251  TCTGTAGCTT TCTCTTYCTT GGAGAAAGGA AAAGACCCAA GGGGTTGGCA
991
992   1301  GCRATATGTG AAAAAATTCA GAATTTATGT TGTCTAATTA CAAAAAGCAA
993
994   1351  CTTCTAGAAT CTTTAAAAAT AAAGGACGTT GTCATTAGTT CTTTGGTTTG
995
```

```
 996    1401  TATYATTCTA AAACCTTCCA AATCTTAAAT TTACTTTATT TTAAAATGAT
 997
 998    1451  AAAATGAAGT TGTCATTTTA TAAACCTTTT AAAAAGATAT ATATATATGT
 999
1000    1501  TTTTCTAATG TGTTAnAGTT CATTGGAACA GAAAGAAATG
1001
1002
```

```
1003  (2) INFORMATION FOR SEQ ID NO:14:
1004
1005      (i) SEQUENCE CHARACTERISTICS:
--> 1006          (A) LENGTH: 21781 base pairs
1007              (B) TYPE: nucleic acid
1008              (C) STRANDEDNESS: single
1009              (D) TOPOLOGY: linear
1010
1011     (ii) MOLECULE TYPE: DNA (genomic)
1012
1013    (iii) HYPOTHETICAL: NO
1014
1015     (iv) ANTI-SENSE: NO
1016
1017     (vi) ORIGINAL SOURCE:
1018          (A) ORGANISM: Homo sapiens
1019
1020     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14
1021
1022        1 cctttacgt catcggggag actgggtggc aatcgcagcc cgagagacgc atggctcttt
1023       61 ctgccctcca tcctctgatg taccttgatt tcgtattctg agaggctgct gcttagcggt
1024      121 agcccttgg tttccgtggc aacggaaaag cgcgggaatt acagataaat taaaactgcg
1025      181 actgcgcggc gtgAGCTCGC TGAGACTTCC TGGACCCCGC ACCAGGCTGT GGGGTTTCTC
1026      241 AGATAACTGG GCCCCTGCGC TCAGGAGGCC TTCACCCTCT GCTCTGGGTA AAGvvvvvaa
1027      301 aaataaagga cgctgtcgay tagttcttct ggtttgtatt attctmaaac cttccaaatc
1028      361 ttcaaattta cttyatttta aaatgatgaa aatgaagttg tcattttata aacctttaa
1029      421 aaagatatat atatatgttt ttctaatgtg tTAAAGTTCA TTGGAACAGA AAGAAATGGA
1030      481 TTTATCTGCT CTTCGCGTTG AAGAAGTACA AAATGTCATT AATGCTATGC AGAAAATCTT
1031      541 AGAGTGTCCC ATCTGgtaag tcagcacaag agtgtattaa tttgggattc ctatgattat
1032      601 ctcctatgca aatgaacaga attgacctta catactaggg aagaaaagac atgtctagta
1033      661 agattaggct attgtaattg ctgattttct taactgaaga actttaaaaa tatagaaaat
1034      721 gattccttgt tctccatcca ctctgcctct cccactcctc tcctttcaa cacatcctgt
1035      781 ggtccgggaa agcagggctc tgtcttgatt ggttctgcac tgggcaggat ctgttagata
1036      841 ctgcatttgc tttctccagc tctaaavvvv vvvvvvvvaa atgctgatga tagtatagag
1037      901 tattgaaggg atcaatataa ttctgttttg atatctgaaa gctcactgaa ggtaaggatc
1038      961 gtattctctg ctgtattctc agttcctgac acagcagaca tttaataaat attgaacgaa
1039     1021 cttgaggcct tatgttgact cagtcataac agctcamagt tgaacttatt cactaagaat
1040     1081 agctttattt ytaaataaat tattgagcct catttatttt ctttttctcv vvvvTCTGGA
1041     1141 GTTGATCAAG GAACCTGTCT CCACAAAGTG TGACCACATA TTTTGCAAvv vvvatgtggc
1042     1201 tccattatta gcttttgttt ttgtccttca taacccagga aacacctaac tttatagaag
1043     1261 ctttactttc ttcaattaag tgagaacgaa aatccaactc catttcattc tttctcagag
1044     1321 agtatatagt tatcaaaagt tggttgtaat catagttcct ggtaaagttt tgacatatat
1045     1381 tatctttttt ttttttgag acaagtctcg ctctgtcgcc caggctggag tgcagtggca
1046     1441 tgaggcttgc tcactgcacc tccgcccccg agttcagcga ctctvvvvvv vvvvvvvtga
1047     1501 gatctagacc acatggtcaa agagatagaa tgtgagcaat aaatgaacct taaatttttc
```

fix remaining sequence as the format is displayed in SEQUENCE 1.

number here

delete

```
1048        1561 aacagctact ttttttttt tttttgaga cagGGkCTTA CTCTGTTGTC CCAGCTGGAG
1049        1621 TACAGw-G-T GCGATCATGA GGCTTACTGT TGC-TTGAC- TCCTA-GGCT CAAGCGATCC
1050        1681 TATCACCTCA GTCTC-CAAG -TAGCT-GGA CTgtaagtgc acaccaccat atccagctaa
1051        1741 attttgtgtt ttctgtagag acggggtttc gccatgtttc ccaggctggt cttgaacttt
1052        1801 gggcttaacc cgtctgccca cctaggcatc ccaaagtgct aggattacag gtgtgagtca
1053        1861 tcatgcctgg ccagtatttt agttagctct gtcttttcaa gtcatataca agttcatttt
1054        1921 ctttaagtta gttaacaacc ttatatcatg tattcttttc tagcavvvvv vvvvvvvvtg
1055        1981 tgatcataac agtaagccat atgcatgtaa gttcagtttc atagatcatt gcttatgtag
1056        2041 tttaggtttt tgcttatgca gcatccaaaa acaattagga aactattgct tgtaattcac
1057        2101 ctgccattac tttttaaatg gctcttaagg gcagttgtga gattatcttt tcatggctat
1058        2161 ttgccttttg agtattcttt ctacaaaagg aagtaaatta aattgttctt tctttcttta
1059        2221 taatttatag atvvvvvATT TTGCATGCTG AAACTTCTCA ACCAGAAGAA AGGGCCTTCA
1060        2281 CAGTGTCCTT TATGTAAGAA TGATATAACC AAAAGgatat aannngaant gatgctaggt
1061        2341 tggaagcaac cacagtagga aaagtagaaa ttatttaata acatagcgtt cctataaaac
1062        2401 cattcatcag aaaaatttat aaaagagttt ttagcacaca gtaaattatt tccaaagtta
1063        2461 ttttcctgaa agttttatgg gcatctgcct tatacaggtn ttgvvvvvvv vvvvvvggta
1064        2521 ggcttaaatg aatgacaaaa agttactaaa tcactgccat cacacggtta tatacagatg
1065        2581 tcaatgatgt attgatttag aggttttcta ctgttgctgc atcttatttt tatttgttta
1066        2641 catgtctttt cttattttag tgtccttaaa aggttgataa tcacttgctg agtgtgtttc
1067        2701 tcaaavvvvv GAGCCTACAA GAAAGTACGA GATTTAGTCA ACTTGTTGAA GAGCTATTGA
1068        2761 AAATCATTTG TGCTTTTCAG CTTGACACAG GTTTGGAGTA TGCAAACAGC TATAATTTTG
1069        2821 CAAAAAAGGA AAATAACTCT CCTGAACATC TAAAAGATGA AGTTTCTATC
1070  ATCCAAAGTA
1071        2881 TGGGCTACAG AAACCGTGCC AAAAGACTTC TACAGAGTGA ACCCGAAAAT CCTTCCTTGC
1072        2941 AGaaaccatt tgttttcttc ttcttcttct tcttcttttc ttttttttt ctttttttt
1073        3001 tcwtgagagt ggagtcyttg gctcgtssyg gcgccgaggc taggaagcag ctcctcctgc
1074        3061 gctcgtacgc cacgctagta ggtagcgtgg gcakacacgg cacgcgcacc atggccaggc
1075        3121 taatttttgt attgttvvvv vvvvvvvvva tgatggagat cttaaaaagt aatcattctg
1076        3181 gggctgggcg tagtagcttg cacctgtaat cccagcactt cgggaggctg aggcaggcag
1077        3241 ataatttgag gtcaggagtt tgagaccagc ctggccaaca tggtgaaacc catctctact
1078        3301 aaaaatacaa aaattagctg ggtgtggtgg cacgtacctg taatcccagc tactcgggag
1079        3361 gcggaggcac aagaattgct tgaacctagg acgcggaggt tgcagcgagc caagatcgcg
1080        3421 ccactgcact ccagcctggg ccgtagagtg agactctgtc tcaaaaaaga aaaaaaagta
1081        3481 attgttctag ctgggcgcag tggctcttgc ctgtaatccc agcactttgg gaggccaagg
1082        3541 cgggtggatc tcgagtccta gagttcaaga ccagcctagg caatgtggtg aaaccccatc
1083        3601 gctacaaaaa atacaaaaat tagccaggca tggtggcgtg cgcatgtagt cccagctcct
1084        3661 tgggaggctg aggtgggagg atcacttgaa cccaggagac agaggttgca gtgaaccgag
1085        3721 atcacgccac cacgctccag cctgggcaac agaacaagac tctgtctaaa aaaatacaaa
1086        3781 taaaataaaa gtagttctca cagtaccagc attcattttt caaaagatat agagctaaaa
1087        3841 aggaaggaaa aaaaaagtaa tgttgggctt ttaaatactc gttcctatac taaatgttct
1088        3901 taggagtgct ggggttttat tgtcatcatt tatccttttt aaaaatgtta ttggccaggc
1089        3961 acggtggctc atggctgtaa tcccagcact ttgggaggcc gaggcaggca gatcacctga
1090        4021 ggtcaggagt gtgagaccag cctggccaac atggcgaaac ctgtctctac taaaaataca
1091        4081 aaaattaact aggcgtggtg gtgtacgcct gtagtcccaG ctactcggga ggctgaggca
1092        4141 ggagaatcaa ctgaaccagg gaggtggagg ttgcagtgtg ccgagatcac gccactgcac
1093        4201 tctagcctgg caacagagca agattctgtc tcaaaaaaaa aaaacatata tacacatata
1094        4261 tcccaaagtg ctgggattac atatatatat atatatatat atTatatata tatatatata
1095        4321 tatatgtgat atatatgtga tatatatata acatatatat atgtAatata tatgtgatat
1096        4381 atatataata tatatatgtA atatatatgt gatatatata tatacacaca cacacacata
1097        4441 tatatgtAtg tgtgtgtAca Cacacacaca caaattagcc aggcatagtt gcacacgctt
1098        4501 ggtagaccCa gctactcagg aggCtgaggG aggagaATct cttgaActta ggaggcggag
1099        4561 gttgcagtga gctgagattg cgccactgca ctccagcctg ggtgacagag caggactctg
1100        4621 tacacccccc aaaacaaaaa aaaaagttat cagatgtgat tggaatgtat atcaagtatc
```

1101	4681	agcttcaaaa tatgctatat taatacttca aaaattacac aaataataca taatcaggtt
1102	4741	tgaaaaattt aagacaacms aaraaaaaaw ycmaatcaca matatcccac acattttatt
1103	4801	attmctmctm cwattatttt gwagagmctg ggtctcacyc ykttgctwat gctggtcttt
1104	4861	gaacyccykg ccycaarcar tcctsctcca bcctcccaar gtgctgggga twataggcat
1105	4921	garctacccg ctcccagccc magacaattt tagtgtgwgc aaaattcctg ggcatttttt
1106	4981	cmaggchdnm wayayvtwds mtgcwhrtga hbggctgatg wgtcaattta ttttgyccat
1107	5041	ggtgtcaakt ytctcttcag gaggaaaagc acamgaactg gccaayaatt gcttgactgt
1108	5101	tctttaccaw actgttwarc agGAAACCAr TCTCAGTGTC CCAACTCTCT AACCTTGGAA
1109	5161	CTTGTGAGAA CTCTGAGGAC AAAGCAGCGG ATACAACCTC AAAAGACGTC TGTCTACATT
1110	5221	GAATTGggta agggtctcag gtttttaag tatttaataa taattgctgg attccttatc
1111	5281	ttatagtttt gccamaaatc ttggtcataa tttgtatttg tggtaggcag cwttgggaag
1112	5341	tgaattttat gagccctatg gtgagttata aaaaatgtaa aagacgcagt tcccaccttg
1113	5401	aagaatctta ctttaaaaag ggagcaaaag aggccaggca tggtggctca cacctgtaat
1114	5461	cccagcactt tgggaggcca aagtgggtgg atcacctgag gtcgggagtt cgagaccagc
1115	5521	ctagccaaca tggagaaact ctgtctgtac caaaaaataa aaaattagcc aggtgtggtg
1116	5581	gcacataact gtaatcccag ctactcggga ggctgaggca ggagaatcac ttgaaCccgg
1117	5641	gaggtggagg ttgcggtgaa ccgagatcgc accattgcac tccagcctgg gcaaaaataG
1118	5701	cgaaactcca tctaaaaaaa aaaaagagag caaaagaaag amtmtctggt tttaamtmtg
1119	5761	tgtaaatatg ttttTggaaa gatggagagt agcaataaga aaaaacatga tggattgcta
1120	5821	cagtatttag ttccaagata aattgtacta gatgaggaag ccttttaaga agagctgaat
1121	5881	tgccaggcgc agtggctcac gcctgtaatc ccagcacttt gggaggccga ggtgggcgga
1122	5941	tcacctgagg tcgggagttc aagaccagcc tgaccaacat ggagaaaccc catctctact
1123	6001	aaaaaaaaaa aaaaaaaaat tagccggggt ggtggcttat gcctgtaatc ccagctactc
1124	6061	aggaggctga ggcaggagaa tcgcttgaac ccaggaagca gaggttgcag tgagccaaga
1125	6121	tcgcaccatt gcactccagc ctaggcaaca agagtgaaac tccatctcaa aaaaaaaaaa
1126	6181	aAAgagctga Atcttggctg ggcaggatgg ctcgtgcctg taatcctaac gctttggaag
1127	6241	accgaggcag aaggattggt tgagtccacg agtttaagac cagcctggcc aacatagggg
1128	6301	aaccctgtct ctatttttaa aataataata cattttggc cggtgcggtg gctcatgcct
1129	6361	gtaatcccaa tactttggga ggctgaggca ggtagatcac ctgaggtcag agttcgagac
1130	6421	cagcctggat aacctggtga aacccctctt tactaaaaat acaaaaaaaa aaaaaaatta
1131	6481	gctgggtgtg gtagcacatg cttgtaatcc cagctacttg ggaggctgag gcaggagaat
1132	6541	cgcttgaacc agggaggcgg aggttacaat gagccaacac tacaccactg cactccagcc
1133	6601	tgggcaatag agtgagactg catctcaaaa aaataataat ttttaaaaat aataaatttt
1134	6661	tttaagctta taaaaagaaa agttgaggcc agcatagtag ctcacatctg taatctcagc
1135	6721	agtggcagag gattgcttga agccaggagt ttgagaccag cctgggcaac atagcaagac
1136	6781	ctcatctcta caaaaaaatt tctttttaa attagctggg tgtggtggtg tgcatctgta
1137	6841	gtcccagcta ctcaggaggc agaggtgagt ggatacattg aacccaggag tttgaggctg
1138	6901	tagtgagcta tgatcatgcc actgcactcc aacctgggtg acagagcaag acctccaaaa
1139	6961	aaaaaaaaaa aagagctgct gagctcagaa ttcaaactgg gctctcaaat tggattttct
1140	7021	tttagaatat atttataatt aaaaaggata gccatctttt gagctcccag gcaccaccat
1141	7081	ctatttatca taacacttac tgttttcccc ccttatgatc ataaattcct agacaacagg
1142	7141	cattgtaaaa atagttatag tagttgatat ttaggagcac ttaactatat tccaggcact
1143	7201	attgtgcttt tcttgtataa ctcattagat gcttgtcaga cctctgagat tgttcctatt
1144	7261	atacttattt tacagatgag aaaattaagg cacagagaag ttatgaaatt tttccaaggt
1145	7321	attaaaccta gtaagtggct gagccatgat tcaaacctag gaagttagat gtcagagcct
1146	7381	gtgctttttt tttgtttttg tttttgtttt cagTagaaac gggggtctca ctttgttggc
1147	7441	caggctggtc ttgaactcct aacctcaaat aatccaccca tctcggcctc ctcaagtgct
1148	7501	gggattacag gtgagagcca ctgtgcctgg cgaagcccat gcctttaacc acttctctgt
1149	7561	attacatact agcttaacta gcattgtacc tgccacagta gatgctcagt aaatatttct
1150	7621	agttgaatat ctgtttttca acaagtacat ttttttaacc ctttttaatta agaaaacttt
1151	7681	tattgattta ttttttgggg ggaaattttt taGGATCTGA TTCTTCTGAA GATACCGTTA
1152	7741	ATAAGGCAAC TTATTGCAgg tgagtcaaag agaacctttg tctatgaagc tggtattttc
1153	7801	ctatttagtt aatattaagg attgatgttt ctctcttttt aaaaatattt taacttttat

```
1207  GAGGTAGATG
1208      10741 AATATTCTGG TTCTTCAGAG AAAATAGACT TACTGGCCAG TGATCCTCAT GAGGCTTTAA
1209      10801 TATGTAAAAG TGACAGAGTT CACTCCAAAT CAGTAGAGAG TAATATTGAA
1210  GGCCAAA-TA
1211      10861 TTTGGGAAAA CCTATCGGAA GAAGGCAAGC CTCCCCAACT TAAGCCATGT
1212  AACTGAAAAT
1213      10921 CTAATTATAG GnGCATTTGT TTCTGAGCCA CAGATAATAC AAGAGCGTCC CCTCACAAAT
1214      10981 AAATTAAAGC GTAAAAGGAG ACCTACATCA GGCCTTCATC CTGAGGATTT
1215  TATCAAGAAA
1216      11041 GCAGATTTGG CAGTTCAAAA GACTCCTGAA ATGATAAATC AGGGAACTAA CCAAACGGAG
1217      11101 CAGAAtggtc aagtgaTGAA TATTACTAAT AGTGG-TCA- TGAGAATAAA ACAAAAGGTG
1218      11161 ATTCTATTCA GAATGAGAAA AATCCTAACC CAATAGAATC ACTCGAAAAA
1219  GAATCTGCTT
1220      11221 TCAAAACGAA AGCTGAACCT ATAAGCAGCA GTATAAGCAA TATGGAACTC
1221  GAATTAAATA
1222      11281 TCCACAATTC AAAAGCACCT AAAAAGAATA GGCTGAGGAG GAAGTCTTCT
1223  ACCAGGCATA
1224      11341 TTCATGCGCT TGAACTAGTA GTCAGTAGAA ATCTAAGCCC ACCTAATTGT ACTGAATTGC
1225      11401 AAATTGATAG TTGTTCTAGC AGTGAAGAGA TAAAGAAAAA AAAGTACAAC
1226  CAAATGCCAG
1227      11461 TCAGGCACAG CAGAAACCTA CAACTCATGG AAGGTAAAGA ACCTGCAACT
1228  GGAGCCAAGA
1229      11521 AGAGTAACAA GCCAAATGAA CAGACAAGTA AAAGACATGA CAGTGATACT
1230  TTCCCAGAGC
1231      11581 TGAAGTTAAC AAATGCACCT GGTTCTTTTA CTAAGTGTTC AAATACCAGT GAACTTAAAG
1232      11641 AATTTGTCAA TCCTAGCCTT CCAAGAGAAG AAAAAGAAGA GAAACTAGAA
1233  ACAGTTAAAG
1234      11701 TGTCTAATAA TGCTGAAGAC CCCAAAGATC TCATGTTAAG TGGAGAAAGG
1235  GTTTTGCAAA
1236      11761 CTGAAAGATC TGTAGAGAGT AGCAGTATTT CACTGGTACC TGGTACTGAT TATGGCACTC
1237      11821 AGGAAAGTAT CTCGTTACTG GAAGTTAGCA CTCTAGGGAA GGCAAAAACA
1238  GAACCAAATA
1239      11881 AATGTGTGAG TCAGTGTGCA GCATTTGAAA ACCCCAAGGG ACTAATTCAT GGTTGTTCCA
1240      11941 AAGATAATAG AAATGACACA GAAGGCTTTA AGTATCCATT GGGACATGAA
1241  GTTAACCACA
1242      12001 GTCGGGAAAC AAGCATAGAA ATGGAAGAAA GTGAACTTGA TGCTCAGTAT
1243  TTGCAGAATA
1244      12061 CATTCAAGGT TTCAAAGCGC CAGTCATTTG CTCTGTTTTC AAATCCAGGA AATGCAGAAG
1245      12121 AGGAATGTGC AACATTCTCT GCCCACTCTG GGTCCTTAAA GAAACAAAGT
1246  CCAAAAGTCA
1247      12181 CTTTTGAATG TGAACAAAAG GAAGAAAATC AAGGAAAGAA TGAGTCTAAT
1248  ATCAAGCCTG
1249      12241 TACAGACAGT TAATATCACT GCAGGCTTTC CTGTGGTTGG TCAGAAAGAT AAGCCAGTTG
1250      12301 ATAATGCCAA ATGTAGTATC AAAGGAGGCT CTAGGTTTTG TCTATCATCT CAGTTCAGAG
1251      12361 GCAACGAAAC TGGACTCATT ACTCCAAATA AACATGGACT TTTACAAAAC CCATATCGTA
1252      12421 TACCACCACT TTTTCCCATC AAGTCATTTG TTAAAACTAA ATGTAAGAAA AATCTGCTAG
1253      12481 AGGAAAACTT TGAGGAACAT TCAATGTCAC CTGAAAGAGA AATGGGAAAT GAGAACATTC
1254      12541 CAAGTACAGT GAGCACAATT AGCCGTAATA ACATTAGAGA AAATGTTTTT
1255  AAAGAAGCCA
1256      12601 GCTCAAGCAA TATTAATGAA GTAGGTTCCA GTACTAATGA AGTGGGCTCC
1257  AGTATTAATG
1258      12661 AAATAGGTTC CAGTGATGAA AACATTCAAG CAGAACTAGG TAGAAACAGA
1259  GGGCCAAAAT
```

```
1260        12721 TGAATGCTAT GCTTAGATTA GGGGTTTTGC AACCTGAGGT CTATAAACAA AGTCTTCCTG
1261        12781 GAAGTAATTG TAAGCATCCT GAAATAAAAA AGCAAGAATA TGAAGAAGTA
1262    GTTCAGACTG
1263        12841 TTAATACAGA TTTCTCTCCA TCTCTGATTT CAGATAACTT AGAACAGCCT ATGGGAAGTA
1264        12901 GTCATGCATC TCAGGTTTGT TCTGAGACAC CTGATGACCT GTTAGATGAT GGTGAAATAA
1265        12961 AGGAAGATAC TAGTTTTGCT GAnAATGACA TTAAGGAAAG TTCTGCTGTT TTTAGCAAAA
1266        13021 GCGTCCAGAG GGGAGAGCTT AGCAGGnGTC CTAGCCCTTT CACCCATACA CATTTGGCTC
1267        13081 AGGGTTACCG AAGAGGGGCC AAGAAATTAG AGTCCTCAGA AGAGAACTTA
1268    TCTAGTGAGG
1269        13141 ATGAAGAGCT TCCCTGCTTC CAACACTTGT TATTTGGTAA AGTAAACAAT ATACCTTCTC
1270        13201 AGTCTACTAG GCATAGCACC GTTGCTACCG AGTGTCTGTC TAAGAACACA GAGGAGAATT
1271        13261 TATTATCATT GAAGAATAGC TTAAATGACT GCAGTAACCA GGTAATATTG
1272    GCAAAGGCAT
1273        13321 CTCAGGAACA TCACCTTAGT GAGGAAACAA AATGTTCTGC TAGCTTGTTT TCTTCACAGT
1274        13381 GCAGTGAATT GGAAGACTTG ACTGCAAATA CAAACACCCA GGCTCCTTTC TTGATTGGTT
1275        13441 CTTCCAAACA AATGAGGCAT CAGTCTGAAA GCCAGGGAGT TGGTCTGAGT
1276    GACAAGGAAT
1277        13501 TGGTTTCAGA TGATGAAGAA AGAGGAACGG GCTTGGAAGA AAATAATCAA
1278    GAAGAGCAAA
1279        13561 GCATGGATTC AAACTTAGgt attggaacca ggttt--gtg tgtgtggccc agtctattta
1280        13621 tagaagtgag ctaaatgttt atgcttttgg ggagcacatt ttacaaattt ccaagtatag
1281        13681 ttaaaggaac tgcttcttaa acttganaca tgttcctcct aaggtggttn ncatagaaaa
1282        13741 aagtccttca cacagctagg acgtcatctt tggcttgaat gagctttaac atcctantta
1283        13801 ctggtggact tacttctggt ttcattttat taaagaaatc caggtgtccc aagcaaggat
1284        13861 ttatcactgg gtgaaagaaa gtaatccagc cccgcantga gaggaaagca nagaagtgga
1285        13921 gngnttggct tgttncactg acccnctttg gnnttaggga gagactgtcc vvvvvvvvvv
1286        13981 vvvggtgtcc caaagcaagg aatttaatca ttttgtgtga catgaaagta natccagtcc
1287        14041 tgccaatgag aagaaaaagc acagcaagnn ngcagcgttt atagtctgct tttacatctg
1288        14101 aacctctgtt tttgttattt aagGT-AAGG TGAAGCAGCA TCTGGGTGTG AGAGTGAAAC
1289        14161 AAGCGTCTCT -GAAGACTGC TCAGGGCTAT CCTCTCAGAG TGACATTTTA -ACCACTCAG
1290        14221 CAGAGyvvvv vvvvvvvvng ngaatgtaat cctaatattt cncnccna-c tta-aaagaa
1291        14281 -tacc-actc caanggcatc ncaatacatc aatcaattgg gg-aattggg atttt--ccc
1292        14341 tcnctaacat ca-ntggaat aatttcatgg c-attaattg catgaatgtg gttagattaa
1293        14401 aaggtgttca tgcta-gaac ttgtagttcc atacta-ggt -atttcaat- cnttgttcct
1294        14461 caaan-aatt tgtatgatat attntcattt aatggaaa-g cttctcaaag tatttc-att
1295        14521 tt-cttggta ccatnnatcg tttttgaatn ag-gg-Nt=A CCATGCAACA TAACCTGATA
1296        14581 AAGCTCCAGC AGG-AAATGG CTGAACTAGA AGCTGTGTTA GAACAGCATG
1297    GGAGCCAGCC
1298        14641 TTCTAACAGC TACCCTTCCA TCATAAGTGA CTCTTCTGCC CTTGAGGACC TGCGAAATCC
1299        14701 AGAACA-A-A GCACATCAGA -AAAAGGtgt gtattgttgg ccaaacactg atatcttaan
1300        14761 caaaattctt tccttcccct ttatctcctt ctgaagagta aggacctagc tccaacattt
1301        14821 tatgatcctt gctcagcaca tgggtaatta tggagccttg gttcttgtcc ctgctcacaa
1302        14881 ctaatatacc agtcagaggg acccaaggca gtcattcatg ttgtcatctg agavvvvvvv
1303        14941 vvvvvvaaat tctgcctgat atacttgttn caaaaccaat ttgtgtatca tagattgatg
1304        15001 cttttgaana aaatcagtat tctaacctga attatcacta tcagaacaaa gcagtaaagt
1305        15061 agatttgttt tctcattcca tttaaagcag TATTAACTTC ACAGA-AAAG TAGTGAATAC
1306        15121 CCTATAAGCC AGAATCCAGA AGGCCTnTCT GCTGACAAGT TTGAGGTGTC TGCAGATAGT
1307        15181 TCTACCAGTA AAAATAAAGA ACCAGGAGTG GAAgtaagaa acatcaatgt aaagatgctg
1308        15241 tggtatctga catctttatt tatattgaac tctgattgtt aatttttttc accatacttt
1309        15301 ctccagtttt tttgcataca ggcatttata cacttttatt gctctaggat acttcttttg
1310        15361 tttaatccta tataggvvvv vvvvvvvvvg gataagntca agagatattt tgataggtga
1311        15421 tgcagtgatn aattgngaaa atttnctgcc tgcttttaat cttcccccgt tctttcttcc
1312        15481 tncctccctc ccttcctncc tcccgtcctt nccttccttt tccctccctt ccnccttctt
```

dashes are NOT valid.

```
1313      15541 tccntctntc tttcctttct ttcctgtcta cctttctttc cttcctccct tccttttctt
1314      15601 ttctttcttt cctttccttt tctttccttt ctttccttte ctttctttct tgacagagtc
1315      15661 ttgctctgtc actcaggctg gagtgcagtg gcgtgatctc gnctc-actg caacctct--
1316      15721 gtc-tc-cca ggt-tcaagc aattttcctg cctcagcctc ccgagtagct gagatta-ca
1317      15781 ggcgccagcc -ac-cac-ac cc-ag-ctac tgacctgctt ttvvvaaaca gctgggagat
1318      15841 atggtgcctc agaccatccc catgttatat gtcaaccctg acatattggc aggcaacatg
1319      15901 aatccagact tctaggctgt catgcgggct ctttctttgc cagtcattnc tgatctctct
1320      15961 gacatgaact gtntcaggta tgctttggct gcccagcaag tatgaaatgt cctntcacaa
1321      16021 ggggtggcga tggttttctc ctt-ccattt atctttctAG GTCATCCCCT TCTAAATGCC
1322      16081 CATCA-TTA- GATGATAGGT GGTACATG-C ACAGTTGCTC TGGGAGTCTT CAGAA-TAGA
1323      16141 AACTACCCA- CCTCAAGAGG AGCT-CATTA AGGTTGTTGA TGTGGAGGAG CAACAGCTGG
1324      16201 AAGAGTCTGG GCCACACGAT TTGACGGAAA CATCTTACTT GCCAAGGCAA GATCTAGgta
1325      16261 atatttcatc tgctgtattg garcamacac tytgatttta ctctgaatcc tacataaaga
1326      16321 tatt-ctggt tarccaactt -ttagat-gt rctagtctat -catggacac ttttgt-tat
1327      16381 acttaattaa gcccacttta gaaaaatagc tcaagtgtta atcwaggttt acttgwaaat
1328      16441 tattgaaact gttaatccat ctatatttta attaatggtt taactaatga ttttgaggat
1329      16501 gwgggagtck tggtgtactc tamatgtatt atttcaggcc aggcatagtg gctcacgcct
1330      16561 ggtaatccca gtayycmrga gcccgaggca ggtggagcca gctgaggtca ggagttcaag
1331      16621 acctgtcttg gccaacatgg gngaaaccct gtcttcttct taaaaaanac aaaaaaaatt
1332      16681 aactgggttg tgcttaggtg natgccccgn atcctagttn ttcttgnggg ttgagggagg
1333      16741 agatcacntt ggaccccgga ggggngggtg ggggngagca ggncaaaaca cngacccagc
1334      16801 tggggtggaa gggaagccca ctcnaaaaaa nnttnvvvvv vvvvvvvvAG GGAACCCCTT
1335      16861 ACCTGGAATC TGGAATCAGC CTCTT-CTCT GATGACCC-T GAATCTGATC C-TTCTGAAG
1336      16921 ACA-GAGCCC C-A-GAGTCA -GCTCGTGTT GGC-AACA-T ACCATCTTC- AACCTCTGCA
1337      16981 TTGAAAGTT- CCCCAATTGA AA-GTTGCAG AATCTGCCCA GAGTCCAGCT GCTGCTCA-T
1338      17041 ACTACTGATA CTGCTGGGTA TAATGCAATG GAAGAAAGTG TGAGCAGGGA
1339  GAAGCCAGAA
1340      17101 TTGACAGCTT CAACAGAAAG GGTCAACAAA AGAATGTCCA TGGTGG-TGT CTGGCCTGAC
1341      17161 CCCAGAAGAA TTTgtgagtg tatccatatg tatctcccta atgactaaga cttaacaaca
1342      17221 ttctggaaag agttttatgt aggtattgtc aattaataac ctagaggaag aaatctagaa
1343      17281 ancaatcaca gttctgtgta atttaatttc gattactaat ttctgaaaat ttagaayvvv
1344      17341 vvvvvvvvvv ncccnncccc cnaatctgaa atgggggtaa ccccccccca accganacnt
1345      17401 gggtn-gcnt agaga-nttt -aa-tggccc nt-tctgagg -nacanaagc ttaagc-cag
1346      17461 gng-acgtgg ancnat-gng ttgtttnttg tt-tggttac ctccagcctg gtgacaganc
1347      17521 aagctctgtc tngnaaaant aggngagggn naagattcac nntaaatagt tccaggacac
1348      17581 gtgtagaacg tgcaggattg ctacgtaggt aaacatatgc catggtggga taactagtat
1349      17641 tctgagctgt gtgctagagg taactcatga taatggaata tttgatttaa tttcagATGC
1350      17701 TCGTGTACAA GTTTGCCAGA AAACACCACA TCACTTTAAC TAATCTAATT ACTGAAGAGA
1351      17761 CTACTCATGT TGTTATGAAA ACAGAtatac ccaagaacct ttacagaata ccttkcatct
1352      17821 sctgcataaa accacatgag gcgaggcaca gtggcgcatg cctgtawtcg caRcactttg
1353      17881 ggaggccgag gcgggcagat accccgagRt tMSgagtcga gccatcctgg ccagcatggt
1354      17941 gaaaccccgt ctctactaNN aaatggNaaa attaNctggg tgtggtcgcg tgcNcctgta
1355      18001 gtcccagcta ctcgtgaggc tgaggcagga gaatcacttg aaccggggaa atggaggttt
1356      18061 cagtgagcag agatcatNcc cctNcattcc agcctggcga cagagcaagg ctccgtcNcc
1357      18121 Naaaaaataa aaaaaaacgt gaacaaataa gaatatttgt tgagcatagc atggatgata
1358      18181 gtcttctaat agtcaatcaa ttactttatg aaagacaaat aatagttttg ctgcttcctt
1359      18241 acctcctttt gttttgggtt aagatttgga gtgtgggcca ggcacvvvvv vvvvvvvvga
1360      18301 tctat-agct agccttggCg tctagaagat gggtgttgag aagagggagt ggaaagatat
1361      18361 ttcctctggt cttaacttca tatcagcctc ccctagactt ccaaatatcc atacctgctg
1362      18421 gttataatta gtggtgtttt cagcctctga ttctgtcacc aggggtttta gaatcataaa
1363      18481 tccagattga tcttgggagt gtaaaaaact gaggctcttt agcttcttag gacAgcactt
1364      18541 cctgattttg ttttcaactt ctaatccttt gagtgTktct cattctgcag aTGCTGAGTT
1365      18601 TGTGTGTGAA CGGACACTGA AATATTTTCT AGGAATTGCG GGAGGAAAAT
```

```
1366   GGGTAGTTAG
1367        18661 CTATTTCTGG kaagkataat actatwtctc cccycctccc CtAacamcTc agaattgsat
1368        18721 tttacaccta acatttacam ctaAgnyyyt kgctgatgct tgagtctgag tcmcmaaagk
1369        18781 cctttaatTg Taatactaac tactTttatc ttTaatatca cttTgTtcag AtaagCtggt
1370        18841 gaTGCtggga aaatggtctc ttttataact aataggacct aAtctgctcc tagcaaTgtt
1371        18901 agca-tatga gctagggatt tatttaatag tcggcaggaa tccatgtgca rcagncaaac
1372        18961 ttataatgtt taaattaaac atcaactctg tctccagaag gaaactgctg ctacaagcct
1373        19021 tattaaaggg ctgtggcttt agagggaagg acctctcctc tgtcattctt cctgtgctct
1374        19081 tttgtgaatc gctgacctct ctatctccgt gaaaagagca cgttcttctg ctgtatgtaa
1375        19141 cctgtctttt ctatgatctc tvvvvvvvvv vvvvnaaaaa cggggnnggg antgggcctt
1376        19201 aaanccaaag ggcnaactcc ccaaccattn aaaaantgac nggggattat taaaancggc
1377        19261 gggaaacatt tcacngccca actaatattg ttaaattaaa accaccaccn ctgcnccaag
1378        19321 gagggacctt ctccnncaaa ncttattaaa gggcctgngc cttaaagagg gaaggaccct
1379        19381 ctcctctgtc aatcttcctg tgctctnttt gtgaatcgct gacctctcta tctccgtgaa
1380        19441 aaagagcacg ttcttctgct gtatgtaacc tgtcttttct atgatctctt tagggGTGAC
1381        19501 CCAGTCTATT AAAGAAAGAA AAATGCTGAA TGAGgtaagt acttgatgtt acaaactaac
1382        19561 cagagatatt cattcagtca tatagttaaa aatgtatttg cttccttcca tcaatgcacc
1383        19621 actttcctta acaatgcaca aattttccat gataatgagg atcatcaaga attatgcagg
1384        19681 cctgcactgt ggctcatacc tataatccca gcgctttggg aggctgaggc gcttggatcv
1385        19741 vvvvvvvvvv vvtgcacccc aagacggatg ggnnctcccc cnccnctatt ttgatttaga
1386        19801 aantgggttc ccatgngtaa gcngtgcgac ccctgcctca atgactnctc ctaatctccc
1387        19861 aaattctagg atagngggtg anccc-tccn c-tgnc-tga -t-ccntaaa atg-gacgtg
1388        19921 tctgctncnc cact-ccatt gaagga-gct tctctttctc ttatcctgat gggt-gtgnt
1389        19981 -gtttctttc agCATGATTT TGAAGTCAGA GGAGATGTGG TCAATGGAAG AAACCACCAA
1390        20041 GGTCCAAAGC GAGCAAGAGA ATCCCAGGAC AGAAAGgtaa agctccctcc ctcaagttga
1391        20101 cnnaaatctc accccaccac tctgtattcc actccccttt gcagagatgg gccgcttcat
1392        20161 tttgtaagac ttattacata catacacagt gctagatact ttcacacagg ttctttttc
1393        20221 actcttccat cccaaccaca taaataagta ttgtctctac tttatgaatg ataaaactaa
1394        20281 gagatttaga gaggctgtgt aatttggatt cccgtctcgg gttcagatcv vvvvvvvvvv
1395        20341 vvATCTTCAG GGGGCTAGAA ATCTGTTGCT ATGGGCCCTT CACCAACATG CCCACAGgta
1396        20401 agagcctcgg gagaanccag agttccagca ccagcctttg tcttacatag tggagtatta
1397        20461 taagcaaggt cccacgatgg gggttcctca gattgctgaa atgttctaga ggctattcta
1398        20521 tttctctacc actctccaaa caaaacagca cctaaatgtt atcctatggc aaaaaaaaac
1399        20581 tataccttgt ccccttctc aagagcatga aggtggttaa tagttaggat tcagtatgtt
1400        20641 atgtgttcag atggcgttga gctgctgtta gtgccvvvvv vvvvvvvvtt tgagagacta
1401        20701 tcaaacctta taccaagtgg ccttatggag actgataacc agagtacatg gcatatcagt
1402        20761 ggcaaattga cttaaaatcc atacccctac tattttaaga ccattgtcct ttggagcaga
1403        20821 gagacagact ctcccattga gaggtcttgc tataagcctt catccggaga gtgtagggta
1404        20881 gagggcctgg gttaagtatg cagattactg cagtgatttt aGatcctaaa tgtccatttt
1405        20941 agATCAACTG GAATGGATGG TACAGCTGTG TGGTGCTTCT GTGGTGAAGG AGCTTTCATC
1406        21001 ATTCACCCTT GGCACAGvvv vvvvvvvvvv cagagcaaga ccctgtctca aaaacaaaca
1407        21061 aaaaaaatga tgaagtgaca gttccagtag tcctactttg acactttgaa tgctctttcc
1408        21121 ttcctgggga nccagGTGTC CACCCAATTG TGGTTGTGCA GCCAGATGCC TGGACAGAGG
1409        21181 ACAATGGCTT CCATGtttat cactcattac ctggtgcttg agtagcacag ttcttggcac
1410        21241 atttttaaat atttgttgaa tgaatggcta aaatgtcttt ttgatgtttt tattgttatt
1411        21301 tgttttatat tgtaaaagta atacatgaac tgtttccatg gggtgggagt aagatatgaa
1412        21361 tgttcatcac vvvvvvvvvv vvvcagtaat cctnagaact catacgaccg ggcccctgga
1413        21421 gtcgnt-gnt tngagcctag tccngg-ag- aatgaattga cactaatctc tgcttgtgtt
1414        21481 ctctgtctcc agCAATTGGG CAGATGTGTG AGGCGCCTGT GGTGACCCGA GAGTGGGTGT
1415        21541 TGGACAGTGT AGCACTCTAC CAGTGCCAGG AGCTGGACAC CTACCTGATA CCCCAGATCC
1416        21601 CCCACAGCCA CTACTGACTG CAGCCAGCCA CAGGTACAGA GCCACAGGAC CCCAAGAATG
1417        21661 AGCTTACAAA GTGGCCTTTC CAGGCCCTGG GAGCTCCTCT CACTCTTCAG TCCTTCTACT
1418        21721 GTCCTGGCTA CTAAATATTT TATGTACATC AGCCTGAAAA GGACTTCTGG CTATGCAAGG
```

1419 21781 GTCCCTTAAA GATTTTCTGC TTGAAGTCTC CCTTGGAAAT

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
08/348,824	11/29/94	SHATTUCK-EIDENS	D 248841093474

#2

03P1/0104

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

0000

DATE MAILED: 01/04/95

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 $ 860.

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 to complete the basic filing fee.

2. ☐ Additional claim fees of $_____ 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 items omitted at time of execution.

   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 to the oath or declaration is: ☐ missing; ☐ a reproduction; ☐ 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 for returned checks. (37 CFR 1.21(m)).

9. ☐ Your filing receipt was mailed in error because 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 and any questions about this notice to B. DAVENPORT Application Processing Division, Special Processing and Correspondence Branch (703) 308-1202.

A copy of this notice MUST be returned with the response.

GeneDX 1035, pg. 175

MAIL ROOM
74 MAR 6 1995
PAT. & TRADEMARK OFF.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. patent application of:

Donna M. SHATTUCK-EIDENS et al.

Serial No.: 08/348,824

Filed: 29 November 1994

Attn: Application Processing Division, Special Processing and Correspondence Branch

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE MUTATIONS

TRANSMITTAL OF FORMAL DOCUMENTS

Hon. Commissioner of Patents and Trademarks
Washington, D.C. 20231

Sir:

In response to the Notice to File Missing Parts of Application (Filing Date Granted) dated 4 January 1995, Applicants submit herewith the following documents:

1. An executed Declaration and Power of Attorney document;
2. A Petition for Extension of Time (in triplicate);
3. A check from the Assignee of this application covering both the filing fee/surcharge and extension fee above ($970.00);
4. A computer disk of the sequence listing, with an executed Statement Under 37 CFR §1.821(f);
5. Two Assignments, with covering sheerts and authorization to charge the recordation fees to our Deposit Account
6. A Preliminary Amendment; and
7. Copies of PTO-1533 and its attached Notice to Comply.

Please charge any discrepancies in fees to 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
1201 New York Avenue, N.W., #1000
Washington, D.C. 20005
(202) 962-4810

Atty. Docket No. 24884-109347-4
Dated: 6 March 1995
6565.01

#3

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 the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) 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 MUTATIONS

for which an application for Letters Patent of the United States was filed on 29 November 1994 and assigned Serial No. 08/348,824.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment(s) referred to above.

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 application(s) or PCT international application(s) designating the United States of America that is/are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in that/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 application(s) and the national or PCT international filing date of this application:

Ser. No. 08/308,104, filed 16 September 1994

Ser. No. 08/300,266, filed 2 September 1994

Ser. No. 08/289,221, filed 12 August 1994

Ser. No. 08/246,002, filed 20 May 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; Fred W. Hathaway, Registration No. 32,236; Royal W. Craig, Registration No. 34,145; E. Brendan Magrab, Registration No. 36,205; John J. Bettino, Registration No. 36,747; 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.

1-00

Full Name of Sole or First Inventor: Donna M. SHATTUCK-EIDENS

Inventor's Signature: Donna M. Shattuck-Eidens

Date of Signature: Feb 17, 1995

Residence Address: 1874 Texas Street
Salt Lake city, Utah 84108 UT

Post Office Address: SAME

Country of Citizenship: USA

2-00

Full Name of Second Inventor: Jacques SIMARD

Inventor's Signature: Jacques Simard

Date of Signature: February 16th 1995

Residence Address: 4808 Pilet
St-Augustin de Desmaures
Québec, Canada, G3A 1Z9 CA

Post Office Address: SAME

Country of Citizenship: CANADA

MAIL ROOM
74 MAR 6 1995
PAT. & TRADEMARK

03C6
1/13/95 110 — 115

CEIVED
MAR 14 1995
APPLICATION DIVISION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. patent application of:

Donna M. SHATTUCK-EIDENS et al.

Serial No.: 08/348,824

Filed: 29 November 1994

Attn: Application Processing Division, Special Processing and Correspondence Branch

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE MUTATIONS

PETITION FOR EXTENSION OF TIME

Hon. Commissioner of Patents and Trademarks
Washington, D.C. 20231

Sir:

Petition is hereby made for an extension of one (1) month in which to respond to the Notice to File Missing Parts of Application (Filing Date Granted) mailed 4 January 1995.

This Petition is submitted in triplicate, and is accompanied by the requisite formal documents, computer sequence disk, two assignments (with covering sheets), and a check from the Assignee of the application in the amount of $970.00, which includes the filing fee/surcharge and extension fee. If any discrepancies in fees are noted, the same should be charged to Deposit Account 22-0261.

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. 24884-109347-4
Dated: 6 March 1995
6565.01

090 BA 03/10/95 08348824 1 115 110.00 CK

#3

MAIL ROOM
74 MAR 6 1995
PAT. & TRADEMARK OFFICE

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re. application of:

Donna M. SHATTUCK-EIDENS et al.

Serial No. 08/348,824

Filed: 29 November 1994

Attn: Application Processing Division, Special Processing and Correspondence Branch

For: 17q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE MUTATIONS

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, Applicants present 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. 24884-109347-4
Dated: 6 March 1995
1935.01

Application No. 08/348824

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 copy of the marked-up "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) 557-0400

Please return a copy of this notice with your response.

This Raw Listing contains the General Information Section and up to the first 5 pages.

ENTERED

SEQUENCE LISTING

(1) General Information:

(i) APPLICANT: Shattuck-Eidens, Donna M.
Simard, Jacques

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

(iii) NUMBER OF SEQUENCES: 61

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Venable, Baetjer, Howard & Civiletti
(B) STREET: 1201 New York Avenue, N.W., Suite 1000
(C) CITY: Washington
(D) STATE: DC
(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: WordPerfect 5.1/5.2

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/348,824
(B) FILING DATE: 29-NOV-1994
(C) CLASSIFICATION:

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

(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

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/246,602
(B) FILING DATE: 20-MAY-1994

```
47
48     (viii) ATTORNEY/AGENT INFORMATION:
49              (A) NAME: Ihnen, Jeffrey L.
50              (B) REGISTRATION NUMBER: 28,957
51              (C) REFERENCE/DOCKET NUMBER: 24884-109347-3
52
53        (ix) TELECOMMUNICATION INFORMATION:
54              (A) TELEPHONE: 202-962-4810
55              (B) TELEFAX: 202-962-8300
56
57
58   (2) INFORMATION FOR SEQ ID NO:1:
59
60         (i) SEQUENCE CHARACTERISTICS:
61              (A) LENGTH: 5857 base pairs
62              (B) TYPE: nucleic acid
63              (C) STRANDEDNESS: single
64              (D) TOPOLOGY: linear
65
66        (ii) MOLECULE TYPE: cDNA
67
68       (iii) HYPOTHETICAL: NO
69
70        (iv) ANTI-SENSE: NO
71
72        (vi) ORIGINAL SOURCE:
73              (A) ORGANISM: Homo sapiens
74
75        (ix) FEATURE:
76              (A) NAME/KEY: CDS
77              (B) LOCATION: 64..5769
78
79
80        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
81
82   GGCCCCTGCG CTCAGGAGGC CTTCACCCTC TGCTCTGGGT AAAGTTCATT GGAACAGAAA      60
83
84   GAAATGGATT TATCTGCTCT TCGCGTTGAA GAAGTACAAA ATGTCATTAA TGCTATGCAG     120
85
86   AAAATCTTAG AGTGTCCCAT CTGTCTGGAG TTGATCAAGG AACCTGTCTC CACAAAGTGT     180
87
88   GACCACATAT TTTGCAAATT TTGCATGCTG AAACTTCTCA ACCAGAAGAA AGGGCCTTCA     240
89
90   CAGTGTCCTT TATGTAAGAA TGATATAACC AAAAGGAGCC TACAAGAAAG TACGAGATTT     300
91
92   AGTCAACTTG TTGAAGAGCT ATTGAAAATC ATTTGTGCTT TTCAGCTTGA CACAGGTTTG     360
93
94   GAGTATGCAA ACAGCTATAA TTTTGCAAAA AAGGAAAATA ACTCTCCTGA ACATCTAAAA     420
95
96   GATGAAGTTT CTATCATCCA AAGTATGGGC TACAGAAACC GTGCCAAAAG ACTTCTACAG     480
97
98   AGTGAACCCG AAAATCCTTC CTTGCAGGAA ACCAGTCTCA GTGTCCAACT CTCTAACCTT     540
99
```

100	GGAACTGTGA GAACTCTGAG GACAAAGCAG CGGATACAAC CTCAAAAGAC GTCTGTCTAC	600
101		
102	ATTGAATTGG GATCTGATTC TTCTGAAGAT ACCGTTAATA AGGCAACTTA TTGCAGTGTG	660
103		
104	GGAGATCAAG AATTGTTACA AATCACCCCT CAAGGAACCA GGGATGAAAT CAGTTTGGAT	720
105		
106	TCTGCAAAAA AGGCTGCTTG TGAATTTTCT GAGACGGATG TAACAAATAC TGAACATCAT	780
107		
108	CAACCCAGTA ATAATGATTT GAACACCACT GAGAAGCGTG CAGCTGAGAG GCATCCAGAA	840
109		
110	AAGTATCAGG GTAGTTCTGT TTCAAACTTG CATGTGGAGC CATGTGGCAC AAATACTCAT	900
111		
112	GCCAGCTCAT TACAGCATGA GAACAGCAGT TTATTACTCA CTAAAGACAG AATGAATGTA	960
113		
114	GAAAAGGCTG AATTCTGTAA TAAAAGCAAA CAGCCTGGCT TAGCAAGGAG CCAACATAAC	1020
115		
116	AGATGGGCTG GAAGTAAGGA AACATGTAAT GATAGGCGGA CTCCCAGCAC AGAAAAAAAG	1080
117		
118	GTAGATCTGA ATGCTGATCC CCTGTGTGAG AGAAAAGAAT GGAATAAGCA GAAACTGCCA	1140
119		
120	TGCTCAGAGA ATCCTAGAGA TACTGAAGAT GTTCCTTGGA TAACACTAAA TAGCAGCATT	1200
121		
122	CAGAAAGTTA ATGAGTGGTT TTCCAGAAGT GATGAACTGT TAGGTTCTGA TGACTCACAT	1260
123		
124	GATGGGGAGT CTGAATCAAA TGCCAAAGTA GCTGATGTAT TGGACGTCCT AATTGAGGTA	1320
125		
126	GATGAATATT CTGGTTCTTC AGAGAAAATA GACTTACTGG CCAGTGATCC TCATGAGGCT	1380
127		
128	TTAATATGTA AAAGTGACAG AGTTCACTCC AAATCAGTAG AGAGTAATAT TGAAGGCCAA	1440
129		
130	ATATTTGGGA AAACCTATCG GAAGAAGGCA AGCCTCCCCA ACTTAAGCCA TGTAACTGAA	1500
131		
132	AATCTAATTA TAGGNGCATT TGTTTCTGAG CCACAGATAA TACAAGAGCG TCCCCTCACA	1560
133		
134	AATAAATTAA AGCGTAAAAG GAGACCTACA TCAGGCCTTC ATCCTGAGGA TTTTATCAAG	1620
135		
136	AAAGCAGATT TGGCAGTTCA AAAGACTCCT GAAATGATAA ATCAGGGAAC TAACCAAACG	1680
137		
138	GAGCAGAATG GTCAAGTGAT GAATATTACT AATAGTGGTC ATGAGAATAA AACAAAAGGT	1740
139		
140	GATTCTATTC AGAATGAGAA AAATCCTAAC CCAATAGAAT CACTCGAAAA AGAATCTGCT	1800
141		
142	TTCAAAACGA AAGCTGAACC TATAAGCAGC AGTATAAGCA ATATGGAACT CGAATTAAAT	1860
143		
144	ATCCACAATT CAAAAGCACC TAAAAAGAAT AGGCTGAGGA GGAAGTCTTC TACCAGGCAT	1920
145		
146	ATTCATGCGC TTGAACTAGT AGTCAGTAGA AATCTAAGCC CACCTAATTG TACTGAATTG	1980
147		
148	CAAATTGATA GTTGTTCTAG CAGTGAAGAG ATAAAGAAAA AAAAGTACAA CCAAATGCCA	2040
149		
150	GTCAGGCACA GCAGAAACCT ACAACTCATG GAAGGTAAAG AACCTGCAAC TGGAGCCAAG	2100
151		
152	AAGAGTAACA AGCCAAATGA ACAGACAAGT AAAAGACATG ACAGTGATAC TTTCCCAGAG	2160

153
154 CTGAAGTTAA CAAATGCACC TGGTTCTTTT ACTAAGTGTT CAAATACCAG TGAACTTAAA 2220
155
156 GAATTTGTCA ATCCTAGCCT TCCAAGAGAA GAAAAAGAAG AGAAACTAGA AACAGTTAAA 2280
157
158 GTGTCTAATA ATGCTGAAGA CCCCAAAGAT CTCATGTTAA GTGGAGAAAG GGTTTTGCAA 2340
159
160 ACTGAAAGAT CTGTAGAGAG TAGCAGTATT TCACTGGTAC CTGGTACTGA TTATGGCACT 2400
161
162 CAGGAAAGTA TCTCGTTACT GGAAGTTAGC ACTCTAGGGA AGGCAAAAAC AGAACCAAAT 2460
163
164 AAATGTGTGA GTCAGTGTGC AGCATTTGAA AACCCCAAGG GACTAATTCA TGGTTGTTCC 2520
165
166 AAAGATAATA GAAATGACAC AGAAGGCTTT AAGTATCCAT TGGGACATGA AGTTAACCAC 2580
167
168 AGTCGGGAAA CAAGCATAGA AATGGAAGAA AGTGAACTTG ATGCTCAGTA TTTGCAGAAT 2640
169
170 ACATTCAAGG TTTCAAAGCG CCAGTCATTT GCTCTGTTTT CAAATCCAGG AAATGCAGAA 2700
171
172 GAGGAATGTG CAACATTCTC TGCCCACTCT GGGTCCTTAA AGAAACAAAG TCCAAAAGTC 2760
173
174 ACTTTTGAAT GTGAACAAAA GGAAGAAAAT CAAGGAAAGA ATGAGTCTAA TATCAAGCCT 2820
175
176 GTACAGACAG TTAATATCAC TGCAGGCTTT CCTGTGGTTG GTCAGAAAGA TAAGCCAGTT 2880
177
178 GATAATGCCA AATGTAGTAT CAAAGGAGGC TCTAGGTTTT GTCTATCATC TCAGTTCAGA 2940
179
180 GGCAACGAAA CTGGACTCAT TACTCCAAAT AAACATGGAC TTTTACAAAA CCCATATCGT 3000
181
182 ATACCACCAC TTTTTCCCAT CAAGTCATTT GTTAAAACTA AATGTAAGAA AAATCTGCTA 3060
183
184 GAGGAAAACT TTGAGGAACA TTCAATGTCA CCTGAAAGAG AAATGGGAAA TGAGAACATT 3120
185
186 CCAAGTACAG TGAGCACAAT TAGCCGTAAT AACATTAGAG AAAATGTTTT TAAAGAAGCC 3180
187
188 AGCTCAAGCA ATATTAATGA AGTAGGTTCC AGTACTAATG AAGTGGGCTC CAGTATTAAT 3240
189
190 GAAATAGGTT CCAGTGATGA AAACATTCAA GCAGAACTAG GTAGAAACAG AGGGCCAAAA 3300
191
192 TTGAATGCTA TGCTTAGATT AGGGGTTTTG CAACCTGAGG TCTATAAACA AAGTCTTCCT 3360
193
194 GGAAGTAATT GTAAGCATCC TGAAATAAAA AAGCAAGAAT ATGAAGAAGT AGTTCAGACT 3420
195
196 GTTAATACAG ATTTCTCTCC ATCTCTGATT TCAGATAACT TAGAACAGCC TATGGGAAGT 3480
197
198 AGTCATGCAT CTCAGGTTTG TTCTGAGACA CCTGATGACC TGTTAGATGA TGGTGAAATA 3540
199
200 AAGGAAGATA CTAGTTTTGC TGAAAATGAC ATTAAGGAAA GTTCTGCTGT TTTTAGCAAA 3600
201
202 AGCGTCCAGA AAGGAGAGCT TAGCAGGAGT CCTAGCCCTT TCACCCATAC ACATTTGGCT 3660
203
204 CAGGGTTACC GAAGAGGGGC CAAGAAATTA GAGTCCTCAG AAGAGAACTT ATCTAGTGAG 3720
205

206 GATGAAGAGC TTCCCTGCTT CCAACACTTG TTATTTGGTA AAGTAAACAA TATACCTTCT 3780
207
208 CAGTCTACTA GGCATAGCAC CGTTGCTACC GAGTGTCTGT CTAAGAACAC AGAGGAGAAT 3840
209
210 TTATTATCAT TGAAGAATAG CTTAAATGAC TGCAGTAACC AGGTAATATT GGCAAAGGCA 3900
211
212 TCTCAGGAAC ATCACCTTAG TGAGGAAACA AAATGTTCTG CTAGCTTGTT TTCTTCACAG 3960
213
214 TGCAGTGAAT TGGAAGACTT GACTGCAAAT ACAAACACCC AGGCTCCTTT CTTGATTGGT 4020
215
216 TCTTCCAAAC AAATGAGGCA TCAGTCTGAA AGCCAGGGAG TTGGTCTGAG TGACAAGGAA 4080
217
218 TTGGTTTCAG ATGATGAAGA AAGAGGAACG GGCTTGGAAG AAAATAATCA AGAAGAGCAA 4140
219
220 AGCATGGATT CAAACTTAGG TGAAGCAGCA TCTGGGTGTG AGAGTGAAAC AAGCGTCTCT 4200
221
222 GAAGACTGCT CAGGGCTATC CTCTCAGAGT GACATTTTAA CCACTCAGCA GAGGGATACC 4260
223
224 ATGCAACATA ACCTGATAAA GCTCCAGCAG GAAATGGCTG AACTAGAAGC TGTGTTAGAA 4320
225
226 CAGCATGGGA GCCAGCCTTC TAACAGCTAC CCTTCCATCA TAAGTGACTC TTCTGCCCTT 4380
227
228 GAGGACCTGC GAAATCCAGA ACAAAGCACA TCAGAAAAAG CAGTATTAAC TTCACAGAAA 4440
229
230 AGTAGTGAAT ACCCTATAAG CCAGAATCCA GAAGGCCTNT CTGCTGACAA GTTTGAGGTG 4500
231
232 TCTGCAGATA GTTCTACCAG TAAAAATAAA GAACCAGGAG TGGAAAGGTC ATCCCCTTCT 4560
233
234 AAATGCCCAT CATTAGATGA TAGGTGGTAC ATGCACAGTT GCTCTGGGAG TCTTCAGAAT 4620
235
236 AGAAACTACC CACCTCAAGA GGAGCTCATT AAGGTTGTTG ATGTGGAGGA GCAACAGCTG 4680
237
238 GAAGAGTCTG GGCCACACGA TTTGACGGAA ACATCTTACT TGCCAAGGCA AGATCTAGAG 4740
239
240 GGAACCCCTT ACCTGGAATC TGGAATCAGC CTCTTCTCTG ATGACCCTGA ATCTGATCCT 4800
241
242 TCTGAAGACA GAGCCCCAGA GTCAGCTCGT GTTGGCAACA TACCATCTTC AACCTCTGCA 4860
243
244 TTGAAAGTTC CCCAATTGAA AGTTGCAGAA TCTGCCCAGA GTCCAGCTGC TGCTCATACT 4920
245
246 ACTGATACTG CTGGGTATAA TGCAATGGAA GAAAGTGTGA GCAGGGAGAA GCCAGAATTG 4980
247
248 ACAGCTTCAA CAGAAAGGGT CAACAAAAGA ATGTCCATGG TGGTGTCTGG CCTGACCCCA 5040
249
250 GAAGAATTTA TGCTCGTGTA CAAGTTTGCC AGAAAACACC ACATCACTTT AACTAATCTA 5100
251
252 ATTACTGAAG AGACTACTCA TGTTGTTATG AAAACAGATG CTGAGTTTGT GTGTGAACGG 5160
253
254 ACACTGAAAT ATTTTCTAGG AATTGCGGGA GGAAAATGGG TAGTTAGCTA TTTCTGGGTG 5220
255
256 ACCCAGTCTA TTAAAGAAAG AAAAATGCTG AATGAGCATG ATTTTGAAGT CAGAGGAGAT 5280
257
258 GTGGTCAATG GAAGAAACCA CCAAGGTCCA AAGCGAGCAA GAGAATCCCA GGACAGAAAG 5340

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION *US/08/348,824A*

DATE: 04/07/95
TIME: 20:13:42

INPUT SET: S3186.raw

Line	Error	Original Text

```
1154        7861 tttaggttca gggatgtatg tgcagtttgt tatataggta aacacacgac ttgggatttg
1155        7921 gtgtatagat ttttttcatc atccgggtac taagcatacc ccacagtttt ttgtttgctt
1156        7981 tctttctgaa tttctccctc ttcccacctt cctccctcaa gtaggctggt gtttctccag
1157        8041 actagaatca tggtattgga agaaacctta gagatcatct agtttagttc tctcatttta
1158        8101 tagtggagga aatacccttt ttgtttgttg gatttagtta ttagcactgt ccaaaggaat
1159        8161 ttaggataac agtagaactc tgcacatgct tgcttctagc agattgttct ctaagttcct
1160        8221 catatacagt aatattgaca cagcagtaat tgtgactgat gaaaatgttc aaggacttca
1161        8281 ttttcaactc tttctttcct ctgttcctta tttccacata tctctcaagc tttgtctgta
1162        8341 tgttatataa taaactacaa gcaaccccaa ctatgttacc taccttcctt aggaattatt
1163        8401 gcttgaccca ggtttttttt ttttttttt tggagacggg gtcttgccct gttgccagga
1164        8461 tggagtgtag tggcgccatc tcggctcact gcaatctcca actccctggt tcaagcgatt
1165        8521 ctcctgtctc aatctcacga gtagctggga ctacaggtat acaccaccac gcccggttaa
1166        8581 ttgaccattc catttctttc tttctctctt ttttttttt ttttttgaga cagagtcttg
1167        8641 ctctgttgcc caggctggag tacagaggtg tgatctcacc tctccgcaac gtctgcctcc
1168        8701 caggttgaag ccatactcct gcctcagcct ctctagtagc tgggactaca ggcgcgcgcc
1169        8761 accacacccg gctaattttt gtatttttag tagagatggg gtttcaccat gttggccagg
1170        8821 ctggtcttga actcatgacc tcaagtggtc cacccgcctc agcctcccaa agtgctggaa
1171        8881 ttacaggctt gagccaccgt gcccagcaac catttcattt caactagaag tttctaaagg
1172        8941 agagagcagc tttcactaac taaataagat tggtcagctt tctgtaatcg aaagagctaa
1173        9001 aatgtttgat cttggtcatt tgacagttct gcatacatgt aactagtgtt tcttattagg
1174        9061 actctgtctt ttccctataG TGTGGGAGAT CAAGAATTGT TACAAATCAC CCCTCAAGGA
1175        9121 ACCAGGGATG AAATCAGTTT GGATTCTGCA AAAAAGggta atggcaaagt ttgccaactt
1176        9181 aacaggcact gaaaagagag tgggtagata cagtactgta attagattat tctgaagacc
1177        9241 atttgggacc tttacaaccc acaaaatctc ttggcagagt tagagtatca ttctctgtca
1178        9301 aatgtcgtgg tatggtctga tagatttaaa tggtactaga ctaatgtacc tataataaga
1179        9361 ccttctTgta actgattgtt gccctttcgc ttttttttt gtttgtttgt ttgttttttt
1180        9421 ttgagatggg gtctcactct gttgcccagg ctggagtgca gtgatgcaat cttggctcac
1181        9481 tgcaacctcc acctccaaAg gctcaagcta tcctcccact tcagcctcct gagtagctgg
1182        9541 gactacaggc gcatgccacc acaCccggTt aatttttgt ggttttatag agatggggtt
1183        9601 tcaccatgtt accgaggctg gtctcaaact cctggactca agcagtctgc ccacttcagc
1184        9661 ctcccaaagt gctgcagtta caggcttgag ccactgtgcc tggcctgccc tttactttta
1185        9721 attggtgtat ttgtgtttca tcttttacct actggttttt aaatataggg agtggtaagt
1186        9781 ctgtagatag aacagagtat taagtagact taatggccag taatctttag agtacatcag
1187        9841 aaccagtttt ctgatggcca atctgctttt aattcactct tagacgttag agaaataggt
1188        9901 gtggtttctg catagggaaa attctgaaat taavvvvvvv vvvvvvgatc ctaagtggaa
1189        9961 ataatctagg taaataggaa ttaaatgaaa gagtatgagc tacatcttca gtatacttgg
1190       10021 tagtttatga ggttagtttc tctaatatag ccagttggTt gatttccacc tccaaggtgt
1191       10081 atgaagtatg tatttttta atgacaattc agtttktgag taccttgtta tttttgtata
1192       10141 tttcaGCTG CTTGTGAATT TTCTGAGACG GATGTAACAA ATACTGAACA TCATCAACCC
1193       10201 AGTAATAA-T GATTTGAACA CCACTGAGAA GCGTGCAGCT GAGAGGCATC
1194   CAGAAAAGTA
1195       10261 TCAGGGTAGT TCTGTTTCAA ACTTGCATGT GGAGCCATGT GGCACAAATT ACTCATGCCA
1196       10321 GCTCATTACA GCATGAGAAC AGCAGTTTAT TACTCACTAA AGACAGAATG AATGTAGAAA
1197       10381 AGGCTGAATT CTGTAATAAA AGCAAACAGC CTGGCTTAGC AAGGAGCCAA
1198   CATAACAGAT
1199       10441 GGGCTGGAAG TAAGGAAACA TGTAATGATA GGCGGACTCC CAGCACAGAA
1200   AAAAAGGTAG
1201       10501 ATCTGAATGC TGATCCCCTG TGTGAGAGAA AAGAATGGAA TAAGCAGAAA
1202   CTGCCATGCT
1203       10561 CAGAGAATCC TAGAGATACT GAAGATGTTC CTTGGATAAC ACTAAATAGC
1204   AGCATTCAGA
1205       10621 AAGTTAATGA GTGGTTTTCC AGAAGTGATG AACTGTTAGG TTCTGATGAC TCACATGATG
1206       10681 GGGAGTCTGA ATCAAATGCC AAAGTAGCTG ATGTATTGGA CGTCCTAATT
```

Est format

#3

730 — 101
130 — 105





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
08/348,824	11/29/94	SHATTUCK-EIDENS	D 246841093474

03P1/0104
VENABLE BAETJER HOWARD & CIVILETTI
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

DATE MAILED: 0000 01/04/95

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 $ 860.

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 to complete the basic filing fee.

2. ☐ Additional claim fees of $_____ 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 items omitted at time of execution.

   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 to the oath or declaration is: ☐ missing; ☐ a reproduction; ☐ 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 for returned checks. (37 CFR 1.21(m)).

9. ☐ Your filing receipt was mailed in error because 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.

090 BA 03/10/95 08348824 1 101 730.00 CK
090 BA 03/10/95 08348824 1 105 130.00 CK

11. ☐ Other.

Direct the response and any questions about this notice to B. DAVENPORT Application Processing Division, Special Processing and Correspondence Branch (703) 308-1202.

A copy of this notice MUST be returned with the response.

GeneDX 1035, pg. 189

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.
08/348,824	11/29/94	SHATTUCK-EIDENS D	248841093474

18M2/0722
VENABLE BAETJER HOWARD & CIVILETTI
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

REES, EXAMINER	
ART UNIT	PAPER NUMBER
1807	5

DATE MAILED:
07/22/96

Please find below a communication from the EXAMINER in charge of this application.

Commissioner of Patents.

GeneDX 1036, pg. 190

Office Action Summary

Application No.	Applicant(s)
08/348,824	SHATTUCK-EIDENS ET AL.
Examiner	**Group Art Unit**
Dianne Rees	1807

☒ Responsive to communication(s) filed on *Mar 6, 1996* .

☐ 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) *1-18* is/are pending in the application.

Of the above, claim(s) _____ is/are withdrawn from consideration.

☐ Claim(s) _____ is/are allowed.

☒ Claim(s) *1-18* 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.

Claims 1-7, and 9-18 are rejected under 35 U.S.C. § 112, first paragraph, as the disclosure is enabling only for claims directed to defined sequences, and to methods for detection of susceptibility to breast and ovarian cancer caused by alterations at specific target sequences within the BRCA1 gene limited to using assays which employ allele specific probes. See M.P.E.P. §§ 706.03(n) and 706.03(z).

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 isolated DNA sequences "consisting essentially of DNA encoding for a mutant BRCA1 polypeptide said DNA selected from the group consisting of mutations set forth in Table 14. Further claims are drawn to methods of "diagnosing a polymorphism associated with predisposition for breast and ovarian cancer in a human comprising detecting a germline alteration of the wild type BRCA1 gene or its expression products in a human sample, said alteration indicating a predisposition to said cancer and selected from the group consisting of the mutations set forth in Tables 11, 12, 12A and 14. "

The product claims thus encompass sequences which differ in an unspecified manner from those disclosed in Table 14. As most of the mutations in Table 14 are a single base change, the sequences may encompass additional mutations to the ones disclosed, different mutations at the same sites as those

disclosed in Table 14 and might even encompass wild type sequences. The sequences are disclosed as having utility in a specific diagnostic test to determine predisposition for breast and ovarian cancer.

A variety of assays are set forth and claimed to detect polymorphisms including detecting BRCA 1 mRNA, such as by hybridization with a BRCA1 gene probe, by 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, and in situ hybridization. The claims are thus drawn to the detection of specific polymorphisms but the majority of assays claimed provide only general data which will allow one to determine that there is "a polymorphism" without any predictably certainty as to what the nature of the polymorphism is (as discussed above the product claims are further ambiguous as to what the nature of the polymorphism even is). It is highly unpredictable that any unselected polymorphism will be enabled as a screening tool, as recited by the claims- "said alteration indicating a predisposition to said (breast and ovarian) cancer". The large size of the BRCA1 gene makes it a target for many mutations and there is no apparent clustering for these mutations or

association with the severity of breast or ovarian cancer (See Shattuck Eidens et al. JAMA 273:535541, 1995). It is further likely that many polymorphisms in fact will be innocuous and not predispose one to any risk. It is not predictable that any given polymorphism which is linked to breast cancer will have utility as a screening tool for ovarian cancer as although the art teaches that there is a higher overall rate of ovarian cancer in individuals with BRCA1 mutations, 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). 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." Thus detection assays which are not geared to the detection of specific mutations (those for which these criteria have been met) will not predictably provide useful markers for risk of breast cancer of ovarian cancer.

It is therefore the position of the examiner that it would require undue experimentation to make and use the isolated DNA sequences as broadly claimed.

Claim Rejections - 35 USC § 112

3. Claims 1-18 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 "An isolated DNA sequence consisting essentially of DNA encoding for a mutant BRCA1 polypeptide " renders the claim vague and indefinite in that it is unclear how "essentiallY" is defined, and therefore it is not clear what the metes and bounds of the DNA sequences being claimed are. Further the recitation of "said DNA selected from the group consisting of mutations" in that it is unclear how the "sequences" may be selected from a group of mutations per se. The claim might be amended to recite --said DNA selected from a group of DNA sequences consisting of mutations--.

b) Claim 2 is indefinite in the recitation of "A nucleic acid probe complementary to human altered BRCA1 sequences" in that it is not clear what the extent of complementarity or

specificity of these probes are. Do the probes distinguish mutant vs. wild type sequences for example? Does complementary refer to full complementarity? The claim should provide some functional language to define the probes or more specifically recite the sequences encompassed by the probes.

c) Claim 3 is indefinite in the recitation of "the wild type BRCA1 gene" as the term lacks proper antecedent basis in the claim.

d) In claim 6, 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".

e) Claim 8 is indefinite in that the claim recites --wherein the alteration of the wild type BRCA 1 gene by hybridization with an allele specific probe" as the claim does not recite what the probe is to be hybridized to.

d) Claim 10 is indefinite in the recitation that "the alteration of the wild type BRCA1 gene is detected by amplifying all or part of the BRCA1 gene for a specific BRCA1 allele-- in that it is unclear if the claim intends that only the mutant allele is amplified or if both wild type or mutant allele may be amplified in the method. Clarification is requested.

e) Claim 14 is indefinite in that it is unclear where the non-wild type sequences are derived from (are these allele specific oligos or simply nucleic acids isolated from affected individuals?).

Provisional 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).

A statutory type (35 U.S.C. 101) double patenting rejection can be overcome by canceling or amending the conflicting claims so they are no longer coextensive in scope. The filing of a terminal disclaimer cannot overcome a double patenting rejection based upon 35 U.S.C. 101.

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).

Claims 1-18 provisionally rejected under 35 U.S.C. 101 as claiming the same invention as that of claims 1-18 of copending application Serial No. 08409305. This is a *provisional* double patenting rejection since the conflicting claims have not in fact been patented.

Provisional Obviousness Double Patenting

Claims 3-18 are provisionally rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1-16 of copending application Serial No. 08483553. Although the conflicting claims are not identical, they are not patentably distinct from each other because the methods recited in 08348824 meet all of the limitations of claims 1-19 in 08483553.

Claims 3-18 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 08348824 meet all of the limitations of claims 1-28 of 08487002.

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).

5. 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 18, 1996

W. GARY JONES
SUPERVISORY PATENT EXAMINER
GROUP 1800
7/19/96

Notice of References Cited

Application No.	Applicant(s)		
08/348,824	SHATTUCK-EIDENS 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
*	N	WO 95/19369		PCT	HOLT ET AL.	---	
	O						
	P						
	Q						
	R						
	S						
	T						

NON-PATENT DOCUMENTS

*		DOCUMENT (Including Author, Title, Source, and Pertinent Pages)	DATE
*	U	DORLAND'S MEDICAL DICTIONARY cited in 08487002	
*	V	GOLDBERG ET AL. CLINICAL CHEMISTRY 39: 2360-2374 " "	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).)

Notice of References Cited

Application No.	Applicant(s)		
08/348824	SHATTUCK-EIDENS 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 cited in 08487002	1990
x	V	BOYD, NATURE GENETICS 9: 335-336 " "	1995
x	W	MERAJVER ET AL, NATURE GENETICS 9;439-443 " "	1995
x	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)		
08/348824	SHATTUCK-EIDENS 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
*	U	FRIEDMAN ET AL., NATURE GENETICS 8: 1-6 cited in 08487003	1994
*	V	GOLDGAR ET AL. J. OF NAT. CANCER INST, 86: 200-209 cited in 08487002	1994
	W		
	X		

Form PTO 948 (Rev. 10-94) U.S. DEPARTMENT OF COMMERCE - Patent and Trademark Office Application No. 8/348824

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) 11/29/94

A.___ not objected to by the Draftsperson under 37 CFR 1.84 or 1.152.

B.__✓__ 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) 7-8C
   - ___ 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)_____
   - ___ 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
   - ___ Views and enlarged view not labled separatly or properly. Fig(s) 7-8B,8C

   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)
    - ___ Lines, numbers & letters not uniformly thick and well defined, clean, durable, and black (except for color drawings). Fig(s) 1-6, 8B, 8C
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) 1-6
    - ___ 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) 6
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 WMS DATE 6/17/96

Applicant's Copy

GeneDX 1035, pg. 203

OTPE
SEP 10 1996
PATENT & TRADEMARK OFFICE

#6
9-25-96

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

OTPE
SEP 10 1996
PATENT & TRADEMARK OFFICE

In re: patent application of:

Donna M. SHATTUCK-EIDENS et al.

Serial No. 08/348,824 Examiner: D. Rees

Filed: 29 November 1994 Group Art Unit: 1807

For: 17Q-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE MUTATIONS

SEP 13 1996
GROUP 1807

EXPRESS ABANDONMENT

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

Applicants hereby expressly abandon the above-identified application, as of the date given hereinbelow.

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. 24884-109347-03
Dated: 09 September 1996

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.
08/348,824	11/29/94	SHATTUCK-EIDENS D	248941093-74

18M2/1015
VENABLE BAETJER HOWARD & CIVILETTI
1201 NEW YORK AVENUE NW
SUITE 1000
WASHINGTON DC 20005

EXAMINER	
REES,D	
ART UNIT	PAPER NUMBER
1807	7

DATE MAILED: 10/15/96

NOTICE OF ABANDONMENT

This application is abandoned in view of:

1. ☐ Applicant's failure to respond to the Office letter, mailed _____ .
2. ☑ Applicant's letter of express abandonment which is in compliance with 37 C.F.R. 1.138.
3. ☐ Applicant's failure to timely file the response received _____ within the period set in the Office letter.
4. ☐ Applicant's failure to pay the required issue fee within the statutory period of 3 months from the mailing date of _____ of the Notice of Allowance.

   ☐ The issue fee was received on _____ .

   ☐ The issue fee has not been received in Allowed Files Branch as of _____ .

   In accordance with 35 U.S.C. 151, and under the provisions of 37 C.F.R. 1.316(b), applicant(s) may petition the Commissioner to accept the delayed payment of the issue fee if the delay in payment was unavoidable. The petition must be accompanied by the issue fee, unless it has been previously submitted, in the amount specified by 37 C.F.R. 1.17 (l), and a verified showing as to the causes of the delay.

   If applicant(s) never received the Notice of Allowance, a petition for a new Notice of Allowance and withdrawal of the holding of abandonment may be appropriate in view of Delgar Inc. v. Schuyler, 172 U.S.P.Q. 513.

5. ☐ Applicant's failure to timely correct the drawings and/or submit new or substitute formal drawings by _____ as required in the last Office action.
   ☐ The corrected and/or substitute drawings were received on _____ .
6. ☐ The reason(s) below.

STEPHANIE W. ZITOMER
PRIMARY EXAMINER
GROUP 1800

PTO-1432 (REV. 5-83)

PTO/SB/66 (11-96)
App 1 for use through 10/31/99. OMB 0651-0031
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

REQUEST FOR ACCESS OF ABANDONED APPLICATION UNDER 37 CFR 1.14(a)

PROCESSED BY
JAN 8 1998
FIII

In re Application of	
Shattuck - Fidelus et al.	
Application Number	**Filed**
S.N. 348,824	11/29/94
Group Art Unit	**Examiner**

Paper No. #8

Assistant Commissioner for Patents
Washington, DC 20231

I hereby request access under 37 CFR 1.14(a)(3)(iv) to the application file record of the above-identified ABANDONED application, which is: (CHECK ONE)

___ (A) referred to in United States Patent Number 5,693,473, column _____.

___ (B) referred to in an application that is open to public inspection as set forth in 37 CFR 1.11, i.e., Application No. _____, filed _____, on page _____ of paper number _____,

___ (C) an application that claims the benefit of the filing date of an application that is open to public inspection, i.e., Application No. _____, filed _____, or

___ (D) an application in which the applicant has filed an authorization to lay open the complete application to the public.

Please direct any correspondence concerning this request to the following address:

_____

_____

_____

Sarla C. Shroff
Signature

12/23/97
Date

SARLA C. SHROFF
Typed or printed name

FOR PTO USE ONLY

Approved by: _____ (initials)

Unit: _____

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

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.

Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

REQUEST FOR ACCESS OF ABANDONED APPLICATION UNDER 37 CFR 1.14(a)

In re Application of	
Donna M. Shattuck-Eidens	
Application Number	**Filed**
348824	11-29-94
Group Art Unit	**Examiner**

Paper No. 4

Assistant Commissioner for Patents
Washington, DC 20231

I hereby request access under 37 CFR 1.14(a)(3)(iv) to the application file record of the above-identified ABANDONED application, which is: (CHECK ONE)

___ (A) referred to in United States Patent Number 5693473, column _____.

___ (B) referred to in an application that is open to public inspection as set forth in 37 CFR 1.11, i.e., Application No. _____, filed _____, on page _____ of paper number _____,

___ (C) an application that claims the benefit of the filing date of an application that is open to public inspection, i.e., Application No. _____, filed _____, or

___ (D) an application in which the applicant has filed an authorization to lay open the complete application to the public.

Please direct any correspondence concerning this request to the following address:

_____

_____

_____

Calvin Perry
Signature

1-12-98
Date

Calvin Perry
Typed or printed name

FOR PTO USE ONLY

Approved by: _____
(Initials)

Unit: _____

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

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.

REQUEST FOR ACCESS OF ABANDONED APPLICATION UNDER 37 CFR 1.14(a)

RECEIVED
JAN 23 1998
MATRIX CUSTOMER SERVICE CENTER

In re Application of	
D. SHATTUCK-EIDENS ET AL	
Application Number	**Filed**
08/348,824	11/29/94
Group Art Unit	**Examiner**

Paper No. 10

Assistant Commissioner for Patents
Washington, DC 20231

I hereby request access under 37 CFR 1.14(a)(3)(iv) to the application file record of the above-identified ABANDONED application, which is: (CHECK ONE)

___ (A) referred to in United States Patent Number 5693473, column 1,

___ (B) referred to in an application that is open to public inspection as set forth in 37 CFR 1.11, i.e., Application No. _____, filed _____, on page _____ of paper number _____,

___ (C) an application that claims the benefit of the filing date of an application that is open to public inspection, i.e., Application No. _____, filed _____, or

___ (D) an application in which the applicant has filed an authorization to lay open the complete application to the public.

Please direct any correspondence concerning this request to the following address:

_____

_____

_____

TRINC THAE
Signature

1/22/98
Date

BERGER & ALTMAN
Typed or printed name

FOR PTO USE ONLY

Approved by: _____
(Initials)

Unit: _____

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

PTO/SB/68 (04-01)
Approved for use through 10/31/2002. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
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.

REQUEST FOR ACCESS TO AN APPLICATION UNDER 37 CFR 1.14(e)

RECEIVED
APR 30 2003
File Information Unit

In re Application of	
Application Number: 08/348 829	Filed: 11-29-94
Art Unit	Examiner

Paper No. 11

Assistant Commissioner for Patents
Washington, DC 20231

1. ☒ I hereby request access under 37 CFR 1.14(e)(2) to the application file record of the above-identified ABANDONED Application, which is not within the file jacket of a pending Continued Prosecution Application (CPA) (37 CFR 1.53(d)) and is: (CHECK ONE)

☒ (A) referred to in:

United States Patent Application Publication No. _____, page _____, line_____,

United States Patent Number 5673423, column _____, line _____, or

an International Application which was filed on or after November 29, 2000 and which designates the United States, WIPO Pub. No. _____, page _____, line_ ____.

☐ (B) referred to in an application that is open to public inspection as set forth in 37 CFR 1.11(b) or 1.14(e)(2)(i), i.e., Application No._____, paper No. _____, page _____, line _____.

2. ☐ I hereby request access under 37 CFR 1.14(e)(1) to an application in which the applicant has filed an authorization to lay open the complete application to the public.

Signature

4-30-03
Date

Henry Dunk
Typed or printed name

FOR PTO USE ONLY	
Approved by: _____ (initials)	
Unit: _____	

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

PATENT APPLICATION FEE DETERMINATION RECORD

Effective October 1, 1994

Application or Docket Number: 34824

CLAIMS AS FILED - PART I

	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE	SMALL ENTITY FEE	OR	OTHER THAN SMALL ENTITY RATE	OTHER THAN SMALL ENTITY FEE
BASIC FEE				365.00	OR		730.00
TOTAL CLAIMS	18 minus 20 =	* —	x$11=		OR	x$22=	
INDEPENDENT CLAIMS	2 minus 3 =	* —	x38=		OR	x76=	
MULTIPLE DEPENDENT CLAIM PRESENT			+120=		OR	+240=	
			TOTAL		OR	TOTAL	730

* If the difference in column 1 is less than zero, enter "0" in column 2

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)

PACE DATA ENTRY CODING SHEET

U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

1ST EXAMINER DAVE W/ 12-30-94 DATE 12-30-94

2ND EXAMINER [illegible] DATE 3/31/95

APPLICATION NUMBER	TYPE APPL	FILING DATE MONTH	DAY	YEAR	SPECIAL HANDLING	GROUP ART UNIT	CLASS	SHEETS OF DRAWING
08/348824	1	11	29	94	0	1807	435	10

TOTAL CLAIMS	INDEPENDENT CLAIMS	SMALL ENTITY?	FILING FEE	FOREIGN LICENSE	ATTORNEY DOCKET NUMBER
19	2	0	860	N	248/41093477

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	08/308,104	PCT/		09	06	94
13	2	08/308,266	PCT/		09	02	94
13	2	08/221,291	PCT/		09	12	94
13	2	08/246,602	PCT/		05	20	94
			PCT/				

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 - 365-974

JOURNAL
OF THE NATIONAL CANCER INSTITUTE

ISSN 0027-8874
Volume 86
Number 3
Pages 161–248

February 2, 1994



Samuel Broder
Director, National Cancer Institute

Susan Molloy Hubbard
Director, International Cancer Information Center

Julianne Chappell
Chief, Scientific Publications Branch

EDITORIAL BOARD

Daniel C. Ihde
Editor-in-Chief

Susan G. Arbuck
Frank M. Balis
William J. Blot
Peter M. Blumberg
John D. Boice
Bruce A. Chabner
Richard L. Cysyk
Susan S. Ellenberg
Michael A. Friedman
Frank J. Gonzalez

Eric J. Seifter
Book Review Editor

Michael M. Gottesman
Peter Greenwald
Susan M. Hubbard
Hynda K. Kleinman
Barnett S. Kramer
Theodore S. Lawrence
Lance A. Liotta
Marc E. Lippman
Dan L. Longo
Douglas R. Lowy
Susan G. Nayfield
David L. Nelson
Kenneth Olden

Joost J. Oppenheim
John R. Ortaldo
Harvey I. Pass
Oliver W. Press
Alan S. Rabson
Robert H. Shoemaker
Richard M. Simon
Maryalice Stetler-Stevenson
Margaret A. Tucker
J. Paul Van Nevel
Noel S. Weiss

EDITORIAL ADVISORY BOARD

James O. Armitage
Bruce C. Baguley
Laurence H. Baker
William T. Beck
Clara D. Bloomfield
Benjamin Bonavida
George J. Bosl
Barry Brown
C. Norman Coleman
O. Michael Colvin
Thomas H. Corbett
Pelayo Correa
Stephen P. Creekmore
Johanna T. Dwyer
Merrill J. Egorin
Soldano Ferrone
Isaiah J. Fidler
Richard I. Fisher
David FitzGerald
Øystein Fodstad
Antonio Fojo
Rebecca Gelman
Lori J. Goldstein
Harvey M. Golomb
Eliezer Gorelik
Jean L. Grem
Arnold H. Greenberg
Maureen M. Henderson
Gloria H. Heppner
Ronald B. Herberman
E. Carmack Holmes
Waun Ki Hong
William J. Hoskins
Alan N. Houghton
James N. Ingle
David H. Johnson
Randall K. Johnson
V. Craig Jordan
John S. Kovach
Margaret L. Kripke
Mark G. Kris
Donald W. Kufe
Bernard Levin
Brian R. Leyland-Jones
Allen S. Lichter
Guy McClung
Frank L. Meyskens
Anthony B. Miller
Malcolm S. Mitchell
James J. Mulé
C. Kent Osborne
John J. O'Shea
David M. Ota
David F. Paulson
Henry C. Pitot
Igor B. Roninson
Edward A. Sausville
Thomas J. Sayers
David Schottenfeld
Herman A. J. Schut
Richard K. Severson
William R. Shapiro
Roy E. Shore
Paul M. Sondel
Patricia S. Steeg
Herman D. Suit
Sandra M. Swain
Mario Szmol
Raymond Tartle
Ian Tannock
Joel E. Tepper
J. Tate Thigpen
Peter R. Twentyman
Larry M. Weisenthal

A Large Kindred With 17q-Linked Breast and Ovarian Cancer: Genetic, Phenotypic, and Genealogical Analysis

*David E. Goldgar, Patty Fields, Cathryn M. Lewis, Thao D. Tran, Lisa A. Cannon-Albright, John H. Ward, Jeff Swensen, Mark H. Skolnick**

Background: **Mutation of a specific, but as yet unidentified, gene BRCA1 on chromosome 17q results in increased susceptibility to breast and ovarian cancer. It is important to know the effects of this gene in terms of the age-specific risks of these cancers and the potential interaction of this gene with other known risk factors.** ***Purpose:*** **We performed detailed studies on a large multigenerational family, in which there is known 17q-linked breast and ovarian cancer, in order to characterize the effects of the BRCA1 mutation on development of breast and ovarian cancer.** ***Methods:*** **Data from the Utah Population Database were used to identify a family (identified as K2082) with a cluster of premenopausal breast cancer and ovarian cancer at any age. Blood samples from 195 members of the family were obtained and these individuals were genotyped for a series of four chromosome 17q polymorphic markers. Information on reproductive history, cancer incidence and treatment, and lifestyle factors was collected on 72 women in the family by questionnaire or through contact with living relatives.** ***Results:*** **Odds in favor of linkage of breast and ovarian cancer in this family to the BRCA1 region of chromosome 17q are greater than 10^8 to 1. The estimated risks for breast or ovarian cancer because of the BRCA1 mutation in this family are 40% by age 50 years and 90% by age 70. No differences between affected and unaffected older BRCA1 gene carriers were observed for a number of known epidemiologic risk factors for these cancers. The gender of the parent from whom the mutant BRCA1 allele was inherited was significantly associated with phenotypic expression ($P = .04$). A recombinant which places BRCA1 distal to the marker Mfd191 was observed.** ***Conclusions:*** **Women with the BRCA1 mutation are at increased risk of developing breast and ovarian cancer. In our study population, the mutation appears to confer a lower risk of cancer at younger ages than found in previous studies. Continued interaction with family K2082 will be useful in longitudinal follow-up studies and in studies of the psychosocial implications of providing DNA diagnosis of BRCA1. [J Natl Cancer Inst 86:200–209, 1994]**

The localization to the long arm of chromosome 17 of a specific gene (BRCA1) conferring increased susceptibility to breast cancer (*1*), and the subsequent finding that ovarian cancer was also an expression of the BRCA1 locus (*2*), provide tremendous opportunities for detailed studies of the role of a specific gene in the development of these cancers. The large study conducted by the Breast Cancer Linkage Consortium (BCLC) estimated the degree of genetic heterogeneity and the age-specific penetrance of the BRCA1 locus using data from 214 families collected worldwide (*3*). Based on genotypings at a series of marker loci surrounding the BRCA1 locus, it was estimated that 45% of families with breast cancer only and virtually 100% of breast cancer families which also had at least one case of ovarian cancer were linked to 17q. The analysis of the 214 families by the BCLC localized BRCA1 to an approximately 14-centimorgan (cM) interval flanked by the markers Mfd15 (D17S250) and 42D6 (D17S588). Subsequently, the BRCA1 locus has been further localized to a narrower interval flanked by THRA1 and D17S181 (*4-6*).

Claus et al. (*7*) used family history data from the large population-based Cancer and Steroid Hormone (CASH) study of 4370 breast cancer probands to perform a segregation analysis of a hypothesized breast cancer susceptibility locus. This analysis favored an autosomal dominant locus with penetrances for breast cancer of 0.38 by age 40 years and 0.66 by age 60. These penetrance estimates most likely represent the average effects of a number of different loci as well as other familial effects. Analyses of the BCLC set demonstrated that the penetrance of the 17q-linked BRCA1 locus was significantly higher than the estimates from the CASH data set (*3*). However, the estimation of heterogeneity and penetrance in such a diverse collection of families as the BCLC must be viewed in light of the clear sampling bias toward submission of families with earlier onset of breast cancer to the consortium and the pooling of families in which different BRCA1 mutations may be segregating.

Excluding family history, a number of demographic and reproductive risk factors influence breast cancer risk (*8*). The strongest such factor is generally accepted to be age at first pregnancy, where risk of breast cancer is increased ap-

**Correspondence to:* David E. Goldgar, Ph.D., Genetic Epidemiology, 420 Chipeta Way, #180, Salt Lake City, UT 84108.

See "Notes" section following "References."

GeneDX 1035, pg. 215

proximately twofold in a woman whose first full-term pregnancy occurs after age 29 years compared with before age 20. Other reproductive factors such as low parity and early age at menarche also confer increased risk. Among postmenopausal women, late age at menopause and, to a lesser extent, obesity are also implicated. Similarly, a number of reproductive and hormonally related factors convey increased risk for ovarian cancer. Nulliparity and few pregnancies have consistently been related to higher risk for ovarian cancer. A clear relationship exists where increasing numbers of pregnancies decrease the risk for ovarian cancer (*9*). Early menarche and late menopause confer a moderately increased risk (*10*). In contrast to breast cancer, age at first pregnancy does not appear to have a large effect on ovarian cancer.

There are a number of intriguing questions that have not been addressed by previous studies. For example, do the known risk factors for breast and ovarian cancer in the general population affect the expression of the BRCA1 locus in families? Alternatively, is the effect of the BRCA1 locus so large that the relative effects of other risk factors are so small as to be undetectable? Also of interest is the extent to which the BRCA1 locus confers increased risk of other cancers, particularly in males.

In an attempt to address some of these questions and to circumvent many of the aforementioned difficulties in estimating penetrance, we have performed detailed studies on a single large kindred with 17q-linked breast and ovarian cancer. The members of this kindred comprise the descendants of a Utah pioneer who was born in Somersetshire, England, in 1812 and arrived in the Salt Lake Valley circa 1850. One hundred ninety-five members of this kindred have been genotyped for a set of highly polymorphic probes surrounding the BRCA1 locus, allowing us to follow a rare haplotype containing the BRCA1 mutation. Identifying BRCA1 gene carriers enables us to estimate the penetrance of the particular BRCA1 mutation present in this family, examine the interaction of BRCA1 expression with known demographic risk factors, and identify other cancer sites which may represent alternative expression of the BRCA1 locus. This family also provides a large number of meioses in a family whose linkage to 17q is beyond doubt. Thus, there is an increased likelihood of identification of one or more key recombination events which may provide further localization information to aid in the isolation of the BRCA1 locus.

Methods

Ascertainment of K2082

Kindred 2082 was ascertained in 1991 through the Utah Population Database (UPDB) as one of a series of families that met certain criteria for sampling. Specifically, we ascertained clusters of related individuals with premenopausal breast cancer or ovarian cancer at any age. The UPDB is composed of two key linked resources: the Utah Genealogy, consisting of over one million Utah descendants of approximately 10000 Mormon pioneers (*11*), and the Utah Cancer Registry (UCR), which has complete statewide incidence data since 1966 and now contains more than 113000 cases. Links were created between the Utah Genealogy and the UCR whenever unique identity could be established. Currently, over 45000 of the cancer cases in the Utah Cancer Registry and, more recently, the Idaho Cancer Registry, have been linked to the genealogical resource. The UPDB allows us to ascertain large cancer-prone kindreds from a well-defined population.

Kindred 2082 was originally ascertained from the UPDB because of the presence of four family members with ovarian cancer and two members with breast cancer before age 45 years. The pedigree was extended through the following sources: (*a*) cases of breast or ovarian cancer that were diagnosed out of state and reported by family members, (*b*) identification of family members not present in the genealogy but who were diagnosed with these cancers in the UCR, and (*c*) a recent update of the UPDB resource. These studies were approved by the Institutional Review Board of the University of Utah and informed consent was obtained from all study participants. Fig. 1 displays the abbreviated pedigree structure of the family. This includes all descendants of the founder necessary to connect the 47 women who descend from this individual with either breast or ovarian cancer (or both). It is important to recognize that 13 of these family members are also related through another individual because of the intermarriage of K2082 and a second, independently ascertained kindred, K2081; these family members have a K2081 grandfather or great-grandfather in common and a great-grandmother or grandmother who are sisters. The connections of K2082 with K2081 are denoted in Fig. 1. In fact, nine of these individuals were originally ascertained as part of K2081. The connection of these individuals to K2082 was only discovered after the linked haplotype in some members of this kindred was recognized and prompted additional genealogical investigation. Of note also are the two instances of consanguineous marriage within the family, although the closest relationship between married individuals is half first cousins.

Of the 30 individuals with breast cancer depicted in Fig. 1, 21 were recorded in either the Utah or Idaho Cancer Registry, three were verified through examination of medical records, one by death certificate, and the remaining five were verified based upon information supplied by a first-degree relative. Similarly, of the 20 women with ovarian cancer, 12 were verified through a cancer registry, one by medical records, one by death certificate, and the remaining six were based upon either self report (one woman), or from information supplied by a first-degree (three women) or second-degree (two women) relative.

Genotyping of 17q Markers

From 50 mL of blood, nuclear pellets were extracted from white blood cells (*12*) and DNA extracted with phenol and chloroform, precipitated with ethanol, and resuspended in Tris EDTA. For two deceased individuals whose genotype at the relevant 17q loci could not be inferred from those of their children, DNA was extracted from normal tissue sections sliced from tumor samples embedded in paraffin. All markers genotyped were short tandem repeat (STR) loci assayed by polymerase chain reaction (PCR) using standard procedures (*13*). Radiolabeled PCR products were electrophoresed on standard 6% polyacrylamide denaturing sequencing gels (*14*).

Linkage Analysis

Linkage to 17q was analyzed using the age-specific penetrance function derived from the analysis of the CASH study data (*7*) and implemented as 14 liability classes (seven affected, seven unaffected) in the LINKAGE package (*15*) version 5.1. A frequency of 0.003 was assumed for the mutant BRCA1 allele. Details of the implementation of this penetrance function and specific penetrance values can be found in reference (*3*). As in the BCLC analysis, women with ovarian cancer were given the same penetrance as women in the earliest onset breast cancer group (age <30 years). Men were assumed to be in the youngest unaffected liability class and thus were essentially regarded as being of unknown phenotype.

Selection and Localization of Genetic Markers Used in the Study

In order to assess possible linkage to chromosome 17q, family members were genotyped for four STR loci on chromosome 17q. These four markers

GeneDX 1035, pg. 216

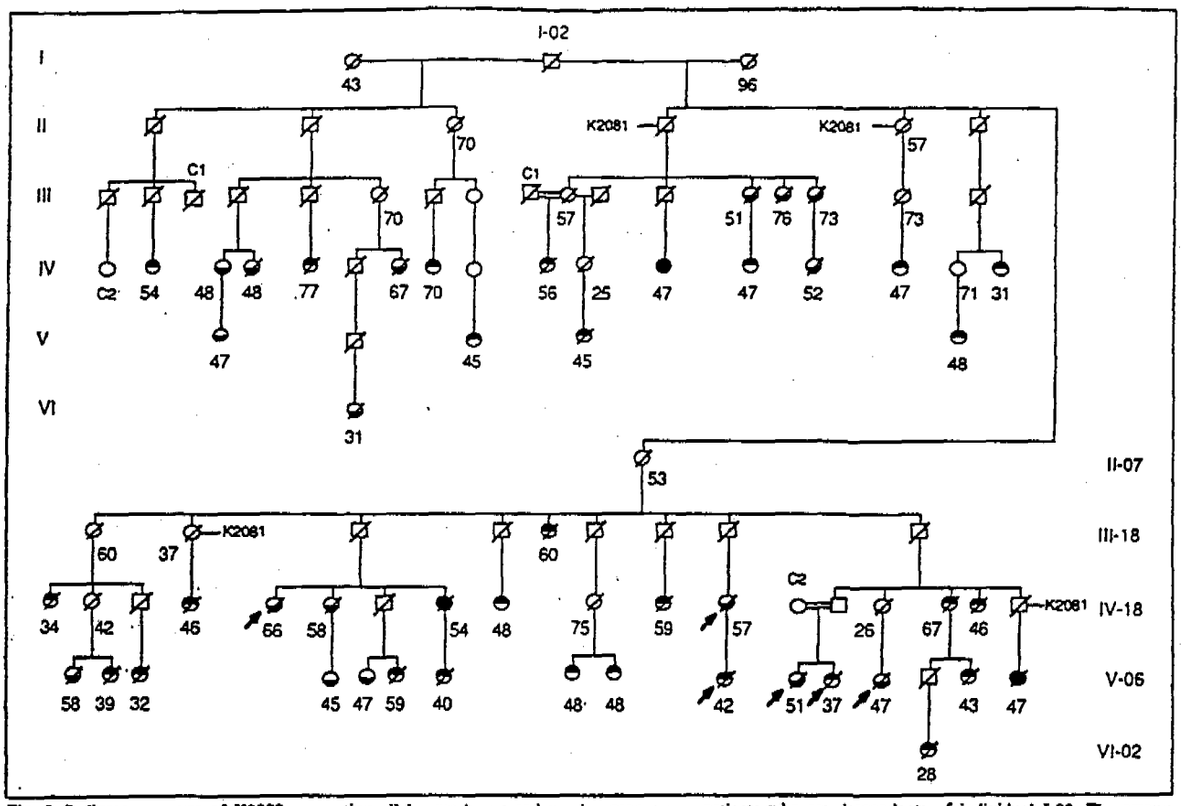


Fig. 1. Pedigree structure of K2082 connecting all known breast and ovarian cancer case patients who are descendants of individual I-02. The arrows designate the six probands leading to the initial ascertainment of K2082. Women diagnosed with breast cancer are denoted by the upper-half shaded symbol, ovarian cancer by the lower-half shaded symbol. Totally filled circles represent women diagnosed with both breast and ovarian cancer. Numbers under each symbol represent age at diagnosis, age at death, or age at last observation for affected, deceased, and living unaffected women, respectively. Pedigree numbering is sequential for each generation from upper left to lower left portion of the diagram; to aid in identification of specific individuals, the number of the first individual in each generation in the lower portion of the figure is shown. Symbols with slashes indicate individuals who are deceased.

(Mfd15, Mfd191, ED2, and Mfd188) were chosen to span the BRCA1 region in approximately 1- to 2-cM intervals. Although closer flanking markers exist, Mfd15 [D17S250 (*16*)] and Mfd188 [D17S579 (*17*)] were chosen because they are both highly polymorphic. Mfd191 (D17S776) is an unpublished STR marker made available to us by James Weber (Marshfield Research Foundation, Marshfield, Wisc.). ED2 is an STR locus that we identified in a yeast artificial chromosome (YAC A167E6 from the Washington University YAC library) that was selected for the presence of RNU2, a gene known to lie in the BRCA1 candidate region.

We have localized ED2 and Mfd191 to the region between Mfd15 and Mfd188. Independent individuals from the Centre d'Etude du Polymorphisme Humain (CEPH, Paris, France) (*18*) panel of families were used to estimate allele frequencies and marker heterozygosities. Table 1 gives details of the markers and two-point LOD scores between the markers and the breast/ovarian cancer phenotype. The LOD (logarithm of the odds) score is defined as the decimal logarithm of the ratio of the likelihood of the observed data under the hypothesis of linkage between the disease and marker loci, to the corresponding likelihood under the assumption of free

Table 1. Genetic markers used in the linkage and haplotype analysis of K2082

Marker*	Locus	No. of alleles observed	HET† (n)	LOD θ_{max}‡	f(LA)§	Ref. No. or primers
Mfd15	D17S250	11	0.82 (76)	4.15 (0.05)	0.15	(*12*)
Mfd191	D17S776	9	0.55 (44)	1.16 (0.15)	0.07	Weber J: personal communication, 1992 (GDB)
ED2	—	13	0.55 (80)	8.52 (0.00)	0.013	5′-CTAAGGAGGTTTCTCTGGAC-3′ 5′-TTCACAACTCAAGGTAAGATAGG-3′
Mfd188	D17S579	12	0.81 (72)	4.86 (0.05)	0.08	(*13*)

*For each marker, maximum LOD scores between the marker and the BRCA1 locus are reported.

†HET = observed heterozygosity of marker in independent CEPH individuals.

‡$LOD\theta_{max}$ = maximum LOD score observed between marker and BRCA1 locus. Values in parentheses = recombination fraction at which maximum was achieved.

§f(LA) = frequency of allele co-segregating with BRCA1 estimated in CEPH independents.

GeneDX 1035, pg. 217

recombination; a LOD score of 3.0 thus corresponds to odds of 1000:1 in favor of linkage. A maximum LOD score was achieved between the breast/ovarian cancer phenotype and the marker EDZ of 8.52, corresponding to odds in favor of linkage of greater than 10^8 to 1. To perform the analyses described here, it was necessary to order the two unlocalized markers with respect to the two established loci Mfd15 and Mfd188. A number of recombinant breakpoints in the panel (*18*) and in K2082 have proven to be reliable and consistent across a large set of markers on 17q. The relevant observations are shown in Table 2. For example, CEPH individual 13292-04 places both Mfd191 and ED2 proximal to Mfd188, while 13294-04 places both markers distal to THRA1.

BRCA1 Carrier Determination

To examine the effects of age and reproductive factors on BRCA1 expression, K2082 family members were classified as BRCA1 gene carriers (+), noncarriers (–), or unknown (U) status. Women were classified as carriers if they or their offspring had a haplotype containing the linked alleles at Mfd191, ED2, and Mfd188. Individuals displaying a recombinant haplotype in this region (e.g., ID "a" in Table 2) were classified as unknown, as were untyped individuals with no typed offspring. Table 3 reports the age distribution among BRCA1 carriers, noncarriers, and those of unknown haplotype status for each cancer phenotype. All individuals in the pedigree whose BRCA1 status could not be unequivocally determined were assigned weights corresponding to the probability of being a gene carrier based on distance to the nearest gene-carrying relative and any genotypic data available in offspring. To avoid confounding the penetrance analysis, all such determinations were made on the basis of genotypic data only; no phenotypic data were considered. In the majority of the statistical comparisons, only women whose BRCA1 status was unequivocal were used; the weighted gene-carrier probabilities were incorporated only into the analysis of penetrance.

Questionnaire and Genealogical Data

A questionnaire detailing demographic and social factors, reproductive history, oophorectomy and hysterectomy, and cancer history was administered to 72 women affected by or at risk for breast and ovarian cancer. Unfortunately, only eight questionnaires were completed by affected individuals (five breast cancer patients and three ovarian cancer patients) because most of the women were deceased by the time of ascertainment. For women who could not be interviewed, we attempted to gather

Table 2. Key recombinants used for fine mapping of new markers from the breast cancer region from CEPH families and K2082*

Family	ID No.	M/P	Mfd15	THRA1	Mfd191	ED2	Mfd188	42D6
CEPH								
13292	04	M	1	1	1	1	0	0
13294	04	M	1	0	0	—	0	—
13294	06	M	0	0	1	—	1	—
1333	06	M	0	0	1	—	1	1
K2082	a	M	1	—	1	0	0	0
K2082	V-18	M	0	—	0	1	1	1
K2082	b	P	0	—	0	1	1	1
K2082	c	M	0	—	0	0	1	1

*a = daughter of IV-19; b = daughter of IV-18; c = daughter of IV-25. M/P indicates a maternal (M) or paternal (P) recombinant. The numeral "1" indicates inherited allele is of grandpaternal origin, while "0" indicates grandmaternal origin. Pedigree references for K2082 refer to Fig. 1.

Table 3. Age and age at onset distribution by cancer phenotype and BRCA1 gene carrier status

Hap*	No.	Age at diagnosis, y					Mean	SD
		<40	40-49	50-59	60-69	≥70		
				Breast cancer†				
+	17	3	8	2	4	0	48.5	10.6
–	10	1	5	2	0	2	52.3	13.1
U	3	1	2	0	0	0	42.3	7.2
Total	30	5	15	4	4	2	49.1	11.3
EXP‡	6479	1.0	4.0	6.5	8.2	10.3	63.5	13.1
				Ovarian cancer				
+	14	0	6	6	1	1	54.4	8.0
–	1	1	0	0	0	0	32.0	—
U	5	0	2	1	1	1	58.0	13.7
Total	20	1	8	7	2	2	54.2	10.5
EXP‡	941	1.0	2.6	4.3	5.8	6.3	62.3	13.4

Hap*	No.	Age/age at death, y					Mean	SD
		<40	40-49	50-59	60-69	≥70		
				Unaffected				
+	34	15	12	4	2	1	42.0	12.7
–	42	10	10	9	6	7	50.7	15.6
U	38	19	11	4	1	3	41.7	13.1
Total	114	44	33	17	9	11	45.1	14.5

*BRCA1 status based on four-locus 17q haplotype: + = BRCA1 gene carrier; – = gene noncarriers; U = unknown status.
†Includes three breast cancers in women with both breast and ovarian cancer.
‡Expected frequency of observed K2082 cases based on age distribution of cases in the UPDB.

GeneDX 1035, pg. 218

information on reproductive history by contacting living first-degree relatives and, failing this, obtaining the information from the Utah Genealogy. The difficulty with the latter approach was that, in many instances, the most recent generation is not present in the database. Accordingly, the number of pregnancies was treated as a missing observation for any cases in which the genealogy showed no offspring.

Results

Estimation of BRCA1 Penetrance in K2082

With the exception of the two breast cancer probands who were required to be diagnosed before age 45 years according to our selection criteria, the ascertainment criteria and the decision to extend and/or genotype a particular branch of K2082 did not depend on the age at the onset of cancer. The two breast cancer probands who are second cousins and were selected because of early onset and whose 17q genotypes define the linked haplotype have been excluded from the analysis. To be totally unbiased, the penetrance analysis should include all other women who are descendants of the founder, I-02 in Fig. 1, who carry the susceptible haplotype, regardless of their cancer phenotype. Because of the nature of our sampling strategies, it was not possible to achieve this goal; however, we believe we have included the vast majority of individuals who are potential (50% risk) haplotype carriers in the penetrance estimation. Failure to include all haplotype carriers could result in a slight overestimation of BRCA1 penetrance in this family.

As a first step in exploring the age-specific penetrance of the 17q breast cancer susceptibility locus, both a Kaplan–Meier estimate of the age of onset distribution and a parametric model (log-normal) were constructed. In the Kaplan–Meier analysis, only women in K2082 who carry the susceptible haplotype with certainty were included. However, in the parametric analysis, women (both unaffected and affected) with breast and/or ovarian cancer who had unknown BRCA1 status were also included by weighting them by their probability of being a BRCA1 gene carrier. In the penetrance analysis, breast and ovarian cancer were considered as equivalent outcomes; women affected with either disease were classified as affected at their age at diagnosis. For the three women with both breast and ovarian cancer, the earliest age at diagnosis was used. Unaffected women were classified as censored at their current age or age at death, if deceased. The SAS (SAS Inc., Cary, N.C.) program LIFEREG was used to estimate the parameters of the assumed log-normal distribution. Fig. 2 shows the Kaplan–Meier step function and the estimated log-normal curve. For comparison, the estimated penetrance of the BRCA1 gene in breast/ovarian cancer families obtained from the analysis of the BCLC data set (*3*) is also shown.

Histology of 17q-Linked Breast and Ovarian Cancers

The distribution of histologic types among the breast and ovarian cancers was compared with that in the UPDB, which has data on 6479 breast cancer cases and 941 ovarian cancer cases. The expected distribution among histologic subtypes of the breast and ovarian cancer cases in K2082 was calculated and compared with the observed data. The expected numbers were calculated based upon the distribution of age at diagnosis of breast or ovarian cancer in the individuals in K2082, which was generally younger than that in the registry as a whole. No striking differences in the histopathology of cases in K2082 and breast/ovarian cancer cases from the general population were evident. There was a slight increase in lobular breast carcinomas among the K2082 cases (three cases observed versus 1.4 expected).

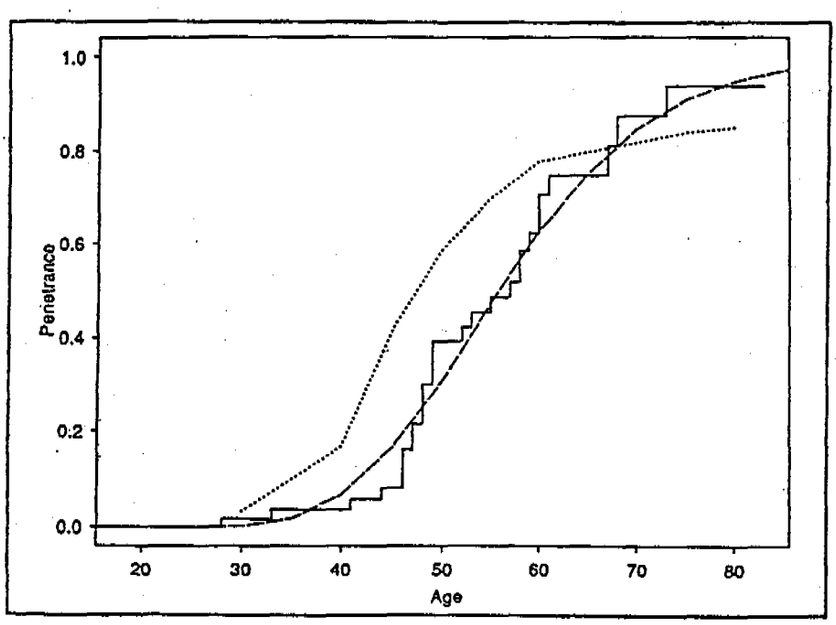


Fig. 2. Estimates of penetrance. Each graph represents cumulative risk of developing breast or ovarian cancer as a function of age. Dashed line: log-normal fit to data from known and possible gene carriers in K2082; solid line: Kaplan–Meier step function estimate using data from K2082 known BRCA1 carriers only; dotted line: penetrance estimates from BCLC data set (*3*).

GeneDX 1035, pg. 219

Incidence of Other Cancer Sites in K2082

Two methods were used to examine the incidence of other types of cancer in K2082. The first method examined all cancers in the UPDB among all descendants of the founder and compared the incidence of each cancer site with that found in the UPDB as a whole. The expectations were calculated based on the birth-year cohort, sex, and birthplace (Utah versus non-Utah) distribution of the descendants. The statistical significance of each site was assessed by computing the *P* value associated with the observed numbers of each type of cancer, assuming a Poisson distribution with parameter equal to the expected number. We have identified 81 cases of cancers among the 1611 descendants of individual I-02 (*see* Fig. 1) found in the UPDB who were born after 1875. This does not represent all cases of cancer among the descendants, of course, because individuals could have been diagnosed elsewhere or not be members of the Utah Genealogy; this is especially true among the younger generations. However, it does provide a relatively unbiased approach to finding excesses of other cancers that may represent alternative expressions of the BRCA1 locus. Although 81 cases of cancers were found (in 73 individuals) compared with 54 expected ($P = .0003$), most of this excess can be attributed to breast (20 observed cases versus eight expected [$P = .0003$]) and ovarian (10 observed cases versus 1.22 expected [$P<.0001$]) cancers. Excluding breast and ovarian cancers, 51 other cases of cancers were found compared with 44.6 expected, which is not statistically significant. Moreover, among 36 distinct sites examined, none were significant ($P<.05$) and only two, endocrine system ($P = .06$) and female genitals ($P = .054$), were found with $P<.10$. The disadvantage of this approach is that subtle effects of the BRCA1 locus may be diluted; only about one-eighth of the descendants eligible to be included in the cancer registry would be expected to be BRCA1 gene carriers.

The second approach to examining the occurrence of other cancers tabulates all cases of cancers found in individuals who have a probability of at least 0.5 of being a BRCA1 gene carrier as determined by the following factors: (*a*) known genotyped or inferred carriers of the identified susceptibility haplotype, (*b*) first-degree relatives of known BRCA1 gene carriers, and (*c*) cases of breast/ovarian cancer of unknown gene status. These cancers are shown in Table 4. This analysis examines more directly the presence of other cancers in BRCA1 carriers but, in contrast to the previous approach, does not provide any standard by which to compare these observations.

Comparison Between BRCA1 Gene Carriers and Noncarriers

In order to examine any possible effects of the BRCA1 mutation that may be manifest prior to the onset of cancer, female BRCA1 gene carriers (+) and noncarriers (–) were

Table 4. Additional cancers found in individuals in K2082 who have at least a 50% risk of being a BRCA1 gene carrier

(a) Cancers found in six unaffected gene carriers and five first-degree relatives who have unknown BRCA1 status

Cancer site or type	Sex	Age at diagnosis, y	Source of diagnosis*	Relationship†	ID No.‡
Lymphoma	Male	34	UCR	Nephew	IV-04
Throat	Male	57	UCR	Brother	V-20
Lymphoma	Male	64	UCR		
Lymphocytic leukemia	Male	68	UCR	Uncle	V-7
Prostate	Male	58	UCR	Brother	IV-23
Throat	Male	38	Fam	Son	IV-18
Brain	Female	21	Fam	Daughter	IV-18
Pituitary	Female	31	UCR	Daughter	IV-10
Thyroid	Male	83	UCR	Brother	III-15
Prostate	Male	69	UCR	Brother	IV-14
Colon	Female	43	Fam	Daughter	IV-4

(b) Cases of breast or ovarian cancer (excluding known sporadic cases)

Cancer site or type	Age at diagnosis, y	Case site	Age, y	Source of diagnosis*	ID No.‡
Ovarian	48	Breast	46	UCR	IV-2
Pituitary	58			UCR	IV-2
Granulocytic leukemia	68	Ovarian	58	UCR	V-21
Meningioma	56	Breast	59	UCR	V-16
Ovarian	47	Breast	45	Fam	V-6
Breast	55	Ovarian	54	UCR	IV-27
Second breast	76	Breast	60	UCR	IV-23
Lymphoma	68	Breast	68	UCR	IV-19

*UCR = Utah Cancer Registry; Fam = family report.

†Family relationship to closest breast or ovarian cancer case patient who is a BRCA1 carrier. By definition, all individuals are first-degree relatives of a known BRCA1 carrier.

‡ID number as shown in Fig. 1. In (*a*) this is the ID of the cancer case patient to which the individual is related; in (*b*) it is the ID of the case patient with the additional cancer.

GeneDX 1035, pg. 220

compared for a number of variables related to fertility and reproduction. No statistically significant differences between the two groups were found for the following variables: (*a*) age at menarche (12.6 years [+] versus 13.0 years [–]), (*b*) age at first pregnancy (22.9 years versus 22.9 years), or (*c*) parity (4.4 years versus 4.3 years). Of the 137 pregnancies in BRCA1 gene carrier women who completed the questionnaire, 24 (18%) resulted in miscarriage. Similarly among the non-gene carriers, there were 28 miscarriages from 186 pregnancies (15%). There was a suggestion of an increased incidence of oophorectomy in gene carriers who did not have ovarian cancer (nine of 29) compared with non-gene carriers (six of 35), although this difference was not statistically significant ($P = .14$).

Effect of Risk Factors on Expression of BRCA1

To examine the effects of known risk factors for breast and ovarian cancer in this family, two sets of analyses were performed. The first analysis compared the following three groups of BRCA1 gene carriers in K2082: 1) unaffected women over 40 years of age, 2) women affected with breast cancer, and 3) women affected with ovarian cancer. Women with both breast and ovarian cancer were included in the ovarian category. Because most of the relevant risk factors have changed over the last century and because affected women were more likely to have been born earlier than unaffected BRCA1 gene carriers, women were divided into three birth cohorts: born before 1920, between 1920 and 1940, and after 1940. Comparisons among groups for age at first live birth, parity, age at last live birth, age at menarche, number of pregnancies, and weight (defined as "usual weight") were performed using a two-factor (cohort and affectation group) analysis of variance allowing comparisons among groups to be adjusted for cohort trends. The comparison between affected groups was subdivided into two orthogonol single degree of freedom contrasts: affected women versus unaffected women and breast cancer versus ovarian cancer. The results of these analyses are shown in Table 5. With one exception, no significant differences for any of the risk factors for either of the comparisons were found. Women affected with breast cancer reported a statistically significant higher weight ($P = .03$) than women with ovarian cancer, but because of the limited interview data, this result is based on only three cases in each group.

A second set of analyses addressed the question of the relationship between age at diagnosis of the 30 BRCA1 gene carriers affected with breast or ovarian cancer and the above identified risk factors. Initially, Pearson correlation coefficients were calculated between age at first pregnancy, parity, age at last pregnancy, and age of onset. The strongest correlations with age of onset were with age at last pregnancy ($r = .57$, $P = .001$) and age at first pregnancy ($r = .52$, $P = .003$); no significant correlation was found with number of live births ($r = .30$, $P = .11$). It was surprising to us that these correlations were positive; the epidemiologic evidence would have predicted a negative correlation. We repeated these correlations, correcting for temporal trends by computing partial correlations adjusting for year of birth. When year of birth was accounted for, all correlations between age at diagnosis and other risk factors were not significant.

For each BRCA1 gene carrier or probable gene carrier (untyped affected first-degree relative of a known gene carrier), we recorded whether the mutant BRCA1 allele was inherited from her mother or father. The frequency of paternal origin of the BRCA1 allele was 11/16 (69%) in women with ovarian cancer, 7/17 (41%) in women with breast cancer only, and 5/19 (26%) in unaffected women over the age of 40 ($P = .04$).

Table 5. Comparison of affected and unaffected BRCA1 gene carriers for reproductive factors*

	All BRCA1 gene carriers							
	Breast cancer (n = 16)		Ovarian cancer (n = 14)		Unaffected† (n = 19)			
	Mean	SD	Mean	SD	Mean	SD	P_{AU}‡	P_{BO}§
Parity	4.6	2.0	4.7	1.5	4.5	1.8	0.33	0.66
Age at first pregnancy, y	22.8	2.9	23.9	3.8	23.0	4.1	0.78	0.65
Age at last pregnancy, y	32.7	6.6	34.6	6.1	32.7	5.3	0.68	0.98
Age or age at diagnosis, y	48.1	10.8	54.4	8.0	51.0	9.5	0.07	0.21
	Interviewed BRCA1 gene carriers							
	Breast cancer (n = 3)		Ovarian cancer (n = 3)		Unaffected† (n = 14)			
	Mean	SD	Mean	SD	Mean	SD	P_{AU}‡	P_{BO}§
Age at menarche, y	13.0	1.7	12.3	0.6	12.5	1.1	0.89	0.37
Gravid	6.3	2.1	5.0	1.0	4.5	1.9	0.27	0.36
Weight	197	25	139	11	157	29	0.38	0.029

*All comparisons are adjusted for cohort effects.
†Includes only unaffected gene carriers 40 years or older.
‡Significance level for testing differences between affected and unaffected gene carriers.
§Significance level for testing differences between breast and ovarian cases.

GeneDX 1035, pg. 221

Sporadic Cases

As shown in Table 3, ten women with breast cancer could be shown not to have inherited the linked haplotype found in the majority of other cases in the family. A single sporadic (i.e., not attributed to BRCA1) ovarian cancer case was identified. Interestingly, the age at diagnosis of this woman was 31 years, 15 years younger than that of the youngest 17q-linked BRCA1 ovarian cancer case. The mean age of diagnosis among the ten apparently sporadic breast cancer case patients was age 52 years compared with 48 years among the BRCA1 carrier case patients (P = .44 by Student's t test). As shown previously, the number of sporadic cases observed corresponds closely to the number expected in a group of descendants of this size (eight cases of breast cancer and one case of ovarian cancer), based on demographically matched figures in the UPDB.

Recombinants in the Region

As shown in Table 2, there were a number of recombination events in the region between Mfd15 and Mfd188 in this family. However, only one of these recombination events took place in an affected individual. This woman (V-18 in Fig. 1) was affected with ovarian cancer at age 45 years and has inherited the linked haplotype at ED2 and Mfd188 from her mother, who was diagnosed with ovarian cancer at age 58 years. She did not inherit the linked allele at either Mfd191 or Mfd15. This provides convincing evidence that the BRCA1 locus is located distal to Mfd191.

Conclusions

K2082, to our knowledge, has the largest LOD score of any 17q-linked family yet identified. Because of delayed age at onset, sex-specificity, and rapid mortality, identifying large kindreds segregating breast and/or ovarian cancer and demonstrating their linkage to 17q can be quite difficult. The expansion of this family to its current size from the initial six probands would have been impossible without the aid of the UPDB. This resource allowed us to identify cases of breast and ovarian cancer in more remote branches of the family and connect these branches to the rest of the kindred. The availability of a dense map of highly polymorphic marker loci surrounding the BRCA1 locus was a key factor in identifying a rare haplotype cosegregating with breast and ovarian cancers in K2082; this in turn allowed the detection of linkage, and differentiation between BRCA1 cases and sporadic cases, in the absence of direct segregation information.

The mean ages at diagnosis of breast or ovarian cancer of the known affected BRCA1 gene carriers in K2082 were 48 and 54 years for breast and ovarian cancer, respectively. Although substantially lower than what one would expect in the population as a whole, these ages at diagnosis are substantially higher than those associated with 17q-linked breast/ovarian cancer in previous studies (*6*). For example, in the two largest breast/ovarian cancer families reported by Feunteun et al. (*6*), the median age at diagnosis of the cases was substantially younger than in K2082. In one family with a 17q LOD score of 3.03, the median ages at diagnosis were 39.5 years and 43 years for the 14 breast cancer patients and the 10 ovarian patients, respectively. Similarly, the second family (LOD = 2.15) had median ages of 31 years and 44 years for the nine breast and three ovarian cancer cases, respectively. This compares with the median age at diagnosis of 48 years of age for the breast cancer cases and 52 years for the ovarian cancer cases in K2082. It is unclear if K2082 is atypical with respect to age at onset or the proportion of cases in which the BRCA1 is expressed as ovarian rather than breast cancer. The selected nature of the pedigrees that are available for comparison makes such determinations difficult. The onset of breast cancer in K2082 appears, however, to be substantially later than most of the linked cases from the BCLC dataset as a whole. This is seen from comparison of the penetrance functions in Fig. 2 in which the penetrance of the 17q mutation in K2082 is lower than the BCLC data until about age 65 years and is slightly higher at older ages, achieving a penetrance by age 80 years for expression as breast or ovarian cancer of 0.95, compared with 0.85 in the BCLC data set. Whether this is a manifestation of preferential submission of early-onset families to the consortium, a consequence of the use of the CASH model which makes linkage difficult to assess in small later-onset families (*4*), or represents a real difference caused by differences in the underlying gene defect in this family is unclear at present. The answer to this question awaits accumulation of more data or, more likely, the identification and cloning of the BRCA1 gene and characterization of the mutations which are found in specific pedigrees.

There were no obvious effects of BRCA1 other than breast and ovarian cancer. There appeared to be a slight excess of cancers of the endocrine system using the observed/expected approach. This was borne out by the presence of two pituitary cancers and a thyroid cancer in individuals who had at least 50% risk of being BRCA1 gene carriers. Other cancers with multiple occurrences in such individuals are non-Hodgkin's lymphoma, leukemia, and brain, throat, and prostate cancers. Non-Hodgkin's lymphomas occurred in three probable BRCA1 carriers: two males and one female who also had late onset breast cancer. It is difficult to interpret this finding in the absence of expected values; however, three cases among the branches of the family which we have studied in detail seems higher than one would expect by chance. Although prostate cancer is extremely common, it is noteworthy that both such case patients were diagnosed at a relatively early age (58 years and 68 years).

No differences in fertility were found between BRCA1 gene carriers and noncarriers. Age at first pregnancy, parity, and miscarriage rate were comparable in the two groups. Among women unaffected with ovarian cancer there was an increased frequency of oophorectomy among gene carriers. Although not statistically significant (P = .14), this observation could indicate that the BRCA1 locus produces some precancerous effects, such as benign cysts. On the other

GeneDX 1035, pg. 222

hand, this finding could be explained in part by BRCA1 gene carriers being more closely related to aggregations of ovarian cancer and thus more cancer-phobic, leading to prophylactic oophorectomy. When queried as to the reason for these surgeries, the majority reported that the oophorectomy was performed in conjunction with hysterectomy, with heavy bleeding the most frequently reported reason for surgery. Three oophorectomies were clearly for prophylactic reasons, while three more were due to other causes, specifically ovarian cysts, a benign ovarian tumor, and hyperplasia. Among these three latter cases, one woman was a certain BRCA1 gene carrier and two women had unknown BRCA1 status. By contrast, the frequency of hysterectomy was equal for the two groups.

We have observed considerable variation in the expression of the BRCA1 locus in K2082. There are at least three gene carriers who have remained cancer-free well into their sixties. Among affected women, after adjusting for the probands which led to the ascertainment of the family, breast and ovarian cancer cases were observed in roughly equal proportions. Moreover, within each cancer type, ages at diagnosis of obligate gene carriers ranged from 28 years to 68 years for breast cancer and 46 years to 73 years for ovarian cancer. It seems unlikely that this degree of variation in expression could be entirely due to chance; yet, when we examined the most likely sources for this variation (i.e., age at first pregnancy, parity, etc.) no effect was found when we controlled for cohort differences. Initially, significant correlations between age at diagnosis of cancer and age at first and last pregnancy were observed. However, when the effect of birth year on both measures is accounted for, these correlations become nonsignificant. A difference in weight between breast cancer case patients and ovarian cancer case patients was observed, but given that it was based on only three cases in each group, more data are obviously needed.

Although our sample size of affected women and older, unaffected BRCA1 gene carriers was not large, it should have been sufficient to detect biologically meaningful effects of reproductive factors. Again, a larger sample size, particularly of a recent cohort, is required to answer this question definitively. The only factor that seemed to have any association with expression of BRCA1 was the parental origin of the mutant allele, with more ovarian cancers resulting from paternal inheritance, while unaffected women were more likely to have inherited BRCA1 from their mothers. However, this result was only marginally statistically significant ($P = .04$) and could have arisen as a result of ascertainment bias. Moreover, when we examined the age at onset of breast or ovarian cancer as a function of maternal or paternal BRCA1 inheritance, no significant difference was found; in fact, cases with paternal inheritance had slightly later mean ages at diagnosis for both breast and ovarian cancer.

The analyses of the 214 families studied by the BCLC (*3*) and the individual papers from consortium members have provided an excellent base for characterizing the BRCA1 locus. Until the BRCA1 gene is identified and its mutations characterized, these studies are hampered by the need to pool data from many families, which may be segregating different BRCA1 mutations and/or represent distinct genetic, demographic, and environmental backgrounds. The advantages of studying a large kindred with well-defined ascertainment are threefold: first, the problem of selection bias which is present in so many linkage sets is minimized; second, one has a relatively homogeneous sample of individuals carrying identical mutations for epidemiologic studies; and third, the identification of individuals who carry the mutant BRCA1 allele can be done with a high degree of certainty. The obvious limitation of this single family approach is that, because this study is of a single mutation in a single family, the generalization of these findings to other mutations is difficult. We cannot know if the BRCA1 mutation in this family is common, rare, or unique. Nonetheless, we believe that, until the BRCA1 locus is cloned, the ongoing study of K2082 presented here can serve as a model for future studies of the effects of the BRCA1 locus. Moreover, the localization of two polymorphic STR markers, Mfd191 and ED2, in the Mfd15–Mfd188 interval and the observation of a clear recombinant event which places BRCA1 distal to Mfd191 will hopefully aid in the eventual isolation of the BRCA1 gene. The study of this kindred, through determination of BRCA1 carrier status in unaffected women and the estimation of age-specific penetrance specific to this kindred, has allowed us to begin a new series of studies focusing on long-term follow-up and the attitudes toward, and effects of, genetic counseling for 17q-linked breast and ovarian cancer.

References

(*1*) Hall JM, Lee MK, Newman B, et al: Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250:1684–1689, 1990

(*2*) Narod SA, Feunteun J, Lynch HT, et al: Familial breast-ovarian cancer locus on chromosome 17q12-q23. Lancet 338:82–83, 1991

(*3*) Easton DF, Bishop DT, Ford D, et al: Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. Am J Hum Genet 52:678–701, 1993

(*4*) Goldgar DE, Cannon-Albright LA, Oliphant A, et al: Chromosome 17q linkage studies of 18 Utah breast cancer kindreds. Am J Hum Genet 52:743–748, 1993

(*5*) Bowcock AM, Anderson LA, Black DM, et al: THRA1 and D17S181 flank an interval of <4 cM for the breast-ovarian cancer gene (BRCA1) on chromosome 17q21. Am J Hum Genet 52:718–722, 1993

(*6*) Feunteun J, Narod SA, Lynch HT, et al: A breast-ovarian cancer susceptibility gene maps to chromosome 17q21. Am J Hum Genet 52:736–742, 1993

(*7*) Claus EB, Risch N, Thompson WD: Genetic analysis of breast cancer in the cancer and steroid hormone study. Am J Hum Genet 48:232–242, 1991

(*8*) Harris JR, Lippman ME, Veronesi U, et al: Breast cancer. N Engl J Med 327:319–328, 1992

(*9*) Daly MB: The epidemiology of ovarian cancer. Hematol Oncol Clin North Am 6:729–738, 1992

(*10*) Parazzini F, Franceschi S, La Vecchia C, et al: The epidemiology of ovarian cancer. Gynecol Oncol 43:9–23, 1991

(*11*) Skolnick M: The Utah genealogical data base: a resource for genetic epidemiology. *In* Banbury Report No 4: Cancer Incidence in Defined Populations (Cairns J, Lyon JL, Skolnick M, et al, eds). New York: Cold Spring Harbor Laboratory, 1980, pp 285–297

(*12*) Bell GI, Selby MJ, Rutter WJ: The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature 295:31–35, 1982

(*13*) Mullis K, Faloona F, Scharf S, et al: Specific enzymatic amplification

GeneDX 1035, pg. 223

of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol 51:263–273, 1986

(14) Weber JL, May PE: Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396, 1989

(15) Lathrop GM, Lalouel JM, Julier C, et al: Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Am J Hum Genet 37:482–498, 1985

(16) Weber JL, Kwitek AE, May PE, et al: Dinucleotide repeat polymorphisms at the D17S250 and D17S261 loci. Nucleic Acids Res 18:464, 1990

(17) Hall JM, Friedman L, Guenther C, et al: Closing in on a breast cancer gene on chromosome 17q. Am J Hum Genet 50:1235–1242, 1992

(18) Daussett J: Le Centre d'Etude du Polymorphisme Humain. La Presse Médicale 15:1801–1802, 1986

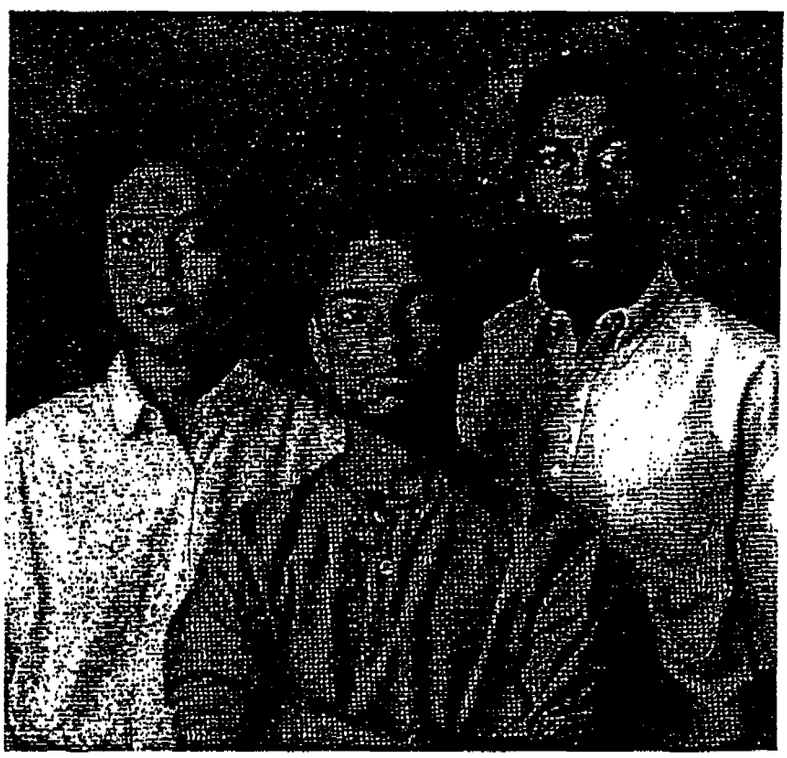
Notes

Affiliations of authors: D. E. Goldgar, P. Fields, C. M. Lewis, T. D. Tran, L. A. Cannon-Albright, J. Swensen, M. H. Skolnick (Genetic Epidemiology Group, Department of Medical Informatics), J. H. Ward (Division of Hematology/Oncology, Department of Internal Medicine), University of Utah, Salt Lake City.

Supported by Public Health Service grants CA48711, CN05222, CA55914 (National Cancer Institute) and RR00064 (National Center for Research Resources) from the National Institutes of Health, Department of Health and Human Services.

The authors would also like to thank the members of K2082 for their extensive cooperation in this study.

Manuscript received May 27, 1993; revised October 22, 1993; accepted November 8, 1993.



If all your children deserved to go to college,
but you could only send one, how would you choose?

A better question is why should you have to choose? Why shouldn't every bright, hard-working kid get a chance to fulfill his or her dreams?

That's why your donation to the United Negro College Fund is so important. As the most experienced source in providing general operating support for historically black colleges, we've enabled thousands of deserving students to get a quality education.

But there are still thousands who urgently need your support. How can we possibly say yes to one and no to another? Send your donation to: UNCF, 500 E. 62nd St., New York, NY 10021.

UNITED NEGRO COLLEGE FUND
A Mind Is A Terrible Thing To Waste.

Ad Council — A Public Service of This Publication

© 1991 United Negro College Fund Inc.

variants. Furthermore, probe cyclization reactions depend on an intramolecular reaction as opposed to reaction between pairs of independent probe molecules as in amplification by the polymerase chain reaction. Thus, there should be fewer problems with nonspecific reactions resulting from interactions between noncognate pairs of probe segments with cyclizable probes. The present probe design should permit the simultaneous analysis of multiple gene sequences in a DNA sample.

In conclusion, the nucleic acid probe presented here permits highly specific detection of nucleotide sequences and, although the target is not amplified, highly sensitive detection is possible through efficient reduction of nonspecific signal. Circularizable probes should be applicable in a number of other contexts, including the detection of specific RNA molecules expressed in tissue sections as T4 DNA ligase can assist in ligation reactions involving RNA strands (*8*). Moreover, immobilized padlock probes could be useful for preparative purposes, such as trapping circular target molecules from solution when screening gene libraries.

REFERENCES AND NOTES

1. U. Landegren, R. Kaiser, J. Sanders, L. Hood, *Science* 241, 1077 (1988); A. M. Alves and F. J. Carr, *Nucleic Acids Res.* 16, 8723 (1988); F. Barany, *Proc. Natl. Acad. Sci. U.S.A.* 88, 189 (1991).
2. D. Y. Wu and R. B. Wallace, *Gene* 76, 245 (1989).
3. A. Jäschke, J. P. Fürste, D. Cech, V. A. Erdmann, *Tetrahedron Lett.* 34, 301 (1993).
4. G. Prakash and E. T. Kool, *J. Am. Chem. Soc.* 114, 3523 (1992); N. G. Dolinnaya *et al.*, *Nucleic Acids Res.* 21, 5403 (1993).
5. The upper faint bands observed in lanes 3 and 4 probably represent small amounts of linear dimer molecules, appearing as a consequence of ligation of one end each of two different probe molecules. This material proved susceptible to exonuclease digestion. The extra lower bands in these lanes were not reproducible between experiments. Small amounts of uncatenated, circular probes appearing in lane 7 most likely were a consequence of endonuclease activity in the exonuclease preparation. With increasing amounts of exonuclease, catenated probes are lost and more free circular probes appear (M. Nilsson *et al.*, unpublished data).
6. J. R. Riordan *et al.*, *Science* 245, 1066 (1989).
7. H. F. Willard and J. S. Waye, *Trends Genet.* 3, 192 (1987); A. G. Matera and D. C. Ward, *Hum. Mol. Genet.* 7, 535 (1992); A. Baldini *et al.*, *Am. J. Hum. Genet.* 46, 784 (1990).
8. N. P. Higgins and N. R. Cozzarelli, *Methods Enzymol.* 68, 50 (1979).
9. T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982).
10. C. Sund, J. Ylikoski, P. Hurskainen, M. Kwiatkowski, *Nucleos. Nucleot.* 7, 655 (1988).
11. D. Pinkel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 85, 9138 (1988).

12. We thank E. Johnsen for technical assistance and T. Hansson for molecular modeling. U. Pettersson offered critical comments on this manuscript. Supported by the Beijer, Procordia, and Borgström foundations; by NUTEK, the Technical and Medical Research Councils of Sweden; and by the Swedish Cancer Fund.

18 July 1994; accepted 1 September 1994

Localization of a Breast Cancer Susceptibility Gene, *BRCA2*, to Chromosome 13q12-13

Richard Wooster,* Susan L. Neuhausen,* Jonathan Mangion,* Yvette Quirk,* Deborah Ford,* Nadine Collins, Kim Nguyen, Sheila Seal, Thao Tran, Diane Averill, Patty Fields, Gill Marshall, Steven Narod, Gilbert M. Lenoir, Henry Lynch, Jean Feunteun, Peter Devilee, Cees J. Cornelisse, Fred H. Menko, Peter A. Daly, Wilma Ormiston, Ross McManus, Carole Pye, Cathryn M. Lewis, Lisa A. Cannon-Albright, Julian Peto, Bruce A. J. Ponder, Mark H. Skolnick, Douglas F. Easton,† David E. Goldgar, Michael R. Stratton

A small proportion of breast cancer, in particular those cases arising at a young age, is due to the inheritance of dominant susceptibility genes conferring a high risk of the disease. A genomic linkage search was performed with 15 high-risk breast cancer families that were unlinked to the *BRCA1* locus on chromosome 17q21. This analysis localized a second breast cancer susceptibility locus, *BRCA2*, to a 6-centimorgan interval on chromosome 13q12-13. Preliminary evidence suggests that *BRCA2* confers a high risk of breast cancer but, unlike *BRCA1*, does not confer a substantially elevated risk of ovarian cancer.

In 1990, a breast cancer susceptibility gene, known as *BRCA1*, was localized to chromosome 17q (*1*). Subsequent studies demonstrated that *BRCA1* accounts for most families with multiple cases of both early-onset breast and ovarian cancer and about 45% of families with breast cancer only, but few if any families with both male and female breast cancer (*2*). Several other genes can confer susceptibility to breast cancer. Germline mutations in the p53 gene on chromosome 17p cause a wide range of neoplasms including early-onset breast cancer, sarcomas, brain tumors, leukemias, and adrenocortical cancer (*3*). Certain rare abnormalities of the androgen receptor appear to be associated with breast cancer in men (*4*), and epidemiological studies have suggested that heterozygotes for the ataxia telangiectasia gene, *AT*, on chromosome 11q are at elevated risk of breast cancer (*5*). However, mutations in *p53* and *AT* can only be responsible for a small minority of breast cancer families that are unlinked to *BRCA1* (*6*).

To localize other genes that predispose to breast cancer, we performed a genomic linkage search using 15 families that had multiple cases of early-onset breast cancer and that were not linked to *BRCA1*. These families were classified according to the number of cases of female breast cancer, male breast cancer, and ovarian cancer (Table 1). In addition to a negative lod score (logarithm of the likelihood ratio for linkage) with markers flanking *BRCA1*, all but one of the families used for this study had at least one breast cancer case diagnosed before age 50 that did not share a *BRCA1* haplotype with other breast cancer cases in the family. The exception, CRC 136, had an obligate sporadic case diagnosed at age 53. Families were genotyped with polymorphic microsatellite repeat markers (*7*, *8*). Typing of the markers *D13S260* and *D13S263* provided provisional evidence for the presence of a susceptibility gene on chromosome 13, which was subsequently confirmed by analysis of additional polymorphisms in the region.

R. Wooster, J. Mangion, Y. Quirk, N. Collins, S. Seal, M. R. Stratton, Section of Molecular Carcinogenesis, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK.
S. L. Neuhausen, K. Nguyen, T. Tran, P. Fields, C. M. Lewis, M. H. Skolnick, D. E. Goldgar, Department of Medical Informatics, University of Utah, Salt Lake City, UT 84108, USA.
D. Ford, D. Averill, G. Marshall, J. Peto, D. F. Easton, Section of Epidemiology, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK.
S. Narod, Department of Medicine, Division of Medical Genetics and Division of Human Genetics, McGill University, Montreal, Canada H3G 1A4.
G. M. Lenoir, International Agency for Research on Cancer, 150 Cours Albert-Thomas, 69372 Lyon Cedex 08, France.
H. Lynch, Department of Preventive Medicine and Public Health, Creighton University School of Medicine, Omaha, NE 68178, USA.
J. Feunteun, Institute Gustav-Roussy, Villejuif, France.
P. Devilee and C. J. Cornelisse, Departments of Pathology and Human Genetics, University of Leiden, 2333 AL Leiden, Netherlands.
F. H. Menko, Department of Clinical Genetics, Free University of Amsterdam, 1007 MB Amsterdam, Netherlands.
P. A. Daly, W. Ormiston, R. McManus, Department of Medicine, Trinity College Medical School, St. James Hospital, Dublin 8, Ireland.
C. Pye and B. A. J. Ponder, CRC Human Cancer Genetics Group, Department of Pathology, University of Cambridge, Cambridge CB2 1QP, UK.
L. A. Cannon-Albright, Department of Internal Medicine, University of Utah, Salt Lake City, UT 84132, USA.

*These authors contributed equally to this study.
†To whom correspondence should be addressed.

GeneDX 1035, pg. 225

wo-point lod scores were calculated a set of closely spaced markers on mosome 13q (Table 2) (9). Ten other kers were typed to confirm the segtion of haplotypes. The order of kers and intervals between them (in imorgans) is 13cen-*D13S283*-(3)-*S221*-(2)-*D13S120*-(2)-*D13S217*-*D13S289/D13S290*-(3)-*D13S260*-(1)-*S171*-(2)-*D13S267*-(2)-*D13S220/S219*-(5)-*D13S218*-(5)-*D13S263*-*D13S155*-(2)-*D13S153*-13qter (*10*). maximum total multipoint lod score markers *D13S260* and *D13S267* was at a location 5 cM proximal to *S260*. However, the admixture test cated significant evidence of heterosity ($P = 0.001$) with an estimated portion of 13q-linked families of 74% 6 CI 35 to 97%). Under the assumpof heterogeneity, the maximum total score was 11.65 and the most likely tion for *BRCA2* was coincident with *S260*. Multipoint lod scores at *S260* for each family are shown in le 2. The haplotypes confirmed cosegtion of chromosome 13q markers with disease (an example from CRC 186 is vn in Fig. 1). Two recombinants place *A2* telomeric to *D13S289*, in breast er cases diagnosed at ages 43 and 39 families IARC 2932 and CRC 186, ctively). One recombinant in a bial breast cancer case diagnosed at 38 and 41 in Utah 107 places the centromeric to *D13S267*. The dise between these two markers is esti- to be 6 cM (7), and these flanking kers place BRCA2 in a physical region ned by 13q12-13.

The proximal part of chromosome 13 in which *BRCA2* is situated commonly shows loss of heterozygosity (LOH) in sporadic breast and ovarian cancers, suggesting that *BRCA2* is inactivated during oncogenesis (*10*). However, the tumor suppressor gene *RB1* is also located in this region and may account for the LOH observed. Indeed, somatic mutations in *RB1* have been reported in sporadic breast cancers (*11*). However, the presence of numerous recombinants between *RB1* [the marker *D13S153* is within the *RB1* gene (*12*)] and the disease in linked families indicates that *BRCA2* is not *RB1*. Other candidate genes within 13q11-14 include members of a family of tyrosine kinase genes that are related to the *FMS* proto-oncogene (*13*) and the *FTE1* gene, which may act as an effector of the *v-fos* oncogene and is a mammalian homolog of a yeast gene involved in protein import into mitochondria (*14*).

Like *BRCA1*, *BRCA2* appears to confer a high risk of early-onset breast cancer in females; previous segregation analysis of the largest BRCA2-linked family (Utah 107) indicated a risk of 87% by age 80 (*15*), which is comparable to the *BRCA1* penetrance. However, other aspects of the *BRCA2* phenotype may differ from the *BRCA1* phenotype. For example, in the two families showing the strongest evidence of

Table 1. Breast cancer families used in the genome search for *BRCA2*. FBC, female breast cancer; MBC, male breast cancer; OvC, ovarian cancer.

Family	Number of FBCs	Number of FBCs under age 50	Number of MBCs	Number of OvCs	Lod score at *BRCA1**	Number of sporadic cases†
CRC 007	7	5	0	0	−1.45	2
CRC 018	5	3	0	0	−0.41	1
CRC 028	3	2	1	0	0.04	1
CRC 135	6	4	0	0	−0.49	1
CRC 136	6	4	0	0	−0.02	1
CRC 186	16	15	1	1	−2.61	7
IARC 2932	15	10	0	0	−2.02	3
Leiden 49	4	4	1	4	−1.11	1
Utah 107	38	25	3	6	−3.57	7
Utah 1001	14	11	0	0	−0.48	2
Utah 1929	4	4	0	0	−0.41	1
Utah 2027	4	4	0	0	−1.14	1
Utah 2043	2	2	1	1	−0.44	1
Utah 2044	9	6	1	4	−1.40	4
Utah 9018	5	5	0	0	−0.53	1

*Multipoint lod score based on *D17S250* and *D17S579*, which flank *BRCA1* in an interval of approximately 6 cM, or closer flanking markers. †Minimum number of cases affected with breast cancer under age 60 or ovarian cancer that do not share a 17q haplotype.

e 2. Two-point and multipoint lod scores for chromosome 13q markers in breast cancer families showing evidence against linkage to *BRCA1*.

mily	Two-point lod scores at recombination fractions of 0.00 and 0.05										Multipoint lod score*
	D13S289		*D13S260*		*D13S267*		*D13S219*		*D13S263*		*D13S260-D13S267*
	0.00	0.05	0.00	0.05	0.00	0.05	0.00	0.05	0.00	0.05	
007	−1.23	−0.81	0.77	0.65	0.22	0.16	0.29	0.24	−0.30	−0.11	0.97
018	0.73	0.62	0.78	0.67	0.26	0.21	0.10	0.08	−0.48	−0.27	0.84
028	0.00	0.00	−0.02	0.00	0.26	0.21	0.02	0.01	−0.67	−0.41	0.15
135	0.43	0.36	0.15	0.12	0.14	0.12	−0.06	−0.04	−0.54	−0.32	0.29
136	−1.26	−0.84	−1.07	−0.75	−0.57	−0.42	−0.28	−0.10	−1.06	−0.71	−1.24
186	−0.30	0.03	1.84	1.60	2.35	2.08	1.00	0.84	4.08	3.67	3.70
2932	−0.03	0.10	1.33	1.22	0.80	0.67	1.62	1.38	−0.72	−0.17	1.93
n 49	−0.65	−0.37	0.08	0.06	−0.21	−0.16	−0.65	−0.25	−0.75	−0.37	−0.44
107	0.24	1.11	1.84	2.11	−0.86	−0.23	0.24	0.20	1.26	1.66	3.48
1001	−0.81	−0.41	−3.25	−1.99	−1.63	−0.96	−0.39	−0.17	−1.91	−1.02	−3.40
1929	−0.46	−0.33	−0.33	−0.26	−0.47	−0.37	−0.16	−0.14	−0.25	−0.19	−0.45
2027	0.39	0.34	0.14	0.13	−0.06	−0.05	0.19	0.16	0.69	0.59	−0.11
2043	0.93	0.81	0.85	0.74	−0.39	−0.27	−0.01	−0.02	−1.13	−0.52	0.86
2044	−0.59	−0.44	1.54	1.37	1.15	0.99	0.88	0.76	−1.29	−0.72	2.11
9018	0.23	0.19	0.16	0.12	0.00	0.00	−0.11	−0.09	−0.86	−0.47	0.00
	−2.39	0.34	4.80	5.78	1.00	1.98	2.67	2.87	−3.95	0.65	

oint lod scores were calculated at *D13S260*, the most likely position of *BRCA2* in the heterogeneity analysis.

GeneDX 1035, pg. 226

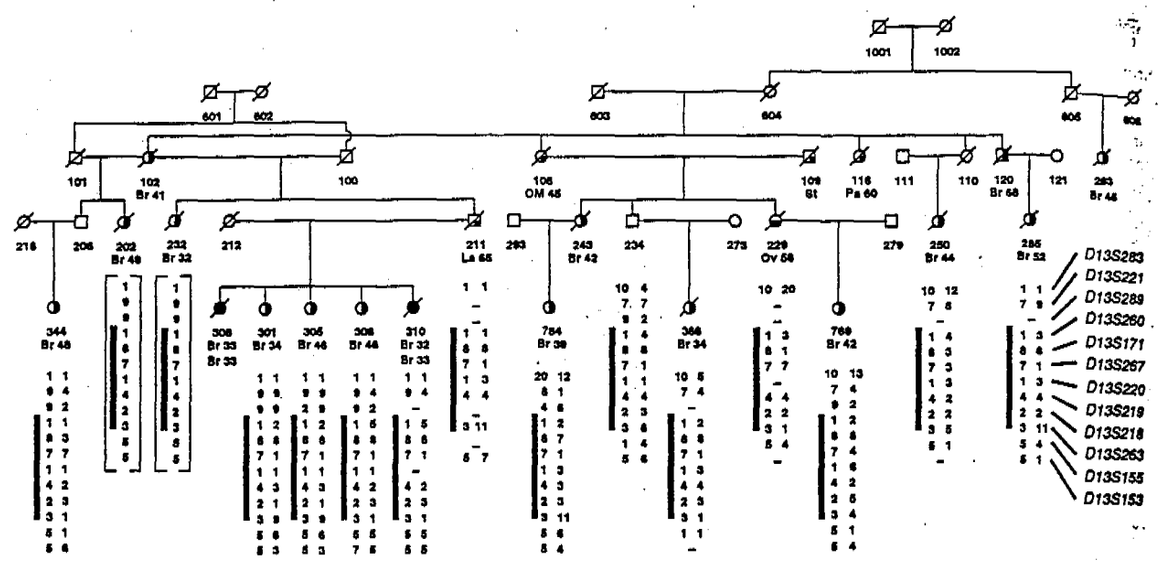


Fig. 1. Pedigree of CRC 186. Half shading (right side), breast cancer; full shading, bilateral breast cancer; half shading (bottom), ovarian cancer; quarter shading, other cancer. Types of cancer: Br, breast; Ov, ovary; Pa, pancreas; La, larynx; St, stomach; OM, ocular melanoma. The number after the cancer type is the age of diagnosis. Unaffected individuals who are potential gene carriers have been omitted. Marker numbers are shown on the right adjacent to the haplotype of individual 285. The black bar indicates the haplotype shared by all affected individuals. Genotypes in square brackets are inferred.

linkage to *BRCA2* (multipoint lod score greater than 3.0), there are 49 reported cases of breast cancer, 39 under age 50, and only 3 ovarian cancers (excluding 5 cases of breast cancer and 4 of ovarian cancer in Utah 107 that do not carry the linked haplotype). This suggests that the risk of ovarian cancer attributable to *BRCA2* may be lower than that for *BRCA1*, which confers an estimated 63% risk by age 70 (*16*). There may also be a difference in the risk of male breast cancer. In the same two families, there were four cases of male breast cancer and three more cases in other families showing some evidence of linkage to *BRCA2*. By contrast, no male breast cancers have been observed in families showing strong evidence of linkage to *BRCA1*. Thus, the risk of breast cancer in men carrying *BRCA2* mutations, though still small, is probably greater than in men carrying *BRCA1* mutations. However, the absolute risk of male breast cancer is still small, and many families where the risk of breast cancer is attributable to *BRCA2* will be characterized by female breast cancer only (for example, IARC 2932).

Although in the majority of families in our data set breast cancer can now be attributed to *BRCA1* or *BRCA2*, it is likely that these genes still do not account for all breast cancer caused by high-risk susceptibility genes (of the order of 5% of all cases). Both the overall evidence for genetic heterogeneity and inspection of haplotypes in individual families indicate that an additional gene (or genes) conferring susceptibility to breast cancer remains to be discovered.

REFERENCES AND NOTES

1. J. M. Hall *et al.*, *Science* 250, 1684 (1990); S. A. Narod *et al.*, *Lancet* 338, 82 (1991).
2. D. F. Easton, D. T. Bishop, D. Ford, G. P. Crockford, and the Breast Cancer Linkage Consortium, *Am. J. Hum. Genet.* 52, 678 (1993); M. R. Stratton *et al.*, *Nat. Genet.* 7, 103 (1994).
3. J. M. Birch *et al.*, *Cancer Res.* 54, 1298 (1994); D. Malkin *et al.*, *Science* 250, 1233 (1990); S. Srivastava, Z. Zou, K. Pirollo, W. Blattner, E. H. Chang, *Nature* 348, 747 (1990).
4. R. Wooster *et al.*, *Nat. Genet.* 2, 132 (1992); J.-M. Lobaccaro *et al.*, *ibid.* 5, 109 (1993).
5. M. Swift, D. Morrell, R. Massey, C. L. Chase, *N. Engl. J. Med.* 325, 1831 (1991).
6. J. Prosser *et al.*, *Br. J. Cancer* 63, 181 (1991); W. Warren *et al.*, *Oncogene* 7, 1043 (1992); R. Wooster *et al.*, *Hum. Genet.* 92, 91 (1993).
7. J. Weissenbach *et al.*, *Nature* 359, 794 (1992); G. Gyapay *et al.*, *Nat. Genet.* 7, 246 (1994).
8. Genotyping was performed independently at the Institute of Cancer Research [R. Wooster *et al.*, *Nat. Genet.* 6, 152 (1994)] and the University of Utah [L. A. Cannon-Albright *et al.*, *Science* 258, 1148 (1994)].
9. Lod scores were computed with the FASTLINK version of the LINKAGE program [G. M. Lathrop, J. M. Lalouel, C. Julier, J. Ott, *Proc. Natl. Acad. Sci. U.S.A.* 81, 3443 (1984); R. W. Cottingham Jr., R. M. Idury, A. A. Schaffer, *Am. J. Hum. Genet.* 53, 252 (1993)], using a model for familial breast cancer derived from the segregation analysis of the Cancer and Steroid Hormone Study (CASH). Under this model, susceptibility to breast cancer is conferred by an autosomal dominant gene with population frequency 0.0033, and the risk of breast cancer by age 70 in carriers is 0.67 compared to 0.05 in noncarriers. Penetrances for ovarian cancer in carriers of the breast cancer susceptibility gene were derived from E. B. Claus, J. M. Schildkraut, W. D. Thompson, and N. J. Risch [*Am. J. Hum. Genet.* 53, A787 (1993)], which estimated a cumulative frequency of 10% by age 60. The effects of age were incorporated into the analyses as in (*2*). Male breast cancer cases were assigned to the same liability classes as female breast cancer cases diagnosed below age 30. All cancers other than breast and ovarian cancer were ignored in this analysis. Allele frequencies were those reported in the Genome Data Base and were based on 56 CEPH (Centre d'Etude du Polymorphisme Humain) chromosomes. Evidence for heterogeneity was evaluated by the admixture model [J. Ott, *Analysis of Human Genetic Linkage* (Johns Hopkins Univ. Press, Baltimore, 1985), pp. 200–203]. Lod scores for linkage to *BRCA1* were computed as in (*5*).
10. C. Lundberg, L. Skoog, W. K. Cavenee, M. Nordenskjold, *Proc. Natl. Acad. Sci. U.S.A.* 84, 2372 (1987); P. Devilee, M. van den Broek, N. Kuipers-Dijkshoorn, R. Kolluri, P. M. Khan, *Genomics* 5, 554 (1989); D. Deng *et al.*, *Cancer Res.* 54, 499 (1994).
11. E. Y.-H. P. Lee *et al.*, *Science* 241, 218 (1988); A. T'Ang, J. M. Varley, S. Chakraborty, A. L. Murphree, Y.-K. T. Fung, *ibid.* 242, 263 (1988).
12. J. Toguchida *et al.*, *Genomics* 17, 535 (1993).
13. O. Rosnet *et al.*, *Oncogene* 8, 173 (1993).
14. C. J. Kho and H. Zarbl, *Proc. Natl. Acad. Sci. U.S.A.* 89, 2200 (1992).
15. D. T. Bishop *et al.*, *Genet. Epidemiol.* 5, 151 (1988).
16. D. F. Easton, D. Ford, D. T. Bishop, and the Breast Cancer Linkage Consortium, unpublished data.

17. All studies were carried out with the full informed consent of the families. We thank the Cancer Research Campaign, the Medical Research Council, the Committee for Clinical Research of the Royal Marsden Hospital, the National Institutes of Health (grants CA-48711, HG-00571, CN-05222, and RR-00064), grant DAMD17-94-J-4260 from the U.S. Army, Myriad Genetics, the European Community Concerted Action on Genetics and Epidemiology of Hereditary Breast Cancer Linkage (PL920890), the Chetwood Aiken fund, and the Ligue Nationale contre le Cancer for their support. We also thank the families for their encouragement and cooperation, the many clinicians who have helped in the identification of cases, in particular the late Dr. E. Gardner, all the members of the International Breast Cancer Linkage Consortium, F. Lennard, M. Ponder, and T. Harrington.

25 July 1994; accepted 26 August 1994

GeneDX 1035, pg. 227

[CANCER RESEARCH 52, 1643–1646, March 16, 1992]

Advances in Brief

The Human Prohibitin Gene Located on Chromosome 17q21 Is Mutated in Sporadic Breast Cancer[1]

Takaaki Sato, Hiroko Saito, Jeff Swensen, Arnold Olifant, Carla Wood, David Danner, Takashi Sakamoto, Kenichi Takita, Fujio Kasumi, Yoshio Miki, Mark Skolnick, and Yusuke Nakamura[2]

Departments of Biochemistry [T. Sat., H. S., T. Sak., K. T., Y. N.,] and Surgery [F. K.], Cancer Institute, Tokyo, 1-37-1, Kami-Ikebukuro, Toshima-ku, Tokyo 170, Japan; Genetic Epidemiology, Department of Medical Informatics, The University of Utah, Salt Lake City, Utah 84108 [J. S., A. O., Y. M., M. S.]; and Laboratory of Molecular Genetics, National Institute of Aging, Baltimore, Maryland 21224 [C. W., D. D.]

Abstract

A gene called "prohibitin" was isolated as a candidate antiproliferating gene in rat liver cells. We have isolated the human homologue of the rat prohibitin gene and mapped it to chromosome 17q12-21 where a gene responsible for hereditary breast cancer was localized. DNA sequence analysis of 2 exons in this gene in 23 sporadic breast cancers, which showed loss of heterozygosity on the long arm of chromosome 17 or developed in patients 35 years old or younger, identified 4 cases of somatic mutation; 2 of these were missense mutations; 1 showed a 2-base deletion resulting in truncation of the gene product due to a frame shift; the other had a C to T transition in an intron adjacent to an intron-exon boundary. These results suggest that this gene may be a tumor suppressor gene and is associated with tumor development and/or progression of at least some breast cancers.

Introduction

Breast cancer is the most common cancer in women. One in 10 Caucasian women and one in 60 Japanese women will develop breast cancer in their lifetime. The highest risk factor for breast cancer is family history (1) which reflects dominantly inherited susceptibility (2–5) caused by at least three susceptibility loci, the p53 locus on chromosome 17p13 (6), a 17q-linked susceptibility locus (7), and one or more unmapped loci. Our studies based on LOH[3] in breast tumors identified 17q21 as one of the commonly deleted regions (8, 9), indicating the possible presence of a tumor suppressor gene, in the same region as a gene responsible for familial breast cancer (7).

The rat prohibitin gene was isolated for its ability to negatively regulate cell proliferation (10). McClung *et al.* (11) isolated a rat prohibitin gene as one of a set of cDNAs derived from mRNAs which were more frequently expressed in normal liver than in regenerating liver. The gene caused an arrest of DNA synthesis by an *in vitro* assay in normal human fibroblast and HeLa cells, further demonstrating its antiproliferative activity (10). Moreover, it showed significant homology to the *Cc* gene, which was considered to be important for development and differentiation of *Drosophila melanogaster* (12). Hence, the function of the prohibitin gene is considered as an antiproliferating factor. The rat cDNA with a complete open reading frame was subsequently cloned (10) and used to isolate a fragment of human genomic DNA which was mapped by *in situ* hybridization to 17q12–17q21 (13). Because the prohibitin gene was assigned to this region, we undertook an analysis of it as a candidate for one of the tumor suppressor genes associated with breast cancer.

Materials and Methods

Materials. Tumor and normal tissues removed by mastectomy from 23 patients with primary breast cancer at the Cancer Institute Hospital, Tokyo, were analyzed for somatic mutations of the prohibitin gene. Extraction of DNAs from tumor and normal tissues was carried out according to the method described previously (8). Allelotype study of these 23 tumors was reported previously (8, 9).

RT-PCR and cDNA Cloning. To isolate human prohibitin cDNA, two oligonucleotides were synthesized for a RT-PCR reaction. Single-stranded cDNA was synthesized using 0.1 μg of human normal liver mRNA, 200 units of Moloney murine leukemia virus-reverse transcriptase (BRL) and a (5′-TCACCCTCAGCAGAGATGAT-3′) primer. Double-stranded cDNAs were then synthesized and amplified with 2.5 units Taq polymerase (Cetus), the above primer, and the (5′-CGCTCTCGACCACGTAATGT-3′) primer for 35 cycles (94°C, 1 min; 55°C, 2 min; and 72°C, 2 min). A 443-base pair product was cloned in pBluescript II SK(−) (Stratagene, La Jolla, CA) and sequenced. A human fetal brain cDNA library (approximately 3.5×10^5 clones) (Clontech, Palo Alto, CA) was screened with the PCR product. Two positive clones were isolated and then sequenced. Intron-exon boundaries were determined by comparing the genomic DNA sequence derived from cosmid and/or phage clones with that of the cDNA clone.

Mutation Analysis. Sequence analysis was undertaken for examining somatic mutations in sporadic primary breast cancer. A pair of tumor and its corresponding normal DNAs was sequenced by the following method. Briefly, PCR products of exon 4 (nucleotides 301 to 443) which is highly conserved in the *Drosophila*, *Cc* gene (12), were amplified using two primers in the introns flanking exon 4 (5′-GTACTCCAGCCTAGGCAAC-3′ and 5′-CAGGAAACTAGCAGCCACAT-3′; 94°C, 1 min; 55°C, 2 min; and 72°C, 2 min, 35 cycles). Products were cloned in pBluescript II SK(−) and at least 100 independent clones were pooled. Both strands were sequenced (14) using two internal primers in introns (5′-ACACTTGTTTTCCTACAG-3′ and 5′-GTGCTCTGGGCTCGAGC-3′) and Sequenase (T7 DNA polymerase; Pharmacia, Piscataway, NJ). The samples found alterations were tested the whole procedure more than twice to confirm the mutations.

Results

Isolation of Human Prohibitin Gene. To examine a possible role in human cancer, human prohibitin cDNA was isolated, based on its homology to the rat cDNA. First, a pair of oligonucleotides derived from the region of rat prohibitin gene, which was conserved in the *Cc* gene of *D. melanogaster* were used to create a 443-base pair RT-PCR product using human normal liver mRNA as a template. This PCR product was then

Received 2/11/92; accepted 2/12/92.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported in part by grants from the Ministry of Education, Culture and Sciences, the Ministry of Health and Welfare, the Vehicle Racing Commemorative Foundation, and NIH Grant CA-48711.

[2] To whom requests for reprints should be addressed, at Department of Biochemistry, Cancer Institute, Tokyo, 1-37-1, Kami-Ikebukuro, Toshima-ku, Tokyo 170, Japan.

[3] The abbreviations used are: LOH, loss of heterozygosity; RT-PCR, reverse transcriptase-polymerase chain reaction; cDNA, complementary DNA.

GeneDX 1035, pg. 228

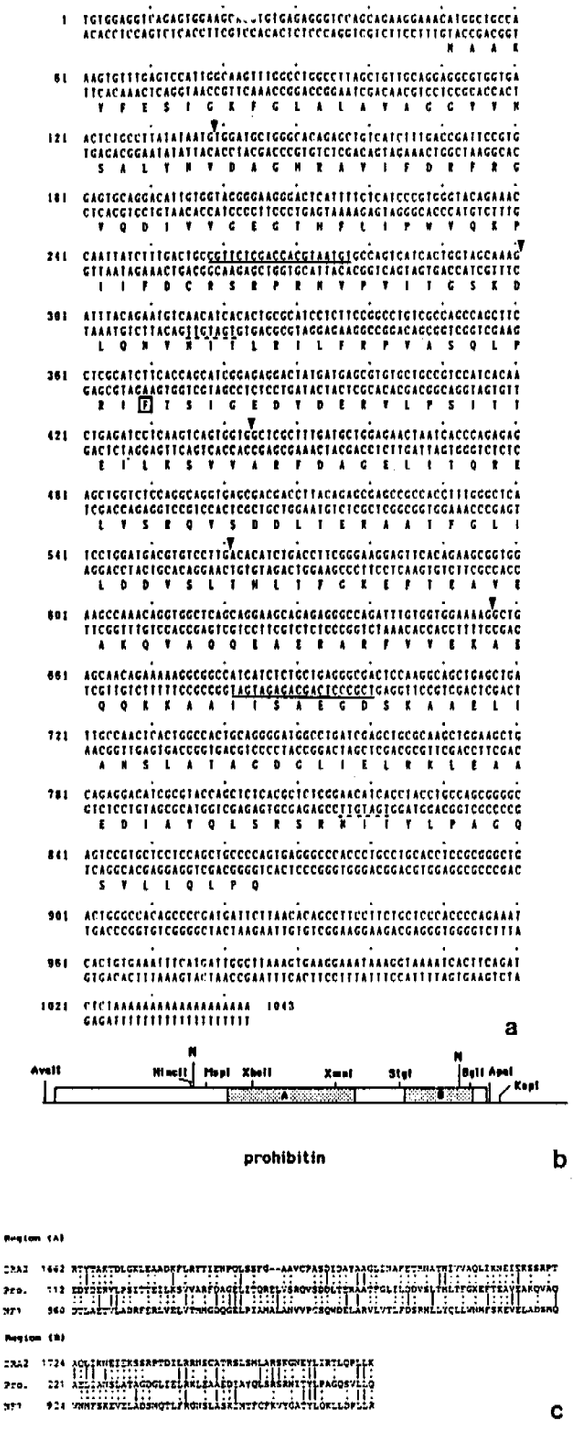


Fig. 1. *a*, human prohibitin nucleotide sequence and predicted amino acids sequence. Nucleotides are numbered at the *left*. Two oligonucleotides used to isolate human cDNA by the RT-PCR method are shown *underlined* and two potential sites of *N*-linked glycosylation are indicated by *dashed lines* above the amino acid sequence. The single amino acid difference at codon 107 between human (phenylalanine) and rat (tyrosine) prohibitin protein is enclosed in a box. *Arrows*, intron-exon junctions. *b*, a schematic representation of the prohibitin gene. *A* and *B* in a *hatched box*, region homologous to *NF1* and *IRA2*, respectively. The 272 residues of human prohibitin share a 99.6% identity with rat prohibitin (4) and residues 117–265 share a 50.7% identity with the *Cc* protein of *D. melanogaster* (12). Two *Ns* indicate the potential *N*-glycosylation sites. *c*, alignment of the amino acid sequence of the human prohibitin and *GAP* catalytic domains of yeast *IRA2* and neurofibromatosis type 1 (*NF1*) protein (15, 16). Residues 221–269 share a 28.6% identity (the *B* region in Fig. 1*b*) with the yeast *IRA2* protein (17) which is a homologue of neurofibromatosis type 1 (*NF1*) (15, 16) and residues 112–190 shared a 21.5% identity with *NF1* (the *A* region in Fig. 1*b*). The regions of similarity were identified by the GCG sequence analysis software package (version 7.0; Madison, WI). *Vertical bars* and *dots*, identical amino acid residues and conservative replacement, respectively. Residue numbers of each protein are shown to the *left*.

Table *Prohibitin gene mutations in sporadic breast cancers*

Tumor	Age (yr)	LOH on ch17q[a]	Codon	Mutation[b] nucleotide	Amino acid
136	38	+	88	GTC-GCC	Val–Ala
342	60	+	105	CGC-CAC	Arg–His
120	35	+	91–92	TCACACT-TCACT	Frame-shift
218	39	+		GTG/gtgagtgaaca-GTG/gtgagtgaata	

[a] LOH was determined by RFLP analysis as described previously (8, 9).
[b] The underlined nucleotides were mutant: lowercase letters represent introns; uppercase letters represent exons.

used to screen human cDNA, and genomic cosmid and phage libraries. The cDNA sequence containing the entire coding region is shown in Fig. 1*a*. This gene was mapped to chromosome 17q21–22 by *in situ* hybridization (13) and has been confirmed to the region 17q12–21 by fluorescent *in situ* hybridization[4] giving most likely assignment of 17q21.

The size of the predicted protein is 272 residues. It was expressed in all rat tissues tested by RT-PCR: liver, heart, colon, brain, kidney, adrenal gland, uterus, thymus, stomach, spleen, pancreas, skeletal muscle, lung, and ovary (data not shown).

Prohibitin is highly conserved (99.6% identical at amino acid level and 90.3% identical at nucleotide level between rat and human). Furthermore, it showed significant homology to the *Cc* gene which is considered to be important in the development and differentiation of *D. melanogaster* (12) and is also homologous to the human *NF1* gene (15, 16) and the yeast *IRA2* gene (17) as shown in Fig. 1, *b* and *c*.

Analysis of Somatic Mutation of the Prohibitin Gene in Sporadic Breast Cancer. We screened for somatic mutations in the prohibitin gene in DNAs from 23 breast tumors; 7 of these tumors developed in patients 35 years old or younger, 11 of the tumors showed significant reduction or loss of one of the polymorphic alleles on the long arm but not on the short arm of chromosome 17, and 5 tumors with LOH on chromosome 17q developed in patients 35 years old or younger. LOH in chromosome 17 was determined by restriction fragment length polymorphism analysis as described previously (8, 9). Furthermore, the *Eco*RI polymorphism of the human prohibitin gene (13) was also tested for the detection of LOH. We first examined the fourth exon since it is highly conserved in the homologous *Drosophila Cc* gene. The exon was amplified by PCR, subcloned, and then sequenced. The results are shown in Table 1 and Fig. 2, *a* and *b*. In one tumor, somatic mutation from CGC to CAC at codon 105 that resulted in amino acid change from arginine to histidine was detected as shown in Fig. 2*a* although a normal G is still observed in tumor because of the contamination of the normal cells. In another tumor, a 2-base deletion was found at codon 90–92 (Fig. 2b) which caused a

[4] K. Okui and Y. Nakamura, unpublished data.

GeneDX 1035, pg. 229

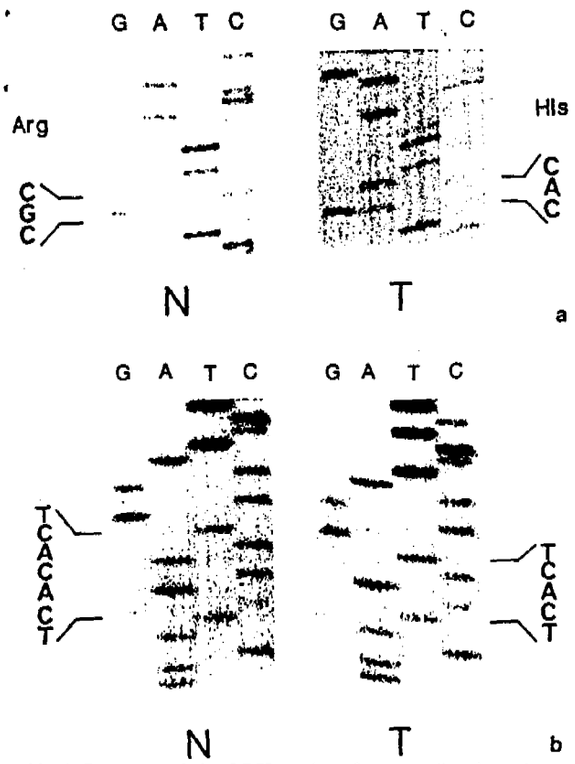


Fig. 2. Sequence analysis of PCR products from sporadic primary breast cancers. DNAs extracted from tumor and its corresponding normal tissues were tested for mutation analysis of exon 4 in the prohibitin gene. PCR primers in introns flanking exon 4 were used to amplify prohibitin nucleotide 301 to 443 as described in "Materials and Methods." *a*, sequence analysis of the T342 PCR product shows a G (in the normal allele) to A (in the mutated allele) transition in codon 105. The autoradiography represents the result from pooled clones. In *b*, the T120 product shows a 2-base deletion from TCACACT (in the normal allele) to TCACT (in the mutated allele) at codon 90–92, resulting in a frame-shift mutation.

frame shift and generated a new stop codon at nearly 70 base pairs downstream. As shown in Table 1, two other somatic mutations have been also detected in this region; one is a missense mutation from valine (GTC) to alanine (GCC) at codon 88 and the other is a C to T change in an intron near the intron-exon boundary. All tumors listed in Table 1 were detected the loss of one of polymorphic bands with one or more restriction fragment length polymorphism markers on the long arm of chromosome 17. However, somatic alteration in exon 5 was detected in none of the 23 tumors (data not shown).

Discussion

In this report, we present the cloning of the human prohibitin gene and its somatic mutations in primary breast cancer. In one case (T120), it is clear that tumor cells could produce no normal prohibitin gene product due to the loss of one allele and frame-shift mutation on the other allele. In another three tumors, it is still uncertain that missense mutations in tumors 136 and 342 caused a significant effect on the prohibitin function or that the point mutation in the intron in tumor 218 induced an abnormal splicing. However, it is notable that a change from arginine to histidine found in tumor 342 is the same as one of the most frequent mutations (at codon 273) observed in the *p53* gene (18).

These results suggested (*a*) that the defect of the prohibitin gene may have a significant role for development and/or progression of at least some breast carcinomas and (*b*) that the region in the prohibitin gene which was evolutionally conserved in the *Drosophila Cc* gene might be one of the important functional domains of this gene. Although the function of this gene or whether it acts in dominantly negative fashion (19, 20) is still unclear, it is suspected that this gene works as a negative growth factor (21) or *GAP* like protein (22–25) since this gene product has homology to the *Drosophila Cc* protein, the yeast *IRA2* protein, and the *NF1* protein. The human gene appears to be a member of a gene family because Southern blot analysis of genomic DNA detected several bands (data not shown). Taken together, the prohibitin gene could be a new tumor suppressor gene in breast cancer. Further analysis of somatic and germ line mutations in the prohibitin gene will be able to reveal the significant association with tumor development and/or progression in primary breast cancer.

Acknowledgments

We are grateful to Drs. Mike Jones, Futoshi Akiyama, Goi Sakamoto, Mutsuko Miyagi, Takashi Imai, and Akira Horii for helpful advice and discussions. We would also like to thank Kiyoshi Noguchi for preparation of the manuscript.

References

1. Anderson, D. E. A genetic study of human breast cancer. J. Natl. Cancer Inst., *48:* 1029–1034, 1972.
2. Bishop, D. T., Cannon-Albright, L., McLellan, T., Gardner, E. J., and Skolnick, M. H. Segregation and linkage analysis of nine Utah breast cancer pedigrees. Genet. Epidemiol., *5:* 151–169, 1988.
3. Go, R. C. P., King, M. C., Bailey-Wilson, J., Elston, R. C., and Lynch, H. T. Genetic epidemiology of breast cancer and associated cancers in high-risk families. J. Natl. Cancer Inst., *71:* 455–461, 1983.
4. Newman, B., Austin, M. A., Lee, M., and King, M. C. Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. Proc. Natl. Acad. USA, *85:* 3044–3048, 1988.
5. Skolnik, M. H., Cannon-Albright, L. A., Goldgar, D. E., Ward, J. H., Marshall, C. J., Schumann, G. B., Hogel, H., McWhorter, W. P., Wright, E. C., Tran, T. D., Bishop, D. T., Kushner, J. P., and Eyre, H. J. Inheritance of proliferative breast disease in breast cancer kindreds. Science (Washington DC), *250:* 1715–1720, 1990.
6. Malkin, D., Li, F. P., Strong, L. C., Fraumeni, J. F., Nelson, C. E., Kim, D. H., Kassel, J., Gryka, M. A., Bischoff, F. Z., Tainsky, M. A., and Friend, S. H. Germ line mutations in familial syndrome of breast cancer, sarcomas, and other neoplasms. Science (Washington DC), *250:* 1233–1238, 1990.
7. Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B., and King, M. C. Linkage of early-onset familial breast cancer to chromosome 17q21. Science (Washington DC), *250:* 1684–1689, 1990.
8. Sato, T., Tanigami, A., Yamakawa, K., Akiyama, F., Kasumi, F., Sakamoto, G., and Nakamura, Y. Allelotype of breast cancer: Cumulative allele losses promote tumor progression in primary breast cancer. Cancer Res., *50:* 7184–7189, 1990.
9. Sato, T., Akiyama, F., Sakamoto, G., Kasumi, F., and Nakamura, Y. Accumulation of genetic alterations and progression of primary breast cancer. Cancer Res., *51:* 5794–5799, 1991.
10. Nuell, M. J., Stewart, D. A., Walker, L., Friedman, V., Wood, C. M., Owens, G. A., Smith, J. R., Schneider, E. L., Dell'Orco, R., Lumpkin, C. K., Danner, D. B., and McClung, J. K. Prohibitin, an evolutionarily conserved intracellular protein that blocks DNA synthesis in normal fibroblasts and HeLa cells. Mol. Cell. Biol., *11:* 1372–1381, 1991.
11. McClung, J. K., Danner, D. B., Stewart, D. A., Smith, J. R., Schneider, E. L., Lumpkin, C. K., Dell'Orco, R. T., and Nuell, M. J. Isolation of a cDNA that hybrid selects antiproliferative mRNA from rat liver. Biochem. Biophys. Res. Commun., *164:* 1316–1322, 1989.
12. Eveleth, D. D., and Marsh, J. L. Sequence and expression of *Cc* gene, a member of the dopa decarboxylase gene cluster of *Drosophila*: possible translational regulation. Nucleic Acids Res., *14:* 6169–6183, 1986.
13. White, J. J., Ledbetter, D. H., Eddy, R. L., Shows, T. B., Stewart, D. A., Nuell, M. J., Friedman, V., Wood, C. M., Owens, G. A., McClung, J. K., Danner, D. B., and Morton, C. C. Assignment of the human prohibitin gene

GeneDX 1035, pg. 230

(PHB) to chromosome ...id identification of a DNA polymorphism. Genomics, *11:* 228–230, ...

14. Nishisho, I., Nakamura, Y., Miyoshi, Y., Miki, Y., Ando, H., Horii, A., Koyama, K., Utsunomiya, J., Baba, S., Hedge, P., Markham, A., Krush, A. J., Petersen, G., Hamilton, S. R., Nilbert, M. C., Levy, D. B., Bryan, T. M., Preisinger, A. C., Smith, K. J., Su, L-K., Kinzler, K. W., and Vogelstein, B. Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. Science (Washington DC), *253:* 665–669, 1991.
15. Buchberg, A. M., Cleveland, L. S., Jenkins, N. A., and Copeland, N. G. Sequence homology shared by neurofibromatosis type-1 gene and *IRA-1* and *IRA-2* negative regulators of the *RAS* cyclic AMP pathway. Nature (Lond.), *347:* 291–294, 1990.
16. Xu, G., O'Connell, P., Viskochil, D., Cawthon, R., Robertson, M., Culver, M., Dunn, D., Stevens, J., Gesteland, R., White, R., and Weiss, R. The neurofibromatosis type 1 gene encodes a protein related to GAP. Cell, *62:* 599–608, 1990.
17. Tanaka, K., Nakafuku, M., Tamanoi, F., Kaziro, Y., Matsumoto, K., and Toh-e, A. *IRA2*, a second gene of *Saccharomyces cerevisiae* that encodes a protein with a domain homologous to mammalian *ras* GTPase-activating protein. Mol. Cell. Biol., *10:* 4303–4313, 1990.
18. Nigro, J. M., Baker, S. J., Preisinger, A. C., Jessup, J. M., Hostetter, R., Cleary, K., Bigner, S. H., Davidson, N., Baylin, S., Devilee, P., Glover, T., Collins, F ... Veston, A., Modali, R., Harris, C. C., and Vogelstein, B. Mutations ... the *p53* gene occur in diverse human tumor types. Nature (Lond.), *342:* 705–708, 1989.
19. Herskowitz, I. Functional inactivation of genes by dominant negative mutations. Nature (Lond.), *329:* 219–222, 1987.
20. Finlay, C. A., Hinds, P. W., and Levine, A. J. The *p53* proto-oncogene can act as a suppressor of transformation. Cell, *57:* 1083–1093, 1989.
21. Tanaka, K., Nakafuku, M., Satoh, T., Marshall, M. S., Gibbs, J. B., Matsumoto, K., Kaziro, Y., and Toh-e, A. *S. cerevisiae* genes *IRA1* and *IRA2* encode proteins that may be functionally equivalent to mammalian *ras* GTPase activating protein. Cell, *60:* 803–807, 1990.
22. Xu, G., Lin, B., Tanaka, K., Dunn, D., Wood, D., Gesteland, R., White, R., Weiss, R., and Tamanoi, F. The catalytic domain of the neurofibromatosis type 1 gene product stimulates *ras* GTPase and complements *ira* mutants of *S. cerevisiae*. Cell, *63:* 835–841, 1990.
23. Ballester, R., Marchuk, D., Boguski, M., Saulino, A., Letcher, R., Wigler, M., and Collins, F. The *NF1* locus encodes a protein functionally related to mammalian GAP and yeast *IRA* proteins. Cell, *63:* 851–859, 1990.
24. Hall, A. *ras* and GAP—who's controlling whom? Cell, *61:* 921–923, 1990.
25. Trahey, M., and McCormick, F. A cytoplasmic protein stimulates normal N-*ras* p21 GTPase, but does not affect oncogenic mutants. Science (Washington DC), *238:* 542–545, 1987.

GeneDX 1035, pg. 231

sequences. However, since the reversion rate in retroviral replication of such a 4-bp insertion is lower than 10^{-7} per insertion (*5*), this explanation is unlikely. Different markers in the two U5 regions may also result from mismatch repair of either the 5′ U5 or the 3′ U5 region. If the 5′ U5 region experienced the repair and resulted in a different marker than the 3′ U5 region, then the minus strand transfer had to be intermolecular (from WH204 RNA to WH13 RNA) to form the 3′ LTR. Suppose the minus strand strong stop DNA synthesis of WH13 RNA never initiated. As a result u5 and pbs were intact. Then it was possible for the minus strand DNA synthesis to continue and copy the U5 marker of WH13 (Bam HI). After completion of DNA synthesis and repair, the Bam HI marker remained in the 5′ U5 and resulted in different markers in the two U5 regions. It is also possible that the 3′ U5 region was repaired to a different marker. Then the minus strand strong stop transfer had to be intramolecular (from WH13 to WH13) to result in the 5′ WH13-like LTR. Suppose a strand displacement event occurred at the end of the 3′ LTR and formed a mismatch at the 3′ U5: the minus strand DNA contained a Bam HI site and the plus strand contained a Cla I site from the displaced fragment. After repair, the Cla I marker remained and resulted in the two different U5 markers. It is most likely that the different markers in the two U5 sequences were generated from mismatch repair of one of the LTR's. However, it is not clear which pathway generated this recombinant provirus.

27. Fisher's exact test was calculated with 21 recombinants. The recombinant with two different markers in the U5 was not included because it is not clear which type of minus strand transfer was used. Calculation including this sample was also performed, and the result indicated the distribution still is not random.
28. H. M. Temin, *Trends Genet.*, in press.
29. We thank Rebecca Wisniewski, Bonnie Fritz, and Kevin Krebsbach for technical assistance; N. Drinkwater, W. Sugden, and M. Susman for critical reading of the manuscript; M. Hannink, K. Iwasaki, V. Pathak, and G. Pulsinelli for discussion and helpful comments. Supported by Public Health Service grants CA-22443 and CA-07175 from the National Cancer Institute; a postdoctoral fellowship from the American Cancer Society (W.-S.H.); and an American Cancer Society Research Professorship (H.M.T.)

13 June 1990; accepted 19 September 1990

Germ Line p53 Mutations in a Familial Syndrome of Breast Cancer, Sarcomas, and Other Neoplasms

DAVID MALKIN, FREDERICK P. LI, LOUISE C. STRONG, JOSEPH F. FRAUMENI, JR., CAMILLE E. NELSON, DAVID H. KIM, JAYNE KASSEL, MAGDALENA A. GRYKA, FARIDEH Z. BISCHOFF, MICHAEL A. TAINSKY, STEPHEN H. FRIEND*

Familial cancer syndromes have helped to define the role of tumor suppressor genes in the development of cancer. The dominantly inherited Li-Fraumeni syndrome (LFS) is of particular interest because of the diversity of childhood and adult tumors that occur in affected individuals. The rarity and high mortality of LFS precluded formal linkage analysis. The alternative approach was to select the most plausible candidate gene. The tumor suppressor gene, p53, was studied because of previous indications that this gene is inactivated in the sporadic (nonfamilial) forms of most cancers that are associated with LFS. Germ line p53 mutations have been detected in all five LFS families analyzed. These mutations do not produce amounts of mutant p53 protein expected to exert a trans-dominant loss of function effect on wild-type p53 protein. The frequency of germ line p53 mutations can now be examined in additional families with LFS, and in other cancer patients and families with clinical features that might be attributed to the mutation.

IN 1969, LI AND FRAUMENI REVIEWED MEDICAL RECORDS and death certificates of 648 childhood rhabdomyosarcoma patients and identified four families in which siblings or cousins had a childhood sarcoma (*1*). These four families also had striking histories of breast cancer and other neoplasms, suggesting a new familial cancer syndrome of diverse tumors (Li-Fraumeni syndrome; LFS). Recently completed prospective studies have confirmed the high risk in family members of the tumor types that comprise LFS (*2*). Since the original description of the syndrome, systematic studies and anecdotal reports have confirmed its existence in various geographic and ethnic groups (*3*). The spectrum of cancers in the syndrome (Table 1) has been determined to include breast carcinomas, soft tissue sarcomas, brain tumors, osteosarcoma, leukemia, and adrenocortical carcinoma. Possible component tumors of LFS are melanoma, gonadal germ cell tumors, and carcinomas of the lung, pancreas, and prostate (*4*, *5*). These diverse tumor types in family members characteristically develop at unusually early ages, and multiple primary tumors are frequent.

To test the hypothesis that the Li-Fraumeni syndrome has a genetic etiology, Williams and Strong (*5*) applied segregation analysis and demonstrated that the observed cancer distribution in families best fit a rare autosomal dominant gene model. This model also predicted that the probability, for the families at risk, of developing any invasive cancer (excluding carcinomas of the skin)

D. Malkin, C. E. Nelson, D. H. Kim, J. Kassel, M. A. Gryka, and S. H. Friend are in the Division of Molecular Genetics, Massachusetts General Hospital Cancer Center, MGH East, Charlestown MA 02129. F. P. Li and C. E. Nelson are in the Clinical Epidemiology Branch, Division of Cancer Etiology, NCI, NIH, Bethesda MD, 20892 and in the Division of Biostatistics and Epidemiology, Dana-Farber Cancer Institute, Boston, MA, 02115. L. C. Strong is in the Division of Pediatrics, University of Texas M. D. Anderson Cancer Center, Houston, TX 77030. J. F. Fraumeni, Jr., is in the Epidemiology and Biostatistics Program, Division of Cancer Etiology, NCI, NIH, Bethesda MD, 20892. F. Z. Bischoff and M. A. Tainsky are in the Department of Tumor Biology, University of Texas M. D. Anderson Cancer Center, Houston TX, 77030. S. H. Friend is also in the Division of Hematology-Oncology, The Children's Hospital, Dana-Farber Cancer Institute, Department of Pediatrics, Harvard Medical School, Boston MA, 02114.

*To whom correspondence should be addressed.

GeneDX 1035, pg. 232

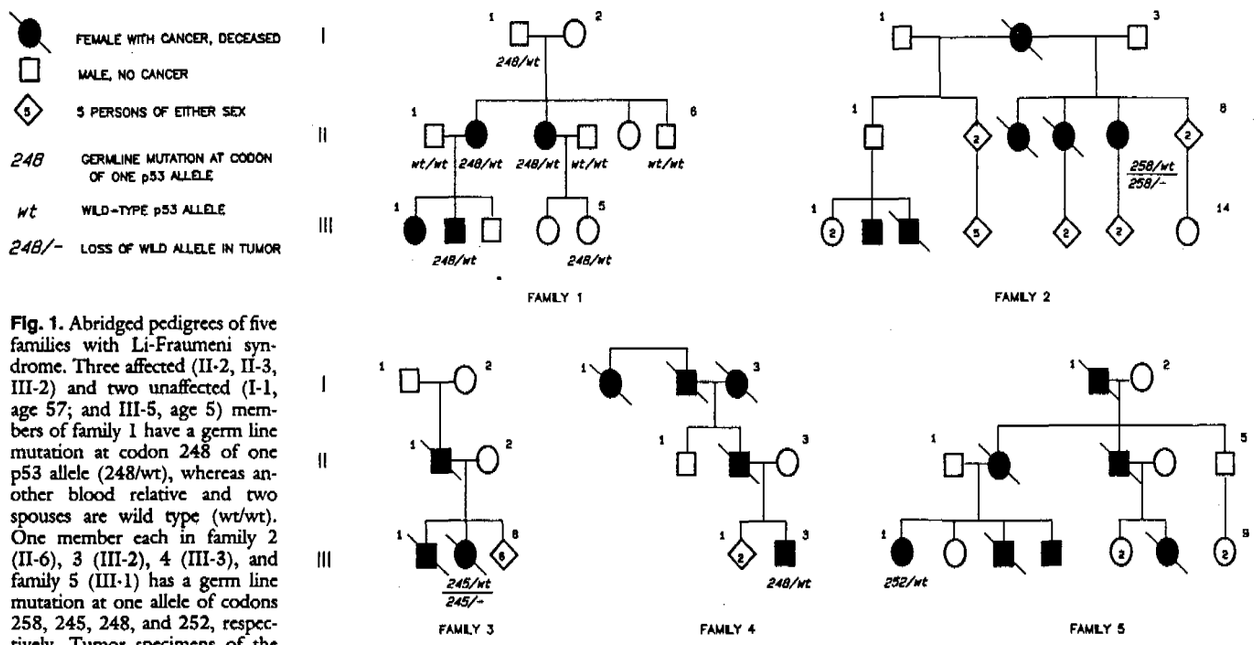


Fig. 1. Abridged pedigrees of five families with Li-Fraumeni syndrome. Three affected (II-2, II-3, III-2) and two unaffected (I-1, age 57; and III-5, age 5) members of family 1 have a germ line mutation at codon 248 of one p53 allele (248/wt), whereas another blood relative and two spouses are wild type (wt/wt). One member each in family 2 (II-6), 3 (III-2), 4 (III-3), and family 5 (III-1) has a germ line mutation at one allele of codons 258, 245, 248, and 252, respectively. Tumor specimens of the single cases in families 2 and 3 show loss of the remaining wild-type allele (258/- and 245/-, respectively). In family 1, II-2 developed unilateral breast carcinoma at age 30; II-3, bilateral breast carcinoma at 28; III-1, soft tissue sarcoma at age 1 and osteosarcoma at age 8; and III-2, brain tumor at age 5. In family 2, I-2 developed carcinoma of ampula of Vater at age 47 and bladder carcinoma at age 53; II-4, ovarian germ cell carcinoma at age 16; II-5, soft tissue sarcoma at age 21; II-6, unilateral breast carcinoma at age 34; III-3, osteosarcoma at age 20; III-4, brain tumor at age 16. In family 3, II-1, developed soft tissue sarcoma at age 58; III-1, osteosarcoma at age 11; and III-2, osteosarcoma at age 19. In family 4, I-1 developed bilateral breast carcinoma at age 68; I-2, melanoma at age 54; I-3, uterine carcinoma at age 66; II-2, melanoma at age 45 and brain tumor at age 52; and III-3, soft tissue sarcoma at age 5 and osteosarcoma at age 21. In family 5, I-1, developed a brain tumor at age 26; II-2, bilateral breast carcinoma at age 27; II-3, brain tumor at age 38; III-1, soft tissue sarcoma at age 2, osteosarcoma at age 15, and unilateral breast carcinoma at age 32; III-3, osteosarcoma at age 22; III-4, leukemia at age 17, and osteosarcoma at age 18; and III-7, brain tumor at age 9.

reaches almost 50 percent by age 30, when only 1 percent of the general population has developed cancer (*6*). More than 90 percent of the gene carriers would develop cancer by age 70. While the syndrome has thus been characterized statistically, identification of the gene and the mechanisms for the striking cancer predisposition has been elusive.

Difficulties in cloning the LFS gene arose from a number of restrictions placed on traditional methods. The accumulation of informative tissue specimens and data has been hampered by the rarity of the syndrome, by the high mortality among affected individuals, and by lack of an unambiguous definition of the syndrome. In addition, the cancers in relatives without the gene often cannot be distinguished from cancers associated with the gene. Linkage studies and searches for constitutional karyotypic alterations in patients with LFS have failed to specify the chromosomal region on which to focus. Furthermore, these families do not have any precancerous conditions or associated malformation syndromes such as aniridia, which was useful in mapping the Wilms tumor gene (WT1) to chromosome 11p13 (*7*). Because of these difficulties, we targeted potential candidate genes for study.

The class of genes most strongly associated with familial tumors has been the tumor suppressor genes. The first tumor suppressor gene to be cloned, the retinoblastoma gene (Rb) (*8*), seemed an unlikely candidate for a germ line mutation in LFS because retinoblastoma has not been observed in these families. The second tumor suppressor gene to be recognized was p53, located on the short arm of chromosome 17, band p13 (*9*). Inactivating mutations of p53 have been associated with sporadic osteosarcomas, soft tissue sarcomas, brain tumors, leukemias, and carcinomas of the lung and breast (*10, 11*). Together, these tumors also account for more than half of the cancers in selected series of LFS families. Furthermore, transgenic mice carrying a mutant p53 gene have an increased incidence of osteosarcomas, soft tissue sarcomas, adenocarcinomas of the lung, and adrenal and lymphoid tumors (*12*). These tumors all have been noted in families with LFS. We therefore decided to test whether families with LFS carry germ line p53 gene mutations.

Germ line p53 gene mutations in LFS families. Skin biopsies were obtained from three affected and five unaffected relatives in family 1 (see Fig. 1). Their fibroblasts were grown in short-term culture, and DNA was extracted by standard techniques (*13*). The polymerase chain reaction (PCR) was used to amplify the genomic region encoding exons 5 through 8 of the p53 gene (*14*). This segment contains a majority of the conserved domains and most of the mutations found in sporadic tumors (*10, 15*). The primers used for amplification were designed to contain Eco RI sites at their 5′ ends so that the 1685-bp amplified fragment could be subcloned for sequencing. Multiple primers were then used to sequence the entire genomic region on both strands (*16*). Each sequencing reaction was repeated on multiple clones. Three members in family 1 (II-1, II-4, and II-6) have the wild-type p53 sequence in six out of six clones (Table 2). These individuals are considered to carry two wild-type p53 alleles and have no evidence of tumors. DNA from separate clones representing the five other family members contained both wild-type p53, and p53 with a C to T mutation at the first nucleotide of codon 248. These individuals are considered to carry both a wild-type and a mutant p53 allele in their fibroblasts. Even though several polymorphisms have been identified in the p53 gene (*17*), none has been detected in the five conserved domains (*10, 11, 15, 18*). These findings are derived from the sequencing of the p53 gene in more than 150 samples of the DNA extracted from

GeneDX 1035, pg. 233

constitutional cells (*10*). The codon 248 mutation, which changes an arginine to tryptophan, occurs within conserved region IV that spans codons 234 to 258 (boxed area, Fig. 2). It has been shown that it is not a polymorphism; in fact, it is the most common p53 mutation in sporadic tumors studied to date (*10, 11, 18*). The three affected individuals tested in this family have the codon 248 mutation. The unaffected grandfather (I-2) and his granddaughter (III-5) also carry the codon 248 mutation and are carriers predicted to be at high risk of developing cancer.

Mutations in the p53 gene are documented to occur in the establishment of cells in long-term culture (*19, 20*). Thus, during in vitro passage of the LFS fibroblasts, the p53 gene might have acquired changes not present in the germ line. Although it is extremely unlikely that all five fibroblast lines acquired the codon 248 mutation, we collected blood from two of the individuals in this family (II-2 and II-3) who have breast cancer and also one mutant p53 allele in their fibroblasts. DNA was extracted from their lymphocytes after four passages in tissue culture. Lymphocytes from both sisters have a C to T mutation at the first nucleotide of codon 248 (Table 2). This is the same mutation found in fibroblasts of these patients.

To support the hypothesis that this p53 mutation represents a germ line mutation and not artifact of PCR, we studied segregation of the p53 chromosome region in family 1. We used YNZ22.1, a highly polymorphic DNA sequence on chromosome 17p that

Table 1. Cancers in 43 families with Li-Fraumeni Syndrome (LFS), by tumor type and age at diagnosis. All families were ascertained through a proband with sarcoma, who is excluded from the tabulations.

Tumor type	Age at diagnosis (years)*			
	0–14	15–44	>44	All ages
Component tumors of LFS				
Breast carcinoma	0	49	11	60
Soft tissue sarcoma	13	12	4	29
Brain tumors	12	15	1	28
Osteosarcoma	6	6	2	14
Leukemia	8	4	2	14
Adrenocortical carcinoma	5	0	0	5
Possible component tumors of LFS				
Lung carcinoma	0	7	12	19
Prostate carcinoma	0	0	8	8
Pancreas carcinoma	0	1	6	7
Melanoma	0	1	2	3
Other tumors				
Colorectal carcinoma	1	3	4	8
Lymphoma	0	5	1	6
Stomach carcinoma	0	3	1	4
Other	5	13	8	26
All cancers				
	50	119	62	231

*Only first cancer was counted in patients with multiple tumors.

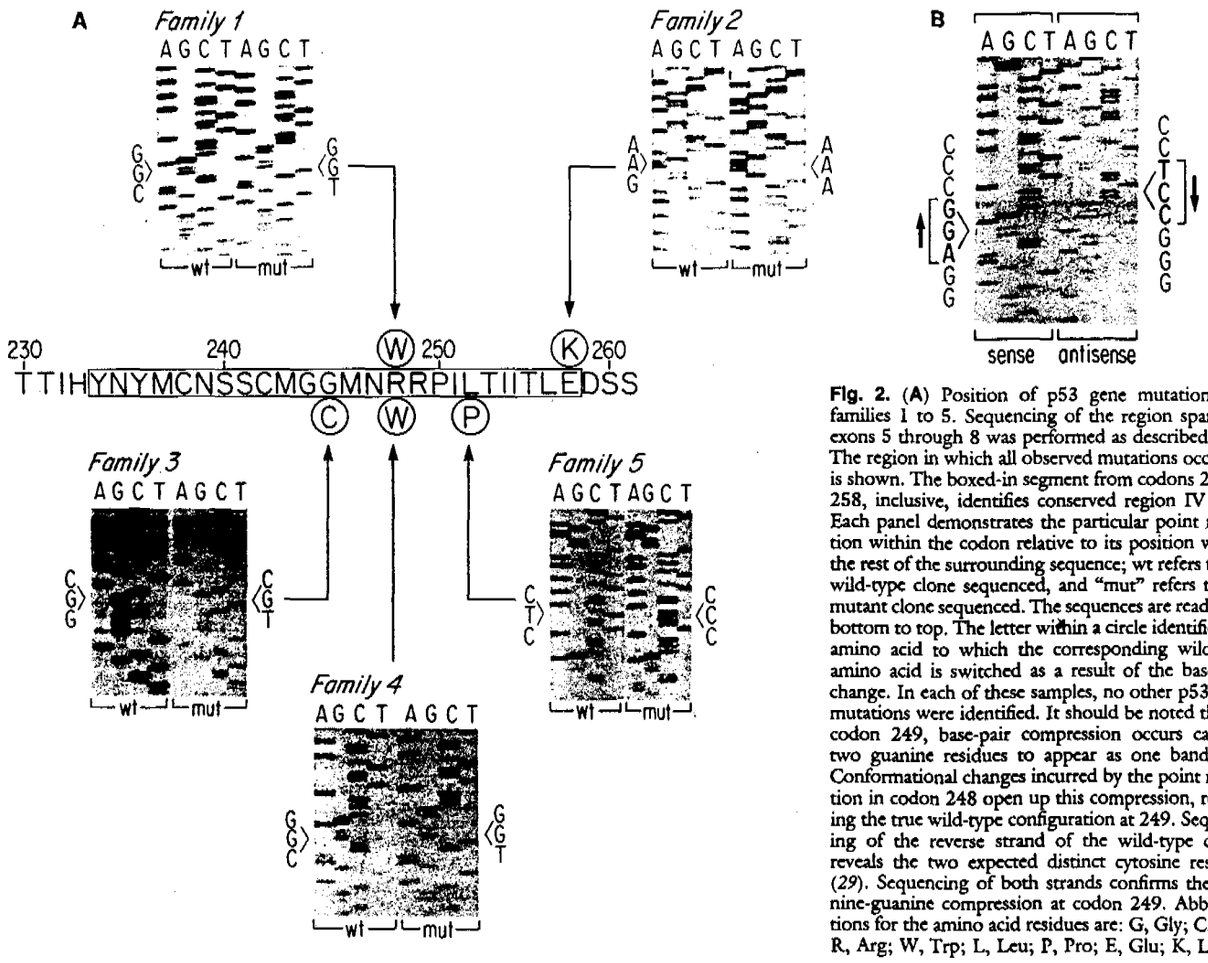


Fig. 2. (**A**) Position of p53 gene mutations for families 1 to 5. Sequencing of the region spanning exons 5 through 8 was performed as described (*16*). The region in which all observed mutations occurred is shown. The boxed-in segment from codons 234 to 258, inclusive, identifies conserved region IV (*15*). Each panel demonstrates the particular point mutation within the codon relative to its position within the rest of the surrounding sequence; wt refers to the wild-type clone sequenced, and "mut" refers to the mutant clone sequenced. The sequences are read from bottom to top. The letter within a circle identifies the amino acid to which the corresponding wild-type amino acid is switched as a result of the base-pair change. In each of these samples, no other p53 gene mutations were identified. It should be noted that in codon 249, base-pair compression occurs causing two guanine residues to appear as one band. (**B**) Conformational changes incurred by the point mutation in codon 248 open up this compression, revealing the true wild-type configuration at 249. Sequencing of the reverse strand of the wild-type clones reveals the two expected distinct cytosine residues (*29*). Sequencing of both strands confirms the guanine-guanine compression at codon 249. Abbreviations for the amino acid residues are: G, Gly; C, Cys; R, Arg; W, Trp; L, Leu; P, Pro; E, Glu; K, Lys.

GeneDX 1035, pg. 234

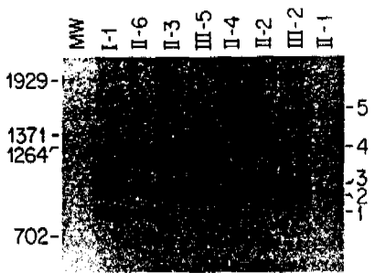


Fig. 3. DNA (Southern) gel blot analysis. Hybridization of the variable number tandem repeat (VNTR) polymorphic probe YNZ22.1 to genomic DNA from the eight members of family 1 indicated. A λBstE II DNA molecular marker is shown in the first lane. Numerical indexing of the VNTR alleles is indicated on the right.

identifies variable number tandem repeat (VNTR) sequences (*21*). DNA from the eight members of family 1 was digested with Hinf I and, after DNA (Southern) blot transfer, was hybridized to the YNZ22.1 probe (*22*). Five polymorphisms for the YNZ22.1 allele were present in this family (Fig. 3). Alleles were numbered from 1 through 5 according to size. Five individuals whose constitutional DNA contains a mutant p53 gene all carry the YNZ22.1 allele No. 3 (Table 2). This demonstrates the co-segregation data confirm that co-segregation of this YNZ22.1 allele and the mutant gene in the family. The segregation data confirm members from three generations of family 1 have the identical germ line p53 gene mutation.

We next examined four additional LFS families (Fig. 1, families 2 through 5) for germ line p53 gene mutations. In family 2, lymphocytes and a portion of the breast cancer were available from member II-6. DNA encoding exons 5 through 8 of the p53 gene was first amplified from the lymphocytes. Mutation of a G to an A at the first nucleotide of codon 258 was found in three of four clones. This results in a switch from a glutamic acid to a lysine at this residue. (Table 2 and Fig. 2). This mutation apparently has not been noted as a polymorphism (*10, 11, 18*). All six clones containing DNA amplified from the breast cancer had the same mutation at codon 258 of the p53 gene. This suggests that a mutant p53 allele is present in this patient's constitutional cells, and that the breast cancer cells lost the remaining wild-type p53 allele. Because the DNA of this patient was uninformative for two p53 polymorphic markers and uninformative for the YNZ22.1 allele, it is difficult to confirm that the cells in the breast tumor had no remaining wild-type p53 protein (*20, 21*).

Osteosarcoma cells, normal tissue adjacent to the tumor, and skin fibroblasts were obtained from individual III-2 in family 3. Both the adjacent normal tissue and skin fibroblasts had a mutation of a C to a T at the first nucleotide of codon 245, which results in cysteine replacing the glycine at this position. The same mutation was found in three of four clones analyzed from the individual's osteosarcoma. In this case, the YNZ22.1 polymorphic probe was informative and showed a 70 percent reduction of one allele in the tumor sample. The partial loss of the wild-type allele is due to admixture of normal tissue in this sample (*20*). This suggests that individual III-2 in family 3 has a germ line p53 mutation at codon 245, and that cells within the tumor lost the remaining wild-type p53 allele.

Lymphocytes from individual III-3 in family 4 were analyzed for p53 gene mutations. Clones containing amplified DNA encoding p53 had the same C to T transition at the first nucleotide in codon 248 as found in family 1. The high frequency of codon 248 mutations in sporadic tumors and its presence in two of the five LFS families suggest a propensity to change at this site. Individual III-1 in family 5 has a germ line mutation at codon 252. In the first position of this codon there is a T to C change, which results in proline replacing leucine at this position.

All p53 germ line mutations noted in the LFS families are located in conserved region IV, where no polymorphisms have been detected and where nucleotide changes have previously been limited

Table 2. Analyses of p53 gene mutations in members of five families with Li-Fraumeni syndrome.

Patient	Tumor type	Age at diagnosis (years)	DNA source	Ratio mut/total clones	Codon	YNZ22.1 VNTR
Family 1						
Affected						
II.2	Breast	30	Fibroblasts	3/4	248	3,4
			Lymphocytes	4/6	248	
II.3	Bilateral breast	28	Fibroblasts	2/6	248	3,4
			Lymphocytes	3/6	248	
III.2	Brain	5	Fibroblasts	2/6	248	3,2
Carriers						
I.2			Fibroblasts	3/6	248	3,1
III.5			Fibroblasts	4/6	248	3,2
Non-carriers						
II.1			Fibroblasts	0/6		2,2
II.4			Fibroblasts	0/6		2,5
II.6			Fibroblasts	0/6		1,4
Family 2						
II.6	Breast	34	Lymphocytes	3/4	258	
			Tumor	6/6	258	
Family 3						
III.2	Osteosarcoma	19	Fibroblasts	2/6	245	1,2
			Normal	2/4		
			Tumor	3/4		2
Family 4						
III.3	Soft tissue sarcoma	5	Lymphocytes	2/4	248	
	Osteosarcoma	21	Lymphocytes	2/6*		
Family 5						
III.1	Soft tissue sarcoma	2	Fibroblasts	1/4	252	
	Osteosarcoma	14	Fibroblasts	1/4		
	Breast	31				

*In samples from family 4, patient III.3, separate PCR reactions were analyzed as indicated.

GeneDX 1035, pg. 235

to tumor cells (*10, 11, 18*). All the mutations described in Table 2 were confirmed by a second PCR reaction and resequencing the gene products. Performing the PCR reactions in duplicate allowed us to overcome the inconsistencies due to the 0.25 percent rate of base misincorporation documented by others and previously confirmed in our laboratory (*10, 23*). Germ line p53 mutations were not identified in seven healthy controls and seven individuals with sporadic tumors who had no family history of cancer.

Possibility of restrictions on germ line p53 mutations. The half-lives of most mutant p53 proteins are markedly increased in comparison to the half-life of wild-type p53 protein. This is primarily a result of a complex they form with the heat shock proteins (hsc-70) (*24*). These mutant p53 proteins attain levels comparable to those known to exert a "trans-dominant loss of function" effect on wild-type p53 (*23, 25*). Other rare mutant p53 proteins that fail to form a complex with hsc-70 have a reduced capacity to override wild-type p53 function in transformation assays (*26*). We therefore asked whether such a trans-dominant loss of function effect of some mutant p53 proteins would restrict the types of germ line p53 mutations found in LFS families. The level of p53 protein was analyzed by immunoprecipitation of cell lysates from normal fibroblasts, LFS fibroblasts, and KHOS-240S cells (Fig. 4). The latter contain a p53 mutation at codon 156. This mutant protein forms a complex with hsc-70 (*27*). The high level of p53 protein found in the KHOS-240S cells is representative of the large amounts of protein found in cells with mutant p53 protein that complexes the hsc-70 proteins (*25*). In contrast, the p53 protein levels in the three LFS fibroblast lines are comparable to those of wild-type p53. The level of p53 protein in fibroblasts from individuals of 12 LFS families (including the four with documented germ line p53 mutations analyzed in Fig. 4), were also tested by immunofluorescence studies. All had low fluorescent intensities comparable to cells containing wild-type p53 protein (*20*). Both the immunoprecipitation and immunofluorescence studies probably indicate a selection against the p53 gene mutations that stabilize the mutant p53 protein that exerts a trans-dominant loss of function effect on the wild-type protein. Screening tests to determine germ line p53 mutations should not rely on methods that assume that the mutations will result in high levels of mutant protein.

Interpretations of the data. The presence of p53 mutations in the sporadic forms of many human tumors suggests that these alterations are an important step in the transformation of diverse cells. All cells in the individuals with LFS that we studied presumably have a single wild-type p53 allele. This finding provides an opportunity to compare the effects of p53 inactivation on the development of cancer in different tissues. The most remarkable feature of this syndrome is the diversity of tumors in the gene carriers. This suggests that numerous tissues are rendered more susceptible to transformation when they have only one wild-type p53 allele. The inactivation of p53 must therefore be at least one of the rate-limiting steps in the transformation of many tissues. However, tumors in LFS are not evenly distributed by site and do not follow cancer frequency patterns in the general population (Table 1). For example, breast carcinoma and colon carcinoma have comparable frequencies in the general population, and frequently have associated p53 mutations (*6*). In the LFS families, breast cancer occurs much more frequently than colon carcinoma (Table 1). The difference suggests that p53 inactivation is a more critical rate-limiting step in breast carcinoma formation than it is in colon carcinoma formation. Other patterns of tumor frequencies among affected individuals with LFS can provide additional clues regarding the differences in tissue-specific growth control pathways.

The development of molecular diagnostic tests to identify carriers of germ line p53 mutations provides a new tool for screening members of LFS families. Recently, an International Working Group on LFS was established, and nearly 100 candidate families have been enrolled. Members of these families can be examined for germ line p53 mutations. This effort should help clarify the spectrum of cancers attributable to germ line p53 gene alterations. In addition, gene carriers will be identified among the unaffected children and older relatives. A determination can also be made of the family members who are in fact, not carriers. The ability to detect germ line p53 mutations creates a responsibility to properly use this information, and to develop strategies for the care of gene carriers at exceptionally high risk of cancer.

The frequency of individuals in the general population who carry germ line p53 mutations is undefined. Li-Fraumeni syndrome might identify only the rare families in which germ line mutations in the p53 gene are highly penetrant and expressed in multiple tissues. The role of p53 germ line mutations in human cancer cannot be fully addressed by assessing its role in familial cancer. In retinoblastoma, for example, 85 percent of germ line mutation carriers, as identified by the presence of bilateral tumors, occur de novo (*28*). A germ line p53 mutation could occur in a substantial population of children and young adults who die of the cancer and do not propagate the gene. We are therefore seeking p53 mutations in individuals and families who display one or more features of LFS, such as multiple primary cancers or breast cancer in several young relatives. Those with a germ line p53 gene mutation are enrolled in a multicenter registry for further analyses.

In conclusion, we have shown that alterations of the p53 gene occur not only as somatic mutations in human cancers, but also as germ line mutations in some cancer-prone families. The frequency of germ line p53 mutations in cancer patients in the general population is unknown and merits further study. Both somatic and germ line mutations also arise in another tumor suppressor gene, the retinoblastoma gene. These finding indicate that the search for additional tumor suppressor genes can proceed along two parallel research

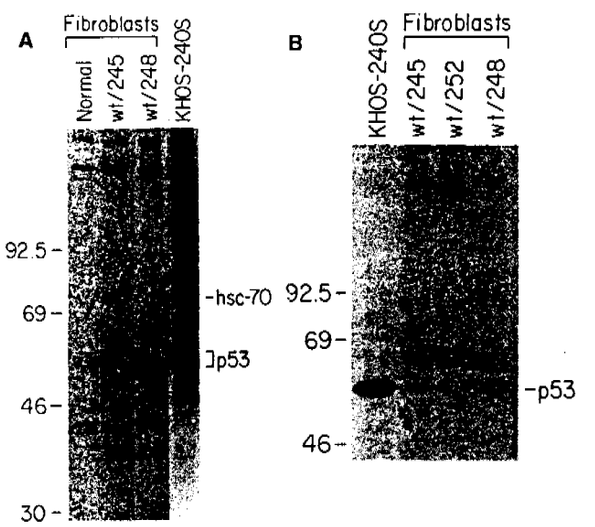


Fig. 4. Amount of p53 protein produced by various p53 alleles. Lysates from [^{35}S]methionine-labeled fibroblasts and KHOS-240S osteosarcoma cells were immunoprecipitated with monoclonal antibody pAb421 (*20, 30*). (**A**) (Lane 1) Normalfibroblasts; lane 2, wt/245 fibroblasts from (III-2) family 3 with a mutant codon 245 p53 allele; lane 3, wt/248 fibroblasts from (II-2) family 1 with a mutant codon 248 p53 allele; lane 4, KHOS-240S cells with a mutant codon 156 p53 allele. (**B**) (Lane 1) KHOS-240S cells as in Fig. 4A; (lane 2), the wt/245 mutant as in Fig. 4A; (lane 3) wt/252 fibroblasts from (III-1) family 5 with a mutant codon 252 p53 allele; (lane 4) wt/248 as in parens.

GeneDX 1035, pg. 236

pathways. One is through analysis of the loss of heterozygosity in tumor samples, and the second is through the identification and study of cancer-prone families.

REFERENCES AND NOTES

1. F. P. Li, and J. F. Fraumeni, Jr., *Ann. Intern. Med.* **71**, 747 (1969).
2. _____, *J. Am. Med. Assoc.* **247**, 2692 (1982); J. E. Garber *et al.*, *Proc Am. Assoc. Cancer Res.* **31**, 210 (1990).
3. J. M. Birch *et al.*, *J. Clin. Oncol.* **8**, 583 (1990); J. M. Birch *et al.*, *Br. J. Cancer* **49**, 325 (1984).
4. F. P. Li *et al.*, *Cancer Res.* **48**, 5358 (1988).
5. W. R. Williams and L. C. Strong, *Familial Cancer, 1st International Research Conference* (Karger, Basel, 1985), p. 151; L. C. Strong *et al.*, *J. Natl. Cancer Inst.* **79**, 1213 (1987).
6. J. L. Young, C. L. Perry, A. J. Asire, Eds., *Natl. Cancer Inst. Monogr.* **57**, KY72 (1981) (NIH publication no. **81-2330**).
7. V. M. Ricarrdi *et al.*, *Pediatrics* **61**, 604 (1978); U. Francke, L. B. Holmes, L. Atkins, V. M. Riccardi, *Cytogenet. Cell Genet.* **24**, 185 (1979).
8. S. H. Friend *et al.*, *Nature* **323**, 643 (1986).
9. C. A. Finlay, P. W. Hinds, A. J. Levine, *Cell* **57**, 1083 (1989); D. N. Eliyahu, D. Michalovitz, S. Eliyahu, O. Pinhasi-Kimhi, M. Oren, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8763 (1989); M. Isobe, B. S. Emanuel, D. Giron, M. Oren, C. Croce, *Nature* **320**, 84 (1986).
10. J. M. Nigro *et al.*, *Nature* **342**, 705 (1989).
11. J. W. Romano *et al.*, *Oncogene* **4**, 1483 (1989); L. M. Mulligan *et al. Proc. Natl. Acad. Sci.*, 5863 (1990); H. Ahuja, M. Bar-Eli, S. H. Advani, S. Benchimol, M. J. Cline, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6783 (1989); J. Bartek, R. Iggo, J. Gannon, D. P. Lane, *Oncogene* **5**, 893 (1990); R. Iggo, K. Gatter, J. Bartek, D. Lane, A. L. Harris, *Lancet* **335**, 675 (1990); S. J. Baker *et al.*, *Science* **244**, 217 (1989); A. G. Menon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5435 (1990); T. Takahashi *et al.*, *Science* **246**, 491 (1989); B. Seizinger, unpublished results.
12. A. Lavigueur *et al.*, *Mol. Cell Biol.* **9**, 3982 (1989).
13. All cell lines were maintained in Dulbecco's modified Eagle medium with glucose at 4.5 g/liter, L-glutamine, and sodium pyruvate. The medium was supplemented with 10 percent inactivated fetal calf serum and antibiotics. Cells were grown at 37°C in a humidified atmosphere containing 5 percent CO_2. Genomic DNA was extracted according to standard techniques (J. Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, NY, ed. 2, 1989). To extract genomic DNA from blood lymphocytes, we first separated the cells from other blood components by passage through a Ficoll-Hypaque gradient. Tumor samples were processed after grossly contaminating normal tissue was removed. The tumor was then gently teased and homogenized into small fragments, which were resuspended in lysis buffer. The DNA was extracted from the resuspended tissue.
14. A 1.7-kb fragment was generated from 0.5 to 2.0 μg of the genomic DNA template. The sense and antisense primers used were, respectively. F1, 5′-ATGC GAATT CCCCT GCCCT CAACA AGAT-3′ starting at the first codon downstream to intron 4, and R2, 5′-TATAG GAATT CGTGG TGAGG CTCCC CTT-3′ starting at the ninth codon upstream of intron 8. Each primer had synthetic Eco RI sites at their 5′ ends to facilitate cloning. The DNA template and primers were added to a PCR buffer (DuPont) containing 50 mM tris-HCl (pH 9.5), 1.5 mM magnesium chloride, and 20 mM ammonium sulfate, to which were added dCTP, dATP, dTTP, and dGTP (at 0.2 mM each), and 0.5 U of Replinase (DuPont). The 100-μl reaction was amplified in a Perkin-Elmer Thermocycler (Cetus) for 35 cycles of 94°C (1 minute) 55°C (2 minutes), and 72°C (3 minutes). The reaction was initiated with one 80-second cycle at 94°C and ended with a 7-minute step at 72°C.
15. T. Soussi *et al.*, *Oncogene* **5**, 945 (1990).
16. Ten reactions were performed for each DNA template, and the PCR products were pooled and digested with Eco RI, fractionated by electrophoresis, and after acid-phenol extraction and purification from low-melt agarose (New England Biolabs), fragments were ligated into a Bluescript vector "PBSK" (Stratagene) that had been digested with Eco RI and treated with calf alkaline phosphatase. After transformation, individual clones derived from standard minipreparations were sequenced in both directions with primers derived from the p53 coding and intron sequences. Analysis of six clones allows detection of both alleles more than 98 percent of the time. The use of either ^{35}S-labeled or ^{32}P-labeled nucleotides with T7 DNA polymerase yielded high autoradiographic resolution on a 12 percent polyacrylamide gel. The primer 5N, 5′-GGTTGGCTCTGACTGTA-3′, started at the first codon of exon 7 and generated the sequence of the region that contained the mutations described in the text.
17. V. L. Buchman, P. M. Chumakov, N. N. Ninkina, O. P. Samarina, G. P. Georgiev, *Gene* **70**, 245 (1988).
18. B. Vogelstein, unpublished results.
19. A. J. Levine, unpublished results.
20. S. H. Friend, unpublished results.
21. Y. Nakamura *et al.*, *Nucleic Acids Res.* **16**, 5707 (1988); U. Masharani, D. Wolf, P. M. Frossard, *ibid.* **16**, 7757 (1988).
22. Southern blotting techniques were performed as described [J. Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, ed. 2, 1989), p. 9.31] for the capillary transfer method, Hybond nylon membranes (Amersham), and nonformamide containing hybridization solutions. The YNZ22.1 hybridization probe (*20*) was labeled by random priming the Klenow fragment of DNA polymerase, and [α-^{32}P]dCTP. Hybridization of the probe to 12 μg of genomic DNA digested with Hinf I was carried out at 65°C for 12 to 48 hours, and the blots were exposed to film for 3 to 15 days.
23. L. Diller *et al.*, *Mol. Cell. Biol.* **10**, 5772 (1990).
24. H. W. Sturzbecher, P. Chumakov, W. J. Welch, J. R. Jenkins, *Oncogene* **1**, 201 (1987); P. W. Hinds, C. A. Finlay, A. B. Frey, A. J. Levine, *Mol. Cell. Biol.* **7**, 2863 (1987); C. A. Finlay *et al.*, *ibid.* **8**, 531 (1988).
25. P. W. Hinds, C. Finlay, A. J. Levine, *J. Virol.* **63**, 739 (1989).
26. P. W. Hinds *et al.*, *Cell Growth Dev.*, in press.
27. J. W. Romano *et al.*, *Oncogene* **4**, 1483 (1989).
28. J. Wiggs *et al.*, *N. Engl. J. Med.* **318**, 151 (1988).
29. The presence of a guanine-guanine compression at codon 248 in the wild-type p53 gene is confirmed by sequencing both strands.
30. In preparation for immunoprecipitations, cell lines were metabolically labeled (at 50 μCi/ml) with methionine-free DME (ICN) in 2.5 ml of methionine-free DNA plus 2 percent dialyzed fetal calf serum for 2 hours. Cells were rinsed in ice-cold PBS twice before lysis in 1 ml of ELB (250 mM NaCl, 0.1 percent NP40, 50 mM Hepes, pH 7.0, 1 mM PMSF (Sigma), 5 mM EDTA, 0.5 mM DTT). The lysate was scraped from the plate, centrifuged, and stored at −80°C. Lysates were first adsorbed with fixed *Staphylococcus* protein A overnight at 4°C, then incubated with PAb122, before precipitation and loading on the 8 percent polyacrylamide gels.
31. We thank M. Dreyfus for data management; J. Garber, J. Schneider, and T. Russell for collecting specimens on families 1, 2, and 3; P. Begin and J. A. Cook for collecting specimens on families 4 and 5; W. W. Nichols for establishing cell lines on these two families; and S. E. Lux IV for important discussions, as well as S. Orkin and K. J. Isselbacher for critical review of the manuscript. Supported in part by the Baylor Fund Sub-Allocation for Research Program, Lucille P. Markey Foundation, and the Merck Foundation; a Research Fellowship of the Medical Research Council of Canada (D.M.); NIH grant 34936 (L.C.S. and M.A.T.); NIH grant 5-T32-CA09299 (F.Z.B); an Akvan T. and Viola D. Fuller American Cancer Society Fellow Junior Research Fellowship (D.H.K.); and a Lucille P. Markey Scholarship (S.H.F.).

24 September 1990; accepted 29 October 1990

GeneDX 1035, pg. 237

BRCA1: More than a hereditary breast cancer gene?

Jeff Boyd

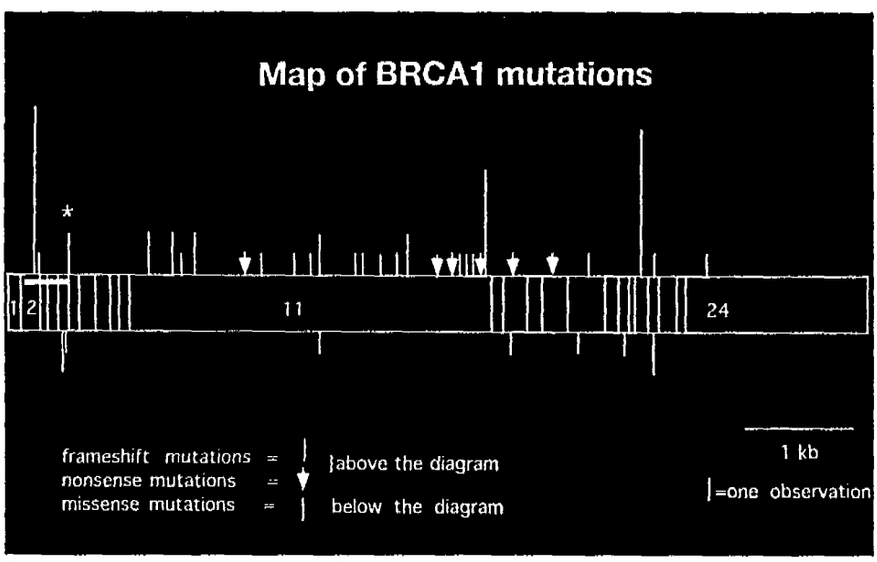
Department of Obstetrics and Gynecology, University of Pennsylvania Medical Center, Philadelphia, Pennsylvania 19104, USA

Following the discovery by Mark Skolnick and colleagues of a hereditary breast and ovarian cancer susceptibility gene, *BRCA1*, some six months ago[1], other laboratories which had been devoted to that arduous search have rapidly confirmed the 'candidate' gene as correct[2–5], and many have set off looking for the next destination in the world of breast cancer genetics — *BRCA2*. As recent history will attest, however, finding a major disease gene is only a beginning, to be followed inevitably by a period of surprises, contradictions and often a frustrating lack of progress. This issue of *Nature Genetics* contains three important — and perhaps surprising — insights into the relevance of *BRCA1* to human breast and ovarian cancer[6–8].

Among the many reasons for great excitement prior to the cloning of *BRCA1* was the expectation that somatic mutations in the gene would be found to contribute to tumorigenesis in the far more common sporadic forms of breast and ovarian carcinoma. This hypothesis was predicated on the observation that loss of heterozygosity (LOH), the molecular genetic hallmark of tumour suppressor gene inactivation, was common in sporadic breast and ovarian cancers in the region of chromosome 17 known to include *BRCA1*. Up to 80% of sporadic breast cancers[9,10] and a similar number of sporadic ovarian cancers[11–13] contain allelic deletions on 17q. Thus the surprise when Futreal and colleagues[14] reported that breast and ovarian cancers selected for 17q LOH did not appear to contain somatic mutations in *BRCA1*. Several mutations were found in their collection of tumours, but all were present in the germlines of individuals who, retrospectively, could be considered at high risk based on remarkable medical or family histories. While the genetics community deftly catalogued a list of germline *BRCA1* mutations in members of high-risk families[5] (Fig. 1), evidence of somatic mutations in sporadic breast and/or ovarian cancers remained conspicuously unreported.

Could it be that *BRCA1* plays a role in the development of hereditary cancers but not sporadic tumours? Although initially unpalatable, such a reality is not without precedent. As Vogelstein and Kinzler recently pointed out in the same context[15], the *WT1* tumour suppressor gene is mutated in most familial Wilms tumours but rarely in its sporadic manifestation; similarly, mutations in the family of genes encoding mismatch repair proteins responsible for hereditary nonpolyposis colorectal cancer have thus far been difficult, but not impossible, to associate with sporadic tumours[16]. Fueling skepticism of the involvement of *BRCA1* in sporadic cancer were additional reports of LOH in sporadic breast and ovarian tumours in which more rigorous deletion mapping implicated at least one, and perhaps two, additional region(s) harbouring putative tumour suppressor genes distal to the *BRCA1* locus on 17q (refs 17–19).

Just when it seemed that the only logical course was to begin rethinking the role of *BRCA1* in sporadic tumours, we learn of reassuring news from Merajver *et al.*[6] and Hosking *et al.*[7], who describe a total of five somatic *BRCA1* mutations occurring in ovarian cancers presumed to be sporadic. All of the mutations are likely to be of functional significance; four are frameshifts resulting in a prematurely truncated protein, while the fifth is a previously described missense mutation in the zinc-binding domain of the protein. The frequency of mutations was relatively low; Merajver *et al.* screened 47 unselected ovarian tumours and Trowsdale *et al.* examined 17 tumours preselected for 17q LOH. Nevertheless, these findings are the first to implicate *BRCA1* directly in sporadic tumorigenesis. The somatic nature of



D. Shattuck-Eidens

Fig. 1 Summary of germline mutations in *BRCA1*. A large collection of researchers has recently compiled an extensive list of *BRCA1* mutations in families with breast and/or ovarian cancer[5]. The positions of these 38 distinct mutations are depicted in relation to the 24 exons of *BRCA1*, with frameshift and nonsense mutations indicated above and missense mutations underneath the transcript. *, A splice site mutation in intron 6. (Not shown is a putative 'regulatory' mutation.) The height of each line designating a mutation is proportional to the number of independent observations. The three most common germline mutations are frameshifts that occur in exon 2 (185delAG); exon11 (4184del4); and exon 20 (5382insC).

A similar spread of mutations is seen among the sporadic *BRCA1* mutations in ovarian tumours which have been identified. These consist of single nucleotide insertions in exons 2 and 22; a missense mutation at Cys 61 (exon 5); a nonsense mutation in exon 19; and a frameshift leading to premature truncation of the BRCA1 protein (exon 23)[6,7].

GeneDX 1035, pg. 238

the mutations is unequivocal based on the presence of only the wild-type sequence in DNA samples from matched normal tissues, and all five are accompanied by somatic LOH in the *BRCA1* region, consistent with the paradigm for tumour suppressor gene inactivation.

Interesting though these findings are, it remains curious that *BRCA1* alterations in sporadic breast cancers have yet to be described, in spite of a mutation search the intensity of which is probably without parallel in cancer molecular genetics. There are several possible interpretations of the available data. Coding region mutations may be prevalent in sporadic breast and ovarian cancers, but have simply been missed. This is unlikely for technical reasons, as many of the same groups describing germline mutations have been searching for mutations in sporadic tumours. A more likley scenario would be that the mutational spectrum in sporadic tumours is distinct from the that of familial cases, and that alterations in noncoding intronic or promoter regions have thus far defied detection by all researchers.

'BRCA1' misnomer?

Alternatively, a higher frequency of mutations in ovarian than in breast cancers may ultimately render '*BRCA1*' (for 'Breast Cancer 1') a misnomer in one sense: the great majority of families with dominant inheritance of ovarian cancer are linked to *BRCA1*[21,22] and at least some fraction of sporadic ovarian cancers (10% or so) now appears to contain *BRCA1* mutations. In contrast, less than half of the families displaying dominant transmission of breast cancer are thought to be linked to *BRCA1*[21], while acquired mutations in sporadic breast tumours are still undescribed. An important task would then be to determine the genetic driving force of the frequent LOH occuring in the *BRCA1* region in sporadic breast cancers, presumably an additional tumour suppressor gene nearby. The situation inspires recollection of the colon cancer story in which *MCC*, originally a candidate gene for familial adenomatous polyposis[23], was found to be mutated in sporadic but not familial tumours and was only about 100 kb distant from the *APC* gene that was ultimately shown to be responsible for the hereditary syndrome[24,25].

Consistent with a much wider potential role for *BRCA1* in breast tumorigenesis, however, are the findings described by Thompson *et al.* in this issue[8], who report that *BRCA1* antisense oligonucleotides stimulate the proliferation of normal and malignant mammary epithelial cells, concomitant with the down-regulation of *BRCA1* expression. In addition, *BRCA1* mRNA levels are significantly decreased in cells from sporadic breast carcinomas compared to normal breast epithelial cells. Since both alleles were typically coordinately expressed in the breast cancers studied, the authors speculate that this down-regulation results from an altered upstream regulatory mechanism. Their finding of one tumour with allele-specific expression, together with a similar case reported by Miki *et al.*[1], suggests that regulatory mutations (in the promoter, for example) are also possible. If confirmed, these findings would greatly expand the relevance of *BRCA1* to human breast cancer. Confirmation or clarification of these observations is critical, however, as they are in direct contrast to those reported by Futreal *et al.*, in which altered *BRCA1* mRNA levels were not observed in any breast or ovarian cancer tested[14].

The spectacular gestational history of *BRCA1* may prove difficult to match with its ultimate impact on human cancer. Nevertheless, it would seem that the gene still has great potential for our understanding of breast and ovarian cancer, and that critical epigenetic functions in pathways of development and tumorigenesis may prove to be among its most important attributes. □

1. Miki, Y. *et al. Science* **266**, 66–71 (1994).
2. Castilla, L. *et al. Nature Genet.* **8**, 387–391 (1994).
3. Simard, J. *et al. Nature Genet.* **8**, 392–398 (1994).
4. Friedman, L. *et al. Nature Genet.* **8**, 399–404 (1994).
5. Shattuck-Eidens, D. *et al. J. Am. med. Assoc.* **273**, 535–541 (1995).
6. Merajver, S.D. *et al. Nature Genet.* **9**, 439–443 (1995).
7. Hosking, L. *et al. Nature Genet.* **9**, 343–344 (1995).
8. Thompson, M.E. *et al. Nature Genet.* **9**, 444–450 (1995).
9. Futreal, P.A. *et al. Cancer Res.* **52**, 2624–2627 (1992).
10. Cropp, C.S. *et al. Cancer Res.* **54**, 2548–2551 (1994).
11. Russell, S.E.H. *et al. Oncogene* **5**, 1581–1583 (1990).
12. Yang-Feng, T.L. *et al. Int. J. Cancer* **54**, 546–551 (1993).
13. Cilby, W. *et al. Cancer Res.* **53**, 2393–2398 (1993).
14. Futreal, P.A. *et al. Science* **266**, 120–122 (1994).
15. Vogelstein, B. & Kinzler, K.W. *Cell* **79**, 1–3 (1994).
16. Liu, B. *et al. Nature Genet.* **9**, 48–55 (1995).
17. Jacobs, I.J. *et al. Cancer Res.* **53**, 1218–1221 (1993).
18. Saito, H. *et al. Cancer Res.* **53**, 3382–3385 (1993).
19. Cropp, C.S. *et al. Cancer Res.* **53**, 5617–5619 (1993).
20. Godwin, A.K. *et al. Am. J. hum. Genet.* **55**, 666–677 (1994).
21. Easton, D.F. *et al. Am. J. hum. Genet.* **52**, 678–701 (1993).
22. Steichen-Gersdorf, E. *et al. Am. J. hum. Genet.* **55**, 870–875 (1994).
23. Kinzler, K.W. *et al. Science* **251**, 1366–1370 (1991).
24. Nishisho, I. *et al. Science* **253**, 665–669 (1991).
25. Joslyn, G. *et al. Cell* **66**, 601–613 (1991).

Progress in prostate cancer

Lisa Cannon-Albright[1] & Rosalind Eeles[2]

[1]*Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah 84132, USA*
[2]*Institute of Cancer Research & Royal Marsden Hospital, Sutton, Surrey SM2 5NG, UK*

Research into the molecular genetics of prostate cancer has lagged noticeably compared to other common cancers, at least until recently. This is surprising considering the impact of prostate cancer in the Western world. The incidence of prostate cancer is rising by 1% per year in Europe and it is the most common malignancy and the second leading cause of male cancer deaths in the United States. In 1994, there were 200,000 new cases and 38,000 deaths from the disease in the United States[1] and it is estimated that approximately one of every 11 American men will eventually develop prostate cancer.

Although the molecular mechanisms involved in the progression of prostate cancer remain relatively unknown, the multistep genetic pathway of carcinogenesis that has been demonstrated to occur in several common cancers offers a promising model. Most notably, Vogelstein and colleagues have demonstrated that the progression from colorectal polyp to malignant colon tumour stems from a complex series of genetic alterations at a number of different loci. Many genes and candidate regions have also been tied to the initiation and/or progression of breast cancer, and there

GeneDX 1035, pg. 239

N-632

3312 E

nature genetics



BRCA1 mutations

Somatic mutations in the *BRCA1* gene in sporadic ovarian tumours

Sofia D. Merajver[1], Trinh M. Pham[1], Rosemarie F. Caduff[2], Martha Chen[1], Ellen L. Poy[2], Kathleen A. Cooney[1], Barbara L. Weber[3], Francis S. Collins[4], Carolyn Johnston[5] & Thomas S. Frank[2]

The *BRCA1* gene on chromosome 17q21 is responsible for an autosomal dominant syndrome of increased susceptibility to breast and ovarian cancer but no somatic mutations in tumours have yet been described. To study the potential role of *BRCA1* in sporadic carcinogenesis, we analysed the genomic DNA of tumour and normal fractions of 47 ovarian cancers for mutations in *BRCA1* using the single-strand conformation polymorphism technique. We now describe somatic mutations in the DNA of four tumours which also had loss of heterozygosity (LOH) at a *BRCA1* intragenic marker. Our data support a tumour suppressor mechanism for *BRCA1*; somatic mutations and LOH may result in inactivation of *BRCA1* in at least a small number of ovarian cancers.

[1] *Department of Internal Medicine,* [2] *Department of Pathology, and* [3] *Department of Obstetrics and Gynecology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA*
[3] *Departments of Genetics and Internal Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA*
[4] *National Center for Human Genome Research, National Institutes of Health, Bethesda, Maryland 20892, USA*

Correspondence should be addressed to S.D.M.

An autosomal dominant susceptibility gene for breast and ovarian cancer, *BRCA1*, was recently isolated[1,2] after an arduous four-year search which began in 1990 with its localization to chromosome 17q21 (ref. 3). Analyses of breast and ovarian kindreds identified an inherited germline mutation in approximately 50–60% of cases[4–6]. Earlier this year, a collaborative study[7] of 327 unrelated patients from breast/ovarian families surveyed 38 distinct *BRCA1* mutations in 80 individuals. No statistically significant association was found between the type or nature of a mutation and the risk of breast or ovarian cancer. Loss of function of the BRCA1 protein appears to confer an 87% lifetime risk of breast cancer and a 44% risk of ovarian cancer in breast/ovarian families[8,9]. The risk of ovarian cancer is highly variable; for example, 45% of site-specific breast cancer families (and 67% of breast cancer families with average age of onset under 45 years) are linked to *BRCA1* but have no cases of ovarian cancer[8]. The risks of prostate and colon cancer are also increased in both breast and breast/ovarian *BRCA1* families[9]. The molecular basis for the heterogeneity of penetrance, organ of origin of the cancer, age of onset, and laterality of the lesions has not yet been elucidated. However, knowledge of the pathways from loss of BRCA1 function to cancer will be important for reliable predictive testing and for the design of targeted therapeutic strategies.

Observations of loss of heterozygosity (LOH) for polymorphic markers widely spaced along chromosome 17q in both familial and sporadic tumours has led to the hypothesis that *BRCA1* is a tumour suppressor gene[10–16]. In Knudson's model of carcinogenesis by a tumour suppressor gene, functional loss of both *BRCA1* alleles in the affected tissue(s) is necessary for malignant transformation to occur. Affected individuals in families are presumed to be born with a germline defect in the *BRCA1* gene, which constitutes the first hit in the path to tumorigenesis. The second defect is acquired in the wild-type allele of the breast or ovarian tissue of these patients, thereby contributing to the development of cancer. Disruption of the second allele can lead to an interstitial chromosomal deletion, detected through loss of alleles at flanking polymorphic markers. Thus, LOH is widely held to be an indicator of the presence of a tumour suppressor gene adjacent to the deleted markers.

Smith *et al.*[11] proved that in familial tumours with LOH, the wild-type allele was lost. In one recent study of familial breast and ovarian tumours[17], 47% of tumours from known linked families showed LOH at closely spaced markers within and flanking *BRCA1*; the minimal LOH region in those samples encompassed the known *BRCA1* gene but not 17q23 (ref. 15), 17q25 (ref. 15), or the region between *D17S846* and *D17S746* (ref. 16) where other putative tumour suppressor genes may be located. LOH was confirmed at the *BRCA1* locus[4] in familial tumours after the kindreds' germline mutations were known, lending further support to the hypothesis that *BRCA1* is a tumour suppressor gene.

The initial search for somatic mutations in *BRCA1* in 12 ovarian and 32 breast sporadic tumours[2] somewhat surprisingly yielded only previously unrecognized germline mutations in 1/12 ovarian and 3/32 breast tumours. In order to investigate the potential role of *BRCA1* in sporadic ovarian carcinogenesis, we have searched for mutations in the transcribed region of *BRCA1* in 47 sporadic ovarian tumours. We now report four somatic mutations in *BRCA1* in one allele and LOH in the

GeneDX 1035, pg. 241

other allele, adding support to a tumour suppressor function for *BRCA1*. Our data suggest that somatic inactivation of *BRCA1* by point mutations in the coding region of one allele in conjunction with LOH may be implicated in about 9% of sporadic ovarian tumours.

Sporadic ovarian tumours

A bank of 47 unselected paraffin-embedded ovarian tumours at first diagnosis was used as the source of samples. All tissue sections were independently reviewed by two pathologists to confirm the diagnosis of invasive carcinoma, evaluate the histologic type (using the classification of the World Health Organization[18]), and identify areas of neoplastic and non-neoplastic predominance. The patients' medical records were reviewed by a physician in order to obtain clinical information. The identity of the patients and their individual histories were not known to the researchers performing the experiments.

Pathologic and clinical data on the patients are listed in Table 1. The tumours used in this study were representative for ovarian cancer cases in the U.S. The average age at diagnosis was 60 years (range: 15–80 years). The most common histologic types of carcinoma were papillary serous (15/47) and endometrioid (15/47), followed by clear cell (7/47), mucinous (3/47), transitional cell (3/47), mixed papillary serous-endometrioid (2/47), carcinosarcoma (1/47), and undifferentiated (1/47). The stage at presentation was known for 42 (90%) of the tumours. Seventeen percent of the tumours were diagnosed at stage I, 4.5% at stage II, 55.5% at stage III, and 13% at stage IV. Four patients had a second malignancy in the breast, but these patients affected with two primaries had no known prior family history of breast or ovarian cancer.

Table 1 Patient and tumour characteristics

Sample	Cancer type	Age at diagnosis	Stage	History of breast cancer
1	Clear cell	51	3	
2	Endometrioid	78	3	
3	Pap serous	87	3	
4	Pap serous	74	3	
5	Mixed common	67	4	
6	Endometrioid	52	3	
7	Mucinous	67	3	
8	Endometrioid	57	4	
9	Clear cell	75	2	
10	Clear cell	68	2	
11	Endometrioid	48	1	
12	Pap serous	52	4	Y
13	Endometrioid	53	3	
14	Clear cell	62	3	
15	Pap serous	34	3	
16	Transitional cell	46	u	
17	Endometrioid	64	u	
18	Pap serous	31	u	
19	Endometrioid	73	u	
20	Clear cell	76	4	
21	Endometrioid	62	3	
22	Pap serous	74	3	
23	Pap serous	69	4	
24	CarcSarc	49	3	
25	Endometrioid	62	3	
26	Pap serous	62	3	
27	Endometrioid	48	u	
28	Clear cell	46	4	
29	Pap serous	64	3	
30	Pap serous	79	3	Y
31	Endometrioid	77	3	
32	Pap serous	80	3	
33	Endometrioid	47	1	Y
34	Endometrioid	55	3	
35	Pap serous	54	1	
36	Transitional cell	58	3	
37	Undifferentiated	75	3	
38	Clear cell	54	1	Y
39	Mucinous	15	1	
40	Endometrioid	49	1	
41	Pap serous	74	3	
42	Pap serous	42	3	
43	Transitional cell	52	1	
44	Mixed common	78	3	
45	Endometrioid	67	3	
46	Mucinous	71	3	
47	Pap serous	44	3	

Sample numbers were used to protect the anonymity of the patients. Histologic tumour types were evaluated by two independent observers. The average age of onset of ovarian cancer in this panel is 60 years. Four individuals had a second malignancy in the breast. u, unknown; Y, yes; Pap, papillary.

DNA extraction of tumour and normal fractions

The virtually complete separation of normal and tumour tissue was required for subsequent single strand conformational polymorphism (SSCP) analysis. Two paraffin tissue blocks, one each representing neoplastic and non-neoplastic ovarian tissue (usually endometrium or uninvolved lymph node) were selected from each case. Non-neoplastic tissue was removed from unstained sections of carcinoma under a dissecting microscope using a serial haematoxylin/eosin-stained (H&E) section as a template guide. This procedure minimized mixing of normal and tumour subpopulations. A typical sample consisted of 1 cm^2 of tissue containing 10^3–10^4 nuclei. Two serial tissue sections each of neoplastic and non-neoplastic tissue were extracted from each specimen[20]. The same DNA templates were used for screening for mutations in the transcribed region of *BRCA1* by SSCP analysis and for LOH experiments.

SSCP of tumour/normal samples

The samples were consecutively arrayed in normal/tumour pairs in a 96-well microtitre plate. SSCP analyses of the transcribed sequence of *BRCA1* was carried out on 47 normal/tumour pairs, a normal DNA sample (positive control), and a sample which contained only the DNA extraction reagents (negative control). Forty-eight overlapping PCR primer pairs were used[14]. The PCR products from each reaction were denatured and electrophoresed in non-denaturing acrylamide gels. All SSCP variants were obtained in duplicate in experiments performed with DNA from two different serial extractions, and sequenced independently. For each variant detected in the normal/tumour panel, we estimated its prevalence in the general population by analysing 80 normal chromosomes.

We detected four different SSCP anomalies which are present only in the tumour DNA of samples 13, 36, 37 and 44, (Fig. 1*a–d*). The corresponding normal (N) and tumour (T) DNA fractions are indicated in each panel. These four variants were not observed in 80 normal chromosomes. The bands which were subsequently sequenced and

GeneDX 1035, pg. 242

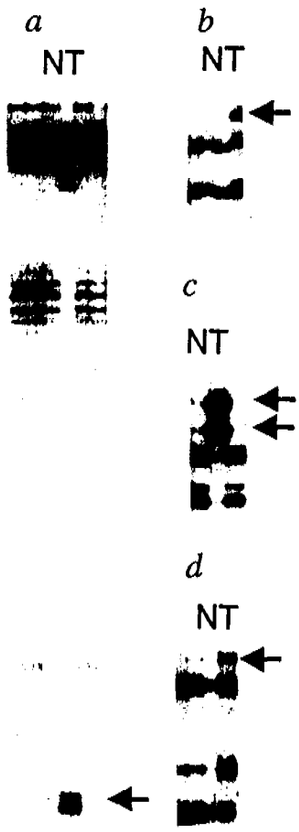


Fig. 1 SSCP analysis of the *BRCA1* gene. SSCP gels of normal/tumour pairs of four samples (13, 36, 37 and 44) demonstrating variants present only in the tumour fractions. *a*, Exon 2 in samples 36N/36T; *b*, exon 5 in 13N/13T; *c*, exon 19 in 44N/44T; *d*, exon 22 in 37N/37T. The band marked with an arrow in (*b*) contains the mutation depicted in Fig. 2.

contained the mutant DNA sequence are indicated by arrows in each panel in Fig. 1.

A germline anomaly in exon 11 of tumour 35 was not detected in 80 normal chromosomes. A polymorphism in exon 11 was detected in four out of 47 tumour/normal pairs and in four out of 80 normal chromosomes. An anomaly in the sequence of Genbank #U15595, which may correspond to a rarely spliced exon 4, was detected in two tumour/normal pairs.

Characterization of somatic *BRCA1* mutations

The DNA samples in the variant and normal bands were sequenced for all samples showing anomalies in the screening SSCP experiments. As an additional control, we sequenced a normal band from another sample which did not have a variant. After cutting the bands from the SSCP gels and re-amplifying the DNA, the products were sequenced with the original SSCP primers. The DNA from the exon 22 variant in sample 37T was cloned into a TA-vector (Stratagene) and sequenced with M13 forward and reverse primers, as the SSCP primers for exon 22 (ref. 1) did not yield readable sequence. As an example, the sequence of exon 5 of the tumour-only variants and normal controls for sample 13T is shown in Fig. 2.

Table 2 summarizes the sequence data on the somatic variants in the *BRCA1* gene found in these tumours. Three mutations are novel and lead to premature truncation of the *BRCA1* (Genbank #U14680) reading frame at codons 16 (sample 36T), 1725 (44T) and 1827 (37T); one mutation in 13T results in a missense alteration of the zinc-finger motif. These four tumours exhibit LOH at one of two intragenic CA-repeat markers, *D17S855* (36T and 37T) or *D17S1323* (13T and 44T) (Fig. 3*a–d*).

As further confirmation of the reduction to homozygosity in these tumours, we digested the 13N and 13T DNA with *Ava*II, taking advantage of the unique *Ava*II site in exon 5 created by the Cys61Gly mutation, which generates fragments of 157 bp and 77 bp in length. Only the tumour DNA showed the predicted restriction pattern; a small amount of normal contamination or undigested DNA was present in the tumour lane (data not shown). The restriction digest and the LOH data confirm that the tumour DNA is hemizygous for the mutated allele.

Other *BRCA1* sequence alterations

We observed two SSCP anomalies in the *BRCA1* coding region and one anomaly in exon 4 in both the germline and tumour DNA of five other samples. In exon 11, we detected a Gln356Arg missense polymorphism in four normal/tumour pairs and in four out of 80 random normals. An A to G alteration at codon 1143 caused a Ser1143Gly missense mutation in samples 35N/35T which was present in only one out of 47 tumours and absent in 80 normal chromosomes; sample 35T had LOH at *D17S1323*.

Two samples, 7 and 45, had a G to T substitution which caused a stop codon to change to a leucine in exon 4; this would avert truncation of the protein at this location if exon 4 were spliced in the *BRCA1* transcript in-frame with exon 3.

Discussion

The notion that *BRCA1* is a tumour suppressor gene derives from previous observations of germline mutations and loss of the wild-type allele in familial breast and ovarian tumours[10,11,17] and of LOH in sporadic tumours[12,14,16]. However, for most tumour suppressor genes involved in familial cancer syndromes, somatic alterations have been found in sporadic tumours of the same kind[20]. Perhaps the most surprising result to emerge from the cloning of *BRCA1* was that Futreal *et al.*[2] did not detect somatic mutations in 32 breast tumours or in an unselected cohort of 12 individuals with ovarian cancer. However, our new data suggest that a larger number of ovarian tumours would have been required to detect an average of at least one somatic mutation.

A more challenging question is posed by the absence of somatic mutations in 32 sporadic breast tumours[2]. It is possible that *BRCA1*-associated tumorigenesis is modulated differently in breast and ovarian tissue, perhaps due to different susceptibility to hormonal or other stimuli.

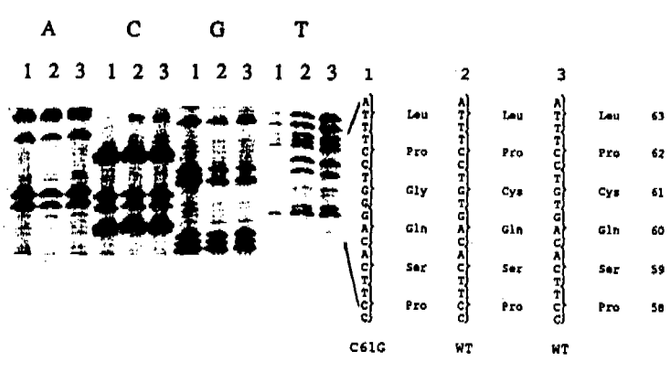


Fig. 2 Autoradiograph of the sequencing gel for exon 5 forward-primed DNA from sample # 13T (lane 1), 13N (lane 2), and another normal control (lane 3). For ease of visualization, the lanes corresponding to the same base were loaded consecutively for the three samples. The amino acid sequence and codon numbers are listed on the left-hand side. The mutant C61G DNA (lane 1) was obtained from the band marked with an arrow in Fig. 1*b*.

GeneDX 1035, pg. 243

Table 2 Somatic mutations and polymorphisms in *BRCA1*

Sample #	LOH	Exon	Nucleotide change	Mutation	Effect on coding sequence	Effect on protein	Present in normal chromosomes
Somatic							
36T	Y	2	GTA→GTCA	151insC	Frameshift	Truncation at codon 16	0/127
13T	Y	5	TGT→GGT	Cys61Gly	Missense	Alter zinc-finger motif	0/127
44T	Y	19	GAA→TAA	Glu1725ter	Nonsense	Truncation at codon 1725	0/127
37T	Y	22	CTT/T→ATT/TT	5502insT	Frameshift	Truncation at codon 1827	0/127
Germline							
35N/35T	Y	11	AGT→GGT	Ser1143Gly	Missense	Altered amino acid sequence	0/80
7N/7T	N	11	CAG→CGG	Gln356Arg	Missense	Altered amino acid sequence	4/80
12N/12T	Y	11	CAG→CGG	Gln356Arg	Missense	Altered amino acid sequence	4/80
18N/18T	N	11	CAG→CGG	Gln356Arg	Missense	Altered amino acid sequence	4/80
45N/45T	Y	11	CAG→CGG	Gln356Arg	Missense	Altered amino acid sequence	4/80

Codon and nucleotide positions are as reported[1]. The mutations are designated by the nomenclature of Beaudet & Tsui[22]. Eighty normal control chromosomes were used to evaluate the frequency of genomic DNA alterations in the general population (data not shown). LOH, Loss of heterozygosity at *D17855* or *D17S1323*; N, normal; T, tumour; Y, yes; N, no.

It would be of great interest to test this hypothesis in cell lines derived from *BRCA1* mutation carriers who choose to undergo prophylactic mastectomy and oopherectomy. *BRCA1* may turn out to play a relatively more important role in sporadic ovarian carcinogenesis than in breast carcinogenesis, perhaps in contrast to *BRCA2* (which appears not to be involved with familial ovarian cancer). In addition, our experience with archival breast tumours[7] indicates that the infiltrating growth pattern of ductal breast carcinomas poses a significant technical challenge to the separation of pure tumour and normal material needed to perform these experiments. This problem is also present, albeit to a lesser degree, in frozen samples and may decrease the probability of detecting somatic mutations in breast tumours. Preliminary results from examination of 33 early-onset sporadic breast tumours yield no mutations (S. Merajver, unpublished data) also in agreement with Futreal *et al.*[2].

The SSCP screening technique is approximately 70% sensitive for single-base changes in a 200 bp DNA fragment[19]. In our study, the specificity of the technique was 100% since a nucleotide change was found upon sequencing each reproducible SSCP variant band. Somatic alterations in *BRCA1* which are likely to be implicated in carcinogenesis were detected in 9% of sporadic ovarian tumours. This number represents a lower limit given the sensitivity of the screening technique and the fact that introns and portions of the 5'- and 3'-untranslated regions were not screened. No correlation was found with any particular histologic type or stage at presentation.

In three out of four cases, the somatic *BRCA1* mutations were novel and led to a shortened protein product. A nonsense mutation resulted in a truncated protein at codon 1725. The frameshift mutation, 151insC, is predicted to result in loss of function of the protein product; the other frameshift mutation caused termination of the new reading frame at codon 1827. Friedman *et al.*[6] have reported that the mutation Tyr1853ter co-segregated with breast cancer in a kindred with seven affected individuals and average age at diagnosis of 32.9 years. If Tyr1853ter (which results in loss of only the last 11 amino acids in the C terminus) is indeed the disease-predisposing mutation in this kindred, and not a rare polymorphism, it is likely that the somatic mutation 5502insT in tumour 37T would also lead to an inactive protein product. Although it is possible in principle that these seemingly 'subtle' alterations in BRCA1 may lead to cancer by a dominant negative effect, the observations of LOH at *BRCA1* in 37T and in the three other tumours with somatic mutations make this hypothesis much less likely. Our data suggest that wild-type BRCA1 was not being expressed in these tumours at the stage at which they were sampled from the patients.

The remaining somatic mutation, Cys61Gly, leads to alteration of the sixth cysteine of the putative zinc-finger domain of BRCA1. This mutation has been reported previously[6,7] to co-segregate with breast cancer in two apparently unrelated kindreds. The LOH and restriction digest experiments in sample 13T confirm that the wild-type allele was lost in this tumour. It is likely, then, that this mutation constitutes the inactivating event for one allele in sample 13T.

One germline missense alteration, Ser1143Gly may represent either a new disease-associated mutation or a rare polymorphism. In the absence of functional analysis of the mutant protein, we cannot resolve this

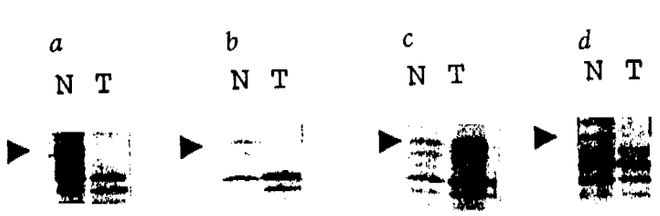


Fig. 3 Loss of heterozygosity in samples with somatic *BRCA1* mutations. All samples with somatic mutations were tested for the two intragenic CA-repeat markers, *D17S855* and *D17S1323*. LOH was present at one of these markers in all samples; all samples were uninformative (data not shown) for the other marker. *a*, 36N/36T with LOH at *D17S855*; *b*, 13N/13T with LOH at *D17S1323*; *c*, 44N/44T with LOH at *D17S1323*; *d*, 37N/37T with LOH at *D17S855*.

GeneDX 1035, pg. 244

question. The number of germline alterations reported here is in fact consistent with the findings of Futreal *et al.*[2]. It is interesting to note, however, that a germline alteration likely to be associated with cancer was found in none of four patients with known breast and ovarian cancer (Table 1). These patients had no evidence of autosomal dominant transmission of a cancer susceptibility allele. They may represent new germline mutations which were not detected by our screening technique. It is more likely, however, that they have coincidental aggregation of two hormone-dependent primary tumours which may be modulated by common risk factors or exposures.

Note added in proof: Another *BRCA1* mutation in an ovarian tumour is reported by Hosking *et al.* in this month's correspondence section.

Methodology

Specimen selection. Forty-seven paraffin-embedded ovarian tumours from the University of Michigan Pathology collection were used in accordance with guidelines from the University of Michigan Human Use Committee for the procurement of archival materials for research. The results of this research were not shared with the patients or their descendants.

DNA extraction. Unstained normal and neoplastic tissue sections were removed with a disposable razor blade from each glass slide and deparaffinized with 100 µl of xylene as described[21]. An equal volume of 100% ethanol was added and the samples pelleted for 10 min at 15,000g, vacuum-dried and digested overnight with 200 ng µl^{-1} Proteinase K in 100 µl of 50 mM tris, pH 8.3. Samples were boiled for 8 min, placed on ice, and spun 5 min to pellet out debris. Samples were extracted in batches of 8–12, each extraction containing a tube without template as a negative control. This method avoids the use of ionic detergents, sonication, or phenol-chloroform separation, all of which yield samples which may be suboptimal for PCR experiments[21].

Each tumour/normal extraction yielded two 200 µl sample with sufficient DNA for 40 reactions. Sequential extractions for a given block were carried out on sections immediately adjacent to the original slide. This procedure resulted in the re-extraction of tumour/normal DNA from regions equivalent to those of the original extraction.

PCR amplification of polymorphic markers. PCR reactions were performed on each tumour/normal pair, a normal human DNA control, and two or more samples from CEPH parents known to be informative for the marker being analysed. DNA extraction reagents and water served as negative control template. The reaction volume was 25 µl. The reaction mixture consisted of 5 µl of DNA template, 2.5 µl 10× *Taq* polymerase buffer, 0.25 µl of 10 mM dNTP mixture, 0.2 µl of *Taq* polymerase, a variable amount of forward and reverse primer ranging from 0.8–1.5 µl of a 10 ng µl^{-1} solution of each, and the remainder in double-distilled water. One of the primers was end-labelled with γ-^{32}P dATP using T_4 kinase. Annealing temperatures were 58 °C for *D17855* and 55 °C for *D17S1323*. PCR conditions were 94 °C × 1 min, annealing temperature × 1 min, 72 °C × 1 min for 35 cycles, and a final extension at 72 °C for 5 min. The PCR products were analysed in 0.4 mm thick 8% denaturing polyacrylamide gels, which were dried and autoradiographed by standard techniques.

SSCP of genomic DNA. Forty-eight *BRCA1* primer pairs amplifying all exons were used. PCR was carried out as outlined above in 10 µl reactions. Direct incorporation of α-^{32}P dCTP was used instead of end-labelling of primers. The PCR products were denatured and analysed on Hydrolink MDE gels (AT Biochem) prepared according to manufacturer's instructions, and run at 6 W for 16 h at room temperature. The gels were dried and autoradiographed by standard methods.

Direct DNA sequencing. After alignment of the autoradiographs with orientation markers to the gels, the variant SSCP bands were scraped directly into 50 ml PCR reactions as described[4]. The DNA was amplified as above with 1 ml of 10 mM forward and reverse primer for 30 cycles. After verification of yield and purity on a 2% agarose gel, the DNA was purified on a G-50 spin column (5 prime-3 prime). Precipitation with ammonium acetate and glycogen and resuspension in 10 ml TE, pH 8.0, 2 ml were used for cycle-sequencing (Stratagene *fmol*), as described[4]. The gels were run at 1,800 V for 3–4 h, dried and exposed at room temperature for 16 h.

Restriction digest of DNA. Following dissection of the variant and normal bands from the dried SSCP gel and amplification with the original primers, 1 µg of DNA was digested with 5 U of *Ava*II and 5 ml of 10× Buffer in a 50 ml reaction for 3 h. The products were analysed on a 4% agarose gel by ethidium bromide staining.

Acknowledgements

We thank our patients for donating their tissues for research and acknowledge helpful discussions with D. Ginsburg, F.J. Couch and M. Erdos. This work was supported by GCRC-MCAP 3M01RR00042-34S1 to S.D.M. from the NIH, by the Harris Foundation (T.S.F), and by the Union Bank of Switzerland (R.F.C.).

Received 3 February; accepted 1 March 1995.

1. Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* **266**, 66–71 (1994).
2. Futreal, P.A. *et al. BRCA1* mutations on primary breast and ovarian carcinomas. *Science* **266**, 120–122 (1994).
3. Hall, J.M. et al. Linkage of early onset breast cancer to chromosome 17q21. *Science* **250**, 1684–1689 (1990).
4. Castilla, L.H. *et al.* Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. *Nature Genet.* **8**, 387–391 (1994).
5. Simard, J. *et al.* Common origins of *BRCA1* mutations in Canadian breast and ovarian cancer families. *Nature Genet.* **8**, 392–398 (1994).
6. Friedman, L.S. *et al.* Confirmation of *BRCA1* by analysis of germline mutations linked to breast and ovarian cancer in 10 families. *Nature Genet.* **8**, 399–404 (1994).
7. Shattuck-Eidens, D. *et al.* A collaborative survey of 80 mutations in the BRCA1 breast and ovarian cancer susceptibility gene. *J. Am. med. Assoc.* **273**, 535–541 (1995).
8. Easton, D.F., Bishop, D.T., Ford, D. & Crockford, G.P. The Breast Cancer Linkage Consortium. Genetic linkage analysis in familial breast and ovarian cancer. Results from 214 families. *Am. J. hum. Genet.* **52**, 678–701 (1993).
9. Ford, D. *et al.* Risks of cancer in *BRCA1* mutation carriers. *Lancet* **343**, 692–695 (1994).
10. Kelsell, D.P., Black, D.M., Bishop, D.T. & Spurr, N.K. Genetic analysis of the *BRCA1* region in a large breast/ovarian family: Refinement of the minimal region containing *BRCA1*. *Hum. molec. Genet.* **2**, 1823–1828 (1993).
11. Smith, S.A., Easton, D.F., Evans, D.G.R. & Ponder, B.A.J. Allele losses in the region 17q12–21 in familial breast and ovarian cancer involve the wild-type chromosome. *Nature Genet.* **2**, 128–131 (1992).
12. Sato, T. *et al.* Allelotype of human ovarian cancer. *Cancer Res.* **51**, 5118–5122 (1991).
13. Lee, J.H., Kavanagh, J.J., Wildrick, D.M., Wharton, J.T. & Blick, M. Frequent loss of heterozygosity on chromosomes 6q, 11, and 17 in human ovarian carcinomas. *Cancer Res.* **50**, 2724–2728 (1990).
14. Eccles, D.M., Cranston, G., Steel, C.M., Nakamura, Y. & Leonard, R.C.F. Allele losses on chromosome 17 in human epithelial ovarian carcinomas. *Oncogene* **5**, 1599–1601 (1990).
15. Jacobs, I.J. *et al.* A deletion unit on chromosome 17q in epithelial ovarian tumors distal to the familial breast/ovarian cancer locus. *Cancer Res.* **53**, 1218–1221 (1993).
16. Cropp, C.S. *et al.* Evidence for involvement of *BRCA1* in sporadic breast carcinomas. *Cancer Res.* **55**, 2548–2551 (1994).
17. Merajver, S.D. *et al.* Germline *BRCA1* mutations and loss of the wild-type allele in early-onset breast/ovarian families. *Clin. Cancer Res.* (in the press).
18. Young, R.H. & Scully, R.E. Pathology of epithelial tumors. *Hem./Onc. Clin. N. Am.* **6**, 739–760 (1992).
19. Orita, M., Suzuki, Y., Sekiya, T.S. & Hayashi, K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874–879 (1989).
20. Knudson, A.G. Antioncogenes and human cancer. *Proc. natn. Acad. Sci. U.S.A.* **90**, 10914–10921 (1993).
21. Frank, T.S. *et al.* Loss of heterozygosity and overexpression of the p53 gene in ovarian carcinoma. *Mod. Pathol.* **7**, 3–8 (1994).
22. Beaudet, A.L. & Tsui, L.-C. A suggested nomenclature for designating mutations. *Hum. Mut.* **2**, 245–248 (1993).

GeneDX 1035, pg. 245

Mutations in the *BRCA1* gene in families with early-onset breast and ovarian cancer

Lucio H. Castilla[1,2*], Fergus J. Couch[3*], Michael R. Erdos[1*], Kent F. Hoskins[3], Kathy Calzone[3], Judy E. Garber[4], Jeff Boyd[5], Matthew B. Lubin[6], Michelle L. Deshano[3], Lawrence C. Brody[1], Francis S. Collins[1] & Barbara L. Weber[3,7]

[1]*National Center for Human Genome Research, Building 49/3A14, National Institutes of Health, Bethesda, Maryland 20892, USA*
[2]*Department of Biology, University of Michigan, Ann Arbor, Michigan 48109, USA*
[3]*Department of Internal Medicine, 1010 BRBI, 422 Curie Boulevard, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*
[4]*Division of Cancer Epidemiology and Control, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA*
[5]*Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*
[6]*Strang Cancer Prevention Center, 428 East 72nd Street, New York, New York 10021, USA*
[7]*Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*

Correspondence should be addressed to B.L.W.

We analysed 50 probands with a family history of breast and/or ovarian cancer for germline mutations in the coding region of the *BRCA1* candidate gene, using single-strand conformation polymorphism (SSCP) analysis on PCR-amplified genomic DNA. A total of eight putative disease-causing alterations were identified: four of these are frameshifts and two are nonsense mutations. In addition, we found two missense mutations, one of which changes the final cysteine of the BRCA1 zinc finger motif to glycine. These data are consistent with a tumour suppressor model, and support the notion that this candidate gene is in fact *BRCA1*. The heterogeneity of mutations, coupled with the large size of the gene, indicates that clinical application of *BRCA1* mutation testing will be technically challenging.

Breast cancer is an extremely common malignancy, affecting one in eight women during their lifetime, with more than 170,000 women being diagnosed in the United States each year. A positive family history has been identified as a major contributor to risk of development of the disease, and this link is particularly striking for early-onset breast cancer. Segregation analysis of families with multiple affected individuals suggested the presence of one or more autosomal dominant susceptibility genes, which together appeared to account for 5–10% of all breast cancers[1,2]. Confirmation of this prediction was achieved by the mapping of a susceptibility gene (*BRCA1*) responsible for familial early-onset breast and ovarian cancer to chromosome 17q12–21 (refs 3,4). Mutations in *BRCA1* are thought to be responsible for 45% of inherited breast cancer and more than 80% of inherited breast and ovarian cancer[5]. More recently a second susceptibility gene, *BRCA2*, was localized to chromosome 13q12–13 (ref. 6). Mutations in *BRCA2* also predispose males to breast cancer, while *BRCA1* seems to carry a higher risk of ovarian cancer than does *BRCA2*. Together, these genes may account for as much as 90% of familial breast cancer, but the existence of additional susceptibility genes is implied by the identification of a few families with a high incidence of breast cancer that fail to show linkage to either *BRCA1* or *BRCA2* (ref. 6).

Although the frequency of *BRCA1* germline mutations in the general population has not yet been determined with accuracy, as many as one in 200 women may harbour *BRCA1* germline mutations. These women are estimated to be at a 85–90% lifetime risk for breast cancer, and a markedly elevated risk of ovarian cancer[5]. A recent collaborative study has indicated that males carrying a mutated *BRCA1* gene are at a three-fold increased risk of prostate cancer, and males and females are at a four-fold increased risk of colon cancer[7].

A novel gene has recently been reported as a "strong candidate" for the 17q-linked *BRCA1* gene[4,9]. The initial screen for germline mutations uncovered five sequence alterations in eight kindreds analysed[4]. Two were frameshift alterations, one was a nonsense mutation (Gln1313ter), and one was identified as a loss of mRNA from the linked allele, probably caused by a regulatory mutation or a mutation that affected the stability of the mRNA molecule. A fifth alteration was a missense mutation. An accompanying paper focusing on primary breast and ovarian carcinomas[9] detected four additional germline alterations in this candidate gene; three of these were missense alterations (one, Met1775Arg, had been found independently in the family study), and one was a nonsense mutation (Asp1541ter).

Over the past four years, we have collected DNA on a number of individuals and families with early-onset breast or breast-ovarian cancer. Here we report the detection of eight different germline mutations in the coding region of the candidate gene for *BRCA1* in these high-risk individuals. The majority of these mutations would be expected to result in loss of function.

Breast/ovarian cancer families studied

Each of the 50 families included in this study has three or more early-onset breast cancer individuals or at least two breast and one ovarian cancer individual. In the initial screen, we included DNA from at least one proband from each family. Ten of these have lod scores (likelihood ratios for linkage) of 1.0 or above for closely linked markers[10]

GeneDX 1035, pg. 246

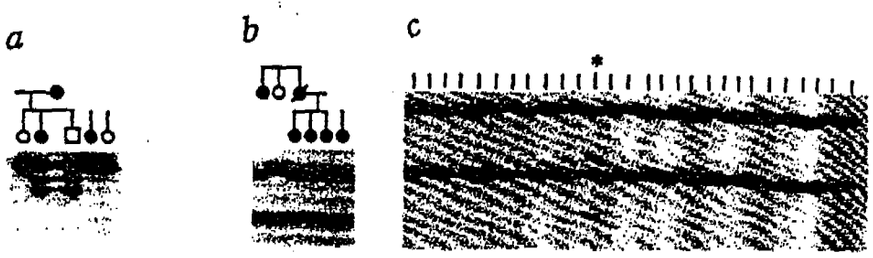


Fig. 1 SSCP analysis of the *BRCA1* gene. Representative SSCP gels of *BRCA1* exons amplified from patient genomic DNA are shown. Filled pedigree symbols indicate a diagnosis of breast cancer. *a*, Amplification of exon 11 (11.15 primer pair) from members of kindred 15. The first individual from the left does not carry the affected haplotype. The unaffected male has been previously shown[10] to carry the affected haplotype. Direct sequencing of the variant bands revealed a 4 bp deletion (Fig. 2a, Table 1). The last two individuals, from kindred 27, represent normal controls for this primer set. The *BRCA1* mutation in this family is in exon 5 (*b*). *b*, Segregation of an exon 5 SSCP variant in kindred 27. DNA from the deceased mother was not available. The last lane contains amplification products from an affected daughter of the deceased mother's first cousin. Variant bands are observed in affected individuals and correlate with segregation of the *BRCA1* haplotype in this family[10]. Sequence of the variant bands revealed a T to G transversion in codon 64. *c*, Scanning of exon 16 by SSCP. Amplification products from 30 individuals are shown. The uppermost band demonstrates a bi-allelic polymorphism. Heterozygotes and homozygotes for both alleles are observed. Amplification products from an individual marked with an asterisk (*) demonstrate two additional bands. The sequence of this band revealed a 19 bp deletion (5085del19).

(F.J.C. *et al.*, manuscript submitted), suggesting linkage to the *BRCA1* locus. For 40 of the families, insufficient samples were available to carry out linkage.

SSCP analysis of high-risk families

We screened for mutations in the entire coding region of the *BRCA1* candidate gene in genomic DNA using the single strand conformation polymorphism (SSCP) technique[11]. To amplify most of the 23 exons, we utilized published primer sequences[8]. Since exon 11 is too large (3,426 basepairs (bp)) to analyse effectively in one piece by SSCP, we designed 19 SSCP primer sets for this exon, each with a length of 200–400 bp (see Methodology). We also amplified exons six and seven with newly designed primers. In all cases, splice junctions were included in the analysis. Each exon was PCR-amplified from genomic DNA and then electrophoresed in a non-denaturing gel, with and without the addition of 10% glycerol. Representative examples of SSCP variants are shown in Fig. 1.

BRCA1 mutation characterization

We initially detected a total of 31 SSCP variants. We further characterized SSCP variants not seen in normal chromosomes. Variant SSCP bands were cut from the gel, re-amplified by PCR, and directly sequenced from both ends using the same primers. In only one instance was sequence alteration not found, probably due to contamination by DNA from the wild-type allele in the sequencing process. Mutations were designated as recommended by Beaudet & Tsui[12]. For the eight mutations described here, clinical information is summarized in Table 1.

Kindred 15 has a 4 bp deletion beginning at nucleotide 3875 (codon 1252) of the reported cDNA sequence (GenBank #U14680), resulting in a stop codon 29 bp downstream from the deletion site. Kindred 27 has a missense mutation at nucleotide position 309. This substitution of a T to G in codon 64 changes a Cys to a Gly. Analysis of tumour DNA in two affected members of this kindred showed that the wild-type allele was lost, and only the Cys64Gly mutant allele remained (data not shown). Kindred 32 has a 40 bp deletion beginning at nucleotide 1294 (codon 392). The subsequent frameshift results in a stop codon 12 bp downstream of the deletion end point, truncating the open reading frame. Kindred S758 has a 19 bp deletion beginning at nucleotide 5085 (codon 1656), generating a frameshift which results in a stop codon 43 bp downstream of the deletion site. Kindred 161 has a nonsense mutation, changing a G to T at nucleotide position 3867. This substitution changes Glu to a stop at codon 1250 (Glu1250ter). Kindred 202 has an insertion of a C after nucleotide position 5438 (codon 1773) generating

Table 1 *BRCA1* mutations in at high-risk kindreds

Codon	Nucleotide change	Coding effect	Mutation	Kindred	Number of affected individuals (Breast cancer/ ovarian cancer)	Average age at diagnosis
64	TGT → GGT	Missense	Cys64Gly	27[a]	12 (8/5)	38
392	Deletes nt 1294–1333	Frameshift	1294del40	32	5 (4/1)	37
1250	GAG → TAG	Nonsense	Glu1250ter	161	4 (3/2)	36
1252	Deletes nt 3875–3878	Frameshift	3875del4	15[a]	9 (9/2)	41
1443	CGA → TGA	Nonsense	Arg1443ter	S753	3 (3/1)	36
1443	CGA → GGA	Missense	Arg1443Gly	51	3 (3/0)	50
1656	Deletes nt 5085–5103	Frameshift	5085del19	S758	4 (4/1)	52
1773	ACC → ACCC	Frameshift	5438insC	202[b]	18 (14/6)	45

Codon and nucleotide positions are as reported by Miki *et al.*[9]. The number of affected individuals includes cases of breast and ovarian cancer. The total number of breast and ovarian cases is greater than the number of affected individuals in kindreds where some individuals have been affected with both cancers. Average age of diagnosis of breast cancer is given. [a,b]Kindreds previously known to have breast disease linked to *BRCA1* with a lod score >1.7[a], or 0.7[b] (F.J.C. *et al.*, manuscript submitted)[10].

GeneDX 1035, pg. 247

Table 2 *BRCA1* primers used for amplification and SSCP analysis of exons 8, 7 and 11

Exon	Sense	Antisense	Product length (bp)
8	Intron 5–ATTTGCCCACATGTCTTTTC	Intron 6–AGTAATGTGCAAACTTCCTGAGT	243
7	Intron 6–ACATAATGTTTTCCCTTGTATTTT	Intron 7–CCACGAGCAAAGACTCC	262
11.0	Intron 10–GTGAATTTTCTGAGACGGATGTA	964–GAGCTGGCATGAGTATTTGTG	169
11.1	883–AGAGGCATCCAGAAAAGTATCAGG	1121–GGGAGTCCGCCTATCATTACAT	239
11.2	1046–ACAGCCTGGCTTAGCAAGGAG	1323–CCCCATCATGTGAGTCATCAGA	278
11.3	1186–AGAAACTGCCATGCTCAGAGAATC	1430–ATGAGGATCACTGGCCAGTAAGTC	245
11.4	1352–TGTATTGGACGTTCTAAATGAGGT	1617–TTGTGAGGGGACGCTCTTGTA	266
11.5	1572–GCATTTGTTACTGAGCCACAGATA	1834–TCTATTGGGTTAGGATTTTTCTCA	263
11.6	1731–CAAACGGAGCAGAATGGTCA	1974–GCCTGGTAGAAGACTTCCTCCTC	244
11.7	1918–TCCACAATTCAAAAGCACCTAAAA	2216–CTCTGGGAAAGTATCGCTGTCAT	299
11.8	2142–GCAACTGGAGCCAAGAAGAGTAAC	2397–TTTGCAAAACCCTTTCTCCACTTA	256
11.9	2281–TTGTCAATCCTAGCCTTCCAAGAG	2504–TTTTGCCTTCCCTAGAGTGCTAAC	224
11.10	2448–TATGGCACTCAGGAAAGTATCTCG	2717–GCGCTTTGAAACCTTGAATGTAT	270
11.11	2635–ACAGTCGGGAAACAAGCATAGAA	2948–TTTGGCATTATCAACTGGCTTATC	314
11.12	2900–AGGCTTTCCTGTGGTTGGT	3205–TTACGGCTAATTGTGCTCACTG	306
11.13	3171–AACATTCCAAGTACAGTGAGCACA	3548–AGATGCATGACTACTTCCCATAGG	378
11.14	3413–TCCTGGAAGTAATTGTAAGCATCC	3737–GGCCCCTCTTCGGTAACC	325
11.15	3686–TCCTAGCCCTTTCACCCATACA	3959–AGATGCCTTTGCCAATATTACCTG	274
11.16	3860–TGCTACCGAGTGTCTGTCTAAGAA	4068–AGAAAGGATCCTGGGTGTTTGTAT	209
11.17	3996–GCTAGCTTGTTTTCTTCACAGTGC	4213–AAGTTTGAATCCATGCTTTGCTCT	218
11.18	4110–CAGGGAGTTGGTCTGAGTGAC	Intron 11–GCTCCCCAAAAGCATAAAC	181

Numbers representing the 5' nucleotide are given at the start of each oligonucleotide sequence.

function mutation places individuals at high risk for breast and ovarian cancer, and loss of the remaining wild type allele provides a necessary second step to initiate the path to tumorigenesis. The surprising feature of *BRCA1*, however, is that so far somatic mutations in sporadic tumours have not been reported[9].

The role of the two missense mutations in causing disease is not as well defined. In the absence of a functional assay, it is impossible formally to rule out a rare "private" linked polymorphism. One of these (Cys64Gly), for example, was found in a large African-American kindred. Even though only a limited number of African-American chromosomes were included in the normal controls, this mutation is likely to be significant, as it results in a mutation of the final cysteine of the predicted $C3HC_4$ zinc-binding RING finger located near the N terminus of the 1,863 amino acid protein[1]. The presence of a RING finger motif suggests that this protein is likely to function as a transcription factor[14]; if so, substitution of this critical amino acid would be expected to alter DNA binding severely. Earlier reports[4,9] identified two examples of Met1775Arg in African-Americans, suggesting this might be a common mutation in this population. Kindred 27 was the only African-American family in our panel.

The goal of this study was to survey a high risk population to identify *BRCA1* mutations. It is perhaps surprising that we have detected mutations in only eight of 50 probands tested. Several circumstances may have contributed to this outcome: (i) We have only studied exons from genomic DNA. Mutations in the promoter, deeper within the introns, or associated with complete or partial deletion of an exon, resulting in loss of one or both PCR primer binding sites, would have been missed. Analysis of cDNA, particularly in concert with identification of transcribed polymorphisms to identify alleles where no mRNA is produced[8] would likely uncover other alterations. (ii) SSCP may have missed some mutations, especially single nucleotide changes, though the efficiency of SSCP with the modifications applied here is expected to be about 90% (ref. 17). (iii) It is quite likely that many of the probands come from pedigrees unlinked to *BRCA1*. Lod scores of 1.0 or greater were obtained in only ten of these kindreds (too few samples being available in the others to obtain significant linkage results), and in most of these ten families, the lod scores were in the range of 1.0–1.5 (ref. 10; F.J.C. *et al.*, manuscript submitted).

The average age of onset for families with (mut^+) and without (mut^-) identified mutations is 41.8 and 45.3 years respectively. There is one family without ovarian cancer in the mut^+ families (12%) as compared with 7/42 in the mut^- group (17%). Small sample size precludes definitive analysis, but these data suggest a trend toward more severely affected families in the mut^+ grouup. However, the mut^- families constitute a group which fit previous descriptions of families likely to have *BRCA1* mutations. Thus, we would not be surprised if more families were found to harbour *BRCA1* mutations that have been missed using the approach we describe.

We found no significant tendency of mutations to cluster in particular exons (though the presence of two different mutations in codon 1443 is of note), and no mutations were found more than once. Both of the nonsense mutations occur in CpG dinucleotides, with mutation of C to T, presumably as a result of deamination of a methylated cytosine[18]. One of the frameshift mutations (3875del4) results in a loss of one copy of a 4 bp repeat (5'–GTCTGTCT–3') in the coding region of exon 11, possibly as a result of slippage during DNA replication.

The consequences of this considerable allelic heterogeneity for molecular diagnosis are obvious: unlike some disorders with an essentially monomorphic mutational basis (such as Huntington disease[19] or achondroplasia[20]), DNA-based detection of *BRCA1* mutations will be technically demanding, with the need to survey a large gene for a wide variety of potential mutations. False negatives are likely to be common unless very robust

GeneDX 1035, pg. 248

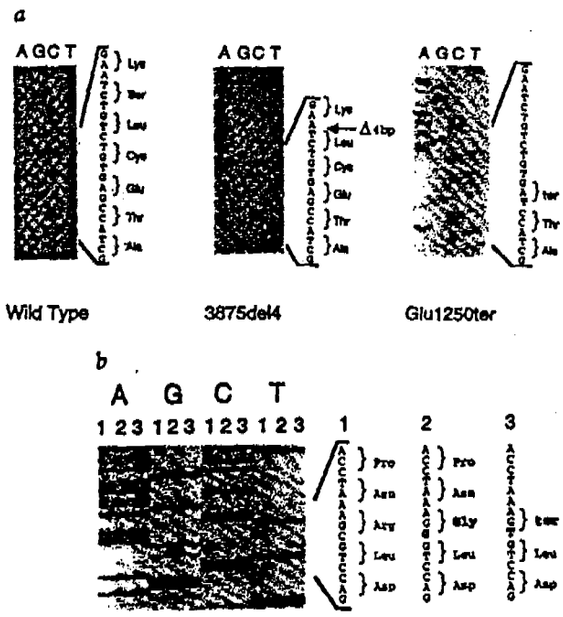


Fig. 2 Sequence analysis of variant SSCP bands. Bands were excised from the gel, re-amplified, and directly sequenced. *a*, Sequence of exon 11 (beginning at codon 1248) with the mutations found in kindreds 15 (3875del14) and 161 (Glu1250ter). The consequences for protein coding are indicated to the right of each sequence. *b*, Sequence results from exon 13 (beginning at codon 1441). To ease mutation identification, the sequencing products were loaded with all the templates for a given nucleotide reaction in adjacent lanes. Sample 1 is wild type, sample 2 is from kindred 51 and sample 3 is from kindred S753. The Arg1443ter and Arg1443Gly mutations are shown.

a shift in the reading frame, and a stop codon in the new reading frame is found at nucleotide position 5603. Kindred 51 has a missense mutation that substitutes a C to G at nucleotide 4446, changing an Arg to a Gly at codon 1443 (Arg1443Gly). Kindred S753 has a nonsense mutation (C to T substitution) at the same nucleotide position (Arg1443ter).

These alterations were not found by SSCP analysis of at least 50 normal chromosomes, making it unlikely that they represent common polymorphisms (data not shown). Representative examples of mutations identified by sequence analysis are shown in Fig. 2. When possible, we also confirmed that the mutations segregated with breast cancer within the family by SSCP analysis. For kindreds 15, 27 and 202, the mutation was shown to segregate with the disease. In kindreds 32, S753, S758 and 161, we did not have access to additional DNA samples from the family, but the fact that these are frameshift and nonsense mutations makes it likely that they are disease-related. Kindred 51, which contains a missense mutation, is the most uncertain, as we cannot formally rule out the possibility that this represents a rare polymorphism.

Additional *BRCA1* sequence changes

In addition to the alterations described above, we identified three other variants that may represent rare polymorphisms. In kindred 178, a synonymous substitution at nucleotide position 710 (Cys197) substitutes a T for a C in the third position of the codon. Since this alteration is close to the 3' end of exon nine, resulting in a sequence change 5'–ATTGCAGgtgagt–3' to 5'–ATTGTAGgtgagt–3', an alternative splice site could potentially be created which would result in a 4 bp deletion in the mRNA molecule. Further experiments will be required to test this hypothesis by sequencing of reverse transcribed cDNA from affected individuals of this kindred.

A missense alteration in kindred 32 substitutes a C for T (Val772Ala). This is likely to be an inconsequential polymorphism in view of the simultaneous frameshift mutation (1294del40) found in this individual. Kindred 62 has a missense alteration at nucleotide position 3238. This substitution of A for G (Ser1040Asn) is likely to be a rare polymorphism since it does not segregate with the disease within this family, and was also found in three of 232 normal chromosomes.

Discussion

A summary of the eight mutations detected in this study, compared to those reported previously[4,9], is shown in Table 1 and Fig. 3. Six are frameshift or nonsense mutations that are predicted to result in a loss of function of the protein product. While a dominant-negative effect of a truncated mutant protein cannot be excluded, the observation of frameshifts early in the coding region (189del11 (ref. 8); 1294del40, this study) and the previous description of a mutant allele in which mRNA expression was completely absent[9] favour the unifying hypothesis that familial early-onset breast and ovarian cancer arises as a consequence of germline loss of function (by whatever means) of this gene.

Previous observations have demonstrated loss of the wild type *BRCA1* allele in tumours arising in pedigrees shown to be linked to 17q21 (refs 13,14) (Merajver, S., personal communication), and we have confirmed this in kindred 27 by demonstrating that only the mutant Cys64Gly allele is found in breast tumours from this family. This gene thus fits nicely with Knudson's two-hit hypothesis for hereditary cancer[15]: a germline loss of

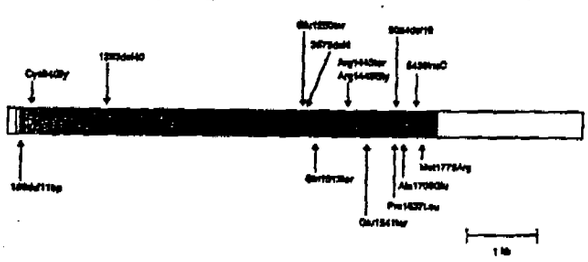


Fig. 3 Schematic diagram of the *BRCA1* cDNA with mutations. Translated regions of the cDNA are indicated by shading and the 23 exons as boxes according to Miki *et al.*[8]. Mutations described in this paper are placed above the cDNA diagram, those reported by Miki *et al.*[8] and Futreal *et al.*[9] are placed below. Mutations have been designated according to the nomenclature recommended by Beaudet & Tsui[12]. An amino acid followed by a codon number designates missense and nonsense mutations. A number without a prefix designates the nucleotide position of insertions (ins) or deletions(del). The number of deleted or inserted nucleotides follows that number.

GeneDX 1035, pg. 249

methods are developed. False positives (missense mutations which are later revealed to be rare benign polymorphisms) will be a potentially serious problem until functional assays for BRCA1 can be established. Furthermore, the considerable challenges of educating and counseling high risk individuals[21,22] about the outcome of molecular testing, especially in the absence of detailed information about genotype-phenotype correlations and the effectiveness of potential interventions, have led several authoritative groups (National Advisory Council for Human Genome Research[23], American Society of Human Genetics[24] and National Breast Cancer Coalition (Visco, F., personal communication)) to recommend that *BRCA1* mutation testing currently be carried out only as part of an approved research protocol, with full informed consent. Our results underscore the wisdom of those statements.

Despite these caveats, the discovery of a major gene for breast and ovarian cancer susceptibility is a major step forward. While the original report[4] was careful to describe this as a "strong candidate", and while the ultimate proof of the question may need to await functional studies, the additional set of mutations that we describe leaves little doubt that this is the long-sought *BRCA1* gene.

Methodology

Patient DNA samples. DNA was isolated from peripheral blood lymphocytes or Epstein-Barr virus transformed lymphoblastoid cell lines using standard methods[25] or with the Purgene DNA isolation kit as directed by the manufacturer (Gentra Systems, Inc.).

SSCP methods. We designed 20 new pairs of PCR primers for *BRCA1* exons 6 and 11 (Table 2). PCR was carried out using standard conditions containing 1 µl (20 ng) of genomic DNA template, 1 µl of 10× PCR buffer (Promega), 0.1 µl of 10 mM dNTP mix (10 mM dATP, 10 mM dGTP, 10 mM dCTP, 10 mM dTTP), 0.2 µl of 10 µM forward primer, 0.2 µl of 10 µM reverse primer, 0.1 µl of Amplitaq DNA polymerase (5 U µl^{-1}), 0.1 µl α^{32}P dCTP and 7.3 µl of ddH_2O. For amplification, each sample was denatured at 94 °C for 1 min and subjected to 2 sets of 20 amplification cycles each. Each cycle of the first set consisted of 30 s denaturing at 92 °C and 40 s annealing and extending at 70 °C (–0.5 °C/cycle). Each cycle of the second set consisted of 30 s at 92 °C, 40 s 60 °C (+1 s/cycle). 50 µl of denaturing stop solution (95% formamide, 10 mM NaOH, 0.25% bromophenol blue) was added to the 10 µl reaction, denatured 5 min at 95 °C, and rapidly cooled on ice. Gels were prepared with 16 ml of Hydrolink MDE gel (AT Biochem), 44.2 ml of ddH_2O, 3.84 ml of 10× TBE buffer (0.9 M Tris, 0.9 M boric acid, 0.02M EDTA, adjusted to pH 8.0), 256 µl of 10% ammonium persulfate and 25.6 µl of TEMED. Duplicate gels were prepared with a supplement of 10% glycerol. 2 µl of each reaction was loaded, and gels were run in 0.6× TBE buffer at 4–8 W for 12–20 h at room temperature. Gels were then dried under vacuum at 80 °C and exposed to autoradiography film at –80 °C or room temperature for 4–20 h.

Direct DNA sequencing. Variant SSCP bands were cut from the gels after alignment with the autoradiograph, and placed directly in the PCR reactions. The reaction contained 41.5 µl ddH_2O, 5 µl 10× PCR buffer, 1 µl 10 mM forward primer, 1 µl 10 mM reverse primer, 1 µl 10 mM dNTP mix and 0.5 µl Amplitaq DNA polymerase. For the re-amplification, the bands were denatured for 1 min at 95 °C and subjected to a 30-cycle amplification. Each cycle was 30 s denaturing at 94 °C, 30 s annealing at 60 °C and 1 min extension at 72 °C. A final extension step for 5 min at 72 °C was added. One tenth of the reaction was run in a 1.5% agarose gel to visualize yield. The amplified DNAs were then column purified with PCR-SELECT II columns (5 PRIME-3 PRIME Inc.), precipitated, and resuspended in 10 µl of ddH_2O. The sequencing reaction mix contained 2 µl DNA, 8 µl ddH_2O, 5 µl 10× sequencing buffer (*fmol*, Promega), 0.5 µl 10 µM primer, 0.5 µl ^{35}S dATP and 1 µl sequencing grade *Taq* polymerase (*fmol*, Promega). Aliquots of 4 µl were placed in each of four tubes, and then 2µ ul of either ddATP, ddGTP, ddTTP and ddCTP were added to each of these reaction tubes. For cycle sequencing, the samples were denatured 2 min at 95 °C, and subjected to 30 cycles (each cycle consisting of 30 s denaturing at 95 °C, annealing 30 s at 60 °C and extending 1 min at 72 °C). After amplification, 3 µl of stop solution (95% formamide, 10 mM NaOH, 0.25% bromophenol blue) were added, denatured for 3 min at 80 °C, placed 2 min on ice, and 3 µl of each reaction were electrophoresed on a 6% sequencing gel for 90 min in 0.5× TBE buffer. Gels were then dried under vacuum at 80 °C and exposed to autoradiography film at or room temperature for 12 h.

Acknowledgements

**The first three authors contributed equally to this work. We express our gratitude to the members of the breast cancer families that participated in this study. We also thank Michelle Walsh, Beth Helmbold, Barbara Biesecker, Dorene Markel, Jeff Struewing and Joan Effiom for their help and contribution to this work. This work was supported by NIH RO1 grants CA-57601 (B.L.W.) and CA-61231 (B.L.W.), and a grant to the University of Michigan Human Genome Center (HG-00209).*

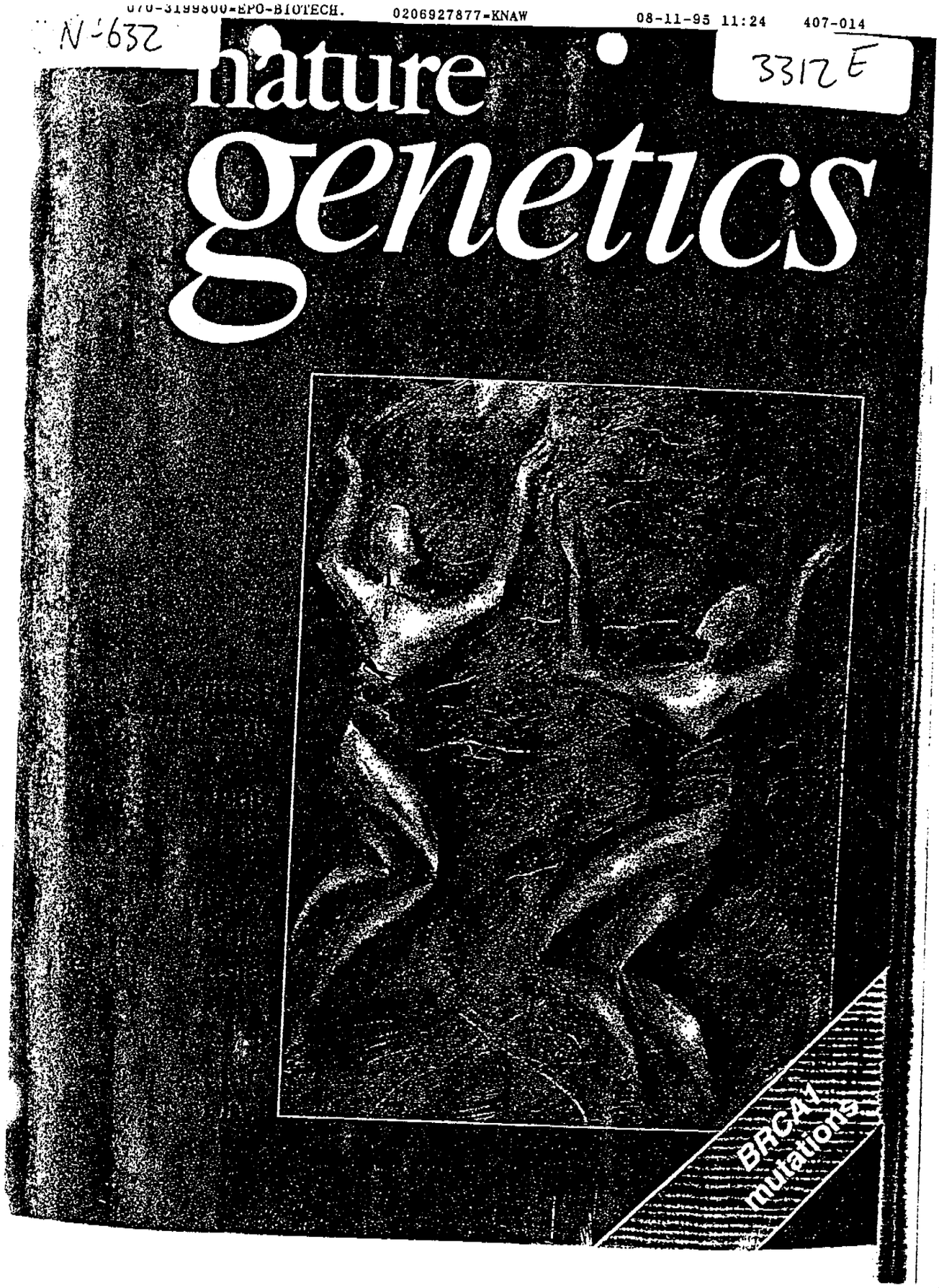
Received 3 November; accepted 8 November 1994.

1. Newman, B., Austin, M.A., Lee, M. & King, M-C. Inheritance of breast cancer: evidence for autosomal dominant transmission in high risk families. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1–6 (1988).
2. Claus, E.B., Risch, N. & Thompson, W.D. Genetic analysis of breast cancer in the cancer and steroid hormone study. *Am. J. hum. Genet.* **48**, 232–241 (1991).
3. Hall, J.M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684–1689 (1990).
4. Narod, S. *et al.* Familial breast-ovarian cancer locus on chromosome 17q12–q23. *Lancet* **338**, 82–83 (1991).
5. Easton, D.F., Bishop, T., Ford, D., Crockford, G.P. & the Breast Cancer Linkage Consortium. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. *Am. J. hum. Genet.* **52**, 678–701 (1993).
6. Wooster, R. *et al.* Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12–13. *Science* **265**, 2088–2090 (1994).
7. Ford, D., Easton, D.F., Bishop, D.T., Narod, S.A. & Goldgar, D.E. Risks of cancer in BRCA1 mutation carriers. Breast Cancer Linkage Consortium. *Lancet* **343**, 692–695 (1994).
8. Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66–71 (1994).
9. Futreal, P.A. *et al.* BRCA1 mutations in primary breast and ovarian carcinomas. *Science* **266**, 120–122 (1994).
10. Chamberlain, J.S. *et al.* BRCA1 Maps proximal to D17S579 on Chromosome 17q21 by genetic analysis. *Am. J. hum. Genet.* **52**, 792–798 (1993).
11. Orita, M., Suzuki, Y., Sekiya, T. & Hayashi, K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874–879 (1989).
12. Beroudet, A.L. & Tsui, L-C. A suggested nomenclature for designating mutations. *Hum. Mut.* **2**, 245–248 (1993).
13. Smith, S.A., Easton, D.F., Evans, D.G.R. & Ponder, B.A.J. Allele losses in the region 17q12–21 in familial breast and ovarian cancer involve the wild-type chromosome. *Nature Genet.* **2**, 128–131 (1992).
14. Kelsell, D.P., Black, D.M., Bishop, D.T. & Spurr, N.K. Genetic analysis of the BRCA1 region in a large breast/ovarian family: refinement of the minimal region containing BRCA1. *Hum. molec. Genet.* **2**, 1823–1826 (1993).
15. Knudson, A.G. Mutation and cancer: statistical study of retinoblastoma. *Proc. natn. Acad. Sci. U.S.A.* **68**, 820–823 (1971).
16. Lovering, R. *et al.* Identification and preliminary characterization of a protein motif related to the zinc finger. *Proc. natn. Acad. Sci. U.S.A.* **90**, 2112–2116 (1993).
17. Vidal-Puig, A. & Moller, D.E. Comparative sensitivity of alternative single-strand conformation polymorphism (SSCP) methods. *BioTechniques* **17**, 490–496 (1994).
18. Duncan, B. & Miller, J. Mutagenic deamination of cytosine residues in DNA. *Nature* **287**, 560–561 (1980).
19. The Huntington Disease Collaboration Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983 (1993).
20. Shiang, R. *et al.* Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia. *Cell* **78**, 335–342 (1994).
21. King, M-C., Rowell, S. & Love, S.M. Inherited breast and ovarian cancer. What are the risks? What are the choices? *J. Am. med. Assoc.* **269**, 1975–1980 (1993).
22. Biesecker, B.B. *et al.* Genetic counseling for families with inherited susceptibility to breast and ovarian cancer. *J. Am. med. Assoc.* **269**, 1970–1974 (1993).
23. National Advisory Council for Human Genome Research. Statement on the use of DNA testing for pre-symptomatic identification of cancer risk. *J. Am. med. Assoc.* **271**, 785 (1994).
24. Statement of the American Society of Human Genetics on genetic testing for breast and ovarian cancer predisposition. *Am. J. hum. Genet.* **55**, i–iv (1994).
25. The National Breast Cancer Coalition. Presymptomatic genetic testing for heritable breast cancer risk. (1994).
26. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular cloning, a laboratory manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).

GeneDX 1035, pg. 250

070-3199800-EPO-BIOTECH. 0206927877-KNAW 08-11-95 11:24 407-014
N-632
3312 E

nature genetics



070-3199800=EPO·)TECH. 0206927877=KNAW 8-11-95 11:24 407-013

article

Confirmation of *BRCA1* by analysis of germline mutations linked to breast and ovarian cancer in ten families

Lori S. Friedman, Elizabeth A. Ostermeyer, Csilla I. Szabo, Patrick Dowd, Eric D. Lynch, Sarah E. Rowell & Mary-Claire King

We provide genetic evidence supporting the identity of the candidate gene for *BRCA1* through the characterization of germline mutations in 63 breast cancer patients and 10 ovarian cancer patients in ten families with cancer linked to chromosome 17q21. Nine different mutations were detected by screening *BRCA1* DNA and RNA by single-strand conformation polymorphism analysis and direct sequencing. Seven mutations lead to protein truncations at sites throughout the gene. One missense mutation (which occurred independently in two families) leads to loss of a cysteine in the zinc binding domain. An intronic single basepair substitution destroys an acceptor site and activates a cryptic splice site, leading to a 59 basepair insertion and chain termination. The four families with both breast and ovarian cancer had chain termination mutations in the N-terminal half of the protein.

Department of Molecular & Cell Biology and School of Public Health, University of California, Berkeley, California 94720, USA

In October 1994, a candidate gene for *BRCA1*, which is responsible for inherited predisposition to breast and ovarian cancers in some families, was isolated by positional cloning[1]. In this report, we confirm that *BRCA1* is this predisposing gene by analysing germline mutations in the families that originally defined the linked phenotype[2], as well as in other families with breast cancer, and often ovarian cancer, linked to chromosome 17q21.

For this purpose, the *BRCA1* gene was screened using single strand conformation polymorphism (SSCP) analysis in 20 *BRCA1*-linked families, using both genomic DNA and cDNA prepared from lymphoblast RNA. Criteria for distinguishing cancer-predisposing mutations from polymorphisms in *BRCA1* were (i) cosegregation of the variant with breast cancer, and with ovarian cancer, if it appeared; (ii) absence of the variant in control chromosomes; and (iii) amino acid substitution in, or truncation of, the BRCA1 protein encoded by the variant sequence.

Detection of mutations in genomic and cDNA

Primers used to screen genomic DNA and cDNA for variation in *BRCA1* by SSCP analysis are indicated in Table 1. All primer pairs were used to screen all 20 families by SSCP analysis. Primers flanking exons 2, 3, 4 (an Alu sequence not found in most cDNA clones), 5, 8, 9, 10 and 12–24 are from the ftp file at morgan.med.utah.edu[1]. New primers were designed to flank exon 1, which is 100 basepairs (bp) 5' of the start site and for which surrounding genomic sequence was reported in Genbank (L182093) as an STS on 17q21 (ref. 3). In order to screen splice junctions of exons 6 and 7, intron 6 was sequenced and primers defined to yield 200–250 bp amplified products flanking exons 6 and 7 independently. Primers were also designed to amplify exon 11 in 250–300 bp fragments, using the *BRCA1* cDNA sequence published as Genbank accession number U14680 (Table 1). New SSCP primers were designed to amplify from cDNA, and are designated c1-c11 in Table 1.

Each variant detected by SSCP analysis was sequenced using several templates: the variant band from the SSCP gel, genomic DNA and/or cDNA from the proband, and genomic DNA from family members. The 18 mutations (M) or polymorphisms (P) detected in genomic and/or cDNA are also indicated in Table 1 next to the primers with which each was observed.

Analysis of *BRCA1* mutations

Two types of *BRCA1* variants fulfilling the criteria for cancer-predisposing mutations are illustrated by families 1 and 84. In family 1, an SSCP pattern reflecting a small deletion cosegregates with breast and ovarian cancer and the *BRCA1*-linked haplotype (Fig. 1). The haplotype analysis predicted that persons #1 and #8 did not inherit the linked chromosome, and the mutation analysis confirmed their *BRCA1* sequence was wild type. Thus their breast cancers at ages 71 and 45 are sporadic. The variant sequence shown has a 2 bp deletion at nucleotides (nts) 2800–2801 (codon 894), creating a stop at codon 901.

In family 84, an SSCP pattern representing a single nucleotide substitution cosegregates with breast cancer and the *BRCA1*-linked haplotype (Fig. 2). The variant genomic sequence has a thymine to guanine mutation at nt 300 of sequence U14680, resulting in a Cys61Gly substitution. This mutation removes the penultimate cysteine of the putative C3HC4 zinc-binding motif[4] — the

GeneDX 1035, pg. 252

070-3199800=EPC UOTECH. 0206927877=FNAW 08-11-95 11:24 407-012

Table 1 *BRCA1* genomic DNA and cDNA primers, polymorphisms and mutations

Primer pair	Tm	Size (bp)	DNA	U14680 bp	Forward primer	Reverse primer	Variant (bp)
exon1	55	315	genomic	1-100	5'- TAG CCC TTG GTT TCC GTG -3'	5'- TCA CAA CGC CTT ACG CCT C -3'	
c1	55	253	cDNA	28-280	5'- ACC AGG CTG TGG GGT TTC -3'	5'- TGG TTG AGA AGT TTC AGC ATG -3'	
exon2	60	250	genomic	101-199	5'- GAA GTT GTC ATT TTA TAA ACC TTT -3'	5'- TGT CTT TTC TTC CCT AGT ATG T -3'	
exon3	60	300	genomic	200-253	5'- TCC TGA CAC AGC AGA CAT TTA -3'	5'- TTG GAT TTT CGT TCT CAC TTA -3'	
c2	55	319	cDNA	212-530	5'- CAA GGA ACC TGT CTC CAC AAA GTG -3'	5'- AAG TCT TTT GGC ACG GTT TCT G -3'	M(300); M(332)
exon4	60	200	genomic	Alu	5'- GTC AAA GAG ATA GAA TGT GAG C -3'	5'- CCC GTC TCT ACA GAA AAC AC -3'	
exon5	60	200	genomic	254-331	5'- CTC TTA AGG GCA GTT GTG AG -3'	5'- TTC CTA CTG TGG TTG CTT CC -3'	M(300)
exon 6	55	200	genomic	332-420	5'- CTT ATT TTA GTG TCC TTA AAA GG -3'	5'- TTT CAT GGA CAG CAC TTG AGT G -3'	M(332)
exon 7	55	250	genomic	421-560	5'- CAC AAC AAA GAG CAT ACA TAG GG -3'	5'- TCG GGT TCA CTC TGT AGA AG -3'	
c3	55	258	cDNA	390-647	5'- TGT GCT TTT CAG CTT GAC ACA GG -3'	5'- CGT CTT TTG AGG TTG TAT CCG CTG -3'	
exon8	60	220	genomic	561-665	5'- TGT TAG CTG ACT GAT GAT GGT -3'	5'- ATC CAG CAA TTA TTA TTA AAT AC -3'	P(intron)
c4	55	298	cDNA	581-878	5'- CCA ACT CTC TAA CCT TGG AAC TGT G -3'	5'- TGC ACG CTT CTC AGT GGT GTT C -3'	
exon9	60	200	genomic	666-712	5'- CCA CAG TAG ATG CTC AGT AAA TA -3'	5'- TAG GAA AAT ACC AGC TTC ATA GA -3'	
exon10	60	220	genomic	713-768	5'- TGG TCA GCT TTC TGT AAT CG -3'	5'- GTA TGT ACC CAC TCT CTT CTT CAG -3'	
exon11AI	55	301	genomic	769-1090	5'- GGA ATT AAA TGA AAG AGT ATG AGC -3'	5'- CTT CCA GCC CAT CTG TTA TGT TG -3'	
exon11A	55	309	both	858-1168	5'- AAC ACC ACT GAG AAG CGT GCA G -3'	5'- CTC ACA CAG GGG ATC AGC ATT C -3'	
exon11B	55	300	both	1066-1365	5'- CAA CAT AAC AGA TGG GCT GGA AG -3'	5'- ACG TCC AAT ACA TCA GCT ACT TTG G -3'	P(1186)
exon11B2	55	209	both	1209-1417	5'- CCT AGA GAT ACT GAA GAT GTT CCT TGG -3'	5'- GCC AGT AAG TCT ATT TTC TCT GAA GAA CC -3'	
exon11C	55	295	both	1299-1593	5'- GGT TCT GAT GAC TCA CAT GAT GGG -3'	5'- TCT GTG GCT CAG TAA CAA ATG CTC -3'	
exon11D	55	254	both	1505-1758	5'- GAA AAC CTA TCG GAA GAA GGC AAG -3'	5'- TCA TCA CTT GAC CAT TCT GCT CC -3'	
exon11D2	55	272	both	1584-1855	5'- GAG CCA CAG ATA ATA CAA GAG CGT C -3'	5'- GCA GAT TCT TTT TCG AGT GAT TCT ATT GGG -3'	
exon11E	55	269	both	1716-1985	5'- ATC AGG GAA CTA ACC AAA CGG AG -3'	5'- CGC ATG AAT ATG CCT GGT AGA AG -3'	
exon11E2	55	410	both	1718-2127	5'- TCA GGG AAC TAA CCA AAC GGA G -3'	5'- CCA TGA GTT GTA GGT TTC TGC TG -3'	
exon11F	55	273	both	1947-2219	5'- AGG CTG AGG AGG AAG TCT TCT ACC -3'	5'- CAG CTC TGG GAA AGT ATC GCT G -3'	
exon11G	55	319	both	2142-2460	5'- GCA ACT GGA GCC AAG AAG AGT AAC -3'	5'- CCT GAG TGC CAT AAT CAG TAC CAG G -3'	P(2201); P(2430); M(2415)
exon11H	55	312	both	2196-2509	5'- CAG CGA TAC TTT CCC AGA GCT G -3'	5'- TCT GTT TTT GCC TTC CCT AGA GTG -3'	P(2430); M(2415)
exon11H2	55	266	both	2335-2620	5'- AAG TGT CTA ATA ATG CTG AAG ACC CC -3'	5'- CCC AAT GGA TAC TTA AAG CCT TCT G -3'	P(2430); M(2415)
exon11I	55	280	both	2485-2764	5'- GCA CTC TAG GGA AGG CAA AAA CAG -3'	5'- CAT TCC TCT TCT GCA TTT CCT GG -3'	P(2430)
exon11I2	55	333	both	2598-2930	5'- GAA GGC TTT AAG TAT CCA TTG GG -3'	5'- CTT ATC TTT CTG ACC AAC CAC AGG -3'	M(2800); M(2863)
exon11J	55	288	both	2716-3003	5'- GCC AGT CAT TTG CTC CGT TTT C -3'	5'- CGT TGC CTC TGA ACT GAG ATG ATA G -3'	M(2800); M(2863)
exon11K	55	305	both	2897-3201	5'- TGC AGG CTT TCC TGT GGT TG -3'	5'- GGC TAA TTG TGC TCA CTG TAC TTG G -3'	
exon11K2	55	298	both	2998-3293	5'- GCA ACG AAA CTG GAC TCA TTA CTC -3'	5'- AAT ACT GGA GCC CAC TTC ATT AGT AC -3'	P(3232)
exon11L	55	270	both	3138-3438	5'- TCA ATG TCA CCT GAA AGA GAA ATG G -3'	5'- CAG GAT GCT TAC AAT TAC TTC CAG G -3'	P(3232); M(3238)
exon11M	55	270	both	3357-3662	5'- TTG AAT GCT ATG CTT AGA TTA GGG G -3'	5'- GAC GCT TTT GCT AAA AAC AGC AG -3'	
exon11N	55	253	both	3552-3804	5'- GTT TGT TCT GAG ACA CCT GAT GAC C -3'	5'- AGT GTT GGA AGC AGG GAA GCT C -3'	P(3667); M(3726)
exon11O	55	289	both	3683-3791	5'- GAG TCC TAG CCC TTT CAC CCA TAC -3'	5'- GTG ATG TTC CTG AGA TGC CTT TG -3'	
exon11P	55	314	both	3857-4170	5'- CGT TGG TAC CGA GTG TCT GTC TAA G -3'	5'- AGC CCG TTC CTC TTT CTT CAT C -3	
c5-exon11R	55	200	genomic	4105-4215	5'- AAA GCC AGG GAG TTG GTC TGA G -3'	5'- GTG CTC CCA AAA GCA TAA A -3'	M(4184)
c5	55	312	cDNA	4105-4416	5'- AAA GCC AGG GAG TTG GTC TGA G -3'	5'- TGA TGG AAG GGT AGC TGT TAG AAG G -3'	M(4184)
exon12	60	220	genomic	4216-4302	5'- GTC CTG CCA ATG AGA AGA AA -3'	5'- TGT CAG CAA ACC TAA GAA TGT -3'	P(intron)
exon13	60	280	genomic	4303-4476	5'- AAT GGA AAG CTT CTC AAA GTA -3'	5'- ATG TTG GAG CTA GGT CCT TAC -3'	
c6	55	287	cDNA	4323-4609	5'- CAT AAC CTG ATA AAG CTC CAG CAG G -3'	5'- GAT GAC CTT TCC ACT CCT GGT TC -3'	
c7	55	290	cDNA	4445-4734	5'- GCG AAA TCC AGA ACA AAG CAC ATC -3'	5'- GCT GTT GCT CCT CCA CAT CAA C -3'	
exon14	60	250	genomic	4477-4603	5'- CTA ACC TGA ATT ATC ACT ATC A -3'	5'- GTG TAT AAA TGC CTG TAT GCA -3'	
exon15	60	250	genomic	4604-4794	5'- TGG CTG CCC AGG AAG TAT G -3'	5'- AAC CAG AAT ATC TTT ATG TAG GA -3'	
c8	55	268	cDNA	4645-4910	5'- ACA TGC ACA GTT GCT CTG GGA G -3'	5'- GGT TGA AGA TGG TAT GTT GCC AAC -3'	
exon16	60	375	genomic	4795-5105	5'- AAT TCT TAA CAG AGA CCA GAA C -3'	5'- AAA ACT CTT TCC AGA ATG TTG T -3'	P(4956)
c9	55	307	cDNA	4837-5143	5'- ATG ACC CTG AAT CTG ATC CTT CTG -3'	5'- GTG ATG TGG TGT TTT CTG GCA AAC -3'	P(4956)
c10	55	290	cDNA	5092-5381	5'- CCC CAG AAG AAT TTA TGC TCG TG -3'	5'- TTC TCT TGC TCG CTT TGG ACC -3'	
exon17	60	350	genomic	5106-5193	5'- GTG TAG AAC GTG CAG GAT TG -3'	5'- TCG CCT CAT GTG GTT TTA -3'	P(intron)
exon18	60	350	genomic	5194-5273	5'- GGC TCT TTA GCT TCT TAG GAC -3'	5'- GAG ACC ATT TTC CCA GCA TC -3'	
exon19	60	220	genomic	5274-5310	5'- CTG TCA TTC TTC CTG TGC TC -3'	5'- CAT TGT TAA GGA AAG TGG TGC -3'	
exon20	60	220	genomic	5311-5398	5'- ATA TGA CGT GTC TGC TCC AC -3'	5'- GGG AAT CCA AAT TAC ACA GC -3'	
c11	55	318	cDNA	5331-5648	5'- GGA GAT GTG GTC AAT GGA AGA AAC -3'	5'- TGC TAC ACT GTC CAA CAC CCA CTC -3'	
exon21	60	275	genomic	5387-5451	5'- AAG CTC TTC CTT TTT GAA AGT C -3'	5'- GTA GAG AAA TAG AAT AGC CTC T -3'	
exon22	60	275	genomic	5452-5526	5'- TCC CAT TGA GAG GTC TTG CT -3'	5'- GAG AAG ACT TCT GAG GCT AC -3'	
exon23	60	250	genomic	5527-5586	5'- CAG AGC AAG ACC CTG TCT C -3'	5'- ACT GTG CTA CTC AAG CAC CA -3'	
exon24	60	275	genomic	5587-5711	5'- ATG AAT TGA CAC TAA TCT CTG C -3'	5'- GTA GCC AGG ACA GTA GAA GGA -3'	M(5677)

The names of 58 primer pairs used to screen 20 *BRCA1*-linked families by SSCP analysis are listed in the first column. T_m indicates the annealing temperature for each PCR primer pair. The size of the PCR product is given in basepairs. The source of DNA template used for each primer pair was genomic DNA, cDNA prepared from lymphocyte RNA, or both genomic DNA and cDNA. Listed in column five are the nucleotides of the *BRCA1* cDNA sequence (U14680) that are analysed by the corresponding primer pair. Nucleotide locations of polymorphisms (P) and mutations (M) are indicated at their position in the *BRCA1* cDNA sequence.

GeneDX 1035, pg. 253

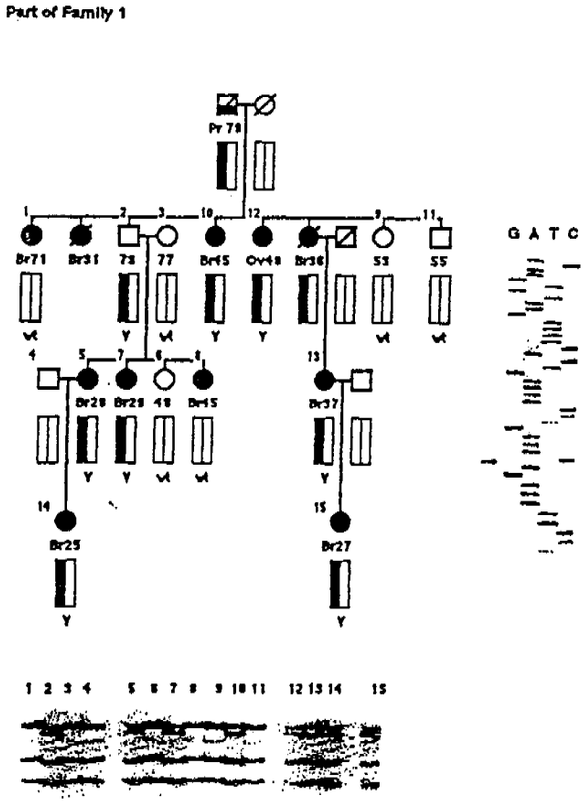


Fig. 1 Frameshift mutation in *BRCA1* linked to breast and ovarian cancer in family 1. Symbols under each relative indicate chromosome 17q21 haplotypes linked (filled rectangles) or not linked (open rectangles) to *BRCA1*. Cancers are indicated as follows: Br, breast cancer; Ov, ovarian cancer; Pr, prostate cancer. Ages at diagnosis are also indicated. V indicates variant sequence and wt indicates wildtype sequence for *BRCA1*, based on SSCP analysis of 15 relatives, shown below the pedigree. The mutant sequence, shown on the right, has a 2 bp deletion (AA) at nts 2800–2801 (arrow), leading to a stop at codon 901. The wild type sequence is TCCTTAAAGAAACAAAGTCCAAAGTC-ACTTTTGAATGTGAACA, with the mutated sequence underlined.

only recognizable domain so far in the BRCA1 protein. This variant does not appear in 120 control chromosomes, suggesting that it is not simply a rare polymorphism.

In all, we identified mutations cosegregating with breast and ovarian cancer and linked to *BRCA1* in ten families (Table 2). None of these mutations has been previously observed (D. Goldgar, personal communication)[1]. The ten families have nine different mutations: the same thymine to guanine substitution at nt 300 (the Cys61Gly mutation) appeared twice, in families 84 and 4. This mutation probably arose independently in the two families, because chromosome 17q21 marker alleles defining the *BRCA1* haplotypes of the two families differ. The ancestry of family 84 is Polish and of family 4 German; insofar as the families are aware, they are not related. The only other missense mutation in our families leads to the substitution of asparagine for serine at codon 1040 in all patients with breast cancer linked to *BRCA1* in family 14. The functional significance of this mutation is currently unknown and might represent either a disease-predisposing allele or a rare polymorphism.

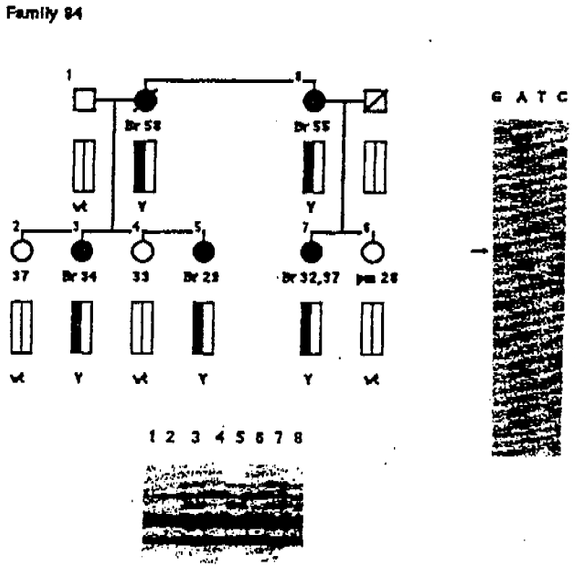


Fig. 2 Missense mutation in the zinc-binding domain of BRCA1 linked to breast cancer in Family 84. Symbols are as described in Fig. 1; individual #6 had a prophylactic mastectomy at age 28, ten years prior to entrance into the study. The mutant sequence, shown on the right, illustrates the T to G substitution at nt 300 (arrow), which leads to replacement of the penultimate Cys in the zinc binding domain with a Gly residue. This Cys61Gly mutation also occurred, probably independently, in Family 4. The wild type sequence is GCTGAAACTTCTCAACCAGAAGAAA-GGGCCTTCACAGTGTGTCCTTTATGTAAGAATGATATA-ACCAAAAGGTAT, with the mutated sequence underlined.

GeneDX 1035, pg. 254

Table 2a Germline mutations cosegregating with breast and ovarian cancer in *BRCA1*-linked families

Family	*BRCA1* nt	Exon	Mutation	Amino acid change	Predicted effect
4	300	5	T to G substitution (TGT→GGT)	Cys61Gly	lose zinc-binding motif
84	300	5	T to G substitution (TGT→GGT)	Cys61Gly	lose zinc-binding motif
82	332	Intron 5	T to G substitution→59 bp insertion	75Stop	protein truncation
3	2415	11	AG deletion	Ser766Stop	protein truncation
1	2800	11	AA deletion	901Stop	protein truncation
102	2863	11	TC deletion	Ser915Stop	protein truncation
14	3238	11	G to A substitution (AGC→AAC)	Ser1040Asn	missense
74	3726	11	C to T substitution (CGA→TGA)	Arg1203Stop	protein truncation
2	4184	11	TCAA deletion	1364Stop	protein truncation
77	5677	24	A insertion	Tyr1853Stop	protein truncation

Table 2b Sites, ages at diagnosis and laterality of cancers associated with *BRCA1* mutations

Family	Breast cancer		Ovarian cancer		Other cancers with with *BRCA1* mutations
	Cases (bilateral)	Avg. age at dx	Cases	Avg. age at dx	
4	10 (1)	39.7			
84	5 (1)	41.6			
82	6 (2)	40.3	2	52.1	Small intestine (dx 46)
3	3 (0)	37.3	4	46.3	
1	16 (0)	37.8	2	54.5	3 Prostate (dx 57, 77, 79)
102	5 (0)	45.2	3	51.7	
14	6 (2)	51.1			
74	6 (1)	45.5			Kidney (dx 55); Colon (dx. 40)
2	5 (1)	38.2			
77	7 (1)	32.9			

However, we have not found the mutation in 120 control chromosomes.

The remaining seven mutations all lead to truncated proteins, mostly due to small insertions or deletions of 1 to 4 bp (Table 2). Four mutations are immediate stops; the others are frameshifts leading to stops after 4, 7 and 9 codons, respectively. In family 3, one AG doublet is deleted from an $(AG)_3$ repeat; the other small length changes appeared to be random throughout the sequence. In family 82, a thymine to guanine substitution occurs within the consensus sequence 5' of the exon 6 splice acceptor site (Fig. 3). This substitution leads to the preferential activation in intron 5 of a cryptic splice acceptor[4]. This generates a mutant transcript containing a 59 bp insertion of intronic sequence 5' of exon 6, causing a stop at codon 75. Both the wild type and mutant transcripts are detected in lymphocyte cDNA with primer pair c2.

One variant observed only at the cDNA level could not yet be classified as a predisposing mutation. In two patients from family 5 with *BRCA1*-linked breast and ovarian cancer, cDNA prepared from lymphoblast RNA lacked exon 3. (No RNA from breast tumour is available for this family.) Exon 3 splice junction sequences are normal in genomic DNA from these patients. The absence of exon 3 has not been observed in lymphoblast cDNA from controls or from other breast cancer families. This case differs from the family 82 variant in that in family 5, an entire exon is

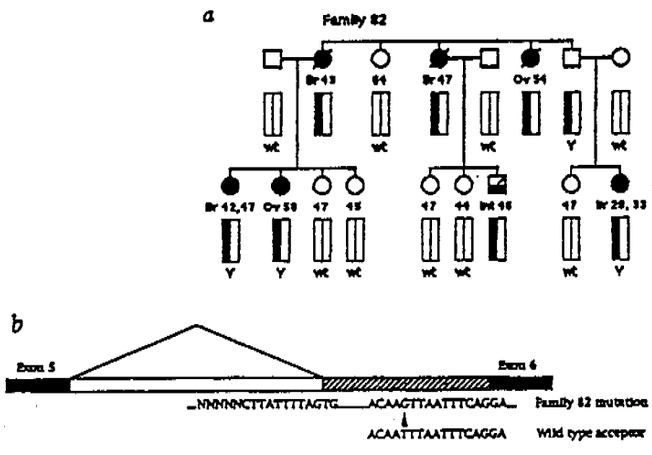


Fig. 3 Aberrant splicing leading to a truncated BRCA1 protein in Family 82. *a*, *BRCA1*-linked patients and unaffected relatives of Family 82. Mutant (*V*) or wildtype (wt) sequence was confirmed in living relatives. (Symbols are as given in Fig. 1). *b*, Aberrant splicing of exons 5 and 6 in Family 82. A cryptic splice acceptor site in intron 5 (open box) leads to insertion of 59 bp (hatched box) into the mutant transcript. Sequences of exons 5 and 6 (filled boxes) are normal. The cryptic acceptor is activated by a T to G substitution in the usual acceptor site. The consensus splice acceptor sequence is $(Py)_{10}NCAG/N$.

GeneDX 1035, pg. 255

Table 3 Neutral polymorphisms in *BRCA1*

BRCA1 nt	Nucleotide variant	Amino acid variant	Allele frequencies
Intron 8	T deleted	non-coding	.67/.33
1186	A/G	Gln356Arg	.82/.13
2201	C/T	silent	
2430	T/C	silent	.67/.33
3232	A/G	Glu1038Gly	.67/.33
3667	A/G	Lys1183Arg	.67/.33
intron 12	CA deleted	non-coding	.95/.05
4956	A/G	Ser1613Gly	.67/.33
intron 16 or 17	not sequenced	non-coding	.69/.31

missing and no intronic sequence has been introduced. It is possible that absence of exon 3 represents alternative splicing characteristic of lymphoblasts we immortalized for this family, rather than having any relationship to cancer predisposition. Alternatively, an inherited mutation in genomic DNA leading to consistent aberrant splicing would strongly suggest that the variant is disease-related.

BRCA1 polymorphisms

Nine polymorphisms were detected in *BRCA1*, six in the coding sequence and three in introns (Table 3). Four polymorphisms alter amino acid residues. Five polymorphisms are in complete disequilibrium with each other, and a sixth is in partial disequilibrium, based on genotypes of 42 chromosomes. Identification of these polymorphisms enabled us to screen for mutations causing loss of *BRCA1* transcript in the ten families for which no mutations were found by SSCP analysis of genomic DNA and cDNA.

Genomic DNAs from probands of these families were genotyped for polymorphisms in the *BRCA1* coding sequence. Probands of five families with undetected mutations were heterozygous for one or more polymorphisms. The cDNA of these five probands were also heterozygous, indicating that mRNA was being expressed from both *BRCA1* alleles. For the remaining five families, patients heterozygous for coding sequence polymorphisms will be identified and the same test carried out.

Discussion

Most of the *BRCA1* mutations we have identified result in premature termination of protein translation. Such truncated proteins could conceivably compete with wildtype protein, thereby acting as dominant mutations. On the other hand, loss of the wildtype allele in the tumours of patients with *BRCA1* germline mutations would suggest the opposite role — that the termination mutation was recessive. Paraffin blocks with adequate malignant cells for loss-of-heterozygosity (LOH) studies were available for one tumour from family 1, five tumours from family 3 and two tumours from family 82. For all eight of these tumors, the wildtype *BRCA1* allele was lost, an observation consistent with recessive mutations[5].

The number of families with *BRCA1* mutations identified is still very small, so associations between genotype and phenotype are not at all certain. Cancer-related mutations occur throughout the *BRCA1* sequence (Table 2). In our series, all four families with both ovarian and breast cancer had truncation mutations in the first half of the protein. However, termination of the BRCA1 protein only 11 amino acids from the C-terminal end was associated with very early-onset breast cancer in family 77. Bilateral breast cancer was associated with mutations throughout the sequence, although ascertainment of bilaterality is biased in very high-risk families in that some women choose bilateral mastectomy following their first diagnosis of breast cancer. In this series, there is no association between age at onset and locale or type of mutation, but these families were selected for early age at breast cancer diagnosis and probably do not represent *BRCA1* patients generally. Better evaluation of outcomes associated with *BRCA1* mutations of different types and at different locations in the sequence will be based on identifying germline *BRCA1* mutations in population-based series of cases not selected for family history.

The mutations found, and not found, in these 20 families illustrate the complexities of screening for inherited *BRCA1* mutations in families, let alone in the general population[4,7]. First, mutations were identified in only 50% of the families in this series, although both genomic DNA and RNA were screened by SSCP analysis and all families have convincing evidence of *BRCA1* linkage[2,4]. The families for which *BRCA1* mutations have not yet been identified are as likely to carry mutations as those successfully screened. The posterior probability of linkage is ≥.95 for families with and without definable mutations. The two groups of families are similar in number and ages at diagnosis of breast and ovarian cancers linked to *BRCA1*; ovarian cancer occurred in four of ten families with mutations found and in five of ten families with mutations still undetected. Additional mutations may prove detectable by other screening techniques such as mismatch cleavage analysis, or by Southern or northern hybridization. However, it is likely that a large fraction of mutations will only be detectable when genomic sequence of the entire gene is available. The importance of genomic sequence is illustrated by the successful detection of two mutations in cDNAs: the 59 bp intronic insertion in family 82 reported here and the loss of transcript in family 2035 previously described[1]. The intronic mutation leading to aberrant splicing in family 82 could only be identified from genomic sequence. In general, mutations in introns and promoter regions can only be identified reliably when sequence of the entire 80 kb *BRCA1* gene is known and critical regions identified therein. The alternative would be to screen breast or ovarian RNA for mutations using multiple techniques, an approach not feasible on a wide scale.

Second, most *BRCA1* variants are not predisposing mutations. Most individuals, including women at high risk of breast cancer, are heterozygous at one or more sites in the *BRCA1* gene. Most of these variants are polymorphisms, some of which will be revealed, after sequencing, to be noncoding or silent. However, in this series, four missense mutations were revealed that altered peptide sequence. Because they occurred at similar frequencies in cases and controls, we concluded that they are polymorphisms. Even by screening cases and controls, some rare polymorphisms may be mistakenly identified as predisposing mutations. The mutation at nt 3238 in family 14 may prove to be of this type, and we would not predict risk for an asymptomatic woman based on that

GeneDX 1035, pg. 256

mutation without determining its biological function.

Third, there are many predisposing mutations in *BRCA1*, and the risks associated with most are not known. The mutations identified so far have high penetrance, because they were identified in families ascertained for very high lifetime risk of breast cancer. Whether other *BRCA1* mutations are associated with more moderate risk, possibly influenced by other genetic or environmental factors, remains to be seen.

Methodology

SSCP. PCR was carried out in 50 µl volumes containing: 50 ng cDNA or genomic DNA, 1× PCR buffer (Boehringer Mannheim), 200 µM dATP, dGTP, dTTP (Boehringer Mannheim), 10 µM dCTP (Boehringer Mannheim), 50 pmoles each primer from Table 1 (Operon), 0.5 µCi ^{32}P-dCTP (NEN Dupont), 1.25 U *Taq* DNA polymerase (Boehringer Mannheim). PCR conditions for primer pairs were 35 cycles of 94 °C, 45 s; annealing temperature (T_a from Table 1) 30 s; 72 °C; 30 s. PCR template was lymphocyte cDNA or genomic DNA from members of *BRCA1*-linked families who carry the predisposing haplotype. Amplified samples were diluted 1:10 in formamide buffer (98% formamide, 10 mM EDTA pH 8, 0.05% Bromophenol blue, 0.05% Xylene cyanol), held at 95 °C for 5 min, then cooled rapidly to 4 °C, and held for 5 min. For each sample, 5 µl was loaded onto an SSCP gel and run at 6 W (constant power) for 16 h in 0.6× TBE at room temperature. (An 80 ml gel solution contains: 0.5× MDE (AT Biochem), 0.6× TBE, 160 µl 25% ammonium persulphate, 38 µl TEMED.) Gels were dried on a vacuum gel dryer and exposed to film for 12–24 h with an intensifying screen. Variant bands were cut out of the gel and rehydrated in 100 µl water.

Sequencing. PCR was carried out in 50 µl volumes containing: 50 ng DNA or cDNA or 1 µl of rehydrated SSCP gel fragment solution, 1× PCR buffer (Boehringer Mannheim), 200 µM dATP, dCTP, dTTP, dGTP (Boehringer Mannheim), 1.25 U *Taq* polymerase (Boehringer Mannheim), 50 pmoles each primer (Operon) that detected the SSCP variant. Cycling conditions were 35 cycles of 94 °C, 45 s, annealing temperature (T_a in Table 1) 30 s, 72 °C, 30 s. Direct sequencing of the double-stranded DNA from the PCR amplification was performed with the USB PCR product sequencing kit, using ^{35}S ATP as the isotope. Sequencing primers were the same as those used to amplify the template. Sequence was run on 6% acrylamide/bis gels at 70 W constant power for 2 h, and exposed to X-ray film for 12–24 h. Mutations were confirmed in multiple members of each family by amplifying and directly sequencing from genomic DNA.

Intron 6 sequence was obtained by amplifying genomic DNA using the primer pair exon 6&7 (ftp file at morgan.med.utah.edu), with the reaction mixture described above. Direct sequencing was done on the 800 bp product as described above.

Acknowledgements

This work was supported in part by NIH grant R01 CA27632. L.S.F. is a Komen Foundation Fellow; C.L.S. is an NIH Fellow; M-C.K. is an American Cancer Society Professor.

Received 8 November; accepted 9 November 1994.

1. Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266, 66–71 (1994).
2. Hall, J.M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250, 1684–1689 (1990).
3. Albertsen, H.M. *et al.* A physical map and candidate genes in the BRCA1 region on chromosome 17q12–21. *Nature Genet.* 7, 472–479 (1994).
4. Sharp, P.A. Split genes and RNA splicing. *Cell* 77, 805–815 (1994).
5. Friedman, L.S. *et al.* The search for BRCA1 and the cloning of 22 genes from chromosome 17q21. *Cancer Res.* (in the press).
6. American Society of Human Genetics. Statement on genetic testing for breast and ovarian cancer. *Am. J. hum. Genet.* 55, 860 (1994).
7. Rowell, S., Newman, B., Boyd, J. & King, M-C. Inherited predisposition to breast and ovarian cancer. *Am. J. hum. Genet.* 55, 861–865 (1994).
8. Arena, J.F. *et al.* Inherited breast cancer: Confirmation of linkage to BRCA2, possible linkage to the estrogen receptor, and the clinical implications of genetic complexity. *J. Am. med. Assoc.* (in the press).

GeneDX 1035, pg. 257

Am. J. Hum. Genet. 52:743–748, 1993

Chromosome 17q Linkage Studies of 18 Utah Breast Cancer Kindreds

D. E. Goldgar, L. A. Cannon-Albright, A. Oliphant, J. H. Ward, G. Linker, J. Swensen, T. D. Tran, P. Fields, P. Uharriet, and M. H. Skolnick

Genetic Epidemiology Group, University of Utah, Salt Lake City

Summary

In this paper we present linkage results from the analysis of 18 Utah breast cancer kindreds, for three 17q markers. Four kindreds had LOD scores greater than 1.0 for at least one of the marker loci. One of these kindreds has a LOD score of 6.07 with D17S579, and we believe it to be the most informative 17q family reported to date. Among the kindreds which appear unlinked to 17q were an early-onset breast cancer family, a large breast-ovarian family, and a kindred with mixed age at onset. Analysis of individual recombinants in the linked families localizes the BRCA1 gene between THRA1 and D17S579 (Mfd188). A comparison of the Cancer and Steroid Hormone Study (CASH) model and a model which assumes a rare dominant susceptibility locus with low penetrance and no phenocopies stresses the difficulties in assessing linkage if the assumptions of the CASH model in terms of age at onset of breast cancer are not appropriate for the BRCA1 locus. A hypothetical breast cancer pedigree is used to calculate gene carrier probabilities under the CASH model, thereby illustrating some of our concerns regarding the use of this model to detect and exclude 17q linkage in breast cancer families.

Introduction

Breast cancer is the most common cancer among American women and has long been recognized to be, in part, a familial disease (Anderson 1972). Studies have shown that a woman's risk of developing breast cancer is increased if one or more first-degree relatives have had breast cancer. A recent analysis of the inherited susceptibility of breast cancer (Claus et al. 1991) in the large data set collected as part of the Cancer and Steroid Hormone Study (CASH) demonstrated a rare dominant susceptibility allele for breast cancer. This population-based study collected family history information on first-degree relatives in a series of 4,730 breast cancer probands diagnosed between the ages of 20 and 54 years. The increased risk due to this susceptibility allele ranged from almost 100-fold in women in their 20s to a modest, twofold increase in women in their 80s. Hall et al. (1990) have shown that breast cancer susceptibility in some kindreds with early age at onset is linked to chromosome 17q. In another study, three of five kindreds with both breast cancer and ovarian cancer confirm the reported linkage (Narod et al. 1991).

Our initial studies of breast cancer were statistical studies of the Utah Population Database, the genealogy of the descendants of the Utah pioneers, linked to the Utah Cancer Registry. This data set allowed us to ascertain large cancer-prone kindreds from a well-defined population. Starting with genealogical records kept in the Utah Genealogical Society, we constructed a genealogy of 1.5 million Utah descendants of approximately 10,000 Mormon pioneers (Skolnick 1980). The Utah cancer registry has been part of the National Cancer Institute's Surveillance, Epidemiology, and End Results system since its inception in 1973. It began statewide efforts in the 1960s, with fairly complete incidence data available since 1966; it currently has more than 113,000 cases. Links were created between the genealogy and the cancer registry whenever unique identity could be established. This resource allows us to obtain informa-

Received June 15, 1992; final revision received December 1, 1992.
Address for correspondence and reprints: David E. Goldgar, Ph.D., Genetic Epidemiology, 420 Chipeta Way, Suite #180, Salt Lake City, UT 84108.
© 1993 by The American Society of Human Genetics. All rights reserved.
0002-9297/93/5204-0012$02.00

GeneDX 1035, pg. 258

tion on a proband's relatives before beginning a study, specifically their cancer status, family size, and location within the state. In this study we examine linkage to 17q markers for 18 kindreds ascertained for multiple cases of breast or ovarian cancer and discuss the implications of the CASH model for detecting or excluding linkage of breast cancer to genetic markers.

Methods

The 18 families which are reported in this paper were ascertained from a number of sources. Fourteen of these families were submitted to the Breast Cancer Linkage Consortium (BCLC) and are included in the analyses reported in the consortium summary paper (Easton et al. 1993), although it should be noted that, because of substantial work done on these families after the final submission of data to the BCLC, there may be differences between the results obtained in the analyses reported here and those in the consortium summary paper. Four additional families were ascertained too recently for BCLC submission. Moreover, genotypic data for the marker D17S579 (MFD188) included in this report were not available for the consortium analysis. These 18 families were ascertained as follows.

Ten families which had three or more cases with inferrable genotypes were selected for linkage studies to 17q markers, from a set of 29 families originally ascertained for a study of proliferative breast disease and breast cancer (Skolnick et al. 1990). Additionally, we included two families (K1001 and K9018) which we have studied since 1980, as part of our breast cancer linkage studies; five families (K2019, K2073, K2079, K2080, and K2082) ascertained from the Utah Population Database for the presence of clusters of breast and/or ovarian cancer; and a self-referred family (K2035) with early-onset breast cancer. These 18 families were investigated and expanded in our clinic and were typed for chromosome 17q markers. Table 1 displays the characteristics of the 18 families included in the present study.

Analysis of these families, as well as other families contributed to the consortium, localized the breast cancer locus between two polymorphic CA-repeat markers: 42D6 (D17S588), a CA repeat isolated in our laboratory; and Mfd15 (D17S250), a CA repeat provided by Jim Weber (Weber et al. 1990). LOD scores of each family, with each of these two markers and with a third marker, Mfd188 (D17S579; Hall et al. 1992), located roughly midway between these two markers, were calculated at a recombination fraction of .001, under two models for the breast cancer locus. The first model was derived by Claus et al. (1991) and has an estimated gene frequency of $q = .0033$, a lifetime risk in gene carriers of about .80, and age-specific risks for breast cancer in nongene carriers. In the second model (affected only [AO]), all cases are assumed to be gene carriers of a rare autosomal dominant gene ($q = .0001$), while all unaffected individuals are assumed to be of unknown phenotype, regardless of age. While the nophenocopies assumption implicit in the AO model is certainly unrealistic for a common disease such as breast cancer, the model is useful for identifying potential recombinants and is better suited to evaluate linkage in families with later age at onset. For the three markers used for the LOD score calculations, allele frequencies were calculated from our own laboratory typings of unrelated individuals in the CEPH panel. In some of the families, additional polymorphic short tandem repeat markers in the region were typed in order to clarify haplotypes; these markers included 26C2 (D17S514; Oliphant et al. 1991), THRA1 (Futreal et al. 1992), and three polymorphisms—NM23, SCG40 (D17S181), and 6C1 (D17S293)—provided to us by Dr. Donald Black (Hall et al. 1992). The order of these loci which was assumed in constructing haplotypes is centromere–26C2–Mfd15–THRA1–Mfd188–SCG40–6C1–42D6–NM23–telomere; this order is based on our own CEPH analysis, as well as on data reported at the 1992 chromosome 17 workshop (Fain 1992).

Results

Table 2 shows the results of the pairwise linkage analysis of each family, with the three markers 42D6, Mfd188, and Mfd15. When a LOD score greater than 1.0 for at least one locus under the CASH model is used as the criterion for linkage to 17q, it would appear that, of these 18 families, 4 (K1901, K1925, K2035, and K2082) are linked to 17q. Three of these kindreds had an informative recombinant, and these will be detailed below.

Kindred 2082 is the largest 17q-linked breast cancer family reported, to date, by any group. The kindred contains 18 cases of breast cancer, 8 cases of ovarian cancer, and 2 cases with both ovarian cancer and breast cancer. The evidence of linkage to 17q in this family is overwhelming; the LOD score with the linked haplotype is greater than 6.0, despite the existence of three cases of breast cancer which appear to be sporadic, that is, they share no part of the linked haplotype from MFD15–42D6. These three sporadic cases were diagnosed with breast cancer at ages 46, 47, and 54 years.

GeneDX 1035, pg. 259

Table 1

Description of the 18 Kindreds

Kindred[a]	No. of Individuals		Breast Cancer				Ovarian Cancer			
				Age at Diagnosis (years)				Age at Diagnosis (years)		
	Total	Typed	No. Affected	Minimum	Median	Maximum	No. Affected	Minimum	Median	Maximum
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[b]	42	19	10	42	53	79	...	...	...	...
1900	70	23	8	45	55	70	1	...	78	...
2080[b]	264	74	22[c]	27	55	92	4	45	53	71
2073[b]	57	29	9	35	57	80	...	...	...	...
1917	16	6	4	43	58	61	...	...	...	...
1920	22	14	3	62	63	68	...	...	...	...
2079[b]	136	18	14	38	66	84	4	52	59	65

NOTE.—Three women diagnosed with both breast cancer and ovarian cancer are counted in both categories.

[a] Kindreds are listed in ascending order of median age at diagnosis of breast cancer.

[b] Not included in collaborative data set.

[c] Includes one case of male breast cancer.

The key recombinant in this pedigree is a woman who developed ovarian cancer at age 45 years and whose mother and aunt had ovarian cancer at ages 58 and 66 years, respectively. She inherited the linked portion of the haplotype for all typed loci distal to Mfd15, including Mfd188, SCG40, 42D6, and NM23, while inheriting unlinked alleles at Mfd15; THRA1 was not informative in this sibship. This recombinant event places BRCA1 distal to Mfd15.

K1901 is typical of early-onset breast cancer families seen by other groups (Hall et al. 1990). The kindred contains 10 cases of breast cancer with a median age at diagnosis of 43.5 years; four cases were diagnosed under age 40 years. The LOD score of this kindred, with the marker 42D6, is 1.5; if the prior probability of 17q linkage for site-specific breast cancer (Easton et al. 1993) is assumed to be .45, this corresponds to a posterior probability of .96 for linkage to this marker. Examination of multilocus haplotypes in this pedigree identifies a recombinant haplotype in an obligate male carrier and his affected daughter, who was diagnosed with breast cancer at age 45 years. Their haplotype for markers Mfd15 and THRA1 differs from that found in all other cases in the family (except one case which could not be completely inferred from her children). The two haplotypes are identical for loci below THRA1, including Mfd188, SCG40, 6C1, and 42D6. Accordingly, kindred 1901 would place the BRCA1 locus distal to THRA1.

Kindred 2035 is similar, in phenotype, to K1901. The median age at diagnosis for the eight cases of breast cancer in this family is 37 years. One case also had ovarian cancer at age 60 years. The breast cancer cases in this family descend from two sisters who were both unaffected with breast cancer until their death in their 8th decade. Each branch contains four cases of breast cancer, with at least one case in each branch having markedly early onset. This family has a LOD score of 2.34, with Mfd15. The haplotypes segregating with breast cancer in the two branches share an identical allele at 26C2, Mfd15, and THRA1, but differ for loci distal to THRA1, i.e., Mfd188, SCG40, 6C1, and

GeneDX 1035, pg. 260

Table 2

LOD Scores of Utah Families, with 42D6, Mfd188, and Mfd15, under Two Genetic Models of Breast Cancer: CASH and AO

	LOD Score[b]					
	42D6 (D17S588)		Mfd188 (D17S579)[c]		Mfd15 (D17S250)	
Kindred[a]	CASH	AO	CASH	AO	CASH	AO
1910	.06	.30	.06	.30	.06	.30
1001	−.30	−1.40	NT	NT	−.52	−3.93
2035	2.34	2.24	.94	−.80	2.34	2.29
2027	−1.22	−2.22	−1.20	−2.23	−1.16	−2.14
9018	−.54	−1.57	−.17	−.18	.11	.13
1925	1.08	.82	.55	.28	−.11	.17
1927	−.41	−1.50	−.35	−1.46	−.44	−1.58
1911	−.27	−1.69	−.43	−1.68	.49	.75
1929	−.49	−2.32	NT	NT	−.49	−2.32
1901	1.50	1.42	.78	.91	.65	−1.31
2082	4.25	3.83	6.07	2.88	2.00	−2.43
2019	−.10	−1.39	−.11	−.58	−.18	−4.06
1900	−.14	−2.69	NT	NT	−.12	−1.67
2080	−.16	−3.77	.76	−4.12	−1.25	−7.81
2073	−.41	−4.97	.63	−.39	−.23	−3.02
1917	−.02	−.30	NT	NT	−.01	.05
1920	−.03	−2.32	NT	NT	.00	.13
2079	.02	−1.89	−.01	−2.10	.01	−1.99

[a] Kindreds are listed in ascending order of median age at diagnosis of breast cancer.
[b] A recombination fraction of .001 between the breast cancer locus and each marker is assumed.
[c] NT = kindred not typed.

NM23. Although the two haplotypes are concordant for marker 42D6, it is likely that the alleles are shared identical by state rather than identical by descent, since the shared allele is the second most common allele at this locus. By contrast, the linked allele shared at Mfd15 has a frequency of .04. This is perhaps 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. For this event not to be a key recombinant requires that a second mutant BRCA1 gene be present in a married-in spouse who also shares the rare Mfd15 allele segregating with breast cancer in both branches of the family. The evidence from this family therefore places the BRCA1 gene proximal to Mfd188.

One additional family shows some evidence of linkage to 17q. Kindred 1911 gives a modest positive LOD score, with Mfd15, under both models but shows evidence of recombination with both Mfd188 and 42D6. Whether this reflects either a recombination event between BRCA1 and the distal markers or lack of informativeness in the key individual cannot be determined.

Among the families which appear not to be linked to 17q, K1001, K2027, and K2080 are of particular interest. K1001 is a large kindred with generally early onset, in which eight cases of breast cancer were typed or could be inferred for markers in the region. There are at least three unique haplotypes present among the breast cancer cases in this family, although the LOD score under the CASH model is essentially uninformative. Linkage to 17q in this family requires that three cases of breast cancer with ages at diagnosis of 49, 54, and 61 years be of sporadic origin. Kindred K2027 contains four cases of early-onset breast cancer (age at diagnosis 34–41 years) with a recombinant which excludes the breast cancer susceptibility locus from the Mfd15-42D6 region of 17q. It appears unlinked to all three markers, even under the CASH model. Kindred K2080 is a large kindred ascertained from the Utah Population Database for the presence of breast and ovarian cancer. Although this kindred has four cases of breast cancer diagnosed between the ages of 27 and 45 years, the kindred is characterized by largely postmenopausal breast cancer, a smaller degree of postmenopausal

GeneDX 1035, pg. 261

ovarian cancer, and a case of male breast cancer. This family gives highly negative LOD scores for all three markers under the AO model but yields a LOD score of 0.76 for Mfd188 under the CASH model. The structure of this family is quite fragmented, and therefore its linkage to 17q is difficult to assess directly. However, of the 14 cases of breast cancer which could be typed or inferred, the allele most often shared among the cases was present seven times for each marker. For Mfd188, the most commonly shared allele is quite rare; this undoubtedly accounts for the observed positive LOD score. In this family, if the bulk of breast cancer cases are due to a single susceptibility locus, the inherited susceptibility does not appear to be due to a gene located on 17q, although the moderate positive LOD score obtained under the CASH model for Mfd188 may warrant further investigation. Although genotyping is in the early stages in family K2079, it shows a phenotypic and linkage pattern similar to that of kindred K2080.

Discussion

In our 18 families, we estimate that 17q-linked, 17q-unlinked, and families which are currently uninformative are present in roughly equal numbers. Surprisingly, of the six families which contain at least one case of ovarian cancer, two are almost certainly linked to 17q, while perhaps three are unlinked and one is largely uninformative. It is noteworthy that the two apparently linked breast-ovarian families have relatively early onset of breast cancer. Thus, the presence of early-onset breast cancer may be a more distinguishing feature of the 17q-linked susceptibility locus than is the presence of ovarian cancer. However, it is important to note that the CASH model under which the data are typically analyzed makes it difficult to assess linkage in later-onset breast cancer families. When the data are analyzed under a model which assumes all cases to be gene carriers, there is no readily apparent relationship between the median age at diagnosis and presence or absence of linkage to 17q. Of the 10 kindreds with median onset age less than 45 years, 5 appear to be unlinked to 17q; in the 8 with later age at diagnosis, there is at least 1 17q-linked family.

Several conclusions emerge from comparing the LOD scores obtained under the two models shown in table 2. First, in the families with both exclusively or primarily early-onset cancer and reasonably large positive LOD scores, there is very little difference in the LOD scores obtained under the two models, suggesting that, under the CASH model, data from old unaffected individuals are not utilized. Second, in families with primarily later onset, there is little linkage information available under the CASH model, even for families which are highly informative under the AO model. Last, and perhaps most significant, when families have a mixed pattern of age at onset (i.e., kindreds K1001, K1911, K1929, K2073, K9018, and K2082), there can be a large discrepancy between the LOD score obtained under the AO model and that obtained under the CASH model. To better separate the effects of not allowing for the phenocopies in the model versus ignoring information from unaffecteds, we recalculated the LOD scores under a model which assumed that all unaffected individuals were of unknown phenotype but otherwise used the CASH model. In our largest linked family (K2082), the LOD score for Mfd188 was reduced from 6.07 to 4.51. Interestingly, for K2035 the LOD score for Mfd188 was reduced, from 0.94 to 0.04, indicating that most of the positive evidence for linkage in this family (which we believe to contain a Mfd188 recombinant) came from the unaffected individuals, under the CASH model. Among the apparently unlinked families, K2080 had large negative LOD scores for all three markers under the AO model but a LOD of 0.76 for Mfd188, while the LOD scores ignoring information coming from unaffected individuals are 0.30, 0.39, and −2.14 for 42D6, Mfd188, and Mfd15, respectively. Thus evidence for linkage is decreased for two of the markers and is slightly increased for the third.

To further illustrate the consequences of the CASH model, consider the small hypothetical pedigree shown in figure 1. This family has a mixed age-at-onset distribution, with ages at diagnosis ranging from 32 to 62 years. We have calculated the genotypic probabilities under the CASH model as implemented in the LINKAGE package (Lathrop et al. 1985), for individuals 4 and 10 in this family, by assuming various ages and ages at diagnosis, respectively. When the age at diagnosis of individual 10 was varied, the age of individual 4 was set at 79 years; when individual 4 was considered, the age at diagnosis of individual 10 was fixed at 51 years. These probabilities are calculated solely on the basis of phenotypes and pedigree structure; no linkage information is incorporated. It appears counterintuitive that individual 10, who is assumed to have an age at diagnosis of 51 years and an affected mother and two early-onset affected cousins, has only a slightly better than 50% chance of carrying a breast cancer susceptibility allele. Note that, if individual 4 were unaffected at 69 years, she would still have a 17% chance of carrying a breast cancer gene. This demonstrates (*a*) why it is diffi-

GeneDX 1035, pg. 262

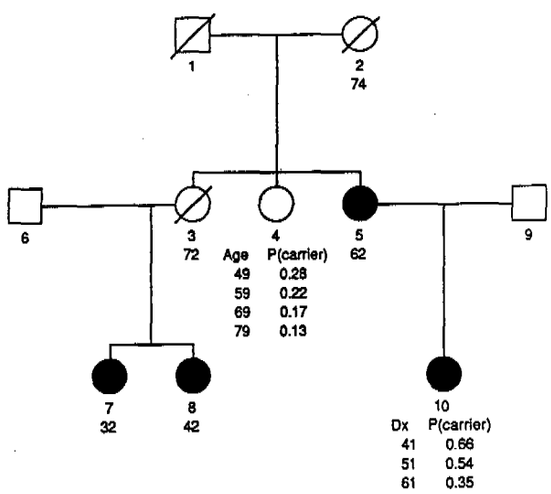


Figure 1 Hypothetical pedigree with familial breast cancer. Beneath each pedigree symbol is the individual number and current age (if unaffected and living), age at death (if deceased), or age at diagnosis (if affected). For individuals 4 and 10 in the pedigree, risks of carrying a breast cancer susceptibility allele under the CASH model are given, for various ages or ages at diagnosis.

cult to exclude linkage in families with either late or mixed age at onset under the CASH model and (*b*) why unaffected individuals contribute little linkage information under this model.

There are at least three distinct susceptibility loci responsible for inherited breast cancer: the p53 locus on 17p, the BRCA1 locus on 17q, and at least one locus responsible for the unlinked residual. It is unlikely that these loci would all have the same phenotypic effects. As the CASH data set represents, in some sense, an average of all of these loci, it is not surprising that the model may at times appear inappropriate for the data in a specific pedigree. We believe that the identification of large extended pedigrees in which linkage can be clearly established and in which key recombinant individuals can be identified will be of great value in refining the localization of the BRCA1 locus, a step which we hope will rapidly lead to its cloning and isolation. The analysis of the pedigrees reported in this paper, in which we have localized the BRCA1 locus between THRA1 and Mfd188, represents a first step in this direction. Once the susceptibility gene is cloned and the mutations are characterized, we will be able to examine more accurately both the phenotypic effects of the 17q-linked breast cancer locus and the interaction of this gene with other risk factors.

Acknowledgments

This work was supported by grants CA-48711, CN-05222, and RR-00064 from the National Institutes of Health.

References

Anderson DE (1972) A genetic study of human breast cancer. J Natl Cancer Inst 48:1029–1034

Claus EB, Risch N, Thompson WD (1991) Genetic analysis of breast cancer in the Cancer and Steroid Hormone Study. Am J Hum Genet 48:232–242

Easton DF, Bishop DT, Ford D, Crockford GP, the Breast Cancer Linkage Consortium (1993) Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. Am J Hum Genet 52:678–701

Fain PR (1992) Third International Workshop on Human Chromosome 17. Cytogenet Cell Genet 60:178–181

Futreal PA, Barrett JC, Wiseman RW (1992) Dinucleotide repeat polymorphism in the THRA1 gene. Hum Mol Genet 1:66

Hall JM, Friedman L, Guenther C, Lee MK, Weber JL, Black DM, King M-C (1992) Closing in on a breast cancer gene on chromosome 17q. Am J Hum Genet 50:1235–1242

Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King M-C (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250:1684–1689

Lathrop GM, Lalouel JM, Julier C, Ott J (1985) Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Am J Hum Genet 37:482–498

Narod SA, Feunteun J, Lynch HT, Watson P, Conway T, Lynch J, Lenoir GM (1991) Familial breast-ovarian cancer locus on chromosome 17q12-q23. Lancet 338:82–83

Oliphant AR, Wright EC, Swensen J, Gruis NA, Goldgar DE, Skolnick MH (1991) Dinucleotide repeat polymorphism at the D17S514 locus. Nucleic Acids Res 19:4794

Skolnick M (1980) The Utah genealogical data base: a resource for genetic epidemiology. In: Banbury rep no 4: Cancer incidence in defined populations. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 285–297

Skolnick MH, Cannon-Albright LA, Ward JH, Marshall CJ, Schumann GB, Hogle H, McWhorter WP, et al (1990) Inheritance of proliferative breast disease in breast cancer kindreds. Science 250:1715–1720

Weber JL, Kwitek AE, May PE, Wallace MR, Collins FS, Ledbetter DH (1990) Dinucleotide repeat polymorphisms at the D17S250 and D17S261 loci. Nucleic Acids Res 18:4640

GeneDX 1035, pg. 263

JOURNAL
OF THE NATIONAL CANCER INSTITUTE

ISSN 0027-8874
Volume 86
Number 3
Pages 161–248

February 2, 1994



Samuel Broder
Director, National Cancer Institute

Susan Molloy Hubbard
Director, International Cancer Information Center

Julianne Chappell
Chief, Scientific Publications Branch

EDITORIAL BOARD

Daniel C. Ihde
Editor-in-Chief

Susan G. Arbuck
Frank M. Balis
William J. Blot
Peter M. Blumberg
John D. Boice
Bruce A. Chabner
Richard L. Cysyk
Susan S. Ellenberg
Michael A. Friedman
Frank J. Gonzalez

Eric J. Seifter
Book Review Editor

Michael M. Gottesman
Peter Greenwald
Susan M. Hubbard
Hynda K. Kleinman
Barnett S. Kramer
Theodore S. Lawrence
Lance A. Liotta
Marc E. Lippman
Dan L. Longo
Douglas R. Lowy
Susan G. Nayfield
David L. Nelson
Kenneth Olden

Joost J. Oppenheim
John R. Ortaldo
Harvey I. Pass
Oliver W. Press
Alan S. Rabson
Robert H. Shoemaker
Richard M. Simon
Maryalice Stetler-Stevenson
Margaret A. Tucker
J. Paul Van Nevel
Noel S. Weiss

EDITORIAL ADVISORY BOARD

James O. Armitage
Bruce C. Baguley
Laurence H. Baker
William T. Beck
Clara D. Bloomfield
Benjamin Bonavida
George J. Bosl
Barry Brown
C. Norman Coleman
O. Michael Colvin
Thomas H. Corbett
Pelayo Correa
Stephen P. Creekmore
Johanna T. Dwyer
Merrill J. Egorin
Soldano Ferrone
Isaiah J. Fidler
Richard I. Fisher
David FitzGerald
Øystein Fodstad
Antonio Fojo
Rebecca Gelman
Lori J. Goldstein
Harvey M. Golomb
Eliezer Gorelik
Jean L. Grem
Arnold H. Greenberg
Maureen M. Henderson
Gloria H. Heppner
Ronald B. Herberman
E. Carmack Holmes
Waun Ki Hong
William J. Hoskins
Alan N. Houghton
James N. Ingle
David H. Johnson
Randall K. Johnson
V. Craig Jordan
John S. Kovach
Margaret L. Kripke
Mark G. Kris
Donald W. Kufe
Bernard Levin
Brian R. Leyland-Jones
Allen S. Lichter
Guy McClung
Frank L. Meyskens
Anthony B. Miller
Malcolm S. Mitchell
James J. Mulé
C. Kent Osborne
John J. O'Shea
David M. Ota
David F. Paulson
Henry C. Pitot
Igor B. Roninson
Edward A. Sausville
Thomas J. Sayers
David Schottenfeld
Herman A. J. Schut
Richard K. Severson
William R. Shapiro
Roy E. Shore
Paul M. Sondel
Patricia S. Steeg
Herman D. Suit
Sandra M. Swain
Mario Szmol
Raymond Tartle
Ian Tannock
Joel E. Tepper
J. Tate Thigpen
Peter R. Twentyman
Larry M. Weisenthal

GeneDX 1035, pg. 264