Determination of Cancer Risk Associated with Germ Line BRCA1 Missense Variants by Functional Analysis

Marcelo A. Carvalho,[1,3] Sylvia M. Marsillac,[1,4] Rachel Karchin,[6] Siranoush Manoukian,[7] Scott Grist,[9] Ramona F. Swaby,[10] Turan P. Urmenyi,[4] Edson Rondinelli,[4,5] Rosane Silva,[4] Luis Gayol,[12] Lisa Baumbach,[12] Rebecca Sutphen,[1,2] Jennifer L. Pickard-Brzosowicz,[2] Katherine L. Nathanson,[11] Andrej Sali,[13] David Goldgar,[14] Fergus J. Couch,[15] Paolo Radice,[7,8] and Alvaro N.A. Monteiro[1]

[1]Risk Assessment, Detection, and Intervention Program, and [2]Lifetime Cancer Screening and Prevention Center, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida; [3]Centro Federal de Educação Tecnológica de Química; [4]Laboratório de Metabolismo Macromolecular Firmino Torres de Castro, Instituto de Biofísica Carlos Chagas Filho, and [5]Departamento de Clínica Médica, Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; [6]Department of Biomedical Engineering, Institute of Computational Medicine, Johns Hopkins University, Baltimore, Maryland; [7]Department of Experimental Oncology, Istituto Nazionale Tumori; [8]Fondazione Italiana per la Ricerca sul Cancro, Institute of Molecular Oncology, Milan, Italy; [9]Department of Haematology and Genetic Pathology, Flinders University Medical Centre, Bedford Park, South Australia, Australia; [10]Department of Medical Oncology, Fox Chase Cancer Center; [11]Division of Medical Genetics, Department of Medicine, Abramson Cancer Center, University of Pennsylvania, Philadelphia, Pennsylvania; [12]The Dr. John T. Macdonald Foundation Center for Medical Genetics and Department of Pediatrics, Miller School of Medicine, University of Miami, Miami, Florida; [13]Department of Biopharmaceutical Sciences, California Institute for Quantitative Biomedical Research, University of California at San Francisco, San Francisco, California; [14]Department of Dermatology, University of Utah, Salt Lake City, Utah; and [15]Mayo Clinic College of Medicine, Rochester, Minnesota

Abstract

Germ line inactivating mutations in *BRCA1* confer susceptibility for breast and ovarian cancer. However, the relevance of the many missense changes in the gene for which the effect on protein function is unknown remains unclear. Determination of which variants are causally associated with cancer is important for assessment of individual risk. We used a functional assay that measures the transactivation activity of BRCA1 in combination with analysis of protein modeling based on the structure of BRCA1 BRCT domains. In addition, the information generated was interpreted in light of genetic data. We determined the predicted cancer association of 22 *BRCA1* variants and verified that the common polymorphism S1613G has no effect on BRCA1 function, even when combined with other rare variants. We estimated the specificity and sensitivity of the assay, and by meta-analysis of 47 variants, we show that variants with <45% of wild-type activity can be classified as deleterious whereas variants with >50% can be classified as neutral. In conclusion, we did functional and structure-based analyses on a large series of *BRCA1* missense variants and defined a tentative threshold activity for the classification missense variants. By interpreting the validated functional data in light of additional clinical and structural evidence, we conclude that it is possible to classify all missense variants in the BRCA1 COOH-terminal region. These results bring functional assays for *BRCA1* closer to clinical applicability. [Cancer Res 2007;67(4):1494–501]

Introduction

The breast and ovarian cancer predisposition gene *BRCA1* displays large allelic diversity with several thousand different alleles documented thus far (Breast Cancer Information Core Database).[16] This is consistent with the expected allelic structure of genes that determine rare monogenic diseases (1). A large portion of the documented alleles have had their disease association determined by inference from genetic and biochemical evidence that indicates that even very small truncations of 11 COOH-terminal amino acids result in a protein with compromised function (2–4). Furthermore, deletion analysis suggests that truncation of only eight amino acids may abrogate function (5). However, a significant number of *BRCA1* alleles, mostly containing missense changes, have proved more difficult to assess for disease association. These are termed unclassified variants (UCV) or variants of uncertain significance. The lack of conclusive genetic information is primarily due to low frequency of each individual UCV.

The problem generated by these UCVs is widely recognized and the need to provide risk assessment to individuals in high-risk families has brought several advances aimed at classifying these variants as follows: (*a*) Bayesian methods to analyze pedigrees (6); (*b*) use of information from interspecies sequence variation (7–9); (*c*) integrated methods combining information from different sources in a comprehensive framework (10, 11); (*d*) functional assays to assess the effect of amino acid changes on protein function (12–16); (*e*) methods based on co-occurrence with a deleterious mutations (17); and (*f*) structure-based analysis to generate computation prediction models (18, 19). Although still far from clinical application, these methods have provided important information.

Here we apply a transcription-based assay to assess the effect of 22 variants (R1443G, V1534M, D1546N, L1564P, P1614S, E1644G, S1655F, L1664P, T1700A, G1706E, V1713A, V1736A, G1738R, G1738E, R1753T, I1766S, L1764P, Q1785H, G1788D, E1794D, V1804D, and P1806A) on the activity of the BRCA1 COOH terminus (amino acids 1,396–1,863) and interpret the findings in light of all available clinical, genetic, and structural information. These UCVs were

[16] http://research.nhgri.nih.gov/bic/

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

M.A. Carvalho and S.M. Marsillac contributed equally to this work.

Conflict of interest: Results from this work may bear on Myriad Genetic Laboratories commercial test for mutations in BRCA1 and BRCA2. D. Goldgar receives royalties from the University of Utah through its license agreement with Myriad Genetics, Inc.

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

doi:10.1158/0008-5472.CAN-06-3297

found in individuals from families with breast or ovarian cancer in which no other deleterious mutation in *BRCA1* or *BRCA2* was found, and thus it is not clear whether they are causally related to disease. We also assessed the function of constructs representing the combination of the common polymorphism S1613G with other rare variants. The transcription-based method, used as a monitor for the integrity of the BRCT domain and its flanking regions, has shown an excellent agreement with existing genetic data (12, 13).

Several findings with significant implication for genetic counseling and risk assessment emerged from this study, which included a meta-analysis of 47 *BRCA1* UCVs. The study determined tentative activity thresholds for functional classification of UCVs, identified regions less tolerant of mutations, and raised the testable hypothesis that there are no moderate-risk *BRCA1* missense variants.

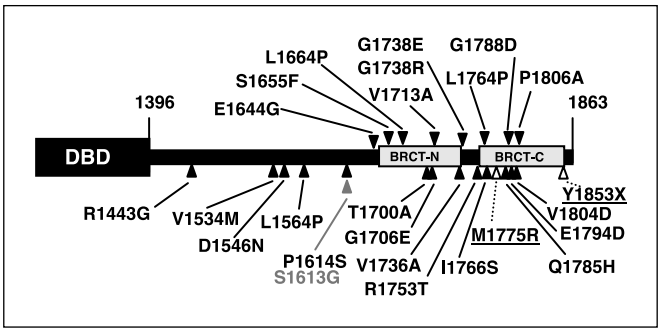

Figure 1. Location of variants in BRCA1. Location of variants (*closed black triangles*) and of negative (*open triangles, underlined*) and positive (*closed gray triangles, gray font*) controls. *Gray boxes,* BRCT domains; *DBD,* GAL4 DNA binding domain.

Materials and Methods

Constructs. Control constructs containing the wild-type (wt) BRCA1, S1613G, M1775R, and Y1853X were previously described (12). Mutations R1443G, V1534M, D1546N, L1564P, P1614S, E1644G, S1655F, L1664P, T1700A, G1706E, V1713A, V1736A, G1738R, G1738E, R1753T, I1766S, L1764P, Q1785H, G1788D, E1794D, V1804D, and P1806A were introduced by splicing by overlapping extension PCR (20) using plasmid p385-BRCA1 as template. Primers sequences are available on request. For each mutation, both products (5′ and 3′ regions) were combined and used as a template for a final round of PCR using 24ENDT and UX13 primers (12). To obtain the double mutants S1613G/V1534M, the same procedure and primers described above were used with the pCDNA3βHA-BRCA1 (gift from Ralph Scully, Beth Israel Deaconess Medical Center, Department of Medicine, Harvard Medical School, Boston, MA), which contains the S1613G polymorphism, as template. For the S1613G/H1402Y, S1613G/L1407P, S1613G/M1628V, and S1613G/T1685I double mutants, previously described primers (12) were used with the pCDNA3βHA-BRCA1 plasmid as template. The final PCR products were then digested with *Bam*H1 and *Eco*R1 and ligated to pLex9 or pGBT9 vectors. All mutations were confirmed by sequencing. To obtain GAL4-DBD fusions in a mammalian expression vector, pGTB9 constructs were digested with *Hin*dIII and *Bam*H1, then a 1.8-kb band was isolated and ligated into equally digested pCDNA3.

Transcription assay in yeast and in mammalian cells. The transcriptional assays were done essentially as described (12, 21). Briefly, *Saccharomyces cerevisiae* strain EGY48 was cotransformed with the effector plasmid pLex9, which contains a fusion of LexA DNA binding domain and BRCA1 amino acids 1,396 to 1,863, with different variants and the plasmid reporter pRB1840, which contains a lacZ gene under the control of one LexA operator (22). At least three individual clones for each variant were tested for liquid β-galactosidase assays using *o*-nitrophenyl-β-D-galactopyranoside and the assays were done in triplicates. Activity was determined as a comparison to wt BRCA1 and S1613G (positive controls) or to M1775R and Y1853X (negative controls). For assays in mammalian cells, we used pG5Luc as a reporter and transfections were normalized with an internal control, phGR-TK (Promega, Madison, WI), which contains a *Renilla* luciferase gene under a constitutive TK basal promoter. Transfections were done with human 293T cells in triplicate using Fugene 6 (Roche, Indianapolis, IN), harvested 28 h posttransfection, and luciferase activity was measured using a dual luciferase assay system (Promega). Western blots were incubated with α-GAL4 DBD mouse monoclonal antibody (Clontech, Mountain View, CA) or α-LexA DBD rabbit polyclonal antibody (Upstate, Charlottesville, VA). We used the SAS application package to calculate confidence intervals for validation.

Structural analysis. For residues located at the BRCT domain, we did an analysis based on the 1t29 structure (23) and a sequence alignment of 13 *BRCA1* orthologues created with SAM-T2K homology search software (24) and subsequently hand-edited. The species used in the alignment are *Homo sapiens* (AAA73985), *Pan troglodytes* (AAG43492), *Gorilla gorilla* (AAT44835), *Pongo pygmaeus* (AAT44834), *Macaca mullata* (AAT44833), *Canis familiaris* (AAC48663), *Bos taurus* (AAL76094), *Monodelphis domestica* (AAX92675), *Mus musculus* (AAD00168), *Rattus norvegicus* (AAC36493), *Gallus gallus* (AAK83825), *Xenopus laevis* (AAL13037), *Tetraodon nigroviridis* (AAR89523). We used the RenderByAttribute routine in the molecular visualization program Chimera to color each residue position according to its percent conservation in the alignment and to visualize patches of conserved residues on the BRCT domain surfaces (25). For each of the eight BRCT UCVs, we computationally replaced the wt side chain with that of the variant and optimized the conformation of the variant backbone and side chain atoms with the mutate model routine in MODELLER (26). Hydrogen bonds in wt and variant protein models were calculated with Chimera FindHBond routine using default parameters. A more detailed description of the method and a computational analysis of 36 BRCT UCVs that integrates features of protein tertiary structure, evolutionary conservation, and amino acid residue properties can be found in a companion paper to our current study (27).

Results

Functional analysis of missense variants. The location of the variants studied, as well as the negative and positive controls, is indicated by arrowheads in Fig. 1. Two known deleterious/high-risk variants, M1775R and Y1853X, were used as negative (i.e., loss-of-function) controls and S1613G (a neutral polymorphism) and wt BRCA1 (amino acids 1,396–1,863) were used as positive controls (12). Because African Americans and Hispanics may disproportionately receive uninformative results (28, 29), we paid particular attention to UCVs found in minority populations. Four variants (P1614S, T1700A, Q1785H, and E1794D) have been documented in African Americans, two (V1713A and G1788D) in Hispanics (ref. 30; Breast Cancer Information Core Database), and two (L1564P and V1804D) in both ethnic groups (30, 31).

Six variants (R1443G, V1534M, D1546N, L1564P, P1614S, and E1644G) that lie upstream of the BRCT domains were investigated for their effect on transcription (Fig. 2). This region displays relatively low conservation across other *BRCA1* orthologues with no recognizable structural motif and was therefore expected to be more tolerant to changes (32). Variants R1443G, P1614S, and E1644G showed transcription activation levels equal or higher than wt BRCA1 whereas V1534M, D1546N, and L1564P had lower activity (between 60% and 80% of the wt activity) in yeast (Fig. 2*A*). In mammalian cells, variants V1534M, D1546N, P1614S, and E1644G showed transcription activation activity comparable to wt (within 1.5 SD) whereas variants R1443G and L1564P had reduced activity (~55% of wt; Fig. 2*B*).

Sixteen UCVs in the BRCT domains (S1655F, L1664P, T1700A, G1706E, V1713A, V1736A, G1738R, G1738E, R1753T, L1764P, I1766S,

G1788D, Q1785H, E1794D, V1804D, and P1806A) were also tested. Variants S1655F, T1700A, G1706E, V1713A, V1736A, G1738R, G1738E, R1753T, L1764P, and 1766S displayed markedly decreased activity with <40% of the wt activity in yeast and <20% in mammalian cells (Fig. 2*A* and *B*). The Q1785H and E1794D variants displayed activity comparable to wt in both yeast and mammalian cells. Variants L1664P and V1804D showed activity comparable to the wt in yeast but ~60% to 80% in mammalian cells, whereas variant P1806A showed activity comparable to the wt in mammalian cells but reduced (64%) in yeast cells. Finally, variant G1788D showed between 30% and 40% activity in yeast or mammalian cells.

Expression levels were comparable for all variants in yeast cells (Fig. 2*C*). In mammalian cells, several variants showed decreased expression levels, suggesting instability of the protein product (Fig. 2*D*). Nevertheless, protein levels in mammalian cells seem to be a poor predictor of overall activity as V1534M or Q1785H, for example, showed very low levels but with activity comparable to wt (Fig. 2*B* and *D*).

As a first approach, we arbitrarily considered >80% and <40% of wt activity as the thresholds to classify the variants as neutral or deleterious, respectively (12). Using this threshold, our data indicate that (*a*) V1534M, P1614S, E1644G, Q1785H, and E1794D do not represent high-risk variants and are likely to be neutral; (*b*) S1655F, T1700A, G1706E, V1713A, V1736A, G1738R, G1738E, R1753T, L1764P, and I1766S represent deleterious/high-risk variants; (*c*) variants R1443G, L1546N, L1564P, L1664P, V1804D, and P1806A do not represent high-risk variants, although we cannot rule out the possibility that they may represent moderate-risk variants; and (*d*) variant G1788D does not represent a neutral variant, although we cannot rule out the possibility that it may represent a moderate-risk variant instead of a deleterious one. Importantly, despite variation in levels in yeast and mammalian assays, no variant presented clearly conflicting results (>80% in one test and <40% in another).

Analysis of double mutants. All previously published assays (3, 12, 13, 33) were done in the context of one molecular haplotype that corresponds to the wt sequence (designated as haplotype 1 in ref. 17). However, there are several frequent *BRCA1* polymorphisms, such as S1613G, and of 10 common *BRCA1* haplotypes, S1613G is present in five (17). Common haplotypes containing S1613G variants do not contribute to disease predisposition, and the presence of S1613G has been determined to have no effect on protein function (12, 33, 34). However, the role of S1613G has not been analyzed in the context of rare haplotypes. To examine whether the co-occurrence in *cis* of the S1613G polymorphism with other variants could affect activity, we combined it with neutral variant H1402Y, located at the coiled-coil motif, because it has been found in combination with S1613G (haplotype 2 in ref. 17). We also combined it with a predicted deleterious variant, L1407P, in

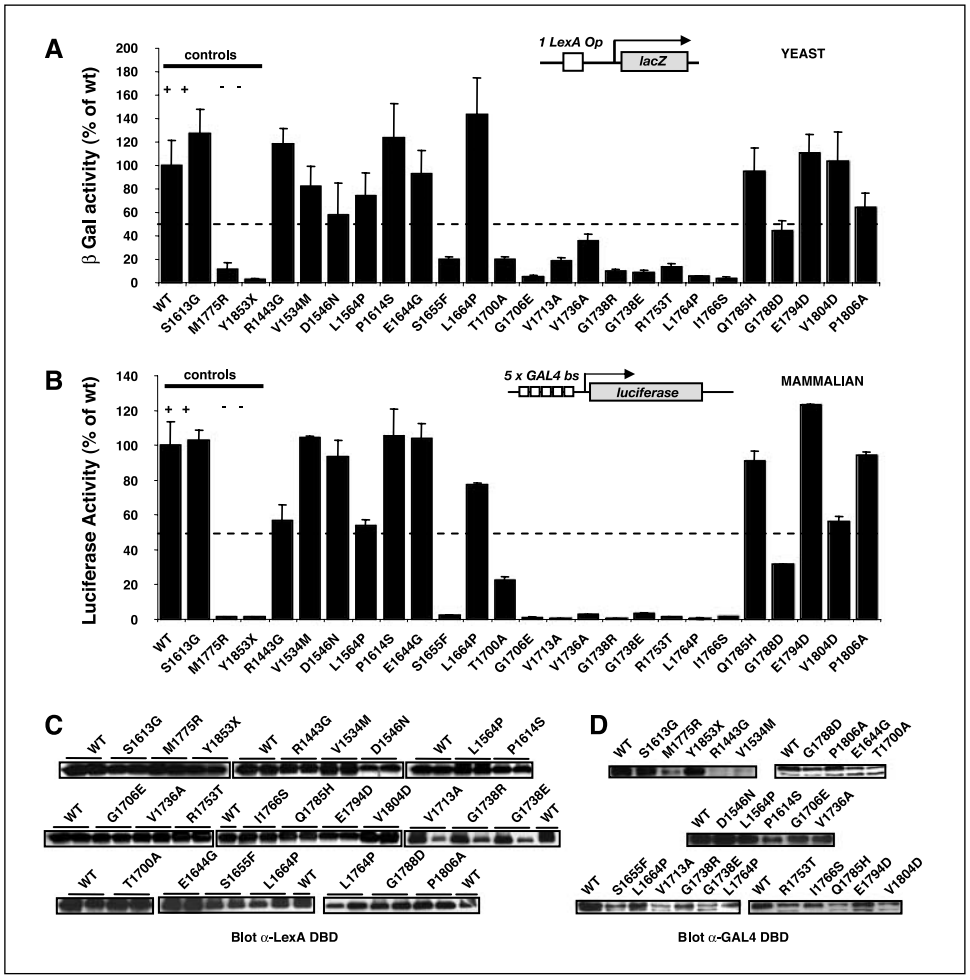

Figure 2. Functional analysis of missense variants in BRCA1. *A,* quantitative transcriptional assay in yeast cells. Cells were cotransformed with a LexA-responsive β-galactosidase reporter gene (*diagram shown above the graph*) and a LexA DBD fusion to residues 1,396 to 1,863 of wt BRCA1, or the same fragment carrying various UCVs. We used the wt and the S1613G neutral polymorphism as positive controls (+). Deleterious mutations M1775R and Y1853X were used as negative controls (−). Three independent yeast clones were tested in triplicates. The activity of the construct with wt BRCA1 was expressed as 100%, with the other results placed on this scale. *B,* quantitative transcriptional assay in mammalian cells. Cells were cotransfected with a GAL4-responsive firefly luciferase reporter gene (*diagram shown above the graph*), a *Renilla* luciferase driven by a constitutive promoter (internal control, not shown), and a GAL4 DNA binding domain (*DBD*) fusion to residues 1,396 to 1,863 of wt BRCA1 (*WT*), or the same fragment carrying various UCVs. Controls are the same as described above but fused to GAL4 DBD. Measurements were done in triplicates and normalized against the internal transfection controls. The activity of the construct with wt *BRCA1* was expressed as 100%, with the other results placed on this scale. To control for possible variations in protein expression levels, samples were analyzed by Western blot with rabbit anti-LexA DBD polyclonal antibody in yeast extracts (*C*) or mouse anti-GAL4 DBD monoclonal antibody in mammalian cell extracts (*D*).

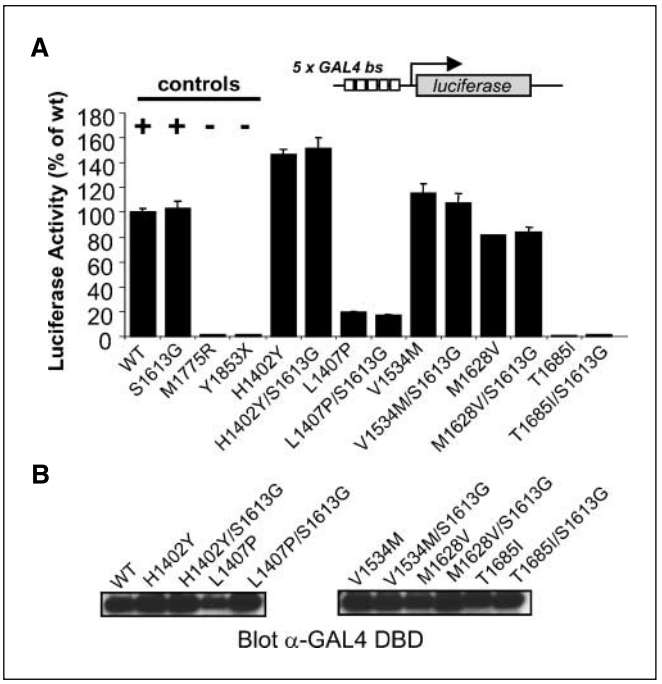

Figure 3. Functional analysis of double mutants in BRCA1. *A,* quantitative assay in human 293T HEK cells done as described in Fig. 2. No difference was found in the activity between pairs of rare variant and rare variant combined with S1613G polymorphism. *B,* protein levels determined by Western blot with mouse anti-GAL4 DBD monoclonal antibody.

the same motif (12). In addition, we arbitrarily chose M1628V, located in the vicinity of the polymorphism, and two other variants, V1534M and T1685I, located at approximately the same distance from the polymorphism. The presence of S1613G did not affect the activity of deleterious (S1613G/L1407P and S1613G/T1685I) or neutral (S1613G/H1402Y, S1613G/V1534M and S1613G/M1628V) variants (Fig. 3).

Clinical data. Pedigrees corresponding to six variants analyzed here illustrate the difficulties in inferring causality even in large kindreds (Supplementary Fig. S1).

V1534M. The proband in this Italian family (M933) was diagnosed with ovarian and bilateral breast cancer and tested positive for the variant. Loss of heterozygosity (LOH) and sequencing analysis in one of the proband's breast tumors revealed loss of the wt allele. We also analyzed DNA from a histologically cancer-free specimen of breast tissue from the sister diagnosed with breast cancer at age 40 years and it revealed wt alleles only. Therefore, the LOH analyses from these two individuals are not conclusive.

D1546N. In family 230, four individuals were tested and two women with early-onset breast cancer do not carry the variant. However, it is possible that the variant derives from the mother's side of the family and the information from the father's side of the family is not relevant to conclude the absence of segregation.

P1614S. The African American family 2593 presents with six cases of breast cancer and one ductal carcinoma *in situ* and also carries the H2116R *BRCA2* UCV. The proband and her sister with ductal carcinoma *in situ* tested positive for the variant, suggesting that this variant could account for the disease.

E1664G. In addition to the proband, we tested the mother who was found not to carry the variant. Excluding the possibility of nonpaternity, this suggests that the variant is neutral because no cancer was reported in the deceased father and his family.

V1736A. Three specimens from two independent tumors from the proband of AUS, who had bilateral breast cancer, tested positive for the variant. Two specimens from the left breast tumor showed LOH of the mutant allele and one specimen from the right breast tumor showed no LOH.

V1804D. For family 1008, only two individuals affected with breast cancer were tested and both were shown to carry the variant. The proband's paternal side on family 922 is Native American. All three individuals affected with breast cancer tested carry the variant.

Structural analysis. To structurally rationalize the loss or retention of BRCA1 transcriptional activation in our panel of missense variants, we have divided the variants located at the BRCT domains into four categories: (*a*) putative disruption of the BRCT fold hydrophobic core; (*b*) putative disruption of the BRCA1 BRCT interaction with phosphorylated protein partners; (*c*) putative disruption of binding sites on the protein surface; and (*d*) no evidence for functional effect (Table 1).

Validation. To evaluate the specificity and sensitivity of the transcription assay, we identified missense variants in the COOH terminus of BRCA1 that have been classified as either deleterious or neutral based on (*a*) the co-occurrence with other deleterious variants because compound heterozygous mutation leads to embryonic lethality in humans with high statistical probability (9, 17) and (*b*) integrated methods using multifactorial likelihood models (9–11, 17).[17] The thresholds for the overall odds for or against causality in the likelihood models are arbitrary. Thus, we choose to use both highly stringent (neutral: overall combined odds >100:1 against causality; deleterious: overall combined odds >1,000:1 for causality; ref. 10) and less stringent (neutral: >10:1 against causality; deleterious: >10:1 for causality) thresholds. Fourteen variants were classified as neutral and 10 variants classified as deleterious using less stringent threshold (Table 2). A more stringent threshold reduces the number of classified neutral and deleterious variants to 13 and 6, respectively.

The transcription assay correctly classified all 24 variants (Table 2; Fig. 4). Because our values for sensitivity and specificity are both 100%, there is no defined upper bound for our confidence intervals. We can exclude with 95% confidence that the sensitivity is not lower than 69% (based on 10 of 10 samples) or 54% (based on 6 of 6 samples; using a more stringent threshold) and the specificity is not lower than 77% (based on 14 of 14 samples) or 75% (based on 13 of 13 samples; using a more stringent threshold).

Missense variants as a group. One important question about risk determination using functional assays is how to interpret quantitative results. It is possible that risk is inversely correlated with protein activity in a continuous fashion (Fig. 4*A*). In that case, variants will present as a continuous series of high-risk, intermediate-risk, and low-risk variants. However, it is also possible that variants will be either high or low risk with virtually no moderate-risk variant found in the population (Fig. 4*B*). To begin to address this problem, we have plotted the activity (and its range of variation) of all the 47 variants tested thus far in the transcription activation assay in a quantitative manner (Fig. 4*C*). Next, we identified a validated deleterious variant that displayed the highest activity (R1699W; Table 2; Fig. 4*C*) and a validated neutral variant that displayed the lowest activity (L1564P; Table 2; Fig. 4*C*).

[17] D. Goldgar, unpublished data.

Table 1. Structure-based analysis of BRCA1 variants in the BRCT domains

Class	Variant	Notes
Hydrophobic core disruption	G1706E	G1706 in BRCT-N is completely inaccessible to solvent. This position is in the α2 helix and sits between two positively charged Lys residues, one of which (K1702) forms hydrogen bonds with SEP6, a phosphorylated Ser critical to BRCA1 phosphopeptide binding. The substitution by Glu introduces a buried negative charge, which is energetically unfavorable.
	V1713A	V1713 lies in the β4 strand of the four-stranded parallel β-sheet in BRCT-N. In BRCT tandem domains, hydrophobic, β-branched amino acid residues are conserved at this position and are probably critical to the stability of the BRCT tandem domain fold. β-branched amino acid residues are generally favored in β sheets and the Ala, although hydrophobic, is not β-branched.
	V1736A	V1736 is completely buried in a tightly packed neighborhood, rich in hydrophobic residues and methyl(ene) groups. The Ala replacement results in loss of a single methyl group and creation of a small cavity in this neighborhood. The destabilizing effects of such cavities have been shown to be correlated with the number of methyl(ene) groups within 6 Å of a mutated position (44), so this cavity is likely destabilizing. Interestingly, both valines in the conserved $W^{1712}VVS^{1715}$ motif, critical for the stability of the BRCT fold, are within 4 Å of this position.
	G1738R, G1738E	G1738 is in the linker region connecting the two BRCT domains and is completely inaccessible to solvent. Our structural modeling indicates that introduction of either an Arg or Glu at this position creates numerous steric clashes. Both Arg and Glu are charged amino acid residues, and placement of a charged residue in a buried position is expected to be destabilizing.
	I1766S	I1766 is completely buried in the hydrophobic core of BRCT-C. The replacement by the polar serine introduces a side chain oxygen atom (OG) that can function as a hydrogen bond donor or acceptor. Inspection of the local environment indicates that there are no available donor or acceptor atoms to satisfy the H-bonding potential of the OG, yielding a destabilized structure.
Phosphopeptide binding disruption	S1655F	S1655 is a critical residue for phospho-Ser recognition, and forms hydrogen bonds with the phospho-Ser pSer0 of CtIP peptide in the BRCA BRCT tandem domain X-ray crystal structure 1y98 (45) and with the phospho-Ser pSer6 of BACH1 peptide in 1t29 and 1t2v (41). Replacement of Ser by Phe breaks hydrogen bonds between S1655 and pSer and also creates steric clashes between pSer and the phenyl ring in all of these structures.
	T1700A	T1700 is part of a hydrogen bonding network in the holo structures of the BRCA1 BRCT domains in complex with BACH1 peptide [PDB 1t29 and 1t2v] (23, 41) and in complex with CtIP peptide [PDB 1y98] (45). The H-bond network includes the phosphorylated Ser (SEP6) of the BACH1 and CtIP, identified as a critical residue in phosphopeptide binding to BRCA1. In these structures, the side chain oxygen OG1 of T1700 forms a H-bond with the side chain OG1 oxygen of S1655, which in turn forms a H-bond with the phosphate oxygen O1P in the phosphorylated Ser SEP6.
	L1764P	L1764 is in the β1 strand of BRCT-C. We predict fold destabilization upon Pro substitution because this position is a key hydrophobic residue in the BRCT superfamily (2) and Pro is not hydrophobic. Furthermore, Pro is compatible with a limited number of backbone conformations, preferring a PHI backbone dihedral angle close to −60 deg. Position 1,764 has a PHI angle of −112 deg that is not compatible with Pro.
Disruption of putative binding sites	R1753T	R1753 is in the linker region that connects the two BRCT domains. The substitution by Thr results in loss of a positively charged side chain in an exposed position. Here, the position occurs in a surface patch highly conserved among BRCA1 orthologues from mammals through *Xenopus* and *Tetraodon*. The surface electrostatic potential, as calculated by DELPHI using united atom AMBER charges, is substantially different for the Arg and Thr variants. The substitution produces an acidic patch on the surface where the Thr is exposed and observable changes in electrostatic potential around neighboring exposed residues in the linker region. The high conservation in this region may be the signature of a BRCA1-specific binding site with a protein partner. Interestingly, the linker region in the 53BP1 BRCT dimer has been identified as a binding site of p53 (46, 47).
No evidence for functional impact	L1664P	L1664 lies at the second turn of an α-helix in BRCT-N (helix α1). In general, Pro amino acid residues past the first turn of a helix are destabilizing because of steric clashes with backbone atoms. However, although helix α1 may be destabilized by P1664, the results of our assay indicate that the effect is not strong enough to disrupt the hydrophobic core of the protein. The Pro side chain bonds to a backbone nitrogen atom and is only compatible with a limited number of backbone conformations, preferring a PHI backbone dihedral angle close to −60 deg. Here, position 1,664 has a PHI angle of −57.5 deg that is compatible with Pro.
	Q1785H	Q1785 in BRCT-C is on a protruding area of the protein surface that is not conserved among BRCA1 orthologues. It lies on the exposed, hydrophilic side of the amphipathic helix α1′ and sticks out into solvent. Although the substitution with His breaks a possible H-bond between the NE2 nitrogen of the Gln and the backbone carbonyl oxygen of Q1781, the Q1781 acceptor can form a compensating H-bond with solvent, or possibly the backbone nitrogen of Q1785.

(*Continued on the following page*)

Table 1. Structure-based analysis of BRCA1 variants in the BRCT domains (Cont'd)

Class	Variant	Notes
	E1794D	E1794 in BRCT-C sticks out into solvent. Substitution by Asp, another negatively charged residue, is not likely to have any functional effect.
	V1804D	V1804 in BRCT-C is on a flexible loop and sticks out into solvent. Introduction of a charged residue on the surface is not likely to affect function unless it is at a binding site. This position is probably not a binding site because it lies on a surface region that is not conserved among BRCA1 orthologues.
	P1806A	P1806 in BRCT-C is found in β-strand β3 and is solvent exposed. Substitution by Ala is not likely to have any functional effect. This position is probably not a binding site because it lies in a surface region that is not conserved among BRCA1 orthologues.

Interestingly, these variants provide a very narrow intermediate range of activity (<50% and >45%). This result suggests that the assay classifies variants either as high or low risk with no variant considered as moderate risk.

Discussion

Reliable classification of *BRCA1* alleles containing missense changes remains a top priority for risk assessment for breast and ovarian cancer. Unfortunately, it seems clear that in the case of most rare variants, no single data source is informative enough to unambiguously classify them into neutral or deleterious (10, 19). To contribute to the classification of these problematic alleles, we have developed a functional test based on transcription activation that has provided a reproducible and standardized way to assess the functional effect of these variants on protein function (12, 13, 21). The transcription assay is a monitor for the integrity of the BRCT domain, and there is emerging evidence that it can also reliably predict functional effect in a region preceding the BRCT domains including the coiled-coil motif (3, 5, 12, 13, 33). Recent experimental evidence has indicated that BRCT domains are specialized motifs that bind phosphorylated peptides (23, 35–41). Importantly, the BRCT region involved in phosphopeptide recognition colocalizes with a region that is critical for transcription activation in the heterologous system we use (18) and may underline a structural basis for the correlation between the transcription assay and the integrity of the BRCT domain.

Table 2. Cross validation of the transcription assay with genetic data

Variants*	Classification	References
H1402Y	Neutral	Judkins et al. (17); Tavtigian et al. (9)
R1443G	Neutral	Overall odds against causality† >50:1
S1512I	Neutral	Tavtigian et al. (9)
V1534M	Neutral	Chenevix-Trench et al. (11)
D1546N	Neutral	Tavtigian et al. (9)
T1561I‡	Neutral	Judkins et al. (17)
L1564P	Neutral	Judkins et al. (17); Tavtigian et al. (9)
S1613G	Neutral	Tavtigian et al. (9)
P1614S	Neutral	Judkins et al. (17)
M1628T	Neutral	Judkins et al. (17); Tavtigian et al. (9)
M1652I‡	Neutral	Tavtigian et al. (9)
A1669S	Neutral	Judkins et al. (17)
R1699W	Deleterious	Overall odds for causality† >100,000:1
R1699Q	Deleterious	Chenevix-Trench et al. (11)
G1706E	Deleterious	Overall odds for causality† >1,000:1
A1708E	Deleterious	Overall odds for causality† >1,000:1
S1715R	Deleterious	Overall odds for causality† >50:1
T1720A	Neutral	Overall odds against causality† >1,000:1
G1738R	Deleterious	Chenevix-Trench et al. (11)
L1764P	Deleterious	Overall odds for causality† >100:1
I1766S	Deleterious	Overall odds for causality† >100:1
M1775R	Deleterious	Overall odds for causality† >1,000:1
G1788V	Deleterious	Overall odds for causality† >1,000:1
V1804D	Neutral	Overall odds against causality† >100:1

*Variants in bold were classified with stringent thresholds (see text).
† D. Goldgar, unpublished results.
‡ Transcriptional assays done in a qualitative manner (5, 33).

One problem with a quantitative approach for classification is at which activity level one should draw a cutoff value. To address this problem, we analyzed all variants tested to date and identified the ones that had been classified using integrated methods. The identification of the neutral variant showing the lowest activity (L1564P) and of the deleterious variant with the highest activity (R1699W) led to the surprising finding that, given the small interval between these variants, there were only two classes (high risk and low risk) with no intermediate (moderate risk) class (Fig. 4). This supports the assumption used in the proposition of the integrated method that variants were either high or low risk (10). Whereas this may reveal a limitation of the assay, it may also reflect a biological property of *BRCA1* alleles. Further studies are needed to investigate this notion with important implications for risk assessment.

Whereas it has been estimated that ~13% of individuals undergoing testing receive uninformative reports due to the finding of an UCV, the picture is likely to be more somber for members of minority populations. In a recent study, it was shown that African Americans had a larger number of UCVs (46%) than Caucasians (12%; ref. 29). Therefore, we analyzed eight variants (L1564P, P1614S, T1700A, V1713A, Q1785H, G1788D, E1794D, and V1804D) that have been found in African Americans and/or Hispanics (30, 31, 42).

Using the threshold of activity defined in Fig. 4 (>50% for neutrals and <45% for deleterious), variants T1700A, V1713A, and G1788D displayed activity compatible with a deleterious classification, whereas L1564P, P1614S, Q1785H, E1794D, and V1804D displayed activity compatible with a neutral classification (Fig. 4). For the L1564P, P1614S, and V1804 variants, the functional results

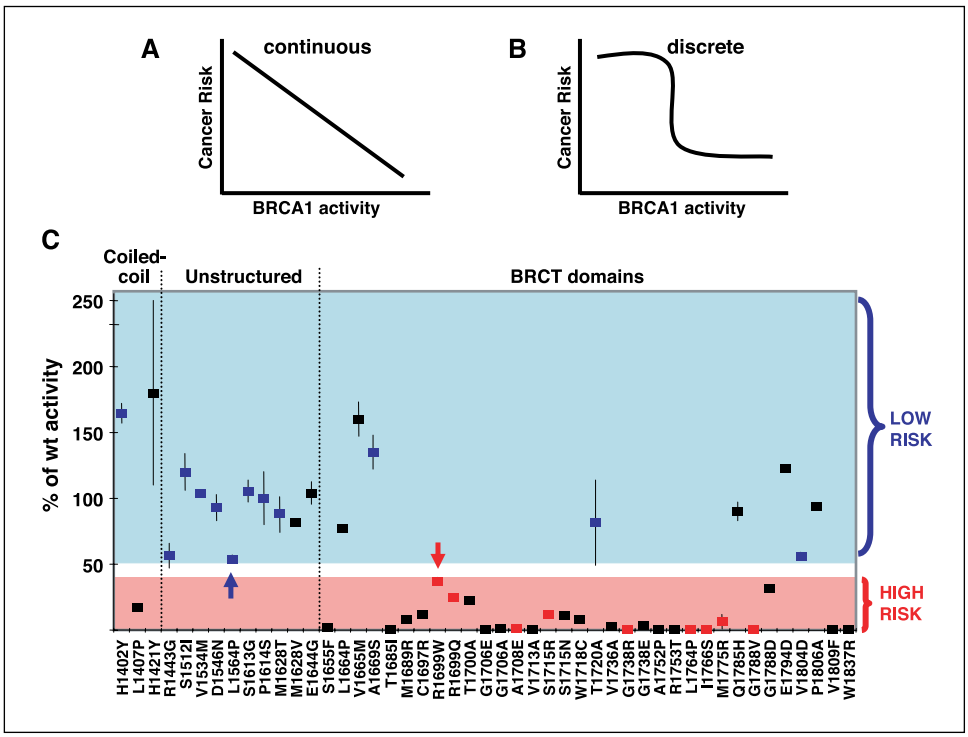

Figure 4. Missense variants and risk. *A,* hypothetical correlation between protein activity and cancer risk assuming a continuous relationship. *B,* hypothetical correlation between protein activity and cancer risk assuming a discrete relationship. *C,* functional analysis of 47 variants tested in a quantitative and standardized fashion plotted as percent of wt activity. *Squares,* average activity of the variant; *bars,* range of interexperiment (replicates in the same experiment) and intraexperiment (assays done at different times) variability. *Red* and *blue squares,* variants that have been classified by genetic and/or integrated methods as deleterious or neutral, respectively (Table 2). *Red* and *blue arrows,* deleterious variant with highest activity and neutral variant with lowest activity, respectively.

are consistent with co-occurrence analyses indicating that they are neutral variants (Table 1; ref. 17). Structural analysis supports a neutral classification for Q1785H and V1804D and a deleterious classification for T1700A, V1713A, and G1788D (Table 2). Interestingly, the T1700 residue is part of the phosphoserine binding pocket in which it makes hydrogen bonding interactions with the serine hydroxyl group, and it may play an important role in binding specificity (35, 36, 41). These results also highlight the difficulty in relying on limited pedigree data that may suggest cosegregation with the cancer phenotype (Supplementary Fig. S1).

We also did functional tests on 14 additional variants and functionally classified six (R1443G, V1534M, D1546N, E1644G, L1664P, and P1806A) as neutral and eight (S1655F, G1706E, V1736A, G1738R, G1738E, R1753T, L1764P, and I1766S) as deleterious (Fig. 4). The V1534M classification is supported by co-occurrence data (Table 2) whereas the D1546N classification is supported by clinical data, specifically from family 230 (Supplementary Fig. S1) in which at least two affected women have been shown not to carry the variant. Structural analysis is consistent with the classification of all BRCT variants (S1655F, L1664P, G1706E, V1736A, G1738R, G1738E, R1753T, L1764P, I1766S, and P1806A; Table 1). However, caution should be exercised in the classification until further evidence for these thresholds is obtained.

S1613G is a common polymorphism that displays a wide geographic occurrence but does not make a significant contribution to breast or ovarian cancer risk (34). Its high allele frequency, ranging from 0.2 to 0.55 in various populations, makes it probable that haplotypes may exist with S1613G in combination with other neutral or deleterious rare variants. The S1613G variant is present in 5 of the 10 most common *BRCA1* haplotypes (17). This raised the question of whether there could be an interaction between these missense changes that confer different properties to these haplotypes. We investigated this by comparing naturally occurring and hypothetical haplotypes containing a combination of neutral and deleterious variants. We found that haplotypes containing Gly^{1613} were no different than haplotypes containing Ser^{1613} in terms of activity in the functional assays. Although we cannot rule out that the transcriptional assay may not be sensitive to detect small changes and that these haplotypes may behave differently *in vivo*, our results suggest that S1613G does not modify the risk conferred by other missense variants in *cis*. This, however, may not hold true for other common polymorphisms.

In conclusion, we have tested a series of common and rare UCVs of BRCA1 with an emphasis on variants found in minority ethnic groups. Combining functional, structural, and clinical information, we classified these alleles as either deleterious or neutral. We have also shown that the S1613G polymorphism does not alter the effect of other rare variants. Importantly, our analysis of 47 variants allowed us to define a tentative threshold of activity for classification and to propose the hypothesis that these rare missense variants confer either high or low risk but not moderate risk. If proved, this hypothesis will have profound implications for genetic counseling.

As a discipline, genetic counseling has seen a tremendous transformation. In the recent past, genetic counseling focused almost exclusively on pediatric syndromes that, by and large, are completely penetrant and do not present confounding phenocopies. Presently, issues of genetic counseling for cancer have come to the forefront. Data from Ontario illustrate this trend, with breast cancer consultations surpassing the number of consultations for all other reasons, including pediatric conditions and fertility issues (43). Many important problems surface about the determination of individual risk of cancer when incomplete penetrance and frequent phenocopies confuse the picture, as is the case with *BRCA1*. Thus, extensive family genotyping, complementary methods for detection, and classification of alleles become paramount.

Acknowledgments

Received 9/5/2006; revised 11/13/2006; accepted 12/15/2006.

Grant support: NIH grant CA92309 (A.N.A. Monteiro); NIH Breast Cancer Specialized Program of Research Excellence Award CA116201-P2 (F.J. Couch); American Cancer Society Research Scholar grant (F.J. Couch); Associazione Italiana per la Ricerca sul Cancro/Fondazione Italiana per la Ricerca sul Cancro, Special Project Hereditary Tumors grant (P. Radice); Italian Ministry of Instruction, University and Research grant RBNE014975 (P. Radice); NIH grant F32 GM072403-02 (R. Karchin); NIH grants R01 GM54762 and U01 GM61390, the Sandler Family Supporting Foundation, IBM, Hewlett Packard, Intel, and NetApps (A. Sali); and the Molecular Imaging and the Molecular Biology cores at the H.L. Moffitt Cancer Center.

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.

We thank the members of the Monteiro Lab for helpful comments; Susan Domcheck, Linda Wadum, Kiley Johnson, and Jenny Mentlick for collecting family data and samples; Carla B. Ripamonti for performing analysis on surgical specimens; and Dana Rollison for help with statistical analysis.

References

1. Smith DJ, Lusis AJ. The allelic structure of common disease. Hum Mol Genet 2002;11:2455–61.
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. Monteiro AN, August A, Hanafusa H. Evidence for a transcriptional activation function of BRCA1 C-terminal region. Proc Natl Acad Sci U S A 1996;93:13595–9.
4. Chapman MS, Verma IM. Transcriptional activation by BRCA1. Nature 1996;382:678–9.
5. 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.
6. Thompson D, Easton DF, Goldgar DE. A full-likelihood method for the evaluation of causality of sequence variants from family data. Am J Hum Genet 2003;73:652–5.
7. Fleming MA, Potter JD, Ramirez CJ, Ostrander GK, Ostrander EA. Understanding missense mutations in the BRCA1 gene: an evolutionary approach. Proc Natl Acad Sci U S A 2003;100:1151–6.
8. Abkevich V, Zharkikh A, Deffenbaugh AM, et al. Analysis of missense variation in human BRCA1 in the context of interspecific sequence variation. J Med Genet 2004;41:492–507.
9. Tavtigian SV, Deffenbaugh AM, Yin L, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet 2006;43:295–305.
10. Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro AN, Tavtigian SV, Couch FJ. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. Am J Hum Genet 2004;75:535–44.
11. Chenevix-Trench G, Healey S, Lakhani S, et al. Genetic and Histopathologic Evaluation of BRCA1 and BRCA2 DNA Sequence Variants of Unknown Clinical Significance. Cancer Res 2006;66:2019–27.
12. Phelan CM, Dapic V, Tice B, et al. Classification of BRCA1 missense variants of unknown clinical significance. J Med Genet 2005;42:138–46.
13. 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.
14. 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 U S A 1997;94:5820–5.
15. Coyne RS, McDonald HB, Edgemon K, Brody LC. Functional Characterization of BRCA1 Sequence Variants Using a Yeast Small Colony Phenotype Assay. Cancer Biol Ther 2004;3:453–7.
16. Morris JR, Pangon L, Boutell C, Katagiri T, Keep NH, Solomon E. Genetic analysis of BRCA1 ubiquitin ligase activity and its relationship to breast cancer susceptibility. Hum Mol Genet 2006;15:599–606.
17. Judkins T, Hendrickson BC, Deffenbaugh AM, et al. Application of embryonic lethal or other obvious phenotypes to characterize the clinical significance of genetic variants found in trans with known deleterious mutations. Cancer Res 2005;65:10096–103.
18. Mirkovic N, Marti-Renom MA, Weber BL, Sali A, Monteiro AN. Structure-based assessment of missense mutations in human BRCA1: implications for breast and ovarian cancer predisposition. Cancer Res 2004;64:3790–7.
19. Monteiro AN, Couch FJ. Cancer risk assessment at the atomic level. Cancer Res 2006;66:1897–9.
20. Higuchi R. Recombinant PCR. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, editors. PCR protocols: a guide to methods and applications. San Diego (CA): Academic Press, Inc.; 1989. p. 177–83.
21. Carvalho MA, Couch FJ, Monteiro AN. Functional assays for BRCA1 and BRCA2. Int J Biochem Cell Biol 2007;39:298–310. Epub 2006 Aug 18.
22. Estojak J, Brent R, Golemis EA. Correlation of two-hybrid affinity data with *in vitro* measurements. Mol Cell Biol 1995;15:5820–9.
23. Shiozaki EN, Gu L, Yan N, Shi Y. Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling. Mol Cell 2004;14:405–12.
24. Karplus K, Karchin R, Barrett C, et al. What is the value added by human intervention in protein structure prediction? Proteins 2001;Suppl 5:86–91.
25. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera-a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–12.
26. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
27. Karchin R, Monteiro ANA, Tavtigian SV, Carvalho MA, Sali A. Functional impact of missense variants in BRCA1 predicted by supervised learning. PLoS Comp Biol. In press 2007.
28. Nanda R, Schumm LP, Cummings S, et al. Genetic testing in an ethnically diverse cohort of high-risk women: a comparative analysis of BRCA1 and BRCA2 mutations in American families of European and African ancestry. JAMA 2005;294:1925–33.
29. Haffty B, Silber A, Matloff E, Chung J, Lannin D. Racial Differences in the incidence of BRCA1 and BRCA2 mutations in a cohort of Early Onset Breast Cancer patients: African American compared to White Women. J Med Genet 2006;43:133–7.
30. Weitzel JN, Lagos V, Blazer KR, et al. Prevalence of BRCA mutations and founder effect in high-risk Hispanic families. Cancer Epidemiol Biomarkers Prev 2005;14:1666–71.
31. McKean-Cowdin R, Spencer FH, Xia LY, et al. BRCA1 variants in a family study of African-American and Latina women. Hum Genet 2005;116:497–506.
32. Szabo CI, Worley T, Monteiro AN. Understanding germ-line mutations in BRCA1. Cancer Biol Ther 2004;3:515–20.
33. Monteiro AN, August A, Hanafusa H. Common BRCA1 variants and transcriptional activation. Am J Hum Genet 1997;61:761–2.
34. Dunning AM, Chiano M, Smith NR, et al. Common BRCA1 variants and susceptibility to breast and ovarian cancer in the general population. Hum Mol Genet 1997;6:285–9.
35. Botuyan MV, Nomine Y, Yu X, et al. Structural basis of BACH1 phosphopeptide recognition by BRCA1 tandem BRCT domains. Structure (Camb) 2004;12:1137–46.
36. Glover JN, Williams RS, Lee MS. Interactions between BRCT repeats and phosphoproteins: tangled up in two. Trends Biochem Sci 2004;29:579–85.
37. Manke IA, Lowery DM, Nguyen A, Yaffe MB. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. Science 2003;302:636–9.
38. Rodriguez M, Yu X, Chen J, Songyang Z. Phosphopeptide binding specificities of BRCA1 COOH-terminal (BRCT) domains. J Biol Chem 2003;278:52914–8.
39. Yu X, Chini CC, He M, Mer G, Chen J. The BRCT domain is a phospho-protein binding domain. Science 2003;302:639–42.
40. Clapperton JA, Manke IA, Lowery DM, et al. Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer. Nat Struct Mol Biol 2004;11:512–8.
41. Williams RS, Lee MS, Hau DD, Glover JN. Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. Nat Struct Mol Biol 2004;11:519–25.
42. Panguluri RC, Brody LC, Modali R, et al. BRCA1 mutations in African Americans. Hum Genet 1999;105:28–31.
43. Andermann A, Narod SA. Genetic counselling for familial breast and ovarian cancer in Ontario. J Med Genet 2002;39:695–6.
44. Serrano L, Kellis JT, Jr., Cann P, Matouschek A, Fersht AR. The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. J Mol Biol 1992;224:783–804.
45. Varma AK, Brown RS, Birrane G, Ladias JA. Structural basis for cell cycle checkpoint control by the BRCA1-CtIP complex. Biochemistry 2005;44:10941–6.
46. 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.
47. 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.

Determination of Cancer Risk Associated with Germ Line BRCA1 Missense Variants by Functional Analysis

Marcelo A. Carvalho, Sylvia M. Marsillac, Rachel Karchin, et al.

Cancer Res 2007;67:1494-1501.