[CANCER RESEARCH 64, 3790–3797, June 1, 2004]

Structure-Based Assessment of Missense Mutations in Human BRCA1: Implications for Breast and Ovarian Cancer Predisposition

Nebojsa Mirkovic,[1] Marc A. Marti-Renom,[2] Barbara L. Weber,[3] Andrej Sali,[2] and Alvaro N. A. Monteiro[4,5]

[1]Laboratory of Molecular Biophysics, Pels Family Center for Biochemistry and Structural Biology, Rockefeller University, New York, New York; [2]Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research, University of California at San Francisco, San Francisco, California; [3]Abramson Family Cancer Research Institute, University of Pennsylvania, Philadelphia, Pennsylvania; [4]Strang Cancer Prevention Center, New York, New York; and [5]Department of Cell and Developmental Biology, Weill Medical College of Cornell University, New York, New York

ABSTRACT

The *BRCA1* gene from individuals at risk of breast and ovarian cancers can be screened for the presence of mutations. However, the cancer association of most alleles carrying missense mutations is unknown, thus creating significant problems for genetic counseling. To increase our ability to identify cancer-associated mutations in *BRCA1*, we set out to use the principles of protein three-dimensional structure as well as the correlation between the cancer-associated mutations and those that abolish transcriptional activation. Thirty-one of 37 missense mutations of known impact on the transcriptional activation function of BRCA1 are readily rationalized in structural terms. Loss-of-function mutations involve nonconservative changes in the core of the BRCA1 C-terminus (BRCT) fold or are localized in a groove that presumably forms a binding site involved in the transcriptional activation by BRCA1; mutations that do not abolish transcriptional activation are either conservative changes in the core or are on the surface outside of the putative binding site. Next, structure-based rules for predicting functional consequences of a given missense mutation were applied to 57 germ-line BRCA1 variants of unknown cancer association. Such a structure-based approach may be helpful in an integrated effort to identify mutations that predispose individuals to cancer.

INTRODUCTION

Many germ-line mutations in the human *BRCA1* gene are associated with inherited breast and ovarian cancers (1, 2). This information has allowed clinicians and genetic counselors to identify individuals at high risk for developing cancer. However, the disease association of over 350 missense mutations remains unclear, primarily because their relatively low frequency and ethnic specificity limit the usefulness of the population-based statistical approaches to identifying cancer-causing mutations. To address this problem, we use here the three-dimensional structure of the human BRCA1 BRCT domains to assess the transcriptional activation functions of BRCA1 mutants. Our study is made possible by the recently determined sequences (3–6) and three-dimensional structures of the BRCA1 homologs (7, 8). In addition, we benefited from prior studies that attempted to rationalize and predict functional effects of mutations in various proteins (9–12), including those of BRCA1 (13, 14).

BRCA1 is a nuclear protein that activates transcription and facilitates DNA damage repair (15, 16). The tandem BRCT domains at the COOH-terminus of BRCA1 are involved in several of its functions, including modulation of the activity of several transcription factors (15), binding to the RNA polymerase II holoenzyme (17), and activating transcription of a reporter gene when fused to a heterologous DNA-binding domain (18, 19). Importantly, cancer-associated mutations in the BRCT domains, but not benign polymorphisms, inactivate transcriptional activation and binding to RNA polymerase II (18–21). These observations suggest that abolishing the transcriptional activation function of BRCA1 leads to tumor development and provides a genetic framework for characterization of BRCA1 BRCT variants.

MATERIALS AND METHODS

The multiple sequence alignment (MSA) of orthologous BRCA1 BRCT domains from seven species, including *Homo sapiens* (GenBank accession number U14680), *Pan troglodytes* (AF207822), *Mus musculus* (U68174), *Rattus norvegicus* (AF036760), *Gallus gallus* (AF355273), *Canis familiaris* (U50709), and *Xenopus laevis* (AF416868), was obtained by using program ClustalW (22) and contains only one gapped position (Supplementary Fig. 1). According to PSI-BLAST (23), the latter six sequences are the only sequences in the nonredundant protein sequence database at National Center for Biotechnology Information that have between 30% and 90% sequence identity to the human BRCA1 BRCT domains (residues 1649–1859).

The multiple structure-based alignment of the native structures of the BRCT-like domains was obtained by the SALIGN command in MODELLER (Supplementary Fig. 2). It included the experimentally determined structures of the two human BRCA1 BRCT domains (Protein Data Bank code 1JNX; Refs. 8, 24), rat BRCA1 BRCT domains (1L0B; Ref. 7), human p53-binding protein (1KZY; Ref. 7), human DNA-ligase IIIα (1IMO; Ref. 25), and human XRCC1 protein (1CDZ; Ref. 13). Structure variability was defined by the root-mean-square deviation among the superposed Cα positions, as calculated by the COMPARE command of MODELLER. The purpose of these calculations was to gain insight into the variability of surface-exposed residues (*left panel* in Fig. 2). In conjunction with observed mutation clustering, these data may point to putative functional site(s) on the surface of BRCT repeats.

Comparative protein structure modeling by satisfaction of spatial restraints, implemented in the program MODELLER-6 (26), was used to produce a three-dimensional model for each of the 94 mutants. The crystallographic structure of the human wild-type BRCA1 BRCT domains was used as the template for modeling (8). The four residues missing in the crystallographic structure (1694 and 1817–1819) were modeled *de novo* (27). All of the models are available in the BRCA1 model set deposited in our ModBase database of comparative protein structure models (28).[6]

For the native structure of the human BRCT tandem repeat and each of the 94 mutant models, a number of sequence and structure features were calculated. These features were used in the classification tree in Fig. 3 (values for all 94 mutations are given in Supplementary Tables 1 and 2).

Buriedness. Accessible surface area of an amino acid residue was calculated by the program DSSP (29) and normalized by the maximum accessible surface area for the corresponding amino acid residue type. A residue was considered exposed if its accessible surface area was larger than 40Å^2 and if its relative accessible surface area was larger than 9% and buried otherwise. A mutation of a more exposed residue is less likely to change the structure and therefore its function.

Received 9/24/03; revised 1/30/04; accepted 3/15/04.

Grant support: This work was supported by Lee Kaplan Foundation, the Fashion Footwear Association of New York/QVC; United States Army award DAMD17-99-1-9389 and NIH CA92309 (A. N. A. M.); the Mathers Foundation, Sandler Family Supporting Foundation, Sun, IBM and Intel (A. S.); and NIH GM 54762 GM61390 (A. S.); and the Breast Cancer Research Foundation (B. L. W.). M. A. M-R. is a Rockefeller University Presidential Postdoctoral Fellow; A. S. is an Irma T. Hirschl Trust Career Scientist; and B. L. W. is an Abramson Investigator.

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

Note: The authors declare that they have no competing financial interests. Supplemental data for this article are available at Cancer Research Online (http: cancerres@aacrjournals.org).

Requests for reprints: Alvaro N. A. Monteiro, H. Lee Moffitt Cancer Center and Research Institute, MRC 3 West, 12902 Magnolia Drive, Tampa, FL 33612. Phone: (813) 745-6321; Fax: (813) 903-6847; E-mail: monteian@moffitt.usf.edu.

[6] http://salilab.org/modbase/.

Table 1 *Functionally characterized missense mutations in the human BRCA1 BRCT domains (amino acid residues 1652–1855)*

Mutation	Allowed residues[a]	Clinical data	Functional data[b]	Comments and references
M1652I	MIF	○	○	Likely amino acid substitution, 'I' allowed in multiple sequence alignment. Frequency in control population equal or higher than in clinical cases (33, 34) and absence of co-segregation in at least one family (33). Activity in transcription assays and in small colony phenotype test comparable to wild-type BRCA1 (35, 36).
M1652K	MIF		●	Charge change, two-class polarity change. No detectable activity in transcription assays (37).
L1657P	L		●	Predicted binding site, large volume change in rigid neighborhood, unlikely amino acid substitution at completely preserved position. No detectable activity in transcription assays (38).
E1660G	EKSC		●	Predicted binding site, large volume change, helix breaker. No detectable activity in transcription assays (38).
V1665M	V		●	Likely amino acid substitution. Activity in transcription assays comparable to wild-type BRCA1 (21).
A1669S	AS	○	○	Likely amino acid substitution, 'S' allowed in multiple sequence alignment. Absence of co-segregation in at least one family and activity in transcription assays comparable to wild-type BRCA1 (21).
H1686Q	H		●	Unlikely amino acid substitution at completely preserved position. No detectable activity in transcription assays at 37°C (38).
D1692N	D		○	Likely amino acid substitution. Activity in transcription assays comparable to wild-type BRCA1 (21). Found in an ovarian cancer case but with no family history (39). However, this mutation has been proposed to interfere with splicing–and is likely to be deleterious.
C1697R	C	●	●	Two-class polarity change, unlikely amino acid substitution at completely preserved position. Segregates with disease in at least one family and had no detectable activity in transcription assays (21).
R1699W	R	●	●	Predicted binding site, large volume change, charge change, two-class polarity change, located at BRCT-repeat interface (8). Segregates with disease in at least one family and had no detectable activity in transcription assays at 37°C (21,41).
R1699Q	R		●	Predicted binding site, large volume change, charge change, located at BRCT-repeat interface (8). No detectable activity in transcription assays at 37°C (21).
K1702E	K		●	Predicted binding site, big volume change, and charge change. No detectable activity in transcription assays (37).
Y1703H	Y		●	Charge change, two-class polarity change, located at BRCT-repeat interface (8). No detectable activity in transcription assays (37).
F1704S	F		●	Large volume change at rigid position, unlikely amino acid substitution at completely preserved position, BRCT-repeat interface (8). No detectable activity in transcription assays (38).
L1705P	LQ		●	Large volume change at rigid position and in rigid neighborhood, helix breaker, BRCT-repeat interface (8). No detectable activity in transcription assays (37).
G1706A	G		○	Not explained. Unlikely amino acid substitution at completely preserved position. Activity in transcription assays comparable to wild-type BRCA1.[c]
A1708E	A	●	●	Large volume change in rigid neighborhood, charge change, unlikely amino acid substitution at completely preserved position, BRCT-repeat interface, known unfolding mutation (8). No detectable activity in transcription assays or small colony phenotype test (18, 19, 36). Found in an early-onset breast tumor, the mutant allele (present in the germ line) was absent in control population and was retained in the tumor (41).
S1715N	S		●	Not explained. Likely amino acid substitution. No detectable activity in transcription assays (21). Found in an early-onset breast cancer case but segregation was not ascertained (42).
S1715R	S	●	●	Large volume change at rigid position and in rigid neighborhood, charge change. Segregates with disease in at least one family and had no detectable activity in transcription assays (21).
S1722F	S		●	Very large volume change at flexible position, unlikely amino acid substitution at completely preserved position, disrupts WXXXC/S motif and 1718–1722 interaction (13). No detectable activity in transcription assays at 37°C (39).
D1733G	DEH		○	Likely amino acid substitution on the surface outside of known functional sites. Activity in transcription assays comparable to wild-type BRCA1 (37).
F1734L	F		●	Unlikely amino acid substitution at completely preserved position. No detectable activity in transcription assays at 37°C (38).
G1738E	G		●	Buried, charge change, unlikely amino acid substitution at completely preserved position. No detectable activity in transcription assays (37).
G1743R	G		●	Unlikely amino acid substitution at completely preserved position. No detectable activity in transcription assays (38).
P1749R	P		●	Charge change, unlikely amino acid substitution at completely preserved position. No detectable activity in transcription assays and partial activity in small colony phenotype test (19, 36). Found in the germ line in an ovarian cancer family and not present in controls but segregation was not determined (43).
A1752P	AS		●	Helix breaker. Greatly decreased transcription activation.[c]
F1761I	FL		●	Not explained. Likely amino acid substitution. No detectable activity in transcription assays (37).
F1761S	FL		●	Very large volume change at flexible position. No detectable activity in transcription assays (37).
M1775E	M		●	Charge change, two-class polarity change, unlikely amino acid substitution at completely preserved position, BRCT-repeat interface (8). No detectable activity in transcription assays (18).
M1775K	M		●	Charge change, two-class polarity change, unlikely amino acid substitution at completely preserved position, BRCT-repeat interface (8). No detectable activity in transcription assays (18).
M1775R	M	●	●	Charge change, two-class polarity change, unlikely amino acid substitution at completely preserved position, BRCT-repeat interface (8). No detectable activity in transcription assays or small colony phenotype test (18, 19, 37). Found in an early-onset breast tumor, the mutant allele (present in the germ line) was absent in control population and was retained in the tumor (41). Segregates with disease (1).
M1775V	M		○	Not explained. Unlikely amino acid substitution at completely preserved position, BRCT-repeat interface (8). Activity in transcription assays comparable to wild-type BRCA1 (18).
L1780P	L		●	Unlikely amino acid substitution at completely preserved position, helix breaker. No detectable activity in transcription assays (37).
P1806A	PLAS		○	Likely amino acid substitution, 'A' allowed in multiple sequence alignment. Activity in transcription assays comparable to wild-type BRCA1 (37).
I1807S	IVL		●	Large volume change at rigid position and in rigid neighborhood. No detectable activity in transcription assays (38).
V1833E	V		●	Charge change, two-class polarity change, and unlikely amino acid substitution at completely preserved position. No detectable activity in transcription assays (37).
A1843T	AS		●	Not explained. BRCT-repeat interface (8). No detectable activity in transcription assays (38). Found in an ovarian cancer case but segregation was not determined (44).

[a] Amino acid residues present at the same position in the *BRCA1* orthologs ('Materials and Methods;' Supplementary Fig. 1). ○ indicates a presumed benign mutation, and ● indicates a presumed cancer-associated mutation.

[b] Transcription activity when fused to a heterologous DNA binding domain; ● denotes loss of function and ○ denotes activity comparable to that of wild-type BRCA1.

[c] C. Phelan, S. Narod and A. Monteiro, unpublished results.

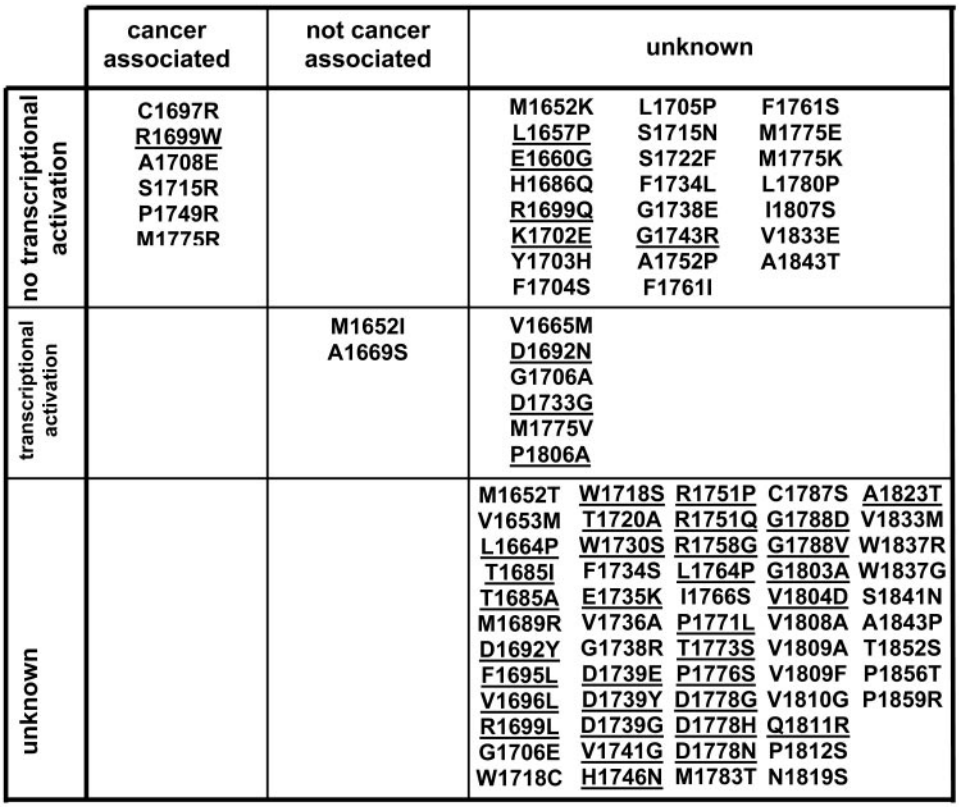

Fig. 1. Classification of the 94 missense mutations in the human BRCA1 BRCT domains analyzed in this study. The mutants were divided into groups depending on information about their cancer association and transcriptional activation function (Tables 1 and 2). The mutations at solvent-exposed positions are *underlined* ("Materials and Methods").

Functional Site. The putative functional site was visually assigned to the surface groove that contains the deleterious mutants (Fig. 2); it consists of residue positions with accessible surface area >9%, as follows: 1654–1657, 1659, 1662, 1663, 1666, 1676, 1678, 1698–1702, 1773, 1774, 1811–1813, 1834, and 1836. A mutation in the functional site is more likely to abolish the function.

Residue Rigidity. A residue isotropic temperature factor (B-factor) was the mean of its atomic B-factors as determined by X-ray crystallography of the native structure; it was also normalized by the mean and standard deviation (SD) of all residue B-factors in the protein chain. A mutation of a less rigid residue at a buried position is less likely to change the structure and therefore its function.

Neighborhood Rigidity. Neighborhood of a residue was defined by the residues that have at least one atom within 5Å of any atom of the central residue. Neighborhood B-factor of a residue was defined to be the mean value of the B-factors of the residues in the neighborhood; it was also normalized by the mean and SD of all neighborhood B-factors for the neighborhoods of the same size. A mutation of a buried residue in a less rigid neighborhood is less likely to change the structure and therefore its function.

Volume Change. Volume change was defined solely by the type of residue substitution (30). A large change in the volume of the amino acid residue type was considered destabilizing to the structure and therefore its function, especially when buried at a rigid position.

Charge Change. Charge change was defined solely by the type of residue substitution (30). Mutations corresponding to changes in charge at buried positions are more likely to change the structure and therefore its function.

Polarity Change. Polarity change was defined solely by the type of residue substitution. Amino acid residue types were classified into three classes according to their hydrophobicity: $0 \equiv$ {LIFWCMVY}, $1 \equiv$ {PATGS}, $2 \equiv$ {HQRKNED} (30). Mutations corresponding to changes in polarity at buried positions are more likely to change the structure and therefore its function.

Mutation Likelihood. Mutation likelihood of a residue substitution at a given MSA position reflects the "background" substitution probabilities in the BLOSUM62 matrix as well as the actually observed substitutions at the given position in the MSA of the BRCA1 BRCT orthologs (Supplementary Fig. 1; Refs. 4–6, 9, 31). The scores are centered on zero, with negative values corresponding to less likely substitutions. Unlikely mutations are more likely to change the structure and function.

Phylogenetic Entropy. Sequence variability was defined for every column in the MSA of the BRCA1 BRCT orthologs (Supplementary Fig. 1) by $\Sigma\, p \ln p$, where p is the observed relative frequency of residue type i (Fig. 2, *right panel*). Mutations at evolutionarily conserved positions are more likely to change the structure and function.

Helix/Turn Breaker. Secondary structure assignment was calculated by DSSP (29). A "helix breaker" mutation was defined as any mutation into the glycine or proline residues at a helical position; a "turn breaker" mutation was defined as any replacement of the glycine or proline residues at a turn position.

Visualization and analysis of the mutations were facilitated by the program DINO.[7]

Transcriptional assays for testing our predictions were performed as described previously (21). The wild-type BRCA1 sequence (residues 1396–1863) fused in frame to the LexA DNA-binding domain was used as the wild-type control and for introducing mutations M1775R, Y1853X, M1689R, and W1718C by site-directed mutagenesis. EGY48 yeast cells were cotransformed with the constructs and pRB1840, which contains a *lac*Z reporter gene under the control of a LexA operator, and tested in liquid β-galactosidase assays. For mammalian transcription assays, BRCA1 constructs were subcloned in frame to GAL4 DNA-binding domain in pCDNA3. GAL4 DNA-binding domain fusion constructs were cotransfected into human 293T cells with pG5Luc, which contains a firefly luciferase gene under the control of five GAL4 binding sites as a reporter for the assay. Transfections were normalized with an internal control phRG-TK (Promega), which contains a *Renilla* luciferase gene under a constitutive thymidine kinase-basal promoter.

RESULTS AND DISCUSSION

We tabulated a total of 94 missense mutations in the human BRCA1 BRCT domains (Tables 1 and 2; Fig. 1). These mutations occurred naturally or were obtained via random mutagenesis. Tran-

[7] http://www.dino3d.org.

Table 2 *Prediction of the cancer phenotype of 57 functionally uncharacterized missense mutations in the human BRCA1 BRCT domains*

Mutation	Allowed residues[a]	Prediction	Comments
M1652T	MIF	●	Large volume change in rigid neighborhood. Found in breast and ovarian cancer families in Finland (45).
V1653M	V	○	Likely amino acid substitution.
L1664P	LITM	●	Helix breaker.
T1685A	T	●	Unlikely amino acid substitution at completely preserved position.
T1685I	T	●	Unlikely amino acid substitution at completely preserved position.
M1689R	MI	●	Charge change, two-class polarity change.
D1692Y	D	●	Unlikely amino acid substitution at completely preserved position.
F1695L	FL	○	Likely amino acid substitution, 'L' allowed in MSA.
V1696L	V	○	Likely amino acid substitution.
R1699L	R	●	Predicted binding site, charge change, two-class polarity change, unlikely amino acid substitution at completely preserved position, BRCT-domain interface (8).
G1706E	G	●	Large volume change at rigid position and in rigid neighborhood, charge change, unlikely amino acid substitution at completely preserved position.
W1718C	W	●	Large volume change at rigid position and in rigid neighborhood, unlikely amino acid substitution at completely preserved position, BRCT signature WXXXS motif (13).
W1718S	W	●	Large volume change at rigid position and in rigid neighborhood, unlikely amino acid substitution at completely preserved position.
T1720A	TIV	○	Likely amino acid substitution.
N1730S	NDS	○	Likely amino acid substitution, 'S' allowed in MSA.
F1734S	F	●	Very large volume change at flexible position, unlikely amino acid substitution at completely preserved position.
E1735K	E	●	Charge change. Found in a family with five breast cancer patients, in first-degree relationship, over two generations (46).
V1736A	V	●	Unlikely amino acid substitution at completely preserved position.
G1738R	G	●	Very large volume change at flexible position, unlikely amino acid substitution at completely preserved position. Found in four unrelated patients from breast-ovarian cancer families. Segregation analysis not done (47).
D1739E	D	○	Likely amino acid substitution.
D1739G	D	●	Unlikely amino acid substitution at completely preserved position.
D1739Y	D	●	Large volume change at rigid position, two-class polarity change, unlikely amino acid substitution at completely preserved position.
V1741G	VI	○	Likely amino acid substitution. Found in combination with another deleterious mutation (E1754X) and therefore likely to represent a benign polymorphism (48).
H1746N	H	●	Unlikely amino acid substitution at completely preserved position.
R1751P	KR	●	Unlikely amino acid substitution at completely preserved position, helix breaker.
R1751Q	KR	○	Likely amino acid substitution. Found in a breast-ovarian cancer family. Segregation analysis not done (49).
R1758G	REG	○	Likely amino acid substitution, 'G' allowed in MSA. Found in two cases with no family history (50).
L1764P	LF	○	Likely amino acid substitution.
I1766S	IV	●	Large volume change at rigid position.
P1771L	PS	○	Likely amino acid substitution.
T1773S	T	○	Likely amino acid substitution, predicted binding site.
P1776S	PT	○	Likely amino acid substitution. Found in an ovarian cancer case but with no family history (39).
D1778G	DG	●	Helix breaker, 'G' allowed in MSA.
D1778H	DG	○	Likely amino acid substitution. Found in a patient with no family history (46).
D1778N	DG	○	Likely amino acid substitution.
M1783T	MI	○	Likely amino acid substitution.
C1787S	C	●	Unlikely amino acid substitution at completely preserved position, BRCT domain interface (8).
G1788D	G	●	Unlikely amino acid substitution at completely preserved position, turn breaker.
G1788V	G	●	Unlikely amino acid substitution at completely preserved position, turn breaker. Found in a bilateral breast and ovarian cancer case with family history (51).
G1803A	GN	○	Turn breaker, out of known and putative binding sites.
V1804D	VTAS	●	Two-class polarity change.
V1808A	V	●	Unlikely amino acid substitution at completely preserved position.
V1809A	VLI	○	Likely amino acid substitution.
V1809F	VLI	○	Likely amino acid substitution.
V1810G	V	●	Unlikely amino acid substitution at completely preserved position.
Q1811R	Q	●	Charge change in putative binding site.
P1812S	P	●	Unlikely amino acid substitution at completely preserved position, predicted binding site. Found in an ovarian cancer case but with no family history (39).
N1819S	NSTR	○	Likely amino acid substitution, 'S' allowed in multiple sequence alignment.
A1823T	ADE	○	Likely amino acid substitution. Found in one patient from a breast-ovarian cancer family. Segregation analysis not done (47).
V1833M	V	○	Likely amino acid substitution. Found in one patient from a breast-ovarian cancer family. Segregation analysis not done (47).
W1837G	W	●	Large volume change at rigid position, unlikely amino acid substitution at completely preserved position, helix breaker, BRCT signature WXXXS motif (13).
W1837R	W	●	Charge change, two-class polarity change, unlikely amino acid substitution at completely preserved position, BRCT signature WXXXS motif (13). Found in an early onset breast cancer case. Proband's father also had breast cancer and the mutation was found to segregate with disease (52).
S1841N	S	●	BRCT signature WXXXS motif (13).
A1843P	AS	●	Helix breaker. Found in ovarian cancer family in Japan. Segregation analysis not done (44).
T1852S	TA	○	Likely amino acid substitution.
P1856T	PQS	○	Likely amino acid substitution.
P1859R	PT	○	Likely amino acid substitution.

[a] Comments list the rationalizations for each of the 57 predictions, as produced by the classification tree (Fig. 3). ○ indicates a predicted benign mutation, and ● indicates a predicted cancer-associated mutation.

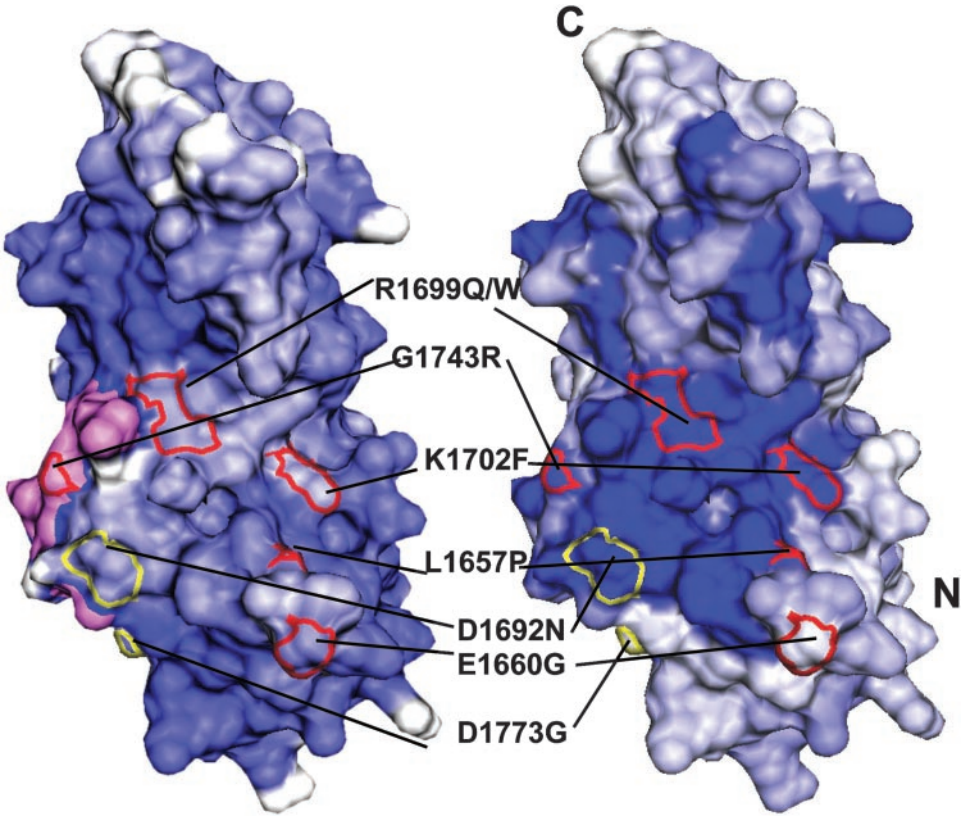

Fig. 2. Missense mutations on the surface of the human BRCA1 BRCT domains. The solvent-exposed residues with mutations that are known to affect and not affect the transcriptional activation activity of BRCA1 are contoured in *red* and *yellow*, respectively (Fig. 1). The only remaining exposed mutation of known functional impact (P1806A) is on the hidden face of the structure. Conservation of structure ("Materials and Methods;" Supplementary Fig. 2) and sequence (see phylogenetic entropy in "Materials and Methods" and Supplementary Fig. 1) are indicated by the intensity of *blue on the left and right panels*, respectively. The region in *violet on the left panel* corresponds to the linker between the two domains. The NH_2- and COOH-termini of the tandem domains are marked on the *right panel*.

scriptional activation function of 37 of 94 mutants has been tested (Fig. 1, *rows 1 and 2*). Eight of the 37 mutants could be tentatively classified either as cancer-associated or benign following established genetic criteria (Ref. 32; Fig. 1, *columns 1 and 2*). The basis of each classification is outlined in Table 1.

For the eight mutations with intact transcriptional activation function, there is no evidence of cancer-association; the frequency in control population is equal or higher than in cancer cases, and there is absence of co-segregation in at least one family (Fig. 1, *row 2*; Table 1). Conversely, none of the 29 tested mutations without transcriptional activation function are thought to be benign, following the same criteria (Fig. 1, *row 2*; Table 1; Ref. 32). These observations indicate that tumor suppression by BRCA1 correlates with its ability to activate transcription.

We first rationalized the functional consequences of the 37 tested mutations in terms of their impact on folding and stability of the native structure and in terms of the integrity of binding sites of the BRCT domains. Next, we used this experience to develop generic rules for prediction of the functional impact of a missense mutation and applied these rules to the remaining 57 functionally uncharacterized mutations listed in Fig. 1 and Table 2. We then tested two of these predictions by functional assays and segregation analysis.

Missense mutations of BRCA1 are likely to affect transcriptional activation in either one of the following two ways: (*a*) when they are exposed to the solvent, they may substantially change the structure or chemical nature of functional sites that bind other molecules, or (*b*) when they are buried in the core, they may prevent folding of the BRCT domains into their native fold or, less likely, affect only the structure of functional sites.

The 37 mutations, including the 8 (Fig. 1, *row 2*) and the 29 mutations (Fig. 1, *row 1*) with and without transcriptional activation function, respectively, were mapped onto the crystallographic structure of the human BRCA1 BRCT domains (Ref. 8; Fig. 2). In addition to this structural mapping, we produced a three-dimensional protein structure model for each mutant and calculated a number of its sequence and structure features ("Materials and Methods"). We used the feature values and their changes between the wild type and the mutant to rationalize the observed effect on transcriptional activation (Table 1; Supplementary Table 1).

The markedly reduced transcriptional activation activity of the 29 mutations (Fig. 1, *row 1*) is readily explained by their likely disruption of the structure of the BRCT domain and/or its binding sites that are involved in transcriptional activation. Twenty-three of these 29 mutations occur at buried positions and involve amino acid residue types with dissimilar charge, hydrophobicity, size, or secondary structure propensity (Fig. 1, *row 1, not underlined*). Thus, they are expected to disrupt the structure and therefore its function. The remaining six mutations resulting in reduced transcriptional activation are exposed (Fig. 1, *row 1, underlined*). Among these six mutations, R1699W and R1699Q replace a conserved residue involved in a salt bridge between the BRCT repeats (8). Another one of the six mutations (G1743R) occurs at a position that is absolutely conserved among all available orthologs (Supplementary Fig. 1). For the remaining three exposed mutations that abolish transcription, we propose that they disrupt interactions with BRCA1 ligand(s). They are located in a groove formed by both BRCT repeats (L1657P and K1702E) and in the ridge that delimits the groove (E1660G). We suggest that this clustering of residues in a single patch on the surface of BRCA1 highlights a binding site with integrity that is necessary for transcription. Moreover, this putative binding site is relatively conserved in sequence and structure among the BRCA1 homologs (Fig. 2; "Materials and Methods"). These results are in agreement with *in vitro* binding studies that determined the binding site of BRCA1 to the RNA polymerase holoenzyme via RNA helicase A (17). Nonconservative substitutions of any amino acid residues that comprise the putative binding site are also likely to affect transcriptional activation. The existence of such a

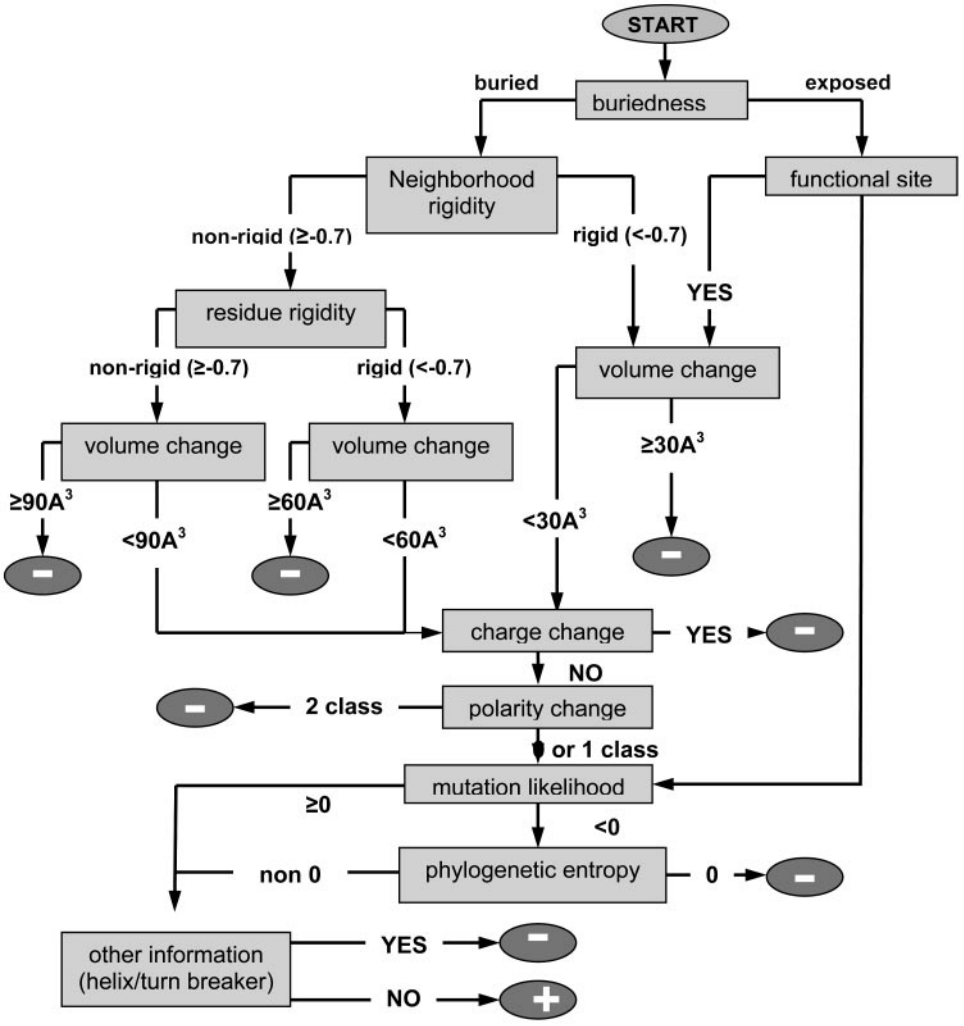

Fig. 3. Classification tree for annotating a given missense mutation in the human BRCA1 BRCT domains. The classification begins at the *top* (Start) and ends at one of the "+" or "-" leaves. The "+" and "-" indicate mutants with intact (benign) and abolished transcriptional activation function (cancer-associated), respectively. This classification tree correctly rationalizes 32 of the 37 functionally characterized variants (Table 1). The tree was also applied to 57 mutants that have not yet been characterized by experiment (Table 2). See "Materials and Methods" and Supplementary Material for a detailed description of the features.

functional region indicates that the tandem BRCT repeats comprise one structural and functional unit, in accordance with previous conclusions (7).

The lack of impact of the eight mutations that do not disrupt transcriptional activation function (Fig. 1, *row 2*) can also be rationalized in structural terms. Three of them are exposed on the face away from the putative binding site proposed above, whereas the remaining five mutations are conservative changes in the core of the fold (Table 1; Fig. 2). Five mutations (G1706A, S1715N, F1761I, M1775V, and A1843T) in the total set of 37, with impact on the transcriptional activation function that is known, could not be readily explained. Possible reasons for the inability to explain these mutants are discussed in detail in Supplementary Table 3.

Because it was straightforward to use the three-dimensional structure of the BRCA1 domains to rationalize the known functional impact of 32 of the 37 mutations, we derived a set of rules for predicting whether a given mutation inactivates transcriptional activation and therefore correlates with predisposition to cancer. These rules rely on the sequence and structural features that proved informative in explaining the presence or absence of a functional impact of the tested mutations. For example, structural neighborhood around a mutated residue can be rigid, and a mutation can introduce a large change (over $30A^3$) in the side chain volume. Such a mutation is likely to destabilize or distort local conformation, thus having an effect on the function of the protein. In contrast, a flexible structural neighborhood is expected to accommodate such a substitution without major perturbations in the structure. Moreover, our rules allow for the possibility that very large changes in the side chain volume (over $90A^3$) can have a significant effect even in a flexible neighborhood. The rules were encoded hierarchically in a "classification tree" scheme (Fig. 3) and are implemented on the web.[8]

Using the classification tree, we predicted the functional consequences of the 57 missense mutations in the BRCT domains that presently lack definitive information about cancer association (Breast Cancer Information Core Database;[9] Table 2; Supplementary Table 2). Our prediction algorithm assigned functional impact to 32 mutants, whereas 25 were predicted to be benign.

To validate our predictions, we tested prospectively two mutants (Fig. 4, *A* and *B*) for which segregation data were available. The W1718C mutation was found in a family with three generations affected where women were diagnosed with either early-onset (<40-years-old) breast cancer or ovarian cancer. It was present in four of four women with cancer and was absent in three of three at-risk women without cancer.[10] The M1689R mutation also segregated with disease in a large family with several generations affected (Fig. 4*C*). Both M1689R and W1718C displayed the loss of function phenotype and segregated with disease in agreement with our predictions (Fig. 4).

[8] http://salilab.org/SNPWeb/.
[9] http://www.nhgri.nih.gov/Intramural_research/Lab_transfer/Bic/.
[10] L. Baumbach-Reardon, personal communication.

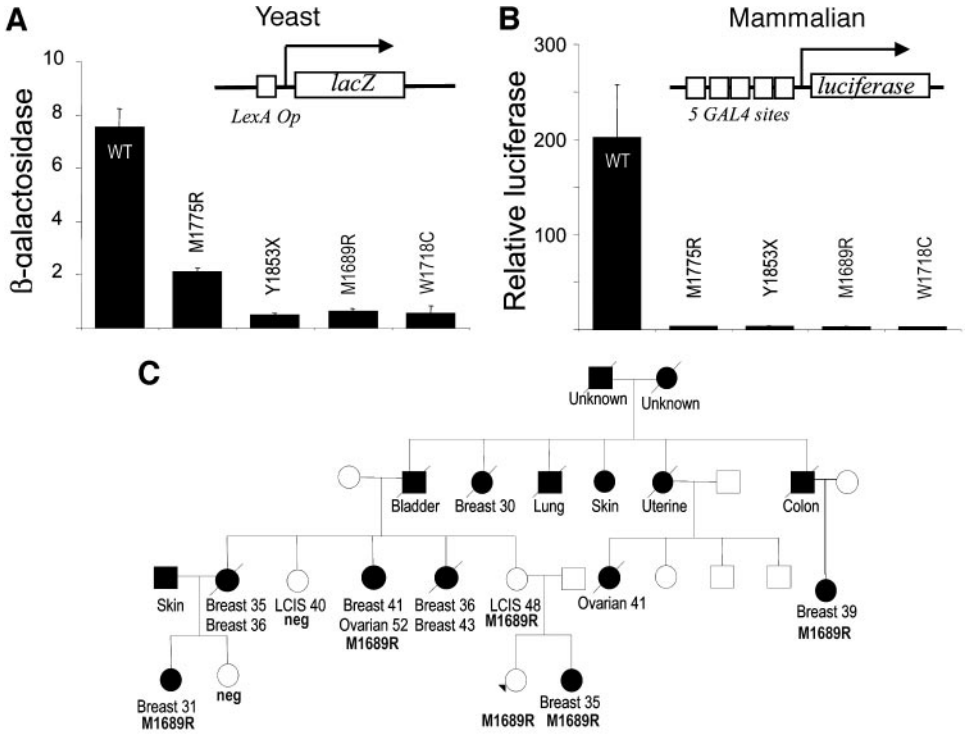

Fig. 4. Transcription assays and segregation analysis confirm predictions for mutations M1689R and W1718C. Structures of the reporter plasmid are depicted on *top of panels A and B*. Cancer-associated mutations M1775R and Y1853X were used as controls. *A*, transcriptional activity in yeast cells. *B*, transcriptional activity in human cells. *C*, pedigree with segregation analysis for the M1689R mutant. *Arrowhead* indicates the proband. The presence (M1689R) or absence (neg.) of M1689R mutation in the germ line of tested individuals is indicated. Site of tumor and age of diagnosis is also indicated. *WT*, wild type.

To illustrate prediction by the classification tree, we used the mutation M1689R that occurs at a buried position in a relatively rigid neighborhood (Table 2, Supplementary Table 2). Mutations at buried and rigid locations are tolerable only if changes in the residue volume, charge, and polarity are sufficiently small; a volume change of more than one methyl group (30Å^3) cannot be accepted. However, the arginine volume is only 18Å^3 larger than that of methionine; therefore, the condition on the maximal allowed volume change is satisfied. At the next checkpoint, the side-chain charge is considered. Because M1689R introduces a positive charge at a buried position, the mutation is considered destabilizing, resulting into the prediction of a loss of function.

The tumor suppressor function of BRCA1 is likely to depend, at least partially, on its transcriptional activity, as suggested previously (18–21) and supported by the mutation set in Table 1. Therefore, the ability to predict which mutations do and do not abolish the transcriptional activation function of BRCA1 can be useful in predicting their cancer association. Our analysis, based primarily on the considerations of protein three-dimensional structure, leads to a prediction of cancer association of the mutations that cannot be easily characterized by other approaches, such as population-based statistical genetics. Although all but two of the present predictions are unverified and thus cannot be used as a sole tool for risk assessment and genetic counseling, this structure-based approach may be helpful in an integrated effort, to identify mutations that predispose an individual to breast and ovarian cancers.

ACKNOWLEDGMENTS

We are grateful to Drs. Andras Fiser and Anthony Brown for helpful discussions and to Dr. Ursula Pieper for help with the server setup.

REFERENCES

1. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science (Wash) 1994;266:66–71.
2. Friedman LS, Ostermeyer EA, Szabo CI, et al. Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. Nat Genet 1994;8:399–404.
3. Szabo CI, Wagner LA, Francisco LV, et al. Human, canine and murine BRCA1 genes: sequence comparison among species. Hum Mol Genet 1996;5:1289–98.
4. Orelli BJ, Logsdon JJ Jr, Bishop DK. Nine novel conserved motifs in BRCA1 identified by the chicken orthologue. Oncogene 2001;20:4433–8.
5. Joukov V, Chen J, Fox EA, Green JB, Livingston DM. Functional communication between endogenous BRCA1 and its partner, BARD1, during Xenopus laevis development. Proc Natl Acad Sci USA 2001;98:12078–83.
6. Bennett LM, Brownlee HA, Hagavik S, Wiseman RW. Sequence analysis of the rat Brca1 homolog and its promoter region. Mamm Genome 1999;10:19–25.
7. Joo WS, Jeffrey PD, Cantor SB, Finnin MS, Livingston DM, Pavletich NP. Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brca1 BRCT structure. Genes Dev 2002;16:583–93.
8. Williams RS, Green R, Glover JN. Crystal structure of the BRCT repeat region from the breast cancer- associated protein BRCA1. Nat Struct. Biol 2001;8:838–42.
9. Chasman D, Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. J Mol Biol 2001;307:683–706.
10. Saunders C, Baker D. Evaluation of structural and evolutionary contributions to deleterious mutation prediction. J Mol Biol 2002;322:891–901.
11. Derbyshire DJ, Basu BP, Serpell LC, et al. Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor. EMBO J 2002;21:3863–72.
12. Ferrer-Costa C, Orozco M, de la Cruz X. Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. J Mol Biol 2002;315:771–86.
13. Zhang X, Morera S, Bates PA, et al. Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. EMBO J 1998;17:6404–11.
14. Huyton T, Bates PA, Zhang X, Sternberg MJ, Freemont PS. The BRCA1 C-terminal domain: structure and function. Mutat Res 2000;460:319–32.
15. Monteiro AN. BRCA1: exploring the links to transcription. Trends Biochem Sci 2000;25:469–74.
16. Venkitaraman AR. Cancer Susceptibility and the Functions of BRCA1 and BRCA2. Cell 2002;108:171–82.
17. Anderson SF, Schlegel BP, Nakajima T, Wolpin ES, Parvin JD. BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A. Nat Genet 1998;19:254–6.
18. Monteiro AN, August A, Hanafusa H. Evidence for a transcriptional activation function of BRCA1 C-terminal region. Proc Natl Acad Sci USA 1996;93:13595–9.
19. Chapman MS, Verma IM. Transcriptional activation by BRCA1 [letter; comment]. Nature (Science) 1996;382:678–9.
20. Scully R, Anderson SF, Chao DM, et al. BRCA1 is a component of the RNA polymerase II holoenzyme. Proc Natl Acad Sci USA 1997;94:5605–10.
21. Vallon-Christersson J, Cayanan C, Haraldsson K, et al. Functional analysis of BRCA1 C-terminal missense mutations identified in breast and ovarian cancer families. Hum Mol Genet 2001;10:353–60.

22. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–80.
23. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–402.
24. Westbrook J, Feng Z, Jain S, et al. The protein data bank: unifying the archive. Nucleic Acids Res 2002;30:245–8.
25. Krishnan VV, Thornton KH, Thelen MP, Cosman M. Solution structure and backbone dynamics of the human DNA ligase IIIalpha BRCT domain. Biochemistry 2001;40: 13158–66.
26. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
27. Fiser A, Do RK, Sali A. Modeling of loops in protein structures. Protein Sci 2000;9:1753–73.
28. Pieper U, Eswar N, Stuart AC, Ilyin VA, Sali A. MODBASE, a database of annotated comparative protein structure models. Nucleic Acids Res 2002;30:255–9.
29. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–637.
30. Bordo D, Argos P. Suggestions for "safe" residue substitutions in site-directed mutagenesis. J Mol Biol 1991;217:721–9.
31. Henikoff JG, Henikoff S. Using substitution probabilities to improve position-specific scoring matrices. Comput Appl Biosci 1996;12:135–43.
32. Couch FJ, Weber BL. Mutations and polymorphisms in the familial early-onset breast cancer (BRCA1) gene. Breast Cancer Information Core. Hum Mutat 1996;8:8–18.
33. Deffenbaugh AM, Frank TS, Hoffman M, Cannon-Albright L, Neuhausen SL. Characterization of common BRCA1 and BRCA2 variants. Genet Test 2002;6:119–21.
34. Greenman J, Mohammed S, Ellis D, et al. Identification of missense and truncating mutations in the BRCA1 gene in sporadic and familial breast and ovarian cancer. Genes Chromosomes Cancer 1998;21:244–9.
35. Monteiro AN, August A, Hanafusa H. Common BRCA1 variants and transcriptional activation [letter]. Am J Hum Genet 1997;61:761–2.
36. Humphrey JS, Salim A, Erdos MR, Collins FS, Brody LC, Klausner RD. Human BRCA1 inhibits growth in yeast: potential use in diagnostic testing. Proc Natl Acad Sci USA 1997;94:5820–5.
37. Hayes F, Cayanan C, Barilla D, Monteiro AN. Functional assay for BRCA1: mutagenesis of the COOH-terminal region reveals critical residues for transcription activation. Cancer Res 2000;60:2411–8.
38. Carvalho MA, Billack B, Chan E, Worley T, Cayanan C, Monteiro AN. Mutations in the BRCT Domain Confer Temperature Sensitivity to BRCA1 in Transcription Activation. Cancer Biol Ther 2002;1:502–8.
39. Janezic SA, Ziogas A, Krumroy LM, et al. Germline BRCA1 alterations in a population-based series of ovarian cancer cases. Hum Mol Genet 1999;8:889–97.
40. Worley T, Vallon-Christersson J, Billack B, Borg A, Monteiro AN. A naturally occurring allele of BRCA1 coding for a temperature-sensitive mutant protein. Cancer Biol Ther 2002;1:497–501.
41. Futreal PA, Liu Q, Shattuck-Eidens D, et al. BRCA1 mutations in primary breast and ovarian carcinomas. Science (Wash D C) 1994;266:120–2.
42. Diez O, Cortes J, Domenech M, et al. BRCA1 mutation analysis in 83 Spanish breast and breast/ovarian cancer families. Int J Cancer 1999;83:465–9.
43. Gayther SA, Harrington P, Russell P, Kharkevich G, Garkavtseva RF, Ponder BA. Rapid detection of regionally clustered germ-line BRCA1 mutations by multiplex heteroduplex analysis. UKCCCR Familial Ovarian Cancer Study Group. Am J Hum Genet 1996;58:451–6.
44. Sekine M, Nagata H, Tsuji S, et al. Mutational analysis of BRCA1 and BRCA2 and clinicopathologic analysis of ovarian cancer in 82 ovarian cancer families: two common founder mutations of BRCA1 in Japanese population. Clin Cancer Res 2001;7:3144–50.
45. Huusko P, Paakkonen K, Launonen V, et al. Evidence of founder mutations in Finnish BRCA1 and BRCA2 families [letter]. Am J Hum Genet 1998;62:1544–8.
46. Garcia-Patino E, Gomendio B, Silva JM, Garcia JM, Espana P, Bonilla F. Missense mutations in the transactivation region of the BRCA1 gene in breast cancer. Cancer Genet Cytogenet 1998;107:80–1.
47. Ladopoulou A, Kroupis C, Konstantopoulou I, et al. Germ line BRCA1 & BRCA2 mutations in Greek breast/ovarian cancer families: 5382insC is the most frequent mutation observed. Cancer Lett 2002;185:61–70.
48. Valarmathi MT, A A, Deo SS, Shukla NK, Das SN. BRCA1 germline mutations in Indian familial breast cancer. Hum Mutat 2003;21:98–9.
49. Stoppa-Lyonnet D, Laurent-Puig P, Essioux L, et al. BRCA1 sequence variations in 160 individuals referred to a breast/ovarian family cancer clinic. Institut Curie Breast Cancer Group [see comments]. Am J Hum Genet 1997;60:1021–30.
50. Newman B, Mu H, Butler LM, Millikan RC, Moorman PG, King MC. Frequency of breast cancer attributable to BRCA1 in a population-based series of American women. J Am Med Assoc 1998;279:915–21.
51. Berchuck A, Heron KA, Carney ME, et al. Frequency of germline and somatic BRCA1 mutations in ovarian cancer. Clin Cancer Res 1998;4:2433–7.
52. Montagna M, Santacatterina M, Corneo B, et al. Identification of seven new BRCA1 germline mutations in Italian breast and breast/ovarian cancer families Cancer Res 1996;56:5466–9.