生物工程学报
journals.im.ac.cn
cjb@im.ac.cn

Chin J Biotech 2008, May 25; 24(5): 851-856
Chinese Journal of Biotechnology ISSN 1000-3061

研究报告

Computational and Structural Investigation of Deleterious Functional SNPs in Breast Cancer *BRCA2* Gene

Rajasekaran R, George Priya Doss, Sudandiradoss C, Ramanathan K, Rituraj Purohit, and Rao Sethumadhavan

School of Biotechnology, *Chemical and Biomedical Engineering*, *Bioinformatics Division*, *Vellore Institute of Technology University*, *Vellore* 632014, *Tamil Nadu*, *India.*

Abstract: In this work, we have analyzed the genetic variation that can alter the expression and the function in *BRCA2* gene using computational methods. Out of the total 534 SNPs, 101 were found to be non synonymous (nsSNPs). Among the 7 SNPs in the untranslated region, 3 SNPs were found in 5′ and 4 SNPs were found in 3′ un-translated regions (UTR). Of the nsSNPs 20.7% were found to be damaging by both SIFT and PolyPhen server among the 101 nsSNPs investigated. UTR resource tool suggested that 2 SNPs in the 5′ UTR region and 4 SNPs in the 3′ UTR regions might change the protein expression levels. The mutation from asparagine to isoleucine at the position 3124 of the native protein of *BRCA2* gene was most deleterious by both SIFT and PolyPhen servers. A structural analysis of this mutated protein and the native protein was made which had an RMSD value of 0.301 nm. Based on this work, we proposed that this most deleterious nsSNP with an SNPid rs28897759 is an important candidate for the cause of breast cancer by *BRCA2* gene.

Key words: *BRCA2* gene, breast cancer, SIFT, PolyPhen, UTRscan, modeled structure

Single nucleotide polymorphisms (SNPs), which are found in every 200–300 bp, represent the most abundant class of genetic variations in the human genome[1]. It is known that 500 000 SNP's fall in the coding regions of the human genome[2] and among these, the non-synonymous SNPs(nsSNPs) causes changes in the amino acid residues. These are likely to be an important factor contributing to the functional diversity of the encoded proteins in the human population[3]. In addition to this, most SNPs in human genome are found in non-coding DNA, including 5′ and 3′ untranslated regions (UTR). Polymorphism in 5′-UTR of the gene may affect transcriptional activity and can therefore be of functional relevance[4]. Polymorphism in the 3′-UTR region may affect gene expression by affecting RNA half-life or influencing ribosomal translation of mRNA[5].

BRCA2 (13q12-q13) is a well-known inherited tumor suppressor gene[6,7]. Mutation of *BRCA2* gene predisposes individuals to breast, ovarian and other cancers[8,9] and almost half the cases of inherited early onset breast cancers have been linked to mutations in *BRCA2* gene. The product of this gene, *BRCA2* protein, interacts with RAD51, essential for homologous recombination, DNA repair, and the maintenance of genome stability. There appear to be a number of different interactions between *BRCA2* and RAD51. One such interaction involves a C-terminal region of *BRCA2*[10]. In addition, *BRCA2* interacts with RAD51 through the eight conserved BRC repeats in *BRCA2* protein[11,12] and mutations within these repeats are associated with cancer predisposition.

Discovering the deleterious SNPs is the main task of pharmacogenomics and pharmacogenetics. Though our literatures survey shows that, there is a wide choice of literature on BRCA2 gene related to breast cancer, yet there are no computational studies undertaken for an in-silico investigation of nsSNPs and UTR SNPs in *BRCA2* gene. We undertook this work mainly to perform a computational analysis of the nsSNPs in *BRCA2* gene and to identify the possible deleterious mutations and proposed a modeled structure for the mutant protein.

In continuation to our previous work on functional SNP analysis of *BRCA1* gene[13], we here report our results on *BRCA2* gene. Out of 101 nsSNPs, the most

Received: November 15, 2006; **Accepted:** December 18, 2006
Corresponding author: R. Sethumadhavan. Tel: +914162202522; Fax: +914162243092; E-mail: rsethumadhavan@vit.ac.in

deleterious functionally significant nsSNPs causing the mutation from aspargine to isoleucine at the position of 3124 in *BRCA2* protein and 6 SNPs, having functional significance in 5′ and 3′ untranslated regions could be a candidate of major concern in the disease of breast cancer caused by *BRCA2* gene.

1 Materials and methods

1.1 Datasets

The SNPs and their related protein sequence for *BRCA2* gene were obtained from the dbSNP[14] http://www.ncbi.nlm.nih.gov/SNP/ for our computational analysis.

1.2 Analysis of functional consequences of coding nsSNPs by sequence homology based method (SIFT)

We used the program SIFT[15] available at http://blocks.fhcrc.org/sift/SIFT.html to detect the deleterious coding non synonymous SNPs. SIFT is a sequence homology based tool which presumes that important amino acids will be conserved in the protein family. Hence, changes at well-conserved positions tend to be predicted as deleterious[16]. We submitted the query in the form of SNPids or as protein sequences. The underlying principle of this program is that, SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence. SIFT is a multistep procedure that, given a protein sequence, (a) searches for similar sequences, (b) chooses closely related sequences that may share similar function, (c) obtains the multiple alignment of these chosen sequences, and (d) calculates normalized probabilities for all possible substitutions at each position from the alignment. Substitutions at each position with normalized probabilities less than a chosen cutoff are predicted to be deleterious and greater than or equal to the cutoff are predicted to be tolerated[15]. The cutoff value in SIFT program is tolerance index of $\geqslant$ 0.05. Higher the tolerance index, less functional impact a particular amino acid substitution is likely to have.

1.3 Simulation for functional change in coding nsSNPs by structure homology based method (PolyPhen)

Analyzing the damaged coding non synonymous SNPs at the structural level are considered to be very important to understand the functional activity of concerned protein. We used the server PolyPhen[17] which is available at http://coot.embl.de/PolyPhen/ for this purpose. Input options for PolyPhen server is protein sequence or SWALL database ID or accession number together with sequence position with two amino acid variants. We submitted the query in the form of protein sequence with mutational position and two amino acid variants. Sequence based characterization of the substitution site, profile analysis of homologous sequences and mapping of substitution site to a known protein 3-dimensional structures are the parameters taken into account by PolyPhen server to calculate the score. It calculates position-specific independent counts (PSIC) scores for each of the two variants, and then computes the PSIC scores difference between them. Higher the PSIC score difference, higher is the functional impact a particular amino acid substitution is likely to have.

1.4 Scanning of UTR SNPs in UTR site

5′ and 3′ UTR regions are involved in various biological processes such as post transcriptional regulatory pathways, stability and translational efficiency[18,19]. We used the program UTRscan[20] which allows to search the user submitted sequences for any of the patterns collected in UTR site. UTRsite is a collection of functional sequence patterns located in 5′ or 3′ UTR sequences. Briefly, two or three sequences of each UTR SNP which have different nucleotide at SNP position are analyzed by UTRscan which looks for UTR functional elements by searching through user submitted sequence data for the patterns defined in the UTRsite and UTR database. If different sequences of each UTR SNP are found to have different functional pattern(s), this UTR SNP is predicted to have functional significance. The internet resources for UTR analysis are UTRdb and UTRsite. UTRdb contain experimentally proved biological activity of functional sequence patterns of UTR sequence from eukaryotic mRNAs[21]. The UTRsite have the data collected from UTRdb and also continuously enriched with new functional patterns.

1.5 Modeling nsSNP locations on protein structure and their RMSD difference

Structural information has been extensively used for studying the effects of nsSNPs[22]. We used the database dbSNP[14] for identifying the nsSNPs onto the structure by using limits option. We confirmed the mutation positions and the mutation residues by this database. The mutation was performed by using SWISSPDB viewer and energy minimization for 3D structures was performed by NOMAD-Ref server[23]. This server use Gromacs as default force field for energy minimization based on the methods of steepest descent, conjugate gradient and L-BFGS methods[24]. We used conjugate gradient method for optimizing the 3D structures. Divergence in mutant structure with native structure is due to mutation, deletions, and insertions[25] and the deviation between the two structures is evaluated by

their RMSD values which could affect stability and functional activity[26].

2 Results and Discussion

2.1 SNP dataset from dbSNP

The *BRCA2* gene investigated in this work was retrieved from dbSNP database[14]. It contained a total of 534 SNPs. Out of which, 101 were non-synonymous SNPs (nsSNPs), 47 were synonymous SNPs, 7 were in non coding regions, which comprises of 3 SNPs in 5′ UTR region and 4 SNPs were in 3′ UTR region. The rest were in the intron region. We selected non-synonymous coding SNPs, 5′ and 3′ UTR region SNPs for our investigation.

2.2 Deleterious and damaged nsSNPs by SIFT and PolyPhen

The conservation level of a particular position in a protein was determined by using a sequence homology based tool; SIFT[15]. Protein sequences of 101 nsSNPs were submitted independently to SIFT program to check its tolerance index. Among the 101 nsSNPs, 35 nsSNPs were found to be deleterious having the tolerance index score of ≤ 0.05. The results are shown in consolidated Table 1.

We observed that, out of 35 deleterious nsSNPs, 14 nsSNPs showed a highly deleterious tolerance index score of 0.007 nsSNPs a tolerance index score of 0.01, followed by 4, 5 and 3 nsSNPs with a tolerance index of 0.02, 0.03 and 0.04 respectively. 11 nsSNPs showed a nucleotide change from A→G, 9 nsSNPs from C→T, 5 nsSNPs from G→T, 4 nsSNPs from A→T, 3 nsSNPs from A→C and 3 nsSNPs showed a nucleotide change from C→G. A→G and C→T nucleotide changes

Table 1 List of nsSNPs that were predicted to be functional significance by both SIFT and PolyPhen

SNPs ID	Nucleotide change	AA change	Tolerance index	PSIC SD	SNPs ID	Nucleotide change	AA change	Tolerance index	PSIC SD
rs11571656	C/T	S976F	0.02	1.890	rs4987046	A/G	Y42C	0.00	Not damaged
rs11571707	C/T	I2490T	0.01	1.743	rs28897716	A/G	D935N	0.03	Not damaged
rs11571769	A/G	A2951T	0.03	1.528	rs41293501	A/C	A2233D	0.00	Not damaged
rs169548	A/G	S2536P	0.01	1.959	rs41293503	A/G	E2236K	0.02	Not damaged
rs2227944	A/T	N987I	0.00	1.908	rs28897729	A/G	V1542M	0.03	Not damaged
rs28897702	G/T	G173V	0.01	1.533	rs41293471	C/T	T207I	0.00	Not damaged
rs28897719	A/T	N1102Y	0.04	2.417	rs41293511	C/G	D2723H	0.00	Not damaged
rs28897728	A/G	G1529R	0.00	2.367	rs41293513	A/C	D2723A	0.00	Not damaged
rs28897737	C/G	S1979R	0.03	1.595	rs1799954	C/T	R2034C	Not deleterious	1.553
rs28897738	C/G	S1982T	0.04	1.648	rs2219594	G/T	H1561N	Not deleterious	2.307
rs28897744	C/T	T2515I	0.01	1.951	rs28897701	C/G	A75P	Not deleterious	1.722
rs28897746	C/T	L2686P	0.00	2.243	rs28897705	G/T	N277K	Not deleterious	2.046
rs28897750	A/C	A2770D	0.05	2.060	rs28897706	A/C	S326R	Not deleterious	1.526
rs28897751	C/T	L2792P	0.00	2.467	rs28897708	C/T	I505T	Not deleterious	1.525
rs28897752	C/T	L2898S	0.02	2.051	rs28897712	C/G	P655R	Not deleterious	1.905
rs28897753	G/T	V2908G	0.01	2.138	rs28897722	A/G	N1228D	Not deleterious	1.631
rs28897758	G/T	L3101R	0.00	1.760	rs28897725	A/C	T1388N	Not deleterious	1.571
rs28897759	A/T	N3124I	0.00	2.769	rs28897726	A/G	E1397K	Not deleterious	1.609
rs28897761	A/G	E3342K	0.00	1.732	rs28897727	G/T	D1420Y	Not deleterious	2.036
rs4987047	A/T	I2944F	0.02	1.734	rs28897733	A/C	K1678T	Not deleterious	1.902
rs766173	G/T	N289H	0.01	1.851	rs28897742	G/T	D2238E	Not deleterious	1.556
rs169547	A/G	A2466V	0.00	Not damaged	rs28897743	A/G	R2336H	Not deleterious	1.613
rs2227943	C/T	S929L	0.01	Not damaged	rs28897745	A/G	D2665G	Not deleterious	2.141
rs11571657	G/T	N1880K	0.05	Not damaged	rs28897754	G/T	K2950N	Not deleterious	1.862
rs28897709	A/G	K513R	0.04	Not damaged	rs28897755	C/T	T3013I	Not deleterious	1.583
rs28897748	C/T	I2719T	0.00	Not damaged	rs28897760	A/C	P3194Q	Not deleterious	1.980
rs41293493	A/G	S1682N	0.03	Not damaged	rs4986859	A/G	N2447D	Not deleterious	1.599

AA change: amino acid change, Tolerance Index and PSIC SD mentioned in bold indicates deleterious and damaged by both SIFT and PolyPhen respectively.

occurred maximum number of times and A→C and C→G nucleotide changes occurred minimum number of times as could be seen from Table 1. The nucleotide change from A→G and C→T accounted for the highest number of deleterious nsSNPs with a SIFT tolerance index of 0.00. This was closely followed by the nucleotide change from G→T and C→T, which showed a tolerance index of 0.01. Also, out of the 14 nsSNPs which showed a SIFT tolerance index of 0.003 of them changed from polar amino acid in the native protein to non-polar amino acid in the mutant type, 3, from non-polar to polar, 2, from negatively charged amino acid to positively charged amino acid, 2, from non-polar to positively charged amino acid and 1 each from non-polar to negatively charged amino acid, aromatic to polar amino acid, positively charged to non-polar amino acid and from non-polar amino acid in native protein to non-polar amino acid in mutant type respectively.

The structural levels of alteration were determined by applying PolyPhen program[17]. Protein sequence with mutational position and amino acid variants associated to 101 ns SNPS, investigated in this work was submitted as input to the PolyPhen server and the results are shown in the consolidated Table 1. A PSIC score difference of 1.1 and above is considered to be damaging. It can be seen that, out of 101 nsSNPs, 40 nsSNPs were considered to be damaging. All the 40 nsSNPs exhibited a PSIC score difference between 1.525 to 2.769.

21 nsSNPs which were observed to be deleterious by the SIFT program also were damaging according to PolyPhen. Hence, we could infer that the result obtained on the basis of sequence details (SIFT) were in good correlation with the result obtained by structural details (PolyPhen), as could be seen from Table 1. It can also be seen that, four nsSNPs, (rs28897728, rs28897746, rs28897751, rs28897759) had a SIFT tolerance index of 0.00 and PSIC score difference≥ 2.00. According to SIFT and PolyPhen results, these four nsSNPs could be considered as significant for the identification of breast cancer due to *BRCA2* gene.

2.3 Functional SNPs in UTR by UTRscan server

UTR SNPs have functional effects on transcriptional regulation, by affecting transcription factor binding sites in promoter regions[27]. We used the UTRscan server[20] for identifying the functionally significant SNPs in untranslated region.

Table 2 shows the list of SNPs in the 5′ and 3′ untranslated regions which are predicted to be of functional significance. We observed that, 2 SNPs in 5′ UTR namely, rs1799943, rs11571836 and 4 SNPs namely, rs11571834, rs11571835, rs7334543 and rs15869 in 3′UTR were predicted to be of functional significance by this server.

15-lipoxygenase differentiation control element (15-LOX-DICE) controls 15-LOX synthesis which catalyses the degradation of lipids and is an important factor responsible for the degradation of mitochondria during reticulocyte maturation. This 15-LOX-DICE exist in this two 5′UTR SNPs and four 3′UTR SNPs which were considered to be of functional significance and hence can be thought to be damaging in the *BRCA2* gene. This result could be thought as an important outcome from this work as there are no reports in the literature which relates the deleterious nature of SNPs with 3′ and 5′ untranslated region of *BRCA2* gene so far.

2.4 Modelling of mutant structure

Information about mapping the deleterious nsSNPs into protein structure was obtained from dbSNP[14]. The available structure for *BRCA2* gene has a PDB id 1iyj. Out of 21 nsSNPs which were found to deleterious both in SIFT and PolyPhen, one nsSNP (rs28897759) showed a SIFT tolerance index of 0.00 and PolyPhen PSIC SD of 2.769 as depicted in consolidated Table 1. This represents the highest deleterious nature among the 21 nsSNPs as could be seen from this Table. Hence we selected this nsSNP for structural analysis. The mutational position and amino acid variant associated with this nsSNP is N→I at the residue position 3124 was mapped in the 1iyj native structure. Mutation at specified position was performed by SWISSPDB viewer to get modeled structure. Then, energy minimizations were

Table 2 List of SNPs (UTR mRNA) that were predicted to be functional significance by UTRscan

SNPs ID	Nucleotide Change	UTR Position	Functional element change
rs1799943	A/G	5′ UTR	15-LOX-DICE → no pattern
rs11571836	A/G	5′ UTR	15-LOX-DICE → no pattern
rs11571834	C/G	3′ UTR	15-LOX-DICE → no pattern
rs11571835	A/C	3′ UTR	15-LOX-DICE → no pattern
rs7334543	A/G	3′ UTR	15-LOX-DICE → no pattern
rs15869	A/C	3′ UTR	15-LOX-DICE → no pattern

performed by NOMAD-Ref server[23] for both the native structure (PDB 1iyj) and mutant modeled structure.

Higher is the RMSD value, more is the deviation between the two structures which in turn change their stability and functional activity. The native structure (1iyj) with Arginine at the residue position 3124 is shown in Fig. 1a, mutation at the residue position 3124 from Arg to Ile for nsSNP (rs28897759) is shown in Fig. 1b, mutant modeled structure with Isoleucine at the residue position of 3124 is shown in Fig. 1c and the superimposed structure between the native protein and mutant with RMSD is shown in Fig. 1d.

This structural analysis portrays that mutant modeled structure posses an RMSD of 0.301 nm as compared to native structure. Moreover this mutant modeled structure had a total energy of 5 355 276.0 Kcal/mole which is many times higher than the total energy of –13 100.54 Kcal/mole of the native structure. Hence by this study, we can without any ambiguity claim that the nsSNP (rs28897759) associated with this mutant structure was highly affected in sequence level (SIFT) and structural level (PolyPhen). Further this mutation could be believed to affect the structures of the native protein in *BRCA2* gene. The result reported by our work in this study is well supported by an experimental study carried out earlier on *BRCA2* gene[28].

Based on the overall results from this study, we could ascertain that the mutation N→I at the residue position 3124 in the *BRCA2* native protein is a potential candidate for the cause of breast cancer by *BRCA2* gene. Apart from this, we have also reported 20 other mutation sites shown in consolidated Table 1, all of which shows a deleterious SIFT and PolyPhen score and are also considered to functionally significant nsSNPs.

3 Conclusions

Breast cancer *BRCA2* gene was investigated in this work by evaluating the influence of functional SNPs through computation methods. In a total of 534 SNPs in *BRCA2* gene, 101 SNPs were found to be non-synonymous. 3 and 4 SNPs were found to be in 5′ and 3′un-translated regions. Out of 101 nsSNPs, 35 of them were found to be deleterious from SIFT and 40 of them were damaging as per the PolyPhen server. 21 nsSNPs were found to be common in both SIFT and PolyPhen server. 2 SNPs in the 5′UTR region and 4 SNPs in the 3′UTR region were found to be of functional significance. Out of 21 major mutation

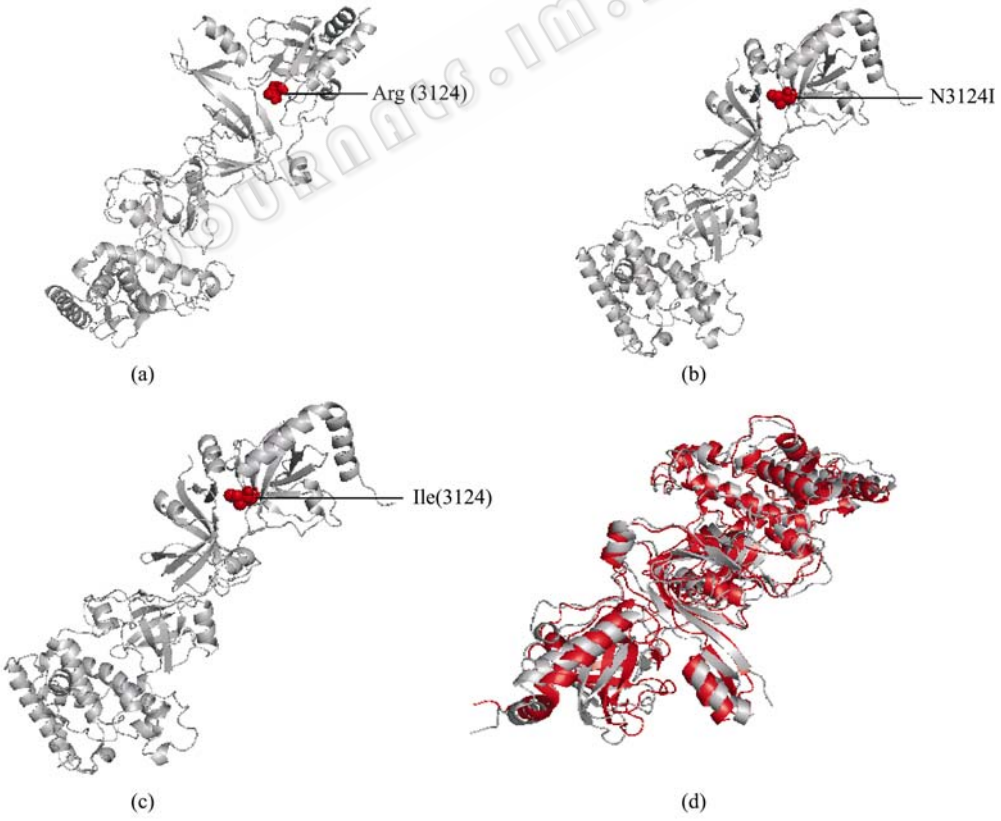

Fig. 1 Native structure of (1iyj) showing Arginine residue at position 3124(a); nsSNP (rs28897759) cause a mutation at residue position 3124 from Arginine to Isoleucine(b); Mutant structure showing Isoleucine residue at position 3124(c); Superimposed structure of native protein 1iyj (grey) with mutant structure 1iyj (red) 3124(N→I) showing RMSD of 0.301 nm(d)

observed in this study, it was found that mutation in the native protein of *BRCA2* gene was from asparagine to isoleucine at the position 3124 showed the most deleterious SIFT and PolyPhen score. The structural analysis reveals an RMSD score of 0.301 nm between the mutant and the native structures of the *BRCA2* protein. Hence we conclude that the nsSNP (rs28897759) associated with this mutation could be an important candidate for the cause of breast cancer by *BRCA2* gene by SIFT, PolyPhen and through structural analysis study.

Acknowledgment

The authors thank the management of Vellore Institute of Technology University for providing the facilities to carry out this work.

REFERENCES

[1] Lee JE, Choi JH, Lee JH, Lee MG. Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis. *Mutat Res*, 2005, **573**: 195–204.

[2] Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variations. *Genomic Res*, 1998, **8**: 1229–1231.

[3] Lander ES. The new genomics, Global views of biology. *Science*, 1996, **274**: 536–539.

[4] Nussbaum R, McInnes R, Willard H. Genetic Variation in Individuals: Mutation and Polymorphism. Thompson & Thompson Genetics in Medicine. 6th ed. Philadelphia: WB Saunders Company, 2001.

[5] Van Deventer SJ. Cytokine and cytokine receptor polymorphisms in infectious disease. *Intensive Care Med*, 2000, **26**: S98–S102.

[6] Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. Identification of the breast cancer susceptibility gene *BRCA2*. *Nature*, 1995, **378**: 789–792.

[7] Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, *et al*. Genetic heterogeneity and penetrance analysis of the *BRCA1* and *BRCA2* genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet*, 1998, **62**: 676–689.

[8] Thorlacius S, Olafsdottir G, Tryggvadottir L, Neuhausen S, Jonasson JG, Tavtigian SV, Tulinius H, Ogmundsdottir HM, Eyfjord JE. A single *BRCA2* mutation in male and female breast cancer families from Iceland with varied cancer phenotypes. *Nat Genet*, 1996, **13**: 117–119.

[9] Gudmundsson J, Johannesdottir G, Bergthorsson JT, Arason A, Ingvarsson S, Egilsson V, Barkardottir RB. Different tumor types from *BRCA2* carriers show wild-type chromosome deletions on 13q12-q13. *Cancer Res*, 1995, **55**: 4830–4832.

[10] Esashi F, Christ N, Gannon J, Liu Y, Hunt T, Jasin M, West SC. CDK- dependent phosphorylation of *BRCA2* as a regulatory mechanism for recombinational repair. *Nature*, 2005, **434**: 598–604.

[11] Chen PL, Chen CF, Chen Y, Xiao J, Sharp ZD, Lee WH. The BRC repeats in *BRCA2* are critical for RAD51 binding and resistance to methyl methanesulfonate treatment. *Proc Natl Acad Sci*, 1998, **95**: 5287–5292.

[12] Wong AK, Pero R, Ormonde PA, Tavtigian SK, Bartel PL. RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene *brca2*. *J Biol Chem*, 1997, **272**: 31941–31944.

[13] Rajasekaran R, Sudandiradoss C, George Priya Doss C and Rao Sethumadhavan. Identification and in silico analysis of functional SNPs of the *BRCA1* gene. *Genomics*, 2007, **90**: 447–452.

[14] Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: The NCBI database of genetic variation. *Nucl Acids Res*, 2001, **29**: 308–311.

[15] Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucl Acids Res*, 2003, **31**: 3812–3814.

[16] Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res*, 2001, **11**: 863–874.

[17] Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*, 2002, **30**(17): 3894–3900.

[18] Sonenberg N. mRNA translation: influence of the 5' and 3' untranslated regions. *Curr Opin Genet*, 1994, **4**(2): 310–315.

[19] Nowak R. Mining treasures from 'junk DNA'. *Science*, 1994, **263**: 608–610.

[20] Pesole G, Liuni S. Internet resources for the functional analysis of 5' and 3' untranslated regions of eukaryotic mRNA. *Trends Genet*, 1999, **15**(9): 378.

[21] Pesole G, Liuni S, Grillo G, Licciulli F, Mignone F, Gissi C, Saccone C. UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res*, 2002, **30**: 335–340.

[22] Cavallo A, Martin AC. Mapping SNPs to protein sequence and structure data. *Bioinformatics*, 2005, **21**(8): 1443–1450.

[23] Lindahl E, Azuara C, Koehl P, Delarue M. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res*, 2006, **34**: W52–W56.

[24] Delarue M, Dumas P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc Natl Acad Sci*, 2004, **101**: 6957–6962.

[25] Han JH, Kerrison N, Chothia C, Teichmann SA. Divergence of interdomain geometry in two-domain proteins. *Structure*, 2006, **14**(5): 935–945.

[26] Varfolomeev SD, Uporov IV. Bioinformatics and molecular modeling in chemical enzymology. Active sites of hydrolases. *Biochemistry*. (Mosc), 2002, 67(10): 1099–1108.

[27] Prokunina L, Alarcn-Riquelme ME. Regulatory SNPs in complex diseases: their identification and functional validation. *Expert Rev Mol Med*, 2004, 1–15.

[28] Kwiatkowska E, Teresiak M, Lamperska KM, Karczewska A, Breborowicz D, Stawicka M, Godlewski D, Krzyzosiak WJ, A. Mackiewicz A. *BRCA2* germline mutations in male breast cancer patients in the Polish population. *Hum Mutat*, 2001, **17**: 73–73.