A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1
Author(s): Yoshio Miki, Jeff Swensen, Donna Shattuck-Eidens, P. Andrew Futreal, Keith Harshman, Sean Tavtigian, Qingyun Liu, Charles Cochran, L. Michelle Bennett, Wei Ding, Russell Bell, Judith Rosenthal, Charles Hussey, Thanh Tran, Melody McClure, Cheryl Frye, Tom Hattier, Robert Phelps, Astrid Haugen-Strano, Harold Katcher, Kazuko Yakumo, Zahra Gholami, Daniel Shaffer, Steven Stone, Steven Bayer, Christian Wray, Robert Bogden, Pri ...
Source: *Science*, New Series, Vol. 266, No. 5182 (Oct. 7, 1994), pp. 66-71
Published by: American Association for the Advancement of Science
Stable URL: http://www.jstor.org/stable/2884716
Accessed: 14/03/2009 21:09

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=aaas.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

American Association for the Advancement of Science is collaborating with JSTOR to digitize, preserve and extend access to *Science*.

http://www.jstor.org

A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene *BRCA1*

Yoshio Miki, Jeff Swensen, Donna Shattuck-Eidens, P. Andrew Futreal, Keith Harshman, Sean Tavtigian, Qingyun Liu, Charles Cochran, L. Michelle Bennett, Wei Ding, Russell Bell, Judith Rosenthal, Charles Hussey, Thanh Tran, Melody McClure, Cheryl Frye, Tom Hattier, Robert Phelps, Astrid Haugen-Strano, Harold Katcher, Kazuko Yakumo, Zahra Gholami, Daniel Shaffer, Steven Stone, Steven Bayer, Christian Wray, Robert Bogden, Priya Dayananth, John Ward, Patricia Tonin, Steven Narod, Pam K. Bristow, Frank H. Norris, Leah Helvering, Paul Morrison, Paul Rosteck, Mei Lai, J. Carl Barrett, Cathryn Lewis, Susan Neuhausen, Lisa Cannon-Albright, David Goldgar, Roger Wiseman, Alexander Kamb, Mark H. Skolnick*

A strong candidate for the 17q-linked *BRCA1* gene, which influences susceptibility to breast and ovarian cancer, has been identified by positional cloning methods. Probable predisposing mutations have been detected in five of eight kindreds presumed to segregate *BRCA1* susceptibility alleles. The mutations include an 11–base pair deletion, a 1–base pair insertion, a stop codon, a missense substitution, and an inferred regulatory mutation. The *BRCA1* gene is expressed in numerous tissues, including breast and ovary, and encodes a predicted protein of 1863 amino acids. This protein contains a zinc finger domain in its amino-terminal region, but is otherwise unrelated to previously described proteins. Identification of *BRCA1* should facilitate early diagnosis of breast and ovarian cancer susceptibility in some individuals as well as a better understanding of breast cancer biology.

Breast cancer is one of the most common and important diseases affecting women. Current estimates indicate that one in eight American women who reach age 95 will develop breast cancer (*1*). Treatment of advanced breast cancer 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.

Y. Miki, J. Swensen, K. Yakumo, C. Lewis, S. Neuhausen, and D. Goldgar are in the Department of Medical Informatics, University of Utah Medical Center, Salt Lake City, UT 84132, USA. D. Shattuck-Eidens, K. Harshman, S. Tavtigian, Q. Liu, W. Ding, R. Bell, J. Rosenthal, C. Hussey, T. Tran, M. McClure, C. Frye, T. Hattier, R. Phelps, H. Katcher, Z. Gholami, D. Shaffer, S. Stone, S. Bayer, C. Wray, R. Bogden, P. Dayananth, and A. Kamb are at Myriad Genetics, 421 Wakara Way, Salt Lake City, UT 84108, USA. P. A. Futreal, C. Cochran, L. M. Bennett, A. Huagen-Strano, J. C. Barrett, and R. Wiseman are at the Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA. J. Ward and L. Cannon-Albright are in the Department of Internal Medicine, University of Utah Medical Center, Salt Lake City, UT 84132, USA. P. Tonin and S. Narod are in the Department of Medical Genetics, McGill University, Montreal, Quebec, H3G 1A4, Canada. P. K. Bristow, F. H. Norris, L. Helvering, P. Morrison, P. Rosteck, and M. Lai are at Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN 46285, USA. M. H. Skolnick is in the Department of Medical Informatics, University of Utah Medical Center, and Myriad Genetics, Salt Lake City, UT 84108, USA.

*To whom correspondence should be addressed.

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 30 (*2*). Breast cancer has been subdivided into two types, early-onset and late-onset, a division that is 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 families with significantly high breast cancer incidence and at least 80% of families with increased incidence of both early-onset breast cancer and ovarian cancer (*3*). Intense efforts to isolate the *BRCA1* gene have proceeded since it was first mapped to chromosome arm 17q in 1990 (*4, 5*). A second locus, *BRCA2*, recently mapped to chromosome arm 13q (*6*), appears to account for a proportion of early-onset breast cancer roughly equal to that resulting from *BRCA1*. Unlike *BRCA1*, however, *BRCA2* may not influence ovarian cancer risk. The remaining susceptibility to early-onset breast cancer is likely attributable to unmapped genes for familial cancer and rare germline mutations in genes such as *TP53*, which encodes the tumor suppressor protein p53 (*7*). It has also been suggested that heterozygote carriers of defective forms of the gene predisposing to ataxia telangiectasia are at higher risk for breast cancer (*8, 9*). Late-onset breast cancer is often familial in origin, although the risks in relatives are not as high as those for early-onset breast cancer (*10, 11*). The percentage of such cases that are due to genetic susceptibility is unknown.

Like many other genes involved in familial cancer, *BRCA1* appears to encode a tumor suppressor, a protein that acts as a negative regulator of tumor growth. Cancer-predisposing alleles typically carry mutations that cause loss or reduction of gene function. Predisposition to cancer is inherited as a dominant genetic trait, whereas the predisposing allele generally behaves as a recessive allele in somatic cells. Thus, a single inherited copy of the mutant allele causes predisposition, and loss or inactivation of the wild-type allele completes one of the steps in progression toward malignancy. When chromosome loss is observed in breast and ovarian tumors from patients who carry *BRCA1* predisposing alleles, the wild-type copy of *BRCA1* is invariably lost while the presumptive mutant allele is retained (*12–14*). This finding supports the hypothesis that *BRCA1* is a tumor suppressor gene and suggests that the functional BRCA1 protein is present in normal breast and ovarian epithelium tissue and is altered, reduced, or absent in some breast and ovarian tumors.

Genetic analysis of recombinant chromosomes in members of large kindreds allowed localization of *BRCA1* initially to a region of 1 to 2 megabases on chromosome 17q (*15–17*) and, subsequently, to a region of about 600 kilobase pairs (kb) (*18*) between markers *D17S1321* and *D17S1325* (*19*). A physical map comprised of overlapping yeast artificial chromosomes (YACs), P1, bacterial artificial chromosomes (BACs), and cosmid clones was generated for this region (*18*).

Identification of a strong BRCA1 candidate gene. Several strategies were used to develop a detailed map of transcripts for the 600-kb region of 17q21 between *D17S1321* and *D17S1325*. Sixty-five candidate expressed sequences (*20*) within this region were identified. Expressed sequences were characterized by DNA sequence, database comparison, transcript size, expression pattern, genomic structure and, most importantly, DNA sequence analysis in individuals from kindreds that segregate 17q-linked breast and ovarian cancer susceptibility. Three expressed sequences eventually were merged into a single transcription unit whose characteristics strongly suggest that it is *BRCA1* (*21*). This transcription unit is located in the center of the 600-kb region (Fig. 1) spanning *D17S855* and will be referred to herein as *BRCA1*.

A combination of sequences obtained from complementary DNA (cDNA) clones, hybrid-selected sequences, and amplified polymerase chain reaction (PCR) products

allowed the construction of a composite, full-length *BRCA1* cDNA. The cDNA clone extending farthest in the 3′ direction contains a polyadenylate tract preceded by a polyadenylation signal. Conceptual translation of the cDNA revealed a single, long open reading frame with a presumptive initiation codon flanked by sequences resembling the Kozak consensus sequence (*22*). This reading frame encodes a protein of 1863 amino acids (Fig. 2A). Smith-Waterman (*23*) and BLAST (*24*) searches identified a sequence near the NH_2-terminus that has considerable similarity to zinc finger domains (*25*) (Fig. 2B). This sequence contains cystine and histidine residues present in the consensus Cys_3-His-Cys_4 (C3HC4) zinc finger motif and shares many other residues with zinc finger proteins in the databases. The *BRCA1* gene is composed of 22 coding exons distributed over roughly 100 kb of genomic DNA (Fig. 3).

Hybridization of RNA blots to labeled fragments of *BRCA1* cDNA revealed a single transcript of 7.8 kb. This transcript is most abundant in testis and thymus, but is also present in breast and ovary (Fig. 4). The cDNA clones derived from the 5′ one-third of *BRCA1* transcripts display a complex pattern of alternative splicing. Four alternative splices were observed downstream of the start codon as independent cDNA clones (P3, P4, B31, and B21 in Fig. 3); three of these patterns were detected in breast cDNA (P3, B31, and B21) and two in ovary cDNA (P3 and B21). In addition, PCR analysis of cDNA samples prepared from breast, ovary, testis, and lymphocyte messenger RNA (mRNA) indicates that there is considerable heterogeneity in splice junction usage near the 5′ end of *BRCA1* transcripts, upstream of the presumptive initiation codon. How this alternative splicing is coordinated with alternative splicing farther downstream, and whether all the splice variants produce proteins with an identical NH_2-terminus, are questions that remain to be explored.

We also probed genomic DNA samples from several different species with *BRCA1* sequences devoid of the zinc finger region. Low-stringency blots revealed strongly hybridizing fragments in tissues from humans, mice, rats, rabbits, sheep, and pigs, but not chickens (Fig. 5). These results suggest that *BRCA1* is conserved in mammals.

Germline BRCA1 mutations in 17q-linked kindreds. Identification of a candidate gene as *BRCA1* requires a demonstration of potentially disruptive mutations in that gene 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 wild-type sequence. The set of DNA samples used in this analysis consisted of DNA from individuals represent-

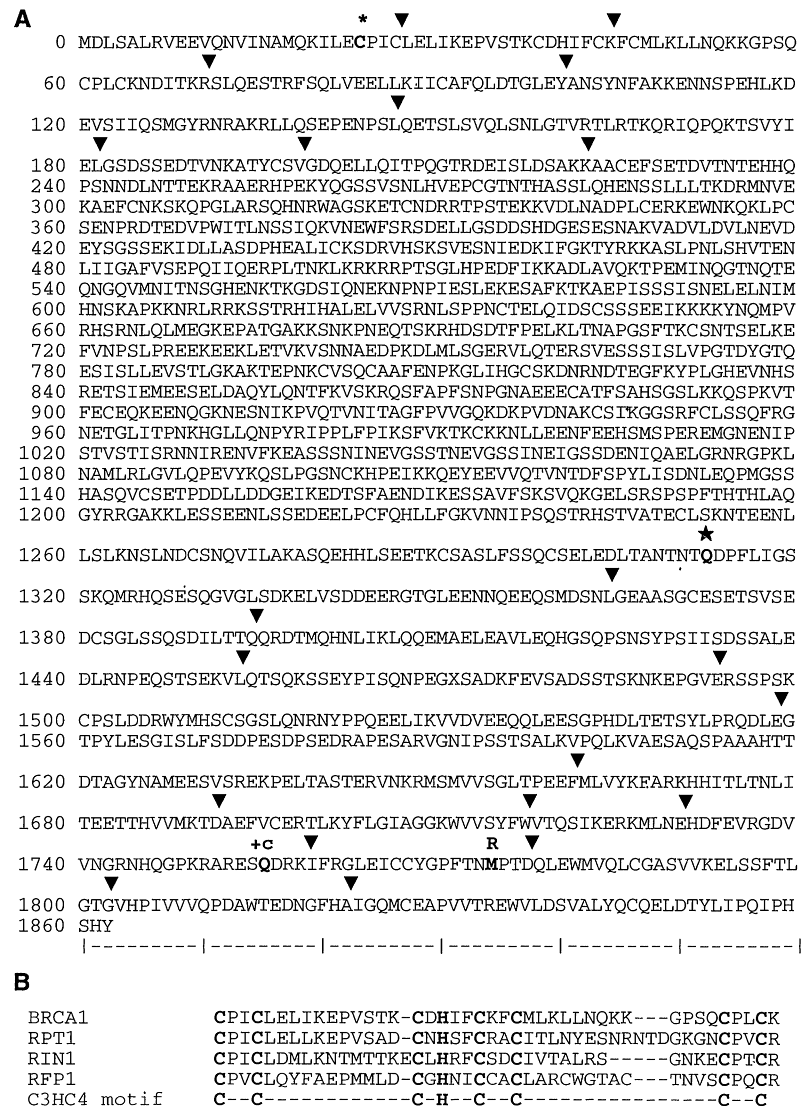

Fig. 2. Predicted amino acid sequences for BRCA1 (*46*). (**A**) Conceptual translation of the *BRCA1* open reading frame, indicating the approximate positions of introns (triangles above sequence) and the locations of germline mutations (boldface residues). The −11 bp deletion in kindred 1901 is shown by an asterisk; the nonsense mutation in kindred 2082 is shown by a star; the frameshift in kindred 1910 is shown by "+c"; and the missense mutation in kindred 2099 is shown by "R". The BRCA1 nucleotide sequence is deposited in GenBank with accession number U14680. PCR primer sequences are available via anonymous FTP at the following internet address: morgan.med.utah.edu in the directory pub/BRCA1; or by fax at the following number: 801-584-3650. (**B**) Alignment of the *BRCA1* zinc finger domain with three other zinc finger domains that scored highest in a Smith-Waterman alignment. RPT1 is a protein that appears to be a negative regulator of the interleukin-2 receptor in mice (*47*). RIN1 is a DNA binding protein that includes a RING finger motif related to the zinc finger (*48*). RFP1 is a putative transcription factor comprising the NH_2-terminal domain of the *RET* oncogene product (*49*). The C3HC4 motif shows the positions of the cystines and the histidine that form the zinc binding pockets.

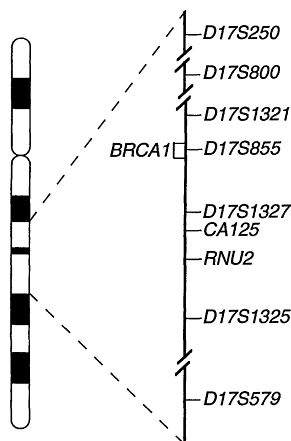

Fig. 1. Schematic map of human chromosome 17. The pertinent region containing *BRCA1* is expanded to indicate the relative positions of two previously identified genes, *CA125* (*34*) and *RNU2* (*45*). *D17S855* is located within *BRCA1*.

ing eight different *BRCA1* kindreds (Table 1). The lod scores (likelihood ratios for linkage) 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 four have low positive or negative lod scores. The latter kindreds were included because they demonstrate haplotype sharing at chromosome 17q21 for at least three affected members. Furthermore, all kindreds in the set display early-onset breast cancer, and four of the kindreds include at least one case of ovarian cancer, both hallmarks of *BRCA1* kindreds. Kindred 2082 has nearly equal incidence of breast and ovarian cancer, an unusual occurrence given the relative rarity of ovarian cancer in the population (*17*). All but two of the kindreds were ascertained in Utah. Kindred 2035 is from the midwestern United States. Kindred 2099 is an African American kindred from the southern United States; all other kindreds are Caucasian.

In the initial screen for predisposing mutations in *BRCA1*, DNA from one individual carrying the predisposing haplotype from each kindred was tested. The 21 coding exons and associated splice junctions were amplified from either genomic DNA samples or cDNA prepared from lymphocyte mRNA (*26*). When the amplified DNA sequences were compared to the wild-type sequence, four of the eight kindred samples were found to contain sequence variants (Table 2). All four sequence variants are heterozygous, and each appears in only one of the kindreds. Kindred 1901 contains an 11–base pair (bp) deletion in exon 2 (Cys24 frameshift to 36 Stop). Kindred 2082 contains a nonsense mutation in coding exon 11 (Gln1313 to Stop) (Fig. 6A). Kindred 1910 contains a single nucleotide insertion in coding exon 20 (Gln1756 frameshift to 1829 Stop) (Fig. 6B), and kindred 2099 contains a missense mutation in coding exon 21 (Met1775Arg). The frameshift and nonsense mutations are likely to disrupt the function of the BRCA1 proteins. The protein encoded by the insertion allele in kindred 1910 would contain an altered sequence beginning 107 amino acids residues from the wild-type COOH-terminus. The effect of the 11-bp deletion in kindred 1901 would be even more dramatic because it occurs at the twenty-fourth codon. This deletion removes the last 11 bp of exon 2 and begins at the first cystine of the zinc finger motif, thereby removing the zinc finger domain. The mutant allele in kindred 2082 would encode a protein missing 548 residues from the COOH-terminus. The missense mutation observed in kindred 2099 is potentially disruptive as it substitutes a large, charged amino acid (Arg) for a small, hydrophobic amino acid (Met). Five common polymorphisms were also identified in the *BRCA1* coding sequence (Table 3).

The individual studied in kindred 2035 is likely to carry a regulatory mutation in *BRCA1*. In her cDNA, two polymorphic sites (PM1 and PM7) appeared homozygous, whereas her genomic DNA revealed heterozygosity at these positions (Fig. 6C). One possible explanation for this observation is that mRNA from her mutant *BRCA1* allele is absent because of a mutation that affects RNA production or stability. We explored this possibility further by examining three additional polymorphic sites (PM6, PM7, and PM2) in the *BRCA1*

Table 1. Kindred descriptions and associated lod scores (*50*). Br, breast cancer; Br <50, breast cancer diagnosed under age 50; Ov, ovarian cancer.

Kindred	Cases (*n*)			Sporadic cases† (*n*)	Lod score	Markers
	Total Br	Br <50	Ov			
2082	31	20	22	7	9.49	*D17S1327*
2099	22	14	2*	0	2.36	*D17S800* and *D17S855*‡
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* and *D17S250*‡
1911	8	5	0	1	−0.20	*D17S250*
1927	5	4	0	1	−0.44	*D17S250*

*Kindred contains one individual who had both breast and ovarian cancer; this individual is counted as both a breast cancer case and as an ovarian cancer case. †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. ‡Both markers were used to calculate multipoint lod scores.

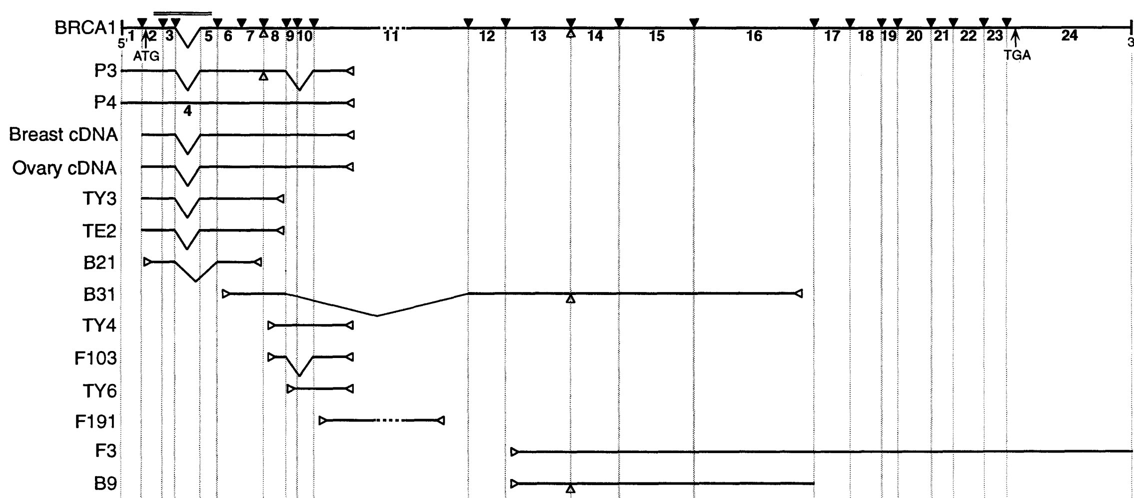

Fig. 3. Diagram of *BRCA1* mRNA, showing the locations of introns and the variants of *BRCA1* mRNA produced by alternative splicing. The top cDNA (*BRCA1*) is the composite used to generate the protein sequence in Fig. 2. Intron locations are shown by filled triangles, and the exons are numbered below the composite cDNA. Alternative mRNAs identified as cDNA clones or in hybrid-selection experiments are shown below the composite. The start codon (ATG) and stop codon (TGA) are indicated. The zinc finger region is denoted noted by a double line. "V" lines connecting exons indicate the absence of internal exons. All exons are drawn proportionally except exon 11 (indicated with a dotted line). Upward-pointing unfilled triangles show the position of a single codon (CAG) found at the start of exons 8 and 14 in some cDNAs. Leftward- and rightward-pointing unfilled triangles represent partial exons in some cDNAs. P3 and P4 are cDNA clones isolated from a placental cDNA library; TY3 and TE2 are 5′ RACE clones from thymus and testis, respectively; B21 and B9 are cDNA clones from a normal breast cDNA library; B31 is a hybrid-selected cDNA clone from breast cDNA; TY4 and TY6 are cDNA clones isolated from a thymus cDNA library; and F191, F103, and F3 are cDNA clones isolated from a fetal brain library. The *BRCA1* variants labeled breast cDNA and ovary cDNA are the major forms detected in these tissues by PCR.

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, her cDNA appeared homozygous. In individuals from other kindreds and in nonhaplotype carriers in kindred 2035, these polymorphic sites appeared 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 potential disruption of protein function, a sequence variant must meet two other criteria to qualify as a candidate predisposing mutation: It must be present in members of the kindred who carry the predisposing haplotype and absent from other members of the kindred, and it must be rare in the general population.

To test for cosegregation of mutations with the corresponding *BRCA1* susceptibility allele, we screened several individuals from kindreds 1901, 1910, 2082, and 2099, including both carriers and noncarriers of the predisposing haplotype (Fig. 6). In each kindred, the corresponding mutant allele was detected only in individuals carrying the *BRCA1*-associated haplotype. In the case of the potential regulatory mutation in 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.

To exclude the possibility that the mutations were simply common polymorphisms in the population, we used allele-specific oligonucleotides (ASOs) for each mutation to screen a set of control DNA samples (*27*). The actual mutation in kindred 2035 has not been identified, so we could not determine its frequency in the general population. Gene frequency estimates in Caucasians were based on random samples from the Utah population. Gene frequency estimates in African Americans were based on 39 samples used in linkage studies and on samples from 20 African Americans from Utah (*28*). None of the four potential predisposing mutations tested was found in the appropriate control population, indicating that they are rare in the general population (Table 2). Thus, both important requirements for *BRCA1* susceptibility alleles are fulfilled by the candidate predisposing mutations: cosegregation of the mutant allele with diseases and absence of the mutant allele in controls, indicating a low frequency in the general population (*29*).

Phenotypic expression of *BRCA1* mutations. The effect of the mutations on the BRCA1 protein correlates 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 a high risk of ovarian cancer comparable to that for breast cancer (*3*). Four of the five kindreds in which *BRCA1* mutations were detected fall into the former category, and the fifth (kindred 2082) falls into the group with high ovarian cancer risk. The *BRCA1* nonsense mutation found in kindred 2082 has an interesting phenotype. Kindred 2082 has a high incidence of ovarian cancer, and the mean age of breast cancer diagnosis is older than that in the other kindreds (*17*). 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 four kindreds that segregate known *BRCA1* mutations have, on average, 1 ovarian cancer for every 10 cases of breast cancer, but have a high proportion of breast cancer cases diagnosed at an early age (late 20s or early 30s). Kindred 1910, which has a 1-bp insertion 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 the potential regulatory *BRCA1* mutation, might also be expected to have a

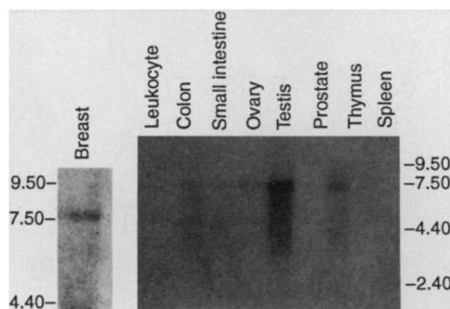

Fig. 4. Tissue expression pattern of *BRCA1*. The blots were obtained from Clontech (Palo Alto, CA) and contain RNA from the indicated tissues. Hybridization conditions were those recommended by the manufacturer, and the probe was a *BRCA1* cDNA fragment corresponding to nucleotides 3575 to 3874. Note that these tissues are heterogeneous and the percentage of relevant epithelial cells in breast and ovary can be variable. Size standards are in kilobases.

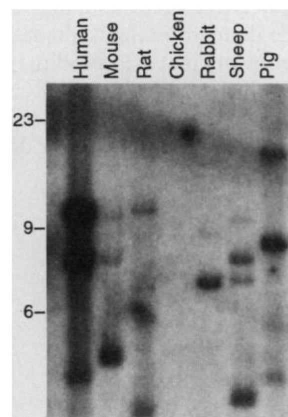

Fig. 5. Blot showing hybridization of a *BRCA1* probe to genomic DNA fragments from various species. DNA was digested with Eco RI, subjected to electrophoresis through a 0.65% agarose gel, and transferred to a nylon membrane, which was then hybridized (*32*) to a probe consisting of random-primed, α-^{32}P–labeled *BRCA1* cDNA sequences comprising a total of 4.6 kb. The probe excluded the zinc finger region. The final wash was at 55°C in ×2 SSPE and 1% SDS for 20 min. Size standards are in kilobases.

Table 2. Predisposing mutations in *BRCA1*. NA indicates not applicable, as the regulatory mutation is inferred and the position has not been identified.

Kindred	Codon	Mutation		Frequency in control chromosomes
		Nucleotide change	Coding effect	
1901	24	−11 bp	Frameshift or splice	0/180
2082	1313	C→T	Gln→Stop	0/170
1910	1756	Extra C	Frameshift	0/162
2099	1775	T→G	Met→Arg	0/120
2035	NA	?	Loss of transcript	NA

Table 3. Neutral polymorphisms in *BRCA1*. For the frequency in control chromosomes, the number of chromosomes with a particular base at the indicated polymorphic site is shown (A, C, G, or T).

Name	Codon location	Base in codon	Frequency in control chromosomes			
			A	C	G	T
PM1	317	2	152	0	10	0
PM6	878	2	0	55	0	100
PM7	1190	2	109	0	53	0
PM2	1443	3	0	115	0	58
PM3	1619	1	116	0	52	0

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 that this *BRCA1* mutant allele has a high penetrance (Table 1). Kindred 1901 displays a phenotypic pattern similar to that of kindred 2035. It is likely that the 11-bp deletion beginning at codon 24 carried in kindred 1901 results in a loss of gene function similar to the effect of the regulatory mutation in kindred 2035.

Although the mutations described in this research article are clearly deleterious, causing breast cancer in women at very young ages, each of the four kindreds with mutations includes at least one woman who carried the mutation but lived until age 80 without developing a malignancy. It will be of utmost importance in future studies to identify other genetic factors or environmental factors that may ameliorate the effects of *BRCA1* mutations. In addition, in three of the eight putative *BRCA1*-linked kindreds, potential predisposing mutations were not found. All of these kindreds have lod scores for *BRCA1*-linked markers of less than 0.55 and thus may not truly segregate *BRCA1*-predisposing alleles. Alternatively, the mutations in these three kindreds may lie in noncoding regions of *BRCA1* and therefore have escaped detection.

The role of BRCA1 in cancer. Most mutant tumor suppressor genes identified to date encode proteins 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 wild-type product (*30*, *31*). A similar dominant-negative mechanism of action has been proposed for some adenomatous polyposis coli (*APC*) alleles that produce truncated molecules (*32*) and for point mutations in the Wilms tumor gene (*WT1*), which alter DNA binding of the WT1 protein (*33*). The nature of three mutations observed in the *BRCA1* coding sequence is consistent with production of either dominant-negative proteins or nonfunctional proteins. All three mutations are located in the COOH-terminal half of the protein. The regulatory mutation inferred in kindred 2035 cannot be dominant-negative; rather, this mutation likely causes reduction or complete loss of *BRCA1* expression from the affected allele. Similarly, the 11-bp deletion in kindred 1901 likely produces a nonfunctional product.

The *BRCA1* protein contains a C3HC4 zinc finger domain similar to domains found in numerous nucleic acid binding proteins. 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 COOH-terminus. Thus, one possibility is that *BRCA1* encodes a transcription factor with an NH_2-terminal DNA binding domain and a COOH-terminal "acidic blob" domain with transactivational activity. Interestingly, the product of another familial tumor suppressor gene, *WT1*, also contains zinc finger domains (*34*), and these are altered by many cancer-predisposing mutations in the gene (*33–35*). The *WT1* gene encodes a transcription factor, and alternative splicing of exons that encode parts of the zinc finger domains alters the DNA binding properties of WT1 (*36*). Some alternatively spliced forms of *WT1* mRNA generate WT1 proteins that act as transcriptional repressors (*37*). Differential splicing of *BRCA1* may alter the zinc finger motif (Fig. 3), raising the possibility that a regulatory mechanism similar to that occurring in *WT1* may apply to *BRCA1*.

The identification of a gene that (i) falls within the interval known from genetic studies to include *BRCA1* and (ii) contains frameshift, nonsense, and regulatory mutations that cosegregate with predisposing *BRCA1* alleles strongly indicates that this gene is *BRCA1*. The observation of potential predisposing mutations in individuals

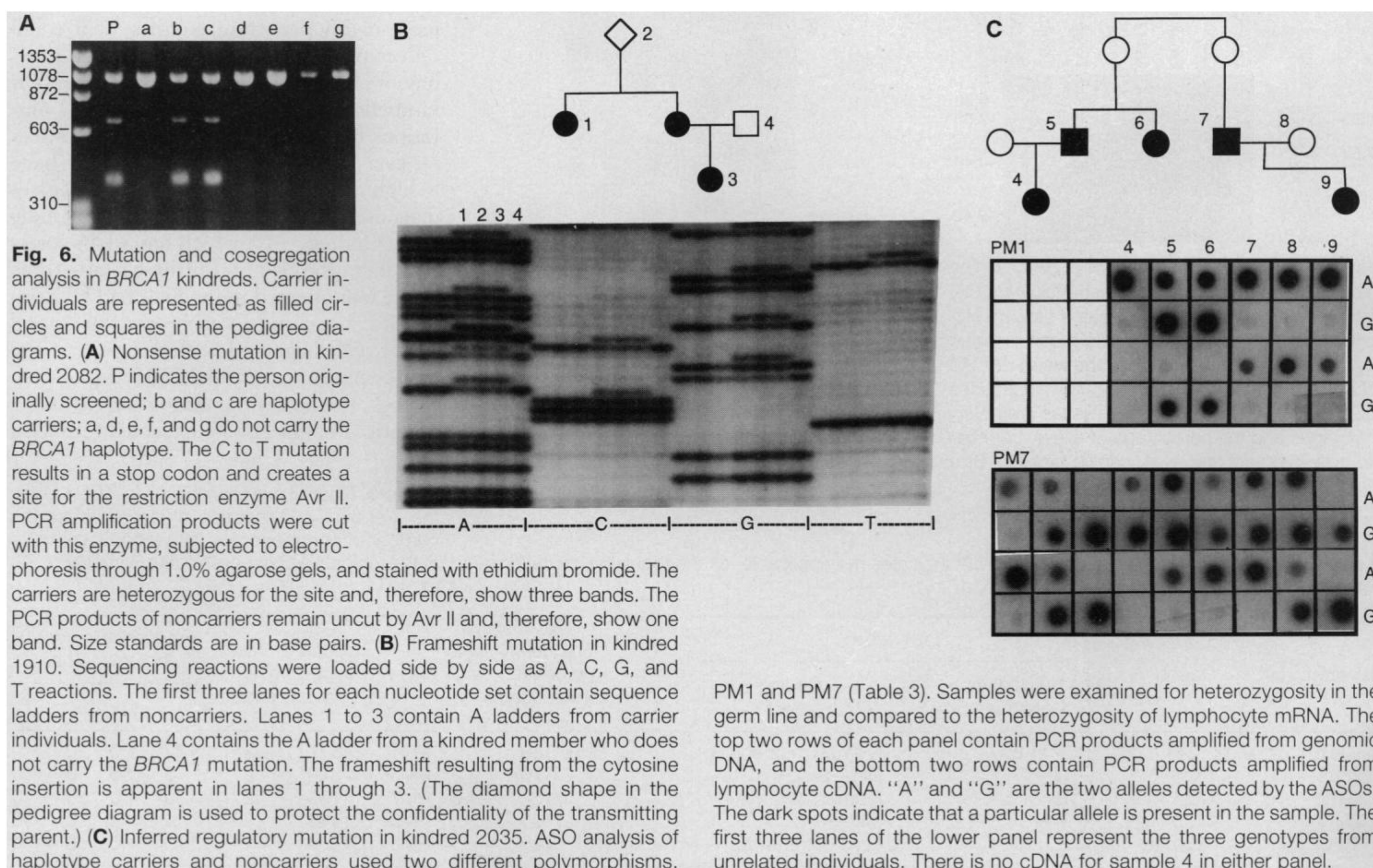

Fig. 6. Mutation and cosegregation analysis in *BRCA1* kindreds. Carrier individuals are represented as filled circles and squares in the pedigree diagrams. (**A**) Nonsense mutation in kindred 2082. P indicates the person originally screened; b and c are haplotype carriers; a, d, e, f, and g do not carry the *BRCA1* haplotype. The C to T mutation results in a stop codon and creates a site for the restriction enzyme Avr II. PCR amplification products were cut with this enzyme, subjected to electrophoresis through 1.0% agarose gels, and stained with ethidium bromide. The carriers are heterozygous for the site and, therefore, show three bands. The PCR products of noncarriers remain uncut by Avr II and, therefore, show one band. Size standards are in base pairs. (**B**) Frameshift mutation in kindred 1910. Sequencing reactions were loaded side by side as A, C, G, and T reactions. The first three lanes for each nucleotide set contain sequence ladders from noncarriers. Lanes 1 to 3 contain A ladders from carrier individuals. Lane 4 contains the A ladder from a kindred member who does not carry the *BRCA1* mutation. The frameshift resulting from the cytosine insertion is apparent in lanes 1 through 3. (The diamond shape in the pedigree diagram is used to protect the confidentiality of the transmitting parent.) (**C**) Inferred regulatory mutation in kindred 2035. ASO analysis of haplotype carriers and noncarriers used two different polymorphisms, PM1 and PM7 (Table 3). Samples were examined for heterozygosity in the germ line and compared to the heterozygosity of lymphocyte mRNA. The top two rows of each panel contain PCR products amplified from genomic DNA, and the bottom two rows contain PCR products amplified from lymphocyte cDNA. "A" and "G" are the two alleles detected by the ASOs. The dark spots indicate that a particular allele is present in the sample. The first three lanes of the lower panel represent the three genotypes from unrelated individuals. There is no cDNA for sample 4 in either panel.

whose early-onset breast or ovarian cancer was not ascertained by family history supports the view that many early-onset cases are due to mutations at the *BRCA1* locus (*38*). The role of *BRCA1* in cancer progression may now be addressed with molecular precision. The large size and fragmented nature of the coding sequence will make exhaustive searches for new mutations challenging. Nevertheless, the percentage of total breast and ovarian cancer caused by mutant *BRCA1* alleles will soon be estimated, and individual mutation frequencies and penetrances may be established. This in turn may permit accurate genetic screening for predisposition to a common, deadly disease. Although such research represents an advance in medical and biological knowledge, it also raises numerous ethical and practical issues, both scientific and social, that must be addressed by the medical community.

Note added in proof: Analysis of kindred 1911 indicates a possible linkage to *BRCA2*, suggesting that early breast cancer in this kindred is not due to a mutation of *BRCA1* (*51*).

REFERENCES AND NOTES

1. American Cancer Society, *Cancer Facts & Figures 1994* (American Cancer Society, Atlanta, GA, 1994), p. 13.
2. E. B. Claus, W. D. Thompson, N. Risch, *Am. J. Hum. Genet.* **48**, 232 (1991).
3. D. F. Easton, D. T. Bishop, D. Ford, B. P. Crockford, Breast Cancer Linkage Consortium, *ibid.* **52**, 678 (1993).
4. J. M. Hall *et al.*, *Science* **250**, 1684 (1990).
5. S. A. Narod *et al.*, *Lancet* **338**, 82 (1991).
6. R. Wooster *et al.*, *Science* **265**, 2088 (1994).
7. D. Malkin *et al.*, *ibid.* **250**, 1233 (1990).
8. M. Swift, L. Sholman, M. Perry, C. Chase, *Cancer Res.* **36**, 209 (1976).
9. M. Swift, P. J. Reitnauer, D. Morrell, C. L. Chase, *N. Engl. J. Med.* **325**, 1831 (1991).
10. L. Cannon-Albright *et al.*, *Cancer Res.* **54**, 2378 (1994).
11. C. Mettlin, I. Croghan, N. Natarajan, W. Lane, *Am. J. Epidemiol.* **131**, 973 (1990).
12. H. S. Smith *et al.*, *J. Cell. Biochem. Suppl.* **14G**, 144 (1993).
13. D. P. Kelsell, D. M. Black, D. T. Bishop, N. K. Spurr, *Hum. Mol. Genet.* **2**, 1823 (1993).
14. S. Neuhausen *et al.*, *Cancer Res.*, in press.
15. A. M. Bowcock *et al.*, *Am. J. Hum. Genet.* **52**, 718 (1993).
16. J. Simard *et al.*, *Hum. Mol. Genet.* **2**, 1193 (1993).
17. D. E. Goldgar *et al.*, *J. Natl. Cancer Inst.* **86**, 200 (1994).
18. S. L. Neuhausen *et al.*, *Hum. Mol. Genet.*, in press.
19. D. E. Goldgar *et al.*, unpublished results.
20. Candidate expressed sequences are defined as DNA sequences obtained by (i) direct screening of breast, fetal brain, lymphocyte, or ovary cDNAs (*39*) or (ii) random sequencing of genomic DNA (*40*) and prediction of coding exons by XPOUND (*41*). 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.
21. Three independent contigs of expressed sequences 1141:1 (649 bp), 694:5 (213 bp), and 754:2 (1079 bp) were isolated by hybrid selection and eventually shown to represent portions of *BRCA1*. When expressed sequence tags (ESTs) for 1141:1 and 754:2 were used as hybridization probes for RNA blots, a single transcript of approximately 7.8 kb was observed in normal breast mRNA, which suggested that they encode different portions of a single gene. Screens of breast, fetal brain, thymus, testis, 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, testis, and breast mRNAs 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.
22. M. Kozak, *Nucleic Acids Res.* **15**, 8125 (1987).
23. T. F. Smith and M. S. Waterman, *J. Mol. Biol.* **147**, 195 (1981).
24. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *ibid.* **215**, 195 (1990).
25. We performed database comparison by using (i) BLAST alignment algorithms (*24*) on the National Center for Biotechnology Information (NCBI) databases; (ii) Smith-Waterman alignment algorithms (*23*) on a MasPar computer to search the SwissProt database (at MasPar and at the European Molecular Biology Laboratory); and (iii) Smith-Waterman algorithms on a Compugen Biocelerator (at the Weizmann Institute) to search the GenBank (nucleotide) and SwissProt (protein) databases.
26. The templates for PCR were lymphocyte cDNA or genomic DNA from members of *BRCA1* kindreds who carried the predisposing haplotype. Sequences of PCR primers used to amplify each exon of BRCA1 are available upon request. The PCR conditions were: one cycle at 95°C (5 min); four cycles at 95°C (10 s), with the annealing temperature (T_{ann}) at 68°C for 10 s, and at 72°C for 10 s; four cycles with T_{ann} = 66°C; four cycles with T_{ann} = 64°C; four cycles with T_{ann} = 62°C; and 30 cycles with T_{ann} = 60°C. The buffer conditions were as described (*42*). The PCR products were purified from 1.0% agarose gels with Qiaex beads (QIAGEN), analyzed by cycle sequencing with [α-^{32}P]deoxy- adenosine triphosphate (*43*), and subjected to electrophoresis on 6% polyacrylamide gels. Polymorphisms were initially detected by eye on both strands and subsequently confirmed by ASO analysis (*27*).
27. PCR products were generated as described (*26*) and quantified after electrophoresis through 2% agarose gels containing ethidium bromide by comparison with DNA standards. PCR product (10 μl) was added to 110 μl of denaturant (7.5 ml of H_2O, 6.0 ml of 1 N NaOH, 1.5 ml of 0.1% bromophenol blue, and 75 ml of 0.5 mM EDTA) and incubated for 10 min at room temperature. Samples (30 μl) were then blotted onto Hybond membrane (Amersham) with a dot-blotting apparatus (Gibco-BRL). The DNA was fixed on the membrane by exposure to ultraviolet light (Stratagene). Prehybridization was carried out at 45°C in ×5 SSPE (0.75 M NaCl, 0.05 M $NaH_2Po_4 \cdot H_2O$, and 0.005 M EDTA) and 2% SDS (*34*). Wild-type and mutant ASOs were labeled by incubation at 37°C for 10 min in a reaction that included 5 μCi of [^{32}P]adenosine triphosphate, 100 ng of ASO, 10 units of T4 polynucleotide kinase (Boehringer-Mannheim), and kinase buffer (*44*). Labeled ASO (20 ng) was used in an overnight hybridization reaction in the same buffer used for prehybridization. Each blot was washed twice in ×5 saline sodium citrate and 0.1% SDS for 10 min at room temperature, and then for 30 min at progressively higher temperatures until nonspecific hybridization signals were eliminated. Blots were exposed to x-ray film for 40 min without an intensifying screen.
28. The African American samples from Utah were from a newborn screening program.
29. A reviewer suggested the possibility that this gene, which we call *BRCA1*, may contain frameshift mutations other than those detected here at a significant frequency in members of the general population.
30. E. Shaulian, A. Zauberman, D. Ginsberg, M. Oren, *Mol. Cell. Biol.* **12**, 5581 (1992).
31. S. Srivastava, S. Wang, Y. A. Tong, Z. M. Hao, E. H. Chang, *Cancer Res.* **53**, 4452 (1993).
32. L. K. Su *et al.*, *ibid.*, p. 2728.
33. M. H. Little *et al.*, *Hum. Mol. Genet.* **2**, 259 (1993).
34. D. A. Haber *et al.*, *Cell* **61**, 1257 (1990).
35. M. H. Little *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4791 (1992).
36. W. A. Bickmore *et al.*, *Science* **257**, 235 (1992).
37. I. A. Drummond *et al.*, *Mol. Cell. Biol.* **14**, 3800 (1994).
38. P. A. Futreal *et al.*, *Science* **266**, 120 (1994).
39. M. Lovett, J. Kere, L. M. Hinton, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9628 (1991); P. A. Futreal *et al.*, *Hum. Mol. Genet.* **3**, 1359 (1994).
40. A. Kamb *et al.*, *Nature Genet.* **8**, 22 (1994).
41. A. Thomas and M. H. Skolnick, *IMA J. Math. Appl. Med. Biol.*, in press.
42. J. Weaver-Feldhaus *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7563 (1994).
43. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, ed. 2, 1989).
44. I. G. Campbell *et al.*, *Hum. Mol. Genet.* **3**, 589 (1994).
45. G. Westin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3811 (1984).
46. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
47. R. Patarca *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2733 (1988).
48. R. Lovering *et al.*, *ibid.* **90**, 2112 (1993).
49. M. Takahashi, Y. Inaguma, H. Hiai, F. Hirose, *Mol. Cell. Biol.* **8**, 1853 (1988).
50. Two of the kindreds referred to here have been studied independently [J. Feunteun *et al.*, *Am. J. Hum. Genet.* **52**, 736 (1993); D. Anderson, unpublished results].
51. D. Goldgar, unpublished results.

52. We are grateful for the cooperation of the individuals from the *BRCA1* kindreds and for the assistance of our clinic coordinators P. Fields, L. Steele, M. MacDonald, and K. Brown and for the help of C. J. Marshall. We thank D. Ballinger, K. Fournier, W. Gilbert, L. Norton, G. Omen, J. Rine, J. Simard, R. Williams, and B. Wold for scientific advice; and F. Bartholomew, H. Brownlee, S. Burgett, J. Collette, B. S. Dehoff, I. L. Jenkins, A. Leavitt, K. Richardson, and K. Rowe for technical support. Linkage study controls were kindly provided by M. Pericak-Vance. Supported in part by NIH grants CA-55914 (M.H.S.), CA-54936, CA-48711, CA-42014, CN-0522, RR-00064, and HG-00571 (D.G.), the National Cancer Institute of Canada, the Canadian Genetic Diseases Network (S.N.), and the Cedars Cancer Institute of the Royal Victoria Hospital (P.T.).

2 September 1994; accepted 14 September 1994

For information about an audio conference on the topic of the breast cancer gene *BRCA1*, see page 15.