Journal of the Indian Institute of Science
A Multidisciplinary Reviews Journal
ISSN: 0970-4140 Coden-JIISAD

REVIEWS

Analyzing BRCA1 Variants of Unknown Significance by Bioinformatics

*Balaiah Meenakumari and Thangarajan Rajkumar**

Abstract | BRCA1 and BRCA2 are the two major genes responsible for predisposition to breast and ovarian cancers. Mutations in BRCA1 gene have a greater risk of developing cancer compared to BRCA2 gene. Many missense variants have been reported for BRCA1 gene, of which only 2% variants are found to have clinical significance. Most of the missense variants are called uncertain/unknown variants for which the clinical significance is not known. The harmfulness of these Variants of unknown significance (VUS) was studied by a bioinformatics approach.

Keywords: *BRCA1, BRCA2, Missense, HBOC and VUS*

1 Introduction

The two main genes responsible for predisposition to breast and ovarian cancers are BRCA1 and BRCA2, often referred to as "breast carcinoma genes". BRCA1 and BRCA2 are tumor suppressor genes and they contribute to many cellular processes including homologous recombination, DNA damage response, cell cycle checkpoint control, ubiquitination, transcriptional regulation, chromatin modification, centrosome duplication and X-chromosome inactivation.[1,2] Both BRCAs are in the class of so-called caretaker genes, which through their multiple functions use a variety of pathways to ensure genomic stability. Mutations in BRCA1 and BRCA2 cause genomic instability, which lead to alterations in additional key genes including Tumor Suppressor Genes and/or oncogenes. Risks of cancer conferred by inherited mutation in BRCA1 are likely to be greater compared to BRCA2 in hereditary breast and ovarian cancers.[3,4]

Genetic variations in BRCA1 are large genomic rearrangements, pathogenic protein truncating mutations, deletions/insertions and some point variants. The point mutations include Missense mutation, non-sense mutation and synonymous mutations. Genetic variations which occur inside the coding region of the gene have effects in regulation and its expression. Even variations occurring in the near splice site of the gene may alter the function of the genes.

1.1 *Effects of the point mutation*

In case of non-sense mutation change in the nucleotide results in amino acid chain termination, and therefore truncation of the polypeptide chain, which in turn leads to change in the function of the gene.

In missense mutation, a substitution of amino acid for the original one may lead to an alteration in structure of the protein and thus its function.

Synonymous mutations do not lead to substitution of the amino acids, but still may cause alteration in mRNA folding and hence translation of proteins.[5]

All these genetic variations somehow affect the structure and function of the protein, but for some of the missense mutations their functional and structural significance was unknown. To date, in the Breast Cancer Information Core Database (BIC database), 567 distinct missense variants have been reported for BRCA1 gene, of which only

Department of Molecular Oncology, Cancer Institute (WIA), Adyar, Chennai 600020, India.
**drtrajkumar@gmail.com*

2% variants are found to have clinical significance. (http://research.nhgri.nih.gov/projects/bic). Most of the missense variants are called uncertain/ unknown variants for which the clinical significance is not known. Our study focuses on the harmfulness of these unknown variants.

1.2 *Structure of BRCA1*

BReast-**CA**ncer susceptibility gene, BRCA1 is an oncogene located on chromosome 17q21, which is of autosomal dominant inheritance pattern. BRCA 1 gene spans 80kb of genomic sequence with 5592 nucleotides and is composed of 22 coding exons distributed over 100 kb of genomic DNA; this gene encodes 1863-residue protein (NCBI amino acid sequence) involved in gene regulation and repair process following DNA damage.[6] More than 200 different germ line mutations associated with cancer susceptibility have been identified. Many disease-predisposing alleles of BRCA1 have loss-of function mutations, the majority of which result in premature truncation of the protein.[2] BRCA1 protein contains an N-terminal RING domain, nuclear localization signals (NLSs), and two C-terminal BRCT domains of ~110 residues. About 36% of all BRCA1 mutations constitute missense mutations (of those, 5.2% are polymorphisms, 7.8% are deleterious, and 87% are unclassified variants), which occur throughout the whole protein sequence.[7]

1.3 *Domain region and their role*

The three domain regions like Zinc ring finger domain (amino acid position 24–64), RAD51 binding domain (757–1064) and BRCT domain (1653–1855) in BRCA1 found to interact with many proteins. Most of the diseases associated missense mutations occur within the N-terminal RING domain and the C-terminal BRCT domain, indicating that these regions have vital tumor suppressor function of BRCA1.[8]

Zinc finger (Znf) domains are relatively small protein motifs which contain multiple finger-like protrusions that make tandem contacts with their target molecule. Some of these domains bind zinc, but many instead bind other metals such as iron or no metal at all. Their binding properties depend on the amino acid sequence of the finger domains and the linker between fingers, as well as on the higher-order structures and the number of fingers. Many proteins containing the RING finger play a key role in the ubiquitination pathway by binding with BRCA1 associated ring domain protein (BARD1). This complex (BRCA1-BARD1) acts as a E3 ubiquitin ligase. Missense variants in this region may affect the ubiquitin activity of this domain.

The carboxyl-terminal BRCT domain acts as a phospho-protein binding domain. Although most BRCT domain-containing proteins participate in DNA-damage checkpoint or DNA-repair pathways, or both, the function of the BRCT domain is not fully understood.[3] The BRCA1 BRCT domain directly interacts with phosphorylated BRCA1-Associated Carboxyl-terminal Helicase (BACH1). This specific interaction between BRCA1 and phosphorylated BACH1 is cell cycle regulated and is required for DNA damage-induced checkpoint control during the transition from G2 to M phase of the cell cycle.[9] Further, the two BRCT domains interact with their respective physiological partners in a phosphorylation-dependent manner. Mutations in this domain region result in the truncation and structural alteration of the protein.

A variety of tools have been used to assess the clinical and functional relevance of the unknown variants in this gene.

2 Materials and Methods

2.1 *Materials*

From 2002 to 2011, 576 Hereditary Breast and/or Ovarian cancer (HBOC) unrelated families have been registered in Hereditary Cancer Clinic, Cancer Institute (WIA), Chennai. Of this, 311 families provided their informed consent and blood sample for mutation analysis. We analysed 150 of these families for mutations in BRCA1 & 2 genes.[4]

2.2 *Methods*

2.2.1 Mutation screening: The sample collection, DNA isolation, PCR-DHPLC (Denaturing High performance Liquid Chromatography) and sequencing were carried out as previously described.[4,10] The sequencing results were analysed to check the variants of unknown Significance by BIC portal. Four variants of unknown significance detected were included to predict the functional effect.

2.3 *Bioinformatics approach to predict the function of variants of unknown significance (VUS)*

The PDB structure for the BRCA1 domains: BRCT region and zinc ring finger region were downloaded from the Protein data bank database, since the complete structure was not present. All the missense mutations falling in the two domain regions were extracted from the BIC database and our sequencing results.

The physical properties for these mutations like Molecular weight, pI, Instability index, aliphatic index, and Grand average hydropathicity was predicted using PROT-PARAM and Scratch protein

predictor tools. Stability and secondary structure prediction of wild type domains were compared using MUPRO and SOPMA. To predict the pathological effect of these mutations, tools like PMUT, SIFT, MutPred and HOPE tools were used.

PMUT: PMUT allows the fast and accurate prediction (approximately 80% success rate in humans) of the pathological character of single point amino acidic mutations based on the use of neural networks. The program also allows the fast scanning of mutational hot spots, which can be obtained by three procedures: (1) alanine scanning, (2) massive mutation and (3) genetically accessible mutations. A graphical interface for Protein Data Bank (PDB) structures, when available, and a database containing hot spot profiles for all non-redundant PDB structures are also accessible from the PMUT server http://mutdb.org/pmut.[11]

Sorting Intolerance From Tolerance (SIFT): SIFT is a sequence homology-based tool that **s**orts **i**ntolerant **f**rom **t**olerant amino acid substitutions and predicts whether an amino acid substitution in a protein will have a phenotypic effect. SIFT is based on the premise that protein evolution is correlated with protein function. Positions important for function should be conserved in an alignment of the protein family, whereas unimportant positions should appear diverse in an alignment. SIFT is available at http://sift.jcvi.org.[12]

POLYPHEN-2: PolyPhen-2 is an automatic tool for prediction of possible impact of an amino acid substitution on the structure and function of a human protein. This prediction is based on a number of features comprising the sequence, phylogenetic and structural information characterizing the substitution. For a given amino acid substitution in a protein, PolyPhen-2 extracts various sequences and structure-based features of the substitution site and feeds them to a probabilistic classifier. Polyphen 2 is available at http://genetics.bwh.harvard.edu/pph2/.[13–15]

MUTPRED: MutPred is a web application tool developed to classify an amino acid substitution (AAS) in human as disease-associated or neutral. MutPred is based upon protein sequence, and which models changes of structural features and functional sites between wild-type and mutant sequences. These changes, expressed as probabilities of gain or loss of structure and function, can provide insight into the specific molecular mechanism responsible for the disease state. MutPred also builds on the established SIFT method but offers improved classification accuracy with respect to human disease mutations. The output of MutPred contains a general score (g), i.e., the probability that the amino acid substitution is deleterious/disease-associated, and top 5 property scores (p), where p is the P-value that certain structural and functional properties are impacted. Certain combinations of high values of general scores and low values of property scores are referred to as hypotheses.

1. Scores with $g > 0.5$ and $p < 0.05$ are referred to as **actionable hypotheses**.
2. Scores with $g > 0.75$ and $p < 0.05$ are referred to as **confident hypotheses**.
3. Scores with $g > 0.75$ and $p < 0.01$ are referred to as **very confident hypotheses**.

Available at http://mutdb.org/mutpred.[16]

HOPE: HOPE collects structural information from a series of sources, including calculations on the 3D protein structure, sequence annotations in Uniprot and predictions from DAS-servers. HOPE combines this information to analyze the effect of a certain mutation on the protein structure. Project HOPE works as an online web server where the user can submit a sequence and mutation. HOPE is available at http://www.cmbi.ru.nl/hope.[17]

3 Results

Of the four VUS in our series, R504H was found in 1 case, L771 L in 47 cases, S1613G in 5 and M1652I in 4 cases. Since the PDB structure was available for zinc finger and BRCT domain, the variants M1652I was included for insilico analysis. In addition to this, 26 missense variants of uncertain/unknown significance from zinc finger domain and 99 from BRCT domain were extracted from BIC portal. The PDB structures ID for the zinc finger domain and BRCT domain were found to be 1JM7 and 1JNX, respectively (Figures: 1 & 2).

The physicochemical properties for the wild type domain region and the mutated region were compared by using the PROT-PARAM and the Scratch protein predictor tool. The molecular weight of the zinc finger domain region was found to be 12802 kda with the theoretical pI 8.4. The Aliphatic index and grand average of hydropathicity is 97.5 and –0.128. The domain region was found to be unstable with the instability index 48.4. The energy minimization for the wild type region was –6102 KJ/mol. Other properties like probabilities of alpha and beta transmembrane proteins, antigenicity property and solubility are

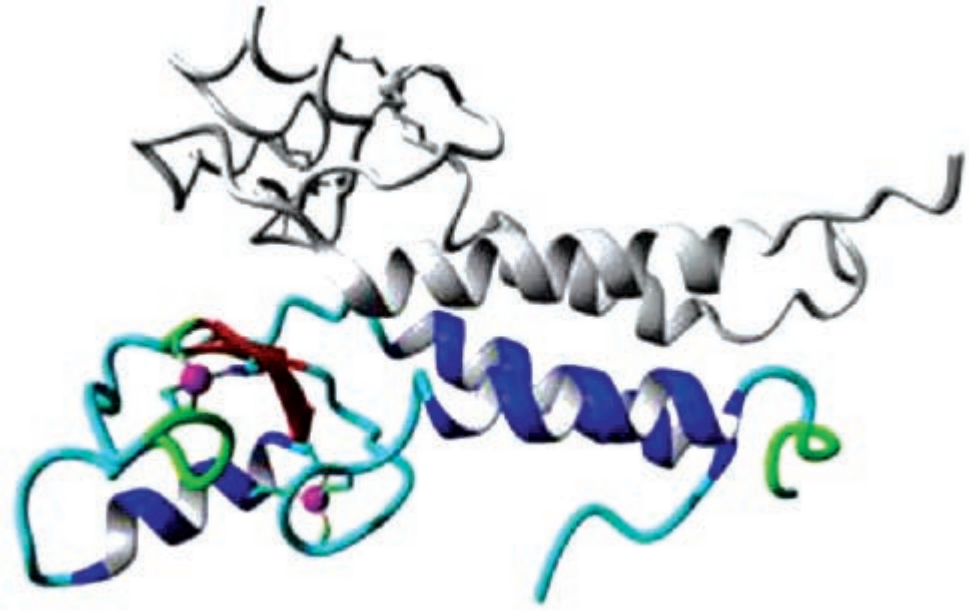

Figure 1: Structure of BRCA1 Zinc finger domain (PDB 1JM7).

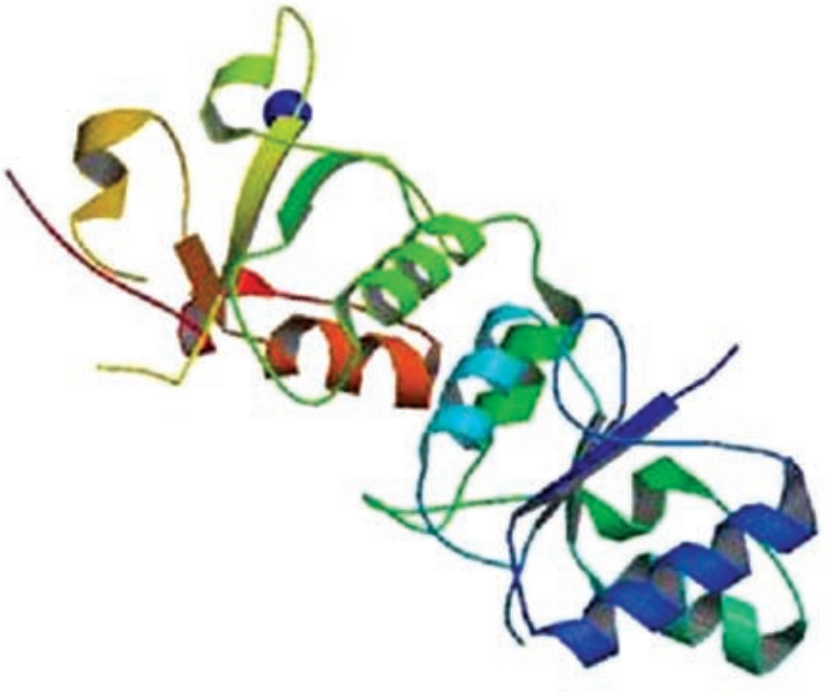

Figure 2: Structure of BRCA1 BRCT domain (PDB 1JNX).

predicted to be within the range from Scratch protein predictor tool.

For the BRCT domain region the molecular weight, pI, aliphatic index, grand average of hydropathicity was found to be 24502 kda with the theoretical pI value of 5.9, aliphatic index was 84.1 and grand average of hydropathicity was –0,164. The domain region was found to be stable with the instability index of 27.7 and energy minimization was found to be –9905 KJ/mol. Scratch predictor tool properties were predicted to be within the range.

For both the zinc finger domain and the BRCT domains the physicochemical properties for the missense mutations were found to show similar

properties, except the energy minimization which was found to deviate from –6000 to –11000 KJ/mol. This shows that at the structural level there is some deviation due to the point mutations. By the Support Vector machine prediction from MUPRO tool the stability for both the domain regions were found to show decreased stability compared to increase stability. Of the 26 VUS in zinc finger domains 12% of the mutations were found to show increased stability and the remaining 88% showed decreased stability. In BRCT domain, of the 100 VUS 20% of them showed increased stability and 80% showed decreased stability.

3.1 *Pathogenic effect of VUS*

The pathogenic effect of the VUS were predicted using the PMUT, SIFT and Polyphen 2 tools. The results were compared and are listed in Table 1 and 2. We found that the prediction results from the three tools used showed 7/26 VUS in zinc finger domain and 20/100 in BRCT domain regions to be potentially pathogenic and affect the structure and function of the protein. The results also showed 3/26 and 13/100 of the variants to have a neutral effect in both the zinc finger and BRCT domain regions, respectively. However, there was much difference in the results from these tools for the other point mutations.

Based on the general score and probability score, the MutPred results are interpreted as actional, confident and very confident hypotheses. In the zinc finger domain, we found 11% actionable hypotheses, 42% confident and 3% very confident hypotheses, resulting in the loss or gain of phosphorylation, glycosylation, ubiquitination, methylation, or change in the catalytic site, or change in the stability and structure of the protein. In the BRCT domain, the predicted MutPred results show 57% actionable hypotheses, 24% confident and 3% very confident hypotheses. Some mutations do not show any hypothesis since the scores are out of range. The prediction results of the MutPred server show that due to the point mutation there is a change in function and structure of protein.

Further analyses were done using the project HOPE tool for the mutations which showed the pathogenic effect for the PMUT, Polyphen and SIFT tools. The HOPE tool gives information about the structural properties of the substituted amino acids based on aspect of contact, structural domain, conservation and amino acid properties. The pathogenic variants included for both the domain regions and the property of the variant is given in Table 3a & 3b.

In the zinc finger domain, wild type amino acids form interactions with the ZN metal ion and are neutral in their charge. However, in the seven probable pathogenic mutations, there is a likelihood of stearic hindrance due to the larger size of the mutant amino acid or change in their charge to positivity, there could be repulsion of the metal ion, resulting in destabilization of the domain. Difference in properties of amino acid disturbs zinc finger domain's interaction with the DNA. Differences in structure might disturb the core structure of domain, thereby affecting the binding properties. Due to the differences in hydrophobicity, loss of hydrophobic interactions in the core of protein occur in four variants.

In case of BRCT domain the wild type residue forms only hydrogen bonds with the other nearer atoms since it does not have any metal ions. But all the 20 variants differ in size and affect the hydrogen bond formation with other atoms. Loss of hydrophobic interaction occurs in nine variants. Due to the loss of hydrogen bonds the protein chain fold disturbs in S1655F, T1685I and T1691I variants. In the variants G1656D and G1788D the protein conformation changes due to unusual torsion angles and results in local structure disturbance.

In L1780P, the wild-type residue is located in an α-helix. Proline disrupts an α-helix when not located at one of the first 3 positions. In case of the mutation at hand, the helix will be disturbed and this can have severe effects on the structure of the protein. In the variant Y1853C, wild-type residue is predicted to be a phosphorylation site. Only serine, threonine and tyrosine residues can be phosphorylated, mutation into cysteine will disturb this modification. Only the variant G1706E was predicted to show no effect upon a mutation. By using the HOPE server we predict that these pathogenic variants may disturb the domains and abolish the function.

4 Discussion

The tools we have used for the prediction were found to be reliable[11,12,14,16,17] since in most of the tools the program retrieves a series of parameters describing the mutation from its internal databases such as Multiple sequence alignment, Protein data bank, BLAST, DSSP, Uniprot, Homology modelling and 3D structure analysis database and so on.

The methods used are based on different aspects and parameters describing the pathogenicity and clue on molecular level about the effect of mutations. By using approach by single method it is not easy for us to predict the pathogenic effect of VUS. Therefore, we need to use multiple methods to compare and rely on the results predicted.

Sequence based methods can be employed to predict physical properties of the protein and the mutations, but in our study there is not much

Table 1: Results of zinc finger domain by PMUT, SIFT, MutPred and Polyphen 2 tools.

S.N.	AA Change	Designation	PMUT	SIFT	POLYPHEN2	MUTPRED Prediction hypothesis
1	Cys to Arg	C24R	pathological	tolerated	probably damage	Confident
2	Cys to Tyr	C24Y	pathological	affect protein function	probably damage	Out of range
3	Leu to Pro	L28P	neutral	affect protein function	probably damage	Out of range
4	Ile to Met	I31M	neutral	tolerated	benign	Actional
5	Glu to Gln	E33Q	neutral	affect protein function	benign	Out of range
6	Thr to Arg	T37R	neutral	tolerated	probably damage	Out of range
7	Thr to Lys	T37K	neutral	tolerated	probably damage	Confident
8	Lys to Asn	K38 N	neutral	tolerated	benign	Out of range
9	Cys to Ser	C39S	neutral	tolerated	probably damage	Confident
10	Cys to Arg	C39R	pathological	tolerated	probably damage	Confident
11	Cys to Tyr	C39Y	pathological	affect protein function	probably damage	Confident
12	His to Arg	H41R	pathological	affect protein function	probably damage	Confident
13	Ile to Val	I42V	neutral	tolerated	benign	Out of range
14	Cys to Ser	C44S	pathological	tolerated	probably damage	Out of range
15	Cys to Tyr	C44Y	pathological	tolerated	probably damage	Confident
16	Cys to Phe	C44F	pathological	affect protein function	probably damage	Out of range
17	Lys to Thr	K45T	neutral	tolerated	possible damage	Actional
18	Lys to Asn	K45 N	neutral	tolerated	possible damage	Actional
19	Cys to Gly	C47G	pathological	tolerated	probably damage	Confident
20	Cys to Phe	C47F	pathological	affect protein function	probably damage	Out of range
21	Leu to Phe	L52F	neutral	tolerated	probably damage	Out of range
22	Cys to Arg	C61R	neutral	tolerated	probably damage	Confident
23	Cys to Tyr	C61Y	pathological	affect protein function	probably damage	Out of range
24	Leu to Phe	L63F	neutral	affect protein function	probably damage	Out of range
25	Cys to Gly	C64G	pathological	tolerated	probably damage	Confident & very confident
26	Cys to Arg	C64R	pathological	affect protein function	probably damage	Confident

difference between wild type and mutated protein. The functional effect of the mutation was predicted by structural information tool. The Polyphen 2 tool gives results based on Position specific independent counts (PSIC) score and the output as Possibly damaging, Probably damaging and Benign effect on structures/ function.[14] Most of the pathogenic mutations are found in BRCT domain compared to zinc finger region. Similar to a study by *Mohammedi L et al* 2009, we also found that the variants S1655F and R1699Q were found to show deleterious effect by using Polyphen 2 tool.[18]

Table 2: Results of BRCT domain by PMUT, SIFT, MutPred and Polyphen 2 tools.

S.N.	AA Change	Designation	PMUT	SIFT	POLYPHEN2	MUTPRED Prediction hypothesis
1	Met to Thr	M1652T	pathological	tolerated	benign	Actional
2	Met to Ile	M1652I	neutral	tolerated	benign	Out of range
3	Val to Met	V1653M	neutral	affect protein function	probably damage	Actional
4	Ser to Phe	S1655F	pathological	affect protein function	probably damage	Confident
5	Gly to Asp	G1656D	pathological	affect protein function	probably damage	Actional
6	Phe to Ser	F1662S	neutral	tolerated	benign	Out of range
7	Met to Leu	M1663 L	neutral	tolerated	benign	Out of range
8	Met to Lys	M1663K	neutral	tolerated	benign	Actional
9	Leu to Pro	L1664P	pathological	tolerated	benign	Actional
10	Val to Met	V1665M	neutral	affect protein function	benign	Out of range
11	Ala to Ser	A1669S	neutral	tolerated	possibly damage	Out of range
12	Glu to Lys	E1682K	pathological	tolerated	benign	Actional
13	Glu to Val	E1682V	neutral	affect protein function	probably damage	Out of range
14	Thr to Ala	T1685 A	neutral	tolerated	possibly damage	Out of range
15	Thr to Ile	T1685I	pathological	affect protein function	probably damage	Actional
16	Met to Thr	M1689T	pathological	tolerated	probably damage	Actional
17	Met to Arg	M1689R	pathological	affect protein function	probably damage	Confident
18	Thr to Lys	T1691K	neutral	tolerated	possibly damage	Actional
19	Thr to Ile	T1691I	pathological	affect protein function	possibly damage	Confident
20	Phe to Leu	F1695 L	neutral	tolerated	possibly damage	Out of range
21	Val to Leu	V1696 L	neutral	affect protein function	possibly damage	Out of range
22	Cys to Arg	C1697R	pathological	affect protein function	possibly damage	Confident
23	Arg to Pro	R1699P	neutral	tolerated	probably damage	Confident
24	Arg to Leu	R1699 L	neutral	affect protein function	possibly damage	Confident
25	Arg to Gln	R1699Q	neutral	tolerated	probably damage	Confident
26	Gly to Ala	G1706 A	neutral	tolerated	possibly damage	Out of range

(*Continued*)

Table 2: Continued.

S.N.	AA Change	Designation	PMUT	SIFT	POLYPHEN2	MUTPRED Prediction hypothesis
27	Gly to Glu	G1706E	pathological	affect protein function	possibly damage	Confident
28	Val to Ala	V1713 A	neutral	affect protein function	possibly damage	Out of range
29	Val to Gly	V1714G	pathological	affect protein function	benign	Out of range
30	Ser to Arg	S1715R	neutral	tolerated	benign	Actional
31	Ser to Cys	S1715C	neutral	tolerated	probably damage	Out of range
32	Ser to Asn	S1715 N	neutral	affect protein function	possibly damage	Out of range
33	Trp to Ser	W1718S	pathological	tolerated	benign	Confident & very confident
34	Trp to Cys	W1718C	pathological	affect protein function	probably damage	Confident
35	Thr to Ala	T1720 A	neutral	tolerated	benign	Out of range
36	Ser to Phe	S1722F	neutral	affect protein function	probably damage	Actional
37	Arg to Gly	R1726G	pathological	tolerated	benign	Out of range
38	Asn to Ser	N1730S	pathological	tolerated	benign	Out of range
39	Asp to Gly	D1733G	neutral	tolerated	possibly damage	Out of range
40	Phe to Ser	F1734S	pathological	affect protein function	benign	Actional
41	Val to Ala	V1736A	neutral	tolerated	possibly damage	Actional
42	Val to Gly	V1736G	pathological	affect protein function	possibly damage	Actional
43	Gly to Arg	G1738R	pathological	tolerated	possibly damage	Actional
44	Gly to Glu	G1738E	pathological	affect protein function	benign	Actional
45	Asp to Tyr	D1739Y	pathological	tolerated	probably damage	Actional
46	Asp to Gly	D1739G	pathological	tolerated	benign	Actional
47	Asp to Glu	D1739E	neutral	affect protein function	benign	Out of range
48	Val to Gly	V1741G	neutral	affect protein function	benign	Actional
49	His to Asn	H1746N	neutral	affect protein function	possibly damage	Out of range
50	Pro to Arg	P1749R	neutral	affect protein function	benign	Out of range
51	Arg to Pro	R1751P	pathological	tolerated	benign	Actional
52	Arg to Gln	R1751Q	neutral	tolerated	probably damage	Actional

(*Continued*)

Table 2: Continued.

S.N.	AA Change	Designation	PMUT	SIFT	POLYPHEN2	MUTPRED Prediction hypothesis
53	Ala to Pro	A1752P	neutral	tolerated	possibly damage	Out of range
54	Ala to Val	A1752V	neutral	affect protein function	possibly damage	Out of range
55	Phe to Ser	F1761S	pathological	affect protein function	probably damage	Actional
56	Gly to Val	G1763V	pathological	affect protein function	benign	Actional
57	Leu to Pro	L1764P	pathological	affect protein function	probably damage	Actional
58	Ile to Ser	I1766S	pathological	affect protein function	possibly damage	Actional
59	Pro to Leu	P1771L	pathological	tolerated	benign	Actional
60	Pro to Arg	P1771R	pathological	tolerated	probably damage	Actional
61	Thr to Ser	T1773S	neutral	tolerated	benign	Actional
62	Thr to Ile	T1773I	pathological	affect protein function	possibly damage	Out of range
63	Met to Arg	M1775R	pathological	affect protein function	probably damage	Out of range
64	Asp to Asn	D1778N	neutral	tolerated	benign	Out of range
65	Asp to Gly	D1778G	pathological	tolerated	possibly damage	Out of range
66	Leu to Pro	L1780P	pathological	affect protein function	probably damage	Actional
67	Met to Leu	M1783L	neutral	tolerated	possibly damage	Out of range
68	Met to Thr	M1783T	pathological	affect protein function	probably damage	Actional
69	Cys to Ser	C1787S	neutral	affect protein function	possibly damage	Very confident
70	Gly to Asp	G1788D	pathological	affect protein function	probably damage	Confident & very confident
71	Ala to Ser	A1789S	neutral	tolerated	possibly damage	Confident
72	Gly to Ala	G1803A	neutral	tolerated	benign	Out of range
73	Val to Asp	V1804D	pathological	tolerated	benign	Out of range
74	Pro to Ala	P1806A	neutral	tolerated	benign	Out of range
75	Val to Ala	V1808A	neutral	affect protein function	benign	Out of range
76	Val to Phe	V1809F	neutral	tolerated	possibly damage	Out of range
77	Val to Ala	V1809A	neutral	affect protein function	benign	Out of range
78	Val to Gly	V1810G	pathological	affect protein function	probably damage	Actional

(*Continued*)

Table 2: Continued.

S.N.	AA Change	Designation	PMUT	SIFT	POLYPHEN2	MUTPRED Prediction hypothesis
79	Gln to Arg	Q1811R	neutral	affect protein function	probably damage	Out of range
80	Pro to Ala	P1812A	neutral	tolerated	benign	Out of range
81	Asp to Gly	D1818G	pathological	tolerated	benign	Out of range
82	Asn to Ser	N1819S	neutral	tolerated	benign	Out of range
83	Gln to His	Q1826H	neutral	tolerated	benign	Out of range
84	Ala to Thr	A1830T	neutral	affect protein function	probably damage	Actional
85	Val to Met	V1833M	neutral	affect protein function	probably damage	Confident
86	Arg to Gln	R1835Q	neutral	tolerated	probably damage	Actional
87	Glu to Lys	E1836K	neutral	tolerated	probably damage	Actional
88	Trp to Gly	W1837G	pathological	tolerated	probably damage	Out of range
89	Trp to Arg	W1837R	pathological	tolerated	probably damage	Actional
90	Trp to Cys	W1837C	pathological	affect protein function	probably damage	Out of range
91	Val to Glu	V1838E	pathological	affect protein function	probably damage	Actional
92	Ser to Arg	S1841R	neutral	tolerated	probably damage	Confident
93	Ser to Asn	S1841N	neutral	affect protein function	probably damage	Out of range
94	Ala to Pro	A1843P	neutral	affect protein function	probably damage	Actional
95	Leu to Arg	L1844R	neutral	tolerated	probably damage	Out of range
96	Asp to Glu	D1851E	neutral	tolerated	probably damage	Out of range
97	Tyr to Cys	Y1853C	pathological	affect protein function	probably damage	Out of range
98	Leu to Pro	L1854P	neutral	affect protein function	probably damage	Actional
99	Pro to Ser	P1856S	pathological	tolerated	benign	Out of range
100	Pro to Arg	P1859R	pathological	affect protein function	possibly damage	Out of range

The variant M1652I of BRCA1 found in four cases (2 in breast and 2 in ovarian cancer) was predicted to be benign with all the used tools. Much evidence is there to show the neutral effect of M1652I variant by both functional and computational methods.[19,20] Therefore the results obtained from our tools are reliable.

Recent study by *Cherbal F et al, 2012* showed the effect of VUS in BRCA1 and BRCA2 gene in the Algerian breast/ovarian cancer families and found that the missense polymorphism have role as the susceptibility breast cancer markers in Algerian cancer families where pathological BRCA1 and BRCA2 mutations were not present.[21] *Hussain et al 2012* studied the role of SNPs in BRAF gene and stated, using HOPE server and others, that we can identify significant structural changes caused by substitution of amino acid for analysing the interaction and conformational change.[22] Our analyses by HOPE also gives significant structural changes upon substitution of amino acids.

Apart from the experimental ways of predicting the effect of VUS, it is also important to analyze the effect by multiple bioinformatics approach. More importantly the effect of mutation by structural analysis could play a major role in predicting the protein function in molecular mechanisms.

Table 3a: HOPE results for Zinc finger domain variants.

S.N.	Variant	Mutated amino acid property
1	C24Y	Loss of hydrophobic interaction with other molecules on the surface of the protein
2	C39Y	Loss of hydrophobic interaction with other molecules on the surface of the protein
3	H41R	Mutation can disturb the interaction with other molecules
4	C44F	The wild-type residue was buried in the core of the protein. The mutant residue is bigger and probably will not fit.
5	C47F	The wild-type residue was buried in the core of the protein. The mutant residue is bigger and probably will not fit.
6	C61Y	Loss of hydrophobic interaction with other molecules on the surface of the protein
7	C64R	Loss of hydrophobic interaction with other molecules on the surface of the protein

Table 3b: HOPE results for BRCT domain variants.

S.N.	Variant	Mutated amino acid property
1	S1655F	The mutation will cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding
2	G1656D	The residue is located on the surface of the protein, mutation of this residue can disturb interactions. Only glycine is flexible enough to make these torsion angles, mutation into another residue will force the local backbone into an incorrect conformation and will disturb the local structure with other molecules or other parts of the protein.
3	T1685I	The mutation will cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding
4	M1689R	The mutation will cause loss of hydrophobic interactions in the core of the protein
5	T1691I	The mutation will cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding.
6	C1697R	The mutation will cause loss of hydrophobic interactions in the core of the protein
7	G1706E	Null
8	W1718C	The mutant residue is smaller than the wild-type residue. The mutation will cause an empty space in the core of the protein
9	V1736G	The mutation will cause loss of hydrophobic interactions in the core of the protein
10	F1761S	The mutation will cause loss of hydrophobic interactions in the core of the protein
11	T1773I	Mutation of this residue can disturb interactions with other molecules or other parts of the protein
12	M1775R	The mutation will cause loss of hydrophobic interactions in the core of the protein
13	L1780P	The mutant residue is smaller than the wild-type residue. The mutation will cause an empty space in the core of the protein
14	M1783T	The mutation will cause loss of hydrophobic interactions in the core of the protein
15	G1788D	The residue is located on the surface of the protein, mutation of this residue can disturb interactions with other molecules or other parts of the protein. The torsion angles for this residue are unusual. Only glycine is flexible enough to make these torsion angles, mutation into another residue will force the local backbone into an incorrect conformation and will disturb the local structure.
16	V1810G	The mutation will cause loss of hydrophobic interactions in the core of the protein
17	W1837C	The mutant residue is smaller than the wild-type residue. The mutation will cause an empty space in the core of the protein
18	V1838E	The mutation will cause loss of hydrophobic interactions in the core of the protein
19	Y1853C	Mutant residue is smaller than the wild-type residue. This will cause a possible loss of external interactions.
20	P1859R	The mutation might cause loss of hydrophobic interactions with other molecules on the surface of the protein.

5 Conclusions

Much work has been carried out to classify the BRCA1 Missense variant of unknown significance by both functional and computational methods. Though the computational methods are less time consuming for predicting the significance of the variant as deleterious or neutral, it needs to be validated by functional approach in order to use as the therapeutic target. Computational methods are also equally important in analysing the

data in these functional assays. To understand the role of unknown significance in cancer risk and to improve diagnostic tests, integrated studies are the best way.

Received 25 August 2012.

References

1. Boulton SJ (2006). Cellular functions of the BRCA tumour-suppressor proteins. Biochemical Society Transactions 34(5): 633–645.
2. Wu W, Koike A, Takeshita T, Ohta T (2008). The ubiquitin E3 ligase activity of BRCA1 and its biological functions. Cell Division 3:1 doi:10.1186/1747-1028-3-1.
3. Bork P, Hofmann K, Bucher P, Neuwald AF, Altschul SF and Koonin EV (1997). A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. FASEB. J. 11: 68–76.
4. Soumittra N, Meenakumari B, Parija T, Sridevi V, Nancy KN, Swaminathan R, Rajalekshmy KR, Majhi U, Rajkumar T (2009). Molecular genetics analysis of hereditary breast and ovarian cancer patients in India. Hereditary Cancer in Clinical Practice 7:13. doi:10.1186/1897-4287-7-13.
5. Kudla G, Murray AW, Tollervey D, and Plotkin JB (2009): Coding-sequence determinants of gene expression in Escherichia coli. Science 324: 255–8.
6. Tavtigan SV, Thomas A, Thomas FS & Sckolnick MK (1998). The BRCA1 gene and its protein products characterization, therapeutics and diagnostics implications. Gynecologic Oncology Associates.
7. Deffenbaugh A (2006). Myriad Genetics, personal communication.
8. Coquelle N, Green R, Glover JN (2011). Impact of BRCA1 BRCT Domain Missense Substitutions on Phosphopeptide Recognition. Biochemistry. 50(21): 4579–4589.
9. Yu X, Chini C.C, He M, Mer G, Chen J (2003). The BRCT domain is a phospho-protein binding domain. Science. 302: 639–642.
10. Rajkumar T, Soumittra N, Nancy NK, Swaminathan R, Sridevi V, Shanta V (2003). BRCA1, BRCA2 and CHEK2 (1100delC) germline mutation in Hereditary Breast and Ovarian cancer families in South India. APJCP 4: 203–208.
11. Ferrer-Costa C, Gelpí JL, Zamakola L, Parraga I, de la Cruz X, Orozco M (2005). PMUT: A web-based tool for the annotation of pathological mutations on proteins. Bioinformatics. 21(14): 3176–8.
12. Ng PC and Henikoff S (2001). Predicting deleterious amino acid substitutions. Genome Res. 11(5): 863–74.
13. Kumar P, Henikoff S, Ng PC (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 4(7): 1073–81.
14. Sunyaev SR, Eisenhaber F, Rodchenkov IV, Eisenhaber B, Tumanyan VG, Kuznetsov EN (1999). PSIC: Profile extraction from sequence alignments with position-specific counts of independent observations. Protein Engg 12(5): 387–394.
15. Ramensky V, Bork P, Sunyaev S (2002). Human non-synonymous SNPs: Server and survey. Nucleic Acids Res 30(17): 3894–3900.
16. Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P (2009). Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 25 (21): 2744–2750.
17. Venselaar H, Tim AH te Beek, Remko KP Kuipers, Maarten L Hekkelman, Gert Vriend (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics 11: 548.
18. Mohammadi L and Vreeswijk MP (2009). A simple method for co-segregation analysis to evaluate the pathogenicity of unclassified variants; BRCA1 and BRCA2 as an example. BMC Cancer 9: 211 doi:10.1186/1471-2407-9-211.
19. Rowling PJ, Cook R, Itzhaki LS (2010). Toward Classification of BRCA1 Missense Variants Using a Biophysical Approach. J Biol Chem. 285(26): 20080–20087.
20. Deffenbaugh AM, Frank TS, Hoffman M, Cannon-Albright L, Neuhausen SL (2002). Characterization of common BRCA1 and BRCA2 variants. Genet. Test 6: 119–121.
21. Cherbal F, Salhi N, Bakour R, Adane S, Boualga K, Maillet P (2012). BRCA1 and BRCA2 unclassified variants and missense polymorphisms in Algerian breast/ovarian cancer families. Dis Markers. 32: 343–53.
22. Hussain MR, Shaik NA, Al-Aama JY, Khan F, Masoodi TA (2012). In silico analysis of single nucleotide polymorphisms (SNPs) in human BRAF gene. Gene. doi:10.1016/j.gene.2012.07.014.

Balaiah Meenakumari MSc., is the Scientific Assistant and a Doctoral student in the Dept. of Molecular Oncology. She is involved with the Muatation analysis for Hereditary cancers.

Thangarajan Rajkumar MD., DM., PhD., DSc., is the Professor and Head, of the Dept. of Molecular Oncology, at the Cancer Institute (WIA), Adyar, Chennai.