Neoplasia 51 (2024) 100989

Contents lists available at ScienceDirect

Neoplasia

journal homepage: www.elsevier.com/locate/neo

Original Research

Preliminary insights on the mutational spectrum of *BRCA1* and *BRCA2* genes in Pakhtun ethnicity breast cancer patients from Khyber Pakhtunkhwa (KP), Pakistan

Hilal Ahmad [a], Asif Ali [b,c,d,**], Roshan Ali [a], Ali Talha Khalil [e,*], Ishaq Khan [a], Mah Muneer Khan [f], Mohammed Alorini [g]

[a] Institute of Basic Medical Sciences (IBMS), Khyber Medical University, Peshawar, KP, Pakistan
[b] Institute of Pathology and Diagnostic Medicine (IPDM), Khyber Medical University, Peshawar, KP, Pakistan
[c] College of Medicine, Gulf Medical University, Ajman, United Arab Emirates
[d] School of Medicine, University of Glasgow, UK
[e] Department of Pathology, Lady Reading Hospital Medical Teaching Institution, Peshawar, KP, Pakistan
[f] Department of Surgery, Khyber Teaching Hospital, Medical Teaching Institution, Peshawar, KP, Pakistan
[g] Department of Basic Medical Sciences, Unaizah College of Medicine and Medical Sciences, Qassim University, Unaizah, Saudi Arabia

ARTICLE INFO

Keywords:
Breast cancer
BRCA1
BRCA2
Novel
Mutations
Molecular simulations
ConSurf
Pakhtun

ABSTRACT

Gene mutations are a source of genetic instability which fuels the progression of cancer. Mutations in *BRCA1* and *BRCA2* are considered as major drivers in the progression of breast cancer and their detection indispensable for devising therapeutic and management approaches. The current study aims to identify novel pathogenic and recurrent mutations in *BRCA1* and *BRCA2* in Pakhtun population from the Khyber Pakhtunkhwa. To determine the *BRCA1* and *BRCA2* pathogenic mutation prevalence in Pakhtun population from KP, whole exome sequencing of 19 patients along with 6 normal FFPE embedded blocks were performed. The pathogenicity of the mutations were determined and they were further correlated with different hormonal, sociogenetic and clinicopathological features. We obtained a total of 10 mutations (5 somatic and 5 germline) in *BRCA1* while 27 mutations (24 somatic and 3 germline) for *BRCA2*. Five and seventeen pathogenic or deleterious mutations were identified in *BRCA1* and *BRCA2* respectively by examining the mutational spectrum through SIFT, PolyPhen-2 and Mutation Taster. Among the SNVs, *BRCA1* $^{p.P824L}$, *BRCA2* $^{p.\ P153Q,\ p.I180F,\ p.D559Y,\ p.G1529R,\ p.L1576F,\ p.E2229K}$ were identified as mutations of the interaction sites as predicted by the deep algorithm based ISPRED-SEQ prediction tool. SAAFEQ-SEQ web-based algorithm was used to calculate the changes in free energy and effect of SNVs on protein stability. All SNVs were found to have a destabilizing effect on the protein. ConSurf database was used to determine the evolutionary conservation scores and nature of the mutated residues. Gromacs 4.5 was used for the molecular simulations. Ramachandran plots were generated using procheck server. STRING and GeneMania was used for prediction of the gene interactions. The highest number of mutations (*BRCA1* 7/10, 70 %) were on exon 9 and (*BRCA2*, 11/27; 40 %) were on exon 11. 40 % and 60 % of the *BRCA2* mutations were associated Grade 2 and Grade 3 tumors respectively. The present study reveals unique *BRCA1* and *BRCA2* mutations in Pakhtun population. We further suggest sequencing of the large cohorts for further characterizing the pathogenic mutations.

Introduction

Breast cancer (BC); a multifactorial disease, is now considered as a well-recognized threat to the health and well-being of women. According to the statistical data, BC is most common malignancy in women in respect of the incidence and mortality. Furthermore, it is predicted that the incidence of BC would increase by 33.8 % by 2040 [1,2]. Only in the year 2020, ~ 2.3 million cases of BC were diagnosed with ~ 685,000 mortalities [3]. While the BC is affecting population all over the world, its incidence, mortalities and survival rates varies in different regions due to changes in risk factors

* Corresponding author at: Department of Pathology, Lady Reading Hospital Medical Teaching Institution, Peshawar, KP, Pakistan.
** Corresponding author at: Institute of Pathology and Diagnostic Medicine (IPDM), Khyber Medical University, Peshawar, KP, Pakistan.
E-mail addresses: draliasif7@gmail.com (A. Ali), alitalha.khalil@lrh.edu.pk (A.T. Khalil).

https://doi.org/10.1016/j.neo.2024.100989
Received 12 December 2023; Accepted 11 March 2024
Available online 27 March 2024

like genetic features, hormonal characteristics, environmental conditions, life style and health facilities [4]. BC survival rates are much high in the west as compared to low middle income countries because of better screening and healthcare facilities [5]. Like other countries, the situation is alarming in Pakistan as the region has the highest incidence of BC in Asia [6]. While, the exact data of the BC in Pakistan is scarce, some reports have indicated an annual incidence of 90,000 cases annually with ~ 16,000 deaths/year [7]. It's predicted that the mortality rates due to breast cancer may spike up to 62 % by 2030 in Pakistan, with patients aged between 30-40 years being most susceptible [8]. Delayed prognosis and management of BC is a major challenge in Pakistan owing to plethora socioeconomic and structural barriers like limited awareness, financial constraints, reluctance in acquiring social support, stigmatization, feminine sensitivity, aversion to male doctors and scarcity of medical resources [7,9]. Previous reports have revealed that ~ 89 % of BC patients are diagnosed at later stage due to lack of awareness [10].

BC is usually linked to a sequence of somatic and germline aberrations in a single cell which leads to the full-fledged malignancy, resulting in dysregulated cell cycle, angiogenesis and suppression of apoptosis causing uncontrolled cellular proliferation. Mutations in high penetrance genes like *BRCA1* (BReast CAncer gene 1) and *BRCA2* (BReast CAncer gene 2) genes increases the risk of having BC by 80 % and are also associated with the hereditary breast cancer [11,12].

BRCA1 and *BRCA2* are the tumor suppressor genes positioned at 17th and 13th chromosome and constitutes 24 and 27 exons respectively. *BRCA1* codes for multidomain protein what constitute 1,863 amino acids. While the *BRCA2* encodes for a 3,418 amino acids multidomain protein [2,13,14]. These encoded proteins are known as Tumor Suppressor Proteins (TSG proteins). The primary role of these TSG proteins encoded by *BRCA1* and *BRCA2* is to act as growth suppressor, maintain the integrity of genetic material by repairing damaged DNA and ensure that the genetic material is preserved. Any mutation in *BRCA1* and *BRCA2* essentially compromise the repair function of these genes, destabilize the DNA and loss of the tumor suppressor activity which eventually leads to accumulation of mutations and progression towards breast cancer [15]. Breast malignancy resulting from mutated *BRCA1* and *BRCA2* are associated with lymphatic penetrance and higher mitotic rates [16]. Previous reports on the genetic mutation in *BRCA1* and *BRCA2* indicated that the mutations may be specific to a region, race and ethnic groups [2].

Despite of their clinical significance, no studies on the Whole Exome Sequencing (WES) of *BRCA1* and *BRCA2* genes are reported from Khyber Pakhtunkhwa region or Pakhtun ethnicity. A detailed analysis of the WES data in the specifically enrolled cohort of Pakhtun ethnicity was carried out for analyzing the significant mutations and their association with clinicopathological findings and social determinants. The advanced Next Generation Sequencing – NGS (Whole Exome) is recognized as a cutting-edge tool for analysis of the mutations. Herein, we present the WES data of 19 confirmed BCs patients with IDC (Invasive Ductal Carcinoma) and 6 paired adjacent normal tissues and identified the mutations, their significance and clinicopathological relevance.

Experimental

Enrollment criteria

Inclusion criteria was patients being diagnosed with TNM Stage 0-IV of the breast cancer with Invasive Ductal Carcinoma. Only FFPE breast tumor blocks with at least 70 % tumor purity were included. Exclusion criteria included patients with secondary tumor in another organ.

Data collection procedure

Patients fulfilling the inclusion criteria were enrolled from selected teaching hospitals (Khyber Teaching Hospital Peshawar, Hayatabad Medical Complex Peshawar). All steps were performed according to the principles of the Helsinki Declaration. This study was approved by the research ethics committee of Khyber Medical University, Peshawar vide: DIR/KMU-EB/VA/000651. The patients were guided on the purposes of the study. Written informed consent (English and Urdu) was obtained and signed from the patients/or their legal guardian. Data including their medical history was recorded.

Tumor biopsy collection and Immunohistochemistry

Formalin (10 %) was used to obtain the tumor biopsy and transferred to the Histopathology section at KMU Peshawar. Gross examination was performed using CAP procedure after 20. General appearance, consistency and color were recorded on a predesigned proforma. FFPE tissue blocks were made in a labeled embedding cassette. Rotary microtome was used to obtain 5 μm sections from the FFPE blocks, labelled and later deparaffinized in xylene. H&E stain was used for the initial confirmation of the invasive ductal carcinoma and the samples were further examined for immunohistochemistry using diverse markers like ER (ERα ISO8430), PR (PR ISO6830), HER2/neu (HER2; A048529) and Ki-67 (IS 62630) is considered important for diagnostics and biomarker expression [17]. FFPE tumor blocks with enriched tumor cell populations were included in the study whereas for normal tissues the FFPE block was prepared from tissue samples at least 2 cm away from the tumor.

Next generation whole exome sequencing (NGS-WES)

Nineteen rich tumor FFPE blocks and six normal FFPE blocks were supplied to Macrogen (Korea) for commercial NGS-WES. Illumina NGS platform with 151 bp paired end reads was used for acquiring the whole exome sequence data. Library preparations was performed using SureSelect V6-(FFPE) reagent kit. An overview of the NGS-WES from the raw data reports as supplied by the Macrogen (Korea) is indicated hereunder;

Preparation of sample and library

The genomic DNA was extracted and monitored for Quality Control. QC was performed on 1 % Agarose gel (30 min at 160V) after loading 10 μL of DNA. Only samples with good bands on gel and with sufficient quantity were considered for library preparation. The genomic library was prepared by 1st random fragmentation and followed by 5′- 3′ end adapter ligation. The resulting fragments were PCR amplified and gel purified. The obtained genomic DNA was quantified using Victor 3 fluorometry which is a fluorescent based quantification method.

Whole exome sequencing

Clonal clusters were amplified by bridge amplification. Illumina SBS technology was used to obtain the sequence data which depends on reversible terminator-based technology for the detection of single bases as they are being incorporated into DNA template strands. The raw files were supplied in the form of FASTQ format.

Sequence analysis

After receiving the FASTQ files, they were examined for QC using (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). FASTQ files were aligned with human reference genome (hg38-UCSC) using Burrows-Wheeler aligner (BWA) tool as reported previously [18]. After alignment, the SAMtools were used for converting the alignment files in to sorted, indexed binary alignment map (BAM). The SAM mpileup tool was applied for obtaining information form the BAM files. Variant Calling Format (VCF) file was obtained using BCF tools and was submitted to ANNOVAR program for creating a final CSV file with complete information. R studio software was used filter out the synonymous single nucleotide variants (SNVs). Different mutation prediction databases like SIFT, Polyphen-2 and Mutation Taster were used identifying the pathogenic mutations. ISPRED-SEQ prediction tool was used for the prediction of the interacting and non-interacting sites. SAAFEC-SEQ server was used to calculate the free

energy and stability of the proteins upon single nucleotide variation. SWIS-MODEL was used to download the available pdb structures and the subsequent visualization was carried out using PyMOL. The network of the interactions of the *BRCA1* and *BRCA2* were retrieved from STRING and GeneMANIA. The mutational spectrum was analyzed in association with the clinicopathological findings and social determinants.

Molecular dynamic simulation

GROMACS 4.5 was used for the simulation of the selected wild and mutant proteins. The water molecules were removed and the OPLS-AA/L all-atom force field was selected. For the solvation process, a generic equilibrated 3-point solvent model i.e. spc216.gro was used. Grompp was used to assemble the binary input file. The MD simulation was finally performed and the comparative structural analysis was performed using tools like gyrate, RMSD and the Gromacs energies (pressure, temperature, density) were obtained. The results were plotted in Origin Pro 8.5. The validity of the wild and mutated models was assessed using Ramachandran plot using PROCHECK server [19].

Evolutionary conservation analysis

For evolutionary conservation prediction of the mutated residue position, the online ConSurf [https://consurf.tau.ac.il/] database was used for conservation prediction. An empirical Bayesian algorithm is applied and the results are colorimetric with scores from 1-9. The exposed and buried nature as well as residue of functional and structural nature are predicted [20].

Mutation mapping and modelling

cBioPortal database was used to create the lollipop plots of *BRCA1* and *BRCA2* while the 3D structure of the mutant proteins was visualized using PyMOL.

A schematic of the study design is indicated in Fig. 1.

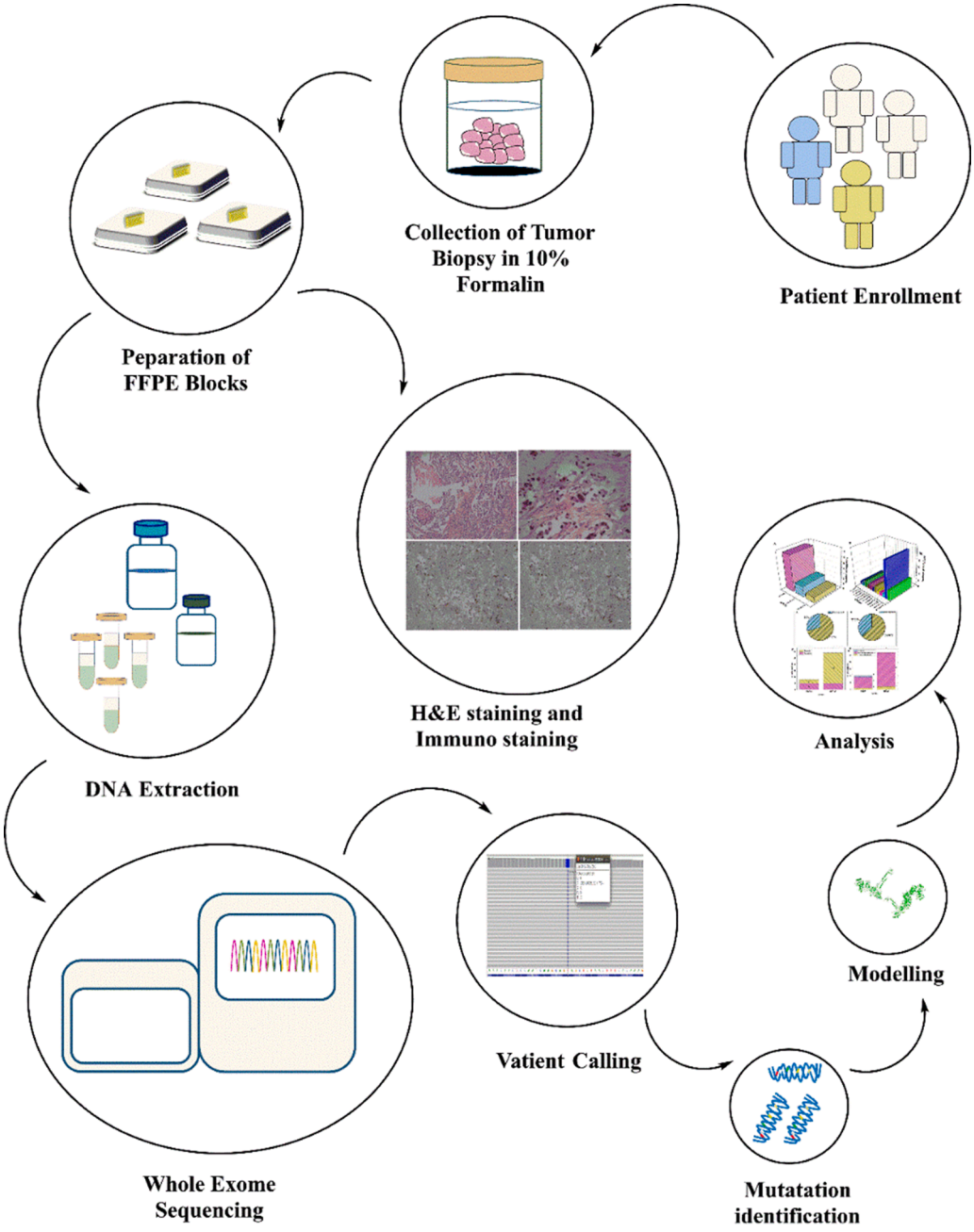


Fig. 1. Schematic of the study design.

Results

Cohort details

The inset Fig. 2 (A-I), reveals the basic details and general demographic charactersitics of the cohort. The average age of the enrolled participants was 46.47 years with mean BMI of 26 kg m^{-2} whereas the average tumor size is 1.96 cm (Fig. 2A). Fig. 2B indicates the age wise distribution of the patients with majority of the patients falling in the age bracket of 31-40 years i.e. 42.18 %. Patients falling within the BMI range of 21-25 were maximum. Fourteen patients were having cancer history in their families (Fig. 2D) while 7 participants have had the history of abortion (Fig. 2E). Investigations of the molecular subtypes revealed in the order of Luminal A > Triple Negative > Luminal B > HER+ as indicated in Fig. 2F. The ER IHC staining revealed that 73.7 % ER positivity, 68.4 % PR positivity and 21 % HER2 positivity as indicated in Fig. 2G. The expression of Ki-67 tumor markers was also studied for active cellular proliferation in the breast cancer. It was revealed that 63 % of the breast tissues has Ki-67 score between 0-30, whereas, 26 % between 31-60 and 10.5 % between 61-90, as indicated in Fig. 2H. The physiological grading revealed that most of the breast cancer tissues were characterized as Grade 2 i.e. 63.15 % of the breast cancers were appearing slightly abnormal, while, 31.57 % of the tissues were characterized as Grade 3 tumors and appeared highly abnormal and were poorly differentiated (Fig. 2I).

Fig. 3A–D summarizes the marital status (Fig. 3C) and physiological profile of the cohort. All patients enrolled in the current study were married except for one. About 57.9 % of the enrolled cohort were postmenopausal (Fig. 3A) whereas the majority i.e. 89.47 % were having a regular menstrual cycle (Fig. 3C). About 78 % of the patients were having their menarche year starting from 13 to 14 years (Fig. 3D).

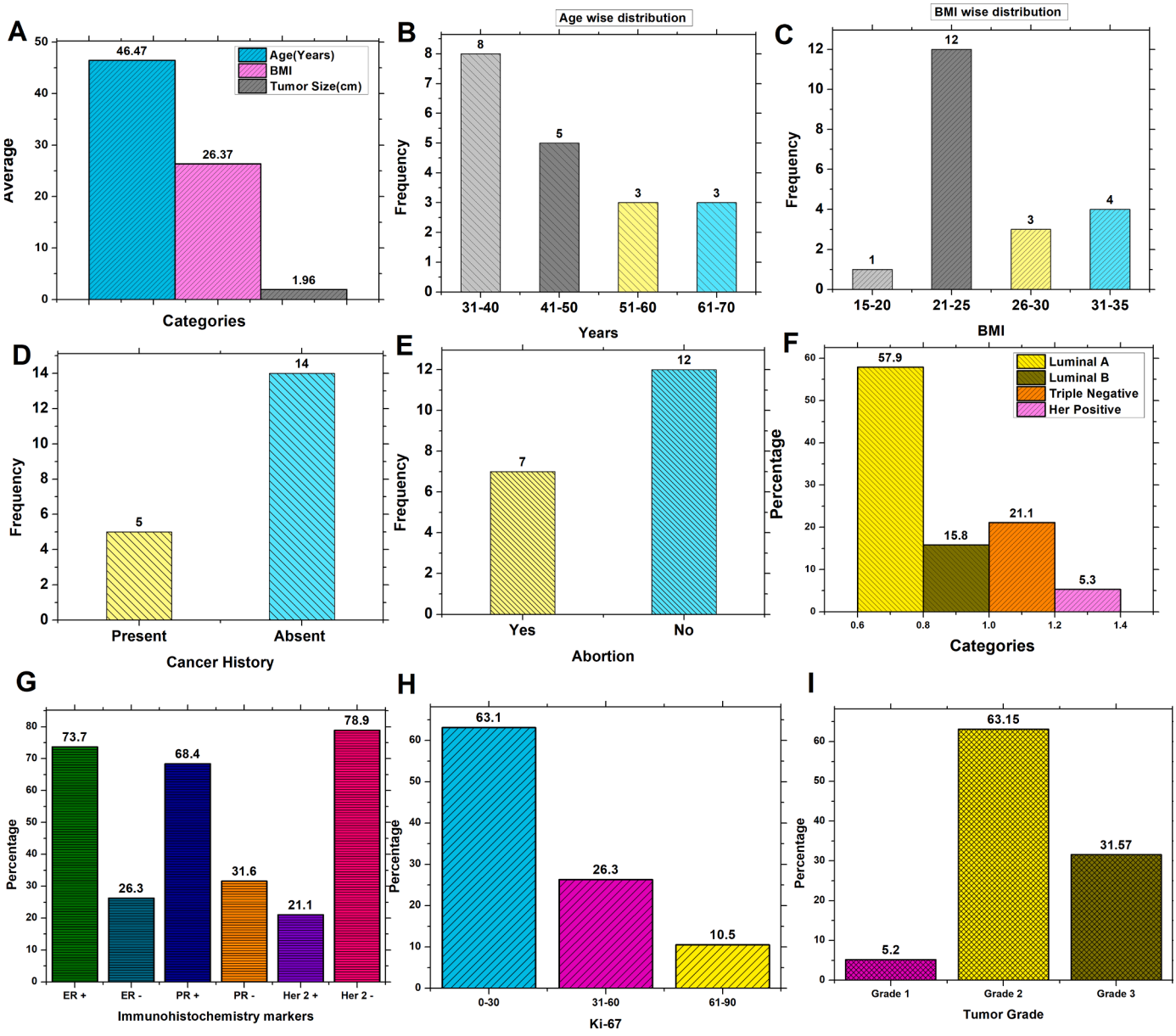


Fig. 2. Stratification of the basic epidemiological data of the enrolled patients; (A): Average Age, BMI and Tumor size (cm); (B): Age wise distribution; (C): BMI wise distribution; (D): History of cancer in the family; (E): History of Abortion; (F): Characterization based on Molecular subtypes; (G): Characterization based on immunohistochemistry markers; (H): Characterization based on Ki-67 marker; (I): Characterization based on tumor grade.

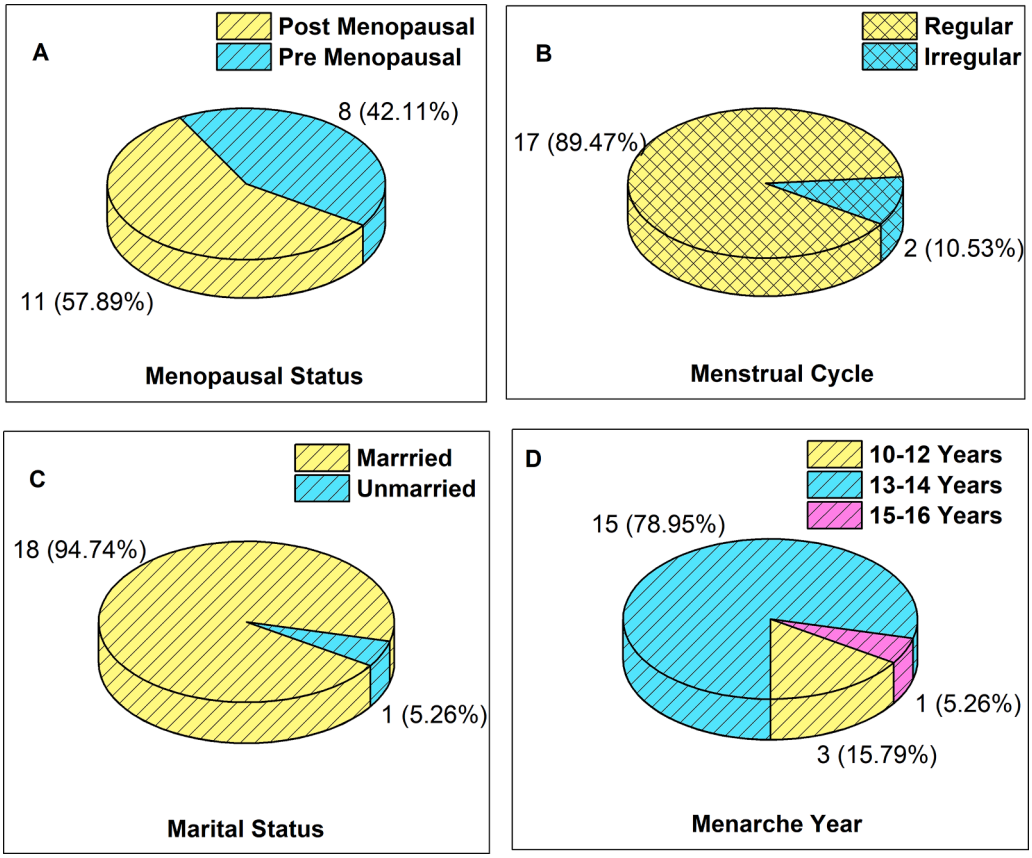


Fig. 3. Stratification of the various hormonal and non-hormonal characteristics of the enrolled patients; (A): Menopausal status; (B): Menstrual Cycle; (C): Marital Status; (D: Menarche year.

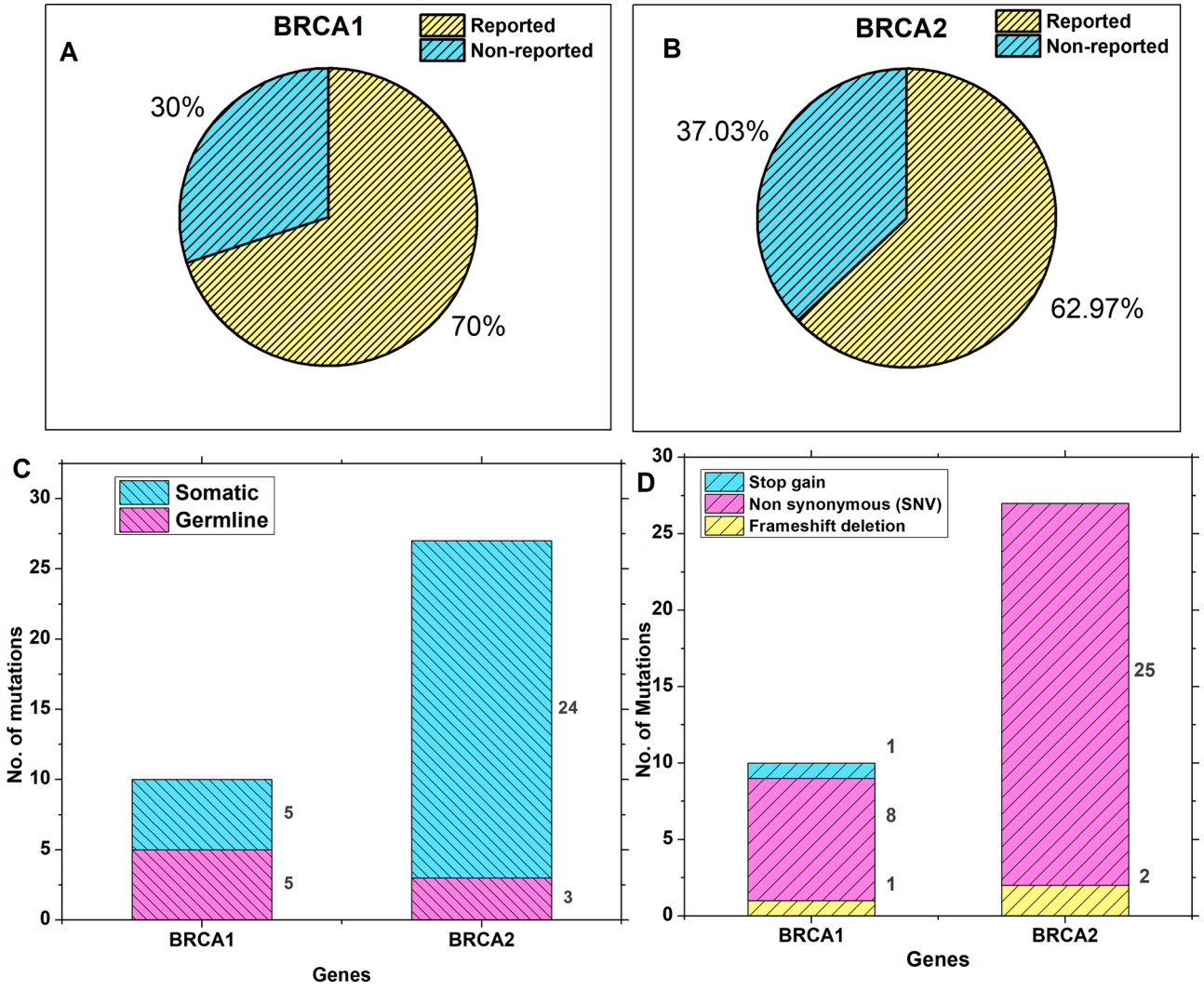


Fig. 4. Summary of genetic mutations; (A): Mutations in *BRCA1*; (B): Mutations in *BRCA2*; (C): Somatic and Germline mutations in *BRCA1* and *BRCA2;* (D): Different types of mutations in *BRCA1* and *BRCA2*.

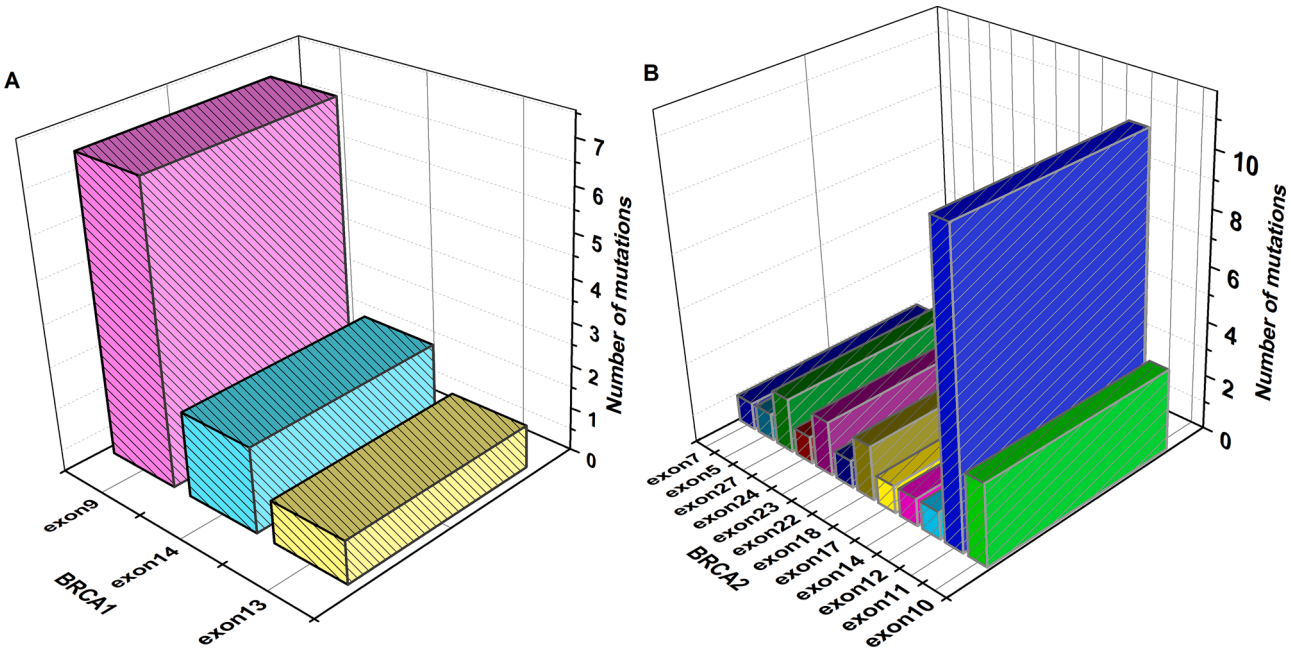


Fig. 5. Mutation spectrum on the exons (A) *BRCA1* and (B) *BRCA2*.

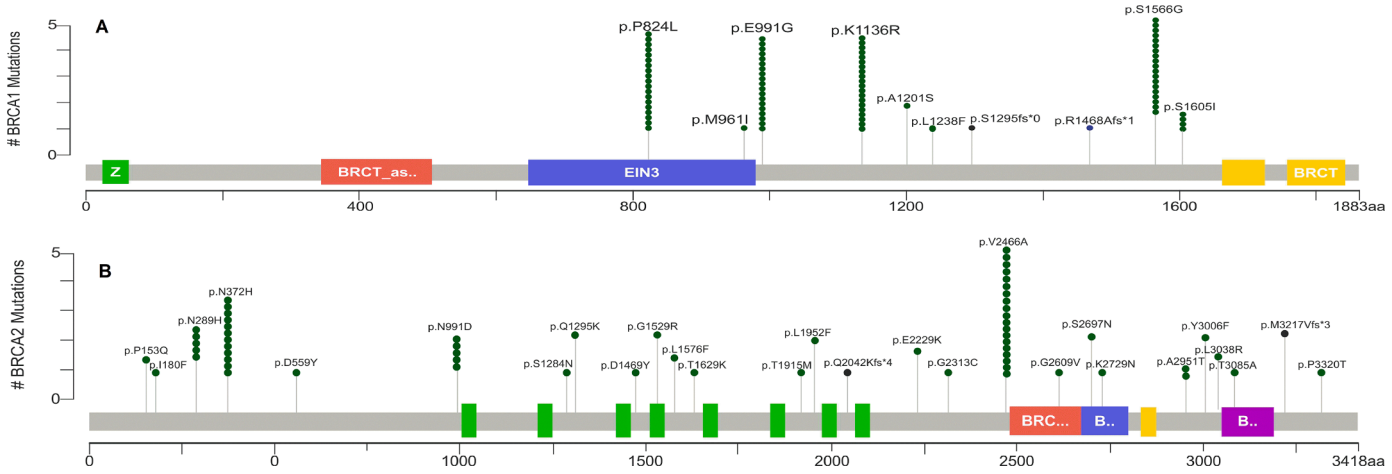


Fig. 6. Lollipop plots showing the distribution of different mutations (A) *BRCA1* (B) *BRCA2*. The plots were obtained by the informatic tool Mutation Mapper and cBioPortal for Cancer Genomics.

Mutational spectrum

The inset Fig. 4A–D summarizes the initial findings regarding the mutational landscape of the enrolled cohort. By searching in the different databases, it was found out that 30 % and 37 % of the total mutations obtained in current study are not reported before for *BRCA1* (Fig. 4A) and *BRCA2* respectively (Fig. 4B). In comparison, the *BRCA2* gene was found to carry excessive mutations as compared to *BRCA1*. The variants present only in tumor tissue samples were classified as somatic mutations while variants present in both tumor and paired normal tissue samples were classified as germline mutations. Out of the total ten mutations, 5 mutations were somatic and 5 mutations were germline. For, *BRCA2*, the number of somatic and germline mutations were 24 and 3 respectively as indicated in Fig. 4C. Further analysis of the mutations revealed that the *BRCA1* has 8 nonsynonymous single nucleotide variants and 1 stop gain and frameshift deletion mutation each (Fig. 4C). The *BRCA2* gene had 2 frameshift deletions while 25 nonsynonymous single nucleotide variants as indicated in Fig. 4D. The exon wise mutations in *BRCA1* and *BRCA2* are indicated in Fig. 5 A-B. A total of 7 *BRCA1* mutations were located on exon 9, while 2 on exon 14 and 1 on exon 13. For *BRCA2*, the highest number of mutations (11) were located on exon 11, followed by 3 mutations on exon 10. Mutations were also located on other exons like exon 5, exon 7, exon 12, exon 14, exon 17, exon 18, exon 22, exon 23, exon 24 and exon 27, as indicated in inset Fig. 5A-B. The inset Fig. 6A-B indicates the lollipop plot of the sequence and location of most significant genetic alterations in the enrolled cancer patients, across *BRCA1* (Fig. 6A) and *BRCA2* (Fig. 6B).

The inset of Figs. 7A–C and 8A–C indicates the in-depth scrutiny of the mutation predications regarding their oncogenic potential. The mutations were screened for pathogenicity across databases like PolyPhen-2, Mutation taster and SIFT. The predictions from SIFT (Sorting Intolerant from Tolerant) encompasses the predictions of amino acid substitution manifesting a phenotypic effect whereas the PolyPhen-2 prediction relies on the allelic replacement. Mutation Taster scores are obtained on the disease-causing potential of the mutations. For *BRCA1*, the SIFT prediction revealed 3 potentially deleterious mutations,

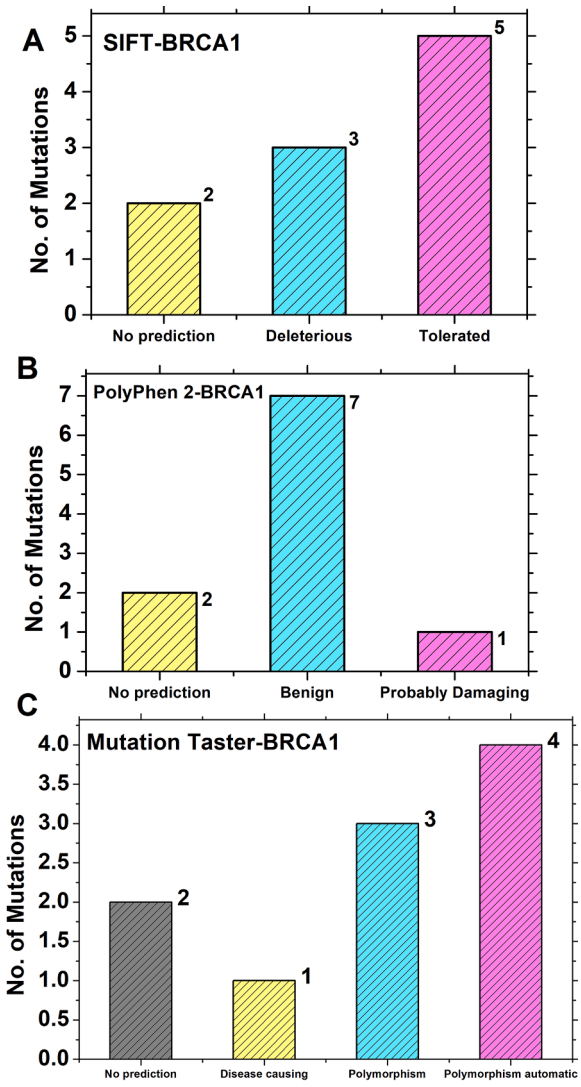


Fig. 7. Mutation pathogenicity in *BRCA1* based on (A): SIFT, (B): Poly phen-2 and (C): Mutation Taster and databases.

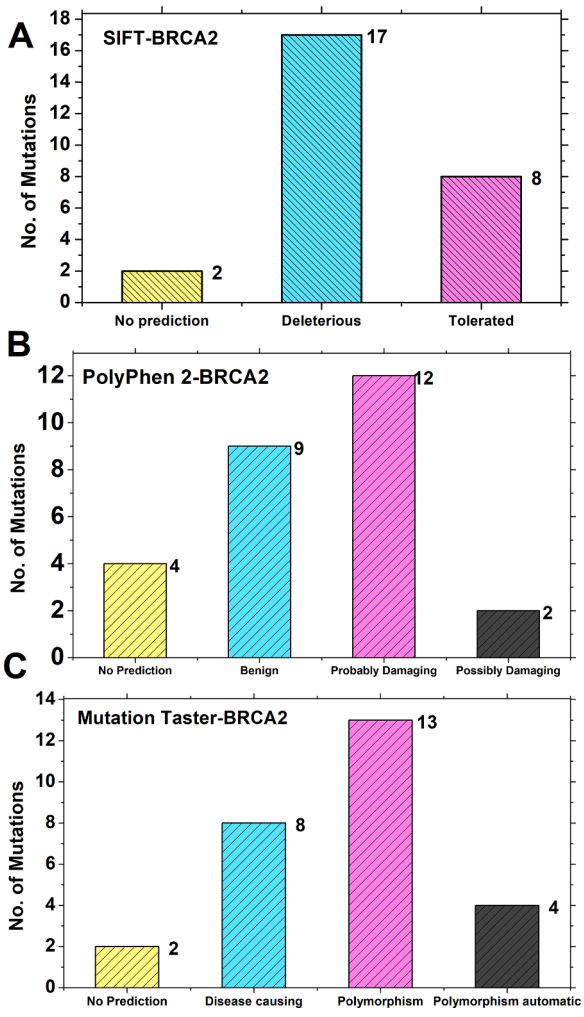


Fig. 8. Mutation pathogenicity in *BRCA2* based on (A): SIFT, (B): Poly phen-2 and (C): Mutation Taster and databases.

PolyPhen-2 prediction revealed 1 potentially damaging mutation and Mutation Taster revealed 1 disease causing mutation. The results for mutation prediction of *BRCA1* mutations are summarized in Fig. 7A–C. For *BRCA2*, a total of 17 potentially deleterious mutations were identified by SIFT predictions. PolyPhen-2 and Mutation Taster revealed 12 probably damaging and 8 disease causing mutations respectively as summarized in Fig. 8A–C.

The inset Fig. 9A–D reveals data retrieved from the ISPRED-SEQ prediction tool (https://ispredws.biocomp.unibo.it/sequence/) which is a deep-learning based web platform for the predication of the interaction and non-interaction sites of a protein. A total of seven interacting site mutations were revealed for *BRCA1* and *BRCA2* as indicated in Fig. 9A. Only single nucleotide variants were considered. It was identified that a total of six interaction sites were revealed for *BRCA2* and one interaction site for *BRCA1*. It was observed that the mutation *BRCA1* $^{p.P824L}$ was an interacting site mutation whereas *BRCA2* $^{p.\ P153Q,\ p.I180F,\ p.D559Y,\ p.G1529R,\ p.L1576F,\ p.E2229K}$ were found to be the interaction site mutations for *BRCA2*. Fig. 9B illustrates the *BRCA1* interacting site mutation (*BRCA1* $^{p.P824L}$) in which Proline is mutated to Leucine. Similarly, Fig. 9C indicates the *BRCA2* interaction site mutations in which Glycine is mutated to Arginine at position 1529 (*BRCA2* $^{p.G1529R}$). Fig. 9D illustrates the interaction site mutation *BRCA2* $^{p.E2229K}$. Illustrations were visualized in PyMOL after downloading the available structures from SWIS-MODEL in pdb format. The inset Fig. 10A-B reveals the mutant proteins of *BRCA1* and *BRCA2* after incorporating all 10 mutations of *BRCA1* and 27 mutations of *BRCA2* on cBioportal (https://www.cbioportal.org/mutation_mapper) and subsequently downloading their pdb structures and visualized using PyMOL. The overall spectrum of mutations is summarized in Tables 1 and 2.

To further investigate the impact of the single nucleotide variations (SNVs) on the protein stability, the SAAFEC-SEQ based online server was used. SAAFEC-SEQ predicts the protein stability and changes in free energy the occurs as a result of SNVs. The results from the SAAFEC-SEQ are summarized in Table 3. All *BRCA1* and *BRCA2* SNVs are shown to have a destabilizing effect on the *BRCA1* and *BRCA1* protein respectively.

The obtained SNVs were further analysed using the ConSurf web-based server. It's proposed that the degree of conservation is linked to the functional and structural importance of the residues. The ConSurf conservation scores (Highest "9″ to lowest "1″) along with the findings of the ConSurf and a screen shot of the mutated position are summarized in Table 4. As can be observed for *BRCA1*, the interacting site mutation i.e. *BRCA1* $^{p.P824L}$ was reported to be highly conserved, buried and of structural nature. Similarly, the mutation *BRCA1* $^{p.S1566G}$ (ConSurf score 8) was found to be conserved and of functional nature. Four *BRCA2* interacting site mutations i.e. *BRCA2* $^{p.P153Q,\ p.I180F,\ p.G1529R,\ p.E2229K}$ were found to be highly conserved (ConSurf score 8 & 9). The details of the prediction and findings are summarized in Table 4.

SNVs can disturb the flexibility of proteins or its small regions which

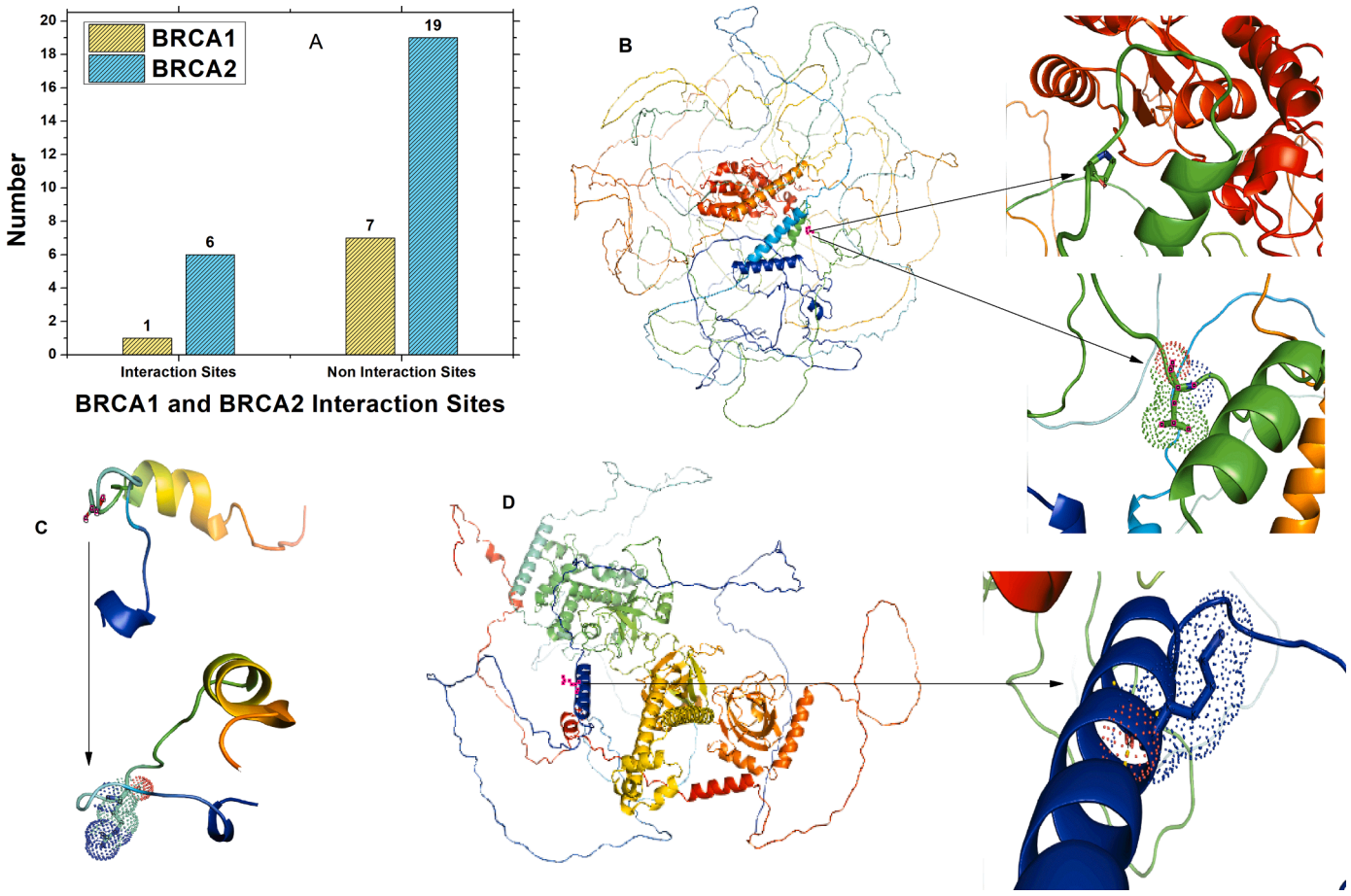


Fig. 9. Summary of the ISPRED-SEQ predictions regarding interacting and non-interacting residues; (A): Number of *BRCA1* and *BRCA2* interacting and non-interacting mutations; (B): Alpha Fold available protein structure of *BRCA1* downloaded from SWIS-MODEL and depiction of interacting mutation; (C/D): Alpha Fold available protein structure of *BRCA2* and illustration of the single nucleotide variations.

may result in shifting of the equilibrium between different conformations, or disturbs the stability and entire conformational dynamics of the protein [21]. Therefore, we analysed selected destabilizing interaction site mutations (*BRCA1*$^{p.P824L}$ and *BRCA2*$^{p.G1529R}$) as confirmed by the ISPRED-SEQ and SAAFEC-SEQ for molecular dynamic measurements as shown in the inset Figs. 11 (A-E) and 12 (A-E). In the inset Figs. 11A and 12A represents the MD simulation measurements related to the radius of gyration (Rg) which is considered important parameter for the folding and unfolding charactersitics of the protein. In both the selected mutations, *BRCA1*$^{p.P824L}$ and *BRCA2*$^{p.G1529R}$, it is observed that the Rg of the mutant protein is relatively higher than the normal protein which indicates that the mutant proteins were disordered or unfolded to an extent as compared to the normal proteins.

The pressure-time graph for *BRCA1* $^{p.P824L}$ and *BRCA2* $^{p.G1529R}$ was plotted against the 100 ps time scale as indicated in Fig. 11B and Fig. 12B respectively. Pressure-time graph revealed an average pressure value of -62 bar and -71 bar for normal and mutant *BRCA1* $^{p.P824L}$. Similarly, for the *BRCA2* $^{p.G1529R}$, the average values of pressure were recorded as -31 bar (normal) and -52 bar (mutant). At the start of the MD trajectory, the pressure of the normal *BRCA1* $^{p.P824L}$ was higher (62 bar) as compared to the mutant (33 bar). The value of the pressure drastically fluctuated during the simulation trajectory. For the *BRCA1* $^{p.P824L}$ mutant, all the pressure values ranged from 90 bar to – 54 bar. For *BRCA1* $^{p.P824L}$ normal, all the maximum and minimum values were 72 bar and -84 bar. The results are indicated in Fig. 11B. In case of the *BRCA2*$^{p.G1529R}$ normal protein, the pressure values ranged between the 395 bar and -340 bar whereas for the mutant protein the pressure values ranged between 348 bar and – 300 bar.

The density (kg/m^3) versus time (ps) obtained for the selected

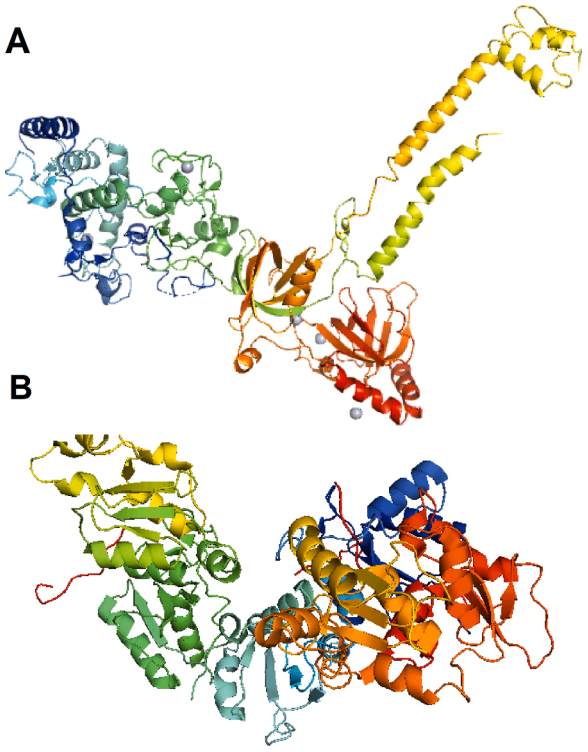


Fig. 10. Molecular modelling analysis of mutant (A) *BRCA1* and (B) *BRCA2* using cBioPortal and subsequent visualization by PyMOL.

Table 1
Mutation spectrum of *BRCA1* gene in patients with breast cancer.

Patients	Mutation Type	Reported	Status	Exon	Mutations		SIFT Prediction	PolyPhen 2 Prediction	Mutation Taster Prediction
19	nonsynonymous SNV	Cosmic	Germline	exon14	*c.A4696G*	*p.S1566G*	D	B	P
19	nonsynonymous SNV	Cosmic	Germline	exon9	*c.A3407G*	*p.K1136R*	T	B	P
19	nonsynonymous SNV	Cosmic	Germline	exon9	*c.A2972G*	*p.E991G*	D	B	P
19	nonsynonymous SNV	Cosmic	Germline	exon9	*c.C2471T*	*p.P824L*	T	B	P
4	nonsynonymous SNV	Cosmic	Germline	exon14	*c.G4815A*	*p.M1605I*	T	B	D
1	frameshift deletion	Novel	Somatic	exon13	*c.4399_4400del*	*p.R1468Afs*1*	.	.	.
1	Stop gain	Novel	Somatic	exon9	*c.3884delC*	*p.S1295fs*0*	.	.	.
1	nonsynonymous SNV	ClinVar	Somatic	exon9	*c.C3712T*	*p.L1238F*	T	B	N
1	nonsynonymous SNV	Novel	Somatic	exon9	*c.G3601T*	*p.A1201S*	D	P	N
1	nonsynonymous SNV	Cosmic	Somatic	exon9	*c.G2883A*	*p.M961I*	T	B	N

Legends: "SIFT": D: Deleterious; T: Tolerated; "PolyPhen-2": D: Probably damaging; P: Possibly damaging; B Benign; "Mutation Taster": A: Disease causing automatic; D: Disease causing; N: Polymorphism; P: Polymorphism automatic

normal and mutated protein sequences of *BRCA1* and *BRCA2* are indicated in Figs. 11C and 12C. The average density values for the wild and mutated *BRCA1* $^{p.P824L}$ were found to be 1006.713 kg/m^3 and 1006.748 kg/m^3 respectively, whereas for *BRCA2* $^{p.G1529R}$ wild and mutated proteins, the average density was calculated as 1003.335 kg/m^3 and 1003.574 kg/m^3 respectively.

The temperature (K) versus time (ps) for the selected mutated and wild proteins is indicated in Figs. 11D and 12D. After plotting the data obtained from simulation, it was revealed that there were fluctuations in temperature for the conformations across 100 ps. The average temperature for wild and mutated *BRCA1* $^{p.P824L}$ was calculated as 300 K and 299 K respectively whereas for *BRCA2* $^{p.G1529R}$, the average values for temperature were obtained as 300.02 K (wild) and 299.95 K for mutated *BRCA2* $^{p.G1529R}$.

Further assessment of the wild and mutated *BRCA1* $^{p.P824L}$ and *BRCA2* $^{p.G1529R}$ was performed using Ramachandran plots obtained from procheck server. For *BRCA1* $^{p.P824L}$ wild protein, the residues in the favorable region were 40.6 % where as 16.2 % and 17 % of the residues

Table 2
Mutation spectrum of *BRCA2* gene in patients with breast cancer.

Patients	Mutation Type	Reported	Status	Exon	Mutations		SIFT Prediction	PolyPhen 2 Prediction	Mutation Taster Prediction
19	nonsynonymous SNV	Cosmic	Germline	exon14	*c.T7397C*	*p.V2466A*	T	.	P
12	nonsynonymous SNV	Cosmic	Germline	exon10	*c.A1114C*	*p.N372H*	T	.	P
5	nonsynonymous SNV	Cosmic	Somatic	exon10	*c.A865C*	*p.N289H*	D	B	P
5	nonsynonymous SNV	Cosmic	Somatic	exon11	*c.A2971G*	*p.N991D*	T	B	P
2	nonsynonymous SNV	Cosmic	Somatic	exon22	*c.G8851A*	*p.A2951T*	D	D	D
1	frameshift deletion	Novel	Somatic	exon27	*c.9649_9650del*	*p.M3217Vfs*3*	.	.	.
1	nonsynonymous SNV	Cosmic	Somatic	exon18	*c.G8187T*	*p.K2729N*	D	D	N
1	nonsynonymous SNV	Cosmic	Somatic	exon11	*c.C5744T*	*p.T1915M*	T	B	N
1	nonsynonymous SNV	Clinvar	Somatic	exon7	*c.A538T*	*p.I180F*	D	D	D
1	nonsynonymous SNV	Clinvar	Somatic	exon23	*c.T9113G*	*p.L3038R*	D	D	D
1	nonsynonymous SNV	Cosmic	Somatic	exon11	*c.C4886A*	*p.T1629K*	D	P	N
1	frameshift deletion	Novel	Somatic	exon11	*c.6123_6124del*	*p.Q2042Kfs*4*	.	.	.
1	nonsynonymous SNV	Cosmic	Somatic	exon11	*c.C3883A*	*p.Q1295K*	D	B	N
1	nonsynonymous SNV	Novel	Somatic	exon11	*c.G3851A*	*p.S1284N*	T	B	N
1	nonsynonymous SNV	Novel	Somatic	exon23	*c.A9017T*	*p.Y3006F*	D	D	D
1	nonsynonymous SNV	Novel	Somatic	exon11	*c.G5856T*	*p.L1952F*	T	B	N
1	nonsynonymous SNV	Cosmic	Somatic	exon17	*c.G7826T*	*p.G2609V*	D	D	D
1	nonsynonymous SNV	Cosmic	Somatic	exon18	*c.G8090A*	*p.S2697N*	T	B	N
1	nonsynonymous SNV	Cosmic	Somatic	exon24	*c.A9253G*	*p.T3085A*	T	B	N
1	nonsynonymous SNV	Cosmic	Somatic	exon11	*c.G4585A*	*p.G1529R*	D	D	D
1	nonsynonymous SNV	Clinvar	Somatic	exon11	*c.G6685A*	*p.E2229K*	D	D	N
1	nonsynonymous SNV	Novel	Somatic	exon5	*c.C458A*	*p.P153Q*	D	D	D
1	nonsynonymous SNV	Cosmic	Somatic	exon11	*c.G4405T*	*p.D1469Y*	D	P	N
1	nonsynonymous SNV	Novel	Somatic	exon27	*c.C9958A*	*p.P3320T*	D	D	N
1	nonsynonymous SNV	Novel	Germline	exon11	*c.C4726T*	*p.L1576F*	D	B	N
1	nonsynonymous SNV	Novel	Somatic	exon10	*c.G1675T*	*p.D559Y*	D	D	N
1	nonsynonymous SNV	Novel	Somatic	exon12	*c.G6937T*	*p.G2313C*	D	D	D

Legends: "SIFT": D: Deleterious; T: Tolerated; "PolyPhen-2": D: Probably damaging; P: Possibly damaging; B Benign; "Mutation Taster": A: Disease causing automatic; D: Disease causing; N: Polymorphism; P: Polymorphism automatic

Table 3
SAAFEC-SEQ calculations for the folding free energy and effect of SNV on protein stability (https://ispredws.biocomp.unibo.it/sequence/).

BRCA1		
Mutation	ddG	Predicted effect
p.S1566G	-0.77	Destabilizing
p.K1136R	-0.56	Destabilizing
p.E991G	-0.83	Destabilizing
p.P824L	-0.68	Destabilizing
p.M1605I	-0.97	Destabilizing
*p.R1468Afs*1*	-	-
*p.S1295fs*0*	-	-
p.L1238F	-1.07	Destabilizing
p.A1201S	-1.30	Destabilizing
p.M961I	-0.83	Destabilizing
BRCA2		
Mutation	ddG	Predicted effect
p.V2466A	-0.75	Destabilizing
p.N372H	-0.64	Destabilizing
p.N289H	-0.58	Destabilizing
p.N991D	-0.47	Destabilizing
p.A2951T	-1.01	Destabilizing
*p.M3217Vfs*3*	-	-
p.K2729N	-0.71	Destabilizing
p.T1915M	-0.62	Destabilizing
p.I180F	-1.07	Destabilizing
p.L3038R	-1.49	Destabilizing
p.T1629K	-0.31	Destabilizing
*p.Q2042Kfs*4*	-	-
p.Q1295K	-0.86	Destabilizing
p.S1284N	-0.38	Destabilizing
p.Y3006F	-1.24	Destabilizing
p.L1952F	-0.27	Destabilizing
p.G2609V	-0.86	Destabilizing
p.S2697N	-0.70	Destabilizing
p.T3085A	-0.63	Destabilizing
p.G1529R	-1.50	Destabilizing
p.E2229K	-0.44	Destabilizing
p.P153Q	-1.01	Destabilizing
p.D1469Y	-0.60	Destabilizing
p.P3320T	-0.90	Destabilizing
p.L1576F	-0.67	Destabilizing
p.D559Y	-0.58	Destabilizing
p.G2313C	-0.35	Destabilizing

were in the allowed and generously allowed regions. In *BRCA1* $^{p.P824L}$ mutant protein, 40.5 % of the residues were in the favorable region. The results are indicated in Fig. 11E and F for wild and mutated *BRCA1* $^{p.P824L}$ respectively. The Ramachandran plot obtained for *BRCA2* $^{p.G1529R}$ wild protein revealed 96.6 % of the residues in the favorable region whereas and for the mutant protein 93.3 % of the residues were in favorable region. Results are indicated in the Fig. 12E and F.

Association of BRCA1 and BRCA2 with socio-genetic and hormonal factors

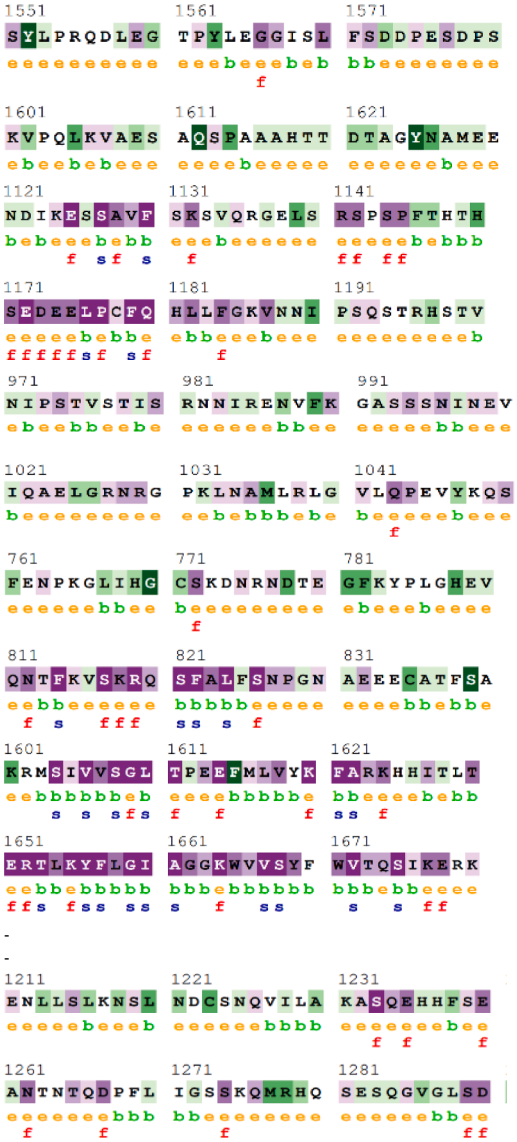
In order to assess the impact of the different sociological and hormonal factors, the mutations were examined in relation to different social and hormonal determinants. The results are indicated in Fig. 13A–D. *BRCA1* $^{p.S1566G,\ p.K1136R,\ p.E991G,\ p.P824L}$ were found in all patients having history of cancer, whereas, 5 other mutations *BRCA1* $^{p.R1468Afs*1,\ p.S1295fs*0,\ p.L1238F,\ p.A1201S,\ p.M961I}$ were not present in patients with cancer history. One mutation *BRCA1* $^{p.M1605I}$ was found to be 20 % prevalent in patients with cancer history and 21.4 % in patients with no cancer history in family. Similarly, in the context of the status of the abortion, the *BRCA1* $^{p.R1468Afs*1}$ was only present in patients with a history of abortion. Furthermore, *BRCA1* $^{p.S1566G,\ p.K1136R,\ p.E991G,\ p.P824L,\ p.M1605I}$ were found in patients both with regular and irregular menstrual cycle. *BRCA1* $^{p.M1605I,\ p.R1468Afs*1,\ p.S1295fs*0,\ p.L1238F,\ p.A1201S,\ p.M961I}$ were not present in the un-married participants. Results are summarized in Fig. 13A–D. The association of *BRCA2* mutations with different sociological and hormonal factors is summarized in Fig. 14A–D. It was observed that (14.81 %; 4/27) mutations were commonly present in patients with or without cancer history, whereas 8/27 mutations (29.62 %) were present specifically in patients with cancer history. Further, 2/27 mutations (7.40 %) were common in married and unmarried participants, and 25/27 mutations i.e. 92.59 % were only present in the married women. The correlation with the abortion status indicated 4/27 (14.81 %) that were found commonly in women with or without abortion history, whereas, 6/27 mutations (22.22 %) were specific to participants with abortion history. The correlation with the regularity of the menstrual cycle indicated that 2/27 (7.40 %) were present in patients with regular or irregular menstrual cycle, whereas, 2/27 (7.40 %) were found to be specific to the irregular menstrual cycle.

Correlation of BRCA1 and BRCA2 with Clinicopathological findings

Results from the clinicopathological findings and their possible association with certain *BRCA1* and *BRCA2* mutations is indicated in Figs. 15A, B and 16A-B respectively. A close look at the mutations reveals that the novel mutations in *BRCA1* i.e. *BRCA1* $^{p.R1468Afs*1,\ p.S1295fs*0,\ p.A1201S}$ were only present in the ER+, PR+, HER2-, Luminal A and Grade 2 patients with breast cancer. Whereas, *BRCA1* $^{p.S1566G,\ p.K1136R,\ p.E991G,\ p.P824L}$ were found commonly in different histopathological parameters. For *BRCA2*, we have obtained 10 novel mutations that are not reported before i.e. *BRCA2* $^{p.M3217Vfs*3,\ p.Q2042Kfs*4,\ p.S1284N,\ p.Y3006F,\ p.L1952F,\ p.P153Q,\ p.P3320T,\ p.L1576F,\ p.D559Y,\ p.G2313C}$. Among these *BRCA2*, 4/10 (40 %) and 6/10 (60 %) were associated with the Grade 2 and Grade 3 breast tumors respectively, and 5/10 (50 %) each were from

Table 4
Summary of the results from the ConSurf webserver (https://consurf.tau.ac.il/consurf_index.php).

BRCA1

Mutation	Conservation Scores (1-9)	Finding	Screenshot from the ConSurf prediction
p.S1566G	8	Highly conserved and exposed residue of functional nature	
p.K1136R	5	Moderately conserved status and exposed residue	
p.E991G	5	Moderately conserved status and exposed residue	
p.P824L	9	Highly conserved and buried residue of structural nature	
p.M1605I	6	Moderately conserved status and buried residue	
*p.R1468Afs*1*	-	-	-
*p.S1295fs*0*	-	-	-
p.L1238F	3	Variable and buried residue	

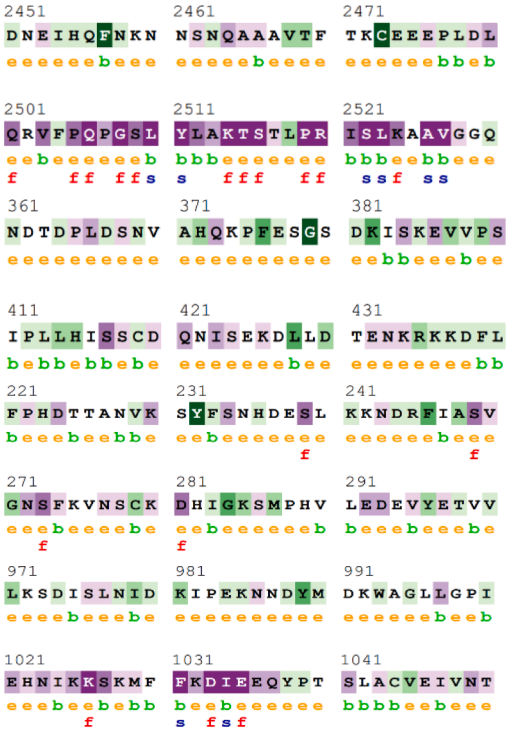
(*continued on next page*)

Table 4 (*continued*)

BRCA1

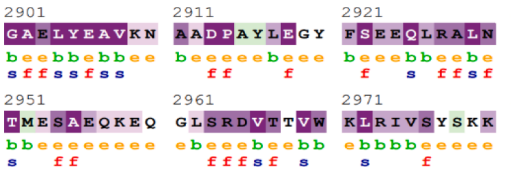
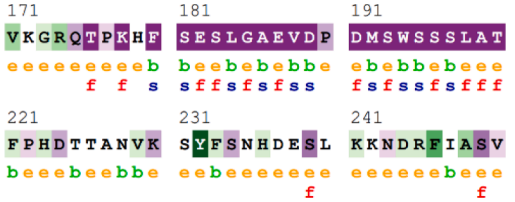
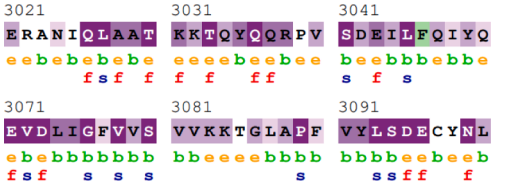
Mutation	Conservation Scores (1-9)	Finding	Screenshot from the ConSurf prediction
p.A1201S	6	Moderately conserved status and buried residue	
p.M961I	1	Highly variable status and exposed residue	

BRCA2

Mutation	Conservation Scores (1-9)	Finding	Screenshot from the ConSurf prediction
p.V2466A	6	Moderately conserved status and buried residue	
p.N372H	2	Variable and exposed residue	
p.N289H	5	Moderately conserved status and exposed residue	
p.N991D	5	Moderately conserved status and exposed residue	

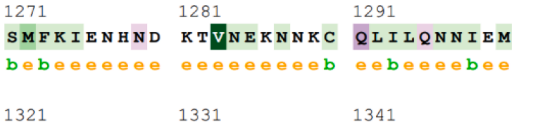
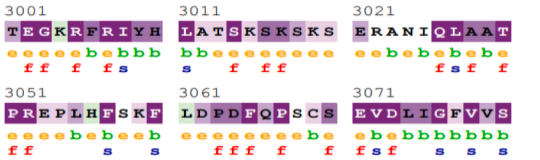
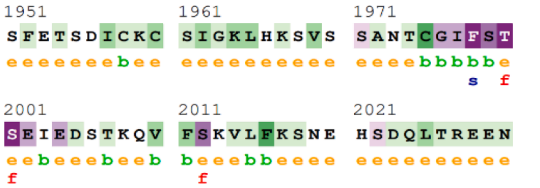
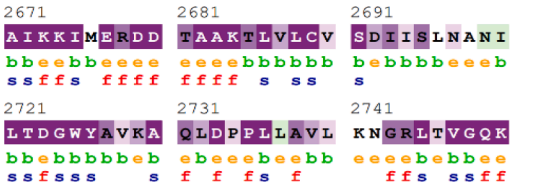
(*continued on next page*)

Table 4 (*continued*)

BRCA1			
Mutation	Conservation Scores (1-9)	Finding	Screenshot from the ConSurf prediction
p.A2951T	9	Highly conserved and buried residue of structural nature	
*p.M3217Vfs*3*	-	-	-
p.K2729N	7	Conserved and exposed residue	
p.T1915M	2	Variable and exposed residue	
p.I180F	9	Highly conserved and buried residue of structural nature	
p.L3038R	8	Conserved and buried residue	
p.T1629K	4	Variable and exposed residue	
*p.Q2042Kfs*4*	-	-	-

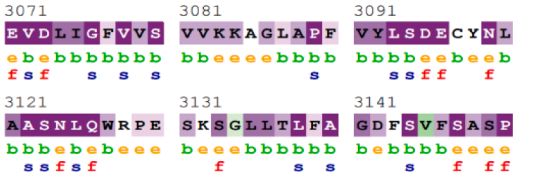
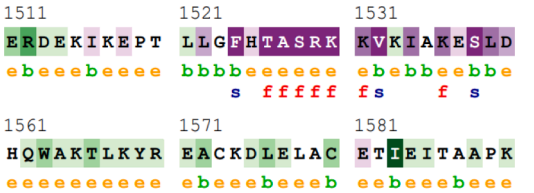
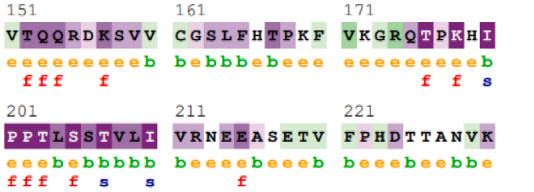
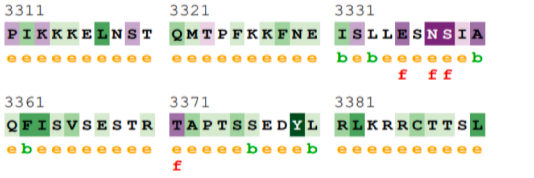
(*continued on next page*)

Table 4 (*continued*)

BRCA1			
Mutation	Conservation Scores (1-9)	Finding	Screenshot from the ConSurf prediction
p.Q1295K	6	Moderately conserved status and exposed residue	
p.S1284N	4	Variable and exposed residue	
p.Y3006F	7	Conserved and buried residue	
p.L1952F	4	Variable and exposed residue	
p.G2609V	9	Highly conserved exposed and functional residue	
p.S2697N	3	Variable and exposed residue	

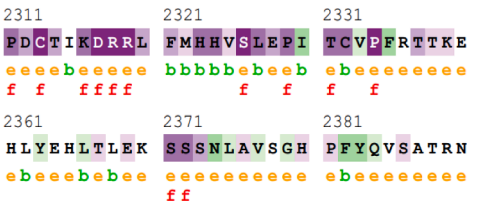
(*continued on next page*)

Table 4 (*continued*)

BRCA1			
Mutation	Conservation Scores (1-9)	Finding	Screenshot from the ConSurf prediction
p.T3085A	5	Moderately conserved and exposed residue	
p.G1529R	9	Highly conserved exposed and functional residue	
p.E2229K	8	Highly conserved exposed and functional residue	
p.P153Q	8	Highly conserved exposed and functional residue	
p.D1469Y	4	Variable and exposed residue	
p.P3320T	5	Moderately conserved and exposed residue	

(*continued on next page*)

Table 4 (*continued*)

BRCA1			
Mutation	Conservation Scores (1-9)	Finding	Screenshot from the ConSurf prediction
p.L1576F	3	Variable and buried residue	
p.D559Y	4	Variable and exposed residue	
p.G2313C	9	Highly conserved exposed and functional residue	

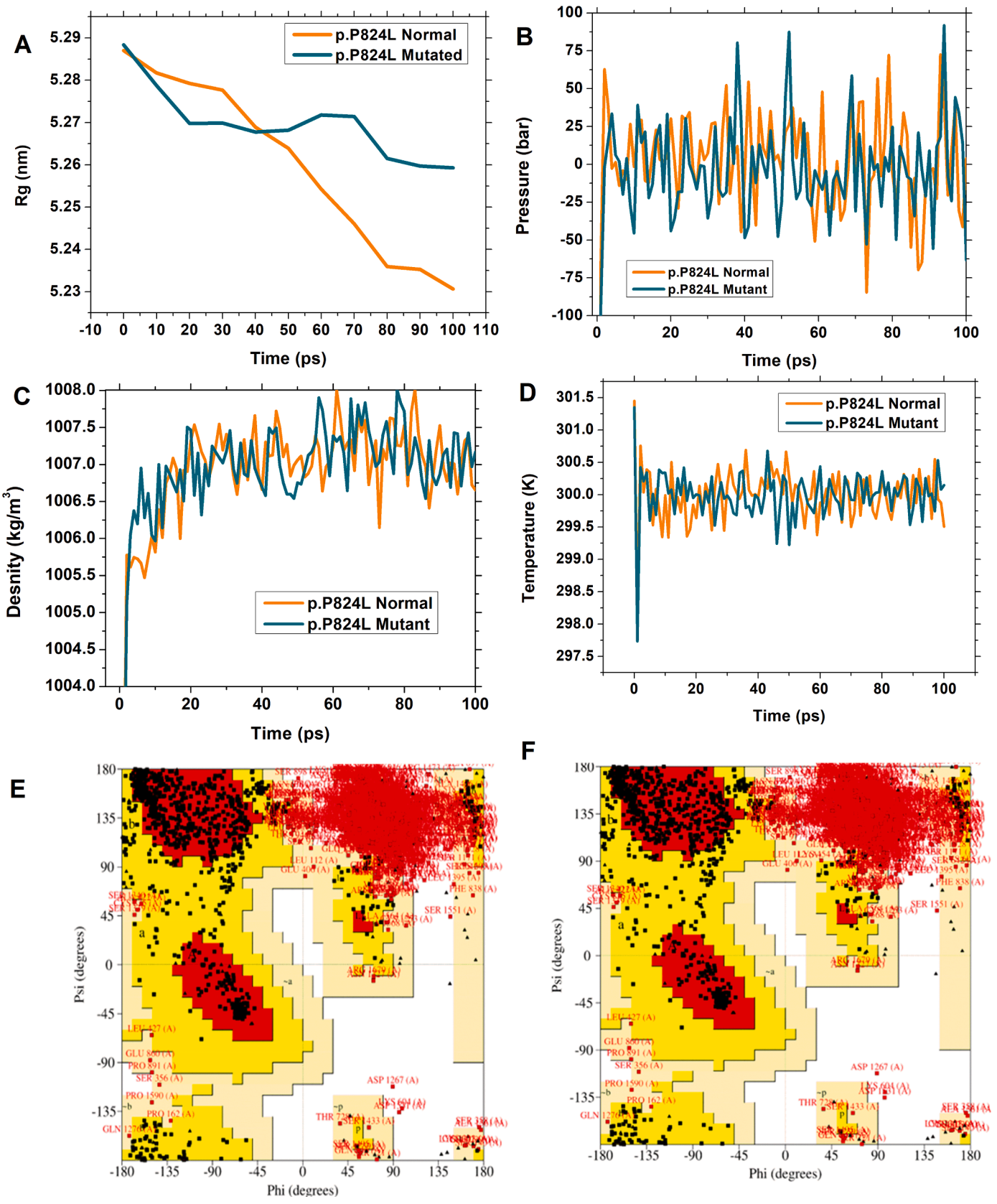


Fig. 11. Comparative molecular dynamic simulation of mutant and normal *BRCA1* $^{p.P824L}$ SNV and their corresponding Ramachandran plots; (A): Radius of gyration; (B): Pressure; (C): Density; (D): Temperature; (E): Ramachandran plot of normal protein; (F): Ramachandran plot of mutant protein.

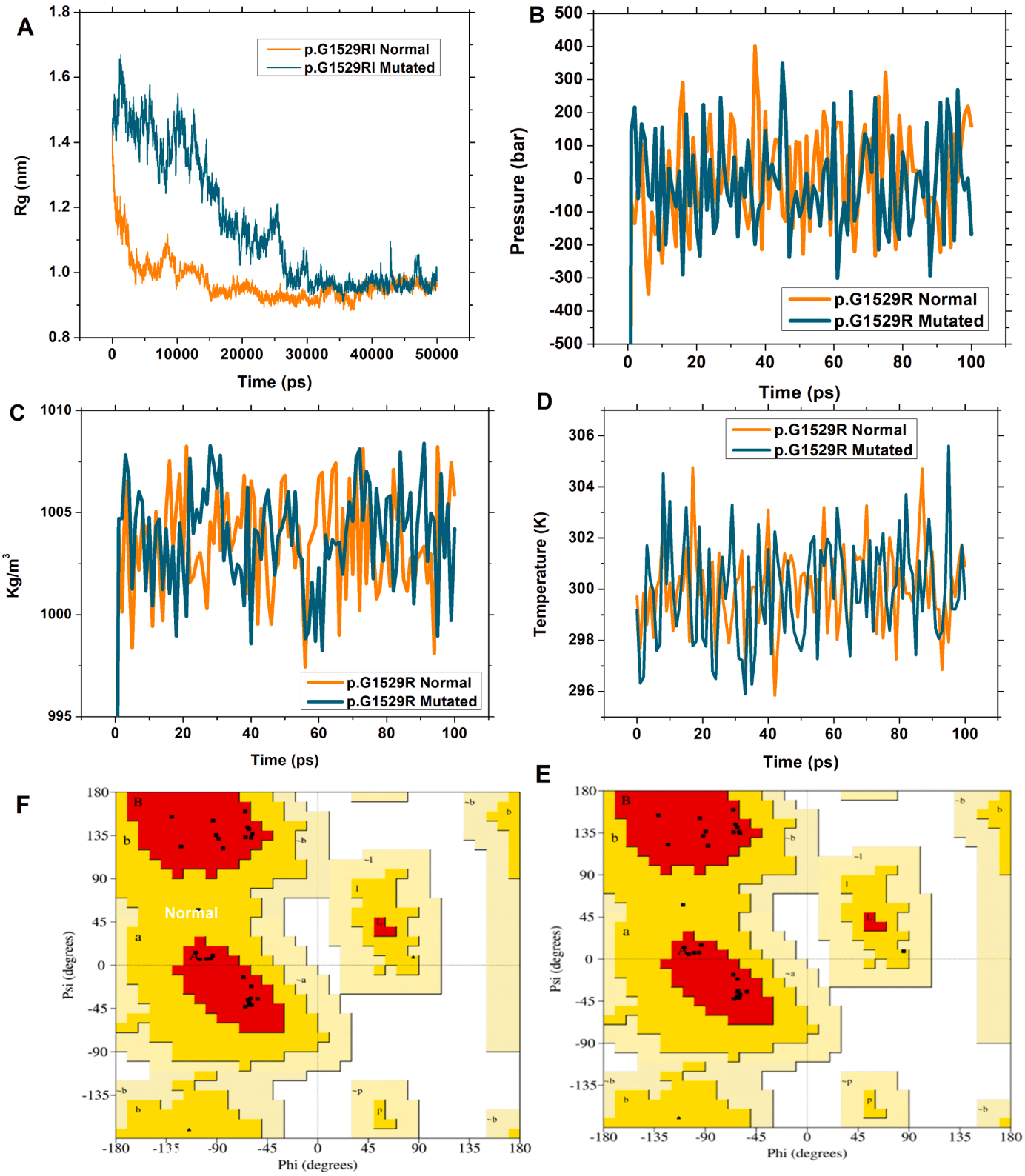


Fig. 12. Comparative molecular dynamic simulation of mutant and normal *BRCA2*$^{p.G1529R}$ SNV and their corresponding Ramachandran plots; (A): Radius of gyration; (B): Pressure; (C): Density; (D): Temperature; (E): Ramachandran plot of normal protein; (F): Ramachandran plot of mutant protein.

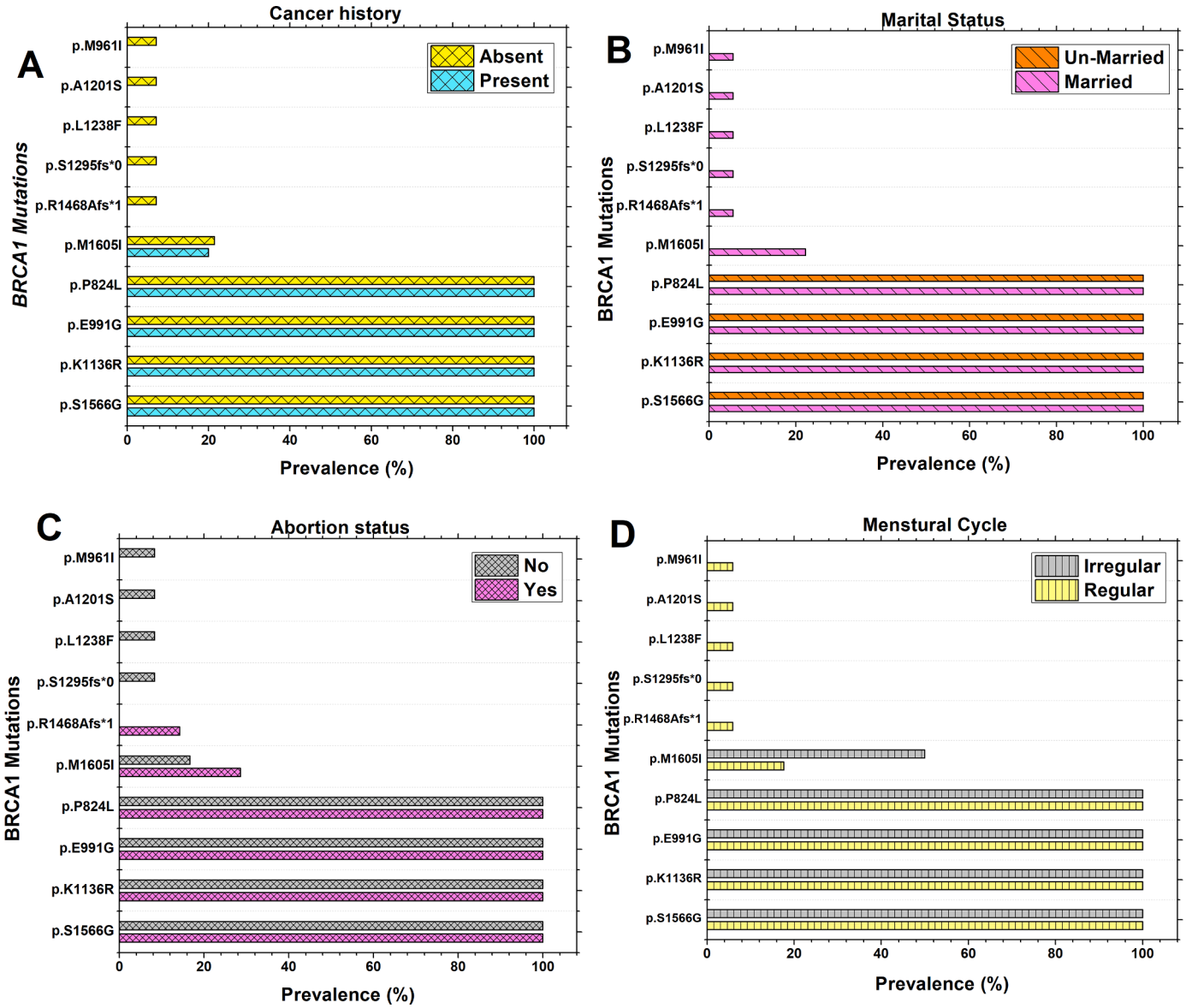


Fig. 13. Association of *BRCA1* mutations with hormonal and non-hormonal factors; (A): Association with (A) Cancer history; (B) Marital status; (C): Abortion status; (D): Menstrual cycle.

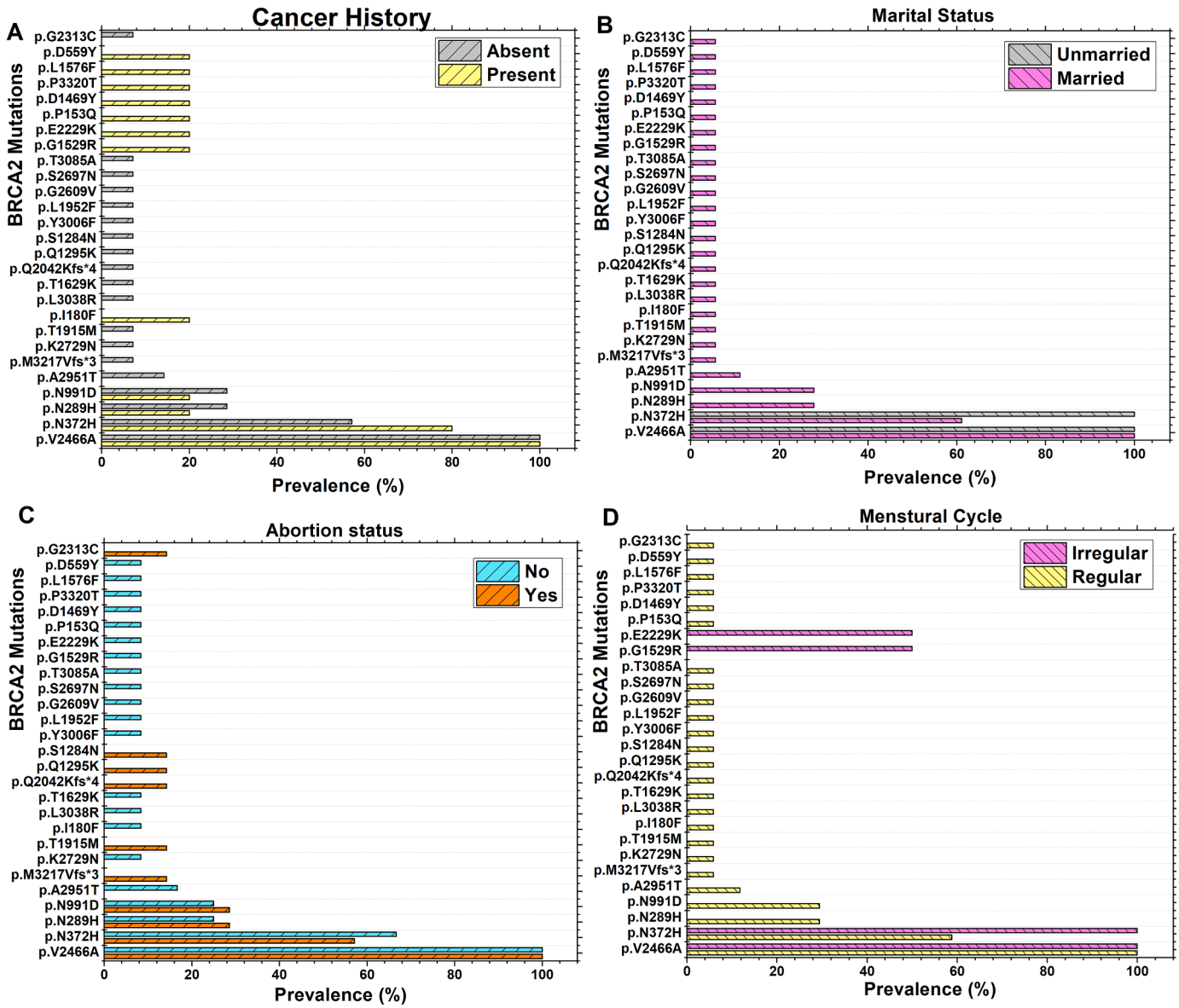


Fig. 14. Association of *BRCA2* mutations with hormonal and non-hormonal factors; (A): Association with (A) Cancer history; (B) Marital status; (C): Abortion status; (D): Menstrual cycle.

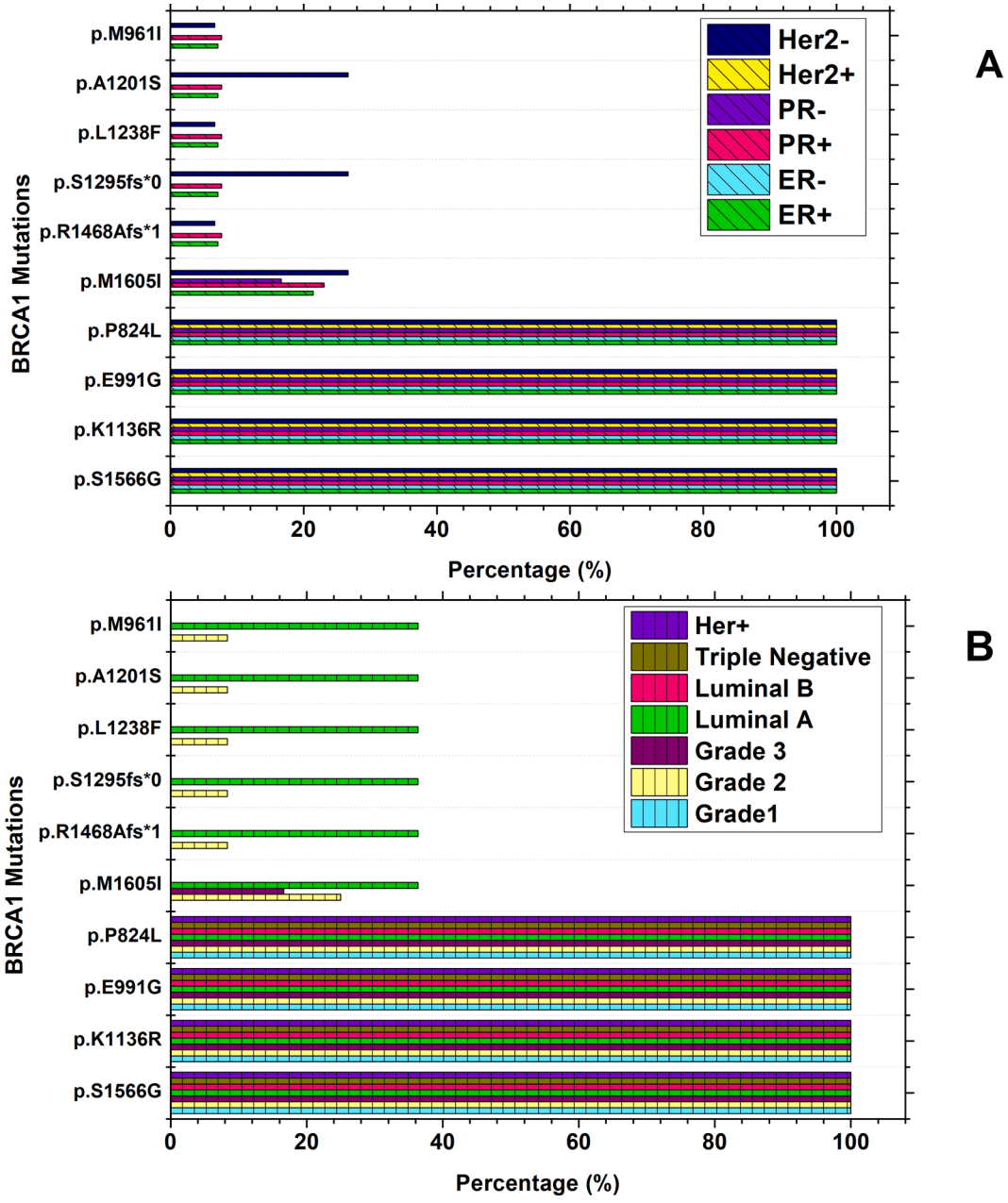


Fig. 15. Clinicopathological association with different *BRCA1* mutations; Association with (A) histochemical markers; (B) Molecular subtypes and tumor grade.

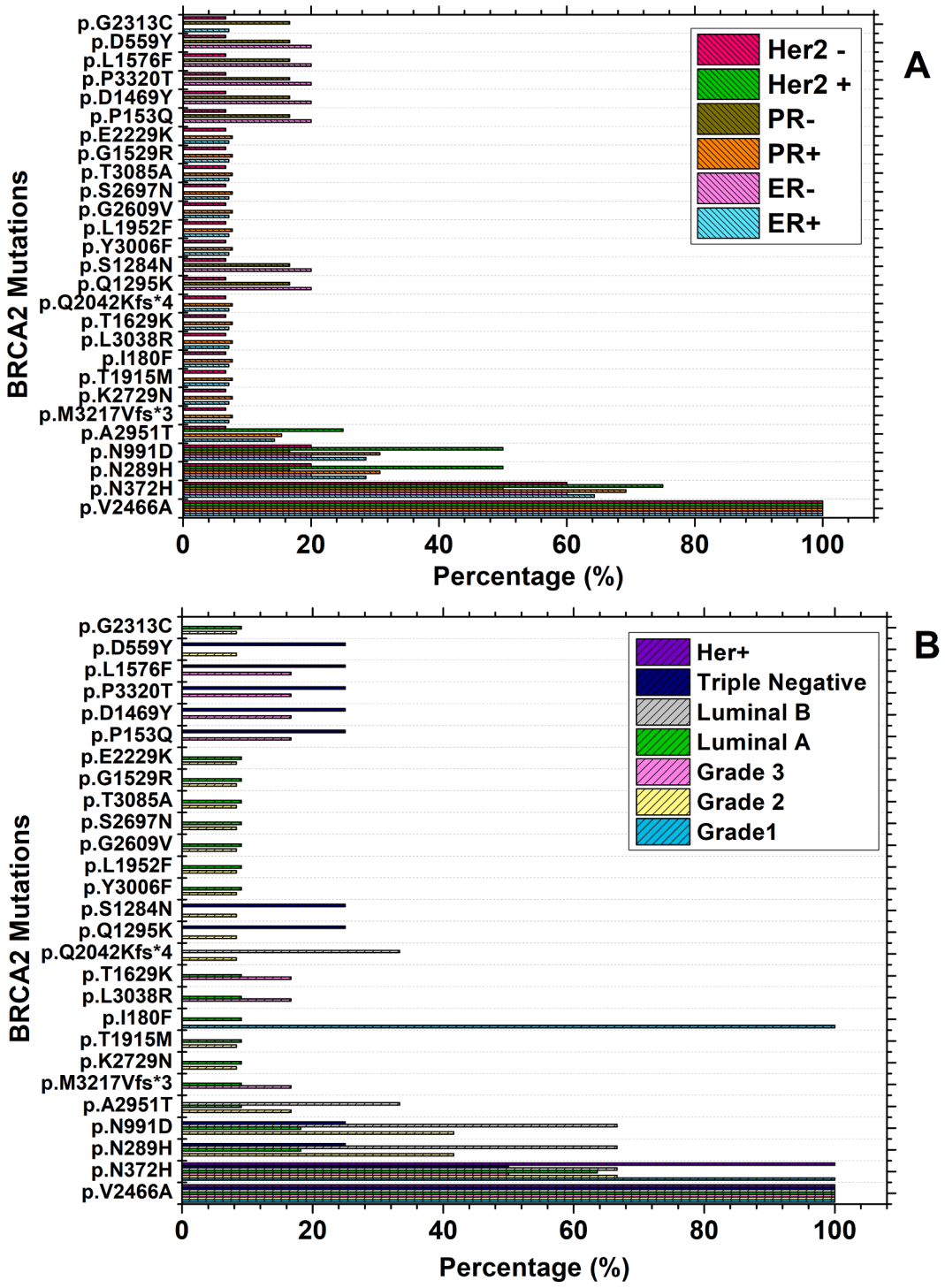


Fig. 16. Clinicopathological association with different *BRCA2* mutations; Association with (A) histochemical markers; (B) Molecular subtypes and tumor grade.

ER+ and ER- tissues. Interestingly, we found out that none of the novel mutations were associated with Her2+ tissues. Only one mutation i.e. *BRCA2* $^{p.Q2042Kfs*4}$ was found to be associated with Luminal B type of breast cancer. Four mutations i.e. *BRCA2* $^{p.M3217Vfs*3,\ p.Y3006F,\ p.L1952F,\ p.G2313C}$ were found to be associated with Luminal A type of cancers. The results are indicated in Fig. 16A, B.

Discussion

The search for genetic biomarkers in breast cancer is one of the areas of considerable research interest. Mutations in the cancer driver genes such as *BRCA1* and *BRCA2* are considered important in the progression of the breast cancer and both of these genes are involved in various DNA repair related pathways. Pathogenic mutations in *BRCA1* and *BRCA2* are known for increasing the risk of different malignancies especially breast and prostate cancers. Our current study is the 1st report of the landscape somatic and germline mutations established from whole exome sequencing in the patients enrolled from Khyber Pakhtunkhwa. Identification of pathogenic mutations in *BRCA1* and *BRCA2* tumor suppresser genes can have significant implications for women at risk and for researchers working in the area of cancer genetics, its management and therapies. It's well established that these tumor suppressors have pivotal functions in repairing the DNA, and mutations in these genes compromises their function which eventually causes cancer. A detailed schematic of the DNA repair related pathways retrieved from the STRING database is indicated in Fig. 17A and C for *BRCA1* and *BRCA2* respectively. The identified reactome pathways were are generated for various genome integrity related functions like sensing double stranded breaks and repair, cell cycle, cell cycle DNA damage check points etc. Similarly, the gene-gene physical interactions for *BRCA1* and *BRCA2* were retrieved from the GeneMania as indicated in Fig. 17C and Fig. 17D respectively. Mutations in *BRCA1* and *BRCA2* can compromise the DNA

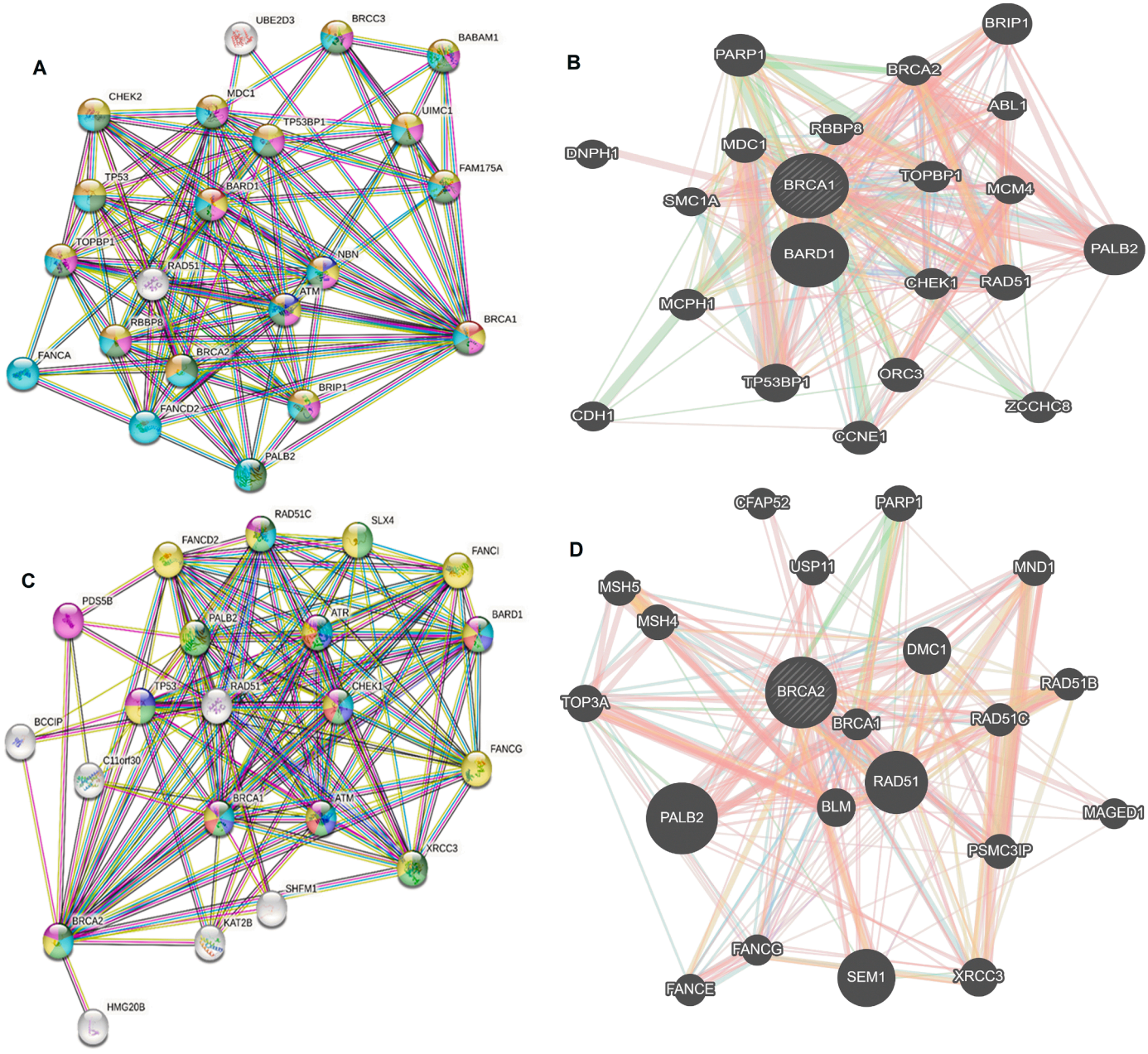


Fig. 17. Various experimentally verified interactions of *BRCA1* and *BRCA2* protein retrieved from STRING and GeneMANIA; (A/B): Various interactions of *BRCA1*; (C/D) Various interactions of *BRCA2*.

repair functions and related pathways eventually increasing the risk of cancer. A total of 10 mutations were identified in *BRCA1* with 08 SNVs and 27 mutations were identified in *BRCA2* with 25 SNVs.

Structurally, the *BRCA1* comprises of 1863 amino acids. The region from 8-96 is the RING finger domain, and from 200-300 is the NLS. The DNA binding domain is located from 452-1079 amino acids [22]. We have identified three mutations in the DNA binding domain i.e. *BRCA1* $^{p.P824L,\ p.M961I,\ p.E991G}$. The DNA binding domain of *BRCA1* has DNA repair activity which is mediated through *BRCA1*-associated surveillance complex (BASC). Another important region in the *BRCA1* is the clusters of serine and threonine sequences which are also referred as known as SQ-cluster domains (SCDs) which are known for *ATM* phosphorylation [22]. We have identified two novel stop gain *BRCA1* $^{p.S1295fs*0}$ and frameshift deletion mutation *BRCA1* $^{p.R1468Afs*1}$ in the SCD domain. The *ATM* protein is involved in repairing of the double stranded DNA [23]. Mutations in the SCD domain may causes inactivation of the *ATM* based repair mechanisms for *BRCA1*.

For *BRCA2*, the structural elements comprise of a relatively less characterized DNA binding domain from 250-500 amino acids [24]. *BRCA2* $^{p.N289H,\ p.N372H}$ mutations were identified in the DNA binding domain. A Total of 9 mutations *BRCA2* $^{p.S1284N,\ p.Q1295K,\ p.D1469Y,\ p.G1529R,\ p.L1576F,\ p.T1629K,\ p.T1915M,\ p.L1952F,\ p.Q2042Kfs*4}$ were identified in the BRC repeats *RAD51* binding domain. Seven mutations *BRCA2* $^{p.G2609V,\ p.S2697N,\ p.K2729N,\ p.A2951T,\ p.Y3006F,\ p.L3038R,\ p.T3085A}$ were identified in the DNA binding domain. *BRCA2* $^{p.P3320T}$ was identified in the NLS (Nuclear Localization Signals) and *RAD51* binding domains. A key role of the *BRCA2* is the repair of the double stranded DNA breaks that leads to gene mutations, chromosomal translocations and deletions which contributes to the cancer progression. The interaction of *BRCA2* and *RAD51* is well established as one of the DNA repair mechanisms. The important region in the *BRCA2* (six BRC repeats), facilitate its interaction with *RAD51*. *RAD51* interaction leads to the highly ordered helical nucleoprotein filaments which mediate the arbitrate the search for homologous sequences on the undamaged sister chromatid, followed by exchange of the DNA strand forming recombination intermediates and covalent linkage [25,26]. *BRCA2-RAD51* complex is therefore important for DNA repair. We obtained 9 mutations in the BRC repeats and propose these mutations as a major driver for cancer progression.

Conclusion

BRCA1 and *BRCA2* genes are well-characterized tumor suppressor genes and the mutations in these two have significant relevance with the progression of breast cancer. The present study was 1st of its kind the included whole exome sequencing of *BRCA1* and *BRCA2* from the FFPE tissue blocks of 19 cancer patients presenting to major tertiary care hospitals from Peshawar, Khyber Pakhtunkhwa. The mutational landscape was identified and corelated with sociogenetic and clinicopathological features. Out of the total spectrum of mutations, 30 % (*BRCA1*) and 37 % (*BRCA2*) mutations were reported novel. For *BRCA1*, 5 somatic and 5 germline mutations were identified, whereas, for *BRCA2*, 24 somatic and 3 germline mutations were identified. The pathogenicity of these mutations was established using different databases. All mutations were predicted to have a destabilizing effect. For *BRCA1*, only one interacting site mutation was identified whereas in *BRCA2*, a total of 6 interacting site mutations were identified. The MD simulations for the selected interacting site mutations confirmed the unfolding of the mutated protein structures. For *BRCA1* $^{p.P824L}$, the average Rg values for normal and mutated proteins were almost similar i.e. 5.260051 nm and 5.26967 nm. However, in case of *BRCA2* significant difference was observed in the average Rg value. The average Rg value for mutated *BRCA2* $^{p.G1529R}$ was found to be 1.14 nm as compared to the 0.97 nm for normal *BRCA2* $^{p.G1529R}$. Preliminary investigation revealed that the one *BRCA1* mutation was present commonly in the patients with abortion history whereas, six *BRCA2* mutations were specific to participants with abortion history. It was also observed that no novel *BRCA2* mutations were present in HER2+ tissues. Overall, we presented preliminary insights on the characterization and correlation of the mutational spectrum of *BRCA1* and *BRCA2* from breast cancer patients from Khyber Pakhtunkhwa which can have valuable implications for the management of the breast cancer patients and designing chemotherapeutic agents.

Our study has few limitations such as limited sample size.

Funding

This research is funded by the Higher Education Commission Pakistan under the National Research Program for Universities vide Project No. 6820/KPK/NRPU/R&D/HEC/2016

Ethical approval

This study has been approved by the research ethics committee of KMU vide. DIR/KMU-EB/VA/000651.

Informed consent statement

Written informed consent was taken from the participants in English and Urdu. The participants were thoroughly explained regarding the aims and objectives of the study.

Data availability statement

The data presented in the study is available in the article.

CRediT authorship contribution statement

Hilal Ahmad: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft. **Asif Ali:** Conceptualization, Funding acquisition, Supervision. **Roshan Ali:** Methodology, Software. **Ali Talha Khalil:** Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Ishaq Khan:** Project administration, Writing – review & editing. **Mah Muneer Khan:** Methodology. **Mohammed Alorini:** Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge the financial support obtained through HEC-NRPU project vide 6820/KPK/NRPU/R&D/HEC/2016. We are also thankful to Dr. Mah Muneer from KTH and Dr. Abid Jamil from HMC for allowing for the collection of the samples.

References

[1] R. Vidra, T.E. Ciuleanu, A. Nemeş, O. Pascu, A.M. Heroiu, N. Antone, et al., Spectrum of BRCA1/2 mutations in Romanian breast and ovarian cancer patients, Int. J. Environ. Res. Public Health 19 (7) (2022) 4314.
[2] Y. Zhang, H. Wu, Z. Yu, L. Li, J. Zhang, X. Liang, et al., Germline variants profiling of BRCA1 and BRCA2 in Chinese Hakka breast and ovarian cancer patients, BMC Cancer 22 (1) (2022) 1–14.
[3] M. Arnold, E. Morgan, H. Rumgay, A. Mafra, D. Singh, M. Laversanne, et al., Current and future burden of breast cancer: global statistics for 2020 and 2040, Breast 66 (2022) 15–23.
[4] Z. Momenimovahed, H. Salehiniya, Epidemiological characteristics of and risk factors for breast cancer in the world, Breast Cancer Targets Ther. (2019) 151–164.
[5] J. Ma, A. Jemal, Breast cancer statistics, Breast Cancer Metastasis Drug Resist. Prog. Prospects (2013) 1–18.

[6] S. Saeed, M. Asim, MM. Sohail, Fears and barriers: problems in breast cancer diagnosis and treatment in Pakistan, BMC Womens Health 21 (1) (2021) 1–10.
[7] I.H. Khaliq, H.Z. Mahmood, M.D. Sarfraz, K.M. Gondal, S. Zaman, Pathways to care for patients in Pakistan experiencing signs or symptoms of breast cancer, Breast 46 (2019) 40–47.
[8] G. Rubi, M. Amir, N. Ashraf, A. Bashir, H. Andaleeb, Breast cancer in Pakistan-an updated overview, Paki. Postgrad. Med. J. 33 (04) (2022) 120–123.
[9] M.R. Sarwar, S. Iftikhar, A. Saqib, Availability of anticancer medicines in public and private sectors, and their affordability by low, middle and high-income class patients in Pakistan, BMC Cancer 18 (1) (2018) 1–11.
[10] F. Gulzar, M.S. Akhtar, R. Sadiq, S. Bashir, S. Jamil, SM. Baig, Identifying the reasons for delayed presentation of Pakistani breast cancer patients at a tertiary care hospital, Cancer Manage Res. 11 (2019) 1087.
[11] P. Economopoulou, G. Dimitriadis, A. Psyrri, Beyond BRCA: new hereditary breast cancer susceptibility genes, Cancer Treat. Rev. 41 (1) (2015) 1–8.
[12] D.G. Evans, A. Shenton, E. Woodward, F. Lalloo, A. Howell, ER. Maher, Penetrance estimates for BRCA1 and BRCA2based on genetic testing in a Clinical Cancer Genetics service setting: Risks of breast/ovarian cancer quoted should reflect the cancer burden in the family, BMC Cancer 8 (1) (2008) 1–9.
[13] J.M. Hall, M.K. Lee, B. Newman, J.E. Morrow, L.A. Anderson, B. Huey, et al., Linkage of early-onset familial breast cancer to chromosome 17q21, Science 250 (4988) (1990) 1684–1689 (1979).
[14] A. Algebaly, R. Suliman, W. Al-Qahtani, Comprehensive study for BRCA1 and BRCA2 entire coding regions in breast cancer, Clin. Transl. Oncol. 23 (2021) 74–81.
[15] A. Mehrgou, M. Akouchekian, The importance of BRCA1 and BRCA2 genes mutations in breast cancer development, Med. J. Islam. Repub. Iran. 30 (2016) 369.
[16] A. Musolino, N. Naldi, M. Michiara, M.A. Bella, P. Zanelli, B. Bortesi, et al., A breast cancer patient from Italy with germline mutations in both the BRCA1 and BRCA2 genes, Breast Cancer Res. Treat. 91 (2005) 203–205.
[17] A. Ali, S. Bell, A. Bilsland, J. Slavin, V. Lynch, M. Elgoweini, et al., Investigating various thresholds as immunohistochemistry cutoffs for observer agreement, Appl. Immunohistochem. Mol. Morphol. 25 (9) (2017) 599–608.
[18] H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows–Wheeler transform, Bioinformatics 26 (5) (2010) 589–595.
[19] M.J. ul Hasnain, F. Amin, A. Ghani, S. Ahmad, Z. Rahman, T. Aslam, et al., Structural and functional impact of damaging nonsynonymous single nucleotide polymorphisms (nsSNPs) on human VPS35 protein using computational approaches, IEEE/ACM Trans. Comput. Biol. Bioinform. 19 (6) (2021) 3715–3724.
[20] S.W. Lim, K.J. Tan, O.M. Azuraidi, M. Sathiya, E.C. Lim, K.S. Lai, et al., Functional and structural analysis of non-synonymous single nucleotide polymorphisms (nsSNPs) in the MYB oncoproteins associated with human cancer, Sci. Rep. 11 (1) (2021) 1–14.
[21] I.N. Smith, S. Thacker, R. Jaini, C. Eng, Dynamics and structural stability effects of germline PTEN mutations associated with cancer versus autism phenotypes, J. Biomol. Struct. Dyn. 37 (7) (2019) 1766–1782.
[22] S.A. Narod, W.D. Foulkes, BRCA1 and BRCA2: 1994 and beyond, Nat. Rev. Cancer 4 (9) (2004) 665–676.
[23] L.S. Stucci, V. Internò, M. Tucci, M. Perrone, F. Mannavola, R. Palmirotta, et al., The ATM Gene in breast cancer: its relevance in clinical practice, Genes 12 (5) (2021) 727 (Basel).
[24] P.R. Andreassen, J. Seo, C. Wiek, H. Hanenberg, Understanding BRCA2 function as a tumor suppressor based on domain-specific activities in DNA damage responses, Genes 12 (7) (2021) 1034 (Basel).
[25] T. Shahid, J. Soroka, E.H. Kong, L. Malivert, M.J. McIlwraith, T. Pape, et al., Structure and mechanism of action of the BRCA2 breast cancer tumor suppressor, Nat. Struct. Mol. Biol. 21 (11) (2014) 962–968.
[26] L. Zheng, S. Li, T.G. Boyer, W.H. Lee, Lessons learned from BRCA1 and BRCA2, Oncogene 19 (53) (2000) 6159–6175.