In Silico-Based Structural Evaluation to Categorize the Pathogenicity of Mutations Identified in the RAD Class of Proteins
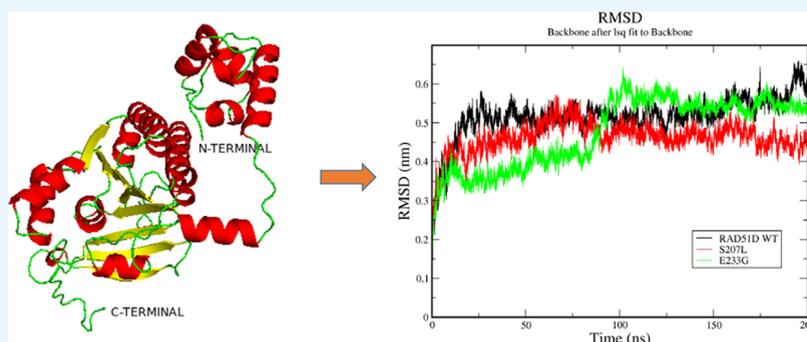
Aaliya Anwaar, Ashok K. Varma, and Reshita Baruah*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: RAD genes, known as double-strand break repair proteins, play a major role in maintaining the genomic integrity of a cell by carrying out essential DNA repair functions via double-strand break repair pathways. Mutations in the RAD class of proteins show high susceptibility to breast and ovarian cancers; however, adequate research on the mutations identified in these genes has not been extensively reported for their deleterious effects. Changes in the folding pattern of RAD proteins play an important role in protein−protein interactions and also functions. Missense mutations identified from four cancer databases, cBioPortal, COSMIC, ClinVar, and gnomAD, cause aberrant conformations, which may lead to faulty DNA repair mechanisms. It is therefore necessary to evaluate the effects of pathogenic mutations of RAD proteins and their subsequent role in breast and ovarian cancers. In this study, we have used eight computational prediction servers to analyze pathogenic mutations and understand their effects on the protein structure and function. A total of 5122 missense mutations were identified from four different cancer databases, of which 1165 were predicted to be pathogenic using at least five pathogenicity prediction servers. These mutations were characterized as high-risk mutations based on their location in the conserved domains and subsequently subjected to structural stability characterization. The mutations included in the present study were selected from clinically relevant mutants in breast cancer pedigrees. Comparative folding patterns and intra-atomic interaction results showed alterations in the structural behavior of RAD proteins, specifically RAD51C triggered by mutations G125V and L138F and RAD51D triggered by mutations S207L and E233G.

1. INTRODUCTION

Breast and ovarian cancers are among the leading causes of cancer deaths around the world. Genetic association of families with a high risk of breast cancer has identified two major susceptibility genes, breast cancer gene 1 (BRCA1) and breast cancer gene 2 (BRCA2).[1,2] The majority of families with autosomal dominant inheritance of breast and ovarian cancer risk can be explained by inherited mutations in the BRCA genes. However, multiple studies have established that BRCA1 and BRCA2 only account for ∼20% of the total familial risk; most families with breast cancer cannot be explained by mutations in these genes alone, suggesting the presence of other susceptibility genes.[3] Furthermore, association of these genes may reveal new drug targets in the genetic background, contributing to the pathogenesis of breast cancer. Additionally, it can lead to genetic testing to identify individuals who are at higher risk of breast cancer. RAD50, also known as RAD50 double-strand break repair protein, RAD51, and its paralogues

RAD51C and RAD51D are of increasing interest for their essential function in maintaining the genomic integrity. The possible role of germline defects in these genes in inherited breast cancer susceptibility has not been studied extensively.

The cell's genomic integrity is maintained by the repair of DNA damage caused by environmental and endogenous DNA-damaging agents. One of the main factors promoting cancer is genetic instability. Malignancy may result from genetic flaws in the DNA damage response and repair pathways. Double-strand break repair and homologous recombination are processes

mediated by RAD genes.[4] RAD50 and RAD51 are two RAD genes that encode proteins with 1312 and 339 amino acids, respectively. Similar to its bacterial counterpart, RecA, hsRad51 catalyzes homologous recombination, which repairs DNA double-strand breaks. Proteins from higher eukaryotes have been found to have a 25 percent amino acid sequence identity with RAD51. The five RAD51 paralogues RAD51B (RAD51L1), RAD51C (RAD51L2), RAD51D (RAD51L3), XRCC2, and XRCC3 play roles in both the early and late stages of homologous recombination (HR) and are required for cellular resistance to double-strand breaks. Although this group shares around 30% amino acid sequence identity, there are no reported functional redundancies, as shown by the embryo lethality brought on by the disruption of three of the five paralogues in mice.[5−7]

The MRN complex consists of MRE11, RAD50, and NBS1.[8,9] The MRN complex is a unique, evolutionarily conserved protein complex that participates in both the HR and non-homologous end-joining (NHEJ) processes. In addition to damage repair, MRN also manages replication stress brought on by oncogenes. The phenotypic variation of these three genes may increase the risk of cancer, particularly breast cancer predisposition. While functionally interacting with other breast cancer risk genes, RAD50 takes part in critical cellular processes for the DNA double-strand break repair mechanism.[10] The human RAD51 gene, HsRAD51, is related to the RecA gene in *Escherichia coli* and is involved in DNA repair and recombination.[11] It has many biochemical similarities to RecA, including the ability to bind to single- and double-stranded DNA, form helical nucleoprotein filaments, participate in DNA strand exchange, and homologous pairing. As per the report by Kato,[12] RAD51's northern blot analysis is comparable to that of BRCA1 and BRCA2. The three genes showed remarkably similar patterns in tissue expression. Additionally, because they are all a part of the same DNA damage response mechanism, any loss in one of them is anticipated to increase the likelihood of cancer-causing mutations. Therefore, it is hypothesized that disruption of the RAD51 pathways may result in cases of hereditary breast and ovarian cancers.[12]

In a remarkable investigation by Meindl et al., monoallelic germline mutations in the RAD51 paralogue RAD51C have been associated with an elevated risk of breast and ovarian cancers.[13] The role of RAD51C as a caregiver and tumor suppressor gene has been confirmed by numerous studies and observations.[14−17] According to biochemical and two-hybrid studies, RAD51C presents itself in the RAD51B/RAD51C/RAD51D/XRCC2 (BCDX2) complex and the RAD51C/XRCC3 (CX3) complex, two distinct complexes composed of RAD51 paralogues.[18−22] The RAD51C HR malfunction may also result in severe chromosomal rearrangements that increase the risk of breast and ovarian cancers. A study found that RAD51C participates in the early phases of homologous recombination by localizing to DNA damage sites both independently and prior to RAD51.[23]

RAD51D attaches to single-stranded DNA following damage and aids in the identification of homology between the damaged and wild-type strands during the repair process through HR via the development of the BCDX2 complex.[24] RAD51D is regarded as a moderate penetrance ovarian cancer susceptibility gene with a lifetime risk of 10−12% when mutated; however, its relationship with breast cancer predisposition is frequently disputed.[25] Numerous studies

have discovered that missense mutations in RAD genes increase the risk of breast and ovarian cancers. Some of these mutations may be neutral, whereas others may have a deleterious effect on the activity of the RAD proteins. Identifying mutations that may interfere with the functioning of these proteins is crucial because the significance of RAD proteins in breast and ovarian cancers is not fully understood. This study aims at identifying the pathogenicity of such missense mutations using bioinformatics-based approaches. Recent advances in bioinformatics have aided in understanding the fundamentals of genetic makeup, such as the functional impact of amino acid residues on the function and structure of the protein.[26−31] These reported deleterious mutations may have a significant impact on the protein's function. Therefore, it is crucial to identify the pathogenicity of mutations identified in the RAD class of proteins.

2. MATERIALS AND METHODS

2.1. Retrieval of RAD Sequences and Mutations from Databases.
UniProt[32] was used to obtain the amino acid sequences for the human-RAD50 (ID Q92878), RAD51 (ID Q06609), RAD51C (ID O43502), and RAD51D (ID O75771) transcripts. The reported RAD mutations were identified from four cancer databases, cBio cancer genomics portal (cBioPortal),[33] catalogue of somatic mutations in cancer (COSMIC),[34] ClinVar,[35] and the genome aggregation database (gnomAD).[36] Furthermore, breast cancer family pedigrees for RAD50 and the RAD51 paralogues RAD51C and RAD51D were also considered for mutation analysis.

2.2. Prediction of Deleterious Effects of Mutations Using Prediction Servers.
Five mutation prediction programs, SIFT,[37] PolyPhen-2,[38] align-GVGD,[39] PROVEAN,[40] and PANTHER,[41] were used to predict the deleterious effects of the missense mutations on the protein function. SIFT uses sequence homology and physical properties to determine if an amino acid substitution affects protein functions. The mutation is predicted as damaging if the toleration index for a particular amino acid substitution is 0.05. PROVEAN is based on the nucleotide sequence, and it predicts the impact of mutations on the protein function. If the prediction score is below −2.5 or above, the nsSNP is predicted to be deleterious or neutral, respectively. PolyPhen-2 is a structural homology-based prediction program. It predicts the effects of a given amino acid substitution based on the protein's structure and function. Align-GVGD, which is based on biophysical characterization, combines protein multiple sequence alignments (MSA) with the biophysical properties of amino acids to accurately predict if a missense mutation is neutral or deleterious. They are categorized into seven classes (C0 as most likely neutral or benign, C15, C25, C35, C45, C55, and C65 as most likely deleterious or pathogenic). Based on the PANTHER-PSEP approach, PANTHER cSNP analyzes the impact of a single amino acid mutation on the protein function. The pathogenicity is categorized as either probably damaging, possibly damaging, or benign.

2.3. Identification of Functional Domains of the RAD Proteins.
The functional characteristics of the proteins and the sites of mutations in RAD proteins were determined using the InterPro[42] tool. Using a database of protein families, domains, and functional sites, InterPro identifies motifs and domains of a protein and, as a result, infer the functional characteristics. Following submission of the protein sequence, the server collects data for the sequences based on the

accession ID provided by the user and explores the conserved domains and other functionally important regions. The results include a graphical representation of the length of the protein and the locations of its domains, active site, motifs, and super families.

2.4. Analysis of the Conserved Amino Acid Residues. The ConSurf[43,44] server was used to predict the structurally and functionally important residues and positions on RAD proteins. To determine whether an evolutionary lineage will retain the same amino acids, the maximum likelihood strategy, often known as the Bayesian method, was used. The method classifies amino acids into three categories: variable (scores between 1 and 3); average (scores between 4 and 6); and conserved (scores between 7 and 9); residues with a score of 9 are regarded as highly conserved. Each amino acid was then classified as functional (f), structural (s), exposed (e), or buried (b). Therefore, the residues with evolutionary scores between 7 and 9 were taken into account.

2.5. Protein Stability Analysis. The I-mutant 3.0 server[45] is used to calculate the free energy of a protein when a mutation is introduced. Using SVM predictions, it predicts a protein's stability for point mutations, and a neural network determines the direction in which the protein's stability shifts after a mutation. Additionally, it forecasts how a mutation will impact the protein's stability. The server determines whether a mutation increases or decreases the protein's stability based on the protein sequence, the substitution, and the position of the residue. It does this by computing the value of free energy change (ΔG) and the reliability index (RI) for each mutation. Negative values of free energy change ($\Delta\Delta G$ or DDG) indicate a decrease in protein stability, and positive values indicate an increase in protein stability.

2.6. Analysis of the Effect of Physicochemical Changes of Mutations on the Proteins. Project HOPE (have (y)our protein explained)[46] is used to predict how point mutations affect our proteins' hydrophobicity, physical and chemical properties, function, and structure. HOPE compiles data from several sources to produce a report on the mutation's impact. The HOPE server does a BLAST[47] search to perform homology modeling[48] using the twinset version of YASARA.[49] It is used to compare the differences between wild-type and mutant amino acids in terms of their sizes, charges, hydrophobicity, and potential interactions that could be brought on by altered residues.

2.7. Molecular Modeling of RAD Proteins. NCBI BLASTP[47] was used against the RCSB protein data bank[50] structures to identify the closely related human RAD homologues.

The comparative ab initio modeling server Robetta[51] was used to build the three-dimensional model structures of RAD50, RAD51, RAD51C, and RAD51D, which employs the algorithm RoseTTA. The query inputs were FASTA sequences of each RAD protein. RoseTTA fold produced five models, and the model with the highest level of confidence was selected.

The models were refined using GalaxyRefine of the GalaxyWEB server.[52] The Ramachandran plot using PRO-CHECK,[53] Verify 3D[54] of SAVES v6.0, and SWISS MODEL's scoring function tool QMEAN server[55] was then used to confirm the stereochemistry of the modeled structures. Mutations were incorporated into the wild-type protein structures using Pymol[56] and visualized using the mutagenesis plugin; the rotamer with the fewer steric conflicts was selected

for the mutated residue to eliminate nonspecific interactions. Energy minimization was also carried out for the atoms within 5A of the mutant.

2.8. Molecular Dynamics Simulation. The GROMACS 2018.1 package[57] was used for molecular dynamics simulations with the AMBER SB99ILDN force field.[58] The AMBER99SB-ILDN force field was selected, as it has been reported to provide results with greater accuracy and consistency.[59] We used pdb2gmx to convert the protein−protein complexes to gromacs files with the -ignh option to remove the hydrogen atoms from the protein structure. The complex was centered at least 1.0 nm from the box's border in the middle of a fixed-volume cubic box filled with SPCE water molecules. Six solvent ions were replaced with six chlorine molecules utilizing genion to guarantee the system's neutrality. The system was put through energy minimization for 50,000 steps with a maximum force of 1000.0 kJ/mol/nm to make sure that there are no steric conflicts and improper geometry. The equilibration process was conducted to prevent the uncontrolled dynamics. In phase I, NVT ensemble, temperature was kept at 310 K with a time constant of 1 ps using a Berendsen thermostat, and all bonds were limited with 50,000 steps. Water molecules were constrained using the SETTLE algorithm,[60] whereas non-water bonds were constrained using the LINCS algorithm.[61] The root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), radius of gyration (R_g), SASA (solvent-accessible surface area), DSSP, and PCA (principal component analysis) data were calculated from the trajectory files produced by the MD simulation for wild types and respective mutants.

The GROMACS software package's essential dynamics (ED) was used to visualize the simulations' overall motions. Proteins' linked motions are extracted through covariance analysis, also known as principal component analysis or ED, to understand the motions that are most essential to a protein's activity. g_covar and g_anaeig, two Gromacs utilities, were applied to analyze the trajectories.

3. RESULTS

3.1. Mutational Datasets Retrieved for RAD Proteins. The FASTA sequences obtained from Uniport for the four proteins RAD50, RAD51, RAD51C, and RAD51D were of 1312aa, 339aa, 376aa, and 328aa, respectively. Four databases, cBioPortal, ClinVar, COSMIC, and gnomAD, were used to retrieve the mutational details. A total of 13,806 mutations for all four proteins were retrieved after the elimination of the redundant mutations across different cancer types. Next, 7127 mutations were identified in RAD50, with 1109 in RAD51, 2943 in RAD51C, and 2627 in RAD51D. In our analysis, 5122 of these mutations, 3018 in RAD50, 233 in RAD51, 1172 in RAD51C, and 699 in RAD51D, were missense mutations.

3.2. Predicted Deleterious Missense Mutations for the RAD Class of Proteins. To predict the functional impact of missense variations of the RAD class retrieved from various databases, five pathogenicity prediction servers SIFT, Poly-Phen-2, align-GVGD, PROVEAN, and PANTHER were used. A total of 569 mutations were predicted to be deleterious for RAD50, with 80 for RAD51, 282 for RAD51C, and 234 for RAD51D. Thus, a total of 1165 mutations were found to be of damaging nature for the RAD class of proteins (RAD50, RAD51, RAD51C, and RAD51D) as predicted by the five prediction tools (Tables S1−S4).

Table 1. RAD51C and RAD51D Mutations with Their Predicted Pathogenicity, Conservation, and Stability

mutant	PROVEAN	PolyPhen-2	SIFT	align GVGD	PANTHER	CONSURF	I-mutant
G125V	damaging	probably damaging	damaging	class C65	probably damaging	highly conserved	decrease
L138F	damaging	damaging	damaging	class C55	possibly damaging	conserved	decrease
S207L	damaging	probably damaging	tolerated	class C65	probably damaging	highly conserved	decrease
E233G	damaging	probably damaging	damaging	class C65	possibly damaging	conserved	decrease

3.3. RAD Domains Identified by the InterPro Server.

Interpro was used to identify the functional domains of RAD50, RAD51, RAD51C, and RAD51D proteins. For the RAD50 protein, it identified ATP-binding cassette domain comprising the Zn-Hook domain, SbcC AAA domain, and superfamily P Loop NTPase (Figure S1). Five functional domains, RecA-like, ATP-binding domain, RAD51 DMC1 RADA domain, AAA+ ATPase, DNA recombination, and repair protein RAD51-like, C-terminal domain, RecA_monomer−monomer_interface domain, and P-loop_NTPase superfamily, were identified for RAD51 (Figure S2). For RAD51C, RecA-like, ATP-binding domain, RAD51 DMC1 RADA domain, DNA recombination and repair protein RAD51-like, C-terminal domain, and superfamily P-loop_-NTPase were identified (Figure S3). In RAD51D, RecA-like, ATP-binding domain, AAA+ ATPase, DNA recombination and repair protein RAD51-like, C-terminal domain, and P-loop_NTPase superfamily were identified (Figure S4).

3.4. Evolutionary Conservation Analysis of Proteins.

The ConSurf server was used to determine the evolutionary conservancy of the amino acid residues. The residues displayed the scores ranging from 1 to 9 where 9 stands for highly conserved sequences. The residues having evolutionary scores in the range of 7−9 were known to be evolutionary-conserved and included for our analysis. Out of 569 pathogenic amino acid mutations in RAD50, 447 were predicted to be in conserved to highly conserved regions. Out of the 80 pathogenic amino acid mutations in RAD51, 56 were predicted to be in conserved to highly conserved regions. Similarly, for RAD51C, out of the 282 pathogenic amino acid mutations, 209 were predicted to be in conserved to highly conserved regions, and lastly, for RAD51D, out of the 234 pathogenic amino acid mutations, 151 were in the conserved to highly conserved regions (Tables S5−S8).

3.5. Mutational Effect on Protein Stability.

Evolutionary-conserved amino acids were further analyzed for the decrease in protein stability. In our observation, 384 out of the 447 mutations showed a decrease in stability in the RAD50 protein; 44 out of the 56 mutations showed a decrease in protein stability in RAD51; 182 out of the 209 mutations showed a decrease in stability in the RAD51C protein; and 122 out of the 151 mutations showed a decrease in the protein stability in RAD51D compared to their respective wild types. The rest of the amino acid mutations showed a rather increased structural stability compared to that of their wild types. Only those mutations that showed decreased stability compared to their wild types were considered for further evaluation (Tables S9−S12).

3.6. Retrieved Familial Pedigrees.

To analyze a rather more refined set of mutants, we further correlated the findings from ConSurf, I-mutant, and selected variants identified in breast cancer pedigrees for RAD50 and the RAD51 paralogues RAD51C and RAD51D. We selected three mutants of RAD50, six mutants of RAD51C, and four mutants of the RAD51D protein that were predicted to be pathogenic and located in highly conserved regions. However, no RAD51 missense mutations associated with familial breast cancer were found in the pedigrees. Hence, we selected D515G,[62] A1216G,[63] and R193W[64] for RAD50; C135Y, L138F, L219S,[65] G125V, D159N,[13] and G153D[66] for RAD51C; and Q115H,[67] S207L,[68] V28M,[69] and E233G[70] for RAD51D from the available repository to correlate with clinical outcomes (Figures S5−S13). Furthermore, based on the availability of RAD structures and relevance of the mutants in populations (Table S13), G125V and L138F for RAD51C and S207L and E233G for RAD51D were selected for further analysis of changes in their physicochemical nature. Tabular depiction of these selected mutants with their predicted pathogenicity, conservation and stability is shown in Table 1.

3.7. Analysis of Mutation-Driven Physicochemical Changes.

Project HOPE was used to evaluate the differences in terms of specific size, charge, hydrophobicity value, and probable interactions that might be induced by mutated residues. It was observed that G125V, S207L, and E233G have increased hydrophobicity compared to that of their wild type counterparts. Again, the mutant E233G was predicted to be smaller in size compared to its wild type counterpart, whereas the rest of the mutants were predicted to be bigger. Table 2 shows the effects of mutations on the physicochemical properties of RAD proteins.

3.8. Modeled Wild Type and Mutants of RAD Proteins.

Since no crystal or solution structures for RAD50, RAD51C, and RAD51D in full length were available and noting the Alphafold models with poor stereochemical assessment, we decided to work on de novo modeling using the Robetta server. Amino acid sequences were submitted to the Robetta server; the RoseTTA fold algorithm generated five models each for RAD51C and RAD51D. Based on the high model confidence value, the best models were further validated for the next set of analysis (Figures 1A and 2A). For RAD51C, the Ramachandran plot statistics of the wild-type protein obtained from the PROCHECK package on SAVES v6.0 showed 94.7% residues at the favored regions, 4.7% allowed region, 0.3% generously allowed region, and 0.9% disallowed region, while RAD51D showed 94.4% residues at the favored regions, 4.5% allowed region, 0.7% generously allowed region, and 0.3% disallowed region (Figures 1B and 2B). With 86.44 and 88.41% of the amino acid residues, respectively, having scored ≥ 0.2 3D/1D profile, both RAD51C and RAD51D structures passed Verify 3D. The model quality estimates were also analyzed using the QMEAN program integrated into the SWISS model workspace. With a Z-score of 0.04 and 0.18 again, both models were clearly within the expected quality range of <1 corresponding to good-quality structures.

Having achieved good-quality models, mutations were incorporated into the modeled structures of RAD51C and RAD51D using the mutagenesis plugin with Pymol choosing rotamers with the fewest steric clashes. Molecular dynamics simulations were then carried out on modeled G125V, L138F, S207L, and E233G mutants and their respective wild types.

Table 2. Effects of Mutations on the Physicochemical Properties of RAD Proteins as Predicted by the HOPE Server

wild-type	amino acid change	amino acid properties	location	effect of polymorphism on the protein
RAD51C	G125V	the variant is larger and more hydrophobic. The mutation is located in a region essential for Holliday junction resolving activity. This area and its function may well be disturbed by the differences in amino acid properties	the variant might interfere with the fundamental nature of the protein since the wild-type is buried in the core	due to the conserved location in which the variant residue is located, this mutation is probably damaging
	L138F	the wild-type amino acid is located in the α-helix and is smaller than the mutant. Because of the mutation, the wild-type residue no longer favors the α-helix as a secondary structure	the mutation may interfere with the protein's core structure since the amino acid is buried	the mutation is located close to a highly conserved position and may sometimes exist without damaging the protein
RAD51D	S207L	the variation is larger and more hydrophobic. The mutation will cause loss of hydrogen bonds in the protein's core, which will interfere with proper protein folding	the variant might interfere with the protein's core since the wild type is buried in the core	the mutant residue is close to a position that is highly conserved and is thus probably damaging to the protein
	E233G	in contrast to the mutation, the wild-type amino acid has a negative charge. The mutant is smaller and more hydrophobic. The protein's core will have an empty space as a result of the mutation. The new residue is not in the proper position to form the same hydrogen bond because of the size difference between the wild-type and mutant residue	the variant might interfere with the fundamental nature of the protein since the wild type is buried in the core	no residues with similar properties have been found at this position and so the mutation is probably damaging to the protein

3.8.1. Wild-Type RAD51C and G125V and L138F Mutants.

MD simulations of 200 ns were performed to understand the differences between the dynamic behavior of wild-type RAD51C and the mutants G125V and L138F. The RMSD backbone plots of the wild type and its mutants were produced to analyze the convergence of MD trajectories. The RMSD plot clearly shows that both native and mutant proteins converged to a stable trajectory after 150 ns when the wild type was stabilized from 100 ns onward (Figure 3A). The wild type exhibited the highest RMSD value of 1.34 nm at the end of the simulation followed by the mutant L138F (1.33 nm) and G125V (0.56 nm) with varying fluctuations throughout the simulation period. Noticeably, L138F showed a higher RMSD value of ∼1.38 nm at 176 ns when compared to that of G125V at 0.97 nm at 41 ns but lower RMSD compared to that of the wild type at 1.48 nm at 111 ns. This graph elucidated that although the mutant L138F exhibited lower stability compared to that of the mutant G125V, the wild type had a greater RMSD value. We examined the RMSF plot, which gives a general overview of the protein's flexible area, to further validate the RMSD results.

To understand the effect of mutation on the dynamic behavior of protein residues, RMSF values of the wild type and the mutants were calculated. it was seen that the wild type exhibited the highest fluctuating pattern among the three, displaying fluctuations at 44, 48, 88, and 250 residue positions (Figure 3B). Following the wild type, the mutant L138F displayed fluctuations at 21, 54, and 353 residue positions. G125V showed lesser fluctuation patterns compared to those of the L138F mutant and the wild type. Additionally, for our residue at position 125, the wild type showed yet again higher flexibility than that of the mutant residue. Comparative residual RMSF values suggested that the wild type was comparatively more flexible, showing higher RMSF values across all regions during the simulation. As a result of the mutant's reduced flexibility, it has a more compact structure than that of the wild type, supporting the findings of the RMSD investigation and the HOPE analysis.

The radius of gyration explains the arrangement of the atoms around the central axis; thus, the more compact structures show lesser radii, and less compact structures have higher radii values. The wild-type value of R_g kept fluctuating for the first 100,000 steps, after which the R_g value was almost constant until the end of the simulation time period. The wild type can be seen unfolding around 75 ns and then refolding around 100 ns compared to the mutant, which unfolded initially but maintained a steady compactness till the end of the time period. As seen from the results obtained, the plot exhibited that the wild type and G125V had similar patterns of compactness throughout the simulation (Figure 3C). For L138F, the R_g change was more significant, and the protein folded extensively at 112 ns before abruptly unfolding. The R_g profile clearly explained that the L138F mutant was more dynamic and less compact than the wild type.

SASA calculates the solvent-accessible surface area of the protein and the buried surface area, that is, the surface area that is not accessible to the solvent. For the first 100 ns of simulation, the SASA value of the wild type was seen to be higher compared to that of the mutants. The values peaked at around 150 ns for the mutant L138F and then gradually came to a constant level for all the three proteins toward the end of the simulation. The wild type was shown to have a slightly higher overall SASA value than that of the mutant L138F,
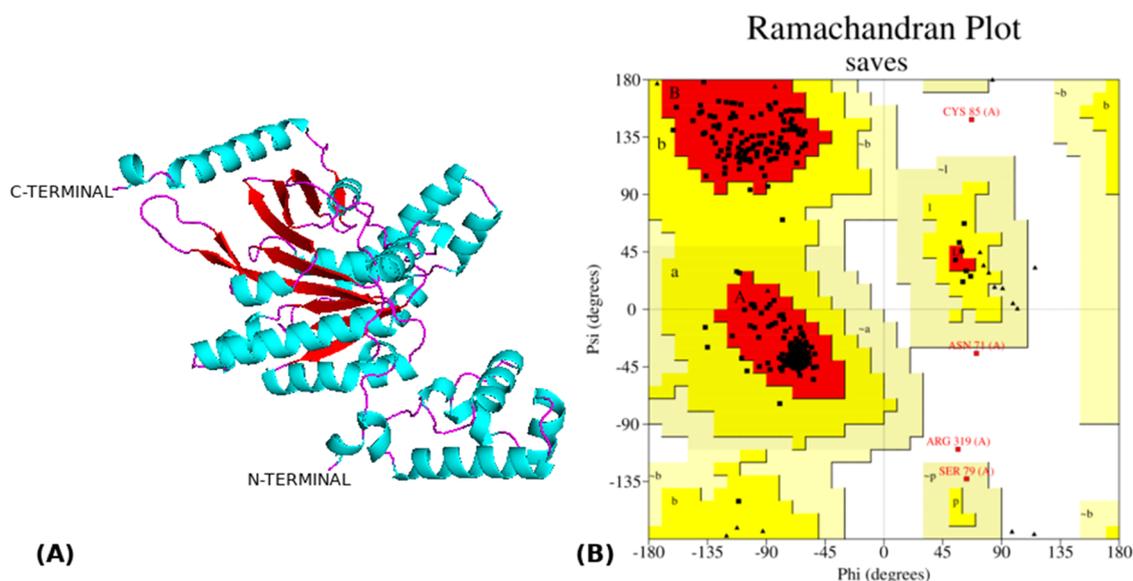
Figure 1. (A) Structure of RAD51C as modeled by Robetta and (B) Ramachandran plot statistics.
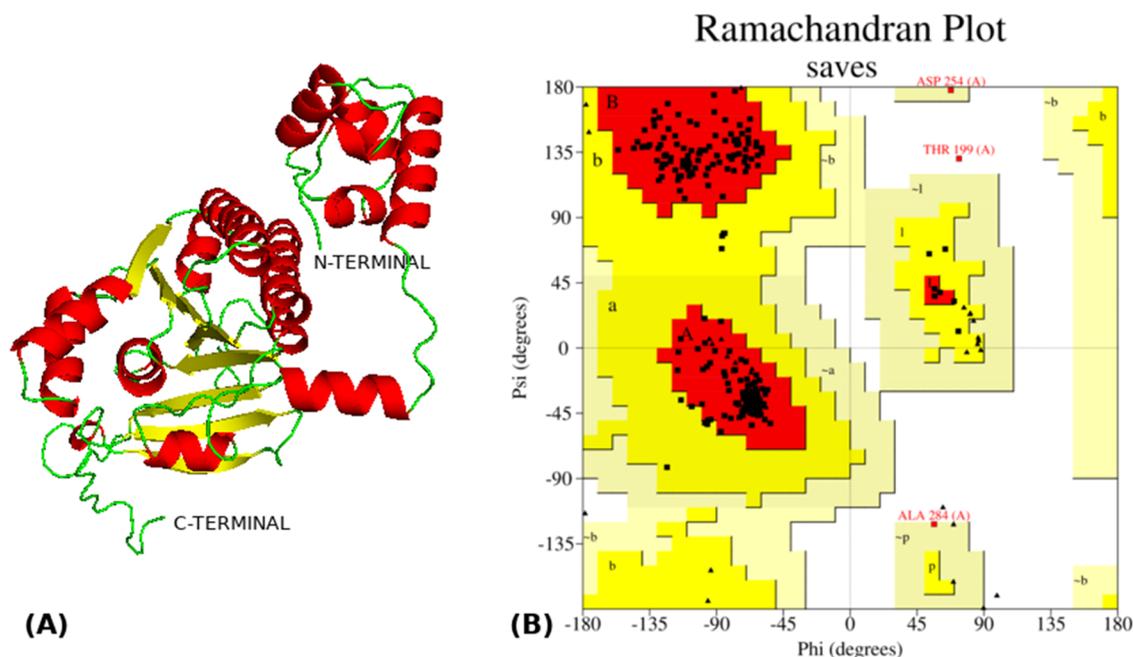

Figure 2. (A) Structure of RAD51D as modeled by Robetta and (B) Ramachandran plot statistics.

followed by G125V, indicating that the wild type destabilized the hydrophobic core (Figure 3D). As a result, residues of the wild type interacted with the solvent more frequently than those of the mutants.

To comprehend the changes in the protein's secondary structures during the simulation period, DSSP analysis was performed. A number of coils, turns, and bends in the native and mutated proteins were observed. The mutation L138F was seen to cause a relative decrease in the number of turns in the protein than the mutation G125V (Figure 3F,G). The turns contribute significantly to secondary structures and also add to the protein's stability. This loss of turn may suggest that the L138F mutation would cause a decrease in stability compared to that of the wild type and the G125V mutation.[71]

Principal component analysis (PCA) was used to fully understand the global or collective motions of the wild type and the mutants G125V and L138F. The projection of trajectories obtained onto the first two principal components indicated the motions. The graphs were plotted with eigenvector 1 on the x-axis and eigenvector 2 on the y-axis, which revealed that the collective motions of the atoms in the wild type and L138F were similar and higher compared to those of G125V (Figure 3H). RAD51C and L138F occupied a larger region of phase space, especially along the first principal component (PC1).

3.8.2. RAD51D Wild Type with S207L and E233G Mutants. The backbone RMSD values of the wild type and mutants fluctuated within 0.28−0.66 nm throughout the simulation period, whereas the wild type stabilized from 50 to 150 ns
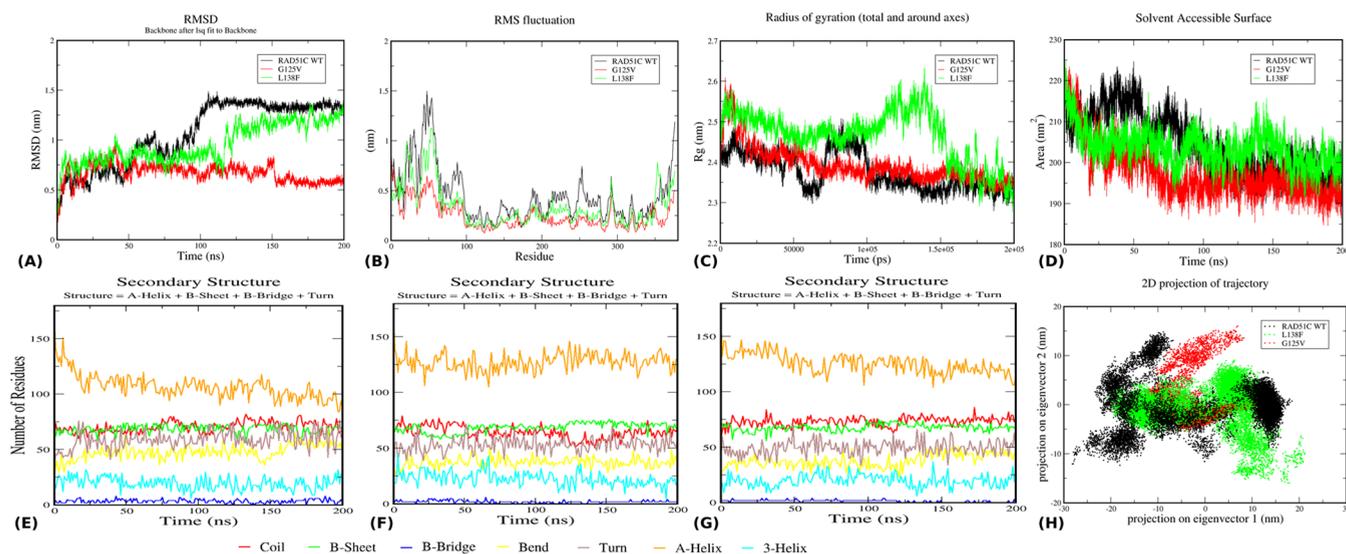
Figure 3. Structural stability, flexibility, and dynamics analysis of the RAD51C wild type (WT) and mutants G125V and L138F. (A) Backbone RMSD of the WT, G125V, and L138F, (B) RMSF of the WT, G125V, and L138F, (C) radius of gyration (R_g) of WT, G125V, and L138F, (D) solvent-accessible surface area (SASA) of the WT, G125V, and L138F, (E−G) secondary structure analysis of the WT, G125V, and L138F, and (H) principal component analysis for the comparative eigenvector projection profile of the WT, G125V, and L138F.
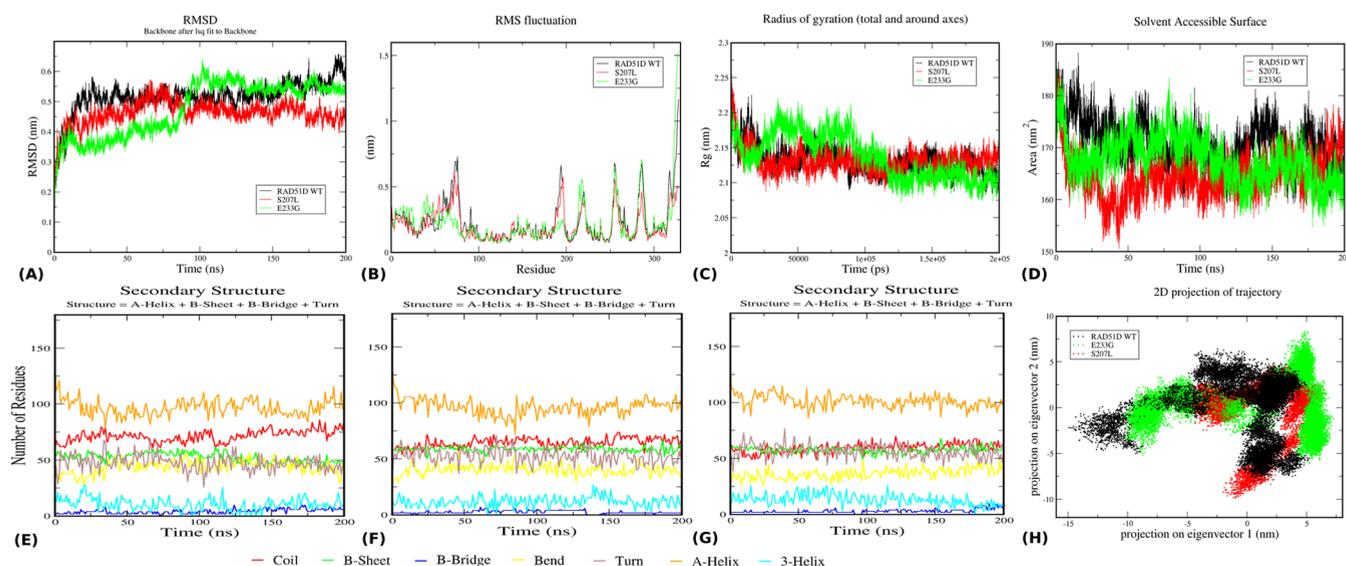

Figure 4. Structural stability, flexibility, and dynamic analysis of the RAD51C wild type (WT) and mutants S207L and E233G. (A) Backbone RMSD of the WT, S207L, and E233G, (B) RMSF of the WT, S207L, and E233G, (C) radius of gyration (R_g) of the WT, S207L, and E233G, (D) solvent-accessible surface area (SASA) of the WT, S207L, and E233G, (E−G) secondary structure analysis of the WT, S207L, and E233G, and (H) principal component analysis for the comparative eigenvector projection profile of the WT, S207L, and E23.

within 0.5 nm (Figure 4A). The wild type exhibited the highest RMSD value of 0.57 nm at the end of the simulation followed by the mutants E233G (0.55 nm at 200 ns) and S207L (0.42 nm at 200 ns) with varying fluctuations throughout the simulation period. When compared to S207L, which had an RMSD of 0.57 nm at 66 ns, E233G had a higher RMSD value of 0.64 nm at 101 ns but a lower RMSD of 0.66 nm at 195 ns than that of the wild type. Additionally, RMSD values for E233G first decreased from 12 to 80 ns and then sharply climbed to 0.64 nm at 101 ns. Average RMSD values of native, S207L, and E233G mutants are 0.2−0.6, 0.24−0.5, and 0.25−0.65 nm, respectively. Although this graph showed that the mutant E233G displayed a lower stability than that of the

mutant S207L, the wild type had a higher RMSD value and was therefore more unstable.

The wild type and the mutant RMSF values fall between 0.07 and 1.16, 0.07 and 0.58, and 0.0 and 1.54 nm, respectively. During the simulated time period, similar variations were seen between the wild type and S207, while E233G showed the most significant number of fluctuations among the three, indicating comparatively higher flexibility in its residues. The wild type showed less variations compared to E233G at residue positions 76, 91, 193, and 328. The mutant E233G structure had comparatively more flexibility than that of the wild type and the mutant S207L, supporting the findings of the RMSD investigation.

The wild type's R_g range is between 2.08 and 2.22 nm, while the R_g range for S207L and E233G is between 2.08 and 2.25 and 2.07 and 2.2 nm, respectively. With a modest increase in the first 25,000 steps, the R_g value for the S207L mutant is comparable to that of the wild type. Folding and unfolding are similar for native and mutant S207L. On the contrary, significant changes in R_g were observed for the E233G mutant, which exhibited an unfolding pattern during 25−95 ns, followed by a steady dip at ∼100 ns, showing refolding of the protein and then maintaining a steady compactness along with the wild type and the S207L mutant (Figure 4C).

The SASA of the native and the two mutants S207L and E233G ranged from 158 to 188.2, 150 to 184, and 155 to 185 nm², respectively. The wild type was more exposed to the solvent, as evidenced by the fact that the wild type had a slightly higher SASA value than that of the mutant E233G, followed by that of S207L (Figure 4D). The SASA for the mutant E233G initially decreased during the initial 45 ns, fluctuating for about 100 ns, after which it converged with that of the mutant S207L until the end of the simulation. The mutants' decreased SASA values imply that a relatively lesser number of residues of the mutants interacted with the solvent compared to the wild type.

From the secondary structure analysis, it was observed that the mutant E233G exhibited a considerable decrease in coils and bends, forming residues compared to the wild type (Figure 4E,G). This loss may suggest that the mutations, particularly E233G, would cause a significant change to the secondary structural elements of the wild type, thus affecting the overall stability of the protein.[72]

The graphs were plotted with eigenvector 1 on the *x*-axis and eigenvector 2 on the *y*-axis. The plot depicted that the collective motions of the atoms in the wild type and E233G were similar and higher compared to those of S207L, along eigenvector 1 and eigenvector 2 projections in the subspace (Figure 4H). It was evident from the results that more dynamics were observed in the wild type and the mutant E233G.

RAD51D's ATPase domain contains the conserved ATP binding Walker A and B motifs, which are essential for its role in HR. These motifs are thought to be involved in the binding of RAD51C and XRCC2, respectively. In a study by Rivera et al., the variant S207L was further shown to be located in the Walker B motif, hence impairing DSB repair by HR.[68] Rodrguez's model revealed that changing to glycine in the 233 position could affect the Rad51D structure in two ways, whereas the glutamic acid in the 233 conserved position is located in the middle of helix F, and its replacement to glycine reduces the structural stability. Second, glutamic acid-233 may form an electrostatic interaction with arginine-186, mimicking the semiburied salt bridge found between aspartic acid-130 and lysine-177 in the RecA structure. Substituting glycine for glutamic acid-233 eliminates the salt bridge, which is conserved across the phylogenetic scale in RecA, and may reduce protein stability and affect the Rad51D structure.[70]

4. DISCUSSION

The RAD class of proteins is crucial for preservation of genomic integrity; therefore, we examined the significance of RAD mutations in hereditary susceptibility to breast cancer. The execution of the functions of RAD proteins critically depends on changes in their structural conformation during biomolecular interactions. Thus, evaluating the molecular cause and effect of these mutations can provide insights into the many forms of cancers and their therapeutic approaches. Here, we used computational studies to analyze the most detrimental missense mutations and their effects on the structural and functional properties of the RAD class of proteins.

The functional analysis of large dataset mutations may offer a critical clue for diagnosis and treatment due to the growing recognition of RAD genes in breast, ovarian, and other related cancers. Therefore, a combination of sequence- and structure-based techniques was utilized for the profiling of highly functional mutations in coding regions to look into potential associations between genetic mutation and phenotypic changes. The main goal of the proposed computational investigation was to identify the single amino acid substitution using the functional consequences of mutations in RAD proteins. To increase the level of confidence in this study, we used five in silico SNP prediction tools, SIFT, PolyPhen-2, PROVEAN, align-GVGD, and PANTHER, to screen a total of 5122 nonredundant missense mutations for RAD50, RAD51, RAD51C, and RAD51D from 13,806 mutations retrieved from cBioPortal, COSMIC, ClinVar, and gnomAD databases. The mutations showing deleterious effects from at least four tools were selected, which resulted in the identification of 1165 pathogenic missense mutations. These in silico-based prediction tools are published in widely accepted journals.[73−80] It is also crucial to ascertain how our mutations affect the overall structure of RAD proteins. Using the ConSurf web server and I-mutant tool, we selected those deleterious mutations that were located in the conserved regions and affecting the overall protein stability.

RAD proteins have recently been grouped as intermediate penetrance genes for hereditary breast cancer susceptibility. Therefore, to correlate our findings, the reported family pedigrees were used to correlate the pathogenesis of RAD mutations and the associated phenotypic traits. Point mutations are known to cause differences in the physicochemical and molecular characteristics of the respective wild type. The HOPE server predicted the mutations G125V and L138F for RAD51C and the mutations S207L and E233G for RAD51D exhibiting overall alterations in protein physico-chemical properties.

Since no template structures for our proteins were available and the Alphafold models were of poor stereochemical assessment, molecular modeling of wild-type structures for RAD51C and RAD51D was performed using Phyre2, the SWISS homology modeling tool, and the RoseTTA fold de novo algorithm of the Robetta server using both homology and de novo modeling. PROCHECK, QMEANS, and VERIFY-3D servers were used to predict the stereochemical geometry. To explore the effects of the mutations on RAD structures, the identified amino acid substitutions G125V, L138F, S207L, and E233G were mutated into the modeled protein structures using Pymol's mutagenesis plugin.

The dynamic behavior of the wild-type and mutant models was analyzed and evaluated through MD simulation. Fluctuations of RAD51D and its mutants S207L and E233G were observed during the 200 ns simulation period at 310 K. The wild type displayed the highest RMSD values, followed by the mutant E233G, indicating more conformational changes in the wild type than in the mutants. The residue-based root-mean-square fluctuation of the backbone exhibited a more flexible E233G mutant. The radius of gyration was calculated

to assess the degree of compactness and folding patterns between the RAD51D and mutant structures along the three Cartesian axes. While the mutant S207L and wild type displayed similar folding and unfolding patterns throughout the simulation time, the mutant E233G showed a significant increase in R_g, which in turn suggested a loss in compactness. The PCA analysis showed that the wild type and E233G cover more area in the two vectors' projections compared to that of the mutant S207L, while the SASA plot indicated that the wild type is more exposed to the solvent than the mutants. Additionally, to comprehend the changes in the protein's secondary structures during the simulation period, DSSP analysis was performed, which showed that the mutation E233G causes a significant decrease in the coil and bend formation, thus affecting the stability of the protein. This finding was supported by the clinical study by Rodriguez et al., where the substitution of glycine for glutamic acid at position 233 was suggested to have an adverse effect on the stability of the RAD51D structure. Glu at 233 position established an electrostatic contact with Arg186 and simulates the semiburied salt bridge between Asp130 and Lys177 observed at locations in the RecA structure. The conserved salt bridge is destabilized when Glu233 is replaced with glycine, thus reducing the stability of the RAD51D structure.[70] It is observed that the RAD51D variant S207L is the pathogenic missense, which is conserved across different species in the ATPase domain, which is known to impede DSB repair by HR.[66]

For RAD51C, we compared the wild type to the two mutants G125V and L138F. The mutants significantly differed in structural conformation from the wild type; however, the wild type appeared to be less stable than the mutants based on its higher RMSD. L138F was more flexible than G125V but less flexible than the wild type. The R_g profile provided additional evidence that the L138F mutant was less compact and more dynamic than the wild type. The SASA plot, however, indicated that compared to the two mutants, the wild type interacted with the solvent more frequently. Areas were covered by both the wild type and the mutant L138F in the two vector projections, indicating that they were more flexible than S207L. The same was evident from the DSSP analysis, which demonstrated that the mutation L138F causes loss of secondary structure turns. The L138F variant has been shown to be functionally impaired in DNA repair-related assays by Meindl et al. and categorized as likely pathogenic in different familial breast cancer studies. Additionally, G125V and L138F were both predicted to impair the function of the RAD51C protein.[13,81] It has also been observed that changing a protein's flexibility or SASA can result in abnormal protein folding and aggregation and a loss of thermodynamic stability.[82] The vast majority of disease-causing mutations, as per the study reported by Steward et al.[83] and Ye, Li et al.,[84] are located in solvent-accessible locations; therefore, it is worth studying these mutations, which could help understand a disease's progression.[83,84] It has been reported that RAD51C's HR mechanism can result in gross chromosomal rearrangements. RAD51C-mutated chicken DT40 and hamster cells have shown spontaneous chromosomal aberrations similar to BRCA genes. The mutated variants G125V and L138F were unable to reverse the hypersensitivity of RAD51C chicken DT40 cells to MMC, and hence, the expression of these missense mutants did not restore normal RAD51 focus formation in RAD51C mutant fibroblasts, indicating that these RAD51C missense mutants have an obvious HR defect.[13]

Time-dependent dynamics and simulation of protein molecules provide precise insights into fluctuations and changes in the protein trajectory. Molecular dynamics is rapidly being used to study the structure, dynamics, and thermodynamics of proteins, nucleic acids, and their complexes. One significant drawback of this study is the dearth of neutral- or low-score mutations accessible for investigation. SNPs interact with consensus sequences to alter RNA processing in addition to modifying the amino acids. In general, it is reasonable to draw the conclusion that in silico approaches continue to be a reliable tool to quickly analyze the anticipated effects of mutations. However, prediction will be more accurate when the in silico results are validated by in vitro and in vivo experimentations. Given the significant number of missense mutations in RAD genes, association studies on genetic variants with functional importance should be conducted.

5. CONCLUSIONS

We observed that the selected mutations, namely, G125V and L138F of RAD51C and S207L and E233G of RAD51D, had brought about an overall change in the protein structure, leading to changes in stability compared to that of their wild-type counterparts. Majority of deleterious mutations alter the size, charge, hydrogen bonds, and other properties of amino acids along with changes in a protein's flexibility, leading to faulty protein folding and loss of thermodynamic stability. The multimodel computational predictions reported here evaluated the possibility of using in silico tools in predicting the functional SNPs. Although there are certain limitations, results from experimental research can be correlated to those of MD simulation analysis to classify the pathogenicity of mutations.

■ ASSOCIATED CONTENT

ⓈSupporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c07802.

> Domain identification of the RAD50 protein using the Interpro tool; domain identification of the RAD51 protein using the Interpro tool; domain identification of the RAD51C protein using the Interpro tool; domain identification of the RAD51D protein using the Interpro tool; family pedigree with the identified RAD50 D515G variant in a breast cancer patient; family pedigree with the identified RAD50 A1216G variant in a breast cancer patient; family pedigree with the identified RAD50 A1216G variant in a breast cancer patient; family pedigree with the identified RAD50 R193W variant in a breast cancer patient; family pedigree with the identified RAD51C C135Y, L138F, and L219S variants in a breast cancer family; family pedigree with identified RAD51C G125V, L138F, and D159N variants in a breast cancer family; family pedigree with identified RAD51C G153D variant in a breast cancer family; family pedigree with the identified RAD51D Q115H variant in an ovarian cancer family; family pedigree with the identified RAD51D S207L variant in an ovarian cancer family; family pedigree with the identified RAD51D V28M variant in an ovarian cancer family; mutations on the RAD50 protein predicted as pathogenic by the prediction servers; mutations on the RAD51 protein predicted as pathogenic by the

prediction servers; mutations on the RAD51C protein predicted as pathogenic by the prediction servers; mutations on the RAD51D protein predicted as pathogenic by the prediction servers; evolutionary conservation of the conserved residues of the RAD50 protein as predicted by the ConSurf server; evolutionary conservation of the conserved residues of the RAD51 protein as predicted by the ConSurf server; evolutionary conservation of the conserved residues of the RAD51C protein as predicted by the ConSurf server; evolutionary conservation of the conserved residues of the RAD51D protein as predicted by the ConSurf server; effect on the stability of the RAD50 protein as predicted by the I-Mutant 3.0 server; effect on the stability of the RAD51 protein as predicted by the I-Mutant 3.0 server; effect on the stability of the RAD51C protein as predicted by the I-Mutant 3.0 server; effect on the stability of the RAD51D protein as predicted by the I-Mutant 3.0 server; and RAD51C and RAD51D mutations reported in populations (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Reshita Baruah − *Advanced Centre for Treatment, Research and Education in Cancer, Navi Mumbai 410210 Maharashtra, India;* ⓞ orcid.org/0000-0002-6103-0436; Email: reshitabaruah@gmail.com

Authors

Aaliya Anwaar − *Advanced Centre for Treatment, Research and Education in Cancer, Navi Mumbai 410210 Maharashtra, India*

Ashok K. Varma − *Advanced Centre for Treatment, Research and Education in Cancer, Navi Mumbai 410210 Maharashtra, India; Homi Bhabha National Institute, Mumbai 400094 Maharashtra, India;* ⓞ orcid.org/0000-0001-6091-5315

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.2c07802

Author Contributions

A.A. performed the experiments, analyzed the data, generated the figures and tables, and wrote the first draft. A.K.V. designed the experiments. R.B. analyzed the data, made the figures, and wrote the paper.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Miki, Y.; Swensen, J.; Shattuck-Eidens, D.; Futreal, P. A.; Harshman, K.; Tavtigian, S.; Liu, Q.; Cochran, C.; Bennett, L. M.; Ding, W.; et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **1994**, *266*, 66−71.

(2) Wooster, R.; Bignell, G.; Lancaster, J.; Swift, S.; Seal, S.; Mangion, J.; Collins, N.; Gregory, S.; Gumbs, C.; Micklem, G.; et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature* **1995**, *378*, 789−792.

(3) Easton, D. F. How many more breast cancer predisposition genes are there? *Breast Cancer Res.* **1999**, *1*, No. 14.

(4) Tsukamoto, Y.; Kato, J.; Ikeda, H. Effects of mutations of RAD50, RAD51, RAD52, and related genes on illegitimate recombination in *Saccharomyces cerevisiae*. *Genetics* **1996**, *142*, 383−391.

(5) Kawabata, M.; Kawabata, T.; Nishibori, M. Role of recA/RAD51 family proteins in mammals. *Acta Med. Okayama* **2005**, *59*, No. 31987.

(6) Lim, D. S.; Hasty, P. A mutation in mouse rad51 results in an early embryonic lethal that is suppressed by a mutation in p53. *Mol. Cell. Biol.* **1996**, *16*, 7133−7143.

(7) Smiraldo, P. G.; Gruver, A. M.; Osborn, J. C.; Pittman, D. L. Extensive chromosomal instability in Rad51d-deficient mouse cells. *Cancer Res.* **2005**, *65*, 2089−2096.

(8) Dolganov, G. M.; Maser, R. S.; Novikov, A.; Tosto, L.; Chong, S.; Bressan, D. A.; Petrini, J. H. Human Rad50 is physically associated with human Mre11: identification of a conserved multiprotein complex implicated in recombinational DNA repair. *Mol. Cell. Biol.* **1996**, *16*, 4832−4841.

(9) Lamarche, B. J.; Orazio, N. I.; Weitzman, M. D. The MRN complex in double-strand break repair and telomere maintenance. *FEBS Lett.* **2010**, *584*, 3682−3695.

(10) Alblihy, A.; Shoqafi, A.; Toss, M. S.; Algethami, M.; Harris, A. E.; Jeyapalan, J. N.; Abdel-Fatah, T.; Servante, J.; Chan, S.; Green, A.; Mongan, N. P.; Rakha, E. A.; Madhusudan, S. Untangling the clinicopathological significance of MRE11-RAD50-NBS1 complex in sporadic breast cancers. *NPJ Breast Cancer* **2021**, *7*, No. 143.

(11) Ogawa, T.; Yu, X.; Shinohara, A.; Egelman, E. H. Similarity of the yeast RAD51 filament to the bacterial RecA filament. *Science* **1993**, *259*, 1896−1899.

(12) Kato, M.; Yano, K.; Matsuo, F.; Saito, H.; Katagiri, T.; Kurumizaka, H.; Yoshimoto, M.; Kasumi, F.; Akiyama, F.; Sakamoto, G.; Nagawa, H.; Nakamura, Y.; Miki, Y. Identification of Rad51 alteration in patients with bilateral breast cancer. *J. Human Genet.* **2000**, *45*, 133−137.

(13) Meindl, A.; Hellebrand, H.; Wiek, C.; Erven, V.; Wappenschmidt, B.; Niederacher, D.; Freund, M.; Lichtner, P.; Hartmann, L.; Schaal, H.; Ramser, J.; Honisch, E.; Kubisch, C.; Wichmann, H. E.; Kast, K.; Deissler, H.; Engel, C.; Müller-Myhsok, B.; Neveling, K.; Kiechle, M.; Mathew, C. G.; Schindler, D.; Schmutzler, R. K.; Hanenberg, H. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat. Genet.* **2010**, *42*, 410−414.

(14) Gresner, P.; Gromadzinska, J.; Twardowska, E.; Rydzynski, K.; Wasowicz, W. Rad51C: a novel suppressor gene modulates the risk of head and neck cancer. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2014**, *762*, 47−54.

(15) Tumiati, M.; Hemmes, A.; Uusivirta, S.; Koopal, S.; Kankainen, M.; Lehtonen, E.; Kuznetsov, S. G. Loss of Rad51c accelerates tumourigenesis in sebaceous glands of Trp53-mutant mice. *J. Pathol.* **2015**, *235*, 136−146.

(16) Thompson, E. R.; Boyle, S. E.; Johnson, J.; Ryland, G. L.; Sawyer, S.; Choong, D. Y.; Chenevix-Trench, G.; Trainer, A. H.; Lindeman, G. J.; Mitchell, G.; James, P. A.; Campbell, I. G. Analysis of RAD51C germline mutations in high-risk breast and ovarian cancer families and ovarian cancer patients. *Hum. Mutat.* **2012**, *33*, 95−99.

(17) Schnurbein, G.; Hauke, J.; Wappenschmidt, B.; Weber-Lassalle, N.; Engert, S.; Hellebrand, H.; Garbes, L.; Becker, A.; Neidhardt, G.; Rhiem, K.; Meindl, A.; Schmutzler, R. K.; Hahnen, E. RAD51C deletion screening identifies a recurrent gross deletion in breast cancer and ovarian cancer families. *Breast Cancer Res.* **2013**, *15*, No. R120.

(18) Masson, J. Y.; Tarsounas, M. C.; Stasiak, A. Z.; Stasiak, A.; Shah, R.; McIlwraith, M. J.; Benson, F. E.; West, S. C. Identification

and purification of two distinct complexes containing the five RAD51 paralogs. *Genes Dev.* **2001**, *15*, 3296−3307.

(19) Dosanjh, M. K.; Collins, D. W.; Fan, W.; Lennon, G. G.; Albala, J. S.; Shen, Z.; Schild, D. Isolation and characterization of RAD51C, a new human member of the RAD51 family of related genes. *Nucleic Acids Res.* **1998**, *26*, 1179−1184.

(20) Liu, N.; Lamerdin, J. E.; Tebbs, R. S.; Schild, D.; Tucker, J. D.; Shen, M. R.; Brookman, K. W.; Siciliano, M. J.; Walter, C. A.; Fan, W.; Narayana, L. S.; Zhou, Z. Q.; Adamson, A. W.; Sorensen, K. J.; Chen, D. J.; Jones, N. J.; Thompson, L. H. XRCC2 and XRCC3, new human Rad51-family members, promote chromosome stability and protect against DNA cross-links and other damages. *Mol. Cell* **1998**, *1*, 783−793.

(21) Braybrooke, J. P.; Spink, K. G.; Thacker, J.; Hickson, I. D. The RAD51 family member, RAD51L3, is a DNA-stimulated ATPase that forms a complex with XRCC2. *J. Biol. Chem.* **2000**, *275*, 29100−29106.

(22) Schild, D.; Lio, Y. C.; Collins, D. W.; Tsomondo, T.; Chen, D. J. Evidence for simultaneous protein interactions between human Rad51 paralogs. *J. Biol. Chem.* **2000**, *275*, 16443−16449.

(23) Badie, S.; Liao, C.; Thanasoula, M.; Barber, P.; Hill, M. A.; Tarsounas, M. RAD51C facilitates checkpoint signaling by promoting CHK2 phosphorylation. *J. Cell Biol.* **2009**, *185*, 587−600.

(24) Yokoyama, H.; Sarai, N.; Kagawa, W.; Enomoto, R.; Shibata, T.; Kurumizaka, H.; Yokoyama, S. Preferential binding to branched DNA strands and strand-annealing activity of the human Rad51B, Rad51C, Rad51D and Xrcc2 protein complex. *Nucleic Acids Res.* **2004**, *32*, 2556−2565.

(25) Osher, D. J.; De Leeneer, K.; Michils, G.; Hamel, N.; Tomiak, E.; Poppe, B.; Leunen, K.; Legius, E.; Shuen, A.; Smith, E.; Arseneau, J.; Tonin, P.; Matthijs, G.; Claes, K.; Tischkowitz, M. D.; Foulkes, W. D. Mutation analysis of RAD51D in non-BRCA1/2 ovarian and breast cancer families. *Br. J. Cancer* **2012**, *106*, 1460−1463.

(26) Rajendran, V.; Purohit, R.; Sethumadhavan, R. In silico investigation of molecular mechanism of laminopathy caused by a point mutation (R482W) in lamin A/C protein. *Amino Acids* **2012**, *43*, 603−615.

(27) Rajendran, V.; Gopalakrishnan, C.; Sethumadhavan, R. Pathological role of a point mutation (T315I) in BCR-ABL1 protein-A computational insight. *J. Cell. Biochem.* **2018**, *119*, 918−925.

(28) Kamaraj, B.; Al-Subaie, A. M.; Ahmad, F.; Surapaneni, K. M.; Alsamman, K. Effect of novel leukemia mutations (K75E & E222K) on interferon regulatory factor 1 and its interaction with DNA: insights from molecular dynamics simulations and docking studies. *J. Biomol. Struct. Dyn.* **2021**, *39*, 5235−5247.

(29) Al-Subaie, A. M.; Kamaraj, B. The Structural Effect of FLT3 Mutations at 835th Position and Their Interaction with Acute Myeloid Leukemia Inhibitors: In Silico Approach. *Int. J. Mol. Sci.* **2021**, *22*, No. 7602.

(30) Rampogu, S.; Lemuel, M. K.; Lee, K. W. Virtual screening, molecular docking, molecular dynamics simulations and free energy calculations to discover potential DDX3 inhibitors. *Adv. Cancer Biol.: Metastasis* **2022**, *4*, No. 100022.

(31) Ali, M. Z.; Farid, A.; Ahmad, S.; Muzammal, M.; Al Mohaini, M.; Alsalman, A. J.; Hawaj, M. A.; Alhashem, Y. N.; Alsaleh, A. A.; Almusalami, E. M.; Maryam, M.; Khan, M. A. In Silico Analysis Identified Putative Pathogenic Missense nsSNPs in Human SLITRK1 Gene. *Genes* **2022**, *13*, No. 672.

(32) Apweiler, R.; Bairoch, A.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L. S. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* **2004**, *32*, D115−D119.

(33) Gao, J.; Aksoy, B. A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S. O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; Cerami, E.; Sander, C.; Schultz, N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signaling* **2013**, *6*, No. pl1.

(34) Bamford, S.; Dawson, E.; Forbes, S.; Clements, J.; Pettett, R.; Dogan, A.; Flanagan, A.; Teague, J.; Futreal, P. A.; Stratton, M. R.; Wooster, R. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **2004**, *91*, 355−358.

(35) Landrum, M. J.; Lee, J. M.; Benson, M.; Brown, G. R.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; Karapetyan, K.; Katz, K.; Liu, C.; Maddipatla, Z.; Malheiro, A.; McDaniel, K.; Ovetsky, M.; Riley, G.; Zhou, G.; Holmes, J. B.; Kattman, B. L.; Maglott, D. R. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **2018**, *46*, D1062−D1067.

(36) Karczewski, K. J.; Francioli, L. C.; Tiao, G.; Cummings, B. B.; Alföldi, J.; Wang, Q.; Collins, R. L.; Laricchia, K. M.; Ganna, A.; Birnbaum, D. P.; Gauthier, L. D.; Brand, H.; Solomonson, M.; Watts, N. A.; Rhodes, D.; Singer-Berk, M.; England, E. M.; Seaby, E. G.; Kosmicki, J. A.; et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **2020**, *581*, 434−443.

(37) Ng, P. C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **2003**, *31*, 3812−3814.

(38) Adzhubei, I. A.; Schmidt, S.; Peshkin, L.; Ramensky, V. E.; Gerasimova, A.; Bork, P.; Kondrashov, A. S.; Sunyaev, S. R. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248−249.

(39) Mathe, E.; Olivier, M.; Kato, S.; Ishioka, C.; Hainaut, P.; Tavtigian, S. V. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res.* **2006**, *34*, 1317−1325.

(40) Choi, Y.; Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **2015**, *31*, 2745−2747.

(41) Thomas, P. D.; Campbell, M. J.; Kejariwal, A.; Mi, H.; Karlak, B.; Daverman, R.; Diemer, K.; Muruganujan, A.; Narechania, A. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **2003**, *13*, 2129−2141.

(42) Mitchell, A. L.; Attwood, T. K.; Babbitt, P. C.; Blum, M.; Bork, P.; Bridge, A.; Brown, S. D.; Chang, H. Y.; El-Gebali, S.; Fraser, M. I.; Gough, J.; Haft, D. R.; Huang, H.; Letunic, I.; Lopez, R.; Luciani, A.; Madeira, F.; Marchler-Bauer, A.; Mi, H.; Natale, D. A.; et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **2019**, *47*, D351−D360.

(43) Ashkenazy, H.; Erez, E.; Martz, E.; Pupko, T.; Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **2010**, *38*, W529−W533.

(44) Berezin, C.; Glaser, F.; Rosenberg, J.; Paz, I.; Pupko, T.; Fariselli, P.; Casadio, R.; Ben-Tal, N. ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics* **2004**, *20*, 1322−1324.

(45) Capriotti, E.; Fariselli, P.; Casadio, R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* **2005**, *33*, W306−W310.

(46) Venselaar, H.; Te Beek, T. A.; Kuipers, R. K.; Hekkelman, M. L.; Vriend, G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinf.* **2010**, *11*, No. 548.

(47) McGinnis, S.; Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **2004**, *32*, W20−W25.

(48) Krieger, E.; Nabuurs, S. B.; Vriend, G. Homology modeling. *Methods Biochem. Anal.* **2003**, *44*, 509−523.

(49) Krieger, E.; Vriend, G. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics* **2014**, *30*, 2981−2982.

(50) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(51) Kim, D. E.; Chivian, D.; Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **2004**, *32*, W526−W531.

(52) Heo, L.; Park, H.; Seok, C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* **2013**, *41*, W384−W388.

(53) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.

(54) Eisenberg, D.; Lüthy, R.; Bowie, J. U. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.* **1997**, *277*, 396−404.

(55) Benkert, P.; Tosatto, S. C.; Schomburg, D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins* **2008**, *71*, 261−277.

(56) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, 2002.

(57) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(58) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950−1958.

(59) Ouyang, Y.; Zhao, L.; Zhang, Z. Characterization of the structural ensembles of p53 TAD2 by molecular dynamics simulations with different force fields. *Phys. Chem. Chem. Phys.* **2018**, *20*, 8676−8684.

(60) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952−962.

(61) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(62) Kuusisto, K. M.; Bebel, A.; Vihinen, M.; Schleutker, J.; Sallinen, S. L. Screening for BRCA1, BRCA2, CHEK2, PALB2, BRIP1, RAD50, and CDH1 mutations in high-risk Finnish BRCA1/2-founder mutation-negative breast and/or ovarian cancer individuals. *Breast Cancer Res.* **2011**, *13*, No. R20.

(63) Mighri, N.; Hamdi, Y.; Boujemaa, M.; Othman, H.; Ben Nasr, S.; El Benna, H.; Mejri, N.; Labidi, S.; Ayari, J.; Jaidene, O.; Bouaziz, H.; Ben Rekaya, M.; M'rad, R.; Haddaoui, A.; Rahal, K.; Boussen, H.; Boubaker, S.; Abdelhak, S. Identification of Novel BRCA1 and RAD50 Mutations Associated With Breast Cancer Predisposition in Tunisian Patients. *Front. Genet.* **2020**, *11*, No. 552971.

(64) Tommiska, J.; Seal, S.; Renwick, A.; Barfoot, R.; Baskcomb, L.; Jayatilake, H.; Bartkova, J.; Tallila, J.; Kaare, M.; Tamminen, A.; Heikkilä, P.; Evans, D. G.; Eccles, D.; Aittomäki, K.; Blomqvist, C.; Bartek, J.; Stratton, M. R.; Nevanlinna, H.; Rahman, N. Evaluation of RAD50 in familial breast cancer predisposition. *Int. J. Cancer* **2006**, *118*, 2911−2916.

(65) Osorio, A.; Endt, D.; Fernández, F.; Eirich, K.; de la Hoya, M.; Schmutzler, R.; Caldés, T.; Meindl, A.; Schindler, D.; Benitez, J. Predominance of pathogenic missense variants in the RAD51C gene occurring in breast and ovarian cancer families. *Hum. Mol. Genet.* **2012**, *21*, 2889−2898.

(66) Clague, J.; Wilhoite, G.; Adamson, A.; Bailis, A.; Weitzel, J. N.; Neuhausen, S. L. RAD51C germline mutations in breast and ovarian cancer cases from high-risk families. *PLoS One* **2011**, *6*, No. e25632.

(67) Loveday, C.; Turnbull, C.; Ramsay, E.; Hughes, D.; Ruark, E.; Frankum, J. R.; Bowden, G.; Kalmyrzaev, B.; Warren-Perry, M.; Snape, K.; Adlard, J. W.; Barwell, J.; Berg, J.; Brady, A. F.; Brewer, C.; Brice, G.; Chapman, C.; Cook, J.; Davidson, R.; Donaldson, A.; et al. Germline mutations in RAD51D confer susceptibility to ovarian cancer. *Nat. Genet.* **2011**, *43*, 879−882.

(68) Rivera, B.; Di Iorio, M.; Frankum, J.; Nadaf, J.; Fahiminiya, S.; Arcand, S. L.; Burk, D. L.; Grapton, D.; Tomiak, E.; Hastings, V.; Hamel, N.; Wagener, R.; Aleynikova, O.; Giroux, S.; Hamdan, F. F.; Dionne-Laporte, A.; Zogopoulos, G.; Rousseau, F.; Berghuis, A. M.; Provencher, D.; et al. Functionally Null RAD51D Missense Mutation Associates Strongly with Ovarian Carcinoma. *Cancer Res.* **2017**, *77*, 4517−4529.

(69) Yang, C.; Arnold, A. G.; Catchings, A.; Rai, V.; Stadler, Z. K.; Zhang, L. The RAD51D c.82G>A (p.Val28Met) variant disrupts normal splicing and is associated with hereditary ovarian cancer. *Breast Cancer Res. Treat.* **2021**, *185*, 869−877.

(70) Rodríguez-López, R.; Osorio, A.; Ribas, G.; Pollán, M.; Sánchez-Pulido, L.; de la Hoya, M.; Ruibal, A.; Zamora, P.; Arias, J. I.; Salazar, R.; Vega, A.; Martínez, J. I.; Esteban-Cardeñosa, E.; Alonso, C.; Letón, R.; Urioste Azcorra, M.; Miner, C.; Armengod, M. E.; Carracedo, A.; González-Sarmiento, R.; et al. The variant E233G of the RAD51D gene could be a low-penetrance allele in high-risk breast cancer families without BRCA1/2 mutations. *Int. J. Cancer* **2004**, *110*, 845−849.

(71) Marcelino, A. M.; Gierasch, L. M. Roles of beta-turns in protein folding: from peptide models to protein engineering. *Biopolymers* **2008**, *89*, 380−391.

(72) Truebestein, L.; Leonard, T. A. Coiled-coils: The long and short of it. *BioEssays* **2016**, *38*, 903−916.

(73) Dong, C.; Wei, P.; Jian, X.; Gibbs, R.; Boerwinkle, E.; Wang, K.; et Liu, X. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **2015**, *24*, 2125−2137.

(74) Kerr, I. D.; Cox, H. C.; Moyes, K.; Evans, B.; Burdett, B. C.; van Kan, A.; McElroy, H.; Vail, P. J.; Brown, K. L.; Sumampong, D. B.; Monteferrante, N. J.; Hardman, K. L.; Theisen, A.; Mundt, E.; Wenstrup, R. J.; Eggington, J. M. Assessment of in silico protein sequence analysis in the clinical classification of variants in cancer risk genes. *J. Community Genet.* **2017**, *8*, 87−95.

(75) Ernst, C.; Hahnen, E.; Engel, C.; Nothnagel, M.; Weber, J.; Schmutzler, R. K.; Hauke, J. Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Med. Genom.* **2018**, *11*, No. 35.

(76) Pshennikova, V. G.; Barashkov, N. A.; Romanov, G. P.; Teryutin, F. M.; Solov'ev, A. V.; Gotovtsev, N. N.; Nikanorova, A. A.; Nakhodkin, S. S.; Sazonov, N. N.; Morozov, I. V.; Bondar, A. A.; Dzhemileva, L. U.; Khusnutdinova, E. K.; Posukh, O. L.; Fedorova, S. A. Comparison of Predictive *In Silico* Tools on Missense Variants in *GJB2, GJB6,* and *GJB3* Genes Associated with Autosomal Recessive Deafness 1A (DFNB1A). *Sci. World J.* **2019**, *2019*, No. 5198931.

(77) Montenegro, L. R.; Lerário, A. M.; Nishi, M. Y.; Jorge, A. A. L.; Mendonca, B. B. Performance of mutation pathogenicity prediction tools on missense variants associated with 46,XY differences of sex development. *Clinics* **2021**, *76*, No. e2052.

(78) Thusberg, J.; Olatubosun, A.; Vihinen, M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum. Mutat.* **2011**, *32*, 358−368.

(79) Galehdari, H.; Saki, N.; Mohammadi-asl, J.; Rahim, F. Meta-analysis diagnostic accuracy of SNP-based pathogenicity detection tools: a case of UTG1A1 gene mutations. *Int. J. Mol. Epidemiol. Genet.* **2013**, *4*, 77−85.

(80) Leong, I. U.; Stuckey, A.; Lai, D.; Skinner, J. R.; Love, D. R. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. *BMC Med. Genet.* **2015**, *16*, No. 34.

(81) Somyajit, K.; Subramanya, S.; Nagaraju, G. Distinct roles of FANCO/RAD51C protein in DNA damage signaling and repair: implications for Fanconi anemia and breast cancer susceptibility. *J. Biol. Chem.* **2012**, *287*, 3366−3380.

(82) Jiajia, C.; Bairong, S. Computational Analysis of Amino Acid Mutation: A Proteome Wide Perspective. *Curr. Proteomics* **2009**, *6*, 228−234.

(83) Steward, R. E.; MacArthur, M. W.; Laskowski, R. A.; Thornton, J. M. Molecular basis of inherited diseases: a structural perspective. *Trends Genet.* **2003**, *19*, 505−513.

(84) Ye, Y.; Li, Z.; Godzik, A. Modeling and analyzing three-dimensional structures of human disease proteins. *Biocomputing 2006* **2006**, 439−450.