Modelling and simulation of mutant alleles of breast cancer metastasis suppressor 1 (BRMS1) gene

Muhammad Muzammal Adeel, Muhammad Qasim*, Usman Ali Ashfaq, Muhammad Shareef Masoud, Mahmood ur Rehman, Muhammad Tahir ul Qamar & Muhammad Rizwan Javed

Department of Bioinformatics and Biotechnology, Government College University (GCUF), 38000, Faisalabad, Punjab, Pakistan; Muhammad Qasim – Email: qasemawan@gmail.com; *Corresponding author

Abstract:
Computational tools occupy the prime position in the analysis of large volume of post-genomic data. These tools have advantage over the wet lab experiments in terms of high coverage, cost and time. Breast cancer is the most common cancer in females worldwide. It is a genetically heterogeneous disorder and many genes are involved in the pathway of the disease. Mutations in metastasis suppressor gene are the major cause of the disease. In this study, the effects of mutations in breast cancer metastasis suppressor 1gene upon protein structure and function were examined by means of computational tools and information from databases.This study can be useful to predict the potential effect of every allelic variant, devise new biological experiments and to interpret and predict the patho-physiological impact of new mutations or non-synonymous polymorphisms.

Key words: Breast cancer; BRMS1; Mutation analysis; Homology modeling

Background:
In now a days, computers are as likely to be used by biologists as by any other highly trained professionals, more specifically in field of bioinformatics; which is focused on making predictions about biological systems and to analyze biological data related to different diseases like cancer [1]. As in computational biology tools are used to predict if two proteins interact or not, if prediction is accurate then computational biology could further be used to analyze biological data obtained from a wet lab experiment. This field can be further broken down into molecular modeling and bioinformatics. Several bioinformatics methods are applied for the different mutational disease analysis [2]. Many of them are based on protein sequence, but several are structure-based, as the latter are more reliable and provide more information. In this work, we have built a homology model of mutated BRMS1 gene applying the most updated available methods of Homology modelling through MODELLER, [3] and have investigated the effects of mutations of BRMS1 using different software, including SNAP it is a neural-network based method which evaluates the single amino acid substitution effects on protein functions [4]. I-Mutant2.0 is a web server used for the prediction of protein stability changes upon single-site mutations. It works on Support Vector Machine based method [5]. PolyPhen2.0.9 it uses structural and comparative evolutionary considerations and predicts the impacts of amino acid substitutions on the stability and function of human proteins [6]. IUPred predicts the disorder tendency of particular amino acid [7], PrDOS gives the information about the disorder region of particular protein [8] and HNB servers is a protein function annotation tool having consortium of three different tools (SMART, miniPEDANT and STRING) that collectively involves in functional domain prediction along with the alignment to different protein databases [9].

Worldwide, breast cancer (BC) is the most common cancer affecting women, and its prevalence and death rate are expected to increase by 50% between 2002 and 2020 [10]. This expected increase of cancer rate is extremely high in developing countries and in less than 20 years, these rates are likely to reach 55% increase in occurrence and 58% in death rate [11]. In 2008, total 48,034 new cases of breast cancer were diagnosed, out of which 47,693 (>99%) were in females and 341 (less than 1%) in males [12]. Among Asian countries such as in Pakistan the incidence rate of breast cancer is increasing alarmingly. In Pakistani females breast cancer is considered as most common

malignancy accounting for 34.6% all female cancer, its rate is 2.5 times higher than neighboring countries like Iran and India **[13].**

In other developing countries, breast cancer shows a major affliction among younger women. In women having age less than 45 years, >20% cases of BC occurrence and >20% of BC mortality are observed but on the other hand in developed countries these statistics are less than 10% and 20%, respectively **[14].** However, the definition of "young woman" may differ for breast cancer and in most articles woman having age less than 35, 40 or 45 years is considered as "young" **[15].**

Different risk factors are associated with breast cancer such as sex, age, family history, early menarche and late menopause **[16].**Many reproductive hormones are also considered as influential factor to increase breast cancer risks by affecting cell proliferation, which result in increase the probability of DNA damage and the elevation of cancer growth **[17].** Out of many known risk factors breast cancer is specifically associated with the metastasis. Metastasis is the process which involves in the propagation of cancerous cells and establishment of secondary tumors away from the primary tumor **[18].** The propagation of cancerous cells away from primary tumor is through the circulatory system. In metastasis process these tumor forming cells get attached to other neighboring cells or proteins **[19].** Normally metastasis is suppressed by Breast cancer Metastasis Suppressor 1 (BRMS1) gene. The breast cancer metastasis suppressor 1 (BRMS1) gene is mapped on human chromosome 11q13 **[20]** and suppresses the metastatic probability of tumor forming cells in vitro without affecting the growth of these cells **[21].** Different studies have shown that BRMS1 gene not only suppresses metastasis in xenograft models of breast cancer **[20]** but also in melanoma **[22, 23],** ovarian cancer **[24],** bladder cancer **[25]** and lung cancer **[26].**

It is also suggested that BRMS1 is a nuclear protein containing a damaged leucine zipper motif and coiled-coiled domains, may involving as a component of transcriptional complex **[20].** Inhibition of metastasis may also take place through the interaction of BRMS1 to histone deacetylase **[27, 28].** Recent studies have shown that the mutation in BRMS1 gene leads to its malfunctioning and results in breast cancer. Different genetic mutations such as (F71L, R163Q, D154H and E135K) **[2]** are responsible for the change in expression of BRMS1 gene which is associated with the breast cancer. This approach allowed a preliminary characterization of BRMS1 gene mutations, with prediction of their potential molecular pathogenic effect.

Methodology:
F71L, R163Q, D154H and E135K mutations in family BRMS1 Gene were examined by using various bioinformatics approaches. The effect of single amino acid substitution on protein structure, function and stability were analyzed.

Mutational Screening
Twenty two mutations reported within BRMS1 gene were retrieved from COSMIC database it is a catalogue of somatic mutation in cancer database that contains the information about the mutations in human genome that are associated with the cancer **[2].** These mutations were analyzed through different Bioinformatics tools. Out of twenty two only four mutations

were selected as they were the part of functional domains **Table 1 (see supplementary material).** For the domain analysis of BRMS1 gene, protein function annotation feature of the HNB server was used **[29].** The server shows that whether the mutating amino acids are the part of the functionally important domain or not. Furthermore, 3D modeling of the domain was done to study the single amino acid change in structure as described in next sections.

Comparative modeling
The target BRMS1 protein structure (PDB ID Q9HCU9) and sequence was retrieved from Uniprot KB **[30].** Already reported mutations were induced manually in normal BRMS1 protein sequence. Different structure prediction programs such as the BuildModel option of FoldX version 3.0 beta **[31]**, Pcons **[32]**, Genesilico **[33]**, I- Tasser **[34]** and RaptorX **[35]**, MODELLER **[3]** were used to predict the mutated structures of BRMS1. The program MODELLER was mainly used for this purpose because it yielded fine results and it generated a model having all non-hydrogen atoms. MODELLER satisfied all spatial restraints by the implementation of comparative protein structure modeling **[36].** Qmean **[37]**, PROCHECK **[38]** and ProsA web (https://prosa.services.came.sbg.ac.at/prosa.php) server **[39]** were used for structural assessment. In order to analyze the similarity, Chimera was used to superimpose all mutated and normal BRMS1 protein structures and the Root Mean Square Deviation (RMSD) values were also computed **[40].**

Mutation analysis
NetSurfP (http://www.cbs.dtu.dk/services/NetSurfP/) was used to find out the surface accessible area of all four selected mutations **[41].** The effect of single amino acid substitution on protein functionality and stability was tested using different methods including SNAP **[5]**, which predicts the effects of single amino acid substitution on protein functionality and I-Mutant2.0 **[42]**, which finds the influence of single amino acid changes on protein stability. SNAP server gives output in the form of numerical values that show the reliability of the prediction. These output values range from -100 to 100, showing the neutral to abnormal functioning, respectively. I-Mutant2.0 server predicts the AAG (Gibbs free energy) values before and after mutation. This AAG value is estimated through the equation: AGf wt - AGf mut. Where, wt stands for wild type and mut is mutated protein. Poly-Phen2.0.9 (Polymorphism and Phenotype) (http:// genetics. bwh.harvard.edu/pph2/) server was used to find out whether the mutations were structurally damaging or not **[43].** IUPred **[44]** and PrDOS **[9]** servers were also used to study the effects of mutations and structural disorder on BRMS1 protein.

Results & Discussion:
BRMS1 protein structure prediction through comparative modeling
To investigate the changes in structure of mutated BRMS1 protein, models were generated through homology modeling using MODELLER as shown in (**Figure 1**). For each mutation, MODELLER predicts top scoring models with almost 100% confidence score. However, for the identification of the errors in mutated models number of algorithm's comparison can be helpful e.g. I-Tasser, Gensilico.

BIOINFORMATION

The predicted models were evaluated using different algorithms and methods. Qmean score of generated models of all selected mutations was in the range of 0-1, indicating reliability of the models. Mutated BRMS1 model's evaluation by Procheck for stereo chemical quality showed that structures predicted by MODELLER were best quality models and had 83.3% of residues in favored regions, 14.5% of residues in allowed regions, and 0.9% residues in disallowed regions for F71L. For R163Q mutation, 84.2% of residues were in favored regions, 11.3% of residues in allowed regions, and 2.3% residues in disallowed regions. For D154H mutated model, 85.5% of residues were in favored regions, 10.9% of residues in allowed regions, and 1.4% residues in disallowed regions. While for E135K mutation, 85.5% of residues were in favored regions, 10.9% of residues in allowed regions, and 1.4% residues in disallowed regions. This ensured the quality of the predicted model as >90% residues in favorable regions, provide a good quality model. All the models were also validated using Swiss model **[45]** and ProsA-web **[39]**, which gave a negative Z-score indicating a good quality of mutated models. RMSD values were calculated using Chimera, where high RMSD value shows the maximum deviation of mutated and normal BRMS1 structure **Table 2 (see supplementary material).**

Figure 1: 3D-Protein structures of wild type (PDB ID Q9HCU9) and mutated BRMS1 proteins **1A-1D**) Wild type BRMS1 protein; **2A-2D**) Mutated BRMS1 proteins with F71L, E135K, D154H andR163Q mutations, respectively. **3A-3D)** Superimposed structures of wild type and mutated proteins withF71L, E135K, D154H and R163Q mutations, respectively. Mutated residue position is shown in red color with labeling. All the structures are manipulated using Chimera.
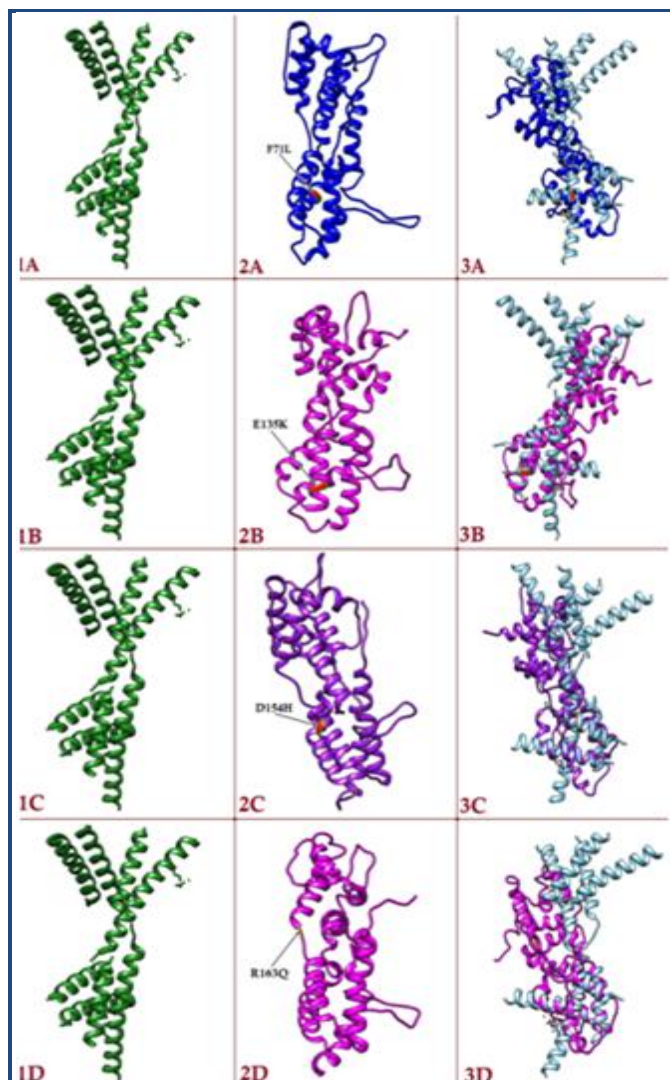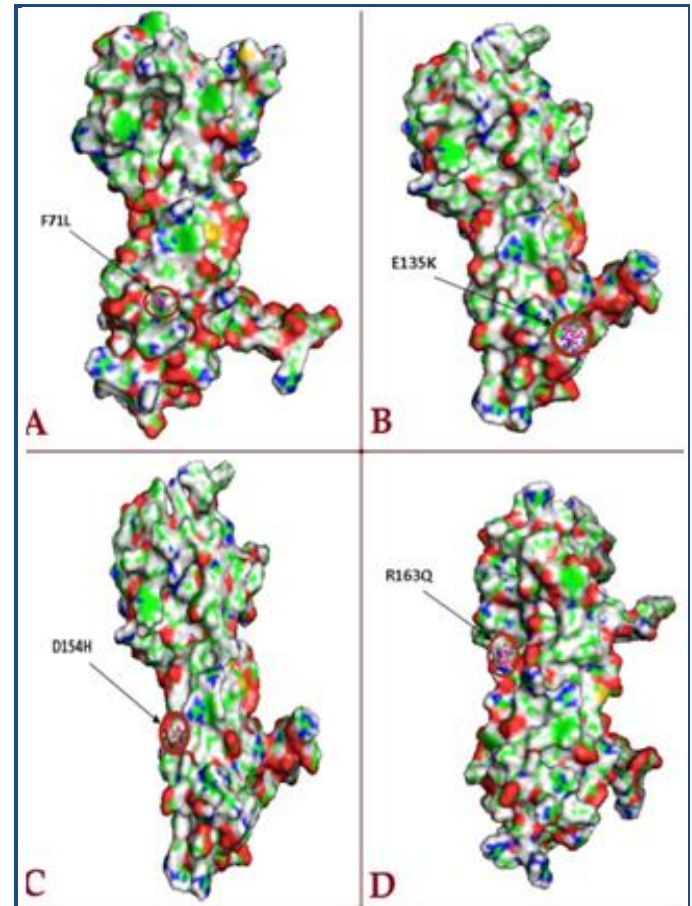

Figure 2: Electrostatic surface potential of mutated protein structures: **A)** Mutated protein structure with F71L mutation; **B)** with E135K mutation; **C**) with D154H mutation and **D)** with R163Q mutation. Red spots on the surfaceare showing the potential change and buried mutated amino acids are represented by labeled red circles.

Mutation analysis
Surface accessibility was checked by using NetSurfP **[40].** NetSurfP server provided information about the exposed and buried amino acids of the protein. Buried amino acid residues are considered to be more important because they are involved in the formation of core interactions that are necessary for protein stability **[41].** Therefore, it was determined whether the mutated residues were buried or exposed. Residues number 71 was buried while other residues number 163,154 and 135 were found to be exposed shown in **(Figure 2)**. Any mutation in these residues may result in conformational changes of the protein and may alter stability of the protein.

Impact of mutations on function and stability
After investigating the SNAP output, it was analyzed that whether mutations F71L, R163Q, D154H and E135K can affect the function of the protein. The predictions for F71L and D154H

mutation were non-neutral showing the change in the function with the reliability value of 2 and expected accuracy of 70%.While, mutations R163Q and E135K were found to be neutral. I- Mutant2.0 server was used for the prediction of change on protein stability. The output provided the DDG score of -1.91 for F71L, -0.70 for E135K, 0.20 for D154H and 0.20 for R163Q. A score below 0 shows decrease in stability and a score above 0 shows an increase in stability. Thus, all these mutations may have effect on the stability of the structure indicated by their negative score. The scale of damaging effect provided by PolyPhen2.0.9 shows a score starting from 0.00 to 1.00 where a score close to 1 is considered as a potential damage. The output of PolyPhen-2.9.0 server provided scores for all selected mutations in the range of 0.994-1.00. Mutation F71L and E135K had score of 1 showing damaging effect, while 0.49 for R163Q has less damaging effect as compared to D154H which also found to have a damaging effect with score of 0.806 **Table 3 (see supplementary material).**

Domain analysis
Functional domains of BRMS1 gene were predicted through HNB server (http://dag.embl- heidelberg.de/hnb cgi/show overview page.pl?taskId= notask). It was found that there are two coiled-coil domains in BRMS1 protein. One is from 51 to 81 amino acids and other is from 147 to 180 amino acid residues.

Protein disorder analysis
Mutations are associated with the disorder of a protein which ultimately causes the loss of a regular secondary structure. Disordered regions were predicted using IUPred server, which predicted that mutation F71L have disorder tendency of 0.4017, R163Q have 0.3948, D154H have 0.2436 and E135K have 0.1611, respectively. Disordered tendency for mutating residues was calculated at the scale of 0-1.00. The mutating residues showed low disorder tendency. Additionally, PrDOS server showed region 1-52, 185-201 and 235-246 as disordered regions. Mutating residues were not found in the disordered region of the protein.

Conclusion:
Our results have shown that missense mutations F71L, R163Q, D154H and E135K have strong structural, conformational and functional effects on mutated protein. Moreover it also affects the stability of protein therefore these mutations are pathogenic.

References:
[1] Ismaeel AG & Ablahad AA, IRACST 2013 3.
[2] Forbes SA *et al._Nucleic Acids Res*. 2011 **39**: D945 [PMID: 20952405]
[3] Andrej S & Blundell TL, *J Mol Bio*. 1993 **234:** 779 [PMID: 8254673]
[4] Thusberg J & Vihinen M, *HumMutat* 2006 **27**: 1230 [PMID: 16986121]
[5] Bromberg Y & Rost B, *Nucleic Acids Res*. 2007 **35**: 3823 [PMID: 17526529]
[6] Emidio C *et al. Nucleic Acids Res*. 2005 **33**: 306
[7] Adzhubei I *et al. Curr Protoc Hum Genet*. 2013 **Chapter 7**: Unit7.20. [PMID: 23315928]
[8] Dosztanyi Z *et al. J Mol Biol*. 2005 **347**: 827 [PMID: 15769473]
[9] Ishida T & Kinoshita K, *Nucleic Acids Res.* 2007 **35**: W460 [PMID: 17567614]
[10] Anderson BO *et al. Cancer* 2008 **113**: 2221 [PMID: 18816619]
[11] Anderson BO *et al. lancet oncol*. 2011 **12**: 387 [PMID: 21463833]
[12] Jemal A *et al. CA Cancer J Clin*. 2011 **61:** 134. [PMID: 21296855]
[13] Shaukat U *et al. Asian Pac J Cancer*. 2013 **14**: 5625 [PMID: 24289553]
[14] Knaul F *et al. SaludPublicaMex*. 2009 **51**: 220 [PMID: 19967277]
[15] Gabriel CA & Susan MD, *Breast Cancer Res*. 2010 **12**: 212 [PMID: 21067532]
[16] Althuis MD *et al. Cancer Causes Control* 2003 **14:** 151 [PMID: 12749720]
[17] Cady B *et al. CA Cancer J Clin*. 1998 **48:** 49 [PMID: 9449933]
[18] Germanov E *et al. Int JMolMed*. 2006 **18**: 1025 [PMID: 17089005]
[19] Woodhouse EC *et al. Cancer* 1997 **80:** 1529 [PMID: 9362419]
[20] Seraj MJ *et al. Cancer Res*. 2000 **60:** 2764 [PMID: 10850410]
[21] Steeg PS, *Journal of the National Cancer Institute* 2004 **96**: E4
[22] Shevde LA *et al. Exp Cell Res*. 2002 **273**: 229 [PMID: 11822878]
[23] Li J *et al. Oncogene* 2011 **30:** 896 [PMID: 20935672]
[24] Zhang S *et al. Int J Gynecol Cancer*. 2006 **16:** 522 [PMID: 16681721]
[25] Seraj MJ *et al. Clin Exp Metastasis*. 2000 **18**: 519 [PMID: 11592309]
[26] Smith PW *et al. Cancer Lett*. 2009 **276:** 196 [PMID: 19111386]
[27] Meehan WJ & Danny RW, *Clin Exp Metastasis.* 2003 **20**: 45 [PMID:12650606]
[28] Meehan WJ *et al. J Biol Chem*. 2004 **279**: 1562 [PMID:14581478]
[29] http://dag.embl-heidelberg.de/hnb cgi/show input page.pl
[30] http://www.uniprot.org/uniprot/Q9HCU9
[31] Schymkowitz J *et al. Nucleic Acids Res*. 2005 **33:** W382 [PMID: 15980494]
[32] Wallner B & Elofsson A, *Protein Sci*. 2006 **15**: 900 [PMID: 16522791]
[33] Kurowski MA *et al. Nucleic Acids Res*. 2003 **31**: 3305 [PMID: 12824313]
[34] Roy A *et al. NatProtoc*. 2010 **5**: 725
[35] Peng J & Xu J, *Proteins* 2011 **79**: 1930 [PMID: 21465564]
[36] Eswar N *et al. Curr Protoc Bioinformatics*. 2006 **Chapter 5**: Unit 5.6 [PMID: 18428767]
[37] Benkert P *et al. Proteins* 2008 **71**: 261 [PMID: 17932912]
[38] Laskowski RA *et al. Journal of applied crystallography* 1993 **26**: 283
[39] Wiederstein M *et al. Nucleic Acids Res*. 2007 **35**: W407 [PMID: 17517781]
[40] Yang Z *et al. J Struct Biol*. 2012 **179**: 269 [PMID: 21963794]
[41] Petersen B *et al. BMC Struct Biol*. 2009 **9:** 51 [PMID: 19646261]
[42] Capriotti E *et al. BMC Bioinformatics*. 2008 **9**: S6 [PMID: 18387208]
[43] Adzhubei IA *et al. NatMethods*. 2010 **7:** 248 [PMID: 20354512]

BIOINFORMATION

[44] Dosztanyi Z *et al. Bioinformatics* 2005 **21**: 3433 [PMID: 15955779]

[45] http://swissmodel.expasy.org/workspace/index.php?func =tools structureassessment1&us erid=USERID&token=T OKEN

BIOINFORMATION

Supplementary material:

Table 1: Genetic Mutation Information in COSMIC Database

Mutation ID	Mutation	Mutation Type	CDS Mutation	GRCh37	CDD
COSM930719	p.F71L	Missense	c.213C>A	11:66109016..66109026	NP056214.1
COSM930715	p.R163Q	Missense	c.488G>A	11:66108287..66108297	NP056214.1
COSM689912	p.R154Q	Missense	c.460G>C	11:66108315..66108325	NP056214.1
COSM194375	p.E135K	Missense	c.403G>A	11:66108468..66108478	NP056214.1

Where:**CDD** = Conserved Domain Database, **GRCh37**=Genome Reference Consortium Human genome build 37, **CDS =** Conserved Domain Sequence

Table 2: Statistical Evaluation of Mutated Protein Structures

Mutations	Qmean Score	Procheck Ramachandran	RMSD(Å)	Z score
F71L	0.175	83.3%	0.851	-2.11
R163Q	0.163	84.2%	0.698	-1.71
D154H	0.175	85.5%	0.812	-1.95
E135K	0.175	85.5%	0.812	-1.95

Table 3: Effect of mutations on BRMS1 Protein Stability and Structural Damage

Mutation	Functional mutation	Stability	Damaging
F71L	🟥	🟥	🟥
R163Q	🟥	🟥	🟥
D154H	🟩	🟥	🟥
E135K	🟥	🟥	🟥

Where:Red boxes are for high effect and green for low/no effect.