International Journal of
Molecular Microbiology

Research Article 2025 | Volume 8 | Issue 1 | 79-102

Open Access

Article Information

Received: October 11, 2025

Accepted: November 1, 2025

Published: November 10, 2025

Keywords

HBV,
NCBI,
nsSNPs,
STAT4,
SNPnexus.

Authors' Contribution

AN and KR conceived and designed the study. All authors contributed equally to the conception, design, analysis, and writing of this manuscript.

How to cite

Sardar, N., Noman, A., Ramzan, K., Bilal, K., Islam, A., Ali, M.Z., Tahir, H.U., Akram, N., Siddique, A.H., Usman, M., 2025. Exploring the Impact of STAT4 Non-Synonymous Mutations on Hepatitis B Virus Susceptibility: A Bioinformatics Approach. Int. J. Mol. Microbiol., 8(1): 79-102.

***Correspondence**

Ali Noman
Email: alinoman1907@gmail.com

Possible submissions

Submit your article

Exploring the Impact of STAT4 Non-Synonymous Mutations on Hepatitis B Virus Susceptibility: A Bioinformatics Approach

Nimra Sardar[1], Ali Noman[2*], Kainat Ramzan[3], Ibtsam Bilal[3], Amina Islam[4], Muhammad Zeeshan Ali[4], Habib Ullah Tahir[4], Nimra Akram[5], Ali Hamza Siddique[5], Muhammad Usman[6]

[1]Center of Excellence in Molecular Biology, University of the Punjab, Lahore, Punjab, Pakistan.
[2]Department of Zoology, Faculty of Life Sciences, University of Okara, Punjab, Pakistan.
[3]Department of Biochemistry, University of Okara, Punjab, Pakistan.
[4]Department of Microbiology and Molecular Genetics, University of Okara, Punjab, Pakistan.
[5]Department of Molecular Biology, University of Okara, Punjab, Pakistan.
[6]Faculty of Science and Technology, University of the Central Punjab, Lahore, Punjab, Pakistan.

Abstract:

Hepatitis B virus (HBV) remains a significant global health challenge, with an estimated 257 to 350 million individuals living with chronic HBV infection worldwide. Based on the results of genome-wide association studies (GWAS), the STAT4 gene was identified as a viable target for further study into its association with HBV-related liver diseases. The *STAT4* signaling pathway is renowned for its critical role in interferon gamma-mediated antiviral response, emphasizing its importance in HBV infection. The present study employed bioinformatics tools to discover potentially damaging nsSNPs and assess the negative consequences of *STAT4* gene mutations. The structural prediction, template refinement, and blind protein-ligand docking studies were assessed using various computational approaches to gain a comprehensive understanding of the functional consequences of the STAT4 gene mutations. From a total of 4749 nsSNPs obtained from the NCBI database, 15 missense SNPs were recognized as detrimental by 13 different bioinformatic tools, including SNPnexus>PolyPhen-2>Predict SNP>Mutpred2>PANTHER>SNP& GO>Meta SNP, and SuSpect, among others. Protein stability was assessed using iStable, Mu Pro, CUPSAT, and Dynamut2 databases. The InBio Discover tool was used to study protein interactions, while the GeneMANIA tool was used to investigate gene-gene interactions. Mutation 3D and PTM sites were assessed to confirm SNPs' detrimental nature. To validate nsSNPs such as D668G (rs751205891), C539Y (rs774187563), R508C (rs780829180), T336S (rs770753645), and P331Q (rs1248978329), protein modeling, structural validation, and protein-ligand interaction studies were performed. These nsSNPs could serve as targets for future *STAT4*-related disease research. Despite the inherent drawbacks of computational tools, the studies executed in the present work provide valuable perceptions and may serve as a resource for future in vitro and in vivo studies on STAT4's involvement in the immune response to HBV infection.

Scan QR code to visit this journal.

INTRODUCTION

The intricate distinction between phenotype and genotype is a key difficulty in studying genomics. The primary concern is identifying genetic polymorphism linked with human hereditary disorders, which remains a target in human and medical genomics. DNA variations are genetic biomarkers, and substantial research efforts have focused on elucidating the interactions between genetic mutations and their associated phenotypic effects (Claussnitzer *et al.*, 2020; Kumar *et al.*, 2022). A Single Nucleotide Variant (SNV) is a broader term that refers to any variation in a single nucleotide (A, T, C, or G) in the genome, which may occur anywhere in the genome, including coding regions, non-coding regions, or intergenic regions (Dou *et al.*, 2024). Single-nucleotide polymorphism (SNP) is a specific type of SNV that is common within a population. To be classified as an SNP, the variant must occur in at least 1% of the population (Naranjo-Galvis *et al.*, 2023). Missense or non-synonymous (nsSNPs) mutation involves a substitution of amino acid sequences and is involved in many inherited disorders; not all nsSNPs are harmful. Many nsSNPs are neutral or have no significant effect on protein function. The impact of an nsSNP depends on factors such as its location in the protein, the properties of the substituted amino acid, and the specific role of the protein in biological processes (Emadi *et al.*, 2020).

These nsSNPs have the potential to negatively impact protein composition, stability, and activity (Lander, 1996; Ng and Henikoff, 2002) by altering the charge, form, hydrophobicity, protein synthesis, and interaction between proteins (Collins *et al.*, 1998; Kucukkal *et al.*, 2015). Humans have made tremendous progress in revealing hundreds of millions of SNPs that were linked to complex clinical problems and phenotypic features associated with an array of well-known disorders (Arshad *et al.*, 2018; Krawczak *et al.*, 2000). Multiple analyses have shown that nearly half of all genetic aberrations correspond to at least one type of variation (Radivojac *et al.*, 2010). Over a decade, genome-wide association studies (GWAS) had an immense effect on the research of major hereditary diseases, revealing several important contrasts between various mammalian traits and pathologies (Tam *et al.*, 2019). Studies have identified an association between the Signal transducer and activator of transcription 4 (*STAT4*) mutations and an increased risk of hepatitis B virus, also known as HBV, and hepatocellular tumor genesis (Shi *et al.*, 2019). The *STAT4* signaling cascade has significance for interferon gamma-mediated antiviral responses (Lu *et al.*, 2015). Prior studies in varied ethnic communities have looked at the impact of the *STAT4* variant (rs7574865) on HBV-induced hepatocellular carcinoma (HCC) incidence (Li *et al.*, 2019).

Hepatitis B virus (HBV) remains a significant global health challenge, with an estimated 257 to 350 million individuals living with chronic HBV infection worldwide (Tan *et al.*, 2021). Despite the availability of effective HBV vaccines since 1982, the disease continues to have a high prevalence, particularly in Africa and the Western Pacific, including China (Zeng *et al.*, 2021). The global prevalence of Hepatitis B surface antigen (HBsAg), a key indicator of HBV infection, is estimated to be around 3.6%, with endemic areas experiencing the highest burden (Sheena *et al.*, 2022). Recent studies have explored the correlation between the STAT4 mutation (rs7574865) and the progression of HBV in individuals with persistent infection. However, the findings of such studies were conflicting and untrustworthy (Liao *et al.*, 2014; Shi *et al.*, 2019). Furthermore, HBV transmission and HCC development are complex events involving numerous mutations in a variety of biological processes (Jiang *et al.*, 2021).

The activity of STAT4 in HBV infection is thought to promote the production of Th1 and Th2 cytokines, leading to increased hepatic inflammation (Wang *et al.*, 2014). The STAT4 gene (*SLEB11* and *DPMC)*, located on chromosome 2q32.2-q32.3, encodes a 748-amino-acid protein critical for the Janus kinase (JAK)-STAT pathway (Qi *et al.*, 2022). This pathway influences immune cell activation and response to therapies like IFN-alpha, making STAT4 mutations relevant to understanding clinical variability in chronic HBV treatment

outcomes. Given the role of STAT4 in hepatic inflammation and immune regulation, its mutations may influence susceptibility to HBV infection and disease progression (Jiang *et al.*, 2016). The identification of functional mutations in a disease-related gene is a substantial issue that frequently necessitates examining a large number of SNPs in an ideal genome. The present study was undertaken to identify potentially damaging nsSNPs using computational methods and to evaluate the damaging nature of mutational variations in the human *STAT4* gene. Using *In-Silico* tools to analyze *STAT4* variations seeks to foster large-scale investigations and the invention of personalized therapies for diseases associated with these variants. Additionally, the study examines the discovery of possible drug targets and their related binding interactions with ligands, to address the future *STAT4* challenges. The future challenges of STAT4 research in HBV include understanding its role in antiviral immunity and liver inflammation, as it regulates cytokine responses critical to HBV clearance. Identifying damaging nsSNPs in STAT4 associated with impaired immune responses could help develop personalized therapies. Additionally, designing specific inhibitors or modulators to balance STAT4 activity without exacerbating liver damage remains a significant therapeutic challenge.

MATERIALS AND METHODS

Data retrieval

The genomic details, including gene names, aliases, HGNC ID, NCBI accession number, ENSEMBL ID, genomic coordinates, gene type, transcript counts, variation counts, UniProtKB ID, reference sequence status, and gene expression was obtained from various databases such as NCBI (https://www.ncbi.nlm.nih.gov/gene), ENSEMBL (http://asia.ensembl.org), HGNC (https://www.genenames.org), and Gene Card (https://www.genecards.org). Figure (1) shows the entire workflow of the present study. The data for the human *STAT4* gene and its sequence in FASTA format were gathered from the NCBI database (Katsaouni *et al.*, 2023; Rozario *et al.*, 2021). The ENSEMBL and NCBI databases were used to extract *STAT4* gene variants. These variations were validated further using the dbSNP database (https://www.ncbi.nlm.nih.gov/snp) (McLaren *et al.*, 2016). The collected SNP results were saved as a standard comma-separated values (CSV) Excel file (Azmi *et al.*, 2023).

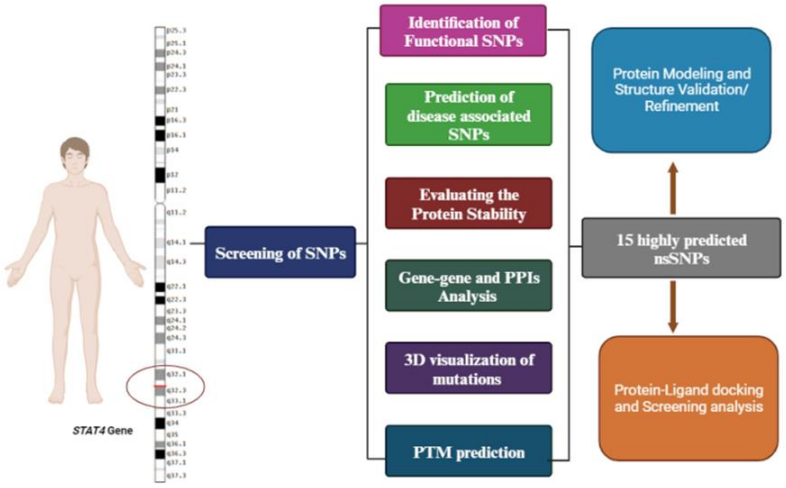


Fig. 1. A visual representation showing different computational tools employed in the current study.

Evaluating deleterious effects of protein-coding SNPs

To explore the functional consequences of damaging SNPs, SIFT>Polyphen>PPh2>Predict SNP>Mutpred2> PANTHER were used (Rozario *et al.*, 2021). These confirmed the precision of our findings, and we categorized such SNPs as high-risk as they were predicted to be detrimental (Hossain *et al.*, 2020). For SNPNEXUS (https://www.snp-nexus.org), an integrated program that include Sorting Intolerant from Tolerant (SIFT) and Polymorphism Phenotyping (PolyPhen) is accessible (Hasnain *et al.*, 2020). The SIFT algorithm assigns a probability score to assess their deleteriousness. SNPs with a probability score below 0.05 are considered detrimental, whereas those with a probability score equal to or greater than 0.05 are tolerated (Honnalli and Adiga, 2023). PolyPhen results are classified as benign, probably and possibly detrimental and a score of 1 has a detrimental effect on the protein (Hasnain *et al.*, 2020; Jahandideh and Zhi, 2014; Mahmud *et al.*, 2016). The common nsSNPs from SIFT and Polyphen were evaluated for the present study. PolyPhen2 (PPh2) is available at https://genetics.bwh.harvard.edu/pph2, and it predicts the consequence of an amino acid change based on certain empirical principles on its sequence (Mahmood *et al.*, 2021).

The input query, which comprised both wild and substituted amino acid sites, was submitted in FASTA format. The substitution is assessed as probably, possibly damaging, or benign, with scores ranging from 0 (benign) to 1 (damaging) and based on position-specific independent score differences, with 1 being the most deleterious (Adzhubei *et al.*, 2013). The PANTHER tool can be found at https://www.pantherdb.org, which examines substitutions using position-specific evolutionary conservation scores computed from the alignment of several proteins (Ahmad *et al.*, 2023). The most predicted mutations were further examined by the MutPred2 web server (http://mutpred.mutdb.org) to sort out deleterious or neutral variations in the protein sequence (Pejaver *et al.*, 2020). The PredictSNP online tool (https://loschmidt.chemi.muni.cz/predictsnp) collects input data from several tools to estimate the impact of an amino acid substitution (Mohkam *et al.*, 2022) and provides a consensus prediction with higher accuracy and precision (Girmay *et al.*, 2022).

Forecasting disease-linked nsSNPs

Using protein functional annotations, SNPs & GO determines whether a genetic variant is associated with disease and the tool is available at https://snps.biofold.org/snps-and-go/snps-and-go.html (Capriotti *et al.*, 2017). To offer more details regarding the impact of mutations on a protein sequence, we used the SuSPect web server (http://www.sbg.bio.ic.ac.uk/suspect) to determine whether a protein residue is prone to deleterious mutations (de Souza Albuquerque *et al.*), and provide results that range from 0 (neutral) to 100 (deleterious) (Ittisoponpisan and Jeerapan, 2021). Meta-SNP are available at http://snps.biofold.org/meta-snp, a web-based tool that forecasts the impact of SNPs using the support vector machine method (Jabuk and Jaralla, 2023). It was optimized to forecast disease-associated point mutations and scored with > 0.5 identifying Deleterious variants (Capriotti *et al.*, 2013).

Analysis of protein stability

I-Stable (http://predictor.nchu.edu.tw/iStable) is an integrated tool for evaluating protein stability variations in a protein sequence. The use of meta-predictions to improve results in either increasing or decreasing protein stabilities (Chen *et al.*, 2013). MU Pro (http://mupro.proteomics.ics.uci.edu) evaluates amino acid variations and generates a DDG value with a score less than 0 showing that the mutation reduces protein stability and a score greater than 0, implying the variation increases protein stability **(Hasnain *et al.*, 2020; Wang *et al.*, 2020b)**. The selected nsSNPs were examined using the DynaMut2 server (https://biosig.lab.uq.edu.au/dynamut2). It displays the change in stability (G) calculated in kcal/mol, and a negative G value implies that the influence is destabilizing, whereas a positive number indicates that the effect is stabilizing (Rodrigues *et al.*, 2021). Cologne University

Protein Stability Analysis Tool (CUPSAT) available at https://cupsat.brenda-enzymes.org, is used to predict stable protein variations resulting from single amino acid mutations (Saini *et al.*, 2018). The CUPSAT server is suitable for a broad spectrum of proteins with PDB or FASTA input formats (Choudhury *et al.*, 2022).

Gene–gene and protein-protein interactions

The GeneMANIA server (https://genemania.org) used a large quantity of functional association data to build a biological network interaction involving the top 20 genes associated with our *STAT4* gene (Mustafa *et al.*, 2020). Using a massive store of biological correlation details may uncover new genes associated with a set of input genes (Baralić *et al.*, 2022; Chetta *et al.*, 2020). The InBio Discover platform, which can be found at https://zs-revelen.com, was used to create a network of high-confidence protein-protein interactions (PPIs). It exploits the inBio-Map, an extensive overview of human protein biology with over 6 million traceable entries. It is based on expected trustworthy networks that take scientific proof, pathways, and other curated data to provide helpful insights into the challenging landscape of cellular protein interface (Li *et al.*, 2017).

Identification of mutation clusters

Mutation3D (http://mutation3d.org) is used for locating clusters of structural changes and offers the analysis of variants in a variety of forms, allowing for uniform access to mutation clusters derived from a massive dataset of over 975,000 somatic variations reported by 6811 tumor genomic research. Users enter a protein of interest with related mutations, and the result is a protein model with its amino acid clusters. It is a useful tool for scholars studying the regional distribution and potential functional impact of mutations in protein structure (Meyer *et al.*, 2016).

Estimating post-translational modifications (PTM)

Understanding PTM data enables more knowledge of the disease association, etiology, and proteins that undergo covalent modifications. These entail the use of functional groups, such as phosphorylation, acetylation, methylation, or ubiquitination all are diverse types of PTMs (Hasan and Khatun, 2018). When amino acid sequences are submitted as input, MusiteDeep (https://www.musite.net) is the first deep-learning algorithm to predict phosphorylation sites (Wang *et al.*, 2020a; Wang *et al.*, 2017). PhosphoSitePlus (PSP), available at https://www.phosphosite.org, provides data for the study of mammalian PTMs, which contributes to a better understanding of the regulatory mechanisms in cellular processes (Ramazi and Zahiri, 2021). To limit the projected outputs of bioinformatic tools, we developed a criterion. Given the enormous number of nsSNPs, only the most probably damaging variants found by all 13 servers were used for further study, including protein modeling and protein-ligand interaction.

Protein structure assessment and modeling

For protein modeling, SWISS-MODEL (https://swissmodel.expasy.org) distinguishes itself as a completely computerized protein structural homology-modeling tool. It employs the UniProtKB database for precise target-template alignment, enabling researchers to build homology models for specific proteins of interest (Rozario *et al.*, 2021). As input queries, the FASTA format was submitted and visualized by the UCSF Chimera tool accessible at https://www.cgl.ucsf.edu/chimera (Verma *et al.*, 2023). Galaxy Refine is used to refine our predicted models and can be found at https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE. This tool improved the accuracy and quality of the structural models developed throughout our inquiry (Seok *et al.*, 2021). SAVES (https://saves.mbi.ucla.edu) has been used to validate the predicted protein 3D structure with a high stereochemical value of the Ramachandran plot, and the ProCheck proves the most favorable zones. The Ramachandran plot generated by the ProCheck program, with over 90% of residue in the most favorable regions, is considered good quality (Colovos and Yeates, 1993; Mahmud *et al.*, 2016). The program TM-align (https://zhanggroup.org/TM-align) was used for comparing the structures of

native and mutant proteins. It calculates the template modeling score (TM-score) and the root-mean-square deviation (RMSD) that ranges from 0 to 1, with 1 indicating perfect similarity between two structurally related components (Carugo and Pongor, 2001; Zhang and Skolnick, 2004; Zhang and Skolnick, 2005).

Protein-Ligand docking analysis

To investigate the consequences of deleterious nsSNPs on STAT4 binding affinity, we used a molecular docking method with the PyRx virtual screening tool available at https://pyrx.sourceforge.io (Ebrahim *et al.*, 2022). We built a suitable target protein from the STAT4 structure coupled with active protein. The active variables were fixed as the grid size of the center (XYZ axis) to dock the ligands, with the 10 highest exclusive computed for each ligand. The PDB files of the ligands and proteins were converted to PDBQT format using the Auto Dock tools (Ferrari and Patrizio, 2021). The molecular structure of the ligand molecules was retrieved from PubChem available at http://pubchem.ncbi.nlm.nih.gov (Singh *et al.*, 2021). The docking result and binding interaction between ligand and receptor proteins were displayed using Discovery Studio Visualizer software (https://discover.3ds.com/discovery-studio-visualizer-download) and PyMOL (https://pymol.org/2) (Shukla *et al.*, 2023).

RESULTS

Screening of nsSNPs

The dataset classifies 4749 SNPs associated with the STAT4 gene based on their genomic location and functional type. Of these, 2500 SNPs are found in the coding regions, with 619 exonic SNPs, including 358 synonymous mutations and 692 non-synonymous mutations, which may have a more significant impact on protein function. The remaining 2250 non-coding SNPs are primarily distributed in intronic regions, 58 SNPs are located in untranslated regions (UTRs), with 56 in the 3' UTR and 2 in the 5' UTR. There are also 122 SNPs in the 5' upstream region and 650 in the 3' downstream region.

Moreover, Table (1) presents the STAT4 protein information. For further in silico analysis, missense or Nonsynonymous SNPs (nsSNPs) were selected and predicted by SIFT and Polyphen. Table (2) provides detailed information on all the nsSNPs, while Figure (2) visually represents the percentage distribution of all SNPs. Subsequently, additional criteria were employed to distinguish between disease-causing harmful SNPs and missense SNPs with unknown significance.

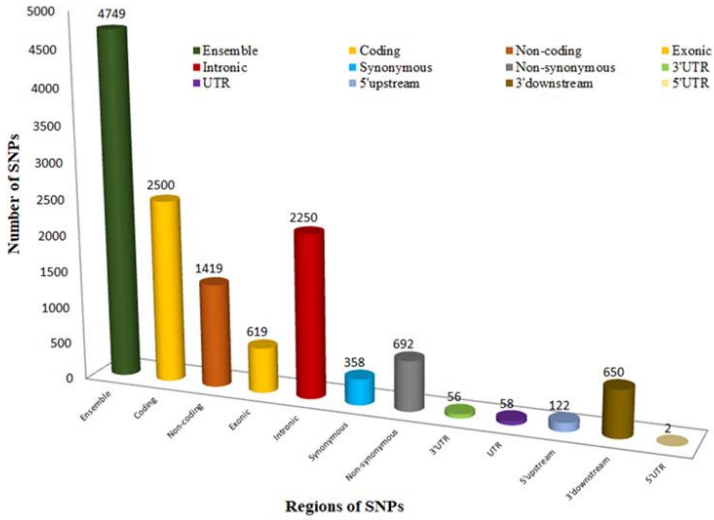


Fig. 2. The distribution of all SNPs in the human *STAT4* gene.

Table 1. *STAT4* proteins retrieve information.

Description	Protein Information
Recommended name	Signal Transducer And Activator Of Transcription 4 (STAT4)
Amino Acids	748
Gene ID	6775
Ensembl ID	ENSG00000138378
Location	2q32.2-q32.3
Exon count	25
NCBI Nucleotide	NC_000002.12
Base Pairs	191029576..191151596
Mass (Da)	85941
Primary accession	Q14765
Organism	Homo Sapiens
FASTA sequence	>sp|Q14765|STAT4_HUMAN Signal transducer and activator of transcription 4 OS=Homo sapiens OX=9606 GN=STAT4 PE=1 SV=1 MSQWNQVQQLEIKFLEQVDQFYDDNFPMEIRHLLAQWIENQDWEAASNNETMATILLQNL LIQLDEQLGRVSKEKNLLLIHNLKRIRKVLQGKFHGNPMHVAVVISNCLREERRILAAAN MPVQGPLEKSLQSSSVSERQRNVEHKVAAIKNSVQMTEQDTKYLEDLQDEFDYRYKTIQT MDQSDKNSAMVNQEVLTLQEMLNSLDFKRKEALSKMTQIIHETDLLMNTMLIEELQDWKR RQQIACIGGPLHNGLDQLQNCFTLLAESLFQLRRQLEKLEEQSTKMTYEGDPIPMQRTHM LERVTFLIYNLFKNSFVVERQPCMPTHPQRPLVLKTLIQFTVKLRLLIKLPELNYQVKVK ASIDKNVSTLSNRRFVLCGTNVKAMSIEESSNGSLSVEFRHLQPKEMKSSAGGKGNEGCH MVTEELHSITFETQICLYGLTIDLETSSLPVVMISNVSQLPNAWASIIWYNVSTNDSQNL VFFNNPPPATLSQLLEVMSWQFSSYVGRGLNSDQLHMLAEKLTVQSSYSDGHLTWAKFCK EHLPGKSFTFWTWLEAILDLIKKHILPLWIDGYVMGFVSKEKERLLLKDKMPGTFLLRFS ESHLGGITFTWVDHSESGEVRFHSVEPYNKGRLSALPFADILRDYKVIMAENIPENPLKY LYPDIPKDKAFGKHYSSQPCEVSRPTERGDKGYVPSVFIPISTIRSDSTEPHSPSDLLPM SPSVYAVLRENLSPTTIETAMKSPYSAE

Identification of the effect of deleterious SNPs

To evaluate the functional consequences of deleterious *STAT4* nsSNPs, a sequential analysis was conducted using SIFT, Polyphen, PPh-2, Predict SNP, Mutpred2, and PANTHER. The SIFT and PolyPhen programs generated a bar graph distribution of SNPs, as shown in Figure (S1), to visually represent the predicted functional impacts of the identified missense mutations. Notably, 28 nsSNPs were identified as posing a high risk of affecting protein function. A detailed dataset for these 28 commonly found nsSNPs is provided in Table (2). In PPh2 analysis, 27 nsSNPs were identified as potentially detrimental. The T177A (rs758709109) variant, among the 28 nsSNPs assessed, was anticipated to be highly possibly damaging, with a 0.995 PSIC score. Predict SNP analysis identified 24 SNPs as highly deleterious, while N479S, K343Q, R241Q, and T177A were deemed to have a neutral effect on the *STAT4* protein, as detailed in Table (2). MutPred2 results indicated that 26 out of the 28 nsSNPs were predicted to be deleterious to the STAT4 protein. The results showed prediction scores ranging from 0.945 to 0.502, with corresponding P values greater than 0.05, which suggests a likely deleterious impact on the protein. Notably, the L307F (0.325) and T177A (0.209) variants exhibited P scores less than 0.5. Utilizing the PANTHER tool, predictions were made regarding the influence of nsSNPs on protein function. For 26 nsSNPs, the estimated scores indicated a probability of adverse effects greater than 0.5. Additionally, two SNPs showed a possibly damaging prediction.

Screening of disease-associated nsSNPs

To identify disease-associated nsSNPs with significant potential to alter the structure or function of the *STAT4* protein, SNP& GO, Meta SNP, and SuSpect tools were employed. Each algorithm utilizes distinct parameters to evaluate nsSNPs as disease-related or neutral. The SNP & GO algorithm highlighted 14 nsSNPs associated with the disease, while the remaining 14 nsSNPs were classified as neutral. A summary of the prediction results is provided in Table (3).

According to Meta-SNP, R70C, T177A, R241Q, and K343Q were predicted to have a neutral effect on the *STAT4* protein, based on a prediction score threshold of <0.5 for neutrality and >0.5 for disease causation. In the case of the SuSpect Server, 25 nsSNPs were identified as highly disease-causing. However, three specific variants, D668G (rs751205891), G618A (rs776987023), and N479S (rs1376354446) were classified as neutral. The comprehensive details of these predictions are provided in Table (3).

Table 2. Different computational tools confirmed detrimental nsSNPs.

		SIFT		Polyphen		Polyphen2		Predict SNP	Mutpred2	Panther	
Variation ID	Mutations	Score	Effect	Score	Effect	Prediction	Score	Effect	Score (0.5)	Effect	Pdel
rs751205891	D668G	0	D	0.997	PD	PD	1	D	0.828	PD	0.85
rs776987023	G618A	0	D	0.998	PD	PD	0.999	D	0.568	PD	0.74
rs771192197	G605R	0	D	1	PD	PD	1	D	0.945	PD	0.89
rs770306554	L597S	0	D	1	PD	PD	1	D	0.88	PD	0.85
rs3024933	R584W	0	D	1	PD	PD	1	D	0.698	PD	0.74
rs774187563	C539Y	0	D	0.947	PD	PD	0.997	D	0.682	PD	0.74
rs1314004125	M517R	0	D	0.928	PD	PD	0.99	D	0.852	PD	0.74
rs376947712	R508H	0	D	0.909	PD	PD	1	D	0.56	PD	0.85
rs780829180	R508C	0	D	0.971	PD	PD	1	D	0.633	PD	0.85
rs1376354446	N479S	0	D	0.982	PD	PD	1	N	0.506	PD	0.74
rs759785386	D476Y	0	D	0.979	PD	PD	0.998	D	0.632	PD	0.5
rs1256812727	N471H	0	D	1	PD	PD	1	D	0.81	PD	0.85
rs544508292	P450A	0	D	0.998	PD	PD	1	D	0.682	PD	0.89
rs1274749529	K343Q	0	D	0.995	PD	PD	1	N	0.759	PD	0.74
rs770753645	T336S	0	D	0.988	PD	PD	0.999	D	0.502	PD	0.85
rs1248978329	P331Q	0	D	0.979	PD	PD	0.996	D	0.739	PD	0.89
rs1424401939	P325L	0	D	0.914	PD	PD	0.998	D	0.666	PD	0.74
rs1399751509	C323Y	0	D	1	PD	PD	1	D	0.81	PD	0.74
rs548245892	L307F	0	D	1	PD	PD	1	D	0.325	PD	0.85
rs764656850	G248R	0	D	1	PD	PD	1	D	0.696	PD	0.85
rs764990697	R241Q	0	D	0.995	PD	PD	1	N	0.711	PD	0.74
rs1280348818	R240W	0	D	0.997	PD	PD	1	D	0.633	PD	0.85
rs758709109	T177A	0	D	0.995	PD	Possibly	0.569	N	0.209	PD	0.74
rs751076320	R70L	0	D	0.97	PD	PD	1	D	0.64	Possibly	0.5
rs1468059700	R70C	0	D	0.995	PD	PD	1	D	0.502	Possibly	0.5
rs761161672	D42G	0	D	0.998	PD	PD	1	D	0.79	PD	0.74
rs1207353579	Q41E	0	D	0.973	PD	PD	0.999	D	0.658	PD	0.74
rs867270496	R31Q	0	D	0.996	PD	PD	1	D	0.844	PD	0.85

*D; Deleterious, N; Neutral, PD; Probably damaging.

Table 3. Screening of possible disease-causing nsSNPs of STAT4 gene.

Variation ID	Mutations	SNPs & GO		Meta SNP		Suspect	
		Effect	RI	Effect	Score	Prediction	Score
rs751205891	D668G	Disease	0	Disease	2	Neutral	16
rs776987023	G618A	Neutral	4	Disease	1	Neutral	15
rs771192197	G605R	Disease	7	Disease	6	Disease	67
rs770306554	L597S	Disease	4	Disease	5	Disease	96
rs3024933	R584W	Disease	1	Disease	1	Disease	56
rs774187563	C539Y	Disease	6	Disease	5	Disease	53
rs1314004125	M517R	Disease	1	Disease	0	Disease	48
rs376947712	R508H	Disease	4	Disease	4	Disease	57
rs780829180	R508C	Disease	6	Disease	6	Disease	78
rs1376354446	N479S	Neutral	1	Disease	3	Neutral	24
rs759785386	D476Y	Disease	2	Disease	4	Disease	39
rs1256812727	N471H	Disease	4	Disease	7	Disease	65
rs544508292	P450A	Disease	2	Disease	4	Disease	82
rs1274749529	K343Q	Neutral	7	Neutral	6	Disease	35
rs770753645	T336S	Neutral	1	Disease	3	Disease	53
rs1248978329	P331Q	Neutral	1	Disease	4	Disease	77
rs1424401939	P325L	Neutral	3	Disease	0	Disease	38
rs1399751509	C323Y	Disease	0	Disease	4	Disease	49
rs548245892	L307F	Neutral	1	Disease	2	Disease	37
rs764656850	G248R	Disease	3	Disease	3	Disease	27
rs764990697	R241Q	Neutral	6	Neutral	7	Disease	22
rs1280348818	R240W	Neutral	1	Disease	2	Disease	46
rs758709109	T177A	Neutral	6	Neutral	6	Disease	28
rs751076320	R70L	Neutral	2	Disease	2	Disease	30
rs1468059700	R70C	Neutral	2	Neutral	1	Disease	46
rs761161672	D42G	Neutral	6	Disease	4	Disease	41
rs1207353579	Q41E	Neutral	6	Disease	1	Disease	51
rs867270496	R31Q	Disease	0	Disease	6	Disease	9

Characterization of Protein Stability Changed by Mutations

The structural impact of 28 potential nsSNPs was assessed using the Mu-Pro servers, and the findings related to protein stability are detailed in Table (4). Notably, MU Pro identified K343Q (rs1274749529) and P325L (rs1424401939) showing an increased impact on protein stability. According to i-stable, 11 nsSNPs were found to enhance protein stability, while 32 nsSNPs were associated with a reduction in protein expression, as indicated in Table (4). The CUPSAT server analysis revealed that 8 mutations had a stabilizing effect, while 20 NSSNPs were destabilizing. Furthermore, the DynaMut2 server was utilized to calculate general dynamic traits of the highest deleterious 23 nsSNPs. Predictions for Δ entropy energy and ΔΔG by ENCoM, comparing the wild-type and mutant STAT4 protein, indicated that four mutants exhibited a stabilizing effect, while 24 were destabilizing. A comprehensive overview of these structural assessments is presented in Table (4).

Table 4. Protein Stability was analyzed by using MUpro, iStable, CUPSAT, and Dynamut2

Variation ID	Mutations	Mu-Pro		i-Stable		CUPSAT		DynaMut2	
		Effect	Score	Stability	Score	Stability	Score	Stability	Score (kcal/mol)
rs751205891	D668G	Decrease	-1.7024511	Decrease	0.7138	Destabilizing	-0.79	Destabilizing	-0.41
rs776987023	G618A	Decrease	-0.70397415	Increase	0.654043	Destabilizing	-2.12	Destabilising	-0.27
rs771192197	G605R	Decrease	-1.0906461	Increase	0.676865	Destabilizing	-0.08	Destablising	-0.83
rs770306554	L597S	Decrease	-2.2844317	Decrease	0.726237	Destabilizing	-2.01	Destabilising	-3.23
rs3024933	R584W	Decrease	-0.6662067	Decrease	0.758422	Stabilizing	1.01	Stabilising	0.12
rs774187563	C539Y	Decrease	-1.2030191	Decrease	0.731439	Stabilizing	2.54	Destabilising	-0.79
rs1314004125	M517R	Decrease	-1.0711499	Increase	0.568745	Destabilizing	-2.75	Destabilising	-0.79
rs376947712	R508H	Decrease	-1.1683329	Increase	0.695355	Destabilizing	-1.99	Destabilising	-0.94
rs780829180	R508C	Decrease	0.90240299	Decrease	0.603023	Destabilizing	-18.5	Destabilising	-0.01
rs1376354446	N479S	Decrease	-0.9309748	Decrease	0.7341	Destabilizing	-0.43	Destabilising	-0.11
rs759785386	D476Y	Decrease	-0.5489519	Increase	0.750551	Destabilizing	-1.39	Stabilising	0.37
rs1256812727	N471H	Decrease	-0.9882099	Decrease	0.677074	Destabilizing	-8.16	Destabilising	-1.12
rs544508292	P450A	Decrease	-1.2378193	Decrease	0.740649	Destabilizing	-4.59	Destabilising	-1.51
rs1274749529	K343Q	Increase	0.3000945	Increase	0.73934	Destabilizing	-9.07	Destabilising	-0.77
rs770753645	T336S	Decrease	-0.395409	Decrease	0.641611	Destabilizing	-3.54	Destabilising	-1.09
rs1248978329	P331Q	Decrease	-0.7923609	Decrease	0.778394	Destabilizing	-7.82	Destabilising	-1.1
rs1424401939	P325L	Increase	0.2452688	Increase	0.634104	Destabilizing	-0.47	Destabilising	-0.47
rs1399751509	C323Y	Decrease	-0.5832999	Decrease	0.699872	Stabilising	2.81	Destabilising	-0.85
rs548245892	L307F	Decrease	-0.83214687	Decrease	0.803803	Destabilizing	-0.35	Destabilising	-1.51
rs764656850	G248R	Decrease	-0.3619475	Decrease	0.685776	Destabilizing	-1.35	Destabilising	-0.87
rs764990697	R241Q	Decrease	-0.8471517	Decrease	0.772587	Stabilising	0.86	Destabilising	-0.7
rs1280348818	R240W	Decrease	-0.5462583	Decrease	0.686031	Stabilising	3.35	Destabilising	-1.14
rs758709109	T177A	Decrease	-1.7647132	Decrease	0.672871	Stabilising	0.35	Destabilising	-0.26
rs751076320	R70L	Decrease	-0.2161614	Decrease	0.578442	Destabilizing	-1.48	Stabilising	0.45
rs1468059700	R70C	Decrease	-0.7118859	Decrease	0.592308	Destabilizing	-1.71	Stabilising	0.2
rs761161672	D42G	Decrease	-1.4731837	Decrease	0.692628	Destabilizing	-0.89	Destabilising	-0.04
rs1207353579	Q41E	Decrease	-1.0127879	Decrease	0.705554	Stabilising	0.95	Destabilising	-0.01
rs867270496	R31Q	Decrease	-1.0198301	Decrease	0.763682	Stabilising	1.2	Destabilising	-1.54

Gene-gene and PPIs Analysis

GeneMANIA was employed to construct the gene-gene interaction network for the STAT4 target gene along with the closest 20 genes, as illustrated in Figure (3A). The detailed interaction data can be found in Supplementary Table (S1). To gain a more comprehensive understanding, a network of Protein-Protein Interactions (PPIs) was also constructed using the inBio-Map resource, as depicted in Figure (3B). This PPI network prediction identified 26 interacting proteins and a total of 116 interactions. This approach helps unravel the complex web of interactions among proteins associated with STAT4, providing insights into its potential functional relationships within cellular pathways.

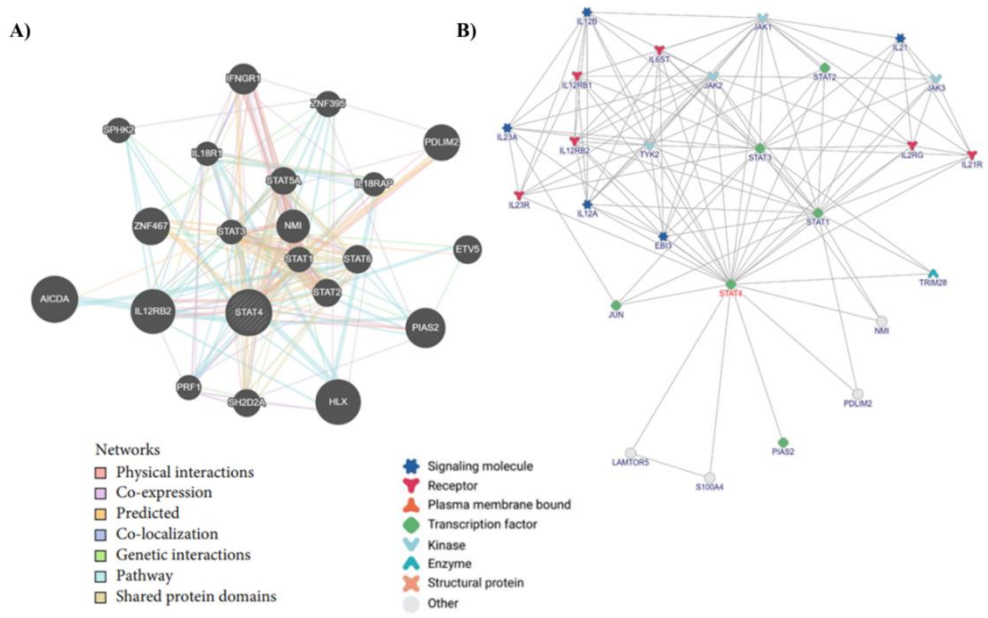


Fig. 3. A) Gene-Gene interaction and **B)** PPI interface network for STAT4 protein identified in Bio-Discover. Network correlations are shown as blue lines.

3D visualization of STAT4 mutation

To visualize the locations of the 28 nsSNPs on the STAT4 protein, the outputs from the Mutation3D server were loaded onto PyMol software. The result gives a 3D representation of the human STAT4 protein (Figure 4), wherein the mutated residues were highlighted in red. The R240W, R241Q, P240A, N471H, N479S, and D476Y nsSNPs formed a clustered mutation (colored red), meanwhile, the remaining 22 SNPs were represented as scattered mutations (colored blue). This visualization provides a clear spatial representation of the distribution of these mutations within the STAT4 protein structure.

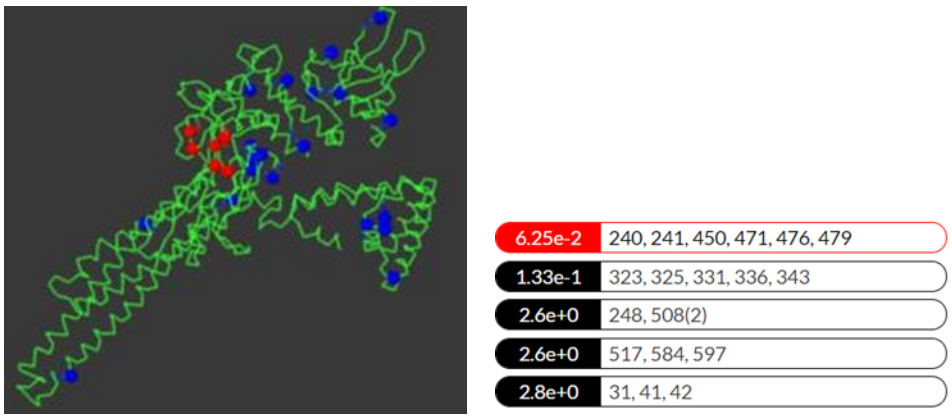


Fig. 4. Mutation3D predicted structural modeling of variant sites in STAT4 protein.

Prediction of post-translational modification

The Musite-deep server was employed to predict post-translational modification sites associated with our candidate SNPs. Protein sequences in FASTA format were submitted as input, and the results revealed that only R508H and R508C were associated with methylarginine. Notably, N-

linked glycosylation at N471H was examined, as outlined in Table (5). Moreover, Phosphosite Plus analysis exposed phosphorylation and ubiquitylation events on the STAT4 gene, as illustrated in (Figure 5). This information provides insights into potential PTM modifications that may influence the functionality of the STAT4 protein. Figure (S2) illustrates the 15 high-risk nsSNPs identified as detrimental to the structure and/or function of the STAT4 protein. This analysis was conducted using 13 in silico tools, which collectively evaluated the potential impacts of these variants.

Table 5. PTM Prediction of STAT4 protein by Musite Deep algorithm.

Mutations	PTM score	Cutoff=0.5
D668G	-	-
G618A	-	-
G605R	-	-
L597S	-	-
R584W	Methylarginine:0.027	None
C539Y	S-palmitoyl_cysteine:0.04	None
M517R	-	-
R508H	Methylarginine:0.817	Methylarginine:0.817
R508C	Methylarginine:0.817	Methylarginine:0.817
N479S	N-linked_glycosylation:0.036	None
D476Y	-	-
N471H	N-linked_glycosylation:0.917	N-linked_glycosylation:0.917
P450A	Hydroxyproline:0.042	None
K343Q	Ubiquitination:0.257;SUMOylation:0.128;N6-acetyllysine:0.187;Methyllysine:0.036;Hydroxylysine:0.018	None
T336S	Phosphothreonine:0.067;O-linked_glycosylation:0.065	None
P331Q	Hydroxyproline:0.038	None
P325L	Hydroxyproline:0.034	None
C323Y	S-palmitoyl_cysteine:0.071	None
L307F	-	-
G248R	-	-
R241Q	Methylarginine:0.046	None
R240W	Methylarginine:0.033	None
T177A	Phosphothreonine:0.152;O-linked_glycosylation:0.065	None
R70L	Methylarginine:0.02	None
R43L	-	-
R43H	-	-
R70L	Methylarginine:0.02	None
R43C	-	-
R70C	Methylarginine:0.02	None
D42G	-	-
Q41E	Pyrrolidone_carboxylic_acid:0.124	None
R4Q	-	-
R31Q	Methylarginine:0.045	None

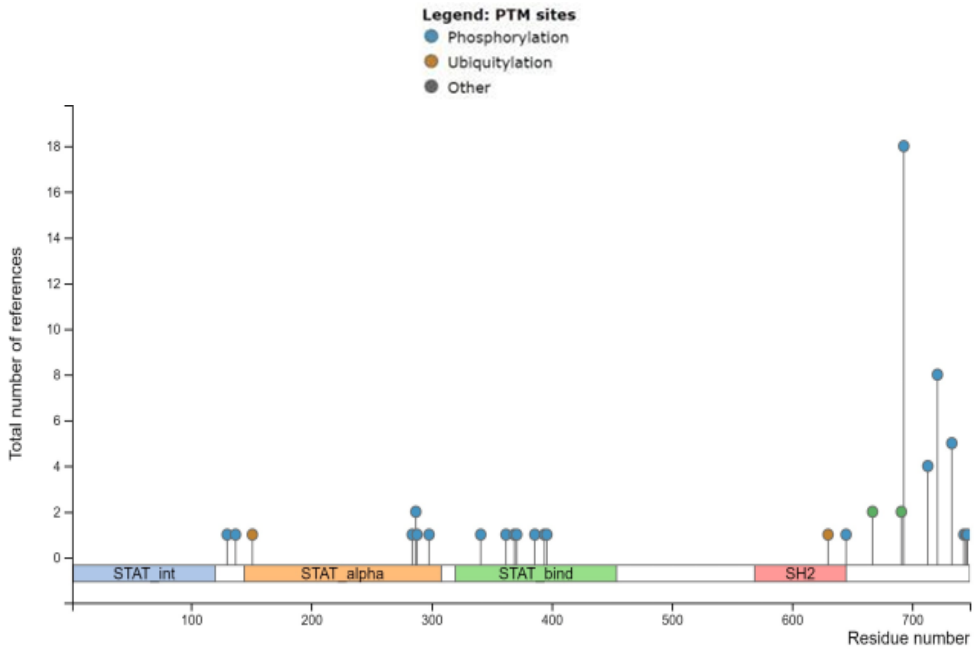
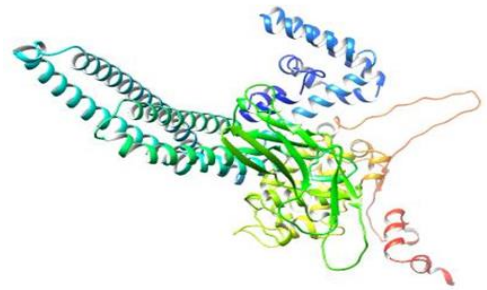


Fig. 5. PhosphoSite Plus anticipates PTM sites on the STAT4 gene.

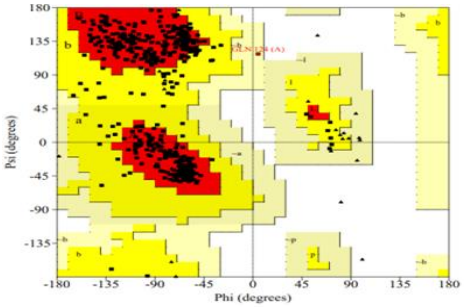
Prediction of target protein modeling

The prediction scores reveal that 15 highly conserved nsSNPs, out of the 28 evaluated on the STAT4 protein, were identified as significant protein conformational modifications. This evaluation was conducted using 13 different in silico tools, which provided a comprehensive analysis of the potential structural and functional impacts of these mutations. For comparative homology modeling, sequences with at least >30% similarity and identity were chosen. We obtained 50 templates, all displaying 100% sequence identity with STML ID (Q14765.1.A), for the query sequence. We modeled the 3D structure of the STAT4 protein using the template ID Q14765.1.A (range: 1-748aa; coverage: 1.00), specifically the Alphafold DB model of STAT4_HUMAN (Organism: Homo sapiens) for the query sequence as presented in Figure (6A). The results, derived from a template with a model quality refined by Galaxy Refine, were further validated using QMEAN with a value of -0.44. The proteins mentioned earlier were downloaded along with their respective PDB files and subjected to mutation using PyMol.

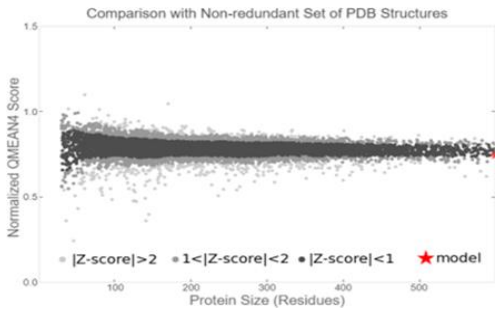
The high RMSD value of 0.33 for mutants D668G, C539Y, R508C, T336S, and P331Q suggests that these mutations lead to substantial structural deviations from the native protein as outlined in Table (6), implying that the mutations might have a considerable impact on the protein conformation. To validate the modeled framework, SAVES was employed, and the Ramachandran plot evaluation was conducted to examine the secondary structure. The resulting structure adhered to all constraints imposed by potential energy calculations. A significant majority of the amino acid residues in the *STAT4* protein (91.50%) were located in a highly favorable region, as illustrated in Figure (6B). The comparison plot indicates the quality of the model in comparison with experimental structures of similar sizes. The x-axis shows the protein length. The y-axis is the normalized QMEAN score. The *STAT4* model is represented as a red star as shown in Figure (6C). For comprehensive details, refer to Table (6) for the complete predicted results.



A.



B.



C.

Residues in most favoured regions [A,B,L]	622	91.5%
Residues in additional allowed regions [a,b,l,p]	57	8.4%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.1%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	680	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	29	
Number of proline residues	37	
Total number of residues	748	

Fig. 6. A) The crystal structure of the human *STAT4* protein **B)** Statistics of Ramachandran plot. The most favored, additional allowed, generously allowed, and disallowed regions are colored in red, yellow, light yellow, and white respectively. **C)** Model validation by QMEAN.

Table 6. The STAT4 gene structural identification and TM Score

Templates	ERRAT	PROCHECK				Verify	TM Align	
	Score	Core	Allow	Generously	Disallowed	Score	TM Score	RMSD
Q14765.1.A	97.4926	91.50%	8.40%	0.10%	0.00%	72.33%		
D668G	97.9885	96.50%	3.40%	0.10%	0.00%	71.66%	0.99881	0.33
G605R	95.8213	96.60%	3.40%	0.00%	0.00%	67.78%	0.99892	0.31
L597S	95.5137	96.30%	3.70%	0.00%	0.00%	69.52%	0.99893	0.31
C539Y	97.4063	96.20%	3.80%	0.00%	0.00%	71.39%	0.9988	0.33
M517R	95.279	95.90%	4.10%	0.00%	0.00%	70.19%	0.99886	0.32
R508H	95.1498	96.60%	3.40%	0.00%	0.00%	72.06%	0.99883	0.32
R508C	95.8333	96.60%	3.40%	0.00%	0.00%	70.32%	0.9988	0.33
N471H	97.971	95.90%	4.10%	0.00%	0.00%	70.45%	0.99882	0.32
P450A	96.4235	95.70%	4.30%	0.00%	0.00%	71.39%	0.99895	0.31
T336S	95.9302	96.30%	3.70%	0.00%	0.00%	73.40%	0.99877	0.33
P331Q	95.8153	95.70%	4.30%	0.00%	0.00%	70.32%	0.9988	0.33
C323Y	97.2779	96.80%	3.20%	0.00%	0.00%	71.66%	0.99886	0.32
G248R	96.4183	95.30%	4.70%	0.00%	0.00%	71.66%	0.99894	0.31
D42G	97.3761	96.30%	3.70%	0.00%	0.00%	70.05%	0.99885	0.32
R31Q	95.8092	95.90%	4.10%	0.00%	0.00%	71.39%	0.9989	0.31

Molecular docking by PyRx

To investigate ligand-protein interactions, molecular docking was performed using the PyRx tool. All 19 selected ligands from the PubChem database were docked with STAT4, resulting in ten distinct conformations for each ligand, characterized by their binding affinity (-Kcal/mol) as depicted in Figure (7). The docking results revealed that these binding affinities are indicative of the compounds' activity levels, and detailed affinities for all compounds can be found in Supplementary Table (S2).

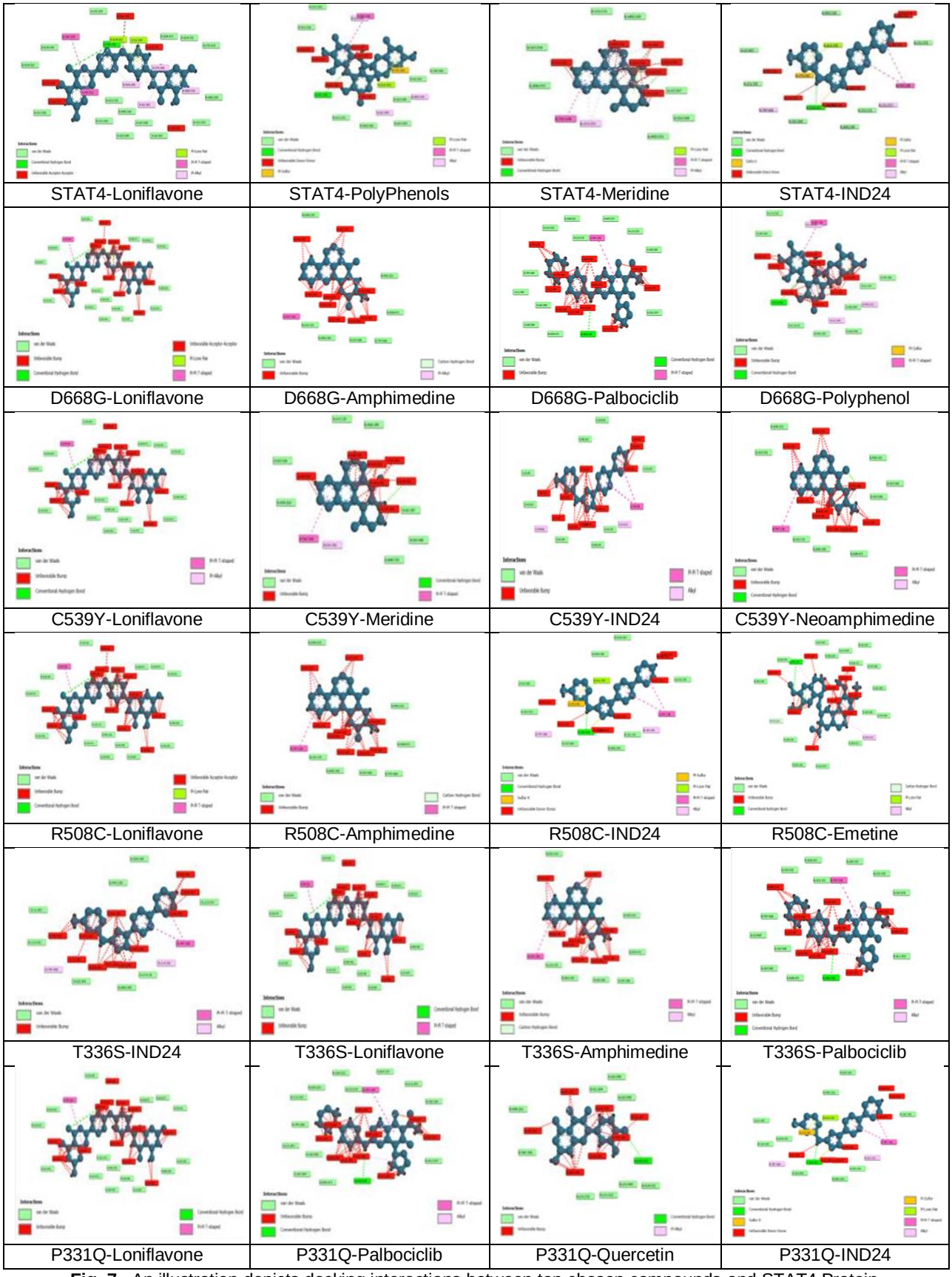


Fig. 7. An illustration depicts docking interactions between top chosen compounds and STAT4 Protein.

The top compounds with strong binding affinities, including Amphimedine, Emetine, IND24, Loniflavone, Meridine, Neoamphimedine, Palbociclib, Polyphenol, and Quercetin, were selected for further investigation. These compounds were then docked with both native and mutated protein complexes. The interactions were explored using Discovery Studio, which provides a 2D representation of all docking interactions. All of the chosen ligands had binding free energies greater than -3 Kcal/mol. The greatest binding energy revealed that the STAT4 protein successfully docked with Loniflavone is -10.5 Kcal/mol. A Loniflavone ligand was fixed in the STAT4 binding pocket sites by forming the conventional hydrogen bond with residues LYS239, GLN140, GLN257, LEU255, LEU131, LEU251, PRO250, GLY248, GLY249, ILE467, LEU332, ARG330, CYS323, GLN321, ASN471; and hydrophobic interactions (Alkyl, pi-alkyl, cation, anion, halogens) with TRP238, GLY254, ASN253, HIS252, ARG240, GLN242, ARG241, ILE244, GLN243, CYS246, ALA245, ILE247, TPR464, PRO331 residues. Figure (7) depicts the interacting residues obtained during docking. Table (7) shows the differences in mutant and wild ligand-protein residue interactions that indicate changed functional properties caused by mutations.

Table 7. Molecular docking interaction between molecules of selected protein-ligand complexes.

Protein	Ligand	Docking Score (-Kcal/mol)	Hydrogen Bond Interaction	Hydrophobic Interactions
Wild	Loniflavone	-10.5	LYS239, GLN140, GLN257, LEU255, LEU131, LEU251, PRO250, GLY248, GLY249, ILE467, LEU332, ARG330, CYS323, GLN321, ASN471	TRP238, GLY254, ASN253, HIS252, ARG240, GLN242, ARG241, ILE244, GLN243, CYS246, ALA245, ILE247, TPR464, PRO331
	PolyPhenols	-10.1	LEU255, GLY254, LYS239, LEU131, PRO250, GLY248, GLY249, ILE247, TRP464	TRP38, LEU251, ASN253, HIS252, ARG240, ARG241, GLN242, GLN243, ILE244, ALA245, CYS246, PRO331
	Meridine	-9.6	LEU131, ARG240, GLY254, ASN253, PRO331, GLY248, ILE247,	HIS252, TRP238, LEU251, ARG241, GLN242, GLN243, ILE244, ALA245, CYS246
	IND24	-9.6	PRO250, LEU332, GLY248, GLN243, ARG240, LEU131, LEU255, ILE467	GLY254, ASN253, HIS252, TRP238, LEU251, ARG241, ILE247, PRO331, ALA245, CYS246, TRP464
D668G	Loniflavone	-10.6	LYS239, GLN140, GLN257, LEU255, LEU131, LEU251, PRO250, GLY248, GLY249, ILE467, LEU332, ARG330, CYS323, GLN321, ASN471	TRP238, GLY254, ASN253, HIS252, ARG240, GLN242, ARG241, ILE244, GLN243, CYS246, ALA245, ILE247, TPR464, PRO331
	Amphimedine	-10.1	ASN253, LEU131, ARG240, ASN471, GLY248, TRP464, PRO331	LEU251, HIS252, ARG241, ALA245, GLN242, GLN243. ILE244, ILE247, CYS246
	Palbociclib	-9.1	GLN321,LEU332, LEU131, TRP464, ILE467, GLY248, GLY249, ASN471, ARG240, ASP237, LEU255, GLY254, GLU319	PRO331, TRP238, CYS246, ILE247, GLN243, ILE244, ALA245, ARG241, GLN242, LEU251, HIS252, ASN253
	Polyphenol	-9	LEU255, GLY254, LYS239, LEU131, PRO250, GLY248, GLY249, ILE247, TRP464	TRP238, LEU251, ASN253, HIS252, ARG240, ARG241, ALA245, GLN242, GLN243, CYS246, ILE244, PRO331
C539Y	Loniflavone	-10.6	LYS239, GLN140, GLN257, LEU255, LEU131, LEU251, PRO250, GLY248, GLY249, ILE467, LEU332, ARG330, CYS323, GLN321, ASN471	TRP238, GLY254, ASN253, HIS252, ARG240, GLN242, ARG241, ILE244, GLN243, CYS246, ALA245, ILE247, TPR464, PRO331
	Meridine	-9	LEU131, ARG240, GLY254, ASN253, PRO331, GLY248, ILE247,	HIS252, TRP238, LEU251, ARG241, GLN242, GLN243, ILE244, ALA245, CYS246

	IND24	-8.9	GLN140, PRO250, ILE467, LEU332, GLY248, ARG240, LEU131, LEU255	PRO331, TRP464, CYS246, ILE247, GLN243, ARG241, LEU251, TRP238, ALA245, HIS252, ASN253, GLY254
	Neoamphimedine	-8.9	ASN253, GLY254, LEU131, ARG240, ASN471, GLY248, GLY249, PRO331	LEU251, HIS252, ARG241, ALA245, GLN242, GLN243. ILE244, ILE247, CYC246
R508C	Loniflavone	-11	GLN140, GLN257, LEU255, LYS239, LEU131, LEU251, PRO250, GLY248, GLY249, ILE467, ASN471, GLN321, ARG330, LEU332, CYS323	TRP238, GLY254, ASN253, HIS257, ARG240, ARG241, GLN242, GLN243, ILE244, CYS246, ALA245, ILE247, TRP464, PRO331
	Amphimedine	-8.6	LEU131, ARG240, ASN471, PRO331, GLY248, TRP464	TRP238, HIS252, ARG241, GLN242, ALA245, ILE244, LEU251, GLN243, ILE247, CYS246,
	IND24	-8.5	ILE267, LEU332, GLY248, GLN243, PRO250, GLN140, ARG240, LEU131, LEU255	PRO331, ILE247, TRP464, CYS246, ALA245, ARG241, LEU251, TRP238, HIS252, ASN253, GLY254,
	Emetine	-8.4	GLN140, PRO250, GLN132, ARG320, GLY254, ARG241, GLU319, LYS239, ARG240, LEU449, GLN321, ASN471, VAL497, TRP500, GLY349, GLY248, TRP464	ASN253, HIS252, LEU251, TRP238, ILE244, GLN243, GLN242, ALA245, CYS246, PRO331
T336S	IND24	-9.5	ILE2467, LEU332, GLY248, PRO250, GLN140, ARG240, LEU131, LEU255	PRO331, ILE247, GLN243, CYS246, ILE244, GLN242, ARG241, LEU251, TRP238, GLY254, ASN253, HIS252
	Loniflavone	-9.2	GLN240, GLN257, LEU255, LYS239, LEU251, LEU131, PRO250, GLY249, GLY248, ILE467, ASN471, GLN321, CYS323, ARG330, LEU332	TRP238, GLY254, ASN253, HIS257, ARG240, ARG241, GLN242, GLN243, ILE244, CYS246, ALA245, ILE247, TRP464, PRO331
	Amphimedine	-9	ASN253, LEU131, ARG240, ASN471, GLY248, TRP464, PRO331	TRP38, LEU251, HIS252, ARG241, GLN242, ILE244, ALA245, CYS246, ILE247, GLN243,
	Palbociclib	-9	LEU332, TRP464, ILE467, GLY248, GLY249, ASN471, GLN321, LEU131, ARG240, ASP237, LEU255, GLY254, GLU319	TRP238, PRO331, CYS246, ILE247, GLN243, ILE244, ALA245, ARG241, GLN242, LEU251, HIS252, ASN253
P331Q	Loniflavone	-8.9	GLN140, GLN257, LEU255, LYS239, LEU131, LEU251, PRO250, GLY248, GLY249, ILE467, ASN471, GLN321, ARG330, LEU332, CYS323	TRP238, GLY254, ASN253, HIS252, ARG240, GLN242, ARG241, ILE244, GLN243, CYS246, ALA245, ILE247, TPR464, GLN331
	Palbociclib	-8.5	GLN331, LEU332, TRP464, ILE467, GLY248, GLY249, ASN471, GLN321, LEU131, ARG240, ASP237, LEU255, GLY254, GLU319	CYS246, ILE247, GLN243, ILE244, ALA245, GLN242, ARG241, LEU251, TRP238, HIS252, ASN253,
	Quercetin	-8.4	ASN253, TRP238, ILE244, LEU131, LEU332, LEU449, GLY248, GLY249, GLN243, GLN321,	HIS252, LEU251, ARG241, GLN242, ALA245, CYS246, ILE247,
	IND24	-8.3	ILE467, LEU332, GLN331, GLY248, GLN243, ARG240, LEU131, PRO250, GLN140, LEU255	TRP464, ILE247, CYS346, ILE244, GLN242, ARG241, LEU251, TRP238, HIS252, GLY254, ASN253, ALA245

DISCUSSION

Generally, SNPs represent genetic variations caused by a single nucleotide alteration in a genome sequence. While some SNPs are multiallelic, the vast majority are bi-allelic, with two different bases present at a single DNA location. Such variants must have a minimum probability level in the population to be classified as SNPs, which is often greater than 1% (Vallejos-Vidal *et al.*, 2020). Considering that nsSNPs are identified by a single amino acid mutation inside the coding regions, resulting in genetic disorders. Recent research has focused substantially on nsSNPs, and large amounts of data are available in public repositories like as

the dbSNP NCBI, Ensembl, and GnomAD databases (Subbiah *et al.*, 2023). Polymorphisms are recognized as potentially valuable biomarkers for disease detection or prognostic due to their high frequency, accessibility, low genotyping costs, and capacity to conduct association research utilizing analytical and computational methods (Srinivasan *et al.*, 2016). The latest study has associated *STAT4* mutations with the transmission of HBV and HCC malignancy (Shi *et al.*, 2019). Hepatitis comprises a major mammalian viral infection that causes a variety of serious and persistent complications of liver function (Li *et al.*, 2019). Multiple studies have examined the link between STAT3 (rs1053004, rs2293152) and STAT4 (rs7574865) variants with the possibility of HBV and chronic HCC in different ethnic groups; however, the outcomes remain unclear and inconsistent (Liao *et al.*, 2014; Shi *et al.*, 2019). The function of a protein is closely linked to its tertiary framework, and any modification in the order of amino acids may impact its structure, leading to diseases.

The purpose of this study was to use bioinformatics tools to discover potentially detrimental nsSNPs in the *STAT4* gene, as well as to investigate the negative nature of *STAT4* mutational changes. Using in-silico tools to assess genetic variants in *STAT4* will allow for larger-scale research and the discovery of targeted treatments for associated diseases. It is preferable to use multiple tools and get an agreement by comparing results from various sources. Additionally, bioinformatics innovations should be validated in the laboratory using a variety of in-vitro and in-vivo techniques. Using the SNPNexus tool, 4749 SNPs in the human STAT4 gene were predicted in the present study. The SIFT and Polyphen both chose 28 nsSNPs for additional *In-Silico* research. Table (1) has complete information about STAT4 protein. As indicated in Table (2), a stepwise analysis was performed to determine the functional consequences of damaging SNPs in the STAT4 gene using SIFT > Polyphen >PPh 2 >Predict SNP>Mutpred2, and PANTHER. Three servers, SNP& GO, Meta SNP, and SuSpect, were applied to find disease-associated nsSNPs with a strong potential influence on the STAT4 protein. The K343Q (rs1274749529) and P325L (rs1424401939) mutations have a higher impact on protein stability, according to MU Pro. Out of 28 nsSNPs, i-stable found 11 nsSNPs to increase protein stability while 32 nsSNPs were shown to decrease protein production. The CUPSAT server study found that 8 mutations were stabilizing, whereas the DynaMut server calculated that 24 were destabilizing. Studies have suggested that certain SNPs may affect the activity of the STAT4 protein and its ability to regulate immune responses, thus contributing to disease susceptibility.

The mutations N471H, D476Y, and N479S in the STAT4 gene have been shown to potentially influence immune function and the regulation of T-cell responses, making individuals more susceptible to Multiple Sclerosis (MS) and Crohn's disease by altering the function of the STAT4 protein, disrupting cytokine signaling pathways and affecting immune cell recruitment, which plays a key role in the development of these conditions. Additionally, the R240W and R241Q mutations in STAT4 have been investigated for their association with diseases such as Systemic Lupus Erythematosus (SLE) and Rheumatoid Arthritis (RA), suggesting that variations in STAT4 can have broad implications for immune-mediated disorders (Beltrán Ramírez *et al.*, 2016; Bravo-Villagra *et al.*, 2024; Nageeb *et al.*, 2018).

The structural assessments suggest that the mutations identified, particularly those linked to methylation and glycosylation modifications, could significantly influence the STAT4 protein's function and stability. The *STAT4* target gene association network was built using GeneMANIA and the InBio Discover platform. The SNPs R240W, R241Q, P240A, N471H, N479S, and D476Y produced a clustered mutation (colored red) that typically correlates with tumors in humans, as represented by Mutation 3D. The PTM analysis showed that mutations such as R508H and R508C, which are associated with methylarginine, could potentially disrupt the protein regulatory functions, while the N-linked glycosylation at N471H may affect its stability and interactions, contributing to pathogenicity.

Using 13 computational tools, we found that 15 nsSNPs are highly detrimental to the *STAT4* protein. The quality and validity of the STAT4 3D model confirmed through SWISS-MODEL and SAVES, offer confidence to the structural predictions, highlighting the potential impacts of these mutations. The high RMSD values for mutants like D668G, C539Y, R508C, T336S, and P331Q indicate substantial structural deviations, which may disrupt normal protein function. The docking analysis further revealed that these mutations might alter STAT4 binding affinity with its ligands, suggesting functional consequences. Moreover, docking compounds with both natural and mutant STAT4 proteins revealed potential therapeutic candidates, with interactions depicted in 2D models offering promising avenues for drug development targeting the mutant STAT4 form.

This study provides a comprehensive In Silico evaluation of functional and structural nsSNPs in the STAT4 protein, contributing new insights to the understanding of these mutations. Further research should be conducted to validate the conclusions of this study and explore the potential clinical implications of STAT4 mutations in different populations. Furthermore, functional and structural studies are needed to explain the potential mechanisms behind the link between nsSNPs and HBV disease susceptibility. The use of bioinformatic approaches for assessing *STAT4* genetic variations will aid in the planning of large-scale investigations and the development of targeted therapeutics for diseases caused by these variations. The fact that individual network knowledge remains elusive could be attributed to genetic, technical, and computational problems that biological networks of a specific disease continue to encounter. The findings of the study could be useful in the study of prospective therapies and diagnostic methods that require both mutational confirmation and substantial experimental studies. Further experiments could include functional assays such as luciferase reporter assays to assess the impact of selected nsSNPs on STAT4 gene expression and activity. Additionally, in vitro studies using cell lines could help evaluate the effect of these mutations on immune cell signaling, cytokine production, and the response to HBV infection. Moreover, in vivo studies using animal models of HBV infection could be conducted to observe how these mutations influence disease progression. Finally, genotype-phenotype association studies in diverse human populations could strengthen the findings and clarify the role of these nsSNPs in HBV susceptibility.

CONCLUSION

A *STAT* is a transcription variable that stimulates the transcription of genes in response to a variety of distinct cytokines. The purpose of the study was to discover potentially damaging nsSNPs in the *STAT4* gene using bioinformatics tools to investigate the destructive nature of the mutational alterations in the *STAT4* gene. According to the trajectory analysis and stepwise prediction of Deleteriousity of nsSNPs (SNPNexus > Predict SNP> PPh2> PANTHER> SNP & GO> Dynamut2> Meta SNP> iStable> SNP & GO), 13 nsSNPs with a mutational influence on the STAT4 function and structure were shown to be extremely detrimental. Using the SWISS Model, we built a 3D model of the *STAT4* and refined it with ERRAT and PROCHECK programs, and the findings demonstrate that it is reliable. The *STAT4* model is also used for docking with ligand compounds. The D668G, C539Y, R508C, T336S, and P331Q variants were docked with chosen drugs with significant binding affinities such as Amphimedine, Emetine, IND24, Loniflavone, Meridine, Neoamphimedine, Palbociclib, Polyphenol, and Quercetin and docked with native and mutant structures and visualized by Discovery studio. Based on the findings of this study, future genome association studies will be able to identify perilous nsSNPs linked to specific HBV patients. The results imply that utilizing computational approaches to determine target nsSNPs can be a viable alternative option. The previously mentioned non-reported nsSNPs can be confidently regarded as important contributors to HBV-related diseases. Wet lab tests are required to determine the specific effects of these nsSNPs on the protein's structure and function. Finding phenotypic or disease-related variation through

genomic research is a challenging scenario that necessitates innovative approaches.

CONFLICT OF INTEREST

Authors hereby declare that they have no conflict of interest.

REFERENCES

Adzhubei, I., Jordan, D.M., Sunyaev, S.R., 2013. Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet., 76(1): 7.20. 1-7.20. 41.

Ahmad, S.U., Ali, Y., Jan, Z., Rasheed, S., Nazir, N.U.A., Khan, A., Rukh Abbas, S., Wadood, A., Rehman, A.U., 2023. Computational screening and analysis of deleterious nsSNPs in human p 14ARF (CDKN2A gene) protein using molecular dynamic simulation approach. J. Biomol. Struct. Dyn., 41(9): 3964-3975.

Arshad, M., Bhatti, A., John, P., 2018. Identification and in silico analysis of functional SNPs of human TAGAP protein: A comprehensive study. PloS one, 13(1): e0188143.

Azmi, M.B., Khan, W., Azim, M.K., Nisar, M.I., Jehan, F., 2023. Identification of potential therapeutic intervening targets by in-silico analysis of nsSNPs in preterm birth-related genes. Plos one, 18(3): e0280305.

Baralić, K., Živančević, K., Božić, D., Jennen, D., Buha-Đorđević, A., Antonijević-Miljaković, E., Đukić-Ćosić, D., 2022. Potential genomic biomarkers of obesity and its comorbidities for phthalates and bisphenol A mixture: In silico toxicogenomic approach. Biocell., 46(2): 519-533.

Beltrán Ramírez, O., Mendoza Rincón, J.F., Barbosa Cobos, R.E., Alemán Ávila, I., Ramírez Bello, J., 2016. STAT4 confers risk for rheumatoid arthritis and systemic lupus erythematosus in Mexican patients. Immunol. Lett., 175: 40-3.

Bravo-Villagra, K.M., Muñoz-Valle, J.F., Baños-Hernández, C.J., Cerpa-Cruz, S., Navarro-Zarza, J.E., Parra-Rojas, I., Aguilar-Velázquez, J.A., García-Arellano, S., López-Quintero, A., 2024. STAT4 Gene Variant rs7574865 Is Associated with Rheumatoid Arthritis Activity and Anti-CCP Levels in the Western but Not in the Southern Population of Mexico. Genes., 15(2): 241.

Capriotti, E., Altman, R.B., Bromberg, Y., 2013. Collective judgment predicts disease-associated single nucleotide variants. BMC Genom., 14(3): 1-9.

Capriotti, E., Martelli, P.L., Fariselli, P., Casadio, R., 2017. Blind prediction of deleterious amino acid variations with SNPs&GO. Hum. Mut., 38(9): 1064-1071.

Carugo, O., Pongor, S., 2001. A normalized root-mean-spuare distance for comparing protein three-dimensional structures. Protein Sci., 10(7): 1470-1473.

Chen, C.-W., Lin, J., Chu, Y.-W., 2013. iStable: off-the-shelf predictor integration for predicting protein stability changes, BMC Bioinform., pp. 1-14.

Chetta, M., Di Pietro, L., Bukvic, N., Lattanzi, W., 2020. Rising roles of small noncoding RNAs in cotranscriptional regulation: in silico study of miRNA and piRNA regulatory network in humans. Genes., 11(5): 482.

Choudhury, A., Mohammad, T., Anjum, F., Shafie, A., Singh, I.K., Abdullaev, B., Pasupuleti, V.R., Adnan, M., Yadav, D.K., Hassan, M.I., 2022. Comparative analysis of web-based programs for single amino acid substitutions in proteins. PloS one, 17(5): e0267084.

Claussnitzer, M., Cho, J.H., Collins, R., Cox, N.J., Dermitzakis, E.T., Hurles, M.E., Kathiresan, S., Kenny, E.E., Lindgren, C.M., MacArthur, D.G., North, K.N., Plon, S.E., Rehm, H.L., Risch, N., Rotimi, C.N., Shendure, J., Soranzo, N., McCarthy, M.I., 2020. A brief history of human disease genetics. Nat., 577(7789): 179-189.

Collins, F.S., Brooks, L.D., Chakravarti, A., 1998. A DNA polymorphism discovery resource for research on human genetic variation. Genome Res., 8(12): 1229-1231.

Colovos, C., Yeates, T.O., 1993. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci., 2(9): 1511-1519.

de Souza Albuquerque, I., da Silva, I.F., Pondé, L.S., de Alencar, A., A preliminary Analysis of the Impact of Missense Variants of Unknown Significance in the APOA1 Gene using prediction softwares.

Dou, J., Tan, Y., Kock, K.H., Wang, J., Cheng, X., Tan, L.M., Han, K.Y., Hon, C.-C., Park, W.-Y., Shin, J.W., Jin, H., Wang, Y., Chen, H., Ding, L., Prabhakar, S., Navin, N., Chen, R., Chen, K., 2024. Single-nucleotide variant calling in single-cell sequencing data with Monopogen. Nat. Biotechnol., 42(5): 803-812.

Ebrahim, E., Teklu, T., Tajebe, F., Wondmagegn, T., Akelew, Y., Fiseha, M., 2022. Association of cytotoxic T-lymphocyte antigen-4 gene polymorphism with type 1 diabetes mellitus: In silico analysis of biological features of CTLA-4 protein on Ethiopian population. Diabetes, Metab. Syndr. Obes.: Targets Ther., 2733-2751.

Emadi, E., Akhoundi, F., Kalantar, S.M., Emadi-Baygi, M., 2020. Predicting the most deleterious missense nsSNPs of the protein isoforms of the human HLA-G gene and in silico evaluation of their structural and functional consequences. BMC Genet., 21(1): 94.

Ferrari, I.V., Patrizio, P., 2021. Development and Validation Molecular Docking Analysis of Human serum albumin (HSA). bioRxiv.

Girmay, S., Ahmad, H.I., Zahra, Q.A., 2022. A simulation analysis and screening of deleterious nonsynonymous single nucleotide polymorphisms (nsSNPs) in sheep LEP gene. BioMed Res. Int., 2022.

Hasan, M.M., Khatun, M.S., 2018. Prediction of protein post-translational modification sites: an overview. Ann Proteom Bioinform., 2: 049-57.

Hasnain, M.J.U., Shoaib, M., Qadri, S., Afzal, B., Anwar, T., Abbas, S.H., Sarwar, A., Talha Malik, H.M., Tariq Pervez, M., 2020. Computational analysis of functional single nucleotide polymorphisms associated with SLC26A4 gene. Plos one, 15(1): e0225368.

Honnalli, N., Adiga, U., 2023. Evaluation of Gene Polymorphisms In Patients With Urinary Oxalate Stones: Cross-sectional Study. Kor. J. Physiol. Pharmacol., 27(4): 114-125.

Hossain, M.S., Roy, A.S., Islam, M.S., 2020. In silico analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions. Sci. Rep., 10(1): 14542.

Ittisoponpisan, S., Jeerapan, I., 2021. In silico analysis of glucose oxidase from Aspergillus niger: Potential cysteine mutation sites for enhancing protein stability. Bioeng., 8(11): 188.

Jabuk, S.I., Jaralla, E.M., 2023. In silico Predictions of Thr136Arg Missense Variant Shows a Remarkable Negative Impact on the Biological Activity of Enterotoxin Type A. Al-Anbar Med. J., 19(1): 54-61.

Jahandideh, S., Zhi, D., 2014. Systematic investigation of predicted effect of nonsynonymous SNPs in human prion protein gene: a molecular modeling and molecular dynamics study. J. Biomol. Struct. Dyn., 32(2): 289-300.

Jiang, D.K., Wu, X., Qian, J., Ma, X.P., Yang, J., Li, Z., Wang, R., Sun, L., Liu, F., Zhang, P., 2016. Genetic variation in STAT4 predicts response to interferon-α therapy for hepatitis B e antigen-positive chronic hepatitis B. Hepatol., 63(4): 1102-1111.

Jiang, Y., Han, Q., Zhao, H., Zhang, J., 2021. The Mechanisms of HBV-Induced Hepatocellular Carcinoma. J. Hepatocell. Carcinoma., 8: 435-450.

Katsaouni, N., Llavona, P., Khodamoradi, Y., Otto, A.-K., Körber, S., Geisen, C., Seidl, C., Vehreschild, M.J., Ciesek, S., Ackermann, J., 2023. Dataset of single nucleotide polymorphisms of immune-associated genes in patients with SARS-CoV-2 infection. Plos one, 18(11): e0287725.

Krawczak, M., Ball, E.V., Fenton, I., Stenson, P.D., Abeysinghe, S., Thomas, N., Cooper, D.N., 2000. Human gene mutation database—a biomedical

information and research resource. Hum. Mutat., 15(1): 45-51.

Kucukkal, T.G., Petukh, M., Li, L., Alexov, E., 2015. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. Curr. Opin. Struct. Biol., 32: 18-24.

Kumar, S., Ashraf, M., Pannu, U., Mehta, S.C., 2022. Single Nucleotide Polymorphism in Interleukin-6 and Interleukin-8 Genes of Equines. J. Equine Vet. Sci., 117: 104058.

Lander, E.S., 1996. The new genomics: global views of biology. Sci., 274(5287): 536-539.

Li, M., Zhao, Y., Chen, X., Fu, X., Li, W., Liu, H., Dong, Y., Liu, C., Zhang, X., Shen, L., 2019. Contribution of sex-based immunological differences to the enhanced immune response in female mice following vaccination with hepatitis B vaccine. Mol. Med. Rep., 20(1): 103-110.

Li, T., Wernersson, R., Hansen, R.B., Horn, H., Mercer, J., Slodkowicz, G., Workman, C.T., Rigina, O., Rapacki, K., Stærfeldt, H.H., 2017. A scored human protein–protein interaction network to catalyze genomic interpretation. Nat. Methods., 14(1): 61-64.

Liao, Y., Cai, B., Li, Y., Chen, J., Tao, C., Huang, H., Wang, L., 2014. Association of HLA-DP/DQ and STAT4 polymorphisms with HBV infection outcomes and a mini meta-analysis. PLoS One, 9(11): e111677.

Lu, Y., Zhu, Y., Peng, J., Wang, X., Wang, F., Sun, Z., 2015. STAT4 genetic polymorphisms association with spontaneous clearance of hepatitis B virus infection. Immunol. Res., 62: 146-152.

Mahmood, M.S., Irshad, S., Kalsoom, U., Batool, H., Batool, S., Butt, T.A., 2021. In silico analysis of missense Single Nucleotide Variants (SNVs) in HBB gene associated with the β-thalassemia. Gene Rep., 23: 101019.

Mahmud, Z., Malik, S.U.F., Ahmed, J., Azad, A.K., 2016. Computational analysis of damaging single-nucleotide polymorphisms and their structural and functional impact on the insulin receptor. BioMed Res. Int., 2016.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., Cunningham, F., 2016. The ensembl variant effect predictor. Genome Biol., 17(1): 1-14.

Meyer, M.J., Lapcevic, R., Romero, A.E., Yoon, M., Das, J., Beltrán, J.F., Mort, M., Stenson, P.D., Cooper, D.N., Paccanaro, A., 2016. mutation3D: cancer gene prediction through atomic clustering of coding variants in the structural proteome. Hum. Mutat., 37(5): 447-456.

Mohkam, M., Golkar, N., Nabavizadeh, S.H., Esmaeilzadeh, H., Berenjian, A., Ghahramani, Z., Gholami, A., Alyasin, S., 2022. In Silico Evaluation of Nonsynonymous SNPs in Human ADAM33: The Most Common Form of Genetic Association to Asthma Susceptibility. Comput. Math. Methods Med., 2022.

Mustafa, M.I., Murshed, N.S., Abdelmoneim, A.H., Abdelmageed, M.I., Elfadol, N.M., Makhawi, A.M., 2020. Extensive in silico analysis of ATL1 gene: discovered five mutations that may cause hereditary spastic paraplegia type 3A. Scientifica., 2020.

Nageeb, R.S., Omran, A.A., Nageeb, G.S., Yousef, M.A., Mohammad, Y.A.A., Fawzy, A., 2018. STAT4 gene polymorphism in two major autoimmune diseases (multiple sclerosis and juvenile onset systemic lupus erythematosus) and its relation to disease severity. Egypt. J. Neurol., Psychiat. Neurosurg., 54(1): 16.

Naranjo-Galvis, C.A., McLeod, R., Gómez-Marín, J.E., de-la-Torre, A., Rocha-Roa, C., Cardona, N., Sepúlveda-Arias, J.C., 2023. Genetic Variations in the Purinergic P2X7 Receptor Are Associated with the Immune Response to Ocular Toxoplasmosis in Colombia. Microorganisms., 11(10): 2508.

Ng, P.C., Henikoff, S., 2002. Accounting for human polymorphisms predicted to affect protein function. Genome Res., 12(3): 436-446.

Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K.A., Lin, G.N., Nam, H.-J., Mort, M.,

Cooper, D.N., Sebat, J., Iakoucheva, L.M., 2020. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. Nat. Commun., 11(1): 5918.

Qi, X., Li, F., Zhang, Y., Zhu, H., Yang, F., Li, X., Jiang, X., Chen, L., Huang, Y., Zhang, J., 2022. STAT4 genetic polymorphism significantly affected HBeAg seroconversion in HBeAg-positive chronic hepatitis B patients receiving Peginterferon-α therapy: A prospective cohort study in China. J. Med. Virol., 94(9): 4449-4458.

Radivojac, P., Vacic, V., Haynes, C., Cocklin, R.R., Mohan, A., Heyen, J.W., Goebl, M.G., Iakoucheva, L.M., 2010. Identification, analysis, and prediction of protein ubiquitination sites. Proteins: Struct. Funct. Bioinf., 78(2): 365-380.

Ramazi, S., Zahiri, J., 2021. Post-translational modifications in proteins: resources, tools and prediction methods. Database, 2021: baab012.

Rodrigues, C.H., Pires, D.E., Ascher, D.B., 2021. DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations. Protein Sci., 30(1): 60-69.

Rozario, L.T., Sharker, T., Nila, T.A., 2021. In silico analysis of deleterious SNPs of human MTUS1 gene and their impacts on subsequent protein structure and function. PloS one, 16(6): e0252932.

Saini, S., Jyoti-Thakur, C., Kumar, V., Suhag, A., Jakhar, N., 2018. In silico mutational analysis and identification of stability centers in human interleukin-4. Mol. Biol. Res. Commun., 7(2): 67.

Seok, C., Baek, M., Steinegger, M., Park, H., Lee, G.R., Won, J., 2021. Accurate protein structure prediction: what comes next. Biodesign., 9(3): 47-50.

Sheena, B.S., Hiebert, L., Han, H., et al., 2022. Global, regional, and national burden of hepatitis B, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Gastroenterol. Hepatol., 7(9): 796-829.

Shi, H., He, H., Ojha, S.C., Sun, C., Fu, J., Yan, M., Deng, C., Sheng, Y., 2019. Association of STAT3 and STAT4 polymorphisms with susceptibility to chronic hepatitis B virus infection and risk of hepatocellular carcinoma: a meta-analysis. Biosci. Rep., 39(6): BSR20190783.

Shukla, S., Rawat, P., Sharma, P., Trivedi, P., Ghous, F., Bishnoi, A., 2023. Spectroscopic characterization, molecular docking and machine learning studies of sulphur containing hydrazide derivatives. Phys. Chem. Chem. Phys., 25(40): 27677-27693.

Singh, J., Malik, D., Raina, A., 2021. Molecular docking analysis of azithromycin and hydroxychloroquine with spike surface glycoprotein of SARS-CoV-2. Bioinform., 17(1): 11.

Srinivasan, S., Clements, J.A., Batra, J., 2016. Single nucleotide polymorphisms in clinics: Fantasy or reality for cancer? Crit. Rev. Clin. Lab. Sci., 53(1): 29-39.

Subbiah, U., Ajith, A., Subbiah, H.V., 2023. Prediction of Deleterious Non-Synonymous Single Nucleotide Polymorphism of Cathelicidin. Curr. Pharmacogenomics. Person. Med. (Formerly Current Pharmacogenomics), 20(2): 92-105.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., Meyre, D., 2019. Benefits and limitations of genome-wide association studies. Nat. Rev. Genet., 20(8): 467-484.

Tan, M., Bhadoria, A.S., Cui, F., Tan, A., Van Holten, J., Easterbrook, P., Ford, N., Han, Q., Lu, Y., Bulterys, M., Hutin, Y., 2021. Estimating the proportion of people with chronic hepatitis B virus infection eligible for hepatitis B antiviral treatment worldwide: a systematic review and meta-analysis. Lancet. Gastroenterol. Hepatol., 6(2): 106-119.

Vallejos-Vidal, E., Reyes-Cerpa, S., Rivas-Pardo, J.A., Maisey, K., Yáñez, J.M., Valenzuela, H., Cea, P.A., Castro-Fernandez, V., Tort, L., Sandino, A.M., 2020. Single-nucleotide polymorphisms (SNP) mining and their effect on the tridimensional protein structure prediction in a set of immunity-related expressed sequence tags (EST) in Atlantic salmon (Salmo salar). Front. Genet., 10: 1406.

Verma, A.K., Jaiswal, G., Sultana, K.N., Srivastava, S.K., 2023. Computational

studies on coumestrol-ArlR interaction to target ArlRS signaling cascade involved in MRSA virulence. J. Biomol. Struct. Dyn., 1-19.

Wang, D., Liu, D., Yuchi, J., He, F., Jiang, Y., Cai, S., Li, J., Xu, D., 2020a. MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. Nucleic Acids Res., 48(W1): W140-W146.

Wang, D., Zeng, S., Xu, C., Qiu, W., Liang, Y., Joshi, T., Xu, D., 2017. MusiteDeep: a deep-learning framework for general and kinase-specific phosphorylation site prediction. Bioinform., 33(24): 3909-3916.

Wang, Y., Feng, D., Wang, H., Xu, M.-J., Park, O., Li, Y., Gao, B., 2014. STAT4 knockout mice are more susceptible to concanavalin A–induced T-cell hepatitis. Am. J. Pathol., 184(6): 1785-1794.

Wang, Z., Huang, C., Lv, H., Zhang, M., Li, X., 2020b. In silico analysis and high-risk pathogenic phenotype predictions of non-synonymous single nucleotide polymorphisms in human Crystallin beta A4 gene associated with congenital cataract. Plos one, 15(1): e0227859.

Zeng, Z., Liu, H., Xu, H., Lu, H., Yu, Y., Xu, X., Yu, M., Zhang, T., Tian, X., Xi, H., 2021. Genome-wide association study identifies new loci associated with risk of HBV infection and disease progression. BMC Med. Genom., 14: 1-12.

Zhang, Y., Skolnick, J., 2004. SPICKER: a clustering approach to identify near-native protein folds. J. Comput. Chem., 25(6): 865-871.

Zhang, Y., Skolnick, J., 2005. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res., 33(7): 2302-2309.