Received: 11 October 2025 | Revised: 01 December 2025 | Accepted: 02 January 2025
DOI: https://doi.org/10.69547/TSFJB.040101

TSF JOURNAL OF BIOLOGY

EISSN: 3006-0478
PISSN: 3006-0869

Website link: www.tsfjb.com

Computational Screening of Non-Synonymous SNPs from Human *SQLE* Gene to Assess Their Association with Multiple Cancer Types

Mahnoor Mushtaq, Sana Batool, Naeem Mahmood Ashraf, Arslan Hamid, Hina Batool

To cite: Mushtaq M, Batool S, Ashraf NM, Hamid A, Batool H. Computational Screening of Non-Synonymous SNPs from Human *SQLE* Gene to Assess Their Association with Multiple Cancer Types. TSF Journal of Biology.2026; 4(1):01-24.

Article QR

Published online: 21 January 2026
Submit your article to this journal: https://tsfjb.com/ojs/index.php/tsfjb/about/submissions
View archive: https://tsfjb.com/ojs/index.php/tsfjb/issue/archive

Research Article

Computational Screening of Non-Synonymous SNPs from Human *SQLE* Gene to Assess Their Association with Multiple Cancer Types

Mahnoor Mushtaq[1] | Sana Batool[2] | Naeem Mahmood Ashraf[3] | Arslan Hamid[4] | Hina Batool[5]*

[1]Department of Life Sciences, University of Management and Technology, Lahore, Pakistan

[2]School of Biological Sciences, University of the Punjab, Lahore, Pakistan

[3]Department of Biochemistry and Biotechnology, University of the Punjab, Lahore, Pakistan

[4]University of Bonn, Germany

[5]Department of Biochemistry and Molecular Biology, Dalhousie University, Nova Scotia, Canada

***Correspondence**
Hina Batool
Email: hn701981@dal.ca

Keywords:
Cancer, cholesterol biosynthesis, functional assessments, non-synonymous SNPs, molecular dynamics simulations, mutants

ABSTRACT

Background: Cancer is a complex disease triggered by various genetic alterations. Genes act as central drivers of oncogenesis, with mutations disrupting their normal function to initiate this process. Recently, SQLE, an oncogene central to cholesterol and steroid biosynthesis, has been repeatedly implicated across multiple cancer types by genome-wide association studies. **Objectives:** This study aims to identify pathogenic non-synonymous SNPs (nsSNPs) in SQLE and evaluate their potential association with cancer using an integrated computational pipeline. **Methods:** Homology-based tools (SIFT, PROVEAN, PolyPhen-2, Mutation Assessor), machine learning-based tools (SuSPect, SNPs&GO, PhD-SNP, MutPred, PMut), and consensus-based predictors (Meta-SNP, PredictSNP, Condel, PON-P2) were used to screen nsSNPs and identify variants predicted to be pathogenic within evolutionarily conserved regions. Protein stability was assessed using sequence-based (I-Mutant, MUpro, iStable2.0) and structure-based (SDM, DynaMut, PoPMuSiC, FoldX) tools. Molecular dynamics simulations were performed to examine the structural and functional consequences of the prioritized mutations. **Results:** Two variants, E153G and L280R, were predicted as highly pathogenic and destabilizing, and both mapped to conserved regions of the SQLE protein. Molecular dynamics simulations showed altered conformational behavior in the mutant models relative to the wild-type structure, consistent with a potential disruption of protein-protein interactions. **Conclusions:** E153G and L280R are computationally predicted to affect SQLE structure, stability, and interaction potential, suggesting a plausible contribution to oncogenic pathways across multiple cancers. These findings need further experimental validation and pathway mapping before any functional or clinical significance can be established.

1. INTRODUCTION

Cancer is a leading cause of death globally [1]. Recently, in 2020, the World Health Organization (WHO) reported 10 million deaths due to cancer, among which breast cancer, lung cancer, and colon rectum cancer were associated with the highest mortality rates [2].Genetic mutations play a central role in different cancer types, changing the expression of key proteins involved in the cancer-associated pathways or affecting their structural conformations. Most of these mutations are missense single-nucleotide

polymorphisms (SNPs) that lead to changes in the protein's amino acid sequence, resulting in structural and functional destabilization[3]. Genetic mutations are, therefore, the main drivers to induce the various types of cancer [4]. So far, several genes have been reported to contribute to cancer development, while the involvement of novel genes is constantly being identified [4].

Recently, the *SQLE* gene has been under focus for its involvement in diverse cancer types [5], including breast cancer, colorectal cancer, hepatocellular carcinoma [6], [7], [8], esophageal squamous cell carcinoma, and prostate cancer [9], [10]. *SQLE* encodes squalene epoxidase, which catalyzes one of the rate-limiting steps in cholesterol biosynthesis and is considered the therapeutic target [9]. Squalene epoxidase is a FAD-linked monooxygenase that oxidizes squalene to 2,3-oxidosqualene in the presence of NADPH and molecular oxygen [11], [12]. The 2, 3-oxidosqualene then converts to lanosterol and cholesterol. Cholesterol is crucial for membrane formation, cell proliferation, and cell differentiation. Moreover, its involvement in tumor cell proliferation, survival, and treatment resistance has also been reported. The genetic mutations in *SQLE* may contribute to cancer by upregulating the activity of squalene epoxidase, which can act as a prognostic factor for metastasis [13].

The advancements in the field of computational biology have made it possible to identify the pathogenic and damaging mutations for a candidate gene. Furthermore, the possible phenotypic impacts of these mutations can also be assessed. The genetic elements predicted to have an impact on the protein structure and function would be effective in establishing the underlying molecular mechanism of various diseases. However, these mutations need experimental validation before their use in the development of diagnostic and management strategies [14]. As, addressing this research gap is mandatory before these variants applied into diagnostic or clinical development.

The present study involves the screening of potentially deleterious missense SNPs from the *SQLE* gene using various computational algorithms. The cancer-associated mutations were screened and subjected to 3D structure modeling followed by molecular docking and MD simulation analysis. This study presents a pipeline through which highly significant missense SNPs associated with cancer can be screened. The study would, therefore, contribute to the development of early diagnostic and management strategies for cancer.

2. METHODOLOGY

All steps involved in the completion of this study are shown in Figure 1.

2.1 DATA RETRIEVAL

The protein sequence of *SQLE* was obtained from Uniprot (https://www.uniprot.org/) using accession ID Q14534**.** The missense SNPs associated with the *SQLE* were retrieved from the dbSNP database (https://www.ncbi.nlm.nih.gov/snp/). The dbSNP categorizes the SNPs into frameshift, deletion, indel, insertion, initiator codon variant, intron, missense, non-coding transcript variant, and synonymous [15]. Of all these mutations, only the missense mutations were selected for the downstream analysis because these mutations change the amino acid sequence of the protein, affecting its 3D structure and function [16].

2.2 FUNCTIONAL ASSESSMENT OF SELECTED MUTATIONS

For the functional assessments of the selected missense mutations from SQLE, three different algorithms; sequence homology, machine learning, and consensus-based algorithms were

considered. More than one computational tool was used for each of these algorithms to increase the prediction accuracies. The missense mutations commonly marked as deleterious by all tools based on one algorithm were considered for further screening through the next algorithm. The computational tools used in our study along with their algorithms, threshold, and predicted outcomes are enlisted in Table I. If variant has score less than the respective threshold mentioned in Table I, this variation is neutral having tolerated effect of substitution and if it exceeds the threshold score the variation determined to be damaging having effect on protein structure and function.

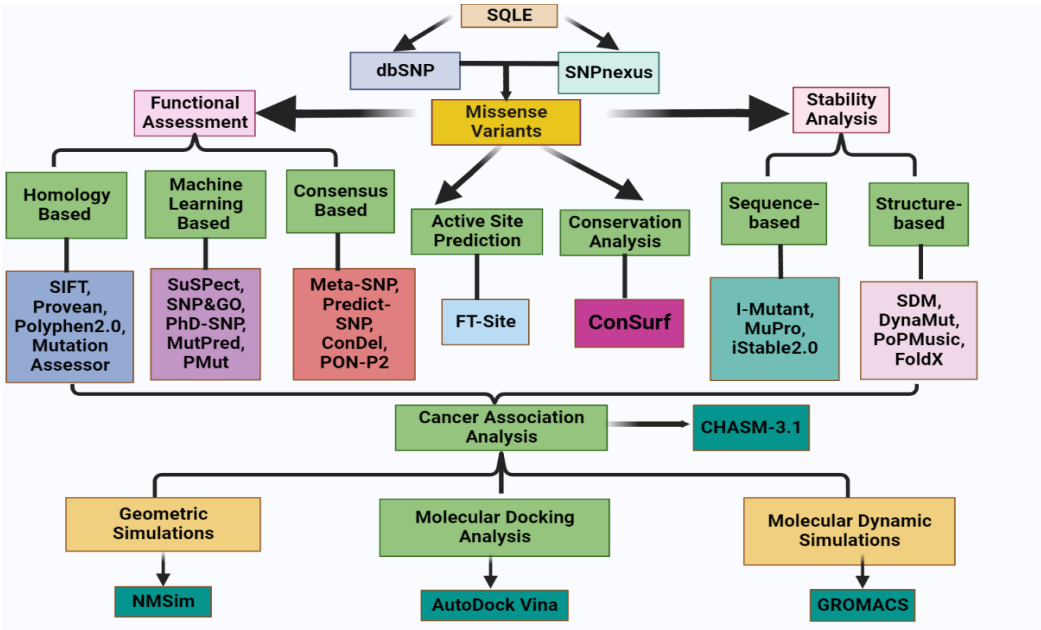

Figure 1. Schematic representation of stepwise evaluation of missense SNPs of human gene SQLE by various computational tools

Table 1. All tools with respective algorithms, threshold, and predicted outcomes

Type of assessment	Algorithms	Tools	Prediction Methods	Threshold	Prediction Outcome
Functional assessments of mutations	Sequence Homology Based	SIFT	Alignment Score	0.05	Tolerated/ Damaging
		Provean	Alignment Score	-2.5	Neutral/Damaging
		PolyPhen 2	Naïve Bayes	0.2	Possibly Damaging/ Probably Damaging
		Mutation Assessor	Multiple Sequence Alignment	0	Neutral/Damaging

Type of assessment	Algorithms	Tools	Prediction Methods	Threshold	Prediction Outcome
	Machine Learning Based	SuSpect	Support Vector Machine	50 (color code)	Neutral/Disease
		SNP&GO	Support Vector Machine	0.5	Neutral/Disease
		PhD-SNP	Support Vector Machine	0.5	Neutral/Disease
		MutPred	AlphaMax Algorithm	0.8	Neutral/Disease
		PMUT	Neural Network	80	Neutral/Disease
	Consensus-Based	Meta-SNP	Consensus Deleteriousness Score	0.5	Neutral/Disease
		Predict SNP	Consensus Score	Color Code	Neutral/Disease
		ConDel	Consensus Deleteriousness Score	0.5	Neutral/Disease
		PON-P2	Random Forest	0.795	Neutral/Pathogenic
Evolutionary conservation analysis	Homology based	CONSURF	Empirical Bayesian Method	8	Scoring 1-9 for low to high conservation
Protein stability analysis	Sequence-Based	I-Mutant3.0	Support Vector Machine (SVM)	0	Stable/Unstable
		MuPro	ANN, SVM	0	Stable/Unstable
		iStable2.0	Meta Predictor	0	Increase/Decrease
	Structure-Based	SDM	ESST and mCSM (mutation cutoff scanning matrix)	0	Stable/Unstable
		DynaMut	Meta Predictor based on normal mode analysis	0	Stabilizing/ Destabilizing
		PoPMusic	ANN and Statistical Potential	0	Stabilizing/ Destabilizing
		FoldX	Linear Regression	0	Stabilizing/ Destabilizing

2.2.1 Homology-based Analysis

The missense mutations obtained from dbSNP were initially screened using four sequence homology-based tools, SIFT (Sorting Intolerant From Tolerant) (https://sift.bii.a-star.edu.sg), PROVEAN (The Protein Variation Effect Analyzer) (http://provean.jcvi.org), POLYPHEN2 (Polymorphism Phenotyping V2) (http://genetics.bwh.harvard.edu/pph2/), and Mutation Assessor (http://mutationassessor.org). The sequence homology algorithm is based on multiple sequence alignment to classify the missense variations as tolerated or damaging to the protein [17]. The mutations commonly marked as deleterious by these four

tools were used for the screening through the next algorithm [18].

2.2.2 Machine Learning-based Analysis

The missense variants characterized as damaging by four homology-based tools were subjected to tools based on supervised learning algorithms for further screening. For machine learning based assessments, SuSPect (http://www.sbg.bio.ic.ac.uk/suspect), SNP&GO (http://snps.biofold.org/snps-and-go), PhD-SNP (https://snps. biofold.org/phd-snp/phd-snp.html), MutPred (http://mutpred.mutdb.org), and PMut (http://mmb.irbbarcelona.org/PMut/) were used. These tools are based on an algorithm that is trained on large datasets to differentiate potentially deleterious variants from benign [19]. All the variants having a predicted score more than the threshold are considered to be disease-associated while those having a score less than the threshold are predicted to be neutral. Again, the mutations commonly marked as deleterious by all these tools were selected for the downstream analysis.

2.2.3 Consensus-based Analysis

After the sequence homology and machine learning-based analysis, the selected pathogenic missense mutations were further screened using a consensus-based algorithm that classifies the missense mutations as damaging or neutral by applying multiple prediction methods. Four tools including Meta-SNP (https://snps.biofold.org/meta-snp/), Predict-SNP (https://loschmidt.chemi.muni.cz/predictsnp, CONsensus DELeteriousness (ConDel) (https://bbglab.irbbarcelona.org/fannsdb/), and PON-P2 (http://structure.bmc.lu.se/PON-P2/) were used consensus-based analysis [20], [21], [22], [23], [24]. After the functional assessment of selected missense SNVs through the consensus-based algorithm the missense mutations commonly marked as deleterious by all these tools were selected.

2.3 Evolutionary Conservation Analysis

As the missense mutations occurring in the evolutionarily conserved regions of the protein have a significant impact on the structure, folding, and function of the protein, therefore, the selected pathogenic missense SNVs were analyzed by the ConSurf (http://consurf.tau.ac.il) web server to identify the variants that are present in the evolutionarily conserved regions of SQLE. ConSurf identifies the mutations occurring in the conserved regions of protein by comparing the sequence with the homologous sequences and assigning the residues a conservation score ranging from 1 to 9 with the conserved residues having a higher conservation score [25].

2.4 Protein Stability Analysis

After the screening of pathogenic and evolutionary conserved missense mutations, the effect of these mutations on protein stability was assessed using the protein sequence and structure-based stability analysis tools. Firstly, support vector machine algorithm-based servers used to perform the sequence-based stability analysis by I-Mutant (https://folding.biofold.org/cgi-bin/i-mutant2.0.cgi), MUpro (http://mupro.proteomics.ics.uci.edu) and iStable2.0 (http://ncblab.nchu.edu.tw/iStable2). These tools use the amino acid sequence of proteins for the stability analysis. The second algorithm considers the structure-based protein stability approach to analyze the structural features, flexibility, and thermo-dynamical differences between the native and mutants [26]. For the structure-based protein stability analysis, SDM (http://marid.bioc.cam.ac.uk/sdm2/), DynaMut (http://biosig.unimelb.edu.au/dynamut/), PoPMusic (https://soft.dezyme.com), and FOLDX (http://foldxsuite.crg.eu/) were used.

2.5 Cancer-association Analysis

CHASM-3.1 was used to filter the variants associated with multiple cancer types by server

CRAVET (www.cravat.us), which evaluates variants in VEST4 and CHASM-3.1. It gives a score as a p-value to all variants where less than 0.05 for CHASM-3.1 predicts the probability of variants as pathogenic passengers of cancer, while more than 0.05 p-value predicts it to be benign cancer-associated variants [27].

2.6 Active Site Prediction

After evolutionary conservation analysis, the amino acid residues belonging to the active site of SQLE were identified using the FTSite server (https://ftsite.bu.edu) [28], [29].

2.7 Structural Modeling of Native and Mutated Protein Models

The 3D model of SQLE was fetched from PDB using PDB ID 6C6R (10.2210/pdb6C6R/pdb), that was validated by RAMPAGE (https://zlab.umassmed.edu › rama). While, mutant models were generated through UCSF Chimera (https://www.cgl.ucsf.edu/chimera/), followed by the energy minimization and stabilization of protein models by YASARA force field. After structure modeling, the wild-type and mutated protein models were superimposed using BIOVIA Discovery Studio to mark the structural variation among these models [30].

2.8 Geometric Simulation Analysis

The conformational transitions in native and mutated SQLE models were determined using the NMSim server (http://www.nmsim.de). The conformational changes are determined in three different steps. Initially, the server uses a graph theoretical approach to disintegrate the protein into flexible and rigid clusters, then dynamic properties of proteins are explored using a coarse-grained approach. Finally, the coarse-grained model is further extended by introducing new conformations in the protein. NMSim provides the results in the form of a trajectory file for all generated conformations and the graphs showing root mean square fluctuations (RMSF), root mean square deviations (RMSD), radius of gyration (Rg), and H-bonding pattern of protein models [31].

2.9 Molecular Docking Analysis and MD Simulation Analysis

The binding affinities of native and mutated SQLE models with the substrate squalene and coenzyme NADPH were assessed using molecular docking analysis. For this purpose, the structures of squalene and NADPH were downloaded from PubChem [32]. The protein models were docked with squalene and NADPH using AutoDock Vina followed by the assessments of binding affinities in terms of kJ/mol [33]. The docked complexes with lower binding energies depict the higher binding affinities with the substrate [34].

The docked complexes of native and mutated protein models were used for molecular dynamics simulation using a GROMACS software package. The dynamic pattern of these complexes was compared by applying the CHARMM36 force field on proteins and CGenFF on targeted compounds. The system was equilibrated for 100 ps followed by a 5 ns production phase. Eventually, the root mean square deviation (RMSD), the number of hydrogen bonds (HB), and compound interaction energies (Coul-SR and LJ-SR) were calculated for these models [35].

2.10 Pathway Analysis

Finally, the pathway analysis was performed using a literature review, KEGG (http://www.genome.jp/kegg/), and Reactome (https://reactome.org) databases [36], [37].

3. RESULTS

3.1 Data Retrieval

The protein sequence of SQLE comprising 574 amino acids was retrieved from the UniProt database. For this sequence, a total of 450

missense, 198 synonymous mutations, 14 inframe deletions, 3 initiator codon variants, and 1 inframe insertion were reported in dbSNP (Supplementary file, S1). Among all these mutations the 450 missense variations were selected for the downstream analysis.

3.2 Functional Assessment of Selected Mutation

For screening the pathogenic missense mutations, the selected 450 mutations were subjected to sequential evaluation by multiple computational tools based on different algorithms. The most damaging disease-causing missense SNPs were filtered, while neutral variants were removed at each step.

3.2.1 Homology-based Analysis

The missense mutations selected from the dbSNP were initially subjected to the sequence homology-based tools (SIFT [17], [38], Provean [39], PolyPhen2.0 [40], and Mutation Assessor [38] for screening. Among these tools, SIFT and PolyPhen2.0 filtered 221 missense mutations as deleterious, likewise, 176 and 246 mutations were marked to be deleterious by Provean and Mutation Assessor, respectively. A total of 105 variants commonly marked as deleterious by all four tools were considered for further analysis ignoring all other mutations.

3.2.2 Machine Learning-based Analysis

The 105 deleterious missense mutations identified by homology-based tools were further subjected to machine learning-based tools including SuSPect [19], SNP&GO [41], PhD-SNP [42], MutPred [43], and PMut [44]. Among these databases SuSPect marked all variants as deleterious, while SNP&GO, PhD-SNP, MutPred, and PMut, respectively marked 54, 82, 34 and 49 mutations as pathogenic at the set threshold (Table 2.1). The 25 missense mutations were commonly marked as deleterious by all machine learning-based tools and were subjected to consensus-based analysis.

3.2.3 Consensus-based Analysis

The consensus classifier are meta-predictors having background algorithm based on combination of both sequence homology and machine learning. META-SNP [45], Predict SNP [46], ConDel [47], and PON-P2 [48] were used for the consensus-based assessments of the 25 missense variations. Meta-SNP, Predict SNP, and ConDel marked all 25 mutations as pathogenic, while PON-P2 pointed to 18 variants as deleterious. Finally, the 18 missense mutations being commonly reported as pathogenic by all consensus-based tools were selected for further characterization.

3.3 Evolutionary Conservation Analysis

As evolutionarily conserved regions have a crucial impact on protein structure and function, the missense mutations in these regions may have significant detrimental effects on protein structure and function [49]. For the screening of evolutionarily conserved mutations, the ConSurf web server was used, which marked 13 out of 18 missense mutations as occurring in the conserved regions of squalene epoxidase (Supplementary file; S2, S3).

3.4 Protein Stability Analysis

The thirteen pathogenic and evolutionarily conserved missense mutations were further evaluated using a combination of sequence and structure-based protein stability predictors. The sequence-based stability predictors, I-Mutant and MuPro, marked twelve, while iStable2.0 marked ten mutants as destabilizing mutations are mentioned in Table II. A total of ten missense mutations were commonly marked, as destabilizing mutations by all three sequence-based protein stability predictors, which were further examined for their impact on protein structure using structure-based protein stability prediction tools. Among the

four structure-based stability predictors, FoldX and PoPMusic reported all ten variants as destabilizing, while SDM and DynaMut, respectively, marked 8 and 9 variants as destabilizing are listed in Table III. Finally, the six common mutations were selected as the destabilizing mutations are mentioned in graph in Figure 2.

3.5 Cancer Association Analysis

To check the possible involvements of six selected mutations in cancer, the CRAVAT server was used. The prediction method CHASM-3.1 showed the significant involvement of the selected mutations in head and neck cancer, breast cancer, colon cancer, rectum cancer, lung cancer, and prostate cancer as shown in Figure 3; B. The involvement of a mutation with a particular cancer type was based on the CHASM p-value, where a p-value less than 0.05 indicated a strong correlation with a particular cancer type.

Table 2. List of sequence-based destabilizing nsSNPs of SQLE gene

Sr#	Mutants	I-Mutant (kcal/mol)	MuPro (kcal/mol)	iStable2.0 (kcal/mol)
1	E153G	-0.17	-1.55589	-1.14
2	L280R	-2.66	-1.69711	-2.083
3	G286R	-1.44	-0.65403	-0.727
4	R343P	-1.24	-1.16084	-0.5894129
5	P415S	-1.89	-1.01433	-1.902
6	G419R	-0.90	-0.68242	-0.0401
7	R458T	-1.20	-1.54186	-0.661
8	L473H	-3.07	-2.17646	-2.18
9	P516R	-1.74	-0.98577	-0.393
10	H522R	-0.04	-1.05962	-0.043

*These enlisted ten mutants predicted to be destabilizing by I-Mutant, MuPro, and iStable2.0 having outcome as DDG values in kcal/mol, where negative values indicating the mutants to be destabilizing.

Table 3. List of structure-based destabilizing nsSNPs of SQLE gene

Sr#	rsID	Mutants	SDM (kcal/mol)	PoPMusic (kcal/mol)	DynaMut (kcal/mol)	FoldX (kcal/mol)
1	rs376060828	E153G	-1.26	0.19	-1.204	3.328
2	rs768230188	L280R	-2.97	2.11	-1.359	4.267
3	rs1815000059	G286R	-2.48	1.87	-0.154	4.99
4	rs1815004106	R343P	-2.66	0.89	-0.927	1.88
5	rs1471656048	R458T	-1.79	1.98	-2.002	2.54
6	rs1815150137	L473H	-0.6	1.65	-0.921	2.56

*These enlisted six mutants predicted to be destabilizing by SDM, PoPMusic, DynaMut, and FoldX having outcome as DDG values in kcal/mol, where negative values indicating the mutants to be destabilizing.

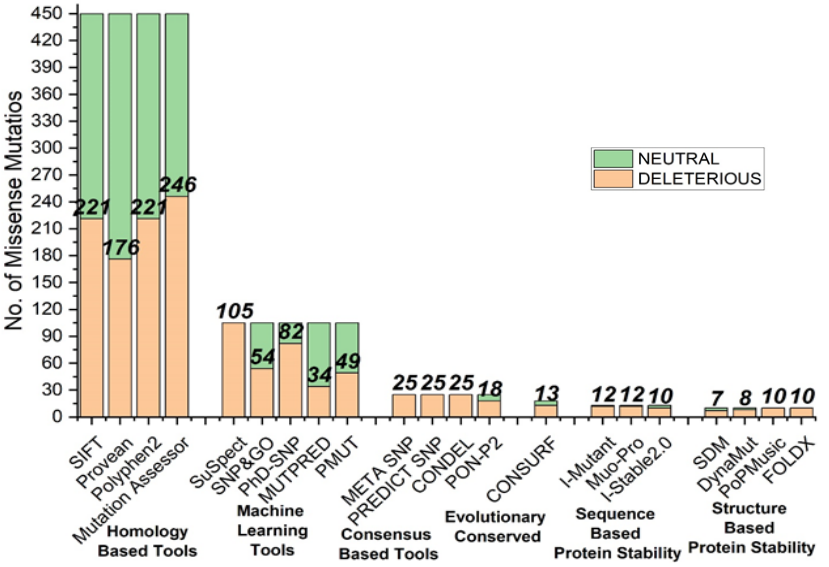

Figure 2. The screening of deleterious nsSNPs of SQLE by a combination of various computational tools such as functional assessment, conservation analysis, and stability assessment of all variants that are categorized according to their algorithms. The pathogenic nsSNPs selected by the first step were subjected to the next analysis. The stepwise analysis by these algorithms shortlisted only the potentially deleterious mutations for the final analysis

3.6 Active Site Analysis

The enzyme squalene epoxidase is a flavoprotein with covalently attached FAD. The active site of this enzyme contains two binding sites; one for substrate squalene and the second for coenzyme NADPH [12]. The active site analysis by FTSite shows the significant amino acids along with their mutations (E153G, G286R, and P415S) belong to the active site SQLE (Supplementary file, S4). Finally, among them, E153G and G286R were further analyzed to predict their impact on the protein structure, conformation, and protein-protein interactions. So, E153G was selected because of its presence in active sites and strong association with most cancer types.

3.7 Structural Modeling of Native and Mutated Protein Models

The tertiary structure of SQLE was retrieved from the PDB (https://www.rcsb.org) using PDB ID 6C6R. The structure of SQLE is a homodimer with each chain attached to the FAD (Supplementary file, S5). The 3D structures of two mutants were generated by UCSF Chimera followed by energy minimization and structure validation through RAMPAGE. Both mutant structures showed 98% residues in the allowed regions of the Ramachandran plots.

3.8 Geometric Simulation Analysis

The conformational transitions of native and mutated SQLE models were assessed using an online geometric simulation tool NMSim. The conformational transitions in the mutated models were assessed in terms of root mean square deviations (RMSD) and root mean square fluctuations (RMSF). The mutated protein models showing significant deviations from the native protein conformations were selected as the conformationally unstable

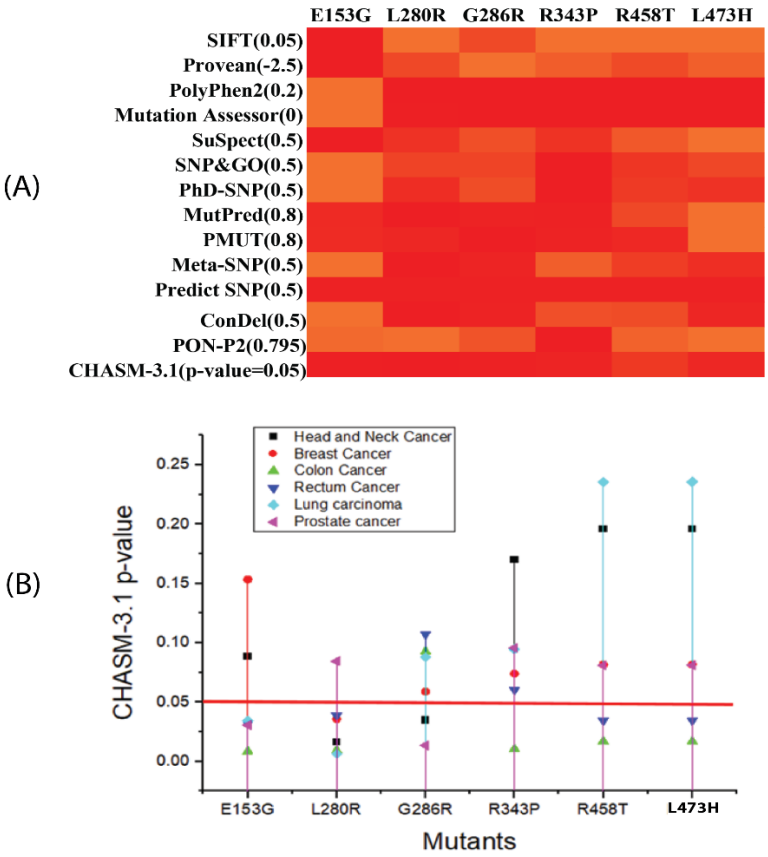

Figure 3. (A) Categorical evaluation of nsSNPs as cancer driver variants. The assessment tools are mentioned on the y-axis with the user-defined threshold of each tool mentioned in parenthesis while the selected six variants scrutinized from this comprehensive computational study are shown on the x-axis. In this heat map, the red color represents all the variants that have scored more than the modified threshold, and therefore, are marked as pathogenic mutations, while the increase in color intensity represents the highest score, and the decrease in score depicts the decrease in color intensity. (B) Cancer-associated variants by CHASM3.1 concerning their p-values. The selected six variants have involvement in multiple cancer types where the different shapes representing different types of cancer majorly drive due to mutation in the SQLE gene. While E153G and L280R indicate maximum cancer involvement in more cancer types having CHASM-3.1 predicted score as p-value less than 0.05. The CHASM-3.1 p-value of less than 0.05 indicates significant involvement in cancer.

models. The RMSD and RMSF analysis of two selected mutants showed significant variations from the wild-type model as shown in Figure 5 (A, B).

3.9 Molecular Docking and MD Simulation Analysis

Initially, the 3D structure of the native model of SQLE was docked sequentially with

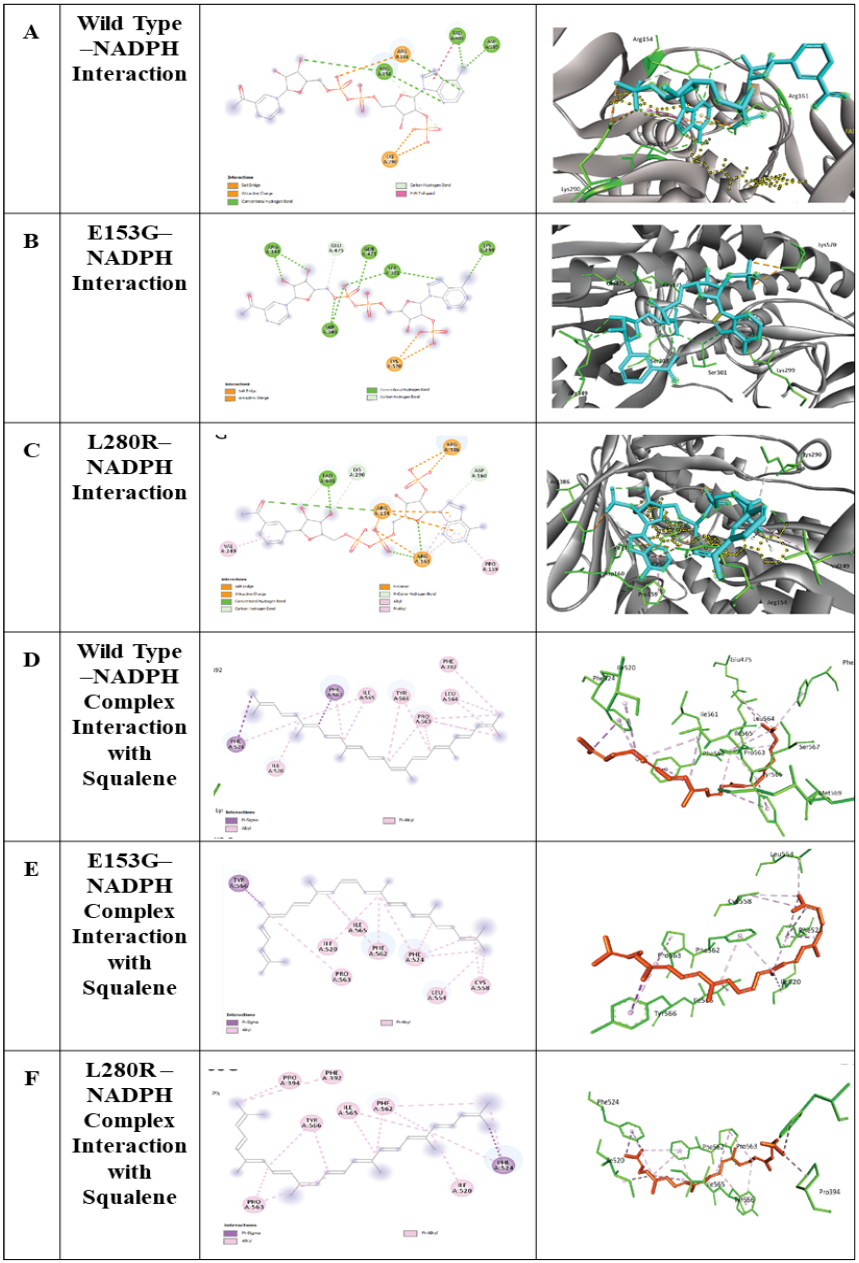

Figure 4. 2D and 3D Docked complexes of wild-type and mutant models to compare inter- and intra-residual interactions. The interacting residues are highlighted in green, cofactor FAD in yellow, NADPH in blue, and squalene in orange. Conventional hydrogen bonds are present in SQLE and NADPH interaction, while alkyl and pi-alkyl interactions are shown with the squalene interaction.

NADPH and squalene (Supplementary file, S5) to assess the molecular interactions and binding affinities. Later the molecular interaction pattern of the protein with ligand and cofactor was compared with that of the mutated proteins. Firstly, SQLE is docked with the cofactor NADPH having prominent interactions with hydrogen bonding (HB), attractive charge, and carbon-hydrogen bonding. The wildtype SQLE showed HB interactions with amino acids R154, D285, and FAD (cofactor) and attractive charge, and carbon HB interaction with D160, R161, V249, G286, L287, F288, L290, H384, R386, S387, M388 amino acids. While the first mutant E153 docked at the same pattern, almost all these interactions were lost and new interactions such as L299, S301, S303, R349, and Q471 were observed as HB while carbon-hydrogen bonds of L570, and E475 were observed. When the second mutant (L280R) docked with NADPH, some previous interactions were retained, some changed as mentioned in Table 4, and R161 occurred as a new HB.

The docked complex of SQLE and NADPH have a binding affinity of -7.1 kcal/mol for wild-type while the prominent change has occurred in the binding affinity of L280R i.e. -8.1 kcal/mol.

Secondly, the SQLE-NADPH complex was docked with substrate squalene, which has alkyl and pi-alkyl as stable interactions. The wildtype (SQLE-NADPH) complex is docked with squalene, where F392, I520, F524, F562, P563, L564, I565, and Y566 are involved in alkyl, pi-alkyl interactions. While in E153G, some interactions were retained, and L554, C558, and P563 were observed as new interactions, in L280R, only one new interaction, P394, occurred, as shown in Table 4. The docked complex of SQLE-NADPH-squalene revealed the change in binding affinity as wild type has -6.8 kcal/mol and the mutants E153G and L280R have shown no significant change as -6.8 and -6.7 kcal/mol, respectively. The details of all other molecular interactions are present in the figurative description in Figure 4. As hydrogen bonding significantly contributes to forming stable protein-ligand complexes, the comparison of hydrogen bonds is described in Table IV, which shows significant variations in the wild-type and mutant models' molecular interactions.

Furthermore, MD simulations were performed to validate the docking results as shown in Figure 5 (C, D). Wild-type and L280R have low RMSD values and show stable interactions, while E153G showed deviation from wild-type, showing its instability and disrupting interaction with NADPH. Wild-type SQLE-NADPH complex docked with squalene showed stable interaction, E153G again showed unstable interaction with squalene as well, while L280R showed moderate stability, indicating some disruptions. Conclusively, the E153G mutant showed disruptive interaction with both NADPH and squalene, suggesting it might affect the protein's overall functionality and stability, and L280R showed stable interactions with NADPH and comparatively less stable with squalene as compared to the wild-type. While the Gibbs free energy changes (ΔG) values as predicted binding affinity of E153G and L280R showed a stronger binding affinity to NADPH and squalene as compared to wild type as shown in Figure 5 (E, F). This indicates that a positively charged arginine at position 280 may tightly bind to NADPH in a non-productive manner, cause hindrance in release of cofactor or disrupt necessary conformational transitions. Conclusively, rather than stronger predicted binding, the overall catalytic activity may get impaired, giving a mechanistic rationale for damaging effect of this mutation. Therefore, the docking results must be interpreted cautiously and require experimental validation.

3.10 PATHWAY ANALYSIS

The expression of many cholesterogenic enzymes, including SQLE, is controlled by the transcription factor SREBP2 [11]. The protein product of SQLE, squalene monooxygenase, catalyzes squalene's oxygenation to squalene 2,3-epoxide and its further oxygenation to dioxidosqualene. The cancer-associated mutations normally increase the expression of the SQLE gene, leading to cholesterol accumulation [12]. The cholesterol accumulation promotes cancer cell growth by disrupting lipid metabolism, cholesterol biosynthesis, and regulation. The overexpression of SQLE is reported in many cancers, including lung cancer, breast cancer, esophageal cancer, colorectal cancer, nasopharyngeal cancer, leukemia, and head and neck cancer. The pathway, having a main reaction and its association with multiple cancer types, is depicted in Figure 6.

Table 4. Comparison of wild-type and mutant docking outcomes based on interactions

Wild-type	E153G	L280R
Hydrogen bonding Interaction of Protein with NADPH		
R154 D285 FAD A 601	L299, S301 S303, R349 Q471	R161 FAD 601
Attractive charge, Carbon Hydrogen bonding Interaction of Protein with NADPH		
D160, R161 V249, G286, L287, F288 L290, H384 R386, S387 M388	L570 E475	R154, P159 D160 R161 V249 L290 R386
Alkyl, pi-alkyl Interactions of SQLE-NADPH complex with Squalene		
F392 I520 F524 F562 P563 L564 I565 Y566	Y566 I520 F524 L554 C558 F562 P563 I565	F392 P394 F524 F562 P563 I565 Y566

*List of the amino acids involved in molecular interactions (hydrogen bonding, attractive charges, carbon-hydrogen bonding, alkyl, and pi-alkyl interaction) during complex formation in sequential docking. Amino acids written in black represent the interacting amino acids in the wild type, while red represents the changed interacting amino acids due to the mutation.

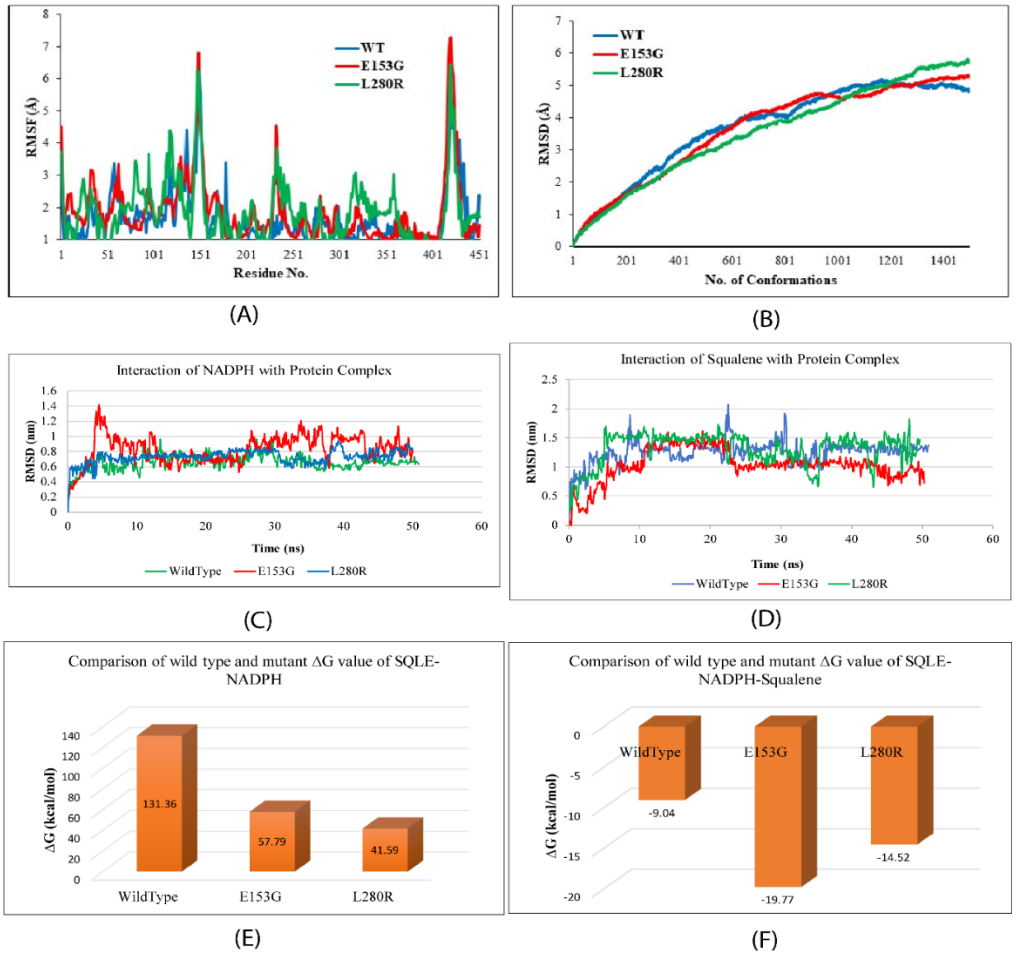

Figure 5. (A) Root Mean Square Fluctuation of native and mutant models by NMSim. Where blue-colored variations are of wild-type model while the selected two mutations are depicted in red and green for E153G and L280R. In this RMSF graph, E153G and L280R (with red and green color) show the highest variation as compared to the wild-type. (B) Root Mean Square Deviation of native and mutant models by NMSim. In this graph, the blue line represents the wild-type and the red and green color represent the root mean square deviation of E153G and L280R. Where both the mutants deviate and exceed the deviation slightly from the wild-type model. (C) Root Mean Square Deviation of Wild-type and Mutant Models by MD Simulations (GROMACS). Wild-type (green) is showing stable interaction with NADPH, E153G (red) is showing deviation from 0.5 to 1.5 nm suggesting its less stable interactions with more conformational changes in protein complex bound with NADPH, and L280R (blue) showing relatively stable interaction with NADPH. (D) Wild-type complex showed stable interaction with squalene, E153G reported the highest RMSD values with significantly unstable interactions, but L280R showed relatively stable interaction as compared to E153G. (E) Gibbs free energy change (ΔG) values of SQLE-NADPH complex and mutants model are shown as predicted binding affinity, where wild type

demonstrated a high value, while both mutants exhibited lower values. The strong binding affinity of mutants supports the increased enzymatic activity of SQLE due to these mutations. (F) ΔG values of SQLE-NADPH-Squalene complex and mutants model, where wild type showed stable interactions, however, the mutants E153G and L280R showed lower ΔG values and stronger binding affinity.

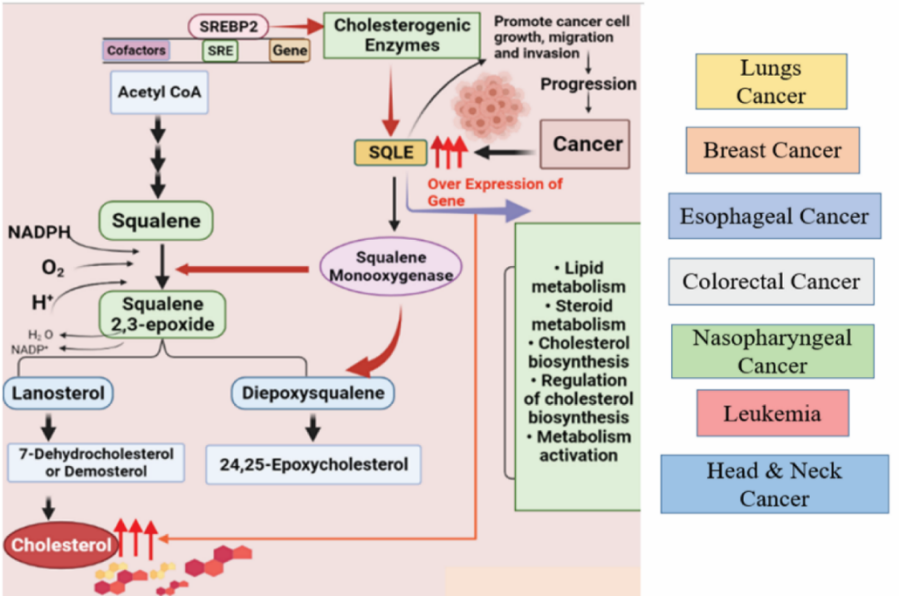

Figure 6. Pathway analysis and its relation with cancer

4. DISCUSSION

Cancer is a complex, multi-gene disease that has spread globally and is a major threat to the public [50]. In recent years, the genetic mutations in cancer target genes have been reported to guide cancer therapeutics [51]. Among these mutations, single-nucleotide variations are the most common mutations that result in amino acid substitutions and lead to changes in protein structure, stability, and interaction with the ligand, cofactor, or other proteins, thus affecting its molecular function [39], [52]. Single-nucleotide variations in the coding region of the DNA cause several alterations that have a great association with complex diseases and deadly consequences [53]. As these damaging non-synonymous mutations have a prominent role in multiple diseases, major medical research is focused on analyzing their impact. Recently, the damaging nsSNPs of *P53* have been investigated, and its detrimental effect on oncogenesis has been determined by computational methods, differential expression, and mutational spectrum [52]. In another study, the *CHK2* gene has a significant role in ovarian and breast cancer, as this gene codes for serine-threonine kinase and has a prominent role in the cell cycle and DNA repair. Its nsSNPs are determined computationally in the conserved region, this study predicts its significance for personalized medicine [54]. Lately *SQLE* gene has been reported as a candidate gene due to its involvement in multiple cancer types [6]. The rate-limiting enzyme (SQLE) in the cholesterol biosynthesis pathway is crucial for multiple

lipid metabolic pathways. Any alteration that leads to enzymatic malfunctioning ultimately causes disturbance in the respective cholesterol pathway, and accumulation of cholesterol induces multiple types of cancer [10]. The functional assessment of variants by functional assay is the best technique but it is costly and time-consuming as well, so the computational approach can be exploited by using various *in silico* tools working with different algorithms as the best and most reliable technique. Therefore, the present study was designed based on computational tools to prioritize the most detrimental mutations in the conserved regions of SQLE.

In this study, there is a comprehensive *in silico* assessment of missense SNPs of the *SQLE* gene to filter the potentially pathogenic variants to understand the etiology of cancer in a better manner. Many studies highlighted the importance of missense mutations as it has a great impact on the structure, folding, and function of proteins [55], [56], [57]. The current study reveals that missense mutations affect the structure, stability, and function of native proteins. Because the molecular functions of a protein depend on its interaction with substrate, cofactor, and other proteins, the changes in molecular interactions may affect the molecular functions of these proteins. We reported SNPs that are pathogenic, evolutionary conserved, highly destabilize the protein structure, and affect the functionality of gene products; these are predicted to be involved in the onset of cancer. [12] and [11] illustrated the structural and functional significance of *SQLE,* and this study is expanded to the computational characterization of missense SNPs. For this purpose, multiple computational tools with different algorithms are used to screen highly pathogenic missense variants, as all the variants are disease-associated, and some are neutral as well [54]. To evaluate our hypothesis, firstly, 450 missense mutations retrieved from dbSNP were annotated in functional assessment categorized as homology (SIFT, Provean, PolyPhen2.0, Mutation Assessor), machine learning (SuSPect, SNP&GO, PhD-SNP, MUTPRED, PMUT), and consensus-based (META-SNP, PREDICT SNP, ConDEL, PON-P2) analysis, conclusively predicted 18 variants as disease-associated, respectively. These functional assessments are based on three different approaches each based on further four different algorithms that increase the accuracy of the prediction [39], [58] to designate the variants to be neutral or highly disease-associated as a change in sequence may alter the geometry, charge, and hydrophobicity of protein [59]. The evolutionary conserved amino acids create a balanced profile of protein and its tendency to mutate naturally, which maintains the structural and functional integrity of macromolecules [60]. Collectively, 18 variants declared pathogenic by the combination of the above three algorithms are subjected to conservation analysis by ConSurf, where 13 pathogenic variants were substitutions for evolutionary conserved amino acids. Currently, it is suggested that the disease-associated pathogenic variants may affect the stability of the proteins [61], so after functional and conservation assessment, it is important to determine the thermodynamic and stability changes. Macromolecule's stability has prime importance for regulation, activity, and function, while alteration in DDG in mutated models disturbs the stability and structure is more conserved, so the final mutant selection is sequence and structure-based, where destabilizing variants are opted for the next analysis [23]. These 13 missense mutations were analyzed based on sequence-based stability (I-Mutant, MuPro, I-Stable2.0) and structure-based protein stability (SDM, DynaMut, PoPMusic, FOLDX). This evaluation anticipates 6 variants (E153G, L280R, G286R, R343P, R458T, L473H) as highly destabilizing. The online servers or

databases can predict the association of variants to the cancer and the great risk of developing multiple types of cancer [62]. These six variants were subjected to CRAVET software, where cancer-associated, two mutations involved in multiple types of cancers (head and neck, colon, rectum, prostate, breast, and lung cancer) with the highest proportion are scrutinized by CHASM-3.1.

For enzyme or protein functioning, its structure and specifically active site and binding residues determine the nature of interaction [63]. Moreover, the binding pocket of protein by FTSite assessed, having E153, G286, and P415 amino acids with a prominent role in the active site that elaborated the importance of E153 and G286 substitution that may alter the interaction profile [64]. The structural impact of these mutations is checked by mapping them onto the 3D structure as the structural changes in protein are analyzed by stability and conformational changes that further confirm the detrimental impact of the mutation [39]. As amino acids in binding pockets have significance in protein structure, the structure is further validated by generating mutant models to compare the interaction pattern of native and mutant models with ligands. NMSim is an online server that significantly evaluates and validates the stable protein structures after the introduction of mutations and predicts the conformations of protein geometrically [65]. Additionally, these mutants, verified by NMSim geometric simulations, have great fluctuations in mutants (E153 and L280) and results predicted on parameters as RMSF and RMSD, where both mutants showed deviation from the wild-type model. The accurate receptor molecule is mandatory for binding in a structurally correct binding pocket [66], while changes in receptor molecules alter the interactions. The top candidates E153G and L280R selected, are significant as they showed allele frequency in different databases as, E153G showed "G=0.000004 (1/264690, TOPMED), G=0.00008 (1/12044, GO-ESP), G=0.00000 (0/10680, ALFA)" and L280R showed "G=0.0000196 (27/1374050, GnomAD_exomes), G=0.000004 (1/264690, TOPMED), G=0.00019 (12/63076, ExAC) (- 1 less), G=0.00006 (3/51624, ALFA)". Their docking analysis along with wild-type further authenticates the binding of SQLE with its ligand (NADPH and squalene) by Autodock Vina, MGL Tool, and BIOVIA Discovery Studio. These alterations in interactions were noted to elaborate the loss of interactions and change in binding energies due to these mutations based on hydrogen bonding, carbon-hydrogen bonding, attractive charges, alkyl, and pi-alkyl interactions. Most of the amino acids lost their interactions, and further, new amino acids showed interactions, and both mutant substitutions provide rigid and bulky amino acids that provide different favorable interactions. The changes in the mutant models of protein highlighted the malfunctioning and instability of the protein [18]. The loss of any hydrogen bonding, negatively charged molecule and introduction of glycine in E153G induced flexibility that destabilize the protein structure by affecting its stability[67], and changes in alkyl, pi-alkyl interactions as present in protein and squalene can affect the conformation and packaging of protein [63]. The L280R showed deviation from wild-type binding affinity when docked with NADPH as compared to E153G. L280R introduces a positively charged, bulky residue into a hydrophobic core that provide steric clashes and destabilize the domain packaging. The selected mutations (E153G and L280R) were validated and authenticated by molecular dynamic simulations showing their prominent role in cancer association, as E153G showed instability in NADPH and SQLE interaction while L280R showed moderate stability. This result was determined at 5 ns production phase, so there is a research gap to perform their MD simulations at 100 ns for future studies.

Altogether, the disruption in the interaction pattern of wild-type protein might induce cancer due to identified mutations. These mutations have functional characterization as pathogenic, evolutionary conserved, and destabilizing the significant structure of protein and more importantly changing the conformations and interactions with ligands in the docked complex. So, these variants can be possible factors behind the proliferation of abnormal cells and cancer progression.

5. CONCLUSION

As a result, the computational pipeline utilized in this study, futuristic insights gained by examination of potentially disease-associated or pathogenic candidates of the SQLE gene involved in cancer progression. A variety of computational tools with different algorithms are employed to predict the deleterious, evolutionary conserved, and destabilizing SNPs that disturb cholesterol biosynthesis. Two variants, E153G and L280R, censored, are likely to cause cancer as they are pathogenic by functional assessment, evolutionary conserved, and alter the stability of the protein, having an association with cancer. These variants were further authenticated by geometric simulations, comparative docking analysis, and MD simulations, which reveal unstable docked complex formation. These mutations may be regarded as real mutants computationally and considered to be oncogenic.

Author Contributions

MM: Conceptualisation, Methodology, Writing — Original Draft. SB & AH: Investigation, Formal Analysis, Writing — Review & Editing. NMA & HB: Supervision, Project Administration. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Generative AI and AI-Assisted Technology Disclosure

The authors did not use any generative AI or AI-assisted tools in the preparation of this manuscript.

Funding Statement

This research received no external funding.

Supplementary Materials Statement

Supplementary files are available upon request

Acknowledgment

The authors are thankful for the online databases used in this research. The research was self-supported by the authors.

Conflict of Interests

Authors declare no conflict of interests.

REFERENCES

1. Martínez-Reyes I, Chandel NS. Cancer metabolism: looking forward. Nat Rev Cancer. 2021;21(10):669-680. https://doi.org/10.1038/s41568-021-00378-6.
2. World Health Organization. Cancer. Accessed February 24, 2022. https://www.who.int/news-room/fact-sheets/detail/cancer
3. Mohammed F, Odintsova E, Chidgey M. Missense mutations in desmoplakin plakin repeat domains have dramatic effects on domain structure and function. Int J Mol Sci. 2022;23(1):529. https://doi.org/10.3390/ijms23010529.
4. Porta-Pardo E, Valencia A, Godzik A. Understanding oncogenicity of cancer driver genes and mutations in the cancer genomics era. FEBS Lett. 2020;594(24):4233-4246. https://doi.org/10.1002/1873-3468.13781.

5. Zhang L, et al. Squalene epoxidase: its regulations and links with cancers. Int J Mol Sci. 2024;25(7). https://doi.org/10.1002/1873-3468.13781.10.3390/ijms25073874.

6. The role of SQLE in head and neck squamous cell carcinoma. Accessed March 1, 2022. https://escholarship.org/uc/item/8jb7n687

7. Wu K, Lin F. Lipid metabolism as a potential target of liver cancer. J Hepatocell Carcinoma. 2024;11:327-346. https://doi.org/10.2147/JHC.S450423.

8. Qiao S, et al. Targeting SQLE-mediated cholesterol metabolism to enhance CD8(+) T cell activation and immunotherapy efficacy in hepatocellular carcinoma. J Immunother Cancer. 2025;13(9). https://doi.org/10.1136/jitc-2025-012345.

9. Qin Y, et al. A novel long non-coding RNA lnc030 maintains breast cancer stem cell stemness by stabilizing SQLE mRNA and increasing cholesterol synthesis. Adv Sci (Weinh). 2021;8(2):2002232. https://doi.org/10.1002/advs.202002232.

10. Liu Y, Fang L, Liu W. High SQLE expression and gene amplification correlates with poor prognosis in head and neck squamous cell carcinoma. Cancer Manag Res. 2021;13:4709-4723. https://doi.org/10.2147/CMAR.S305719.

11. Padyana AK, et al. Structure and inhibition mechanism of the catalytic domain of human squalene epoxidase. Nat Commun. 2019;10(1):1-10. https://doi.org/10.1038/s41467-018-07928-x.

12. Brown AJ, Chua NK, Yan N. The shape of human squalene epoxidase expands the arsenal against cancer. Nat Commun. 2019;10(1):2-5. https://doi.org/10.1038/s41467-019-08866-y.

13. Chua NK, Brown AJ. The degron architecture of squalene monooxygenase and how specific lipids calibrate levels of this key cholesterol synthesis enzyme. Adv Exp Med Biol. 2021:1-12. https://doi.org/10.1007/5584_2020_583.

14. Sagatova AA. Strategies to better target fungal squalene monooxygenase. J Fungi (Basel). 2021;7(1):1-13. https://doi.org/10.1002/1873-3468.13781.10.3390/jof7010049.

15. Day INM. dbSNP in the detail and copy number complexities. Hum Mutat. 2010;31(1):2-4. https://doi.org/10.1002/humu.21149.

16. Iqbal S, et al. Comprehensive characterization of amino acid positions in protein structures reveals molecular effect of missense variants. Proc Natl Acad Sci U S A. 2020;117(45):28201-28211. https://doi.org/10.1073/pnas.2002660117.

17. Yazar M, Özbek P. In silico tools and approaches for the prediction of functional and structural effects of single-nucleotide polymorphisms on proteins: an expert review. OMICS. 2021;25(1):23-37. https://doi.org/10.1089/omi.2020.0141.

18. Mahmood MS, Irshad S, Kalsoom U, Batool H, Batool S, Butt TA. In silico analysis of missense single nucleotide variants (SNVs) in HBB gene associated with the β-thalassemia. Gene Rep. 2021;23:101019. https://doi.org/:10.1016/j.genrep.2021.101019.

19. Yates CM, Filippis I, Kelley LA, Sternberg MJE. SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. J Mol Biol. 2014;426(14):2692-2701. https://doi.org/10.1016/j.jmb.2014.04.026.

20. Capriotti E, Altman RB, Bromberg Y. Collective judgment predicts disease-associated single nucleotide variants. BMC Genomics. 2013;14(suppl 3):S2. https://doi.org/10.1186/1471-2164-14-S3-S2.

21. Bendl J, et al. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput Biol. 2014;10(1). https://doi.org/10.1371/journal.pcbi.1003440.

22. González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet. 2011;88(4):440. https://doi.org/10.1016/j.ajhg.2011.03.004.

23. Khan S, Vihinen M. Performance of protein stability predictors. Hum Mutat. 2010;31(6):675-684. https://doi.org/10.1002/humu.21242.

24. Niroula A, Urolagin S, Vihinen M. PON-P2: prediction method for fast and reliable identification of harmful variants. PLoS One. 2015;10(2):1-17. https://doi.org/10.1371/journal.pone.0117380.

25. 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(suppl 2):W529-W533. https://doi.org/10.1093/nar/gkq399.

26. Sanavia T, Birolo G, Montanucci L, Turina P, Capriotti E, Fariselli P. Limitations and challenges in protein stability prediction upon genome variations: toward future applications in precision medicine. Comput Struct Biotechnol J. 2020;18:1968-1979. https://doi.org/10.1016/j.csbj.2020.07.011.

27. Wong WC, Kim D, Carter H, Diekhans M, Ryan MC, Karchin R. CHASM and SNVBox: toolkit for detecting biologically important single nucleotide mutations in cancer. Bioinformatics. 2011;27(15):2147-2148. https://doi.org/10.1093/bioinformatics/btr357.

28. Ngan CH, Hall DR, Zerbe B, Grove LE, Kozakov D, Vajda S. FTSite: high accuracy detection of ligand binding sites on unbound protein structures. Bioinformatics. 2012;28(2):286-287. https://doi.org/10.1093/bioinformatics/btr651.

29. Kozakov D, et al. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. Nat Protoc. 2015;10(5):733-755. https://doi.org/10.1038/nprot.2015.043.

30. Kemmish H, Fasnacht M, Yan L. Fully automated antibody structure prediction using BIOVIA tools: validation study. PLoS One. 2017;12(5):e0177923. https://doi.org/10.1002/1873-3468.13781. 10.1371/journal.pone.0177923.

31. Krüger DM, Ahmed A, Gohlke H. NMSim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins Nucleic Acids Res. 2012;40(Web Server issue). https://doi.org/10.1093/nar/gks478.

32. Kim S, et al. PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res. 2021;49(D1):D1388-D1395. https://doi.org/10.1093/nar/gkaa971.

33. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455-461. https://doi.org/10.1002/jcc.21334.

34. Yair A, et al. AutoDock4 and Vina virtual screening software: comparison of performance and results. 2021.

35. Justino GC, Nascimento CP, Justino MC. Molecular dynamics simulations and analysis for bioinformatics undergraduate students. Biochem Mol Biol Educ. 2021;49(4):570-582. https://doi.org/10.1002/bmb.21512.

36. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44(D1):D457-D462. https://doi.org/10.1093/nar/gkv1070.

37. Gillespie M, et al. The Reactome pathway knowledgebase 2022. Nucleic Acids Res. 2022;50(D1):D687-D692. https://doi.org/10.1093/nar/gkab1028.

38. Yilmaz A, Çetin İ. In silico prediction of the effects of nonsynonymous single nucleotide polymorphisms in the human catechol-O-methyltransferase (COMT) gene. Cell Biochem Biophys. 2020;78(2):227-239. https://doi.org/10.1007/s12013-020-00905-6.

39. Alzahrani FA, et al. Investigating the pathogenic SNPs in BLM helicase and their biological consequences by computational approach. Sci Rep. 2020;10(1):1-22. https://doi.org/10.1038/s41598-020-69033-8.

40. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;(suppl 76). https://doi.org/10.1002/0471142905.hg0720s76.

41. Application of pathogenicity scores as diagnostic and prognostic markers for MPS disorders: in-silico analysis in MPS I. Accessed February 12, 2023. https://europepmc.org/article/ppr/ppr594701

42. Arshad M, Bhatti A, John P. Identification and in silico analysis of functional SNPs of human TAGAP protein: a comprehensive study. PLoS One. 2018;13(1):e0188143. https://doi.org/10.1371/journal.pone.0188143.

43. Choudhury A, et al. Structural genomics approach to investigate deleterious impact of nsSNPs in conserved telomere maintenance component 1. Sci Rep. 2021;11(1). https://doi.org/10.1038/s41598-021-89450-7.

44. López-Ferrando V, Gazzo A, De La Cruz X, Orozco M, Gelpí JL. PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Res.* 2017;45(W1):W222-W228. https://doi.org/10.1093/nar/gkx313.

45. Halilunnisa U, Hussain M, Ahamed IBA, Ibrahim S, Ahamed B. Structural impact of pathogenic SNPs on β-tubulin using molecular dynamics study. *J Biomol Struct Dyn.* 2022. https://doi.org/10.1080/07391102.2022.2130986.

46. Qorri E, et al. A comprehensive evaluation of the performance of prediction algorithms on clinically relevant missense variants. *Int J Mol Sci.* 2022;23(14). https://doi.org/10.3390/ijms23147946.

47. Zia A, Shams S, Shah M, Afridi SG, Khan A. Structural consequences of IRS-2 nsSNPs and implication for insulin receptor substrate-2 protein stability. *Biochem Genet.* 2023;61(1):69-86. https://doi.org/10.1007/s10528-022-10247-y.

48. Gerasimavicius L, Livesey BJ, Marsh JA. Loss-of-function, gain-of-function and dominant-negative mutations have

profoundly different effects on protein structure. *Nat Commun.* 2022;13(1):1-15. https://doi.org/10.1038/s41467-022-31686-6.

49. Ben Chorin A, et al. ConSurf-DB: an accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. *Protein Sci.* 2020;29(1):258-267. https://doi.org/10.1002/pro.3779.

50. Soerjomataram I, Bray F. Planning for tomorrow: global cancer incidence and the role of prevention 2020–2070. *Nat Rev Clin Oncol.* 2021;18(10):663-672. https://doi.org/10.1002/1873-3468.13781.10.1038/s41571-021-00514-z.

51. Mattiuzzi C, Lippi G. Current cancer epidemiology. *J Epidemiol Glob Health.* 2019;9(4):217-222. https://doi.org/10.2991/jegh.k.191008.001.

52. Alam S, et al. Computational analysis uncovers deleterious SNPs and mutational spectrum of p53 gene in pan-cancer. *Bull Natl Res Cent.* 2022;46(1). https://doi.org/10.1002/1873-3468.13781.10.1186/s42269-022-00859-0.

53. Khalid Z, Almaghrabi O. Mutational analysis predicting the impact of high-risk SNPs in human secretory phospholipase A2 receptor (PLA2R1). *Sci Rep.* 2020;10(1):1-11. https://doi.org/10.1038/s41598-020-68696-7.

54. Badgujar NV, Tarapara BV, Shah FD. Computational analysis of high-risk SNPs in human CHK2 gene responsible for hereditary breast cancer: a functional and structural impact. *PLoS One.* 2019;14(8):e0220711. https://doi.org/10.1371/journal.pone.0220711.

55. Malhotra S, et al. Understanding the impacts of missense mutations on structures and functions of human cancer-related genes: a preliminary computational analysis of the COSMIC Cancer Gene Census. *PLoS One.* 2019;14(7):e0219935. https://doi.org/10.1371/journal.pone.0219935.

56. Kumar SU, et al. Molecular dynamics, residue network analysis, and cross-correlation matrix to characterize deleterious missense mutations in GALE causing galactosemia III. *Cell Biochem Biophys.* 2021;79(2):201-219. https://doi.org/10.1007/s12013-020-00960-z.

57. Agrahari AK, et al. Understanding the structure-function relationship of HPRT1 missense mutations in association with Lesch–Nyhan disease and HPRT1-related gout. *Comput Biol Med.* 2019;107:161-171. https://doi.org/10.1002/1873-3468.13781.10.1016/j.compbiomed.2019.02.014.

58. Navapour L, Mogharrab N. In silico screening and analysis of nonsynonymous SNPs in human CYP1A2. *Sci Rep.* 2021;11(1):1-15. https://doi.org/10.1038/s41598-021-83696-x.

59. Hossain MS, Roy AS, Islam MS. In silico analysis predicting effects of deleterious SNPs of human RASSF5 gene. *Sci Rep.* 2020;10(1):1-14. https://doi.org/10.1038/s41598-020-71457-1.

60. Ashkenazy H, et al. ConSurf 2016: improved methodology to estimate evolutionary conservation in macromolecules. *Nucleic Acids Res.* 2016;44(W1):W344-W350. https://doi.org/10.1093/nar/gkw408.

61. Iqbal S, Li F, Akutsu T, Ascher DB, Webb GI, Song J. Assessing performance of computational predictors for estimating protein stability changes upon missense

mutations. *Brief Bioinform.* 2021;22(6). https://doi.org/10.1093/bib/bbab184.

62. Karihtala P, et al. Comparison of mutational profiles of neuroendocrine breast tumours and carcinomas. *Oncogenesis.* 2022;11(1):53. https://doi.org/10.1038/s41389-022-00427-1.

63. Makwana KM, Mahalakshmi R. Implications of aromatic–aromatic interactions: from protein structures to peptide models. *Protein Sci.* 2015;24(12):1920. https://doi.org/10.1002/pro.2814.

64. Sharma D, Saini R, Mishra A. Natural phytocompounds targeting L-asparaginase of *Mycobacterium tuberculosis*: a molecular dynamics study. *J Biomol Struct Dyn.* 2022. https://doi.org/10.1080/07391102.2022.2036239.

65. Aslam MS, et al. High yield expression, characterization, and biological activity of IFNα2-Tα1 fusion protein. *J Toxicol Environ Health A.* 2019;50(3):281-291. https://doi.org/10.1080/10826068.2019.1689509.

66. Kadukova M, Grudinin S. Docking of small molecules to farnesoid X receptors using AutoDock Vina. *J Comput Aided Mol Des.* 2018;32(1):151-162. https://doi.org/10.1007/s10822-017-0062-1.

67. Mallamace D, Fazio E, Mallamace F, Corsaro C. The role of hydrogen bonding in the folding/unfolding process of hydrated lysozyme. *Int J Mol Sci.* 2018;19(12):3825. https://doi.org/10.3390/ijms19123825.