

IN-SILICO IDENTIFICATION OF nsSNPs IN BRCA1 TO PREDICT BREAST CANCER SUSCEPTIBILITY

Samavia Mustafa¹, Maliha Ghaffar^{*2}, Syeda Ume Farwa³, Anam Abbas⁴,
Atifa Waheed⁵, Sana Khan⁶, Ayesha Mehar⁷

^{1,*2,3,4,5,6,7,8}University of Okara, Faculty of Life Sciences, Department of Biology, Punjab,
Pakistan

¹samimustafa8660@gmail.com, ^{*2}maliha.ghaffar@uo.edu.pk, ³syedafarwa456@gmail.com,
⁴anamabbas473@gmail.com, ⁵atifawaheed2001@gmail.com,
⁶sanamuhmaddkhan6@gmail.com, ⁷ayeshamehar442@gmail.com

DOI: <https://doi.org/10.5281/zenodo.16931002>

Keywords

BRCA1; Breast Cancer;
nsSNPs; In-Silico Analysis;
DNA Repair, Mutation

Article History

Received on 18 May 2025

Accepted on 23 July 2025

Published on 22 August 2025

Copyright @Author

Corresponding Author:

*Maliha Ghaffar

Abstract

Breast cancer is among the most common cancers affecting women globally, with many inherited genetic variations arising from non-synonymous single-nucleotide polymorphisms that alter the function of tumor-suppressor and DNA-repair genes. This study employs an in-silico approach to identify and characterize potentially deleterious nsSNPs in the BRCA1 gene, a key component of the homologous recombination repair pathway. From NCBI dbSNP and UniProt databases using ID: P38398, 44306 BRCA1 nsSNPs were retrieved and 226 were selected for systematic analysis using SIFT, PROVEAN, FATHMM, and PolyPhen-2. Fifty-six variants consistently predicted as deleterious across all tools were further assessed for protein stability changes via SNPs&GO, I-Mutant, and MuPro. We identified 15 high-confidence deleterious BRCA1 nsSNPs which were further subjected to MutPred2 analysis. Several variants exceeded the pathogenicity threshold (>0.75), indicating strong confidence in their deleterious nature. These mutations were predicted to cause critical molecular disruptions, including altered transmembrane regions and DNA binding, loss or gain of post-translational modifications, structural instability, disruption of secondary structure elements (helix, loop, strand), and changes in solvent accessibility. 3D structural modeling with SWISS-MODEL revealed seven mutations M18R, G98D, V11E, L22S, M18T, C39Y, and C61W predicted to significantly destabilize the BRCA1 protein and impair its DNA repair efficiency. These findings provide valuable insights into the structural and functional implications of BRCA1 variants, forming a basis for future experimental validation in hereditary breast cancer risk prediction.

Introduction

Breast cancer remains the leading cancer type among women in 2020, causing about 2.3 million new diagnoses and 685,000 deaths, making it the top cause of cancer-associated mortality. [1, 2]. The *BRCA1* (Breast Cancer Type 1 Susceptibility) was first discovered in 1994 as a key genetic determinant responsible for hereditary forms of breast and ovarian cancers [3]. While environmental and lifestyle factors significantly contribute to sporadic breast cancer cases, the approximate ratio of inherited breast cancers among all is 5–10% [4]. Among the 20 genes studied, *BRCA1*, *BRCA2*, *PALB2*, *TP53*, *RAD51D*, *CDH1*, *PTEN*, *ATM*, *CHEK2*, *BARD1*, *STK11*, and *NBN* are well-known for their links to breast cancer risk. Some of these, like *BRCA1*, *BRCA2*, *TP53*, *CDH1*, *PTEN*, *STK11*, and *PALB2*, are associated with rare syndromes that strongly increase cancer risk [5]. Among the genes implicated in hereditary breast cancer, *BRCA1* is recognized as a high-penetrance tumor suppressor gene that plays a central role in preserving genomic integrity through DNA repair, cell cycle regulation, and transcriptional control. The *BRCA1* gene is situated on chromosome 17q21, covering roughly 81 kilobases, and produces a protein consisting of 1,863 amino acids. This protein is integral to the homologous recombination repair (HRR) pathway, which accurately repairs double-stranded DNA breaks. The N-terminal RING (Really Interesting New Gene) finger domain (residues 24–65) is one of several important domains of the *BRCA1* protein. While on the other hand a coiled-coil domain (residues 1364–1437), and the BRCT (*BRCA1* C Terminus) domains are also crucial for protein-protein interactions and regulation of the DNA damage response [6-8]. *BRCA1*, crucial in maintaining genetic integrity, is quickly recruited to the damage sites, where it forms complexes with several proteins and

facilitates repair mainly through the HR and Non-Homologous End Joining (NHEJ) pathways [9, 10]. In response to genotoxic stress, *BRCA1* is essential for controlling the cell cycle by activating key checkpoints at the G1/S, S, and G2/M phases [8, 11].

Mutations in the *BRCA1* gene, particularly nsSNPs (Nonsynonymous Single Nucleotide Polymorphisms), can lead to the production of structurally and functionally altered proteins, impairing the gene's role in DNA repair and increasing cancer susceptibility. Over 3,000 variants of *BRCA1* have been identified by scientists since its discovery [12]. These nsSNPs result in amino acid substitutions, which may disrupt the protein's folding, stability, and interactions with other critical repair proteins such as *BARD1* and *RAD51* [13]. Several *BRCA1* nsSNPs located in the RING finger, BRCT domains, and regulatory regions have been linked to increased breast cancer risk. C61G variant in the RING finger domain reduces both E3 ligase activity and *BARD1* binding, impairing *BRCA1*'s tumor-suppressive function [14].

Pathogenic *BRCA1* variants have been shown to increase the lifetime risk for a woman of developing breast cancer by up to 65 percent, highlighting the importance of identifying and characterizing such mutations [15]. Although *BRCA1* mutations have important clinical consequences, certain detected variants are still categorized as Variants of Uncertain Significance (VUS), posing challenges in risk assessment and clinical decision-making. Experimental methods to study VUS are expensive and time-consuming. Using bioinformatics and in-silico approaches, it becomes possible to prioritize high-risk nsSNPs for further experimental validation and clinical evaluation. Combining predictions from multiple tools can improve accuracy in identifying pathogenic mutations, making these methods useful for disease

understanding and therapeutic development [16].

This study focuses on the in-silico identification and analysis of high-risk nsSNPs in the BRCA1 gene to predict their potential association with breast cancer susceptibility. Using an integrated pipeline involving multiple predictive tools like SIFT (Sorting Intolerant From Tolerant), PolyPhen-2 (Polymorphism Phenotyping 2), PROVEAN (Protein Variation Effect Analyzer), FATHMM (Functional Analysis through Hidden Markov Models), SNPs&GO (Single Nucleotide Polymorphisms and Gene Ontology), I-Mutant, MuPro (Mutation Protein Stability Prediction), and MutPred2 (Mutation Prediction 2) along with structural modeling and validation techniques like SWISS-MODEL, ProSA (Protein Structure Analysis), QMEAN (Qualitative Model Energy Analysis) and PROCHECK (Program to check the stereochemical quality of protein structures), we aim to identify nsSNPs that are likely to be deleterious and are more likely to affect protein function and contribute to breast cancer susceptibility. This bioinformatics-driven strategy not only streamlines the identification of pathogenic variants but also provides a foundation for future experimental validation and personalized risk assessment in hereditary breast cancer.

Materials and Methods

Extraction of BRCA1 sequence and nsSNPs from databases

nsSNPs in BRCA1 were found and assessed using in-silico analysis. The gene sequence and SNP data were obtained from NCBI's dbSNP (<https://www.ncbi.nlm.nih.gov/snp>) and UniProt databases (<https://www.uniprot.org>). A total of 44,306 SNPs associated with the human BRCA1 gene (UniProt ID: P38398) were retrieved from the NCBI (National Center for Biotechnology Information) dbSNP database using the keyword "BRCA1". Additionally, 40 to 50 SNPs were obtained through FASTA sequence alignment using the BRCA1 reference sequence. The reference protein sequence used for structural and functional analysis was NP_001394512.1 from NCBI. From the complete list of reported SNPs, only nsSNPs, which result in an amino acid change, were subjected to further analysis, due to their possible effect on the structure and function of protein. These particular variants were chosen for in-silico analysis on the based on their conflicting reports regarding their clinical significance, previously unreported or with uncertain significance on ClinVar (Clinical Genome Variation) (<https://www.ncbi.nlm.nih.gov>).

Table 1: Accession Number and Reference Sequence of Human BRCA1

Databases	Human BRCA1 Gene
UniProt Database	UniProt Accession Number P38398
NCBI Database	Protein Reference Sequence NP_001394512.1
Protein Length	1863 Amino Acids

Functional impact of pathogenic nsSNPs

Multiple prediction tools were used to evaluate the functional consequences of the selected nsSNPs. These included SIFT (Sorting Intolerant From Tolerant) (<https://sift.bii.a-star.edu.sg>), to check the impact of amino acid-based substitution on sequence homology and conservation. 226 BRCA1 nsSNPs were submitted one by one to SIFT which categorizes each substitution as "damaging" or "tolerated/benign" depending on the tolerance index score. The values < 0.05 indicating a damaging effect likely to affect protein function, and scores > 0.05 suggesting a tolerated substitution with minimal functional consequence [17].

Other tools used were PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2>), which assesses the structural and functional impact of mutations [18]; PROVEAN (<http://provean.jcvi.org/index.php>), which estimates whether a mutation is deleterious or neutral [19]; and FATHMM (<http://fathmm.biocompute.org.uk>), which uses hidden Markov models to evaluate the potential disease-association of variants [20].

Prediction of disease-associated nsSNPs and effect on protein stability

About 56 variants that were consistently predicted as deleterious or damaging by all four tools were shortlisted for subsequent detailed analyses. To further analyze the clinical significance of the variants, SNPs&GO (<https://snps-and-go.biocomp.unibo.it>) was used to predict disease relevance by integrating sequence and gene ontology data. The impact of mutations on the stability of protein was analyzed using I-Mutant 2.0 (<https://folding.biofold.org>) and MuPro (<http://mupro.proteomics.ics.uci.edu>), both of which predict changes in protein

stability resulting from point mutations [21, 22].

I-Mutant assesses how nsSNPs influence the stability of proteins like BRCA1, offering $\Delta\Delta G$ values, the direction of stability change, and optional parameters such as Reliability Index (RI) and Solvent Accessibility (RSA). The tool achieves a high prediction accuracy of 80% with structure data and 77% with sequence data, making it effective even when 3D models are unavailable. In addition, MutPred2 (<http://mutpred.mutdb.org>) was used to predict molecular and structural mechanisms potentially disrupted by the nsSNPs, such as loss of catalytic residues, altered secondary structure, or disrupted protein-protein interactions. Mutations are categorized as "actionable hypotheses" if their general score or g-value is larger than 0.5 and their p-value is less than 0.05, and as "confident hypotheses" if their g-value is greater than 0.75, the p-value is less than 0.05 [23, 24].

Protein structural modeling

High-risk nsSNPs identified through functional and stability predictions were subjected to 3D structural modeling using SWISS-MODEL (<https://swissmodel.expasy.org/qmean>). The generated native and mutant protein models were then validated using QMEAN (<https://swissmodel.expasy.org/qmean>), ProSA (<https://prosa.services.came.sbg.ac.at/prosa.php>), and Ramachandran plot analyses through the SAVES v6.0 server (<https://saves.mbi.ucla.edu>), ensuring the accuracy and reliability of the structural predictions. Structural differences between native and mutant forms were further examined using TM-align (<https://zhanggroup.org>) for superimposition and structural deviation analysis. Mutation3D was used for

structural clustering and localization of pathogenic variants. The stereochemical quality of the generated 3D models was assessed using several validation tools like PROCHECK

(<https://saves.mbi.ucla.edu>) for geometry of the models, ERRAT (Error Rate Analysis Tool)

(<http://missense3d.bc.ic.ac.uk/missense3d>) for non-bonded interactions; VERIFY 3D (<https://saves.mbi.ucla.edu>)

for compatibility of the models with known 3D structures.

This integrated computational approach enabled the prioritization of potentially deleterious nsSNPs in *BRCA1*, providing insights into their likely structural and functional impacts, and contributing to our understanding of their role in breast cancer susceptibility. Figure 1 illustrated the methodology we used in the study.

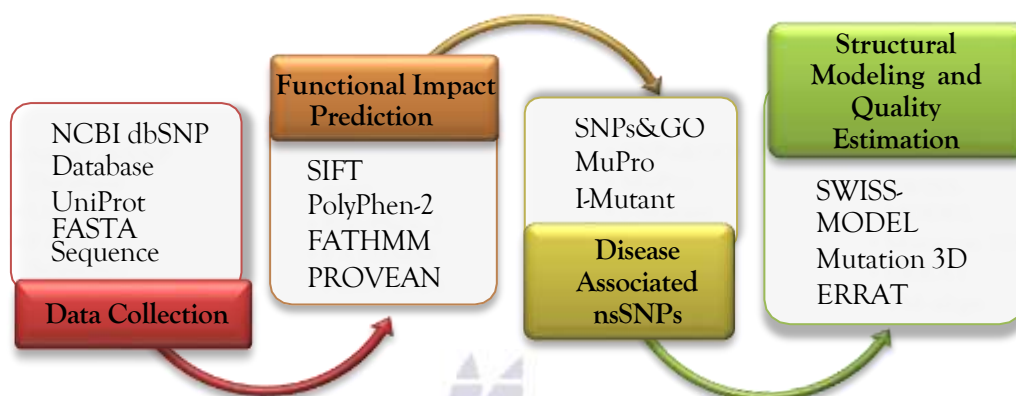


Figure 1: Tools used to predict nsSNPs in BRCA1

Results and Discussions

The NCBI dbSNP and UniProt databases yielded a total of 44,306 SNPs for the human *BRCA1* gene using straightforward keyword searches such as "BRCA1".

Alongside 40 to 50 nsSNPs were identified through sequence alignment. Protein sequences for *BRCA1* were retrieved from UniProt using ID P38398. From these, 226 *BRCA1* nsSNPs were selected for in-silico analysis due to their conflicting or uncertain clinical significance. The 226 *BRCA1* nsSNPs were analyzed, and the results varied across these four tools, with PROVEAN identifying 72 deleterious variants, SIFT 176 non-tolerated, PolyPhen-2 114 likely damaging, and FATHMM 147 damaging nsSNPs, highlighting significant differences in predicted functional impacts. On the other hand, PROVEAN

predicted 154 SNPs as neutral, SIFT marked 50 as tolerated, PolyPhen-2 identified 112 as benign or possibly benign, and FATHMM categorized 79 as tolerated. These widely used computational tools such as SIFT [25], PolyPhen-2 [18], FATHMM [20], and PROVEAN [19] were applied to evaluate the potential functional impact of nsSNPs.

The tools used include SIFT, PolyPhen-2, FATHMM, and PROVEAN. SIFT predicts whether an amino acid substitution affects protein function, with a score ≤ 0.05 considered deleterious. PolyPhen-2 evaluates the possible impact of an amino acid substitution on the structure and function of a protein, where a score > 0.85 indicates a probably damaging mutation. PROVEAN assesses the

biological impact of amino acid substitutions, with scores ≤ -2.5 considered deleterious. According to FATHMM, variants with scores below -1.5 are considered damaging and potentially pathogenic, while those above -1.5 are interpreted as tolerated or functionally neutral.

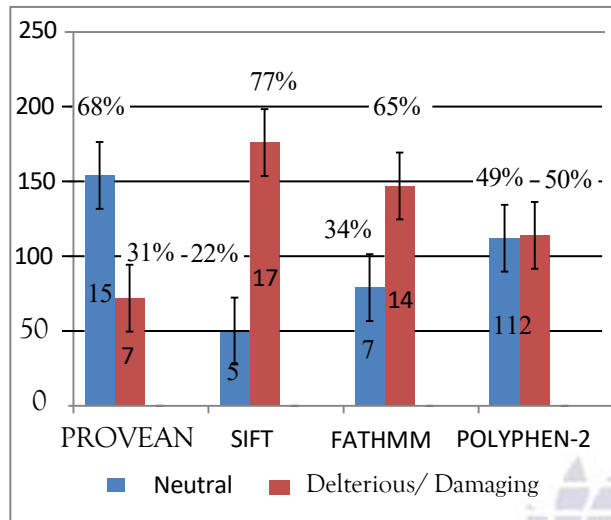


Figure 1: Comparative pathogenicity analysis of functional impact predictions of nsSNPs in BRCA1 using four computational tools.

nsSNPs have gained considerable attention in recent years due to their ability to alter the amino acid sequence of proteins during translation, potentially leading to significant structural and functional changes. According to a study conducted by Arshad et al., [26] 122 nsSNPs were identified and subjected to SIFT and PROVEAN. Out of which 61 nsSNPs were predicted as deleterious. In our study in order to get highly precised results we used a variety of tools to analyze nsSNPs and considered only those nsSNPs which were identified as damaging and non- tolerated by all the applied tools. In contrast, Yadegari and Majidzadeh (2019) performed a large-scale combined analysis on BRCA1 and BRCA2 missense mutations. While their dataset was broader in scope, covering a larger number of variants, our work focused on a targeted set of BRCA1

nsSNPs, allowing for a more detailed structural and functional characterization, including protein stability changes and specific molecular disruptions [16].

Prediction of the effect of disease-associated nsSNPs on protein stability

After analyzing 226 nsSNPs, variants consistently predicted as deleterious or damaging by all four tools were selected for further study, and only 56 high-confidence deleterious variants were selected based on agreement across four prediction tools. All 56 were predicted as disease-associated by SNPs&Go. Most nsSNPs (55 by MuPro and 51 by I-Mutant) were detected to decrease protein stability, indicating likely harmful effects on BRCA1 structure and function that may contribute to breast cancer risk. SNPs&GO classified all variants as disease-associated, with very high-Reliability Index (RI) scores (mostly 9–10), indicating strong confidence in the predictions. Both MuPro and I-Mutant tools predicted a decrease in protein stability for most variants, with MuPro $\Delta\Delta G$ scores mostly negative, I-Mutant supporting these findings with high RI values. A few exceptions, S578F, N473T, G1087A, and K45E showed increased stability, suggesting possible context-specific roles or benign effects. According to Stehr et al. (2011) [27] tumor suppressor mutations predominantly cause protein destabilization and core residue disruption, supporting a loss-of-function mechanism. SNPs&GO, which combines sequence information, evolutionary data, and functional annotations from Gene Ontology, assigns a probability score indicating the likelihood that a variant is disease-related. In this study, only variants with a probability score >0.5 were considered likely pathogenic [16].

Functional Impact Prediction of High-Risk nsSNPs of BRCA1

MutPred2 analysis of BRCA1 nsSNPs revealed several mutations likely to affect protein function by altering stability, structure, or key sites. These findings,

supported by SNPs&Go, MuPro, and I-Mutant, suggest potential impairment of BRCA1's DNA repair role, contributing to breast cancer risk.

Table 2: MutPred2 Results Prediction

BRCA1					
Sr. No	SNP ID	Mutation	MutPred2 Score	Actionable Hypothesis	P Value
1.	rs62625300	F461L	0.539	Altered Ordered Interface	0.03
				Altered Disordered Interface	0.025
				Loss of Acetylation at K459	5.7e-03
				Gain of Relative Solvent Accessibility	0.03
				Altered DNA Binding	9.0e-03
				Loss of Allosteric Site at Y465	0.04
				Loss of Methylation at K459	6.1e-03
				Altered Metal Binding	0.03
2.	rs62625301	T464I	0.541	Altered Disordered Interface	5.1e-03
				Altered Ordered Interface	5.0e-03

				Altered DNA Binding	9.2e-04
				Gain of Relative Solvent Accessibility	0.02
				Loss of Acetylation at K459	0.01
				Loss of Allosteric Site at K467	0.04
				Gain of Methylation at K459	3.6e-03
3.	rs80357027	V1176D	0.587	Altered Disordered Interface	0.03
				Loss of Methylation at K1179	4.9e-03
				Gain of Acetylation at K1179	0.04
				Gain of Ubiquitylation at K1171	0.01
4.	rs80357098	F861C	0.674	Altered Stability	7.2e-03
				Loss of Sulfation at Y856	0.03
5.	rs56100707	P514R	0.620	Altered DNA Binding	1.5e-03
				Gain of Helix	0.02
				Loss of Loop	9.0e-03
				Altered Disordered Interface	0.04
				Gain of Acetylation at K519	8.0e-03
6.	rs80357438	L22S	0.834	Loss of SUMOylation at K519	0.03
				Altered Disordered Interface	0.04
				Altered Metal Binding	0.04
7.	rs80356929	M18R	0.902	Altered Transmembrane Protein	7.1e-03
				Altered Transmembrane Protein	0.01
8.	rs80356860	G1706E	0.767	Altered Stability	0.04
				Gain of Helix	0.01

				Altered Ordered Interface	0.01
				Altered Transmembrane Protein	3.5e-03
9.	rs397509238	E1735K	0.783	Gain of Acetylation at E1735	4.7e-03
				Altered Metal Binding	7.3e-03
				Gain of Allosteric site at R1737	0.03
				Loss of Proteolytic Cleavage at D1739	3.7e-03
				Gain of Ubiquitylation at E1735	0.03
				Altered Stability	0.03
10.	*	M18T	0.876	Altered Transmembrane Protein	0.01
				Altered Stability	0.04
				Gain of N-Linked Glycosylation at N16	0.03
11.	*	H41N	0.634	Altered Metal Binding	0.01
				Altered Transmembrane Protein	0.03
12.	*	C39Y		Altered Disordered interface	5.9e-03
				Altered Metal Binding	0.01
				Altered Transmembrane Protein	0.02
13.	*	C61W	0.820	Loss of Intrinsic Disorder	0.04
				Altered Disordered Interface	0.02
				Altered Metal Binding	0.03
				Loss of Disulfide Linkage at C61	0.03
				Altered Transmembrane Protein	0.03
				Alidone Carboxylic Acid at Q60	0.04
14.	*	G98D	0.604	Altered Disordered Interface	0.04
				Gain of Helix	0.04
				Gain of GPI-anchor amidation at N103	0.02
15.	rs80357017	V11E	0.880	Loss of Relative Solvent Accessibility	9.7e-03
				Gain of Helix	0.02
				Loss of Strand	0.01

MutPred2 delivers a probability score reflecting how likely the mutation is harmful, along with detailed possible functional consequences to help understand its biological impact. It provides specific predicted molecular alterations caused by the mutation, such as loss/gain of phosphorylation, altered stability, or disrupted binding sites, with associated confidence (p-values) [23].

3D Structure Formation using SWISS-MODEL

MutPred2 analysis identified high-risk BRCA1 mutations M18R, V11E, M18T, and C61W with scores >0.8, indicating

strong pathogenic potential. These mutations affect protein stability, DNA binding, and structural features like helices and disulfide bonds. The most significant alteration, T464I, had a p-value of 0.00092, suggesting a strong functional impact possibly affecting phosphorylation sites and protein folding dynamics. Swiss-Model generates BRCA1's 3D structure using template P38398 (amino acids 1-1863), with a GMQE score of 0.42. The model was validated using a Ramachandran plot and a QMEAN Z-score within an acceptable range, indicating moderate quality.

The stereochemical properties of the predicted model were validated using a Ramachandran plot, which showed over 90% of residues in favored regions. Additional validation tools, such as ProSA and ERRAT confirmed the structural integrity of the model. Structural deviation analysis using TM-align revealed significant RMSD values between native and mutant models, particularly for mutations located in the

BRCT domain (amino acids ~1650–1863); known to be involved in phosphopeptide recognition and DNA repair complex formation. Arshad et al. (2021), who used SWISS-MODEL to evaluate nsSNP-driven structural perturbations in BRCA1 and identified C39, H41, L22, C44, C47, and I1707 as critical residues, with mutations at these sites leading to domain-specific structural disruptions [26].

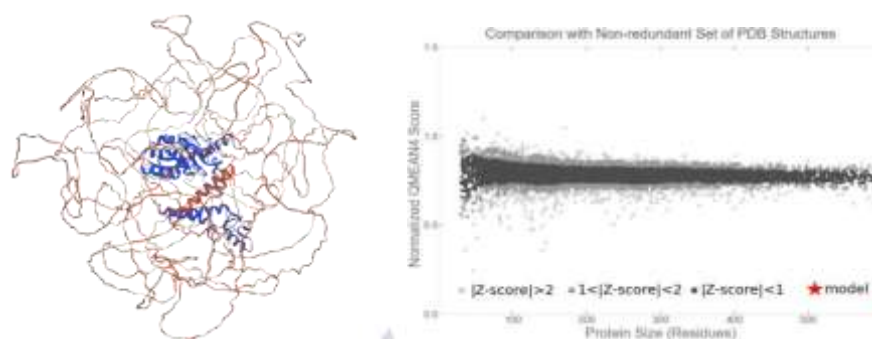


Figure 3: 3D Structure of BRCA1 Protein using SWISS-MODEL

Ramachandran plot analysis using PROCHECK confirmed the stereochemical quality of the refined BRCA1 model. Validation via ProSA-

web gave a Z-score of -4.41, and QMEAN evaluation yielded a score of 0.50, indicating acceptable model quality.

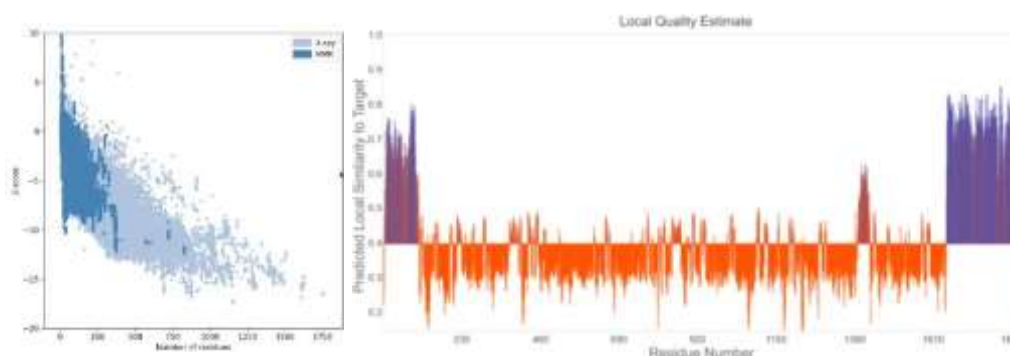


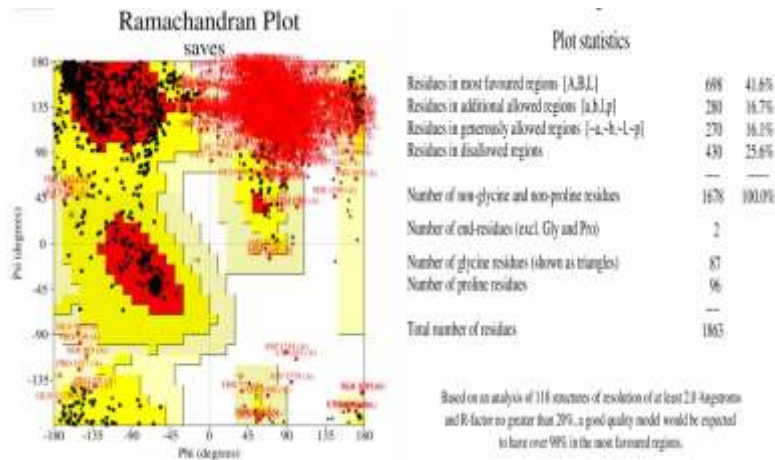
Figure 4: Quality of predicted model by ProSA and QMEAN

The Ramachandran plot analysis of the BRCA1 protein model, constructed using the template P38398, revealed that 41.6% of the amino acid sequence residues are located in the most energetically preferred

regions, and represent conformational angles typically found in stable protein structures. Additionally, 16.7% of the residues fall within allowed regions, which are less ideal but still acceptable

conformations. However, about 25.6% of residues were found in disallowed regions, suggesting that these parts of the model may have strained or energetically unfavorable backbone angles. This distribution indicates that while the overall

structural framework, there are significant portions that require further refinement to improve accuracy and reliability. Figure 5 predicts the visual presentation of the Ramachandran plot with its statistical analysis.



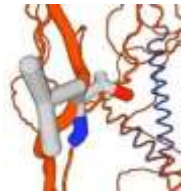
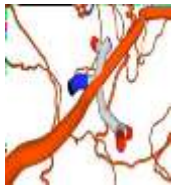
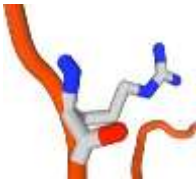
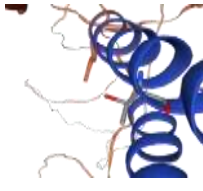
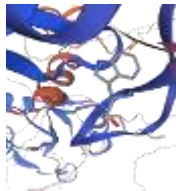
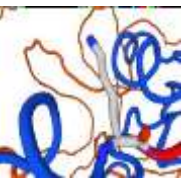
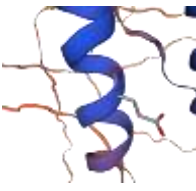
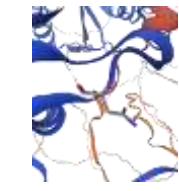
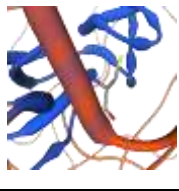
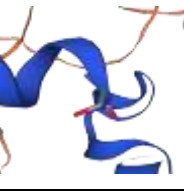
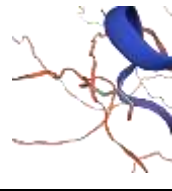
model of BRCA1 has a reasonable

Figure 5: Predicted plot statistics of BRCA1 Ramachandran Plot by PROCHECK

To evaluate the structural impact of deleterious nsSNPs in BRCA1, 3D models of native and mutant proteins were generated using template-based modeling via SWISS-MODEL. Structural changes were examined at the point mutation level, where single nucleotide alterations may lead to loss/gain of phosphorylation, altered stability, or disrupted binding sites. These mutations can impair BRCA1's function in DNA repair, contributing to breast cancer susceptibility.

The mutated BRCA1 variants modeled include F461L, T464I, V1176D, F861C, P514R, L22S, M18R, G1706E, E1735K, M18T, H41N, G98D, V11E, C39Y, and C61W, and their corresponding PDB structures were obtained from SWISS-MODEL. Models were evaluated using QMEAN scores, with both per-residue and global quality estimates, enabling clear discrimination between well- and poorly-modelled regions [28].

Table 3: Structures of BRCA1 gene Mutations

F461L	V1176D	P514R	M18T	C61W
				
M18R	E1735K	V11E	H41N	T464I
				
F861C	L22S	G1706E	C39Y	G98D
				

Structural

Validation and Quality Assessment

Structural validation of BRCA1 protein models with nsSNPs was performed using multiple bioinformatics tools: PROCHECK (Ramachandran plot), ERRAT, and Verify 3D. The native BRCA1 model showed moderate quality with 41.6% residues in favored regions and 25.6% in disallowed regions. Most mutant models maintained similar structural consistency. ERRAT scores ranged from ~46 to 54, indicating acceptable reliability, while Verify 3D scores were low (~14%), reflecting moderate sequence-structure compatibility. Mutation coverage varied as some mutations G98D, V11E, H41N, C61W, and C39Y were fully modeled, others were uncovered, and some clustered in critical domains such as M18R, L22S, and M18T. Mutations such

as M18T, L22S, and M18R were shown clustered located in regions with multiple mutation hotspots, showing PROCHECK core values around 41.5% and ERRAT scores near 52.7, alongside approximately 25.6% disallowed regions. PROCHECK analyzes stereochemical quality by comparing model dihedral angles and geometry against high-quality structures. It is a staple in structural model evaluation [29].

Structural validation of these mutations was carried out by Mutation 3D to map variants such as M18R, M18T, and L22S onto 3D models and identify mutation clusters localized within critical functional domains. Mutation3D's use of complete-linkage clustering ensures that these mutations are closely packed in three-dimensional space, supporting their potential cooperative impact on protein

function [30]. The 3D structure with the location of different mutations embedded

in the protein model is shown in Figure 6.

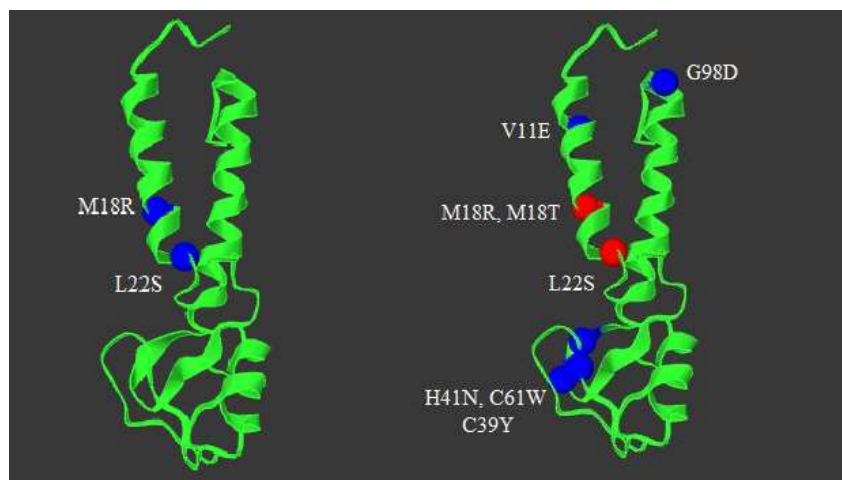


Figure 6: Structural Validation of BRCA1 Mutations by Mutation 3D

Covered mutations, including H41N, C39Y, C61W, G98D, and V11E, reside within well-resolved regions of the BRCA1 structure, enabling more confident assessment. They display PROCHECK core values near 41–42% and ERRAT scores above 51, except C39Y, which shows a slightly lower ERRAT score (~45.8), indicating some structural instability. In contrast, uncovered mutations like F461L, T464I, V1176D, F861C, G1706E, and E1735K occur in regions lacking resolved structural information [31]. While their PROCHECK values are comparable to other categories, their ERRAT scores vary more widely, with E1735K showing a notably low ERRAT score of 46.45, suggesting possible local structural instability.

Figure 6 describes the 3D structure of BRCA1 with key functionally damaging nsSNPs highlighted. The green ribbon shows the protein backbone and red spheres mark clustered mutations (M18R, M18T, and L22S) in a critical domain. Structurally critical domains are essential for preserving the protein's overall conformation; any mutation in these areas can lead to improper folding, instability, or disruption of internal interactions, necessary for the protein's integrity. Functionally critical domains, on the other hand, are involved in the protein's active roles, such as binding to DNA, RNA,

or other proteins, or performing enzymatic or regulatory functions. Blue spheres indicate other important mutations (G98D, V11E, H41N, C61W, and C39Y) spread across the structure. These mutations likely disrupt protein stability and function, affecting BRCA1's role in DNA repair [32]. Mutation3D's use of complete-linkage clustering ensures that these mutations are closely packed in three-dimensional space, supporting their potential cooperative impact on protein function [30].

Notably, several mutations F61L, A63L, F64P, P65Q, N67P, G68A, and A69G are labeled as "clustered", suggesting their proximity in 3D space, which may indicate a structurally or functionally critical domain. Covered mutations are those located in the regions of the protein that are less exposed to the solvent as shown in blue spheres in Figure 6. Uncoverable mutations (not explicitly labeled but implied by the green structure) are located on the protein surface and have increased solvent accessibility.

Collectively, these findings illustrate how integrating *in silico* pathogenicity prediction with 3D structural validation can pinpoint mutation-rich domains that may serve as functional "hotspots" within DNA repair

proteins such as BRCA1 and RAD51D. The identification of spatially clustered and domain-specific deleterious nsSNPs such as M18R, M18T, L22S, G98D, V11E, H41N, C61W, and C39Y in BRCA1 and F61L, A63L, F64P, P65Q, G68A and A69G in RAD51D provides valuable molecular insights into the mechanisms by which these variants could impair homologous recombination, destabilize protein structure, or hinder interactions essential for genome maintenance. By highlighting these high-priority regions, this study not only advances our understanding of the structural underpinnings of breast cancer susceptibility but also lays a strong foundation for targeted experimental validation and the development of precision diagnostics and therapeutics.

Conclusion

This study presents a comprehensive in-silico assessment of nsSNPs in the *BRCA1* gene to evaluate their potential role in breast cancer susceptibility. By integrating multiple predictive algorithms and structural modeling tools, we successfully identified a subset of deleterious nsSNPs that may significantly alter the stability and function of the BRCA1 protein. These high-risk variants, particularly those consistently predicted to be damaging across various computational platforms, warrant further investigation through molecular dynamics simulations and experimental validation. Our findings contribute valuable insights into the mutational landscape of *BRCA1*, reinforcing its pivotal role in maintaining genomic integrity and offering promising targets for genetic screening and risk assessment strategies. Ultimately, the integration of computational predictions with clinical data holds potential to enhance personalized medicine approaches and improve early detection and prevention strategies for hereditary breast cancer.

Limitations

Despite offering a comprehensive in-silico

evaluation of nsSNPs in *BRCA1*, this study has several limitations that must be acknowledged. First, it is difficult to conclusively validate the biological importance of the detected variations because all predictions were computational and lacked experimental validation. The sensitivity and specificity of the predictions made by SIFT, PolyPhen-2, PROVEAN, and MutPred2 may differ, which results in false positives or negatives even if these technologies offer insightful information. Furthermore, only publicly accessible SNPs and protein sequences were included in the datasets used for research, which may have left out uncommon or novel variants that have not yet been cataloged. Furthermore, by concentrating on a group of nsSNPs with questionable relevance, the study may miss other clinically significant mutations that could equally increase the risk of breast cancer. Lastly, while the results are more trustworthy when many prediction techniques are used, the disparities in their outputs demonstrate how challenging it is to come to a definitive conclusion about whether a mutation is hazardous.

Future Directions

The results of this work suggest several possible avenues for advancing our knowledge of how nsSNPs in *BRCA1* contribute to the risk of breast cancer. First and foremost, experimental validation through wet-lab techniques such as mutagenesis studies, protein expression assays, and functional cellular analyses is essential to confirm the in-silico predictions. Integrating clinical and large-scale genomic data from databases like TCGA, gnomAD, and NCBI could help strengthen genotype-phenotype correlations and improve risk prediction. Moreover, characterizing novel or unreported SNPs with uncertain clinical significance could lead to the identification of new pathogenic mutations. Protein-protein interaction analysis may offer insights into how mutations in one gene *BRCA1* could

alter the function of other genes like *BARD1*, *RAD51D*, *BRCA2*, *PALB2*, and *TP53* and disturbing or affecting the DNA repair networks thus playing a role in cancer. Molecular docking approaches could further be employed to identify small molecules that may restore the function of mutated proteins. Including regulatory and non-coding SNPs in future analyses will add another layer of genetic regulation to consider.

Authorship Contributions:

All authors Samavia Mustafa, Maliha Ghaffar, Anam Abbas, Atifa Waheed, Sana Khan and Rabia Yaseen contributed to the study, read and approved the final manuscript and affirm that the work described has not been published previously

References

- Arnold, M., et al., *Current and future burden of breast cancer: Global statistics for 2020 and 2040*. The Breast, 2022. **66**: p. 15-23.
- Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA: a cancer journal for clinicians, 2021. **71**(3): p. 209-249.
- Futreal, P.A., et al., *BRCA1 mutations in primary breast and ovarian carcinomas*. Science, 1994. **266**(5182): p. 120-122.
- Zeng, X., et al., *Screening and predicted value of potential biomarkers for breast cancer using bioinformatics analysis*. Scientific reports, 2021. **11**(1): p. 20799.
- Wang, Y. A., et al., *Germline breast cancer susceptibility gene mutations and breast cancer outcomes*. BMC cancer, 2018. **18**: p. 1-13.
- Wang, Z., et al., *Expression and mutations of BRCA in breast cancer and ovarian cancer: Evidence from bioinformatics analyses*. International journal of molecular medicine, 2018. **42**(6): p.

in any form, whether in print or electronic format.

Acknowledgments: Authors would like to express sincere gratitude to supervisor, Dr. Maliha Ghaffar for invaluable guidance, support, and encouragement throughout. This study was a part of M.Phil. thesis of Samavia Mustafa.

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

3542-3550.

- Miki, Y., et al., *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1*. Science, 1994. **266**(5182): p. 66-71.
- Roy, R., J. Chun, and S.N. Powell, *BRCA1 and BRCA2: different roles in a common pathway of genome protection*. Nature Reviews Cancer, 2012. **12**(1): p. 68-78.
- Jiricny, J., *The multifaceted mismatch-repair system*. Nature reviews Molecular cell biology, 2006. **7**(5): p. 335-346.
- Dietlein, F., L. Thelen, and H.C. Reinhardt, *Cancer-specific defects in DNA repair pathways as targets for personalized therapeutic approaches*. Trends in genetics, 2014. **30**(8): p. 326-339.
- Wu, J., L.-Y. Lu, and X. Yu, *The role of BRCA1 in DNA damage response*. Protein & cell, 2010. **1**: p. 117-123.
- Kobayashi, H., et al., *Hereditary breast and ovarian cancer susceptibility genes*. Oncology reports, 2013. **30**(3): p. 1019-1029.
- Venkitaraman, A.R., *Cancer susceptibility and the functions of BRCA1 and BRCA2*. Cell, 2002. **108**(2): p. 171-182.

- Yuan, Z., et al., *Investigation on BRCA1 SNPs and its effects on mastitis in Chinese commercial cattle*. *Gene*, 2012. **505**(1): p. 190-194.
- Liede, A., et al., *Contribution of BRCA1 and BRCA2 mutations to breast and ovarian cancer in Pakistan*. *The American Journal of Human Genetics*, 2002. **71**(3): p. 595-606.
- Yadegari, F. and K. Majidzadeh, *In silico analysis for determining the deleterious nonsynonymous single nucleotide polymorphisms of BRCA genes*. *Molecular biology research communications*, 2019. **8**(4): p. 141.
- Hu, J. and P.C. Ng, *SIFT Indel: predictions for the functional effects of amino acid insertions/deletions in proteins*. *PloS one*, 2013. **8**(10): p. e77940.
- Adzhubei, I., D.M. Jordan, and S.R. Sunyaev, *Predicting functional effect of human missense mutations using PolyPhen-2*. *Current protocols in human genetics*, 2013. **76**(1): p. 7.20. 1-7.20. 41.
- Sandell, L. and N.P. Sharp, *Fitness effects of mutations: an assessment of PROVEAN predictions using mutation accumulation data*. *Genome biology and evolution*, 2022. **14**(1): p. evac004.
- Shihab, H.A., et al., *Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models*. *Human mutation*, 2013. **34**(1): p. 57-65.
- Capriotti, E., P. Fariselli, and R. Casadio, *I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure*. *Nucleic acids research*, 2005. **33**(suppl_2): p. W306-W310.
- Cheng, J., A. Randall, and P. Baldi, *Prediction of protein stability changes for single-site mutations using support vector machines*. *Proteins: Structure, Function, and Bioinformatics*, 2006. **62**(4): p. 1125-1132.
- Pejaver, V., et al., *Inferring the molecular and phenotypic impact of amino acid variants with MutPred2*. *Nature communications*, 2020. **11**(1): p. 5918.
- Venter, M., et al., *Using MutPred derived mtDNA load scores to evaluate mtDNA variation in hypertension and diabetes in a two-population cohort: The SABPA study*. *Journal of genetics and genomics*, 2017. **44**(3): p. 139-149.
- Kumar, P., S. Henikoff, and P.C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm*. *Nature protocols*, 2009. **4**(7): p. 1073-1081.
- Arshad, S., et al., *In-silico analyses of nonsynonymous variants in the BRCA1 gene*. *Biochemical Genetics*, 2021. **59**(6): p. 1506-1526.
- Stehr, H., et al., *The structural impact of cancer-associated missense mutations in oncogenes and tumor suppressors*. *Molecular cancer*, 2011. **10**(1): p. 54.
- Biasini, M., et al., *SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information*. *Nucleic acids research*, 2014. **42**(W1): p. W252-W258.
- Dym, O., D. Eisenberg, and T. Yeates, *Detection of errors in protein models, in International tables for crystallography volume F: crystallography of biological macromolecules*. 2006, Springer. p. 520-530.
- Meyer, M.J., et al., *mutation3D: cancer gene prediction through atomic clustering of coding variants in the structural proteome*. *Human mutation*, 2016. **37**(5): p. 447-456.
- Colovos, C. and T.O. Yeates, *Verification of protein structures: patterns of nonbonded atomic interactions*. *Protein science*, 1993. **2**(9): p. 1511-1519.

Ramadane-Morchadi, L., et al., ACMG/AMP interpretation of BRCA1 missense variants: Structure-informed scores add evidence strength granularity to the); p. 993-1002.

PP3/BP4 computational evidence. The American Journal of Human Genetics, 2025. 112(5

