

The detailed characterisation of a 400 kb cosmid walk in the *BRCA1* region: identification and localisation of 10 genes including a dual-specificity phosphatase

Karen A. Jones*, Donald M. Black*, Melissa A. Brown, Beatrice L. Griffiths, Hans M. Nicolai, Julie A. Chambers, Marisa Bonjardin, Chun-Fang Xu, Marie Boyd², Robert McFarlane², Bernhard Korn³, Annemarie Poustka³, Michael A. North¹, Leo Schalkwyk¹, Hans Lehrach¹ and Ellen Solomon

Somatic Cell Genetics Laboratory and ¹Genome Analysis Laboratory, Imperial Cancer Research Fund, PO Box 124, Lincoln's Inn Fields, London WC2A 3PX, ²Beatson Institute for Cancer Research, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1BD, UK and ³Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, 69120 Heidelberg, Germany

Received September 13, 1994, Revised and Accepted September 22, 1994

We have produced a detailed physical and transcriptional map of a 400 kb region within the narrowest flanking markers known to contain the hereditary breast and ovarian susceptibility gene, *BRCA1*. The approach described here has avoided the problems of chimaerism, instability and rearrangements commonly observed in yeast artificial chromosomes by converting the YAC clones into ordered chromosome 17-specific cosmid contigs and joining these contigs by cosmid end-walking. A detailed long-range restriction map provided a framework for the cosmid contig assembly and further refines existing physical mapping data. We have used a combined approach towards the isolation of the genes housed within these cosmids. This has resulted in the isolation and precise localisation of eight novel genes, including a novel G protein and an endogenous retrovirus related to the HERV-K family, and the previously described dual-specificity VHR phosphatase and *MOX1* homeobox genes.

INTRODUCTION

Breast cancer is predominantly a sporadic disease affecting about one in ten women in the Western world. One hereditary form of the disease, accounting for a few percent of all cases and characterised by early age of onset, has been linked to a susceptibility gene termed *BRCA1*, on chromosome 17q21 (1). This gene predisposes to both breast and ovarian cancer and it appears that almost all breast cancer families with ovarian cancer cases are linked to *BRCA1* (2).

Since the function of *BRCA1* is unknown, efforts to clone it have relied upon a positional cloning strategy. Extensive characterisations of meiotic breakpoints in breast cancer families have succeeded in narrowing the region to 1.0–1.5 Mb between the markers D17S776 and D17S78 (3–8). Additional fine mapping has been limited by the number of informative meioses available and complicated by the extensive genetic heterogeneity of the disease (9–13). Both breast and ovarian tumours have shown extensive loss of heterozygosity in the region known to

contain *BRCA1*, indicating the presence of one or more tumour suppressor genes (14–21). However, searches for small regions of deletion and other positional clues such as translocations in sporadic or familial tumours have not been successful.

With the aim of identifying the *BRCA1* gene within these defined limits, several tools have been developed, including high density genetic maps (22,23); chromosome 17 hybrid mapping panels (24–26) and fluorescence *in situ* hybridisation (FISH) data, for the physical localisation of polymorphic markers and genes spanning the region (27). These have enabled the assembly of incomplete yeast artificial chromosome (YAC) contigs for the purpose of isolation and analysis of the genes housed within the *BRCA1* region (28). Unfortunately, problems of chimaerism, instability and rearrangements have made YACs a far from ideal substrate for isolating specific genes and constructing comprehensive transcript maps. This work describes identification of a substantial hole in our YAC coverage which has been filled by cosmid end-walking. With the conversion of our YACs into overlapping cosmids, a complete ordered cosmid contig spanning over 400 kb from the gene *IAI.3B* to the polymorphic marker D17S78 has been obtained, together with a detailed long-range physical map constructed by pulsed-field gel electrophoresis analysis. This has enabled the isolation and precise localisation of 10 genes in the region and provides further insight into the characteristics of the chromosomal region surrounding *BRCA1*.

RESULTS

YAC isolation and analysis

In order to clone genomic DNA for detailed characterisation of the *BRCA1* region, large insert YAC clones were isolated with probes and PCR primers corresponding to both known genes and anonymous markers. Several clones were identified but in the region flanking the *RNU2* locus, problems of inter-chromosomal chimaerism, rearrangement and instability prevented the generation of a complete YAC contig (Table 1). A sizeable deletion was noted in one YAC, 17-5-G10 which contained only the markers D17S859 and D17S855 but not *IAI.3B*, *RNU2* and other cosmid markers known to reside between them. In addition a 400 kb YAC, 26D6 which contained the markers D17S858 and D17S859, was unstable and became shortened by internal

*To whom correspondence should be addressed

[†]Present address: Beatson Institute for Cancer Research, Glasgow, UK

Table 1. Results of YAC analysis in the *1A1.3B*–*D17S78* region

YAC name	Library ^a	Probe	Size kb	Features ^b	Markers						
					S855	<i>1A1</i>	<i>RNU2</i>	S858	S859	<i>PPY</i>	p131
106G4	ICRF	<i>1A1</i>	700	unstable		+					
12H4	ICRF	<i>1A1</i>	500	chimaeric		+	+				
A167E6	Wash	<i>RNU2</i>	200	chimaeric		+	+				
176B12	CEPH17	<i>RNU2</i>	800	nd			+				
B196	Wash	<i>RNU2</i>	150	nd			+				
174A12	CEPH17	YAC 2H5 ^c	700	chimaeric				+	+		
175G10	CEPH17	YAC 2H5	nd	chimaeric and deleted	+				+		
26D6	CEPH	<i>PPY</i> /p131	250–400	unstable				(+)	(+)	+	+
B26OE11	Wash	<i>PPY</i>	<100	nd						+	
172H5	CEPH17	YAC 5H2	450	nd						+	+
175H2	CEPH17	<i>PPY</i>	350	nd						+	+
109E5	ICRF	<i>PPY</i> /p131	nd	nd						+	+

Each YAC, indicated in the first column has been tested for each marker in the top row. Presence of a marker is indicated as +. YAC 26D6 lost markers D17S858 and D17S859 on deletion, as indicated by brackets.

^aICRF, ICRF YAC library; Wash, Washington School of Medicine library; CEPH17, CEPH chromosome 17 selected mega YAC library.

^bChimaerism was determined by *in situ* analysis of YAC *Alu* PCR probes.

^c*Alu* PCR probe from YAC 172H5.

deletion to 250 kb when subsequent preparations were made. None of the YACs isolated shared the *RNU2* marker with any of the markers D17S858, D17S859, *PPY* or D17S78 which indicated a gap in YAC coverage in this region.

Cosmid isolation and initial contig construction

To avoid analysis of other genomic regions present in the chimaeric YAC clones, and to isolate non-deleted copies of the DNA that is unstable in the YACs, a gridded flow-sorted chromosome 17 cosmid library, generated at the ICRF (29) was screened with the available YACs. The resulting pools of cosmids were arranged into contigs on the basis of the results of PCR analyses and hybridisations of repeat free cosmid fragments to other members of the cosmid pools (Fig. 1). These results were confirmed by *Pst*I restriction fragment size comparisons and *Alu*-PCR fingerprinting (data not shown).

The 12H4 YAC identified over 200 strong positive clones. Most of these clones were due to the repeated array of U2 RNA genes at this locus(30), as revealed by hybridisation of a duplicate library filter with *RNU2*. The 12H4-positive cosmid clones which were positive for *1A1.3B* were revealed by an additional hybridisation with a *1A1.3B* gene probe. Cosmid B0176 was positive for both *1A1.3B* and *RNU2*, indicating that these genes are less than 40 kb apart, based on the average length of a cosmid insert. A weakly hybridising *RNU2* positive clone, E1132 was found to be negative for *1A1.3B* and did not overlap with cosmid B0176. Hybridisation of three *Eco*RI repeat-free fragments from E1132 back to the cosmid library filters isolated two further cosmids, C1198 and A04131, both of which failed to hybridise to the *RNU2* probe. The cosmids E1132, C1198 and A04131 therefore represent a walk away from the *RNU2* locus in the opposite direction to *1A1.3B* (Fig. 1). The location of cosmid C1198 adjacent to the *RNU2* locus was confirmed by pulsed field gel electrophoresis (PFGE) analysis (data not shown).

Positive cosmid clones identified by screening with YAC 26D6 prior to deletion, were analysed by PCR for the presence of markers D17S858, D17S859, D17S78 and the *PPY* gene. Cosmid A01183 was positive for both D17S78 and *PPY* which confirms the finding that these markers are within 45 kb of one another (31) (Fig. 1). Cosmids D11181 and G08151 were both positive for D17S78 only and represent the most distally extending

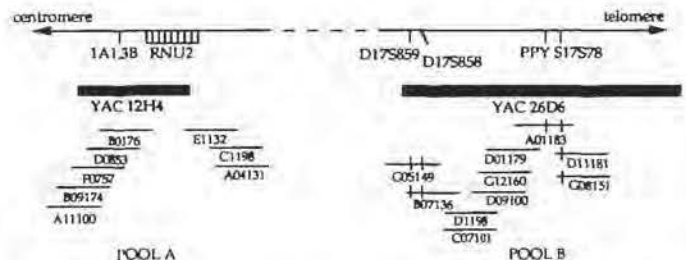


Figure 1. Ordering of cosmid pools generated by chromosome 17 cosmid library screenings using YACs 12H4 and 26D6. Cosmids C1198/A04131 and G05149/B07136 extend the furthest into the gap between the two cosmid pools.

cosmids of the contig. Cosmids G05149 and B07136 were positive for both D17S858 and D17S859 and were shown to link up to the *PPY* and D17S78 cosmids by hybridisation experiments. Cosmid D01179 hybridised strongly with cosmid G12160, D09100, A01183 and itself, but more weakly with cosmids D1198 and C07101. The cosmids D1198 and C07101 in turn hybridised to cosmid B07136. The overlapping cosmid contig of pool B suggest the markers D17S858/D17S859 are roughly 60 kb away from *PPY*/D17S78.

Pulsed field gel electrophoresis analysis of the region *1A1.3B* to *PPY*/D17S78

Analysis of the rare-cutter restriction fragments hybridising to single-copy probes from each of the cosmid pools orientated the cosmid contigs and identified the size of the gap between them. A PFGE Southern filter containing high molecular weight genomic DNA from an immortalised B cell line digested with a number of rare cutter restriction endonucleases in single and double combinations was sequentially hybridised with single-copy probes from the region (see Materials and Methods).

Comparisons of the autoradiographs revealed that *1A1.3B*, *RNU2* and 12E/Bd, a fragment of the D17S858/D17S859-positive cosmid, G05149, reside on the same 750 kb *Not*I fragment and that *1A1.3B* and *RNU2* share the same 380 kb *Mlu*I fragment, but different *Nru*I fragments (Fig. 2). The *Nru*I fragment which contains *RNU2* also hybridises to 12E/Bd, giving the order

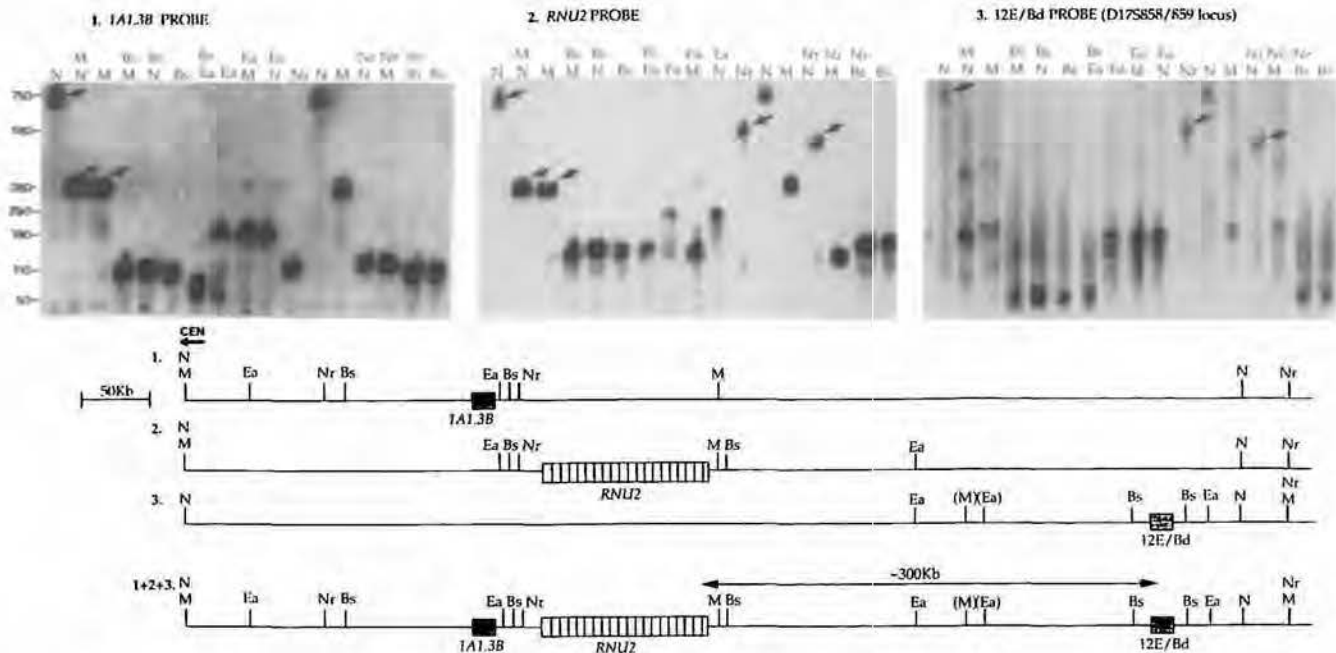


Figure 2. Construction of a physical map from PFGE data using probes *RNU2*, *1A1.3B* and *12E/B*, a fragment of the D17S858/D17S859-positive cosmid, G05149 (see Materials and Methods for probe details). Arrowed bands indicate the restriction fragments which hybridise to more than one probe. Restriction maps were constructed using the information provided by each hybridisation experiment. Superimposing the maps (1+2+3) gives a rough estimation of the distance between *12E/B* and *RNU2* as 300 kb. N, *NorI*; M, *MluI*; Bs, *BssHII*; Ea, *EagI*; Nr, *NruI*; (M) and (Ea), restriction sites showing partial methylation.

1A1.3B–*RNU2*–*12E/Bd*. The probes *PPY* and *p131* hybridised to a smaller 550 kb *NorI* fragment, which indicates that cosmid G05149 is the most proximally extending cosmid of pool B (data not shown). Therefore, combining the cosmid contig analysis and the pulsed-field electrophoresis analysis, the order of the markers is cen–*1A1.3B*–*RNU2*–D17S858/9–*PPY*–D17S78–tel.

The detailed restriction map integrating all of the restriction fragment sizes indicated that the gap between the most distally extending cosmid of pool A, cosmid C1198 and the most proximally extending cosmid of pool B, G05149 was approximately 300 kb in the B cell line tested (Fig. 2).

Generation of a complete cosmid walk between *RNU2* and D17S858

In order to complete the genomic coverage of this region, the cosmids extending the furthest into the gap were used to initiate cosmid walks towards one another. The walking strategy employed the dual opposed SP6 and T7 promoters flanking the cloning site in the cosmid vector, Lawrist 4 (D.Nizetic, unpublished data) to create probes from each end of the cosmid. Neighbouring cosmids were isolated when each end-specific probe was hybridised in turn to the cosmid library, as shown (Fig. 3). The walks proceeded proximally from cosmid G05149 and distally from cosmid C1198 until they overlapped at cosmids C02179 and G124 (Fig. 4B). The distance covered in the cosmid walk correlates roughly with that suggested by the pulsed field electrophoresis analysis above. The ordering of the cosmids throughout the region *1A1.3B* to D17S78 was confirmed by FISH analysis (32). This gave the order as C1198–G05149–D11181 and C1198–G124–G05149 and confirms that the cosmids are contiguous (Fig. 5 and data not shown). In addition, FISH

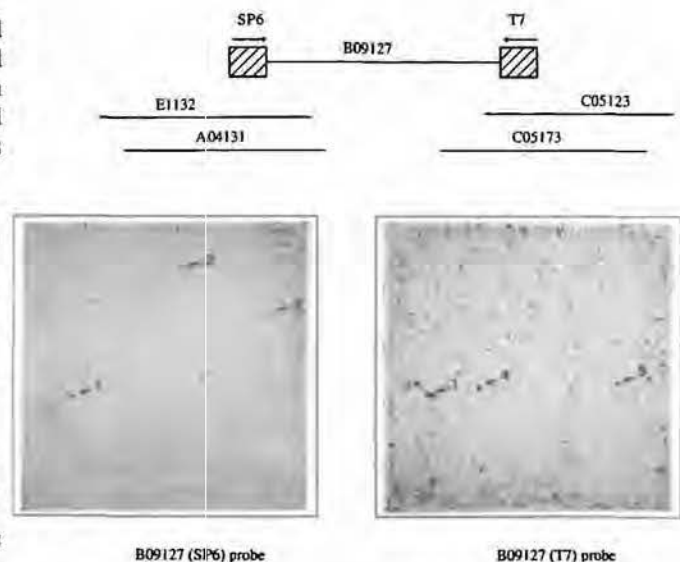


Figure 3. Hybridisation of riboprobes from each end of cosmid B09127 to duplicate filters containing gridded cosmids from the chromosome 17 cosmid library. The T7 riboprobe extends the furthest into the gap as it has hybridised with previously unidentified cosmids, C05123 (5) and C05173 (4), whilst the SP6 riboprobe hybridised to previous cosmids in the walk, E1132 (2) and A04131 (3). Both riboprobes hybridised to the cosmid B09127 (1) from where they originated, serving as an internal positive control.

analysis on chromatin released from nuclei (33) confirmed the overlap of the two cosmid walks at cosmids G124 and C02179 (data not shown).

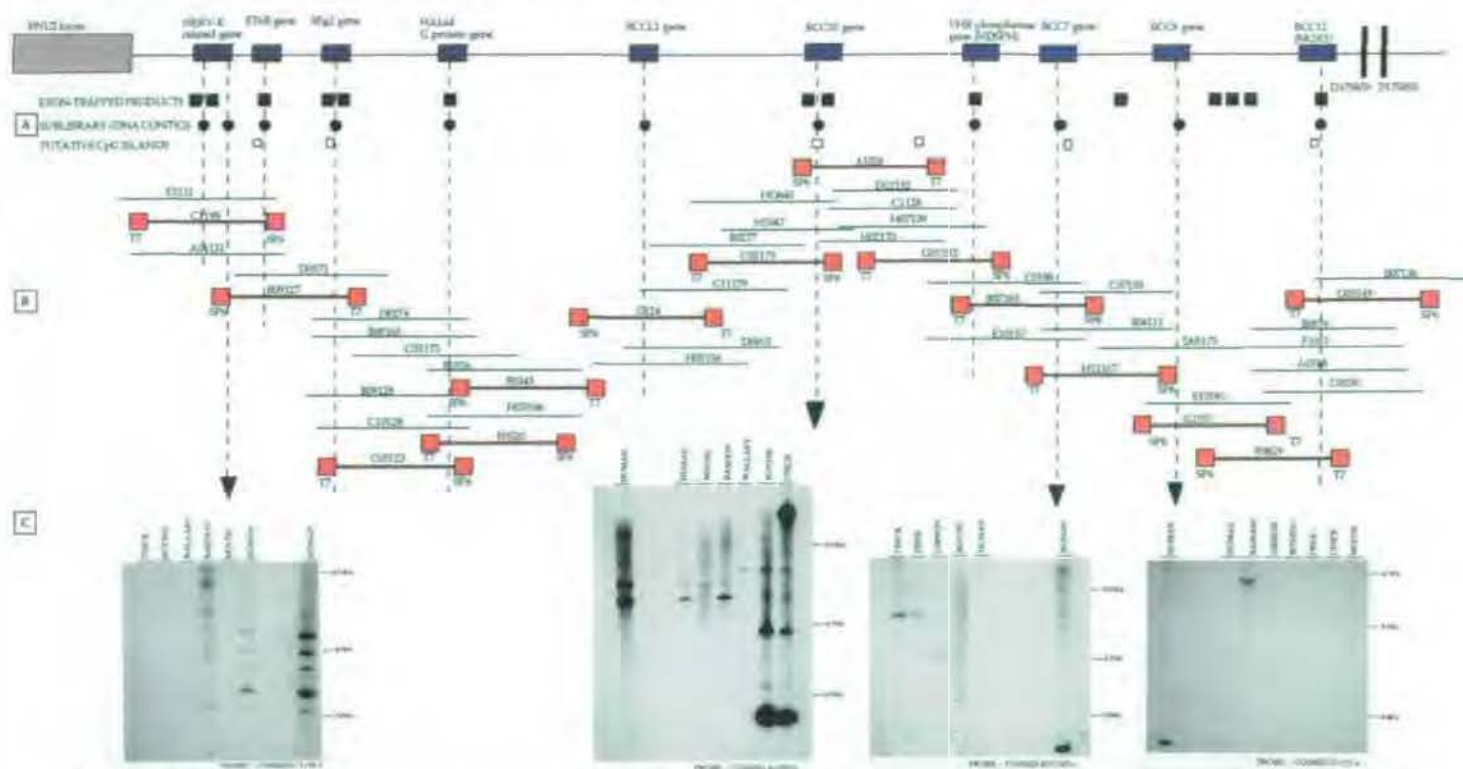


Figure 4. Detailed cosmid clone and transcript map of the region between *RNU2* and *D17S78*. (A) Localisation of 10 genes (blue markers) on the cosmid contig is indicated by dashed vertical lines. The localisation of exon-trapped products (filled squares) and sublibrary cDNA contigs (filled circles) is also given. Open squares indicate the position of *EcoRI* fragments in the cosmid contig which contain putative CpG islands. (B) Cosmid walk between *RNU2* and *D17S78*. Cosmids in red constitute the minimal overlapping set used in the creation of cDNA sublibraries and exon-trapping. The orientation of the SP6 and T7 ends of these cosmids within the walk is shown. (C) Zoo blot results from conserved *EcoRI* fragments in the region which have resulted in the isolation of genes.

Isolation and localisation of genes in the cosmid contig

A minimal set of overlapping cosmids spanning the region *RNU2*–*D17S858/D17S859* (see Fig. 4B) was used as the substrate for the isolation of cDNA clones using a number of techniques. They include direct cDNA selection (34,35), surveying for evolutionary conservation (36), exon trapping (37) and localisation of CpG islands (38,39). Clusters of cDNA clones were localised throughout the cosmid contig and Northern analyses confirmed which of them corresponded to separate gene loci (Fig. 6). The genes isolated include the dual specificity VHR phosphatase (40), the newly identified *MOXI* gene (41), a novel GTP-binding protein (G protein) *HAL64* (Black *et al.*, in preparation), a locus for an endogenous retrovirus and six novel genes with no database sequence homologies to date.

Direct cDNA selection. Overlapping cDNA contigs were identified at the gene loci indicated on the transcript map (Fig. 4A). Members of the cDNA contigs were used as probes to isolate full length cDNA clones from random plated cDNA libraries and sequencing of these clones has identified homologies to other known genes. Nearly all the genes isolated in this region were represented in the subtracted cDNA libraries with the exception of the *BCC8* gene, discovered by a search for evolutionarily conserved sequences and four exon-trapped products.

Evolutionary conservation. Expression-independent methods were also employed to identify genomic fragments corresponding to cDNAs within the *BRCA1* region. All of the *EcoRI* fragments from the minimal set of overlapping cosmids were hybridised

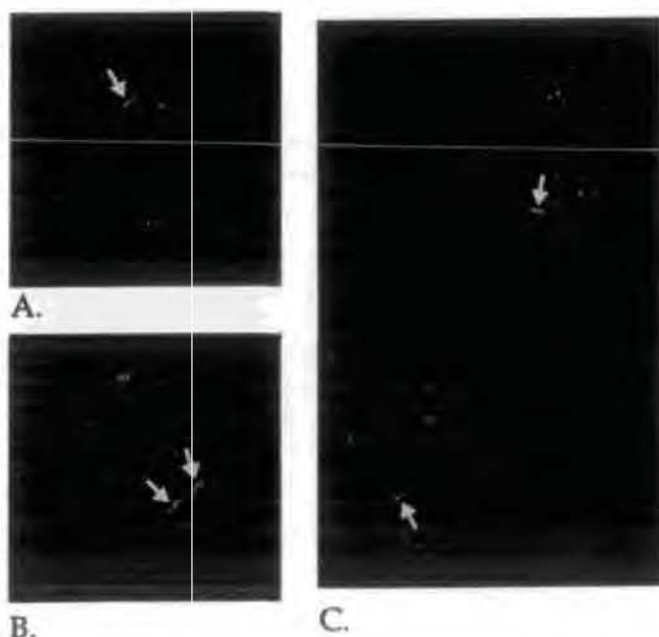


Figure 5. FISH analysis of the cosmid contig spanning the region from *RNU2* to *D17S78*. (A) Cosmid C1198 (red), D11181 (green) and G05159 (red) give the order red-red-green. (B) Cosmid C1198 (red), D11181 (red) and G05149 (green) give the order red-green-red. (C) Cosmid C1198 (red), G124 (green) and G05149 (green) give the order red-green-green. These results confirm that the walk is contiguous and indicate that the cosmid order given by the walk is correct.

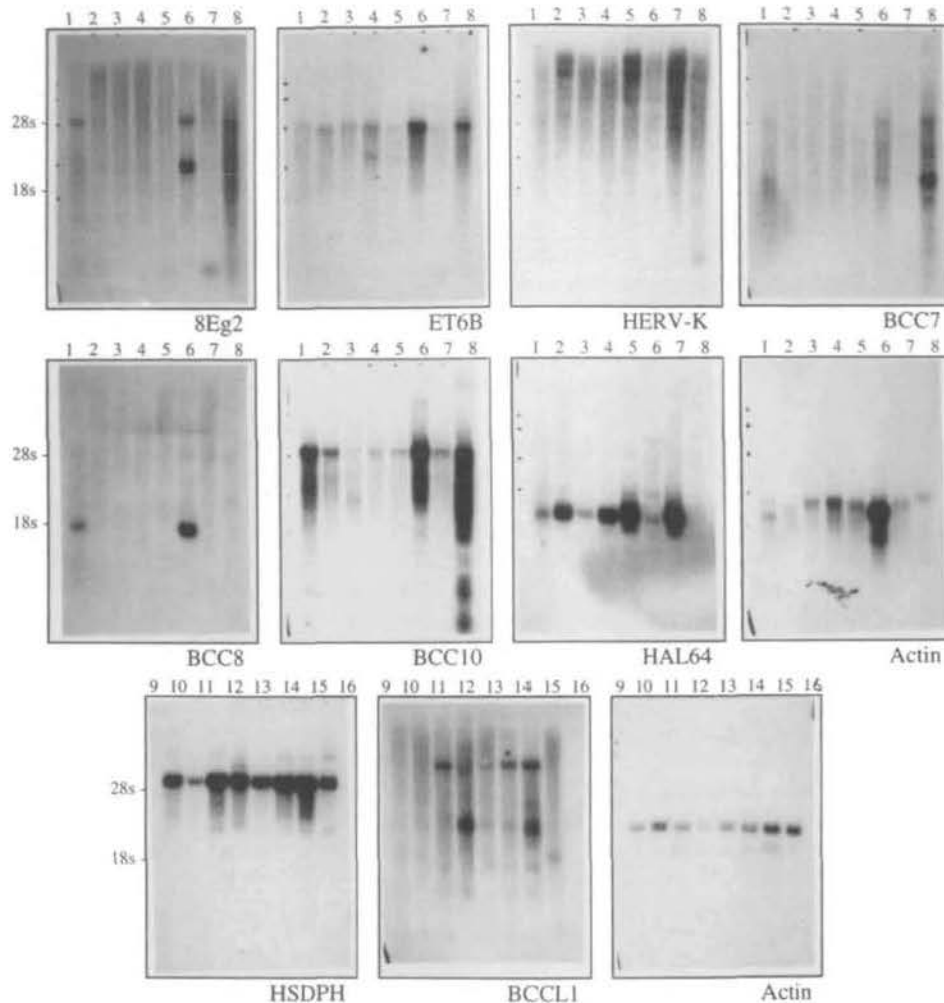


Figure 6. Multiple tissue Northern analyses of the 10 gene loci identified in the region between *RNU2* and *D17S78*. Northern filters (Clontech) contain heart (1), brain (2), placenta (3), lung (4), liver (5), skeletal muscle (6), kidney (7), pancreas (8), spleen (9), thymus (10), prostate (11), testis (12), ovary (13), small intestine (14), colon (15) and peripheral blood leukocyte (16).

to zoo blots containing digested genomic DNA from a variety of species, in a systematic survey for evolutionarily conserved sequences. Six fragments, C1198f, A1028h, A1028c, B07165c, B07165d and G1151a appeared to be clearly conserved (Fig. 4C), while the remaining fragments were too repetitive to detect single bands by Southern hybridisation. Sets of overlapping cDNA clones were identified when each of the conserved fragments were hybridised to the relevant sublibrary (Fig. 4A). These cDNA clones were sequenced and used as probes on large insert cDNA libraries to isolate full length cDNA clones.

Fragment C1198f identified the *env* and *pol* genes of the HERV-K endogenous retrovirus locus, whilst A1028h identified gene BCC10, a ubiquitously-expressed novel gene with no sequence database homologies to date. Fragments c and d of cosmid B07165 both identified another novel gene, BCC7 which is expressed strongly in pancreas and skeletal muscle with message sizes of 1.8 and 4.8 kb (Fig. 6). Fragment G1151a did not hybridise to any clones in the sublibrary suggesting it was not well represented in the pool of cDNA libraries used in the direct selection process. However, the convincing Northern result with this fragment gives strong indication that it contains a gene,

termed *BCC8*, with high tissue specificity (Fig. 6). A cDNA clone isolated from a retina cDNA library with fragment G1151a has shown no sequence database homologies to date. Fragment A1028c did not hybridise with any of the mRNA samples on both types of Northern filters tested, suggesting that either it contains a gene of high tissue specificity not represented by these Northern filters, or it is a fragment of genomic DNA with no function. It has therefore been omitted from the transcript map, but the possibility remains that this is a locus of an additional gene.

Exon-trapping. Exon-trapping was performed with the same sets of cosmids used to create the above-mentioned subtracted cDNA libraries. A total of 14 products were obtained, 10 of which hybridised to clones in the subtracted cDNA libraries. The loci, HERV-K LTR, *ET6B*, *8Eg2*, *HAL64*, *BCC10*, VHR phosphatase and *MOX1* were all identified by exon-trapping. Four products failed to isolate any cDNAs from the subtracted libraries suggesting that they represent genes not selected in the sublibraries. Cloning, sequencing and hybridisation to

conventional cDNA libraries is now in progress to further characterise these products.

Identification of CpG islands. In addition, the minimal cosmid set was screened by restriction enzyme analysis for the presence of CpG islands (38). A putative CG island-containing fragment was defined as a fragment which was cut by both the rare-cutters *SacII* and *BssHII*. All such fragments appear to reside in close proximity to genes isolated in the region (Fig. 4A). Further characterisations of these fragments is underway to confirm they are the 5' regions of these genes.

DISCUSSION

In this paper, we describe a detailed physical and transcriptional analysis of a 400 kb region residing within the narrowest limits known to contain *BRCA1*. This work has identified and located eight novel genes, including a novel G protein, and an endogenous retrovirus related to the HERV-K family, and the previously described dual-specificity VHR phosphatase (40) and *MOX1* homeobox genes (41). All reside within a chromosomal region shown to be frequently deleted in sporadic and familial breast and ovarian tumours and may therefore play a role in breast and ovarian tumourigenesis and progression.

The long-range restriction map described here combined with analyses of YAC cosmid pools gives the order of markers in the *BRCA1* region as *cen-1A1.3B-RNU2-D17S858/D17S859-PPY-D17S78-tel*. This further refines the physical mapping data provided by multicolour fluorescence *in situ* hybridisation with *Alu*-PCR-amplified YAC clone DNA (27). A rough analysis of the restriction fragment data provided by PFGE Southern blot hybridisation with single copy probes in this region suggested that the D17S858/D17S859 locus was approximately 300 kb distal to *RNU2* in the B cell line tested. Although restriction site polymorphisms and differences in methylation could exist between the B cell line used in the PFGE analysis here, and the cell lines used to construct the cosmid and YAC libraries, the distance covered by the cosmid contig correlated roughly with estimates provided by the restriction map. The cosmid contig was confirmed as contiguous by FISH analysis which also agreed the order of the cosmids within the walk, providing further evidence to support the physical mapping data.

The assembly of an ordered cosmid contig has enabled the precise localisation of gene loci in the region. A combination of different gene isolation approaches was employed to make the transcript map as complete as possible. The method of direct cDNA selection (34,35) proved the most efficient and straightforward of the cDNA isolation techniques and has provided cDNA clones at every gene locus on our transcript map with the exception of *BCC8* which appears to be a gene with high tissue specificity (Fig. 6). The use of a pool of cDNA libraries has probably increased the likelihood of isolating most of the genes in this region. In addition, limiting the complexity of the input genomic DNA to a maximum of five cosmids may have also increased the efficiency of the hybrid selection. We have combined this technique with expression-independent gene isolation methods of cross species homology searches, exon-trapping and CpG island hunting. Whilst each of these methods have their limitations (discussed in ref. 42), we believe that the combination of all of them will result in a complete transcript map of the region.

The identification of a complex integration site for the human endogenous retrovirus HERV-K at one end of the transcript map may explain the cloning difficulties experienced in this region; a phenomenon also observed at the G6PD locus of chromosome Xq28 which also contains a HERV-K integration site (43). Many of the cDNA clones isolated from this locus from random plated cDNA libraries and the gridded sublibraries failed to localise to chromosome 17. These clones may represent homologous retroviral inserts from elsewhere in the genome. Sequence analysis of the *BRCA1* region-specific HERV-K cDNA clones, and of subcloned genomic DNA from cosmid C1198, suggest that this locus contains *pol* and *env* genes and a long terminal repeat (LTR). Completion of the sequencing at this locus will indicate whether this HERV-K is transcribed and is capable of encoding proteins. Retrovirus-related sequences exhibit a number of features giving them a potential for involvement in carcinogenesis (reviewed in ref. 44). Proviral integration can cause activation of adjacent cellular protooncogenes by promoter insertion, as observed by induction of lymphomas by the viral promoter-induced activation of the cellular gene, *c-myc* (45). In addition, stimulation of HERV-K expression in the breast carcinoma cell line, T47D by female steroid hormones has been observed. It has been suggested that this enhanced expression may be involved in the aetiology of certain human breast cancers, similar to the MMTV-induction of tumours in mice (46). Hence, the HERV-K locus and its neighbouring genes inside the *BRCA1* region merit further investigation.

Another interesting gene in this region is a new ADP-ribosylation factor, *HAL64*, related to the highly conserved family of GTP-binding proteins with homology to both the *ras* oncogene superfamily and the heterotrimeric G protein α subunits (47). *HAL64* appears most closely related to a processed chicken pseudogene, *CPS1* which has been implicated as an oncogene due to its similarity to the *c-H-ras* oncogene and possible involvement in early embryonic development (48). This gene may represent a new oncogene and further studies are required to see whether it is involved in tumour development.

Another biologically plausible candidate gene, the dual-specificity VHL-related VHR phosphatase was also found to reside within this transcript map. A member of the protein-tyrosine-phosphatases (PTPases), this gene has potential roles in cell signalling, cell growth and proliferation and oncogenic transformation (40). A possible role in the reversal of the effects of protein tyrosine kinases has implicated the PTPases as candidate tumour suppressor genes (49). However, investigations of other PTPases have failed to find any involvement in lung and renal cancers (50,51). We have searched samples of genomic DNA and cDNA in our breast and ovarian cancer families and in breast and ovarian tumours for oncogenic mutations, but have found no alterations to date. However, it may be interesting to see whether any loss of function of this gene within breast and ovarian tumours has any effect on tumour progression.

The remaining genes in this transcript map require further study into possible roles in breast and ovarian cancer. Since they all reside within a region commonly deleted in breast and ovarian tumours, the implications for tumourigenesis and development must be investigated.

MATERIALS AND METHODS

Libraries

YAC clones were isolated from libraries generated at the Imperial Cancer Research Fund (ICRF), Centre d'Etude du Polymorphisme Humain (CEPH) and Washington

School of Medicine (52–54). Screening was carried out by hybridisation to gridded YAC filters and PCR analysis of YAC pools (55). *Alu* PCR products (56) and YAC inserts (57) from the positive YAC clones were used to screen duplicate filters of the chromosome 17 cosmid library containing 20736 clones per membrane, representing a 4-fold genomic coverage (29).

DNA preparation

Genomic DNA for PFGE analysis was prepared from cultured lymphoblastoid cell lines as described (58). Cosmid DNA was prepared by alkaline lysis (59).

Restriction digestion of agarose blocks

Each block was digested in 190 μ l of sterile distilled water and 15–20 units of enzyme (Northumbria Biolabs) with 25 μ l of 10 \times restriction enzyme buffer (Northumbria Biolabs). Digests were carried out for 4 h at 37°C (*Bst*HI digests at 50°C).

Southern transfer and hybridisations

All agarose gels, including pulsed-field gels, were transferred to Hybond-N+ by alkali transfer and hybridised at 65°C according to the manufacturer (Amersham LIFE SCIENCE).

DNA labelling

All probes were labelled by the random priming method using [α -³²P]dCTP (60) and radioactively-labelled probe was separated from unincorporated [α -³²P]dCTP on a Sephadex G50 column. Probes were denatured by boiling and preannealed at 65°C for 90 min with a 5000-fold excess (w/w) of sonicated human placental DNA prior to hybridisation to compete out repetitive elements.

Single copy probes for Southern analysis

Probes used include a 3.7 kb insert from a clone containing the gene *1A1.3B* (61), an *RNU2* gene probe (a gift from A. Weiner, Yale University), a PCR product from the coding region of the gene for pancreatic polypeptide (*PPY*) (62), the probe p131 from the locus D17S78 (63) and 12E/Bd, a fragment of the D17S858/D17S859-positive cosmid, G05149. Between each hybridisation, the blot was stripped and checked by autoradiography for any residual signal.

Pulsed field gel electrophoresis

All gels were made up in 0.5 \times Tris–borate–EDTA electrophoresis buffer with 1% agarose. Gels were run using parameters obtained from the Biorad CHEF DRII CHEFMAPPER programme in order to separate fragments below 1.2 Mb. All gels were run at a circulating temperature of 14°C using the Biorad CHEFII system.

Cosmid walking

Cosmid DNA (1 μ g) was digested to completion using *Rsa*I. An *in vitro* transcription was then performed using the riboprobe II core system kit (Promega) according to manufacturer's instructions. The radioactively labelled riboprobe was then hybridised directly with the cosmid library filters.

Fluorescence *in situ* hybridisation

Interphase nuclei were prepared and FISH was carried out as described previously (32). Antibody detection and signal visualisation were as described (64).

Direct cDNA hybrid selection

A minimal set of fourteen cosmids, which together span the cosmid walk, was divided into three sets of four or five cosmids for the construction of cDNA sublibraries by hybrid selection (34,35,43) using a pool of cDNA libraries from breast cancer cell line ZR75 (a gift from Dr M. Parker) and foetal brain, foetal liver and adult skeletal muscle (Stratagene). Around 500 colonies for each library were picked into microtitre plates and spotted onto filters. Each cosmid of the minimal overlapping set was hybridised in turn to its corresponding sublibrary to identify cosmid-specific cDNA clones. Selected cDNA clones were hybridised back to the sublibrary in order to organise the positive clones into sets of cDNA contigs. Members of each cDNA contig were mapped back to the cosmid contig and to chromosome 17-specific somatic cell hybrid cell lines by Southern hybridisation in order to check their location within the *BRCA1* region. Those cDNA clones not mapping to chromosome 17 or previously identified as repetitive by hybridisation of the sublibrary filters to total human DNA were discarded from further analyses.

Exon trapping

Pools of cosmids were digested to completion with either *Pst*I or a combination of *Bam*HI and *Bgl*II and cloned into pSPL3 (65). DNA from transformants was

purified using Qiagen columns (Diagen) and used to transfect COS-7 cells by electroporation. The cells were grown for 48 h and cytoplasmic total RNA was prepared. First-strand cDNA synthesis and PCR amplification were carried out as described (37). Secondary PCR and cloning of the exon-trapped products was carried out as described (66).

Zoo blot analysis

DNA was prepared from human lymphoblastoid (MOLT-4), mouse lymphoma (EL4), baboon lymphoblastoid (CRL1495), gibbon lymphoma (MLA144), bovine epithelial (MDBK), muntjack deer fibroblast (DM), wallaby kidney (CRL6559) and chick primary fibroblast cell lines. Following Southern hybridisation, filters were washed to a stringency of 1 \times SSC, 0.1%SDS for 15 min at 65°C.

DNA sequence comparison

Sequences were analysed for homologies or identities with known sequences using Intelligenetics and FASTA and TFASTA from the GCG database searching program package (67).

ACKNOWLEDGMENTS

We would like to thank Ranjit Bhogal and Günther Zehetner for help with the cosmid and YAC libraries; Pat Gorman and Denise Sheer for the FISH analysis; Kathy Howe for helpful advice and technical assistance and Anna-Maria Frischau and John Trowsdale for critical reading of the manuscript. We also thank A. Weiner for the *RNU2* probe and M. Parker for the ZR75 cDNA library. M.A.B. is supported by an EMBO postdoctoral research fellowship, and M. Bonjardim was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico of Brazil.

REFERENCES

- Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B. and King, M.-C. (1990) *Science* **250**, 1684–1689.
- Easton, D. F., Bishop, D. T., Ford, D. and Crockford, G. P. (1993) *Am. J. Hum. Genet.* **52**, 678–701.
- Bowcock, A. M., Anderson, L. A., Friedman, L. S., Black, D. M., Osborne-Lawrence, S., Rowell, S. E., Hall, J. M., Solomon, E. and King, M.-C. (1993) *Am. J. Hum. Genet.* **52**, 718–722.
- Chamberlain, J. S., Beohne, M., Frank, T. S., Kioussis, S., Xu, J., Guo, S.-W., Hauser, E. R., Norum, R. A., Helmbold, E. A., Markel, D. S., Keshavarzi, S. M., Jackson, C. E., Calzone, K., Garber, J., Collins, F. S. and Weber, B. L. (1993) *Am. J. Hum. Genet.* **52**, 792–798.
- Porter, D. E., Cohen, B. B., Wallace, M. R., Carothers, A. and Steel, C. M. (1993) *Int. J. Cancer* **53**, 188–198.
- Simard, J., Feunteun, J., Lenoir, G., Tonin, P., Normand, T., The, V. L., Vivier, A., Lasko, D., Morgan, K., Rouleau, G. A., Lynch, H., Labrie, F. and Narod, S. A. (1993) *Hum. Mol. Genet.* **2**, 1193–1199.
- Kelsell, D. P., Black, D. M., Bishop, D. T. and Spurr, N. K. (1993) *Hum. Mol. Genet.* **2**, 1823–1828.
- Goldgar, D. E., Fields, P., Lewis, C. M., Tran, T. D., Cannon-Albright, L. A., Ward, J. H., Swensen, J. and Skolnick, M. H. (1994) *J. Natl Cancer Inst.* **86**, 200–209.
- Schlessinger, D. (1990) *Trends Genet.* **6**, 254–258.
- Narod, S. A. and Amos, C. (1990) *Am. J. Hum. Genet.* **46**, 266–272.
- Zuppan, P., Hall, J. M., Lee, M. K., Ponglikitmongkol, M., King, M.-C. (1991) *Am. J. Hum. Genet.* **48**, 1065–1068.
- Skolnick, M. H. and Cannon-Albright, L. A. (1992) *Cancer* **70**, 1747–1754.
- Sobel, H., Mazoyer, S., Smith, S., Lyonnet, D., Bignon, Y., Narod, S., Ardoin, A., Biron, P., Bobin, J., Bremond, A., Brunatmetigny, M., Chatelain, P., Dargent, D., Dolbeau, D., Frenay, M., Gioanni, J., Guastalla, J., Guerin, D., Kerbrat, P., Namer, M., Peix, J., Plauchu, H., Raudrant, D., Rebattu, P., Rochet, Y., Romestaing, P., Schaefer, R., Schraub, S., Gerard, J., Robert, J., Clavel, M. and Philip, (1993) *Bull. Cancer* **80**, 121–134.
- Foulkes, W., Black, D., Solomon, E. and Trowsdale, J. (1991) *Lancet* **338**, 444–445.
- Futreal, P. A., Soderkvist, P., Marks, J. R., Iglehart, J. D., Cochran, C., Barratt, J. C. and Wiseman, R. W. (1992) *Cancer Res.* **52**, 2624–2627.
- Smith, S. A., Easton, D. F., Evans, D. G. R. and Ponder, B. A. J. (1992) *Nature Genet.* **2**, 128–131.
- Saito, H., Inazawa, J., Saito, S., Kasumi, F., Koi, S., Sagae, S., Kudo, R., Saito, J., Noda, K. and Nakamura, Y. (1993) *Cancer Res.* **53**, 3382–3385.

18. Lindblom, A., Skoog, L., Anderson, T. I., Rotstein, S., Nordenskjöld, M. and Larsson, C. (1993) *Hum. Genet.* **91**, 6–12.
19. Phillips, N., Ziegler, M., Saha, B. and Xynos, F. (1993) *Int. J. Cancer* **54**, 85–91.
20. Jacobs, I. J., Smith, S. A., Wiseman, R. W., Futreal, P. A., Harrington, T., Osborne, R. J., Leech, V., Molyneux, A., Berchuck, A., Ponder, B. A. J. and Bast, R. C. J. (1993) *Cancer Res.* **53**, 1218–1221.
21. Cornelis, R. S., Devilee, P., van Vliet, M., Kuipers-Dijkshoorn, N., Kerssenmacker, A., Bardoel, A., Khan, P. M. and Cornelisse, C. J. (1993) *Oncogene* **8**, 781–785.
22. Anderson, L. A., Friedman, L., Osborne-Lawrence, S., Lynch, E., Weiss, J., Bowcock, A. and King, M.-C. (1993) *Genomics* **17**, 618–623.
23. Albertsen, H., Plaetke, R., Ballard, L., Fujimoto, E., Connolly, J., Lawrence, E., Rodriguez, P., Robertson, M., Bradley, P., Milner, B., Fuhrman, D., Marks, A., Sargent, R., Cartwright, P., Matsunami, M. and White, R. (1994) *Am. J. Hum. Genet.* **54**, 516–525.
24. Black, D. M., Nicolai, H., Borrow, J. and Solomon, E. (1993) *Am. J. Hum. Genet.* **52**, 702–710.
25. Abel, K. J., Boehnke, M., Prahalad, M., Ho, P., Flejter, W., Watkins, M., VanderStoep, J., Chandrasekharappa, S. C., Collins, F. S., Glover, T. W. and Weber, B. L. (1993) *Genomics* **17**, 632–641.
26. O'Connell, P., Albertsen, H., Matsunami, N., Taylor, T., Hundley, J. E., Johnson-Pais, T. L., Reus, B., Lawrence, E., Ballard, L., White, R. and Leach, R. J. (1994) *Am. J. Hum. Genet.* **54**, 526–534.
27. Flejter, W. L., Barcroft, C. L., Guo, S.-W., Lynch, E. D., Boehnke, M., Chandrasekharappa, S., Hayes, S., Collins, F. S., Weber, B. L. and Glover, T. W. (1993) *Genomics* **17**, 624–631.
28. Albertsen, H. M., Smith, S. A., Mazoyer, S., Fujimoto, E., Stevens, J., Williams, B., Rodriguez, P., Cropp, C. S., Slijepcevic, P., Carlson, M., Robertson, M., Bradley, P., Lawrence, E., Harrington, T., Mei Sheng, Z., Hoopes, R., Sternberg, N., Brothman, A., Callahan, R., Ponder, B. A. J. and White, R. (1994) *Nature Genet.* **7**, 472–479.
29. Lehrach, H. (1990) In K. E. Davies and S. M. Tilghman (eds) *Genome Analysis Volume I: Genetic and Physical Mapping*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 39–81.
30. Van Arsdell, S. W. and Weiner, A. M. (1984) *Mol. Cell. Biol.* **4**, 492–499.
31. Chandrasekharappa, S. C., Friedman, L., King, S. E., Lee, Y.-H., Welsh, P., Bowcock, A. M., Weber, B. L., King, M.-C. and Collins, F. S. (1994) *Genomics* **21**, 458–460.
32. Senger, G., Ragoussis, J., Trowsdale, J. and Sheer, D. (1993) *Cytogenet. Cell Genet.* **64**, 49–53.
33. Senger, G., Jones, T. A., Fidlerova, H., Sanseau, P., Trowsdale, J., Duff, M. and Sheer, D. (1994) *Hum. Mol. Genet.* **3**, 1275–1280.
34. Lovett, M., Kere, J. and Hinton, L. M. (1991) *Proc. Natl Acad. Sci. USA* **88**, 9628–9632.
35. Korn, B., Sedlacek, Z., Manca, A., Kioschis, P., Konecki, D., Lehrach, H. and Poustka, A. (1992) *Hum. Mol. Genet.* **1**, 235–242.
36. Monaco, A. P., Neve, R. L., Colletti-Feener, C., Bertelson, C. J., Kurnit, D. M. and Kunkel, L. M. (1986) *Nature* **323**, 646–650.
37. Buckler, A. J., Chang, D. D., Graw, S. L., Brook, D., Haber, D. A., Sharp, P. A. and Housman, D. E. (1991) *Proc. Natl Acad. Sci. USA* **88**, 4005–4009.
38. Linsay, S. and Bird, A. P. (1987) *Nature* **327**, 336–338.
39. Bird, A. P. (1987) *Trends Genet.* **3**, 342–347.
40. Ishibashi, T., Bottaro, D. P., Chan, A., Miki, T. and Aaronson, S. A. (1992) *Proc. Natl Acad. Sci. USA* **89**, 12170–12174.
41. Futreal, P. A., Cochran, C., Rosenthal, J., Miki, Y., Swenson, J., Hobbs, M., Bennett, L. M., Haugen-Strano, A., Marks, J., Barrett, J. C., Tavtigian, S. V., Shattuck-Eidens, D., Kamb, A., Skolnick, M. and Wiseman, R. W. (1994) *Hum. Mol. Genet.* **3**, 1359–1364.
42. Parrish, J. E. and Nelson, D. L. (1993) *GATA* **10**, 29–41.
43. Sedlacek, Z., Korn, B., Konecki, D. S., Siebenhaar, R., Coy, J. F., Kioschis, P. and Poustka, A. (1993) *Hum. Mol. Genet.* **2**, 1865–1869.
44. Leib-Mösch, C., Brack-Werner, R., Werner, T., Bachmann, M., Faff, O., Erfle, V. and Hehlmann, R. (1990) *Cancer Res.* **50**, 5636–5642.
45. Hayward, W. S., Neel, B. G. and Astrin, S. M. (1981) *Nature* **290**, 475–479.
46. Ono, M., Kawakami, M. and Ushikubo, H. (1987) *J. Virol.* **58**, 955–959.
47. Kahn, R. A., Yucel, J. K. and Malhotra, V. (1993) *Cell* **75**, 1045–1048.
48. Alsip, G. R. and Konkel, D. A. (1986) *Nucleic Acids Res.* **14**, 2123–2138.
49. Saito, H. and Streuli, M. (1991) *Cell Growth Different.* **2**, 59–65.
50. Tsukamoto, T., Takahashi, T., Ueda, R., Hibi, K., Saito, H. and Takahashi, T. (1992) *Cancer Res.* **52**, 3506–3509.
51. Gaits, F., Lui, R. Y., Ragab, A., Selves, J., Ragabthomas, J. M. F. and Chap, H. (1994) *Cell. Mol. Biol.* **40**, 677–685.
52. Larin, Z., Monaco, A. P. and Lehrach, H. (1991) *Proc. Natl Acad. Sci. USA* **88**, 4123–4127.
53. Albertsen, H. M., Abderrahim, H., Cann, H. M., Dausset, J., Le Paslier, D. and Cohen, D. (1990) *Proc. Natl Acad. Sci. USA* **87**, 4256–4260.
54. Brownstein, B. H., Silverman, G. A., Burke, D. T., Korsmeyer, S. J., Schlessinger, D. and Olson, M. V. (1989) *Science* **244**, 1348.
55. Green, E. D. and Olson, M. V. (1990) *Proc. Natl Acad. Sci. USA* **87**, 1213–1217.
56. Nelson, D. L., Ledbetter, S. A., Corbo, L., Victoria, M. F., Ramirez-Solis, R., Webster, T. D., Ledbetter, D. H. and Caskey, C. T. (1989) *Proc. Natl Acad. Sci. USA* **86**, 6686–6690.
57. Baxendale, S., Bates, G. P., MacDonald, M. E., Gusella, J. F. and Lehrach, H. (1991) *Nucleic Acids Res.* **19**, 6651.
58. Drumm, M. L., Smith, C. L., Dean, M. D., Cole, J. L., Iannuzzi, M. C. and Collins, F. S. (1988) *Genomics* **2**, 346–354.
59. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular cloning. A Laboratory Manual*. 2nd edition, Cold Spring Harbor Laboratory Press, New York, Vol. 1, 2, 3.
60. Feinberg, A. P. and Vogelstein, B. (1984) *Ann. Biochem.* **137**, 266–267.
61. Campbell, I. G., Nicolai, H. M., Foulkes, W. D., Senger, G., Stamp, G. W., Allan, G., Boyer, C., Jones, K., Bast, R. C. and Solomon, E. (1994) *Hum. Mol. Genet.* **3**, 589–594.
62. Leiter, A. B., Montminy, M. R., Jamieson, E. and Goodman, R. H. (1985) *J. Biol. Chem.* **260**, 13013–13017.
63. Luty, J., Kondoleon, S., van Tuinen, P., Ledbetter, D. H., Vissing, H., Ramsay, A. and Litt, M. (1988) *Nucleic Acids Res.* **16**, 6250.
64. Fidlerova, H., Senger, G., Kost, M., Sanseau, P. and Sheer, D. (1994) *Cytogenet. Cell Genet.* **65**, 203–205.
65. Church, D. M., Stodler, C. J., Rutter, J. L., Murrell, J. R., Trofatter, J. A. and Buckler, A. J. (1994) *Nature Genet.* **6**, 98–104.
66. North, M. A., Sanseau, P., Buckler, A. J., Church, D., Jackson, A., Patel, K., Trowsdale, J. and Lehrach, H. (1993) *Mammalian Genome* **4**, 466–474.
67. Lipman, D. J. and Pearson, W. R. (1985) *Science* **227**, 1435–1441.

Note added in proof

The *BRCA1* gene has recently been identified (Miki *et al.* (1994) Isolation of *BRCA1*, the 17q-linked breast and ovarian cancer susceptibility gene. *Science* **266**, 66–71 and Futreal *et al.* *BRCA1* mutations in primary breast and ovarian carcinomas. *Science* **266**, 120–122).