The complete *BRCA2* gene and mutations in chromosome 13q-linked kindreds

[1]*Myriad Genetics Inc., Salt Lake City, Utah USA*
[2]*Laboratory of Molecular Endocrinology, CHUL Research Center and Laval University, Quebec City, Canada*
[3]*Department of Genetics, Hospital for Sick Children, University of Toronto, Ontario, Canada*
[4]*Deparment of Hematology-Oncology, University of Pennsylvania, Philadelphia, USA*
[5]*Genetic Epidemiology Group, Department of Medical Informatics and Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah, USA*
[6]*Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan, USA*
[7]*Molecular and Cell Biology Research Laboratory, Icelandic Cancer Society, Rekjavik, Iceland*
[8]*Department of Human Genetics, Memorial Sloan-Kettering Cancer Center, New York, New York, USA*
[9]*Unite de Genetique Oncologique, Institute Curie, Paris, France*
[10]*Division of Biology, California Institute of Technology, Pasadena, California, USA*
[11]*Unit of Genetic Epidemiology, International Agency for Research on Cancer, 150 cours Albert Thomas, 63972, Lyon, Cedex 08, France*

Correspondence should be addressed to D.E.G.[11]

S.V. Tavtigian[1], J. Simard[2], J. Rommens[3], F. Couch[4],
D. Shattuck-Eidens[1], S. Neuhausen[5], S. Merajver[6],
S. Thorlacius[7], K. Offit[8], D. Stoppa-Lyonnet[9],
C. Belanger[2], R. Bell[1], S. Berry[1], R. Bogden[1],
Q. Chen[1], T. Davis[1], M. Dumont[2], C. Frye[1],
T. Hattier[1], S. Jammulapati[1], T. Janecki[1], P. Jiang[1],
R. Kehrer[1], J.-F. Leblanc[2], J.T. Mitchell[1],
J. McArthur-Morrison[3], K. Nguyen[5], Y. Peng[4],
C. Samson[2], M. Schroeder[1], S.C. Snyder[1],
L. Steele[5], M. Stringfellow[1], C. Stroup[1],
B. Swedlund[1], J. Swensen[5], D. Teng[1], A. Thomas[1],
T. Tran[1], T. Tran[5], M. Tranchant[2],
J. Weaver-Feldhaus[1], A.K.C. Wong[1], H. Shizuya[10],
J.E. Eyfjord[7], L. Cannon-Albright[5], F. Labrie[2],
M.H. Skolnick[1,5], B. Weber[4], A. Kamb[1]
& D.E. Goldgar[5,11]

Breast carcinoma is the most common malignancy among women in developed countries. Because family history remains the strongest single predictor of breast cancer risk, attention has focused on the role of highly penetrant, dominantly inherited genes in cancer-prone kindreds[1]. *BRCA1* was localized to chromosome 17 through analysis of a set of high-risk kindreds[2], and then identified four years later by a positional cloning strategy[3]. *BRCA2* was mapped to chromosomal 13q at about the same time[4]. Just fifteen months later, Wooster *et al.*[5] reported a partial *BRCA2* sequence and six mutations predicted to cause truncation of the BRCA2 protein. While these findings provide strong evidence that the identified gene corresponds to *BRCA2*, only two thirds of the coding sequence and 8 out of 27 exons were isolated and screened; consequently, several questions remained unanswered regarding the nature of *BRCA2* and the frequency of mutations in 13q-linked families. We have now determined the complete coding sequence and exonic structure of *BRCA2* (GenBank accession #U43746), and examined its pattern of expression. Here, we provide sequences for a set of PCR primers sufficient to screen the entire coding sequence of *BRCA2* using genomic DNA. We also report a mutational analysis of *BRCA2* in families selected on the basis of linkage analysis and/or the presence of one or more cases of male breast cancer. Together with the specific mutations described previously, our data provide preliminary insight into the *BRCA2* mutation profile.

BRCA2 lies near the centre of a 1.4-megabase (Mb) interval flanked by markers *D13S1444* and *D13S310* (F.C. *et al.*, unpublished), completely within a 0.3-Mb homozygous deletion identified in a pancreatic carcinoma xenograft[6] The full-length sequence of the *BRCA2* transcript was assembled by combination of several smaller sequences obtained from hybrid selection, exon trapping, cDNA library screening, genomic sequencing and inter-clone PCR experiments using cDNA as template for amplification ('island hopping'; Fig. 1*a*). The extreme 5' end of the mRNA, including the predicted translational start site, was identified by a modified 5' RACE protocol[7]. The first nucleotide in the sequence (nt 1) is a non-template G, an indication that the mRNA cap is contained in the sequence. A portion of exon 11, which is nearly 5 kilobases (kb) in length, was identified by analysis of roughly 900 kb of genomic sequence in the public domain (ftp://genome.wust1.edu/pub/gscl/brca). This genomic sequence was condensed with our own genomic sequence into a set of 160 sequence contigs. When the condensed sequence was scanned for open reading frames (ORFs), a contiguous stretch of nearly 5 kb was identified which was spanned by long ORFs. This sequence was linked together by island hopping experiments with two previously identified candidate gene fragments (F.C. *et al.*, unpublished)[8]. Our composite *BRCA2* cDNA sequence consists of 11,385 bp but does not include the polyadenylation signal or poly(A) tail.

Conceptual translation of the cDNA reveals an ORF beginning at nt 229 and encoding a protein of 3,418 amino acids. The peptide bears no similarity to other proteins apart from sequence composition. There is no signal sequence at the N terminus, and no obvious membrane-spanning regions. Like BRCA1, the BRCA2 protein is highly charged; roughly one quarter of the residues are acidic or basic. However, there are few clues as to the biochemical function of BRCA2.

The *BRCA2* gene structure was determined by comparing cDNA and genomic sequences. *BRCA2* is composed of 27 exons distributed over roughly 70 kb of genomic DNA (Fig. 1*b*). A CpG-rich region at the 5' end of *BRCA2* extending upstream suggests the presence of regulatory signals often associated with CpG 'islands.' Unlike most human genes, the coding sequence is AT-rich (> 60%). Based on Southern blot experiments, *BRCA2* appears to be unique, with no close homologue in the human genome (data not shown).

Hybridization of labelled cDNA to human multiple tissue northern filters revealed an 11–12 kb transcript detectable in thymus and testis (Fig. 2*a*), suggesting that little of the *BRCA2* mRNA sequence is missing from our composite cDNA. Because the northerns did not include mammary gland RNA, we performed RT-PCR experiments using a *BRCA2* cDNA amplicon that spans the last splice junction on a set of human tissue RNAs (Fig. 2*b*). All of the samples produced positive signals. The highest levels of expression were observed in breast and thymus, with slightly lower levels in lung, ovary and spleen. This pattern of expression is similar to that produced by *BRCA1* amplicons[3].

Individuals from 18 putative *BRCA2* kindreds were screened for *BRCA2* germline mutations by DNA sequence analysis[4]. Twelve kindreds have at least one case of male breast cancer; four have two or more cases; and four include at least one individual affected with ovarian cancer who shares the linked *BRCA2* haplotype. Each of the 18 kindreds has a posterior probability of harboring a *BRCA2* mutation of at least 69% and 9 kindreds have probabilities greater than 90%. Based on these combined probabilities, 16 of the 18 kindreds are expected to segregate *BRCA2* mutations. The entire coding sequence and associated splice junctions were

Fig. 1 *a*, Cloning of *BRCA2*. Sequence-space relationships between the cDNA clones, hybrid selected clones, cDNA PCR products, and genomic sequences used to assemble the *BRCA2* transcript sequence. 2-Br-C:RACE is a biotin-capture RACE product obtained from both human breast and human thymus cDNA. The cDNA clone λ sC713.1 was identified by screening a pool of human testis and HepG2 cDNA libraries with hybrid selected clone GT 713. The sequence 1-Br:CG026→5 kb was generated from a PCR product beginning at the exon 7/8 junction (within λ sC713.1) and terminating within an hybrid selected clone that is part of exon 11. The sequence of exon 11 was corrected by comparison to hybrid selected clones, genomic sequence in the public domain, and radioactive DNA sequencing gels. Hybrid selected clones located within that exon (clone names beginning with mH or GT) are placed below it. The cDNA clones λ wCPF1B8.1, λ wCPF1A5.1, λ wCPF1A5.12, λ wCBF1B6.2, and λ wCBF1B6.3 were identified by screening a pool of human mammary gland, placenta, testis, and HepG2 cDNA libraries with the exon trapped clones wXBF1B8, wXPF1A5, and wXBF1B6. The clone λ wCBF1B6.3 is chimaeric (indicated by the dashed line), but its 5' end contained an important overlap with λ wCPF1A5.1. ⌐➤ denotes the translation initiator. ↑ denotes the translation terminator. *b*, Genomic organization of *BRCA2*. The exons (boxes and/or vertical lines) are parsed across the publicly available genomic sequences (horizontal lines) such that their sizes and spacing are proportional. The name of each genomic sequence is given on the left. The sequences 92M18.00541 and 92M18.01289 actually overlap. Distances between the other genomic sequences are not known. No databases contained genomic sequences overlapping with exon 21. The extent of the peptide sequence published by Wooster *et al.*[5] is indicated below the parsed exons. Exons 1, 11 and 21 are numbered. * denotes two adjacent exons spaced closely enough that they are not resolved at this scale.
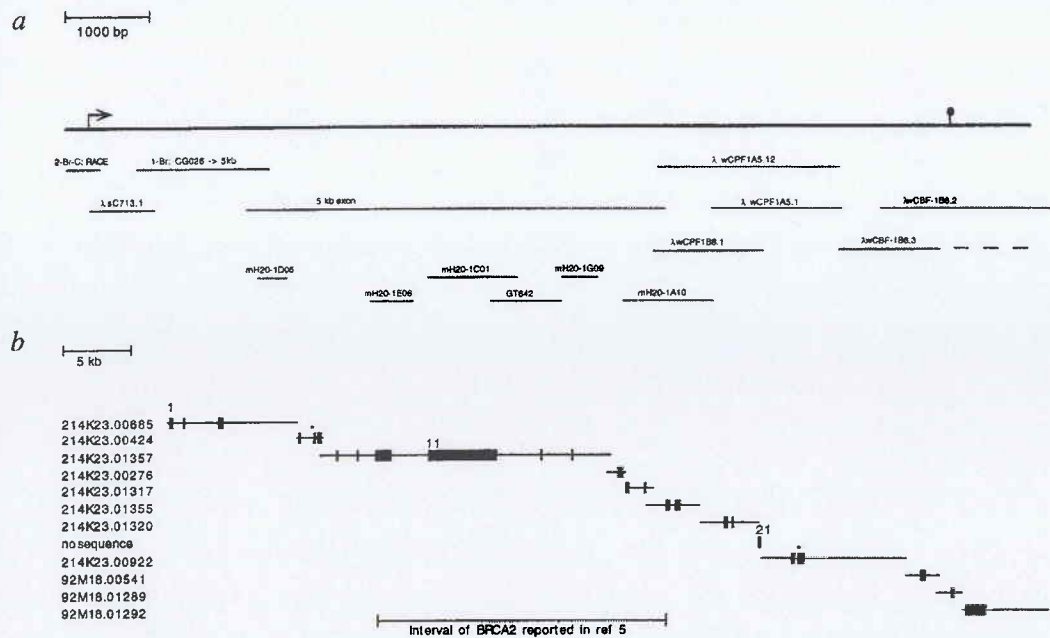
screened for mutations in multiple individuals from 9 kindreds using either cDNA or genomic DNA (Table 1*a*). Individuals from the remaining 9 kindreds were screened for mutations using only genomic DNA (Table 1*b*). These latter screening experiments encompassed 99% of the coding sequence (all exons excluding exon 15) and all but two of the splice junctions.

We identified potentially deleterious sequence alterations in 9 of the 18 kindreds (Table 1). All except one — a deletion of three nucleotides (kindred 1019) — involved nucleotide deletions that altered the reading frame, leading to truncation of the BRCA2 protein. The 3-nt deletion was not observed in 36 unrelated breast cancer cases, hence we have included it in our mutation tally although its effect on BRCA2 function must be proved. All 9 mutations are distinct. In most cases, segregation studies show that the mutations are present in multiple haplotype carriers and absent in noncarriers (data not shown). In addition to these mutations, three silent and three missense substitutions were detected. Based on their frequencies in a set of control chromosomes, we have classified these variants as neutral polymorphisms (Table 1*c*).

Nine of the 18 kindreds were tested for transcript loss. Specific polymorphic sites known to be heterozygous in genomic DNA (Table 1*c*) were examined in cDNA from kindred individuals. The appearance of hemizygosity was interpreted as evidence for a mutation leading to reduction in mRNA levels. Two of the 9 kindreds displayed signs of reduced transcript levels. However, one of these kindreds (1018) contained a previously identi-

fied frameshift mutation, while the second (2367) contained an aberrantly spliced *BRCA2* mRNA that lacked exon 2 (data not shown). The abundance of this mutant transcript was estimated at roughly 20% of wild-type. This implies that some mutations in the *BRCA2* coding sequence may destabilize the transcript in addition to disrupting the protein sequence, similar to *BRCA1* (ref. 9). In no case was a purely regulatory mutation inferred. In summary, 56% of the kindreds (10/18) contained an altered *BRCA2* gene. Half of our kindreds contained microdeletion mutations, mostly frameshifts; none contained missense or nonsense mutations.

BRCA2 is remarkably similar to *BRCA1*. Both genes encode exceptionally large, highly charged proteins; both have many exons; both have a large exon 11 (3,426 bp for *BRCA1* and 4,932 bp for *BRCA2*); both have translational start sites in exon 2; both have coding sequences that are AT-rich; both span approximately 70 kb of genomic DNA; and, both are expressed at high levels in testis. Whether or not *BRCA1* and *BRCA2* participate in the same pathway of tumour suppression in breast epithelium is not known. The different phenotypes of the two mutant genes, particularly the role of *BRCA2* in male breast cancer, suggest that they may not function in the same genetic pathway.

Mutational analysis of *BRCA2* reveals other features in common with *BRCA1*. The distribution of mutations in *BRCA2* appears to be uniform based on our data and data from Wooster *et al.*[5] Mutations have been identified in 6 of the 26 coding exons of *BRCA2* (exons 2, 9, 10, 11, 18 and 23). Nine mutations have been detected
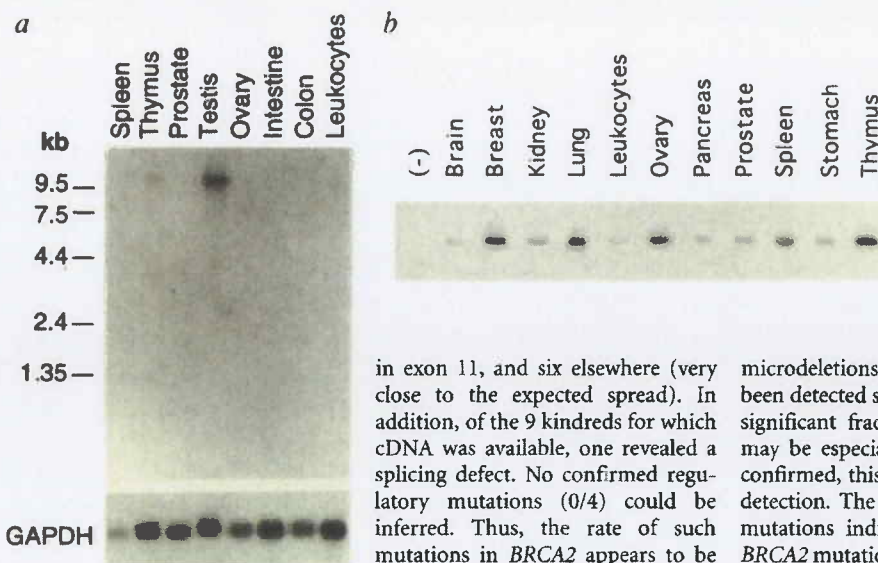
Fig. 2 a, Analysis of *BRCA2* expression. Upper panel: Multiple tissue northern (MTN) filters (Clontech) probed with the 1.55-kb 1-Br:CG026→5-kb PCR product. The 11–12 kb *BRCA2* transcript is detected in testis. Lower panel: the same filter probed with glyceraldehyde-3-phosphate dehydrogenase. *b*, RT-PCR analysis of *BRCA2* expression. A BRCA2 cDNA amplicon spanning the last splice junction was amplified from 0.4 ng or random-primed A⁺ cDNA from the indicated tissues.

in exon 11, and six elsewhere (very close to the expected spread). In addition, of the 9 kindreds for which cDNA was available, one revealed a splicing defect. No confirmed regulatory mutations (0/4) could be inferred. Thus, the rate of such mutations in *BRCA2* appears to be comparable (or lower) to the rate observed in *BRCA1* (10–20%)[10].

The mutation profile of *BRCA2*, however, may differ from *BRCA1*. Of the 15 sequence alterations described so far in *BRCA2* (9 here, 6 from ref. 5), all involve deletions of 1–6 nucleotides. In contrast, microinsertions plus point mutations in *BRCA1* are about as common as

microdeletions[11]. Furthermore, no point mutations have been detected so far in *BRCA2*, whereas they constitute a significant fraction of *BRCA1* variants. Thus, *BRCA2* may be especially vulnerable to deletion mutations. If confirmed, this trait has clear implications for mutation detection. The lack of multiple observations of specific mutations indicates that the number of independent *BRCA2* mutations in the population may be even greater than the number of *BRCA1* mutations.

Approximately two thirds as many mutations were detected in *BRCA2* as expected among our 18 families. This may imply the presence of mutations in regions that are difficult to detect, or in regions that were not screened. For example, only half of the kindreds were screened for regulatory mutations due to lack of cDNA

Table 1 *BRCA2* mutations and polymorphisms

a, Families screened for complete coding sequence (informative cDNA sample)

Family	Number of cancer cases				Lod score	Prior probability	*BRCA2* mutation	exon	codon	effect
	FBC	FBC <50yrs	Ov	MBC						
UT-107	20	18	2	3	5.06	1.00	277 delAC	2	17	termination codon at 29
UT-1018	11	9	0	1	2.47	1.00	982 del4	9	252	termination codon at 275
UT-2044	8	6	4	1	2.13	1.00	4706 del4	11	1493	termination codon at 1502
UT-2367	9	5	1	0	2.09	0.99	SP			deletion of exon 2 in mRNA
UT-2327	13	6	0	0	1.92	0.99	ND			
UT-2388	3	3	1	0	0.92	0.92	ND			
UT-2328	10	4	0	1	0.21	0.87	ND			
UT-4328	4	3	0	0	0.18	0.69	ND			
MI-1016	4	2	0	1	0.04	0.81	ND			

b, Families screened for all exons except 15 (no cDNA sample available)

Family	FBC	FBC <50yrs	Ov	MBC	Lod score	Prior probability	*BRCA2* mutation	exon	codon	effect
CU-20	4	3	2	2	1.09	1	8525 delC	18	2766	termination codon at 2776
CU-159	8	4	0	0	0.99	0.94	9254 del5	23	3009	termination codon at 3015
UT-2043	2	2	1	1	0.86	0.97	4075 delGT	11	1283	termination codon at 1285
IC-2204	3	1	0	4	0.51	0.98	999 del5	9	257	termination codon at 273
MS-075	4	1	0	1	0.50	0.93	6174 delT	11	1982	termination codon at 2003
UT-1019	5	1	0	2	nd	0.95	4132 del3	11	1302	deletion of thr1302
UT-2027	4	4	0	0	0.39	0.79	ND			
MS-036	3	2	0	1	nd	0.90	ND			
UT-2171	5	4	2	0	nd	nd	ND			

c, Common polymorphisms in *BRCA2*

Polymorphism	Description	Effect	Number of chromosomes				
			A	C	G	T	Total
5'UTR-203	TACCAA(G/A)CATTG		9	0	25	0	34
PM-1342	GTA GCA (C/A)AT CAG	His→Asn	24	14	0	0	38
PM-2457	GTA CAA CA(T/C) TCA	His→His	2	0	0	38	40
PM-3199	TAC ATG (A/G)AC AAT	Asn→Asp	37	0	3	0	40
PM-3668	CCT GAA A(A/G)C CAG	Asn→Ser	34	0	6	0	40
PM-4035	GAT TCT GT(T/C) GTT	Val→Val	0	4	0	36	40
PM-7470	ACT AAA TC(A/G) CAT	Ser→Ser	34	0	6	0	40
3'UTR-10,854	AAAAGAA(G/A)CATTTCTA		17	0	15	0	32
3'UTR-11,316	ATTTATTTTTTTTT(T)CAAC	T9 vs. T10	T9: (7)		T10: (29)		36

Mutations and polymorphisms are given by nucleotide position. SP–inferred splice mutation. ND–none detected. nd–not determined. FBC–female breast cancer. Ov–ovaian cancer. MBC–male breast cancer.

letters

Table 2 Primers used to amplify and mutation screen BRCA2 from genomic DNA

Exon	Forward Primer	Reverse Primer	Nested Primer
Exon 2	TGTTCCCATCCTCACAGTAAG*	GTACTGGGTTTTTAGCAAGCA*	
Exon 3	GGTTAAAACTAAGGTGGGA*	ATTTGCCCAGCATGACACA*	
Exon 4	TTTCCCAGTATAGAGGAGA*	GTAGGAAAATGTTTCATTTAA*	
Exon 5	ATCTAAAGTAGTATTCCAACA*	GGGGGTAAAAAAAGGGGAA*	
Exon 6	GAGATAAGTCAGGTATGATT*	AATTGCCTGTATGAGGCAGA*	
Exon 7	GGCAATTCAGTAAACGTTAA*	ATTGTCAGTTACTAACACAC*	
Exon 8	GTGTCATGTAATCAAATAGT*	CAGGTTTAGAGACTTTCTG*	
Exon 9	GGACCTAGGTTGATTGCA*	GTCAAGAAAGGTAAGGTAA*	
Exon 10-1	CTATGAGAAAAGGTTGTGAG*	CCTAGTCTTGCTAGTTCTT*	
Exon 10-2	AACAGTTGTAGATACCTCTGAA*	GACTTTTTGATACCCTGAAATG*	
Exon 10-3	CAGCATCTTGAATCTCATACAG*	CATGTATACAGATGATGCCTAAG*	
Exon 11-1	AACTTAGTGAAAAATATTTAGTGA	ATACATCTTGATTCTTTTCCAT*	TTTAGTGAATGTGATTGATGGT*
Exon 11-2	AGAACCAACTTTGTCCTTAA	TTAGATTTGTGTTTTGGTTGAA*	TAGCTCTTTTGGGACAATTC*
Exon 11-3	ATGGAAAAGAATCAAGATGTAT*	CCTAATGTTATGTTCAGAGAG	GCTACCTCCAAAACTGTGA*
Exon 11-4	GTGTAAAGCAGCATATAAAAAT*	CTTGCTGCTGTCTACCTG	AGTGGTCTTAAGATAGTCAT*
Exon 11-5	CCATAATTTAACACCTAGCCA**	CCAAAAAAGTTAAATCTGACA**	
Exon 11-6	GGCTTTTATTCTGCTCATGGC*	CCTCTGCAGAAGTTTCCTCAC*	
Exon 11-7	AACGGACTTGCTATTTACTGA*	AGTACCTTGCTCTTTTTCATC*	
Exon 11-8	CAGCTAGCGGGAAAAAAGTTA*	TTCGGAGAGATGATTTTTGTC*	
Exon 11-9	GCCTTAGCTTTTTACACAA*	TTTTGATTATATGTCGTTG	TTATTCTCGTTGTTTTCCTTA*
Exon 11-10	CCATTAAATTGTCCATATCTA*	GACGTAGGTGAATAGTGAAGA	TCAAATTCCTCTAACACTCC*
Exon 11-11	GAAGATAGTACCAAGCAAGTC	TGAGACTTTGGTTCCTAATAC*	AGTAACGAACATTCAGACCAG*
Exon 12	GTCTTCACTATTCACCTACG*	CCCCCAAACTGACTACACAA	AGCATACCAAGTCTACTGAAT*
Exon 13	ACTCTTTCAAACATTAGGTCA*	TTGGAGAGGCAGGTGGAT	CTATAGAGGGAGAACAGAT*
Exon 14	TTTATGCTGATTTCTGTTGTAT	ATAAAACGGGAAGTGTTAACT*	CTGTGAGTTATTTGGTGCAT*
Exon 15	GAATACAAAACAGTTACCAGA	CACCACCAAAGGGGGAAA*	AAATGAGGGTCTGCAACAAA*
Exon 16	ATTTCAATTTTATTTTTGCT*	ATGAAATAAAATTACACTCTGTC	TACACTCTGTCATAAAAGCC*
Exon 17	ATGTTTTTGTAGTGAAGATTCT	TAGTTCGAGAGACAGTTAAG*	CAGTTTTGGTTTGTTATAATTG*
Exon 18	CAGAGAATAGTTGTAGTTGTT	AACCTTAACCCATACTGCC*	TTCAGTATCATCCTATGTGG*
Exon 19	TTTTATTCTCAGTTATTCAGTG	GAAATTGAGCATCCTTAGTAA*	AATTCTAGAGTCACACTTCC*
Exon 20	ATATTTTTAAGGCAGTTCTAGA	TTACACACACCAAAAAAGTCA*	TGAAAACTCTTATGATATCTGT*
Exon 21	TGAATGTTATATATGTGACTTTT*	CTTGTTGCTATTCTTTGTCTA	CCCTAGATACTAAAAAATAAAG*
Exon 22	CTTTTAGCAGTTATATAGTTTC	GCCAGAGAGTCTAAAACAG*	CTTTGGGTGTTTTATGCTTG*
Exon 23	TTTGTTGTATTTGTCCTGTTTA	ATTTTGTTAGTAAGGTCATTTTT*	GTTCTGATTGCTTTTTATTCC*
Exon 24	ATCACTTCTTCCATTGCATC*	CCGTGGCTGGTAAATCTG*	
	CTGGTAGCTCCAACTAATC*	ACCGGTACAAACCTTTCATTG*	
Exon 25	CTATTTTGATTTGCTTTTATTATT*	GCTATTTCCTTGATACTGGAC*	
Exon 26	TTGGAAACATAAATATGTGGG*	ACTTACAGGAGCCACATAAC*	
Exon 27	CTACATAATTATGATAGGCTNCG**	GTACTAATGTGTGGTTTGAAA**	
		TCAATGCAAGTTCTTCGTCAGC*	

Primers with a * were used for sequencing.
Primers without a * are replaced by the internal nested primer in the third column for both the second round of PCR and sequencing. For large exons requiring internal sequencing primers, ** signifies the primers used to amplify the exon.

samples. In addition, in only one of the two families in which transcript loss was detected was the actual sequence alteration identified. However, 7/9 kindreds with Lod scores over 0.39 revealed mutations. Thus, it is possible that the assumed prior probabilities for our kindreds were inflated. Several of the kindreds for which mutations have not been defined may not segregate *BRCA2* mutations. Many have been screened without success for *BRCA1* mutations. Therefore, some of the families may represent sporadic clusters; others may be afflicted with breast cancer due to segregation of genes besides *BRCA1* and *BRCA2*.

The characterization of two genes, *BRCA1* and *BRCA2*, that together may account for the vast majority of early-onset hereditary breast cancer, is a major step toward early detection of an important human disease. One of the significant goals ahead is the development of reliable diagnostic tests for *BRCA1* and *BRCA2*. In addition, the definition of other genes that may contribute to breast cancer incidence is an important pursuit.

Methods

Hybrid selection. Two distinct methods of hybrid selection were used. *Method 1: cDNA preparation and selection.* Poly(A)+ enriched RNA from human mammary gland, ovary, testis, fetal brain and placenta tissues and from total RNA of the cell line Caco-2 (ATCC HTB 37) were reverse transcribed using the tailed random primer $RXGN_6$ (5'–CGGAATTCTGCAGATCTA'B'C'N$_6$) and M-MLV Reverse Transcriptase (Life Technologies, Inc.). First strand cDNA was poly(A) tailed, 2nd strand synthesis was primed with the oligo $RXGT_{12}$ (5'–CGGAATTCTGCAGATCT$_{12}$), and then the ds cDNA was expanded by amplification with the primer RXG (5'–CGGAATTCTGCAGATCT). Hybrid selection was carried out for two consecutive rounds to immobilized P1 or BAC DNA as described[12,13]. Groups of two to four overlapping P1 and/or BAC clones were used in individual selection experiments. Hybridizing cDNA was collected, passed over a G50 Fine

Sephadex column and reamplified using the primer RXG. The products were then digested with *Eco*RI, size selected on agarose gels, and ligated into pBluescript (Stratagene) that had been digested with *Eco*RI and treated with calf alkaline phosphatase (Boehringer Mannheim). Ligation products were transformed into competent DH5α *E.coli* cells (Life Technologies, Inc.). *Characterization of retrieved cDNAs.* 200 to 300 individual colonies from each ligation (from each 250 kb of genomic DNA) were picked and gridded into microtitre plates for ordering and storage. Cultures were replica transferred onto Hybond N membranes (Amersham) supported by LB agar with ampicilin. Colonies were allowed to propagate and were subsequently lysed with standard procedures. Initial analysis of the cDNA clones involved a prescreen for ribosomal sequences and subsequent cross screenings for detection of overlap and redundancy. Approximately 10–25% of the clones were eliminated as they hybridized strongly with radiolabelled cDNA obtained from total RNA. Plasmids from 25 to 50 clones from each selection experiment that did not hybridize in the prescreening were isolated for further analysis. The retrieved cDNA fragments were verified to originate from individual starting genomic clones by hybridization to restriction digests of DNAs of the starting clones, of a hamster hybrid cell line (GM10898A) that contains chromosome 13 as its only human material and to human genomic DNA. The clones were tentatively assigned into groups based on the overlapping or nonoverlapping intervals of the genomic clones. Of clones tested, approximately 85% mapped appropriately to the starting clones.

Method 2: (refs 14, 15): *cDNA preparation.* Poly(A)+ enriched RNA from human mammary gland, brain, lymphocyte, and stomach were reverse transcribed using the tailed random primer XN_{12} [5'–(NH$_2$)-GTAGTGCAAGGCTCGAGAACN$_{12}$] and Superscript II reverse transcriptase (Gibco BRL). After 2nd strand synthesis and end polishing, the ds cDNA was purified on sepahrose CL-4B columns (Pharmacia). cDNAs were 'anchored' by ligation of a double stranded oligo RP [5'–(NH2)-TGAGTAGAATTCTAACGGCCGTCATTGTTC annealed to 5'–GAACAATGACGGCCGTTAGAATTCTAC-TCA-(NH$_2$)] to their 5' ends (5' relative to mRNA) using T4 DNA ligase. Anchored ds cDNA was then repurified on sepahrose CL-4B columns. *Selection.* cDNAs from mammary gland, brain, lymphocyte, and stomach tissues were first amplified using a nested version of RP (RP.A: 5'–TGAGTAGAATTC-TAACGGCCGTCAT) and XPCR [5'–(PO4)-GTAGTGCAAG-GCTCGAGAAC] and purified by fractionation on Sepharose CL-4B. Selection probes were prepared from purified P1s, BACs, or PACs, by digestion with *Hinf*1 and Exonuclease III. The single stranded probe was photo-labelled with photobiotin (Gibco BRL) according to the manufacturers recommendations. Probe, cDNA and Cot-1 DNA were hybridized in 2.4 M TEA-Cl, 10mM NaPO4, 1mM EDTA. Hybridized cDNAs were captured on streptavidin-paramagnetic particles (Dynal), eluted, reamplified with a further nested version of RP [RP.B: 5'–(PO4)-TGAGTA-GAATTCTAACGGCCGTCATTG] and XPCR, and size selected on Sepharose CL-6B. The selected, amplified cDNA was hybridized with an additional aliquot of probe and Cot-1 DNA. Captured and eluted products were amplified again with RP.B and XPCR, size selected by gel electrophoresis and cloned into dephosphorylated *Hinc*II-cut pUC18. Ligation products were transformed into XL2-Blue ultra-competent cells (Stratagene). *Analysis.* Approximately 192 colonies for each single-probe selection experiment were amplified by colony PCR using vector primers and blotted in duplicate onto Zeta Probe nylon filters (Bio-Rad). The filters were hybridized using standard procedures with either random primed Cot-1 DNA or probe DNA (P1, BAC, or PAC). Probe positive, Cot-1 negative clones were sequenced in both directions using vector primers on an ABI 377 sequencer.

Exon trapping. Exon amplification was performed using a minimally overlapping set of BACs, P1s and PACs in order to isolate a number of gene sequences from the *BRCA2* candidate region.

Pools of genomic clones were assembled, containing from 100–300 kb of DNA in the form of 1-3 overlapping genomic clones. Genomic clones were digested with *Pst*I or *Bam*H1 + *Bgl*I1 and ligated into *Pst*I or *Bam*H1 sites of the pSPL3 splicing vector. The exon amplification technique was performed[16] and the end products were cloned in the pAMP1 plasmid from the Uracil DNA Glycosylase cloning system (BRL). Approximately 6,000 clones were picked, propagated in 96 well plates, stamped onto filters, and analysed for the presence of vector and repeat sequences by hybridization. Each clone insert was PCR amplified and tested for redundancy, localization, and human specificity by hybridization to grids of exons and dot blots of the parent genomic DNA. Unique candidate exons were sequenced, searched against the databases, and used for hybridization to cDNA libraries.

5' RACE. The 5' end of *BRCA2* was identified by a modified RACE protocol called biotin capture RACE. Poly(A)$^+$ enriched RNA from human mammary gland and thymus was reverse transcribed using the tailed random primer XN_{12} [5'–(NH$_2$)-GTAGTGCAAG-GCTCGAGAACN$_{12}$] and Superscript II reverse transcriptase (Gibco BRL). The RNA strand was hydrolysed in NaOH and first strand cDNA purified by fractionation on Sepharose CL-4B (Pharmacia). First strand cDNAs were 'anchored' by ligation of a double stranded oligo with a 7 bp random 5' overhang [ds UCA: 5'–CCTTCACACGCGTATCGATTAGTCACN$_7$–(NH$_2$) annealed to 5'(PO$_4$)-GTGACTAATCGATACGCGTGTGAAGGTGC] to their 3' ends using T4 DNA ligase. After ligation, the anchored cDNA was repurified by fractionation on Sepharose CL-4B. The 5' end of BRCA2 was amplified using a biotinylated reverse primer [5'(B)-TTGAAGAACAACAGGACTTTCACTA] and a nested version of UCA [UCP.A: 5'-CACCTTCACACGCGTATCG]. PCR products were fractionated on an agarose gel, gel purified, and captured on streptavidin-paramagnetic particles (Dynal). Captured cDNA was reamplified using a nested reverse primer [5'–GTTCGTAATTGTTGTTTTTATGTTCAG] and a further nested version of UCA [UCP.B: 5'-CCTTCACACGCGTATC-GATTAG]. This PCR reaction gave a single sharp band on an agarose gel; the DNA was gel purified and sequenced in both directions on an ABI 377 sequencer.

cDNA clones. Human cDNA libraries were screened with ^{32}P labelled hybrid selected or exon trapped clones. Phage eluted from tertiary plaques were PCR amplified with vector specific primers and then sequenced on an ABI 377 sequencer.

Northern blots. Multiple tissue northern (MTN) filters, which are loaded with 2 µg per lane of poly(A)$^+$ RNA derived from a number of human tissues, were purchased from Clontech. ^{32}P-random-primer labelled probes corresponding to the 1.55 kb PCR fragment 1-Br:CG026→5 kb (exon 7/8 junction into exon 11), and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) were used to probe the filters. Prehybridizations were at 42° C in 50% formamide, 5× SSPE, 1% SDS, 5× Denhardt's mixture, 0.2 mg/ml denatured salmon testis DNA and 2 µg/ml Poly-(A)$^+$. Hybridizations were in the same solution with the addition dextran sulfate to 4% and probe. Stringency washes were in 0.1× SSC/0.1% SDS at 50 °C.

RT-PCR analysis. Poly-(A)$^+$ RNA extracted from 11 human tissues was reverse transcribed using random primes and Superscript II reverse transcriptase (Gibco BRL). Thereafter, 0.4 ng of each cDNA sample was amplified for 20 cycles using the *BRCA2* primers B2#F9833 (5'–CGTACACTGCTCAAATCATTC) and B2#R10061 (5'–GACTAACAGGTGGAGGTAAAG). Samples were diluted 10-fold, and then 2 µl aliquots reamplified for 18 cycles using the primers B2#F9857 (5'–GTACAGGAAA-CAAGCTTCTGA) and B2#R10061.

PCR amplification and mutation screening. All 26 coding exons of *BRCA2* and their associated splice sites were amplified from genomic DNA as described[17]. The DNA sequences of the primers, some of which lie in flanking intron sequence, used for amplification and sequencing appear in Table 2. Some of the exons (2 through 10, 11-5, 11-6, 11-7, and 23 through 27) were amplified by a simple one step method. The PCR conditions for those exons were: single denaturing step of 95 °C (1 min); 40 cycles of 96 °C (6 sec), T$_{ann.}$=55 °C (15 sec), 72 °C (1 min) . Other exons (11-22) required nested reamplification after the primary PCR reaction. In these cases the initial amplification was carried out with the primers in the first two columns of Table 2 for 19 cycles as described above. Nested reamplification for these exons was carried out for 28 or 32 cycles at the same conditions with the primers appearing in the third column of Table 2. The products were purified from 0.8% agarose gels using Qiaex columns (Qiagen). The purified products were analysed by cycle sequencing with α-p^{32}dATP with AmpliCycle^{TM} Sequencing Kit (Perkin Elmer, Branchburg, New Jersey). The reaction products were fractionated on 6% polyacrylamide gels. All (A) reactions were loaded adjacent to each other, followed by the (C) reactions, and so on. Detection of polymorphisms was carried out visually and confirmed on the other strand.

1. Kelsey, J.L. & Gammon, M.D. The epidemiology of breast cancer. *CA Cancer J. Clinic.* **41**, 147-165 (1991).
2. Hall, J.M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684-1689 (1990).
3. Miki, Y. *et al.* A strong candidated for the 17q-linked breast and ovarian cancer susceptibility gene *BRCA1*. *Science* **266**, 66-71 (1994).
4. Wooster, R. *et al.* Localization of a breast cancer susceptibility gene, *BRCA2*, to chromosome 13q12-13 *Science* **265**, 2088-2090 (1994).
5. Wooster, R. *et al.* Identification of the breast cancer gene *BRCA2*. *Nature* **378**, 789-791 (1995).
6. Schutte, M. *et al.* Identification by representational difference analysis of a homozygous deletion in pancreatic carcinoma that lies within the *BRCA2* region. *Proc. Natl. Acad. Sci. USA* **92**, 5950-5954 (1995).
7. Stone, S. *et al.* Complex structure and regulation of the P16 (*MTS1*) locus. *Cancer Res.* **55**, 2988-2994 (1995).
8. Simard, J. *et al.* Generation of a transcription map of the *BRCA2* region. *Am. J. Hum. Genet.*, **57**, (suppl) 1566 (1995).
9. Friedman, L.S. *et al.* Novel inherited mutations and variable expressivity of *BRCA1* alleles including the founder mutation 185 del AG in Ashkenazi Jewish families. *Am. J. Hum. Genet.* **57**, 1284-1297 (1995).
10. Szabo, C.I. & King, M-C. Inherited breast and ovarian cancer. *Hum. Mol. Genet.* **4**, 1811-1817 (1995).
11. Shattuck-Eidens, D. *et al.* A comparative survey of 80 mutations in the *BRCA1* breast and ovarian cancer susceptibility gene: Implication for presymptomatic testing and screening. *JAMA* **273**, 535-541 (1995).
12. Parimoo, S. *et al.* cDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci. USA* **88**, 9623-9627 (1991).
13. Rommens, J.M. *et al.* Towards a transcriptional map of the q21-q22 region of chromosome 7. *In* 'Identification of transcribed sequences', (eds Hochgeschwender, U. & Gardiner, K.) pp. 65-79 (Plenum Press, New York, 1994).
14. Lovett, M., Kere, J. & Hinton, L.M. Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. USA* **88**, 9628-9632 (1991).
15. Hattier, T. *et al.* Analysis of hybrid selection in the search for BRCA1. *Mamm. Genome* **6**, 873–879 (1995).
16. Church, D.M. *et al.* Identification of human chromosome 9 specific genes using exon amplification. *Hum. Mol. Genet.* **2**, 1915 (1993).
17. Kamb, A. *et al.* Analysis of the p16 gene (*CDKN2*) as a candidate for the chromosome 9p melanoma susceptibility locus. *Nature Genet.* **8**, 22–29 (1994).