Vol. 2 No. 3 (2025)
ONLINE ISSN :3007-309X
PRINT ISSN :3007-3081
Journal of Medical & Health Sciences Review

COMPUTATIONAL CHARACTERIZATION OF NON-SYNONYMOUS SNPS OF THE HUMAN PROSTATE CANCER-ASSOCIATED EHBP1 GENE

Fajar Baig[1], Nazia Hadi[1], Muhammad Qayash Khan[2], Tayyaba Iftikhar[3], Naveed Khan[1]

[1]Department of Biotechnology, Abdul Wali Khan University, Mardan, Pakistan
Email: fajarbaigdec@gmail.com , naziahadi123@gmail.com , naveedkhan@awkum.edu.pk
[2]Department of Zoology, Abdul Wali Khan University, Mardan, Pakistan,
Email: qayashkhan@awkum.edu.pk
[3]Department of Pharmacy, Abdul Wali Khan University, Mardan, Pakistan,
Email: t.iftikhar@awkum.edu.pk

ARTICLE INFO:

Keywords:
Prostate cancer; EHBP1; non-synonymous SNPs (nsSNPs); in-silico analysis; protein stability; functional prediction; gene-gene interaction; biomarker.

Corresponding Author:
Naveed Khan,
Department of Biotechnology, Abdul Wali Khan University, Mardan (AWKUM), Pakistan
naveedkhan@awkum.edu.pk

Article History:
Published on 16 September 2025

ABSTRACT

Prostate cancer is one of the frequently reported cancers in men worldwide, and it is the third most prevalent urogenital cancer among males in the Pakistani population. Variations in several oncogenes and tumor suppressor genes have been studied that play a role in prostate cancer. GWAS has identified SNPs in the *EHBP1* gene to be associated with prostate cancer. In this study, we used in-silico approaches for identifying the most damaging non-synonymous SNPs (nsSNPs), playing a significant structural and functional role in EHBP1 protein. Data on non-synonymous SNPs were recruited from Ensembl. Deleterious nsSNPs were identified using SIFT, PolyPhen2, PhD-SNP, fathmm, and SNPs&GO. Structural, functional, stability analysis and conservation profile of nsSNPs were verified using MutPred, I-Mutant, MUpro and ConSurf web server, respectively. STRING was used for protein-protein interaction. GeneMANIA was utilized to check the *EHBP1* gene interaction with other genes. The 3D structures of wild-type and mutant proteins were generated using I-Tasser. Post-translational modification sites were predicted through the MusuitDeep web server. Our study identified 32 most damaging nsSNPs in the *EHBP1* gene. Structural and functional analysis of these nsSNPs manifests that they have a deleterious effect on protein structure and function. Stability analysis showed that 31 of these nsSNPs decrease protein stability and are located in highly conserved regions. Gene-gene interactions revealed a relationship between *EHBP1* and other genes, highlighting its significance in multiple pathways and co-expression patterns. In future, these 32 SNPs provide suitable target variants to be explored, through population-based studies, in diseases associated with the *EHBP1* gene, such as prostate cancer, as well as for their role as novel biomarkers.

INTRODUCTION

Cancer is one of the life-threatening diseases affecting millions of people globally (Akhtar et al., 2019). Prostate cancer is one of the frequently reported cancers in men worldwide, while it is the third most prevalent urinogenital cancer among males in the Pakistani population (Akhtar et al., 2023). Clinical observations and experimental studies have designated various factors that play a role in the pathogenesis of prostate cancer, including geographical region, age, ethnicity, family history of prostate cancer, obesity, hormones like androgens and various genetic variations. Point mutations, miRNAs, Somatic copy number alterations (SCNAS), structural rearrangements and single nucleotide polymorphisms (SNPs) are the genetic alterations that can contribute to tumorigenesis. SNPs are the important biological markers responsible for over 90% of genetic variations affecting gene function, thereby increasing or decreasing susceptibility towards various genetic disorders, including cancer (Beikzadeh et al., 2020).

Variations in several oncogenes (*KRAS* and *BRAF*) and tumor suppressor genes (*PTEN, TP53*, *RB1* and *PIK3CA)* have been studied that have a role in prostate cancer (Wallis and Nam, 2015). Genome-wide association study explores genotypes, thereby recognizing the gene characteristics that are involved in diseases (Liang et al., 2020). GWAS has identified more than 50 susceptibility loci in several genes associated with prostate cancer. Similarly, GWAS has also identified SNPs in the *EHBP1* gene as being associated with prostate cancer (Ao et al., 2015).

The *EHBP1* gene is located on 2p15 chromosomal region, which encodes an adaptor protein regulating vesicular trafficking of endocytes thereby recruiting EPs15-homology containing proteins 1/2 (EHD1/2) and Rab8 family members (Gudmundsson et al., 2008) (Rai et al., 2020). EH domain protein 2 (EHD2) binding protein 1 (EHBP1) was recognized and cloned in 3T3-L1 adipocytes for the first time, containing five NPF motifs and a calponin homology domain. *EHBP1* play a significant role in insulin-stimulated GLUT4 movements and hexose transport, thereby interfering with EHD2 protein, and it also connects clathrin-mediated endocytosis to the actin cytoskeleton. Several studies reported that *EHBP1* gene SNPs significantly contribute to prostate cancer in various ethnic groups, including Caucasians, Africans, and Asian populations (Ao et al., 2015).

This study aimed to analyse the damaging nsSNPs and their effect on the stability of EHBP1 protein using various in-silico tools.

MATERIAL AND METHODS

Enlisting nsSNPs

Complete information of *EHBP1* gene nsSNPs, including SNP Id, protein accession number, and position of residue change, was collected from the Ensembl database.

Identification of damaging nsSNPs

To analyze the damaging and deleterious effects of nsSNPs on the *EHBP1* gene five well established bioinformatics tools were utilized including SIFT (Sorting Intolerant From Tolerant) (Sim et al., 2012), PolyPhen-2 (Polymorphism Phenotyping V2) (Adzhubei et al., 2013), PhD-SNP (Predictor of Human Deleterious SNPs) (Capriotti et al., 2006), SNPs&GO (Single Nucleotide Polymorphism Database & Gene Ontology) (Calabrese et al., 2009) and fathmm (Functional Analysis Through Hidden Markov Models) (Shihab et al., 2013). A list of amino acid substitutions and FASTA sequences was submitted as input in all five tools for prediction. Only those SNPs were selected for further analysis which was predicted to be deleterious or damaging by all five tools.

SIFT

SIFT prediction is based on sequence homology to determine damaging and tolerated SNPs, with a probability score of

<0.05 and ≥ 0.05, respectively (Sim et al., 2012).

PolyPhen-2

PolyPhen-2 predicts the nsSNP as damaging or benign by utilizing substitution sequence, evolutionary properties and structural confirmation. It uses protein sequence, accession number, amino acid change and position as input in order to retrieve results (Adzhubei et al., 2013).

PhD SNP

PhD-SNP predicts the deleterious effect of nsSNPs based on support vector machines (SVM) by utilizing protein sequence and amino acid substitution as input (Capriotti et al., 2006).

SNPS&GO

SNPs&GO predict variants that are related to disease phenotype by utilizing SVM (support vector machine) based techniques. Protein sequence in FASTA format and amino acid substitutions were submitted as input while the output were in the form of disease or neutral phenotype having probability score ≥ 0.5 and < 0.5 respectively (Calabrese et al., 2009).

Fathmm

fathmm predicts functional effects of nsSNPs by combining sequence conservation within Hidden Markov Models in order to represent the alignment of homologous sequences and conserved protein domains with pathogenicity weights (Shihab et al., 2013).

Effect of nsSNPs on the structure and function of the protein

MutPred1.2 (Li et al., 2009) is a bioinformatics web-based tool used for predicting the damaging effect of nsSNPs on the structure and function of proteins. A FASTA sequence of protein, along with amino acid substitutions, was submitted as input.

nsSNPs and protein stability

I-Mutant 2.0 (Capriotti et al., 2006) and MUpro (Cheng et al., 2006) were used for the prediction of the effect that amino acid changes have on the stability of protein structure. These nsSNPs either decrease or increase the stability of the protein. A FASTA sequence of protein, along with the amino acid substitutions, was used as input.

Evolutionary conservation profile

The ConSurf tool was used to check the evolutionary conservation profile of the amino acid present at a specific position in a protein sequence. It checks the conservation status of amino acids based on phylogenetic relationships between homologous sequences (Ashkenazy et al., 2016).

3D Modelling of EHBP1 protein

EHBP1 protein wild and mutant structures were predicted using I-TASSER (Zhang, 2008). Chimaera 1.11 tool was utilized for visualization and analysis of structures generated through I-TASSER (Pettersen et al., 2004). TM-align was utilized for comparison of wild and mutant type protein structures, thereby generating RMSD score (Root mean square deviation) and TM values (Zhang and Skolnick, 2005).

***EHBP1* gene interaction with other genes**

The GeneMania tool was utilized in this study to identify the interactions of the *EHBP1* gene with other genes. This tool generates information about physical interaction, co-localization, common protein domains, co-expression, genetic interaction and pathways involved (Warde-Farley et al., 2010).

Analysis of protein-protein interaction

The STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) tool was used to determine the physical and functional interaction of EHBP 1 protein with other proteins (Jensen et al., 2009).

PTM (post-translational modification) sites prediction in EHBP1

MusiteDeep web server was used to predict the possible post-transcriptional modification sites located in the EHBP1 protein. It is an excellent tool for predicting and visualising PTM sites. It utilizes FASTA protein sequence as an input and predicts results in

real time for a large number of proteins (Wang et al., 2020).

RESULTS

nsSNPs in the EHBP1 gene

A total of 128,205 variants are located in the *EHBP1* gene according to the Ensembl database. Among these 1075 variants are nsSNPs; data of these 1075 nsSNPs were retrieved and included in this study for analysis.

Damaging nsSNPs identification in the EHBP1 gene

Out of the total 1075 nsSNPs found in the *EHBP1* gene, 175 variants were predicted to be damaging by three tools applied for damage predictions; these tools include SIFT, SNPs&GO and PolyPhen-2. SIFT predicted 946 variants affect protein function, while 129 variants were predicted as tolerated. Additionally, SNPs&GO predicted 211 variants as disease-causing, meanwhile 864 nsSNPs were predicted to be tolerated. Furthermore, PolyPhen-2 predicted 546 variants as probably damaging, while the remaining 529 were possibly damaging and benign (Figure 1). Overall, 175 nsSNPs were predicted deleterious by these tools, which were subsequently validated using PhD-SNP and fathmm bioinformatics tools. PhD-SNP predicted 109 non-synonymous variants as disease-causing, while 66 were predicted to be tolerated. Additionally, fathmm predicted 46 nsSNPs as damaging, while the remaining 129 variants were predicted as tolerated (Figure 2). Overall, 32 nsSNPs were predicted to be damaging by both of the tools. Hence, these 32 nsSNPs were predicted to be deleterious by all five tools and utilized for further analysis in this study (Table 1).

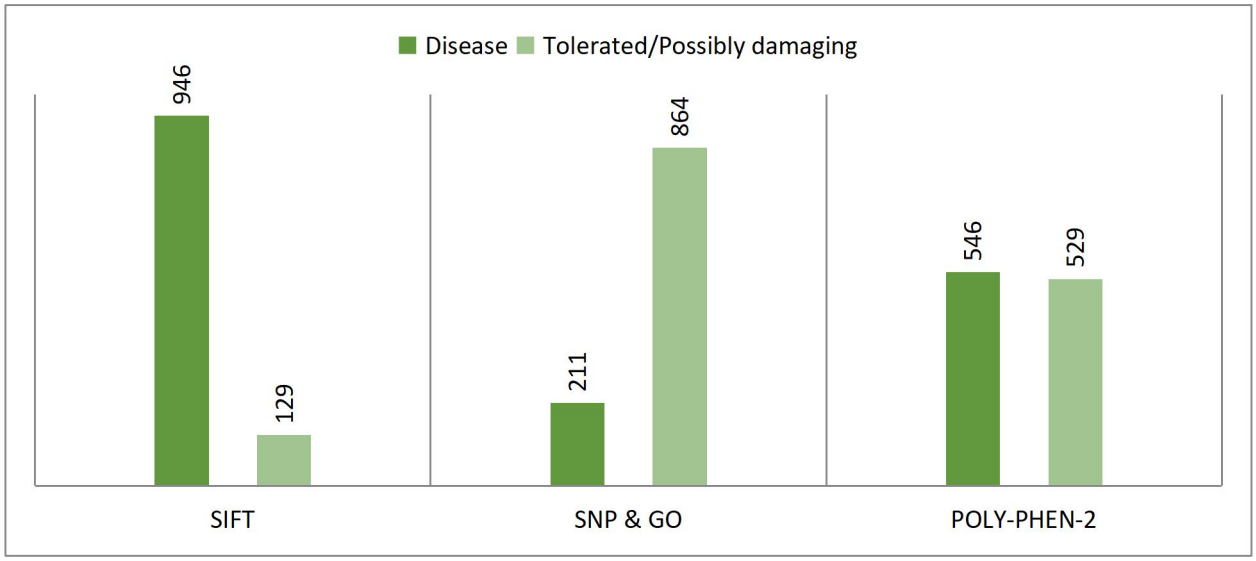


Figure 1. The number of Damaging and Non-damaging nsSNPs identified by SIFT, SNPs&GO and PolyPhen-2

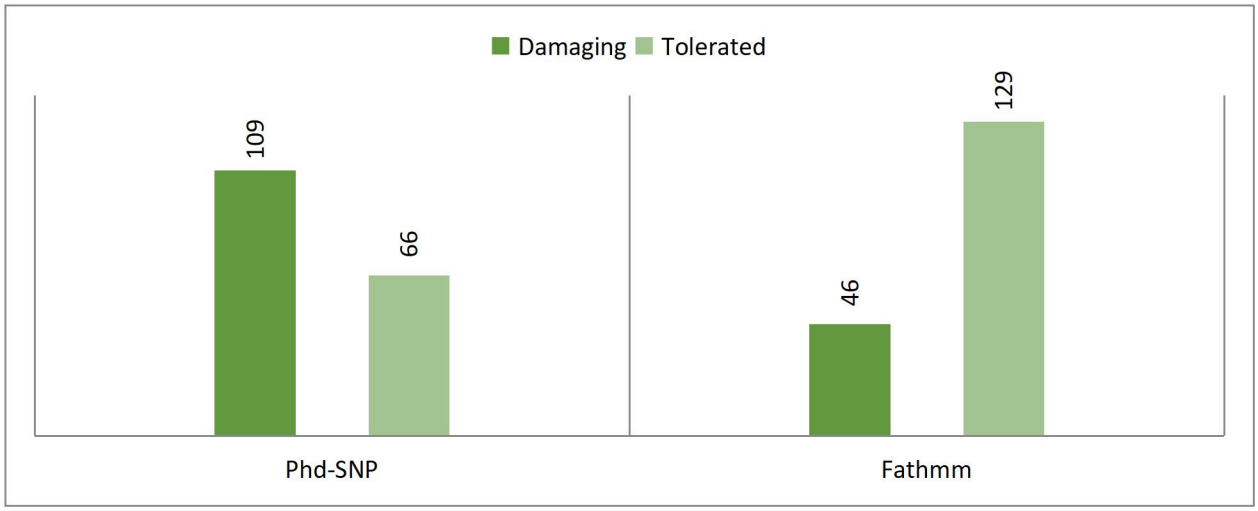


Figure 2. The number of Damaging and Non-damaging nsSNPs identified by PhD-SNP and fathmm

Table 1. The common damaging nsSNPs identified by SIFT, SNPs&GO, PhD-SNP, PolyPhen-2 and fathmm.

S no.	Variant ID	Allele	Amino acid	SIFT score	SNP & GO score	Phd-snp score	Polyphen-2 Score	Fathmm Score
1.	rs1357841649	T/C	W35R	Affect protein function 0.00	Disease 0.807	Disease 4	Probably damaging 1.000	Damaging -1.51
2.	rs1270338713	A/G	D166G	Affect protein function 0.00	Disease 0.733	Disease 6	Probably damaging 0.997	Damaging -1.74
3.	rs910753408	C/T	S448F	Affect protein function 0.00	Disease 0.571	Disease 8	Probably damaging 1.000	Damaging -3.64
4.	rs1296416363	G/T	L449F	Affect protein function 0.01	Disease 0.642	Disease 8	Probably damaging 1.000	Damaging -3.99
5.	rs1324699331	G/A/ **T**	V456F	Affect protein function 0.00	Disease 0.766	Disease 4	Probably damaging 1.000	Damaging -3.82
6.	rs1251754862	T/C	I465T	Affect protein function 0.00	Disease 0.760	Disease 5	Probably damaging 1.000	Damaging -3.72
7.	rs2056839551	C/G	I465M	Affect protein function 0.00	Disease 0.572	Disease 7	Probably damaging 1.000	Damaging -3.80
8.	rs1480624398	A/G	N467S	Affect protein function 0.02	Disease 0.810	Disease 7	Probably damaging 1.000	Damaging -3.72
9.	rs999530556	C/T	T469I	Affect protein function 0.00	Disease 0.829	Disease 7	Probably damaging 1.000	Damaging -3.62
10.	rs1032040422	A/G	T470A	Affect protein function 0.01	Disease 0.585	Disease 1	Probably damaging 0.999	Damaging -3.60
11.	rs1453335398	G/T	G475C	Affect protein function 0.00	Disease 0.860	Disease 8	Probably damaging 1.000	Damaging -5.39
12.	rs1271794992	A/C	L476F	Affect protein function 0.00	Disease 0.644	Disease 9	Probably damaging 1.000	Damaging -3.72
13.	rs781210377	G/T	A480S	Affect protein function	Disease 0.671	Disease 8	Probably damaging 0.994	Damaging -3.80

				0.01				
14.	rs1365369525	A/**G**/T	I481M	Affect protein function 0.00	Disease 0.592	Disease 2	Probably damaging 1.000	Damaging -3.65
15.	rs774252612	C/T	H483Y	Affect protein function 0.00	Disease 0.803	Disease 7	Probably damaging 0.999	Damaging -3.81
16.	rs748018046	A/G	H483R	Affect protein function 0.00	Disease 0.837	Disease 9	Probably damaging 0.999	Damaging -3.81
17.	rs1352187200	T/C	L489S	Affect protein function 0.00	Disease 0.521	Disease 7	Probably damaging 1.000	Damaging -3.69
18.	rs1249602343	A/G	Y492C	Affect protein function 0.01	Disease 0.649	Disease 8	Probably damaging 1..000	Damaging -3.60
19.	rs2057180191	A/G	D509G	Affect protein function 0.00	Disease 0.709	Disease 2	Probably damaging 1.000	Damaging -3.65
20.	rs1267919196	C/T	A512V	Affect protein function 0.01	Disease 0.505	Disease 6	Probably Damaging 1.000	Damaging -3.71
21.	rs1321122729	T/C	I516T	Affect protein function 0.00	Disease 0.636	Disease 5	Probably damaging 1.000	Damaging -3.67
22.	rs768333123	C/A	P522T	Affect protein function 0.00	Disease 0.506	Disease 3	Probably damaging 1.000	Damaging -3.67
23.	rs974000870	G/C	D524H	Affect protein function 0.000	Disease 0.727	Disease 7	Probably damaging 1.000	Damaging -4.43
24.	rs2057184193	A/T	K533N	Affect protein function 0.000	Disease 0.754	Disease 5	Probably damaging 1.000	Damaging -3.81
25.	rs2057184841	T/**C**/G	Y539H	Affect protein function 0.00	Disease 0.774	Disease 8	Probably damaging 1.000	Damaging -3.68
26.	rs2057184841	T/C/**G**	Y539D	Affect protein function 0.00	Disease 0.811	Disease 9	Probably damaging 1.000	Damaging -3.68
27.	rs1296312521	G/T	R544S	Affect protein function 0.01	Disease 0.785	Disease 7	Probably damaging 1.000	Damaging -3.48
28.	rs1339044052	A/G	H546R	Affect protein function 0.02	Disease 0.812	Disease 6	Probably damaging 0.996	Damaging -3.42
29.	rs201434475	G/A/**C**/T	R877P	Affect protein function 0.00	Disease 0.779	Disease 1	Probably damaging 1.000	Damaging -1.61
30.	rs751305005	T/G	V880G	Affect protein function 0.00	Disease 0.550	Disease 0	Probably damaging 1.000	Damaging -1.59
31.	rs1192959984	T/C	L1136P	Affect protein function 0.00	Disease 0.733	Disease 10	Probably damaging 1.000	Damaging -1.58
32.	rs2061893998	C/T	R1189C	Affect protein function 0.00	Disease 0.758	Disease 6	Probably damaging 1.000	Damaging -1.99

EHBP1 stability prediction

I-Mutant and MUpro tools were used to analyze the effect of the selected 32 nsSNPs on stability of *EHBP1* gene. I-Mutant predicted that 30 nsSNPs decrease the stability of protein structure, while two non-synonymous variants (rs910753408 and rs2057184193) were predicted as increasing protein stability. Detailed results of all the nsSNPs along with their RI value (ranging from 0 to 10) are shown in Table 2. MUpro predicted that all 32 nsSNPs decrease the stability of protein structure (Table 2).

Table 2. Effect of nsSNPs on EHBP1 protein stability prediction by I-mutant and Mu-Pro

S NO.	VARIANT ID	AMINOACID	I-mutant	RI	MUpro	DDG value
1.	rs1357841649	W35R	Decrease	6	Decrease	-0.76518782
2.	rs1270338713	D166G	Decrease	5	Decrease	-0.76522613
3.	rs910753408	S448F	Increase	5	Decrease	-0.16446851
4.	rs1296416363	L449F	Decrease	6	Decrease	-1.1913349
5.	rs1324699331	V456F	Decrease	8	Decrease	-0.73142264
6.	rs1251754862	I465T	Decrease	5	Decrease	-2.2223251
7.	rs2056839551	I465M	Decrease	8	Decrease	-0.89107565
8.	rs1480624398	N467S	Decrease	3	Decrease	-1.7244076
9.	rs999530556	T469I	Decrease	9	Decrease	-0.51421395
10.	rs1032040422	T470A	Decrease	10	Decrease	-0.53005735
11.	rs1453335398	G475C	Decrease	7	Decrease	-0.60709109
12.	rs1271794992	L476F	Decrease	9	Decrease	-0.98198009
13.	rs781210377	A480S	Decrease	7	Decrease	-1.0343204
14.	rs1365369525	I481M	Decrease	9	Decrease	-0.9873442
15.	rs774252612	H483Y	Decrease	1	Decrease	-0.34795627
16.	rs748018046	H483R	Decrease	7	Decrease	-0.47959426
17.	rs1352187200	L489S	Decrease	9	Decrease	-2.0765614
18.	rs1249602343	Y492C	Decrease	0	Decrease	-1.114442
19.	rs2057180191	D509G	Decrease	7	Decrease	-0.89676159
20.	rs1267919196	A512V	Decrease	4	Decrease	-0.37118284
21.	rs1321122729	I516T	Decrease	10	Decrease	-2.6001568
22.	rs768333123	P522T	Decrease	9	Decrease	-1.1051465
23.	rs974000870	D524H	Decrease	5	Decrease	-1.4855616
24.	rs2057184193	K533N	Increase	2	Decrease	-1.1214656
25.	rs2057184841	Y539H	Decrease	5	Decrease	-1.1911009
26.	rs2057184841	Y539D	Decrease	0	Decrease	-0.74219838
27.	rs1296312521	R544S	Decrease	8	Decrease	-0.8118642
28.	rs1339044052	H546R	Decrease	7	Decrease	-1.236123
29.	rs201434475	R877P	Decrease	4	Decrease	-1.1992503
30.	rs751305005	V880G	Decrease	10	Decrease	-2.2904057
31.	rs1192959984	L1136P	Decrease	4	Decrease	-1.5850621
32.	rs2061893998	R1189C	Decrease	7	Decrease	-0.87930077

Effect of nsSNPs on the structure and function of EHBP1

To analyse the effect of selected nsSNPs on the structure and functionality of the EHBP1 protein, the Mut-Pred server was utilised. MutPred predictions included Loss of strand, altered disordered interface, metal binding, transmembrane protein, ordered interface, gain or loss of allosteric sites, gain of strand, loss of N-linked glycosylation, gain of helix, loss of loop, loss of sumoylation, loss or gain of ubiquination, loss of ADP-ribosylation, loss or gain of phosphorylation, altered stability, coiled coil, DNA binding, gain of B-factor and loss of catalytic site. Details of each amino acid substitution, along with the MutPred score, affected molecular mechanisms, probability, and P-value, are presented in Table 3.

Table 3. The effect of nsSNPs on structural & functional properties of EHBP1 predicted by MutPred server

Amino acid	MutPred score	Molecular mechanism	Probability	P- value
W35R	0.783	Loss of Strand Altered Disordered interface Altered Metal binding	0.27 0.24 0.23	0.02 0.02 0.04
D166G	0.837	Altered Coiled coil	0.37	0.01
S448F	0.667	Loss of GPI-anchor amidation at N444 Loss of N-linked glycosylation at N444	0.04 0.01	7.1e-03 0.04

L449F	0.459	None	None	None
V456F	0.680	Altered Ordered interface	0.25	0.02
I465T	0.928	Altered Stability	0.52	1.6e-03
		Altered DNA binding	0.34	1.2e-03
		Altered ordered interface	0.28	0.04
		Loss of strand	0.27	0.02
		Loss of Allosteric site at F468	0.20	0.04
		Altered Transmembrane protein	0.16	0.01
		Altered Metal binding	0.10	0.05
		Loss of N-linked glycosylation at N467	0.10	0.01
I465M	0.881	Altered DNA binding	0.31	2.0e-03
		Altered Ordered interface	0.27	8.3e-03
		Gain of Allosteric site at F468	0.21	0.03
		Altered Transmembrane protein	0.16	0.01
		Altered Metal binding	0.11	0.04
		Loss of N-linked glycosylation at N467	0.10	0.01
N467S	0.811	Altered Metal binding	0.28	4.7e-03
		Altered DNA binding	0.27	3.7e-03
		Altered Ordered interface	0.25	0.02
		Loss of Allosteric site at F468	0.21	0.03
		Altered Transmembrane protein	0.18	7.4e-03
		Altered Stability	0.15	0.02
		Gain of Catalytic site at W472	0.11	0.03
		Loss of N-linked glycosylation at N467	0.11	0.01
T469I	0.846	Altered Ordered interface	0.33	7.9e-03
		Altered DNA binding	0.32	1.6e-03
		Gain of Strand	0.27	0.02
		Altered Metal binding	0.26	0.01
		Loss of Allosteric site at R473	0.25	0.01
		Altered Transmembrane protein	0.22	3.0e-03
		Loss of N-linked glycosylation at N467	0.11	0.01
		Loss of Catalytic site at W472	0.10	0.04
T470A	0.752	Altered DNA binding	0.32	1.6e-03
		Altered Ordered interface	0.30	0.02
		Altered Metal binding	0.26	0.01
		Altered Transmembrane protein	0.24	1.7e-03
		Loss of Allosteric site at R473	0.23	0.02
		Gain of N-linked glycosylation at N467	0.11	9.5e-03
		Loss of Catalytic site at W472	0.10	0.04
G475C	0.933	Altered DNA binding	0.34	1.1e-03
		Altered Metal binding	0.27	9.0e-03
		Altered Ordered interface	0.25	0.02
		Loss of Allosteric site at R473	0.24	0.02
		Altered Transmembrane protein	0.23	2.0e-03
		Loss of Catalytic site at W472	0.10	0.04
L476F	0.842	Altered DNA binding	0.29	4.0e-03
		Loss of Helix	0.28	0.02
		Altered Metal binding	0.26	0.01
		Altered Ordered interface	0.25	0.02
		Altered Transmembrane protein	0.23	2.3e-03
		Gain of Allosteric site at R473	0.23	0.02
		Loss of Catalytic site at W472	0.10	0.04
A480S	0.756	Altered Metal binding	0.32	0.01
		Gain of Allosteric site at F485	0.27	6.0e-03
		Altered Transmembrane protein	0.19	6.2e-03
I481M	0.747	Altered Metal binding	0.26	9.5e-03
		Gain of Allosteric site at H483	0.26	7.6e-03
		Altered Transmembrane protein	0.19	6.7e-03
H483Y	0.894	Altered Metal binding	0.78	5.8e-04
		Altered Ordered interface	0.35	9.3e-04
		Gain of Allosteric site at H483	0.24	0.02
		Altered Transmembrane protein	0.13	0.02
H483R	0.919	Altered Metal binding	0.75	6.8e-04
		Gain of Allosteric site at H483	0.29	2.3e-03
		Altered Transmembrane protein	0.13	0.02
L489S	0.629	Gain of Allosteric site at F485	0.25	0.01
		Altered Transmembrane protein	0.18	7.8e-03
		Altered Stability	0.15	0.02

		Gain of Sulfation at Y492	0.02	0.03
Y492C	0.871	Altered Ordered interface Gain of Helix Loss of Loop Altered Transmembrane protein Loss of Sulfation at Y492	0.31 0.27 0.27 0.17 0.01	3.7e-03 0.03 0.03 9.7e-03 0.04
D509G	0.849	Loss of Allosteric site at Y508 Altered Transmembrane protein	0.25 0.11	0.02 0.04
A512V	0.677	Gain of Allosteric site at Y508 Altered Transmembrane protein	0.28 0.10	3.9e-03 0.04
I516T	0.895	Altered Stability Altered Transmembrane protein	0.23 0.11	1.0e-02 0.03
P522T	0.822	Gain of helix	0.29	0.01
D524H	0.912	None	None	None
K533N	0.651	Gain of N-linked glycosylation at K533	0.02	0.03
Y539H	0.883	Altered Ordered interface	0.31	4.2e-03
Y539D	0.959	Altered Ordered interface Gain of Helix	0.32 0.27	3.3e-03 0.04
R544S	0.807	Altered Ordered interface	0.26	0.02
H546R	0.829	None	None	None
R877P	0.745	Altered Disordered interface Loss of Helix Gain of Loop Altered Coiled coil Loss of ADP-ribosylation at R877 Gain of Phosphorylation at S878 Loss of SUMOylation at K881 Gain of Ubiquitylation at K881	0.48 0.31 0.31 0.27 0.26 0.26 0.23 0.17	3.2e-03 4.0e-03 3.5e-03 0.02 8.1e-03 0.02 0.02 0.02
V880G	0.578	Gain of Loop Gain of B-factor Loss of SUMOylation at K881 Altered Disordered interface Loss of Phosphorylation at S878 Loss of ADP-ribosylation at R877 Gain of Ubiquitylation at K881	0.30 0.26 0.25 0.24 0.24 0.20 0.16	5.6e-03 0.01 9.0e-03 0.03 0.04 0.04 0.04
L1136P	0.929	Altered Coiled coil Gain of Intrinsic disorder Altered Disordered interface Altered Stability	0.93 0.32 0.20 0.09	6.1e-04 0.03 0.04 0.05
R1189C	0.896	Altered Disordered interface Loss of Intrinsic disorder Loss of Ubiquitylation at K1188	0.61 0.43 0.22	1.4e-04 0.02 3.3e-03

Evolutionary conservation of EHBP1 protein

To check the evolutionary profile of the EHBP1 variants, the selected nsSNPs were subjected to the ConSurf web server. Among 32 selected variants, ConSurf predicted 12 variants as highly conserved and exposed (W35R, D166G, N467S, T469I, T470A, D509G, P522T, D524H, K533N, H546R, R877P, R1189C). Additionally, 16 variants were predicted as highly conserved and buried (S448F, L449F, V456F, I465T, I465M, G475C, L476F, A480S, I481M, A512V, I516T, Y539H, Y539D, R544S, V880G, L1136P). Furthermore, four variants were predicted as buried among these, three (H483Y, H483R, Y492C) were conserved, while one variant was predicted to be variable (L489S) (Figure 3).

3D-Modelling of EHBP1 protein

The 3D structures of wild type and 32 mutant EHBP1 protein variants were generated through the I-TASSER tool, and Chimaera 1.11 were used for visualisation of the 3D structures of EHBP1 protein (Figure 4). I-TASSER data was submitted to TM-align, and RMSD score and TM-values were generated for all 32 damaging nsSNPs (Table 4).

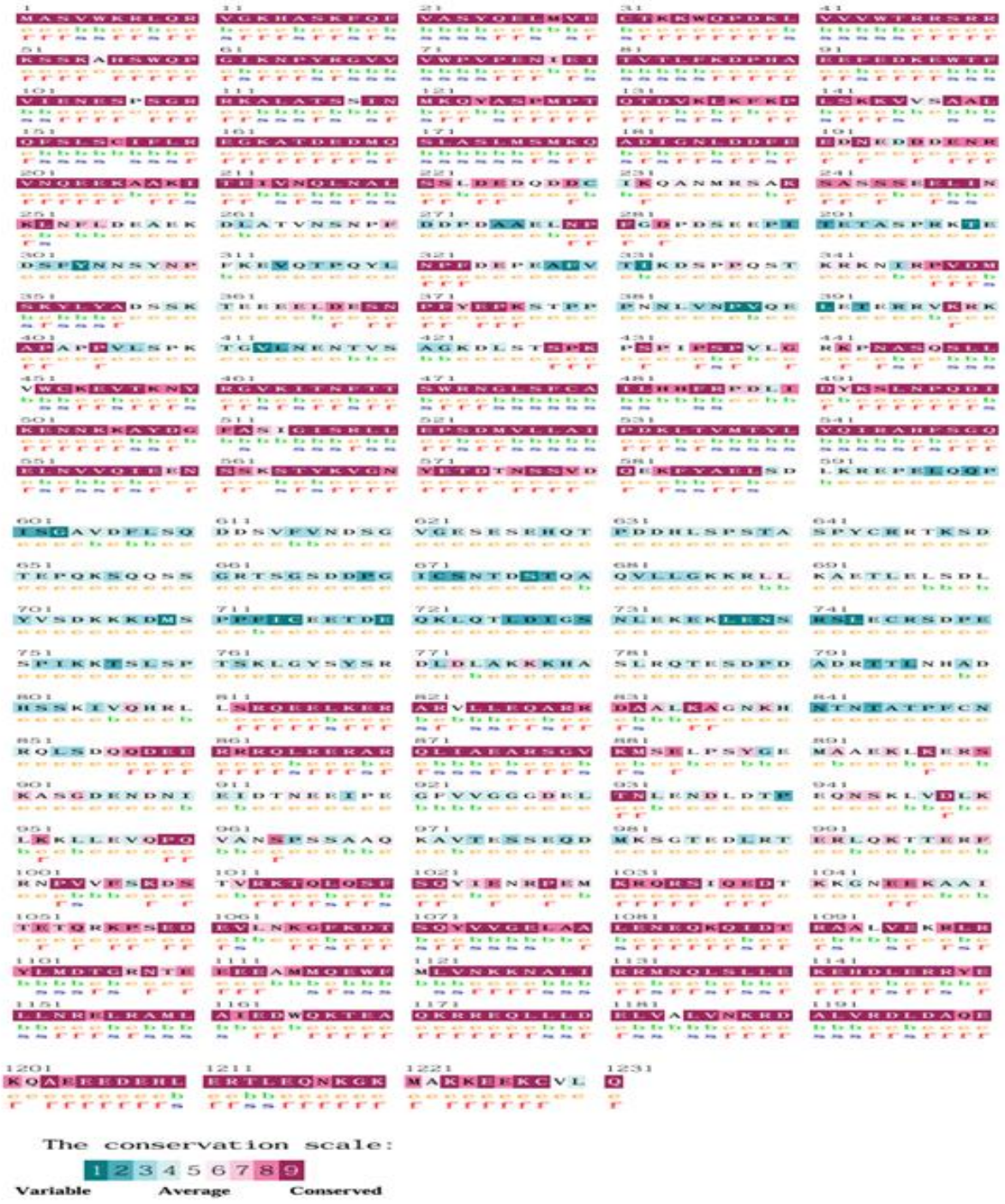
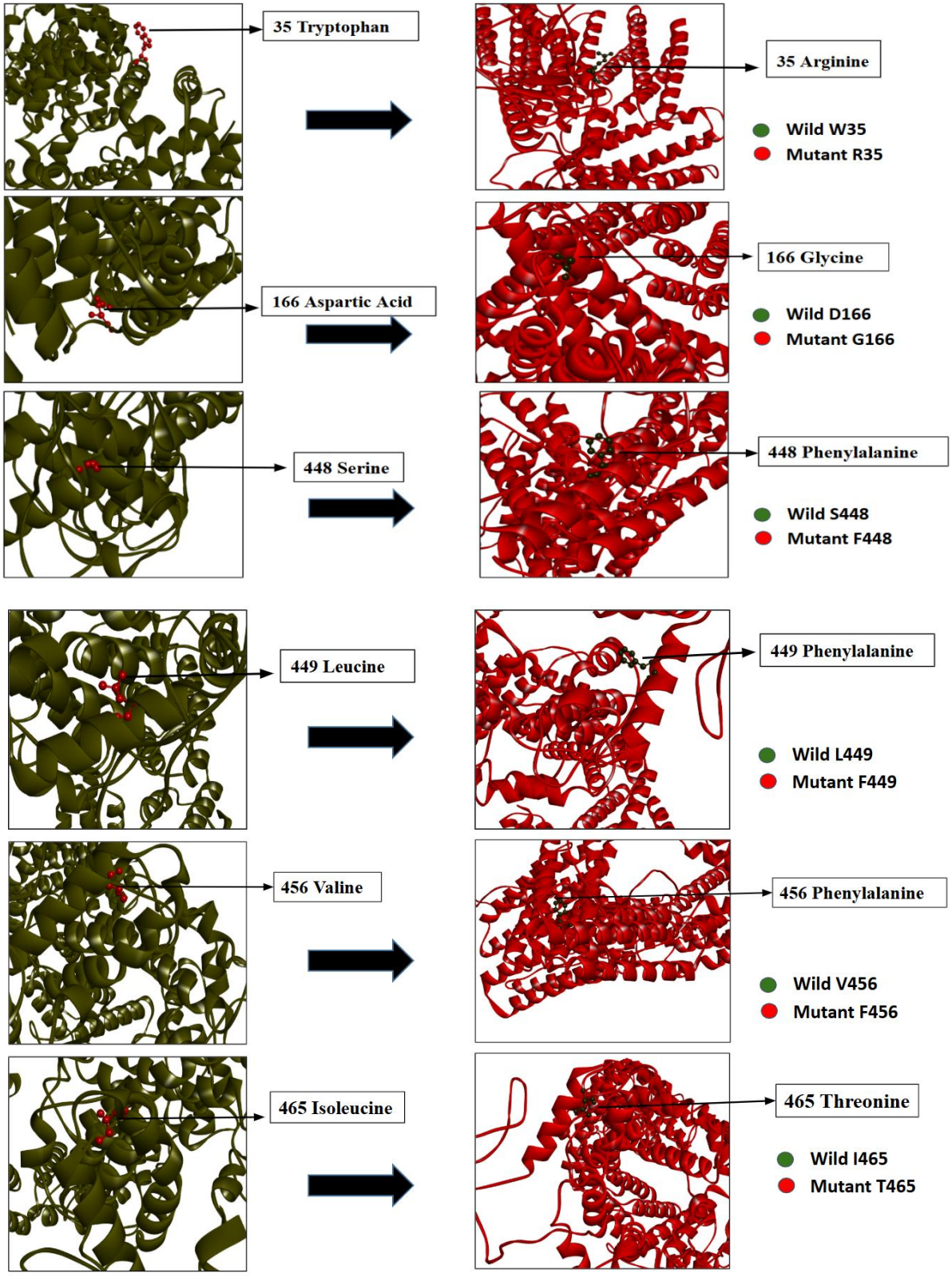
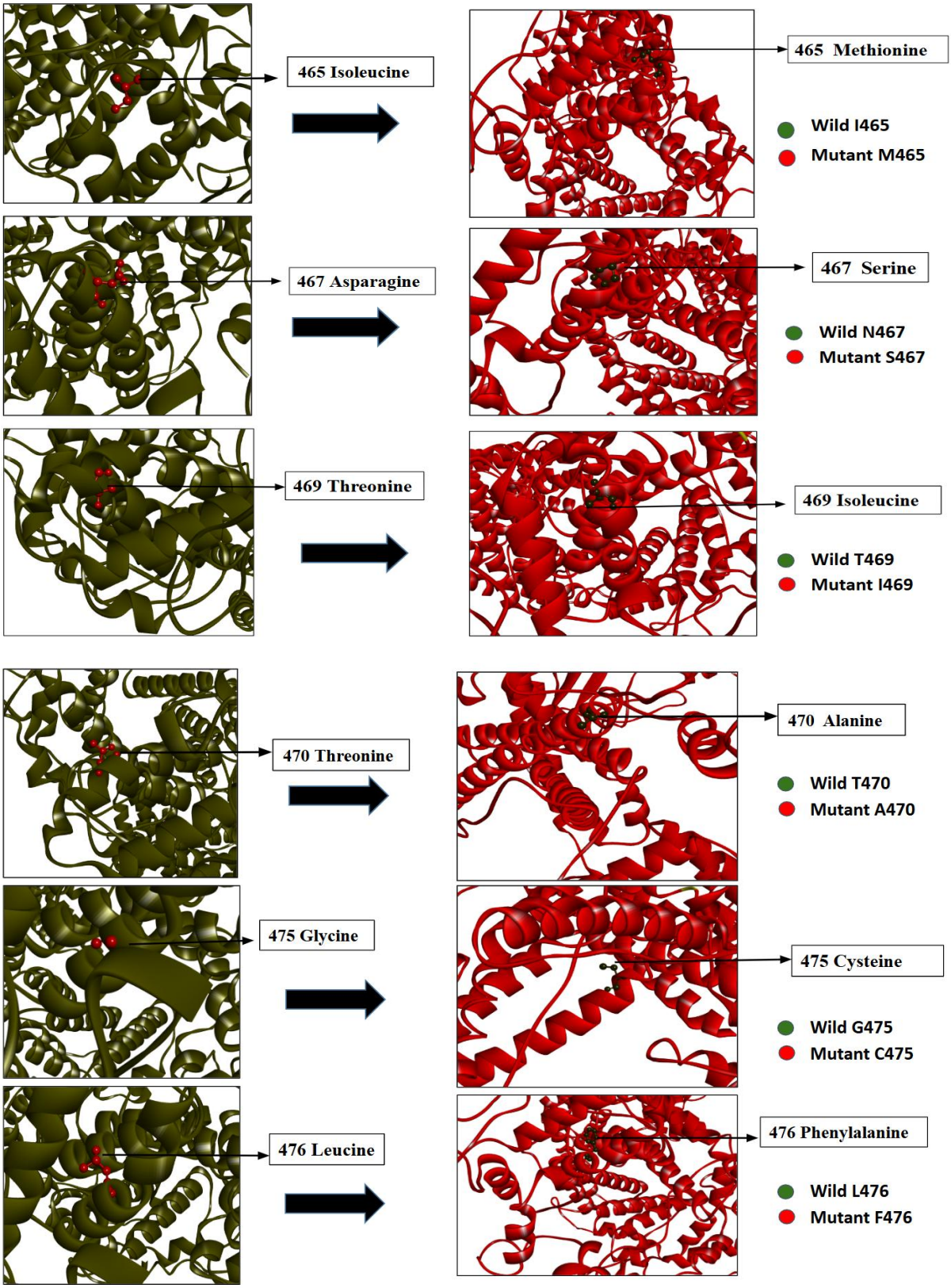
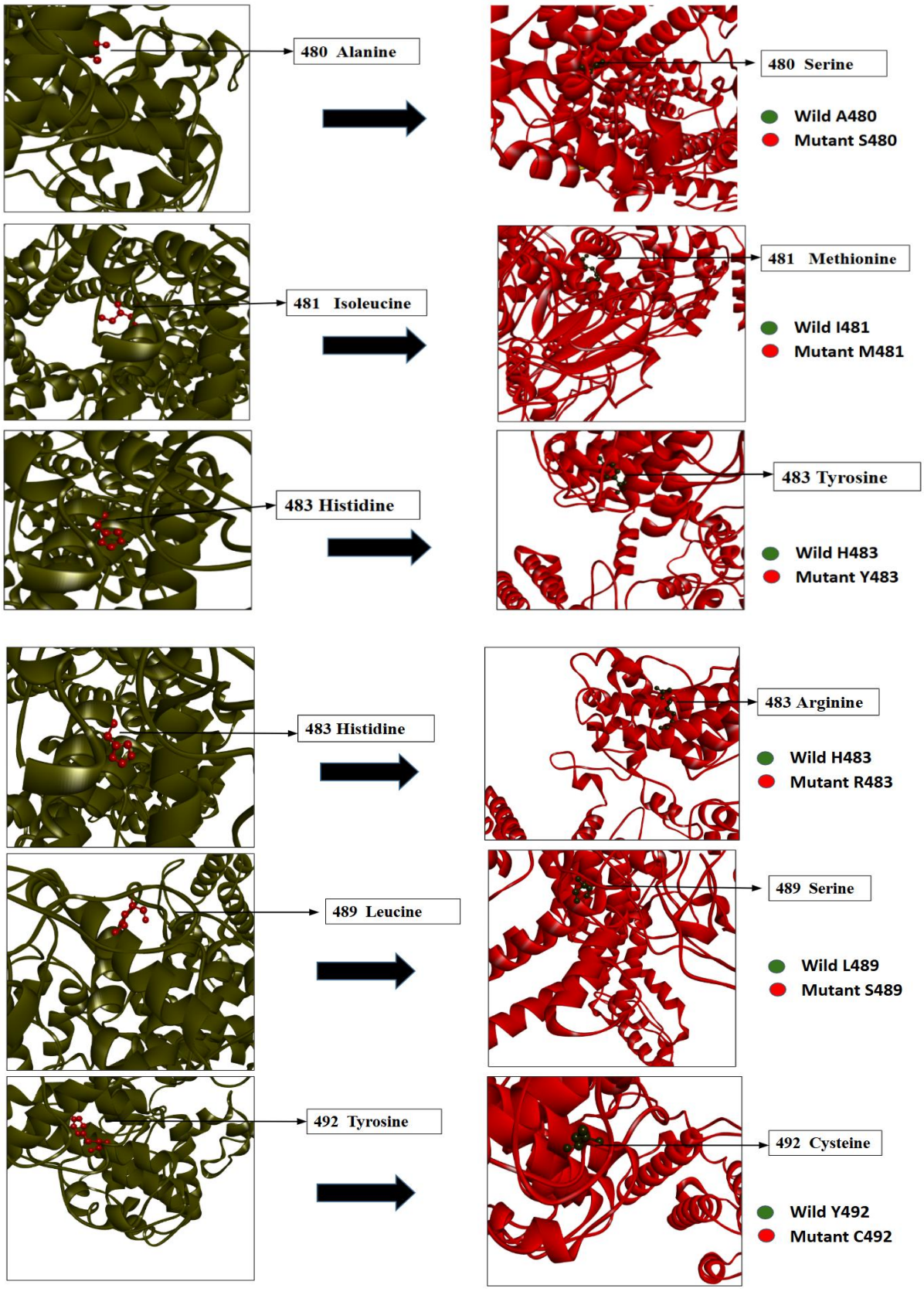
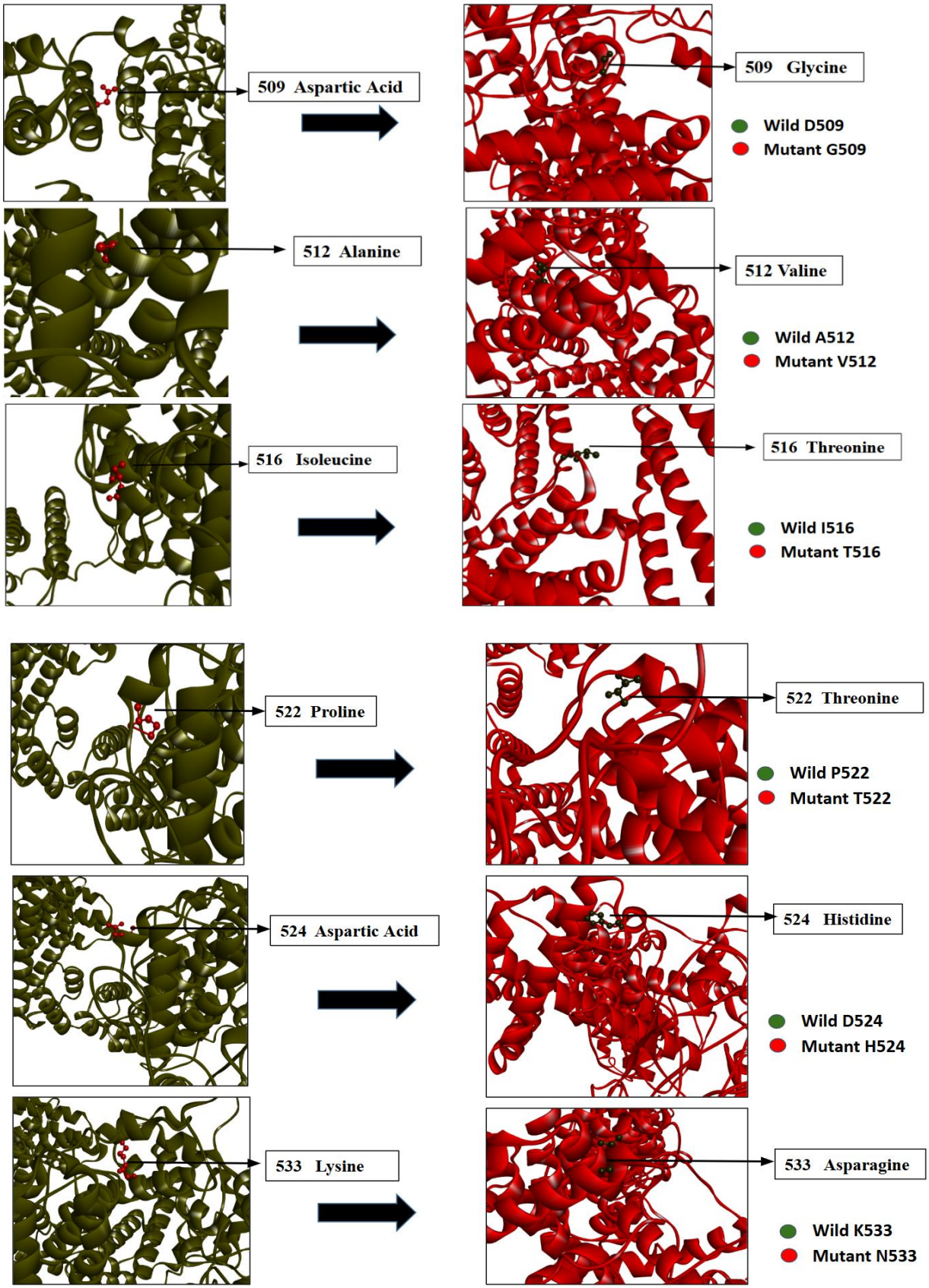


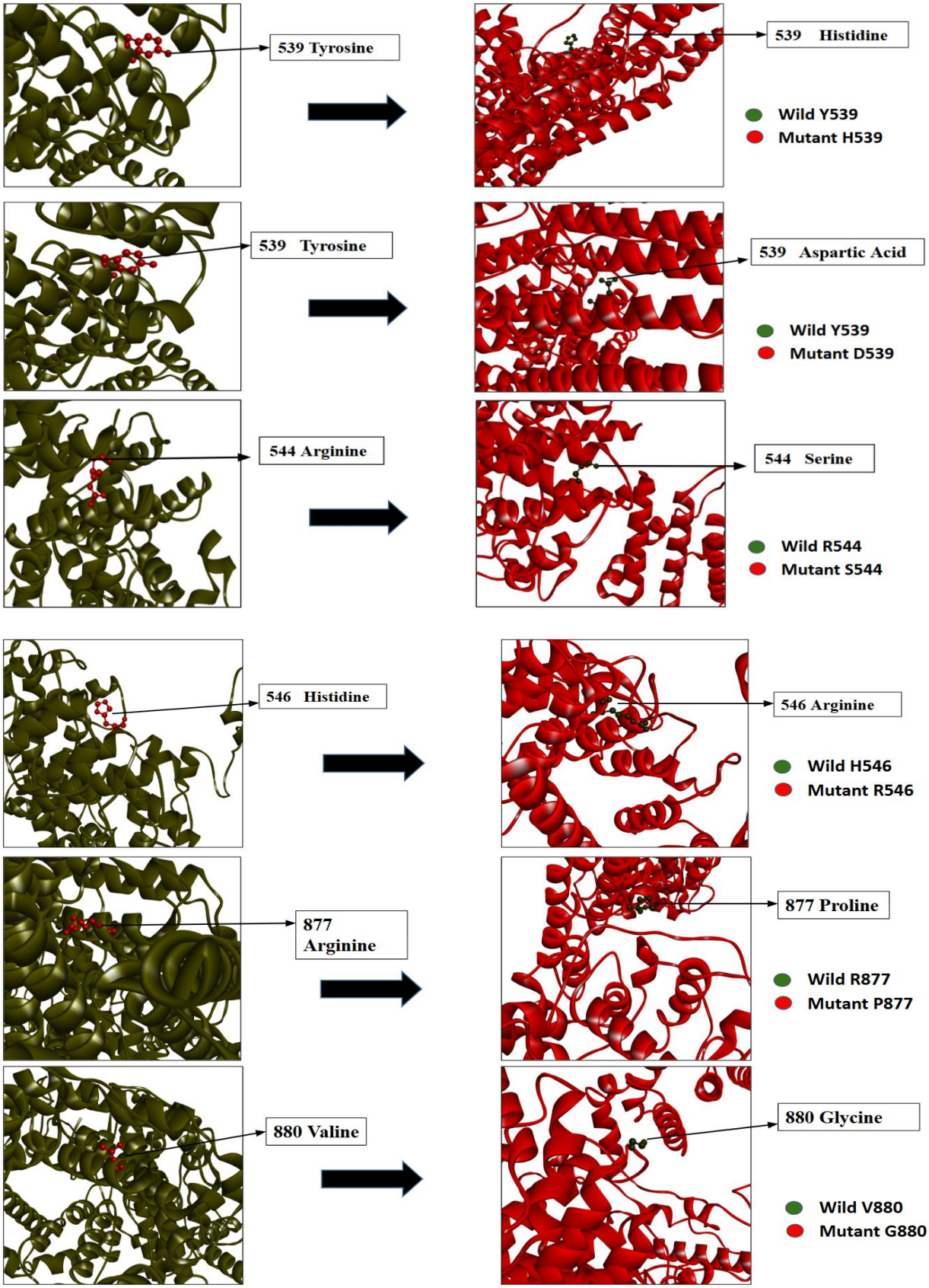
Figure 3. Prediction of ConSurf tool


35 Tryptophan
35 Arginine
Wild W35
Mutant R35
166 Aspartic Acid
166 Glycine
Wild D166
Mutant G166
448 Serine
448 Phenylalanine
Wild S448
Mutant F448
449 Leucine
449 Phenylalanine
Wild L449
Mutant F449
456 Valine
456 Phenylalanine
Wild V456
Mutant F456
465 Isoleucine
465 Threonine
Wild I465
Mutant T465


465 Isoleucine
465 Methionine
Wild I465
Mutant M465
467 Asparagine
467 Serine
Wild N467
Mutant S467
469 Threonine
469 Isoleucine
Wild T469
Mutant I469
470 Threonine
470 Alanine
Wild T470
Mutant A470
475 Glycine
475 Cysteine
Wild G475
Mutant C475
476 Leucine
476 Phenylalanine
Wild L476
Mutant F476


480 Alanine
480 Serine
Wild A480
Mutant S480
481 Isoleucine
481 Methionine
Wild I481
Mutant M481
483 Histidine
483 Tyrosine
Wild H483
Mutant Y483
483 Histidine
483 Arginine
Wild H483
Mutant R483
489 Leucine
489 Serine
Wild L489
Mutant S489
492 Tyrosine
492 Cysteine
Wild Y492
Mutant C492


509 Aspartic Acid
509 Glycine
Wild D509
Mutant G509
512 Alanine
512 Valine
Wild A512
Mutant V512
516 Isoleucine
516 Threonine
Wild I516
Mutant T516
522 Proline
522 Threonine
Wild P522
Mutant T522
524 Aspartic Acid
524 Histidine
Wild D524
Mutant H524
533 Lysine
533 Asparagine
Wild K533
Mutant N533


539 Tyrosine
539 Histidine
Wild Y539
Mutant H539
539 Tyrosine
539 Aspartic Acid
Wild Y539
Mutant D539
544 Arginine
544 Serine
Wild R544
Mutant S544
546 Histidine
546 Arginine
Wild H546
Mutant R546
877 Arginine
877 Proline
Wild R877
Mutant P877
880 Valine
880 Glycine
Wild V880
Mutant G880

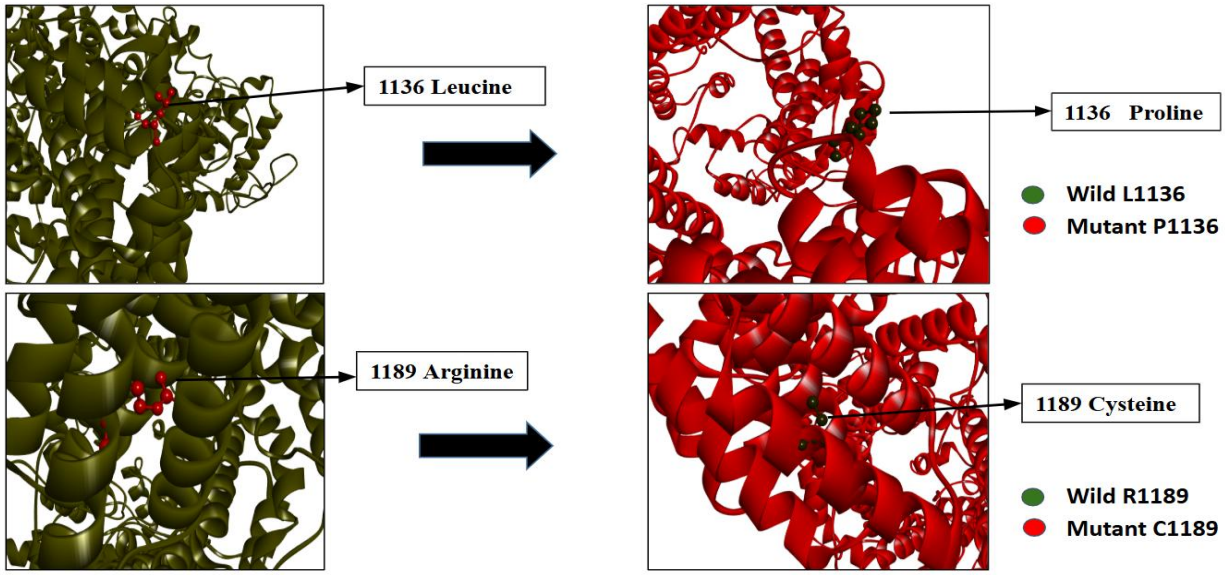


Figure 4. 3D modelling of the wild-type and mutant (each of the 32 SNPs) EHBP1protein

Table 4. TM-Align and RMSD value of deleterious nsSNPs

Amino acids change	TM Align	RMSD	AMINO ACID	TM Align	RMSD
W35R	0.24432	9.66	**L489S**	0.26562	8.85
D166G	0.22818	10.06	**Y492C**	0.87532	2.74
S448F	0.23203	10.22	**D509G**	0.27276	9.20
L449F	0.89150	2.93	**A512V**	0.92010	2.87
V456F	0.23164	9.92	**I516T**	0.22776	10.00
I465T	0.89655	2.83	**P522T**	0.87252	2.99
I465M	0.89540	2.94	**D524H**	0.28540	9.48
N467S	0.88430	2.81	**K533N**	0.27955	9.43
T469I	0.88971	3.12	**Y539H**	0.23522	10.01
T470A	0.23167	9.83	**Y539D**	0.28603	9.99
G475C	0.22563	9.61	**R544S**	0.91539	2.57
L476F	0.85869	2.86	**H546R**	0.78239	3.10
A480S	0.87714	2.73	**R877P**	0.26193	9.22
I481M	0.29648	9.22	**V880G**	0.86112	3.22
H483Y	0.89081	3.00	**L1136P**	0.88047	2.73
H483R	0.88536	2.78	**R1189C**	0.89398	2.43

Protein-Protein interaction of EHBP1

STRING database was utilized to predict the interaction of EHBP1 protein with other proteins. Results showed a complex network of proteins comprised of 11 nodes, 25 edges, and an average node degree of 4.55. Additionally, the PP1 enrichment p-value was 6.91e-05. Local clustering coefficient was 0.853, which indicates the network is highly clustered (Figure 5)

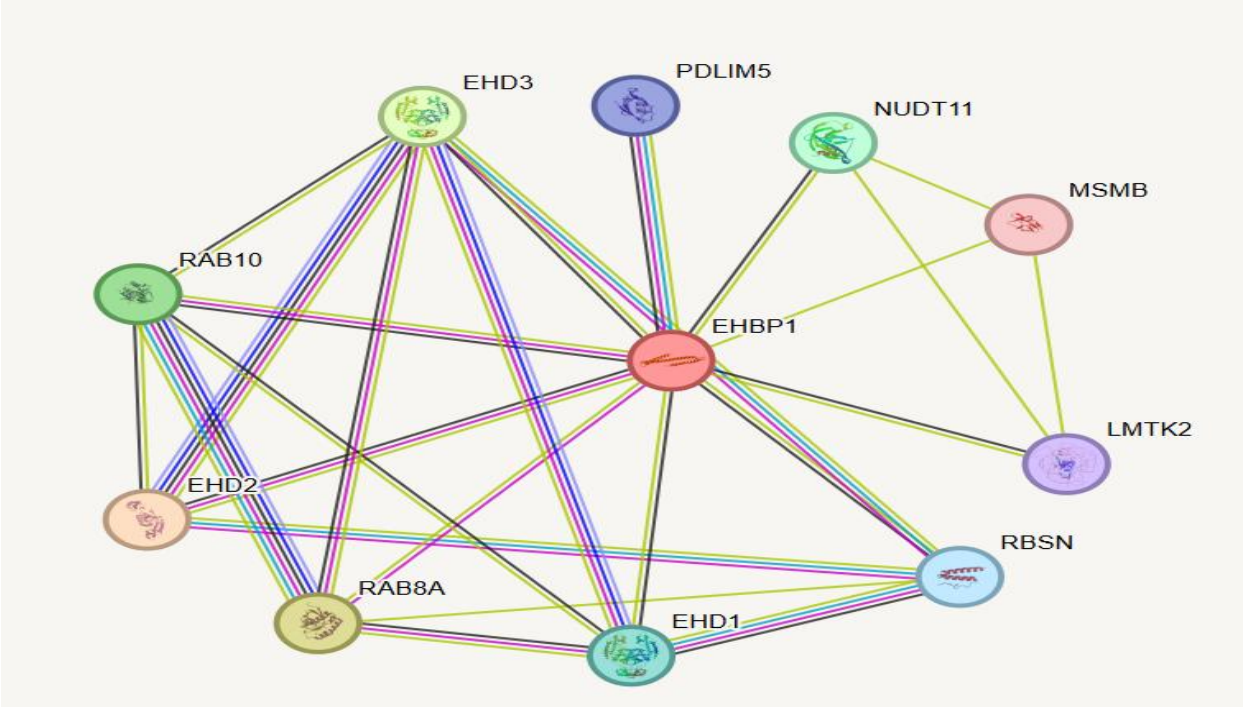


Figure 5. Protein –protein interaction of EHBP1 by STRING database

Gene-Gene interaction of *EHBP*1 gene

The GeneMANIA tool was used to predict the *EHBP1* gene's interaction with other genes. Results showed *EHBP1* gene physically interacts with *KHDRBS2, RPAP2, SOAT 1, RAB10, NDUFS1, EHBP1L1, ACAP2, EHD2, PACSIN1* and *IFT57*. Additionally, its co-expression with *IP6K1, KHDRBS2, SOX5, VPS8, RNF103, KLF12, FAM102B, FAM102A, BMERB1, IFT57, EHD2, BPTF, EHBP1L1, NDUFS1* and *RPAP2*. Moreover, *EHBP1* gene has genetic interaction with *KLF12, RNF103, VPS8, SOX5, KHDRBS2, IP6K1, IP6K3, RPAP2, SOAT1, RAB10, NDUFS1, ACAP2, BPTF, EHD2, PACSIN1, IFT57, BMERB1* and *FAM102A*. Lastly, the *EHBP1* gene was shown to have common protein domains with *FAM102B, FAM102A, BMERB1* and *EHBP1L1* (Figure 6).

Post-Translational Modification Sites Prediction

The MusiteDeep web server was utilized to predict the PTM sites in the EHBP1 protein. Overall, 111 PTM sites were predicted, among which 70 were predicted to be serine phosphorylated, 13 were predicted to be threonine phosphorylated, six were predicted to be tyrosine phosphorylated, four were predicted to be lysine methylated, and 18 were predicted to be lysine ubiquitinated (Figure 7).

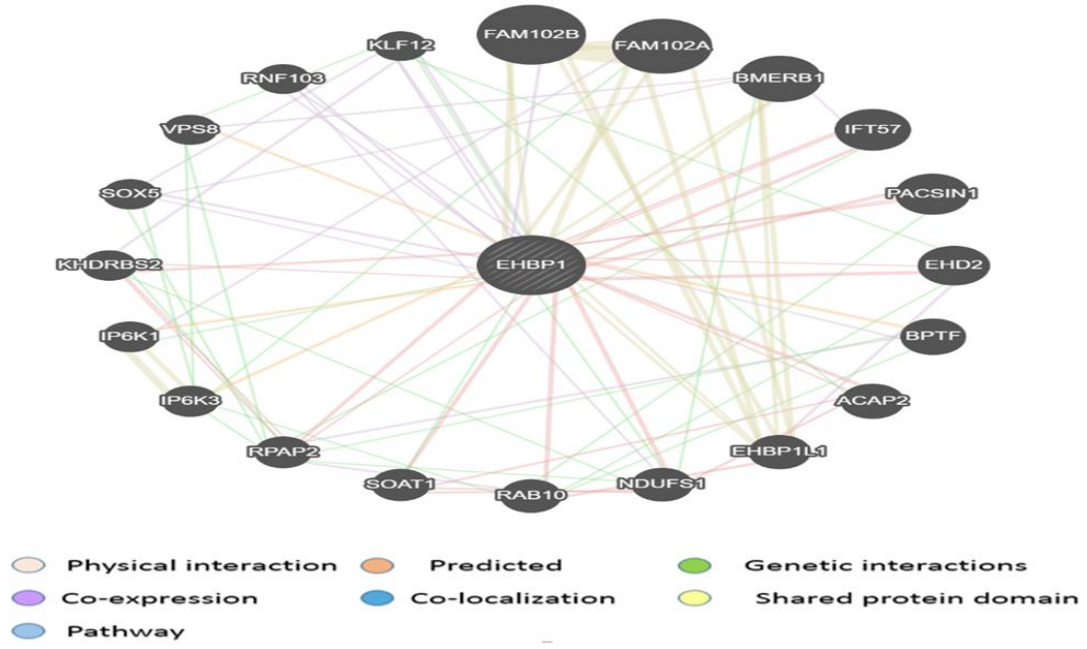


Figure 6. *EHBP1* gene interaction with other genes by GeneMania

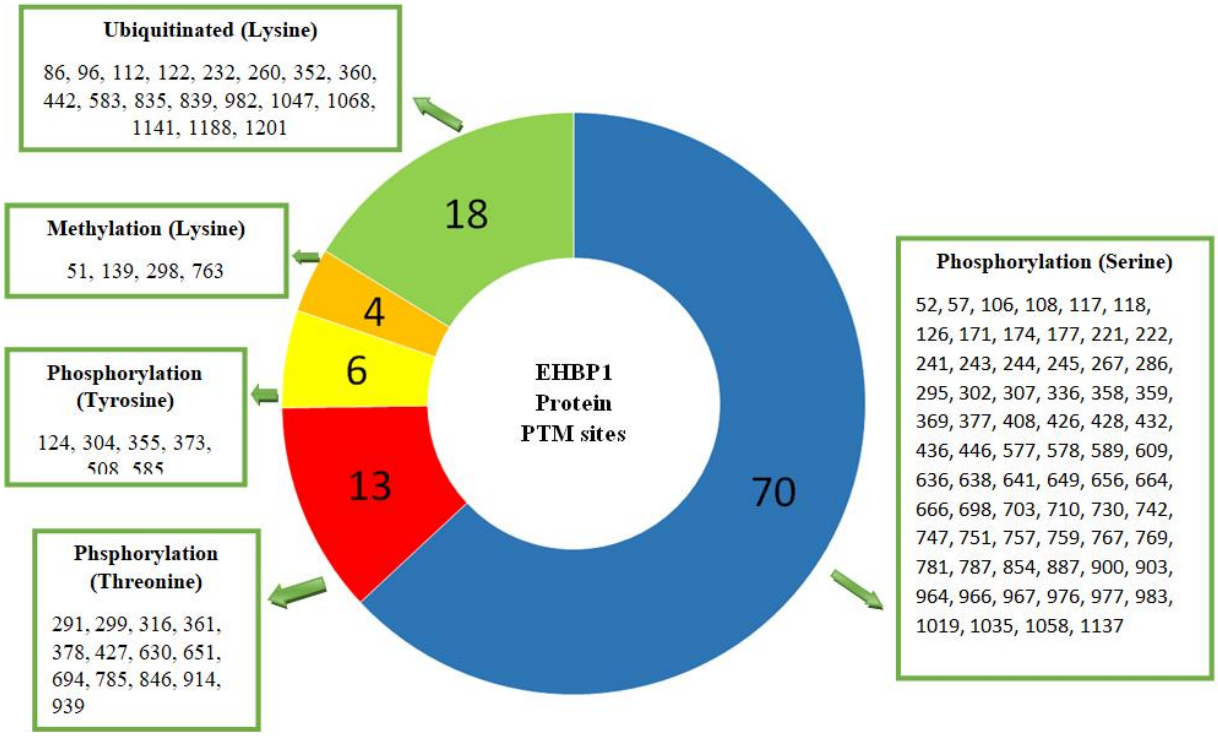


Figure 7. Post-translational modification sites predicted by the MusiteDeep web server

DISCUSSION

Multiple studies have reported an association between EHBP1 variants, most notably rs721048, and an increased risk of prostate cancer in Caucasian cohorts (Gudmundsson et al., 2008; Lindstrom et al., 2011; Tsilidis et al., 2012; Ao et al., 2015). Additional EHBP1 polymorphisms (e.g., rs10496099, rs2710642) have also been linked to high LDL-C (low-density lipoprotein cholesterol), indicating pleiotropic roles for this gene (Liu et al., 2022). Against this background, prioritizing missense variants that are most likely to perturb EHBP1 structure and function is an essential first step toward clarifying potential mechanisms of disease susceptibility.

Using a multi-tool pipeline (SIFT, SNPs&GO, PolyPhen-2, PhD-SNP, and fathmm), we systematically filtered 1,075 EHBP1 missense variants down to 32 nsSNPs that were consistently predicted as deleterious (Table 1). This consensus approach reduces tool-specific bias and increases confidence that the shortlisted variants merit functional follow-up. Notably, many of these substitutions map to evolutionarily conserved positions, either exposed or buried, per ConSurf, supporting their potential structural and functional importance (Ashkenazy et al., 2016).

Protein stability predictions were largely concordant across I-Mutant and MUpro: 30/32 variants were predicted by I-Mutant to decrease stability, with two (S448F and K533N) predicted to increase stability, whereas MUpro predicted reduced stability for all 32 (Table 2). While algorithmic differences can explain these minor discrepancies, the predominant trend toward destabilization suggests that altered EHBP1 folding/robustness may be a common consequence of the prioritized variants.

MutPred analyses further indicate that these nsSNPs could perturb multiple molecular mechanisms, including ordered/disordered interfaces, metal binding, allosteric sites, secondary structure elements, glycosylation and phosphorylation propensities, and features relevant to membrane association (Li et al., 2009). Notably, Y539D had the highest MutPred score (0.959), while variants such as G475C, L1136P, I465T, H483R, R1189C, I516T, and H483Y also scored highly, making them strong candidates for experimental interrogation. Complementing these functional predictions, TM-align/RMSD comparisons between wild-type and mutant models highlighted substantial structural deviations for S448F, D166G, Y539H and I516T (RMSD ≈ 10 Å), consistent with a high perturbation likelihood in localized regions (Zhang & Skolnick, 2005; Zhang, 2008; Pettersen et al., 2004).

Network-level analyses reinforce the potential for broader pathway impacts. STRING revealed a densely connected EHBP1 protein–protein interaction (PPI) network with high local clustering, while GeneMANIA indicated physical interaction, co-expression, genetic interaction, and shared domains with partners including EHD2, RAB10, EHBP1L1, and others. Because EHBP1 coordinates endocytic trafficking and actin cytoskeleton dynamics (Gudmundsson et al., 2008; Rai et al., 2020), variants that destabilize the protein or disrupt interfaces could propagate to vesicular transport, membrane remodeling, and signaling pathways- biological processes plausibly relevant to tumorigenesis and, specifically, prostate-cancer risk (Wallis & Nam, 2015; Liang et al., 2020; Ao et al., 2015).

Finally, MusiteDeep predicted abundant and diverse PTM sites (phosphorylation, ubiquitination, methylation), suggesting that EHBP1 function is likely modulated by complex post-translational control (Wang et al., 2020). Variants that alter PTM motifs or local structure could alter post-translational modifications and hence protein structures and function.

Our findings are based on in-silico predictions and homology-based models; they require

experimental validation in appropriate cellular and biochemical systems, and population-based studies in diverse populations to test associations with prostate cancer and related traits. We also note that our structural assessments used I-TASSER models rather than experimentally determined EHBP1 structures; thus, high-resolution structural work would be valuable to refine variant effect interpretations (Zhang, 2008; Pettersen et al., 2004). Despite these limitations, convergent evidence across conservation, stability, structural deviation, and functional prediction supports the biological relevance of the 32 prioritized nsSNPs.

CONCLUSION

In summary, a consensus in-silico workflow prioritized 32 EHBP1 nsSNPs with a high likelihood of deleterious impact on protein stability, structure, and function. Many lies at conserved residues, displaying large predicted structural deviations, and are implicated in perturbing key molecular mechanisms and interaction interfaces. Given EHBP1's role in processes tied to cancer biology, these variants constitute strong candidates for experimental validation and population-level association testing.

Conflict of Interest: The Authors declared no conflict of interest

Ethical Approval: No ethical approval or informed consent was required as the study did not involve any human or animal subjects or any first-hand data collection

Funding: The work was conducted as part of the research funded by the RIF of Abdul Wali Khan University, Mardan

AUTHORS' CONTRIBUTIONS:

Concept and Design: NK, FB**; Data Retrieval and Analysis:** FB with inputs from NK, NH TI, MQK; **Manuscript drafting:** Fajar Baig, with inputs from Naveed Khan**; Editing and Revision**: Naveed Khan, NH, MQK, TI

REFERENCES

ADZHUBEI, I., JORDAN, D. M. & SUNYAEV, S. R. 2013. Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics,* 76**,** 7.20. 1-7.20. 41.

AKHTAR, M., JAMAL, T., UD DIN, J., HAYAT, C., RAUF, M., UL HAQ, S. M., KHAN, R. S., SHAH, A. A., JAMAL, M. & JALIL, F. 2019. An in silico approach to characterize nonsynonymous SNPs and regulatory SNPs in human TOX3 gene. *Journal of genetics,* 98**,** 104.

AKHTAR, S., HASSAN, F., AHMAD, S., EL-AFFENDI, M. A. & KHAN, M. I. 2023. The prevalence of prostate cancer in Pakistan: A systematic review and meta-analysis. *Heliyon,* 9.

AO, X., LIU, Y., BAI, X.-Y., QU, X., XU, Z., HU, G., CHEN, M. & WU, H. 2015. Association between EHBP1 rs721048 (A> G) polymorphism and prostate cancer susceptibility: a meta-analysis of 17 studies involving 150,678 subjects. *OncoTargets and therapy***,** 1671-1680.

ASHKENAZY, H., ABADI, S., MARTZ, E., CHAY, O., MAYROSE, I., PUPKO, T. & BEN-TAL, N. 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research,* 44**,** W344-W350.

BEIKZADEH, B., ANGAJI, S. A. & ABOLHASANI, M. 2020. Association study between common variations in some candidate genes and prostate adenocarcinoma predisposition through multi-stage approach in Iranian population. *BMC Medical Genetics,* 21**,** 81.

CALABRESE, R., CAPRIOTTI, E., FARISELLI, P., MARTELLI, P. L. & CASADIO, R. 2009. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Human mutation,* 30**,** 1237-1244.

CAPRIOTTI, E., CALABRESE, R. & CASADIO, R. 2006. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics,* 22**,** 2729-2734.

CHENG, J., RANDALL, A. & BALDI, P. 2006. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins: Structure, Function, and Bioinformatics,* 62**,** 1125-1132.

GUDMUNDSSON, J., SULEM, P., RAFNAR, T., BERGTHORSSON, J. T., MANOLESCU, A., GUDBJARTSSON, D., AGNARSSON, B. A., SIGURDSSON, A., BENEDIKTSDOTTIR, K. R. & BLONDAL, T. 2008. Common sequence variants on 2p15 and Xp11. 22 confer susceptibility to prostate cancer. *Nature genetics,* 40**,** 281-283.

JENSEN, L. J., KUHN, M., STARK, M., CHAFFRON, S., CREEVEY, C., MULLER, J., DOERKS, T., JULIEN, P., ROTH, A. & SIMONOVIC, M. 2009. STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic acids research,* 37**,** D412-D416.

LI, B., KRISHNAN, V. G., MORT, M. E., XIN, F., KAMATI, K. K., COOPER, D. N., MOONEY, S. D. & RADIVOJAC, P. 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics,* 25**,** 2744-2750.

LIANG, B., DING, H., HUANG, L., LUO, H. & ZHU, X. 2020. GWAS in cancer: progress and challenges. *Molecular Genetics and Genomics,* 295**,** 537-561.

LINDSTROM, S., SCHUMACHER, F., SIDDIQ, A., TRAVIS, R. C., CAMPA, D., BERNDT, S. I., DIVER, W. R., SEVERI, G., ALLEN, N. & ANDRIOLE, G. 2011. Characterizing associations and SNP-environment interactions for GWAS-identified prostate cancer risk markers—results from BPC3. *PloS one,* 6**,** e17142.

LIU, C.-X., YIN, R.-X., CAO, X.-L., SHI, Z.-H., HUANG, F., WEI, B.-L., DENG, G.-X., ZHENG, P.-F. & GUAN, Y.-Z. 2022. EHBP1, TUBB, and WWOX SNPs, gene-gene and gene-environment interactions on coronary artery disease and ischemic stroke. *Frontiers in Genetics,* 13**,** 843661.

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry,* 25**,** 1605-1612.

RAI, A., BLEIMLING, N., VETTER, I. R. & GOODY, R. S. 2020. The mechanism of activation of the actin binding protein EHBP1 by Rab8 family members. *Nature communications,* 11**,** 4187.

SHIHAB, H. A., GOUGH, J., COOPER, D. N., STENSON, P. D., BARKER, G. L., EDWARDS, K. J., DAY, I. N. & GAUNT, T. R. 2013. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human mutation,* 34**,** 57-65.

SIM, N.-L., KUMAR, P., HU, J., HENIKOFF, S., SCHNEIDER, G. & NG, P. C. 2012. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic acids research,* 40**,** W452-W457.

TSILIDIS, K. K., TRAVIS, R. C., APPLEBY, P. N., ALLEN, N. E., LINDSTROM, S., SCHUMACHER, F. R., COX, D., HSING, A. W., MA, J. & SEVERI, G. 2012. Interactions between genome-wide significant genetic variants and circulating concentrations of insulin-like growth factor 1, sex hormones, and binding proteins in relation to prostate cancer risk in the National Cancer Institute Breast and Prostate Cancer Cohort

Consortium. *American journal of epidemiology,* 175**,** 926-935.
WALLIS, C. J. & NAM, R. K. 2015. Prostate cancer genetics: a review. *Ejifcc,* 26**,** 79.
WANG, D., LIU, D., YUCHI, J., HE, F., JIANG, Y., CAI, S., LI, J. & XU, D. 2020. MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. *Nucleic Acids Research,* 48**,** W140-W146.
WARDE-FARLEY, D., DONALDSON, S. L., COMES, O., ZUBERI, K., BADRAWI, R., CHAO, P., FRANZ, M., GROUIOS, C., KAZI, F. & LOPES, C. T. 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research,* 38**,** W214-W220.
ZHANG, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics,* 9**,** 40.
ZHANG, Y. & SKOLNICK, J. 2005. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic acids research,* 33**,** 2302-2309.