Machines and Algorithms
http://www.knovell.org/mna

Research Article

A Framework for Analysis of SNPs in TAGAP Gene

Haider Usman[1], Qaisar Rasool[1,*], Saad Rasool[1] and Hifssa Aslam[1]

[1]Department of Computer Science, Bahauddin Zakariya University, Multan, 60000, Pakistan

[*]Corresponding Author: Qaisar Rasool Email: qrasool@bzu.edu.pk

Received: 22 November 2023; Revised: 6 December, 2023; Accepted: 20 December 2023; Published: 14 March 2024

AID: 003-01-000033

Abstract: The role of TAGAP in the T cells plays a vital role and the analysis of how the structure and functional impact SNPs have on the TAGAP is the main focus that has been analyzed within the provided studies. Analysis of the genetic maps that are developed within the human body has been analyzed with testing the impact of the mutation on the TAPAG in T cells. The examination of non-synonymous Single Nucleotide Polymorphisms (nsSNPs) on the TAGAP gene functionality and structural characteristics. Four tests have been performed are SIFT, Polyphen-2, PROVEAN, and the stability of the protein was tested through the I-Mutants 3.0. Diabetes Mellitus and Auditory Agnosia are the two diseases that have been analyzed to depict the importance of the TAGAP in the T cells.

Keywords: TAGAP; T cells; Single Nucleotide; Polymorphisms (SNPs); Non-synonymous SNPs (nsSNPs); Protein structure; Protein functionality;

1. Introduction

The majority of autoimmune and infectious diseases are linked to the locus syntenic to the TAGAP genes. Some of these conditions include multiple sclerosis and candidemia, which are related to infections and the immune system. Unlike most studies, very little is known about the role of TAGAP beyond its function in T cells. SNP (Single Nucleotide Polymorphism) refers to the sequence or combination of DNA that arises from the substitution of a single nucleotide. After the change, the sequence of combined DNA is referred to as SNP. This data consists of sequences of four nucleotide bases: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). Simply put, SNP is the most common method used for gene transmutation. According to research, most humans share this mutation, and having a single SNP is common within the human population [1]. Once a study is completed and the association between genetic variations and their phenotypes is understood, the resulting data can help us understand how various diseases are connected to these mutations. These mutations can occur in the coding or non-coding regions of genes. SNPs located in the coding region of genes do not alter the amino acid sequence of newly formed proteins. There are two kinds of SNPs in the coding region. One type, known as synonymous SNPs, has no impact on the shape or sequence of amino acids. The other type, nonsynonymous SNPs, affects the amino acids and can potentially alter the protein sequence. Nonsynonymous SNPs are further divided into nonsense and missense SNPs. These types of SNPs modify the amino acids and can disrupt the function of proteins, causing potential harm [2].

Numerous SNPs occur naturally within the human body, leading to mutations and the development of different genetic maps. This results in various diseases that respond differently to drugs. It is clear that diseases are directly related to these mutations. In other cases, most SNPs are neutral and do not affect

protein function. Therefore, understanding the role of SNPs is essential for unraveling the complex traits and diseases present in humans [3].

This study aims to achieve the following objectives:

- To perform a computational analysis of SNPs associated with TAGAP, focusing on their occurrence and role in the TAGAP locus.
- To investigate how interactions between SNPs and TAGAP predict patient behavior and aid in diagnosing autoimmune, infectious, or mutation-related diseases.
- To evaluate the structural and functional impact of deleterious SNPs on TAGAP proteins, focusing on their role in disease mechanisms.
- To identify functional SNPs with significant relevance that can be utilized for predicting, diagnosing, or treating fungal infections, diabetes, or other mutation-related conditions.

The research hypotheses are as follows:

- SNPs associated with TAGAP significantly influence susceptibility to autoimmune and infectious diseases, such as multiple sclerosis and candidemia.
- Deleterious SNPs in TAGAP alter protein structure and function, contributing to disease pathogenesis.
- Functional SNPs in TAGAP can be computationally identified and utilized to develop predictive models for patient behavior and disease diagnosis.
- Structural changes in TAGAP proteins due to nonsynonymous SNPs are correlated with their impact on immune system functionality and the development of genetic disorders.

In this section, we discuss a brief introduction to the paper, including the study's aim and hypothesis. The second section is a follow-up with the literature review. The methods used to compile the results are detailed in Section 3. The study's findings and results are presented in the fourth part. Section five contains the discussion, while all sections end the paper with conclusions.

2. Literature Review:

A peer-reviewed work by Anne Ndungu et. all [4] that presents multi-SNP models for gene expression analysis. Article describes gene-level investigation of SNP prediction models using TWAS. The research report analyses 43 human tissues using GTEx (Genotype-Tissue Expression Project) multi-SNP gene models and an iterative modeling scheme. Compared to multi-SNP analysis, cis-eQLT variance is lower. About 90% of the 826 gene metabolite pairings analyzed in the paper were dominant of the single eQTL. The framework described in the research was important since more than 90% of colonization signals were created within QTLs in the multi-SNP model. The research found that 8% of TWAS were linked to the causative genes. The results show considerable gaps in the research, and more legitimate SNP analysis employing multi genes is needed.

GWAS (huge genome-wide association studies) have enhanced RA genetic risk factor knowledge. The genetic susceptibility of groups with high Amerindian admixture is unknown. The research aimed to assess prior RA locus reports' generalizability with admixed ancestry in Latin America. In linkage equilibrium, researchers selected about 128 SNPs (single nucleotide polymorphisms) with high RA association in non-Amerindian groups. To genotype roughly 118 SNPs in 313RA participants/487 healthy controls, mid-density polymerase chain reaction was used. The second cohort (250 cases/290 controls) confirmed the connection. The SNP rs2451258 was shown to be linked to RA and 18 additional markers were suggestive. It was upstream of the TAGAP. Haplotype testing showed a strong correlation between neighbouring SNPs and RA near the single transcription activator and transducer4. The study found no replication of earlier genetic link reports to RA. These data suggest that LA population admixture mapping and GWAS can identify novel loci associated with RA. This study can help clinicians and researchers comprehend this disease's pathogen omics. Research on trans-population disparities in RA is also possible [5].

Pehliva et. all [6] studied paediatric TAGAP polymorphism gene patients. There are a number of hypotheses about how the rs1738074 T/C SNP affects TAGAP. No study had focused on Turkish paediatric

patients. The researchers wanted to examine the connection between celiac disease and diabetes mellitus type 1 in Turkish paediatric patients with TAGAP gene SNPs. Researchers used IBM SPSS. This study included 127 paediatric patients and 100 healthy youngsters. We found the polymorphism using an allele-specific polymerase chain reaction. The researchers used IBM SPSS, Arlequin 3.5.2, and Statistics 25.0 for statistical analysis. Researchers may not be biased. The data also showed that 72% of 154 CD patients had C alleles. In addition, 28%, or 60 CD patients, carried the T allele. We also found the C and T alleles in 43.5% and 57.5% of patients with celiac disease and type 1 diabetes, respectively. The control group comprised 67% C alleles and 33% T alleles. The results also showed no significant difference in allele frequencies and genotype between the control group and the patient. The investigation found no significant link between the polymorphism and disease risk.

Rheumatoid arthritis, the most widespread, persistent, and progressive inflammatory illness, damages joints and increases mortality. The C-C motif ligand 21 is a cytokine that is involved in immunological regulation and inflammation. As a result, SNP analysis is the most important in the CCL21 gene because it helps evaluate their functional and structural relevance in finding potential treatment targets for immune-related illnesses like RA. This work identified the most harmful non-synonymous SNPs that affect CCL21 protein function and structure using in silico methods. The main roles in this research may include SNPs&GO, PROVEAN, PolyPhen2, and SIFT. We validated the functional and structural effects, stability, and conservation profile of the other tools using MutPred, I-Mutant, and ConSurf. The results revealed a post-translational modification site. The research is also important for identifying and analyzing human functional SNPs and TAGAP proteins. Chimera v1.11 proposes 3-D protein modeling with I-TASSER. The findings also suggest that non-synonymous SNPs may cause CCL2 protein dysfunction and autoimmune disorders like RA. According to the research, these non-synonymous SNPs may be the most important in examining the CCL21 gene's link to autoimmune disorders like Crohn's Disease (CD) and RA. To determine their suitability for genome editing and pharmacogenomics, these SNPs must be tested in animal studies and tissue samples from diseases [7].

SNPs in the CCR6 gene may cause Lupus nephritis, systemic sclerosis, rheumatoid arthritis, psoriasis, and other autoimmune illnesses. Functional and structural identifications are important polymorphisms for therapeutic and dysfunctional target research. Bioinformatics methods have identified damaged nsSNPs that affect CCR6 function and structure. They utilized PolyPhen2, SNP&GO, SIFT, and PROVEAN to model proteins in 3D in a computer, and Gene MANIA and STRING to guess how genes would interact with each other. The three nsSNPs rs751102128, rs1185426631, and rs1376162684 do the most damage to the CCR6 gene. On the other hand, the seven missense rs139697820, rs1282264186, rs768420505, rs1263402382, rs139697820, rs769360638, and rs1438637216 go back to stop codons. Because of its probable phosphorylation location, rs1376162684's highlighted post-transcriptional alteration is feasible. Gene-gene interactions have demonstrated CCR6's role in several co-expressions and pathways. After that, we can use these ten nsSNPs to study disorders related to CCR6 [8].

SNPs can help DNA develop and repair, according to another study. This study demonstrates that mitochondrial dysfunction repairs DNA and helps it fight the body. These dysfunctions strengthen DNA with SNPs and reactive oxygen and nitrogen species. DNA is composed of SNPs, reactive oxygen, and nitrogen. All of these substances help DNA resist genetic mutations. When its ability increases, DNA will attempt to repair itself and stop mutations. When DNA does not change its amino acid pattern when it becomes too powerful to battle mutations. This increases immunity, DNA strength, and the body's ability to fight cancer. Thus, SNPs and mitochondrial dysfunction strengthen the body, immune system, and DNA repair [9].

3. Methodology

3.1. Collection and compilation of the Database

We retrieved the T cell protein sequence for TAGAP from the protein database website (www.uniprot.org/UniProt/Q8N103). Obtain nsSNP retrieval information for experimental data analysis

from the database.We utilized the National Center for Biotechnology Information (NCBI) at ncbi.nlm.nih.gov/SNP. The NCBI and Entrez retrieval systems have utilized it. The experiment may utilize a variety of computing algorithms. Ensure that dbSNP errors won't affect study outcomes [10]. This classifies nsSNPs as tolerant and neutral using the computing algorithm. Protein stability is crucial for experimental findings, and its stability may be determined via computer data analysis.

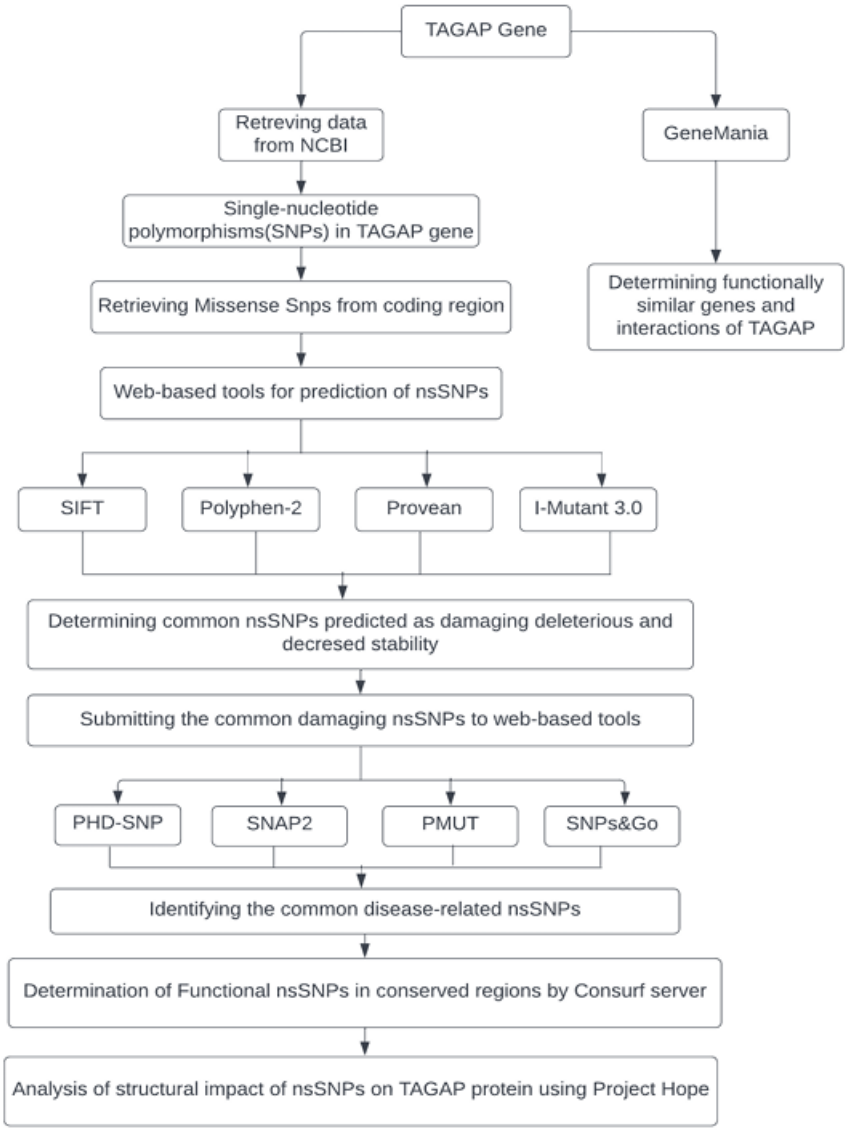

Figure 1: Methodology of the TAGAP Gene Analysis

3.2. Determining the Functional nsSNPs

There were eight distinct types of prediction tools tested to demonstrate the computational usefulness of nsSNPs. First, four prediction techniques were used to show functional nsSNPs, including the following.

- I Mutant (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/)
- SIFT(https://sift.bii.a-star.edu.sg/)
- Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/)
- ROVEAN (http://provean.jcvi.org/index.php)

After analysing nsSNPs that may harm the protein, PhD-SNP (https://snps.biofold.org/phd-snp/phd-snp.html) was used. The last one is SNPs&GO (https://snps.biofold.org/snpsand-go/snps-and-go.html). These simple tools will assist analyze disease and neutral SNPs. Further details on all computational tools are provided above in figure 1.

3.2.1. SIFT

Using Sorting Tolerant from Tolerant (SIFT) server, study reveals harmful coding non-synonyms in SNP. SIFT is mostly used to identify the link between phenotypic variation and mutations [11]. This involves assessing the influence of amino acids on protein and testing their capabilities. The SIFT analyses protein families based on their ability to preserve amino acids, a pre-determined criterion. Damage can be assessed by observing changes in well-conserved places.

The prediction software-defined will get a query first. Co-ordinates will be created from nsSNPs in prediction software. SIFT works by analysing given queries and applying alignment information in numerous ways. The program scores 3.0 for depicting protein sequences based on the MSCS (Median Sequence Conservation Score). It is easier to determine if nsSNPs are harmful or tolerated. The SIFT performance involves several processes, including:

- First, select a comparable sequence.
- Next, choose a similar sequence based on functionality.
- These sequences are analysed to find several alignments.
- Each location's probabilities are compared to probable substitutions.

Next, select a cut-off value to determine if a replacement is harmful or accepted at a certain spot. If the probability cut-off value exceeds the feasible possibilities, the program declares it detrimental for further analysis. A cut-off value of TI > 0.05 is defined for specific algorithms.

3.2.2. PROVEAN

Protein Variation Impact The PROVEAN analyzer tool predicts and determines the influence of amino acid substitutions on protein functioning inside the system software [12]. The initial stage was retrieving nsSNP information from the NCBI database. Clustering and supporting sets will be used to eliminate redundancy. The PROVEAN requires average delta storage computation as the second step. Effective output was achieved by clustering 75% of globally recognized BLAST proteins using captured sequences. Next, a succession of supporting sets will be created from the top clusters in the experimental study. Research indicates that variations in protein similarity can harm the studied protein. Changes in the delta score are evaluated to see if the protein's functioning has changed, aiding in experimental data collecting. A low delta score for a protein modification might have a greater detrimental influence on its functioning. A score of less than 2.5 indicates higher damage to protein functioning, whereas scores above 2.5 indicate neutrality. The purpose of PROVEAN is to collect nsSNPs with scores below 2.5 for future experimentation.

3.2.3. Polyphen-2

The Polyphen-2 test is included to demonstrate how substituting an amino acid affects the protein's functioning and structure. This assay analyses protein sequence alignments from several sources and describes a three-dimensional structure [13]. The Bayesian classifier was used to examine the impact of mutations. Two query types are used in the Polyphen-2 test: gene ID and protein sequences. The Uniport is used to get information, including amino acid replacement. The protein sequence was used in FASTA format. The substitute score will be categorized as 30 through server prediction. The position-specific independence count score (PSIC) is a projected score classification with a value between 0 and 1. Increased PSIC scores indicate more impact on protein analysis through amino acid replacement, and vice versa. Scores vary from 0 to 1 and are divided into three types: Potentially dangerous, perhaps destructive, and benign are categorized based on the score.

3.2.4. PhD-SNP

A similar technique has been used to illustrate how amino acid substitutions might change protein functioning and potentially cause illnesses. Various forms of support vector machines are utilized for PhD SNP testing [14]. The protein sequence was obtained by testing to illustrate the study report results. Retrieving Swiss-PROT codes from the NCBI database and obtaining the protein sequence in FASTA format was crucial. Mutation levels were recorded in the analyzed output table of protein sequences. The output table displays both novel and Wild-type amino acid sets. Two types of nsSNPs exist: neutral and disease-related. A crucial step is to use PhD-SNP to distinguish disease-related nsSNPs from benign ones.

3.2.5. I-Mutant 3.0

I-Mutant 3.0 analyzes protein alterations that may affect stability. Gibbs' free energy change process helps analyze protein heat capacity variations, including temperature transitions [15]. Alterations to protein stability can be identified using I-Mutant 3.0. The format sequence will be retrieved from Uniport to achieve the appropriate format (FASTA). The data input includes the new residue of correlated values and the mutation site to determine the change in free energy, known as Delta Delta G (a measure for forecasting the impact of a single point mutation on protein stability). There are three prediction categories based on DDG value: DDG values below -0.5 kcal/mol indicate severe instability, while values above 0.5 kcal/mol indicate severe stability. The neutral value is shown as 0.5kcal/mol DDG.

3.2.6. SNAP2

A useful tool for analysis is SNAP2, which assigns amino acid substitutions based on their impact and neutrality categories [16]. The impact of SNPs on protein function will be studied utilizing biophysical properties, including structural and evolutionary information. The sequences in the study report will be shown in the same FASTA format as when downloaded from Uniport. SNAP2 employs a neutral network method. The TAPAG gene sequence should be entered into SNAP2 to classify amino acid substitutions and analyze their influence on protein features like neutrality or effect. The Reliability Index RI shows the value from +100 to -100. An analysis will be based on the score, with a strong neutral value of -100 and a strong influence on protein characteristics and functionality at +100.

3.2.7. SNPs&GO

This tool determines the disease-nsSNP connection. Additionally, the method helps determine disease or neutral SNP effects on protein [17]. SVM is the tool's main substitution classification technique. The three-dimensional protein structure and protein sequence profile from the tool that explains the protein's functions are used to retrieve the information. A stringently specified trained and tested approach is used to examine cross validation utilizing a prediction tool. Provide input in FASTA format to determine and inspect output. If the likelihood 32 score exceeds 0.5 in this test, the system evaluates the disease-related impact.

3.2.8. PMut

This free tool aids analysis. The PMut mechanism operates on two levels, obtaining information from local databases, mostly as mutational hotspots. Protein SNPs with specified identities are evaluated in the second step of PMut analysis. The study report requires extensive mutational effort to portray the possible mutations. This tool shows the association between protein illness and amino acid substitution type, as well as the impact of amino acid replacement on protein neutrality. Mendelian mutational analysis is performed using the technique shown in the paper.

3.3. Determination of Functional SNPs that are based on Conserved Region

The test relies on conserved amino acids for experiment evolution. After collecting SNPs from the computational tool, the conserved evolutionary amino acids in the TAPAG gene will be examined to provide the research findings. Next, the TAPAG discussion score will be analysed and shown as study results. The ConSurf server will analyse harmful SNPs for the study report.

3.3.1. ConSurf Server

Certain amino acids include a protein, and particular nucleic acids in RNA or DNA must balance the inherent inclination of the protein to mutate [18]. Macromolecules must also keep their structural and informational identities. The ConSurf service will analyze evolutionarily present amino acids in RNA, DNA, and proteins. The ConSurf Server is used to extract phylogenetic relationships between homologous substance sequences. A unique phylogenetic tree was created using the ConSurf Server, which also analyzes evolutionary conservation. When the score ranges from 1 to 4, factors are named, intermediate scores are depicted, and conserved scores are recorded (7 to 9). The Bayesian calculation approach is used in research report analysis. The ConSurf Server defines protein sequences in FASTA format for prediction as per analysis. A tabular structure helps define the overall conservation score, which is displayed through a color scheme. Protein features linked to functioning and structural complexity are predicted and assessed. To analyze the experimental technique and acquire data, highly conserved amino acids will be chosen.

3.4. Examination of the impact of nsSNPs on protein structural properties

The most important step to follow is an analysis of the impact and effect of mutations of protein structural properties that have been depicted through the deleterious nsSNPs.

3.4.1. Project Hope Utilization

The aim is to compare the wild-type amino acids used in the study report to the mutant amino acids in real time. The program gathers several sources of information to create a three-dimensional picture of protein [19]. This data includes sequence annotations from the study paper. This software program was used to calculate 3D coordinate prediction and services like DAS in the research project. The research uses a specific technique to analyze and illustrate residual traits, whether they are native or new. nsSNPs identified in earlier phases and filtered to meet study objectives are evaluated using this tool. Variants based on various protein sequences have been submitted to Hope software [20]. After submitting the response, the difference between wild-type amino acids and novel residue amino acids was examined. Two approaches have been devised to compare the size and hydrophobicity of new and ancient residues. The difference analysis helps determine the impact of the mutation on protein functioning and structural features.

3.5. Analysis of the TAGAP Gene Interactions

GeneMANIA's database helps display the organization's gene base characteristics. The research project used functional association data for experimentation. Genetic relationship including protein co-localization and co-expression with a comparable pathway. The research study analyzes gene expression by co-expression [21].

Table 1: TAGAP and its functionally similar genes

Gene	Description	Rank
TAGAP	T cell activation RhoGTPase activating protein	N/A
RHOH	ras homolog family member H	1
RHOV	ras homolog family member V	2
RHOT1	ras homolog family member T1	3

RHOF	ras homolog family member F	4
RHOU	ras homolog family member U	5
RHOBTB1	Rho related BTB domain containing 1	6
RHOBTB2	Rho related BTB domain containing 2	7
RHOT2	as homolog family member T2	8
CD69	CD69 molecule	9
RHOD	as homolog family member D	10
RHOJ	as homolog family member J	11
RAC2	Rac family small GTPase 2	12
RHOB	as homolog family member B	13
RHOC	as homolog family member C	14
RHOG	as homolog family member G	15
RHOQ	as homolog family member Q	16
RAC3	Rac family small GTPase 3	17
ARHGEF3	Rho guanine nucleotide exchange factor 3	18
PLEK	pleckstrin	19
DENND1C	DENN domain containing 1C	20

Their expression must be considered to perform the analysis with implementing and subjecting to different situations. A genetic expression is a tool that helps in the prediction of one gene and its functionality that has been associated with another type of gene that has been analyzed in the research report. The perturbation effect of one gene to another gene depict about the functionality that has been associated with one another according to the structure and functionality of the gene that has been analyzed.

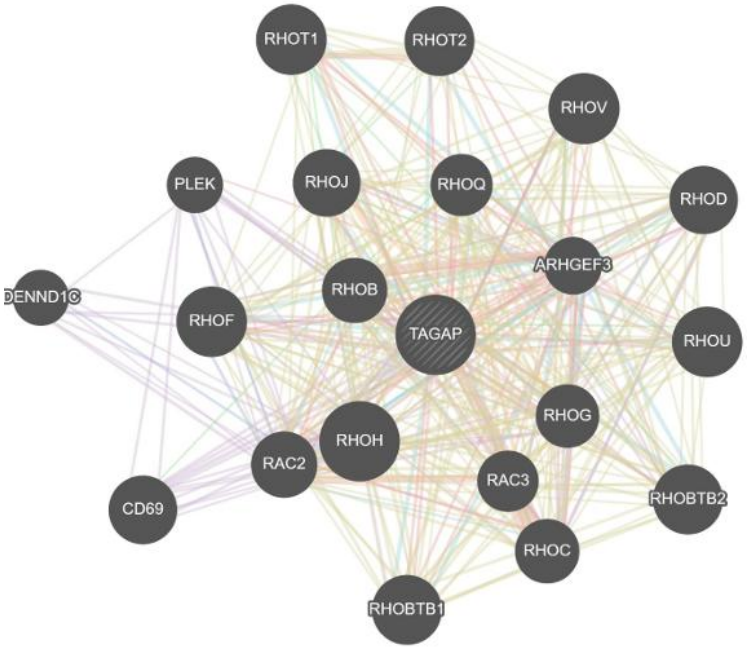

Figure 2: Gene-MANIA Interaction network of TAGAP gene

Studying gene products that are linked may suggest the existence of a related gene. The analytical approach will follow the following procedure to describe the research investigation.

Initially, data will be retrieved from the NCBI website to illustrate experimental data. SNP analysis in the TAGAP gene and missense Snaps from the coding area are required. Four useful online tools for analysis include SIFT, Polyphen-2 PROVEAN, and I-Mutant 3.0. Submit Common harmful nsSNPs to PhD-SNP, PMut, and SNPs&GO tools. The examination of nsSNPs in prevalent diseases led to the creation of the ConSurf Server and Hope Project analysis.

4. Results

4.1. SNPs dataset

Research indicates that database SNP effectively captures SNP interest due to its huge dataset. Research indicates over 100 million SNPs exist globally, with unique or common variances across individuals. These DNA alterations are usually found across genes.

As biological markers, they can help researchers find disease-linked genes. SNPs in a genotype or gene's promoter regions can directly affect sickness by altering gene activity. According to the chart below, the TAGAP gene has 17% nsSNPs, 9% 3'UTR SNPs, 8% 5'UTR SNPs, and 66% other SNPs [22].

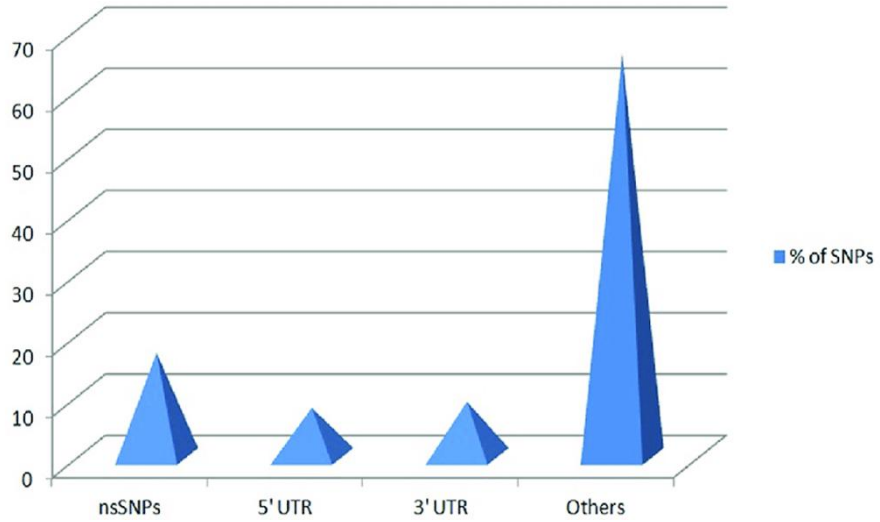

Figure 3: The pyramid in the form of the cluster is showing the SNPs percentage in TAGAP Gene

4.2. Sift, Polyphen-2, PROVEAN, I-Mutant 3.0 results

This study examined the functional effects of nsSNPs using SIFT, PROVEAN, and PolyPhen2. The I-Mutant was used to study the impact of nsSNPs on protein stability. The SIFT results indicate that nsSNPs have an intolerant scoring tolerance index. In Proven, mutations are often considered harmful if the final score is below -2.5 or if nsSNPs may be harmed. Polyphen-2 identifies potentially harmful, likely detrimental, and perhaps benign non-synonymous SNPs. Benign nsSNPs and destructive possibilities are the most accurate predictions compared to the other two. Most estimates are based on the independent count score and position-specific difference, with score 1 being the most detrimental. The research highlights nsSNPs that are prevalent in up to four algorithmic techniques and have a substantial score of zero in SIFT and PolyPhen-2. This is done to emphasize only very harmful SNPs.

TAGAP protein is significant as it is its sole functional domain. Therefore, these nsSNPs may be considered as potential causes of TAGAP dysfunction illnesses, facilitating treatment discovery and development [23].

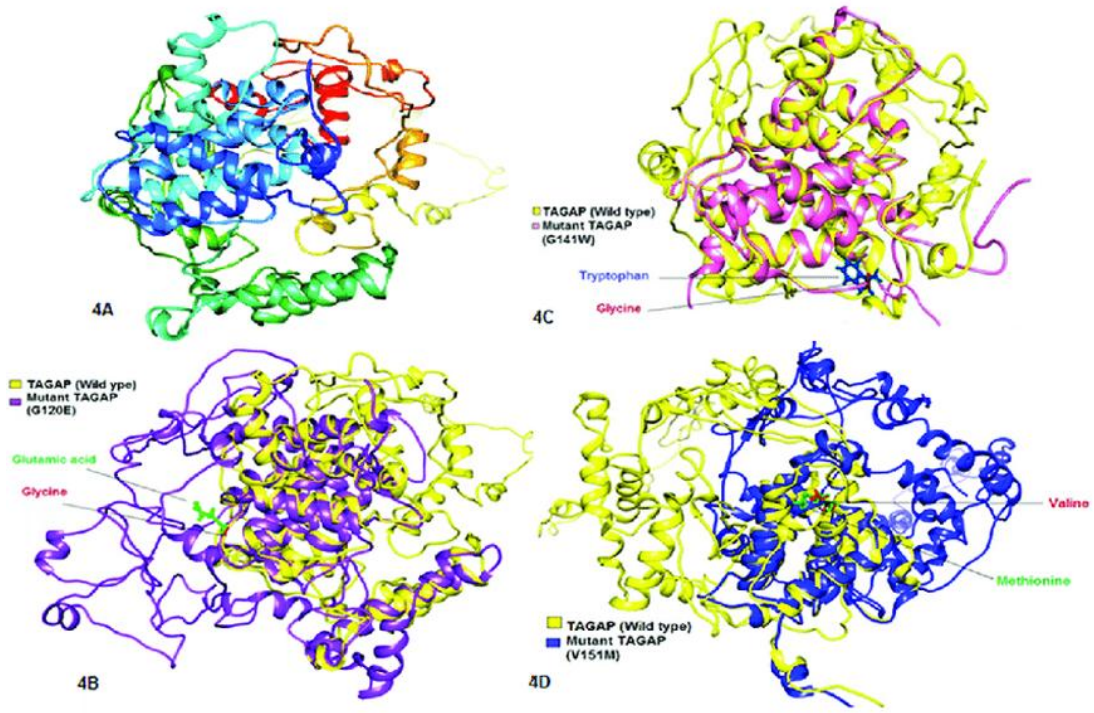

Figure 4: TAGAP structure of a protein with its mutant form [23]

4.3. PhD-SNP, SNAP2, PMut, SNPs&Go results

This study examined the functional effects of nsSNPs using PMut, SNAP2, PhD-SNP, and SNPs&GO. The study utilized several computational methods to identify non-synonymous SNPs that are susceptible to TAGAP protein structure and function, potentially leading to harmful disorders. The computational analysis was conducted using several tools, including Provean, SIFT, PolyPhen-2, PMut, SNPS&GO, and PhD-SNP. In addition, SNAP2, a trained neural network-based tool, uses many criteria to differentiate between sickness and benign alterations. Appreciable accuracy. We utilized the protein sequence as input and received two predictions: effect (high score) or neutral (low score) (negative score). Alternatively, SNPs&GO predicts TAGAP gene mutations using a protein sequence method. A likelihood score above 0.5 implies a disordered impact of the mutation on host protein interactions. PhD-SNP, a tool in SNPs&GO, evaluates data for harmful or neutral mutations. Additionally, PMut studies reveal that neural network intelligence yields 80% correct discoveries of each SNP's compulsive features. It's feasible to predict neutral or disorderly output [24].

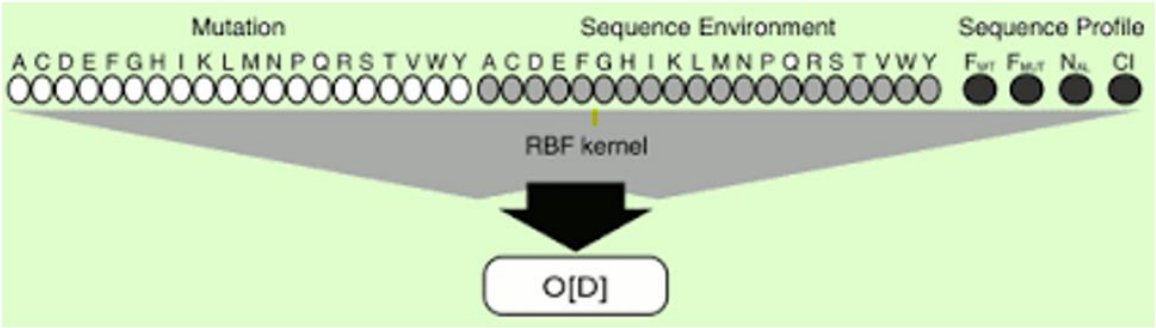

Figure 5: PhD-SNP [24]

4.4. ConSurf results

ConSurf analyzes past acylated acid rates and applies them to the search macromolecule's structure and sequencing. The ConSurf investigation can identify crucial places inside the search macromolecule, as the query interface's slowly shifting sections are crucial for operation. When the search macromolecule's infrastructure is in place, it can distinguish between slowly developing core sites, crucial for structural stabilization, and clusters of slowly changing surface locations, crucial for function [25].

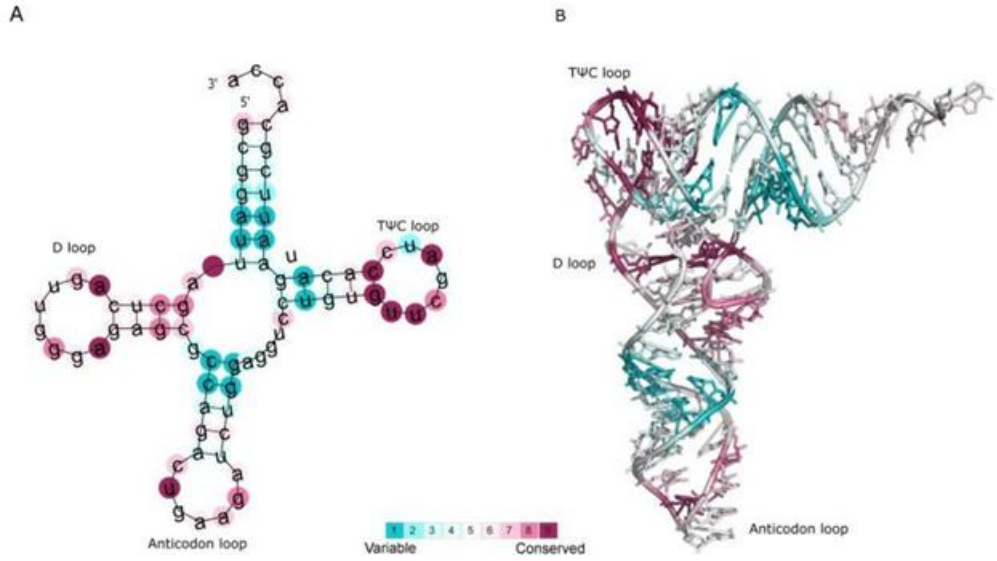

Figure 6: ConSurf Analysis

The ConSurf test in the research report showed the mutation's overall impact. Using the ConSurf service, amino acids evolutionary conversation was examined against 9 of the most damaging nsSNPs in TAGAP protein residues. The ConSurf server findings are shown using the TAGAP protein structure. Using solvent accessibility data, the ConSurf service has produced predictions about functional and structural residues. Based on conserved residue position, two alternatives are shown.

It may be on the protein's inner core or surface as defined by ConSurf. The ConSurf service analyzes the amino acid functionality of proteins to show that they are more conserved than other proteins. The more often nsSNPs are on the conserved reign, the more they damage protein functioning, according to ConSurf server data. Different findings from the ConSurf server show the TAGAP conservation level and structural and functional properties to be studied. The ConSurf service found TAGAP-exposed functional residues G141, E136, T118, N205, V151, and G141 to be highly conserved. The ConSurf server provides a p-value of 0.618 for the T118M mutation. Mutations L100F and F122L had p values of 0.573 and 0.846, respectively. Mutation G120E is 0.902 and N205S is 0.896. G141W and V151W had p-values of 0.663 and 0.676, respectively. The A126T p-value was 0.804 and E136K 0.498.

The TAGAP domain's nsSNP ID assessment findings. Rs748659041 contains amino acid change 100, L to F, and protein stability has deteriorated with RI value 9 and DDG score -0.62, including nsSNP TM Score 1. The given nsSNPs have RM SD 0. For rs368265576, the amino change is 118 T to M, and the results show lower stability as the above nsSNPs. The RI is 6 and the DDG score for that ID is -0.29. TM score is -0.29 and RMSD 0.83. nsSNP study for the ID rs764717611 showed a shift of amino acid 120 G to E, which decreased protein stability with a RI value of just 2 (ConSurf server data). The DDG score for the nsSNP ID is -0.84, with a TM score of 0.78894 and RMSD 1.98.

The stability of rs763380333 must be lowered using a ConSurf web server RI value of 3 to calculate the amino acid change from 122 F to L. The DDG value is -0.6, with a TM score of 0.98778 and RMSD of 0.83. For nsSNPs ID rs780953936, the amino acid change is 126 A to T, indicating a complete loss in stability and RI of 7 due to the ConSurf server. The DDG value is -1.04, with a TM score of 0.7912 and an

RMSD of 1.87. The stability of nsSNP ID rs866898464 is lowered, and the amino acid change is 136 E to K and score of 7 in RI. The ConSurf web server findings for rs866898464 show a DDG score of -1.12 and a Tm score of 0.98778. In ConSurf server data, an amino acid altered from 141 G to W has an ID of rs765146154, indicating low stability and a drop from its original value. The DDG is -0.58 and the TM score of the supplied nsSNPs ID is 0.78894. RMDB scored 1.98. The second final ID assessed by the ConSurf web server, rs777042268, showed decreased instability with 151 V to M. Due to the loss in stability, the total value score is RI 9 and DDG with a value score of -1.53. The RMSD score for the nsSNP ID is 1.99. The last ID evaluated with the ConSurf server is rs778438807, which shows decreased stability and a value of change in the amino acid of 205 N to S. Its RI is 2 with a DDG score of -2.45, TM Score of 0.78851, and RMSD score of 1.99.

Figure 7: Pictorial Representation of ConSurf Result of TAGAP (a)

ConSurf anticipated the amino acid-based TAGAP conversation profile. The highly conserved nsSNP ID rs748659041 with a residual and position of L100 has a conversation score of 8. The conversation score for nsSNP ID rs368265576 is 9, indicating good exposure and conservation. For rs764717611, G120 has a conservation value of 9, while for nsSNP ID analysis, F122 has a conservation score of 9. Position rs780953963 of A126 is highly exposed and has a conservation score of 9.

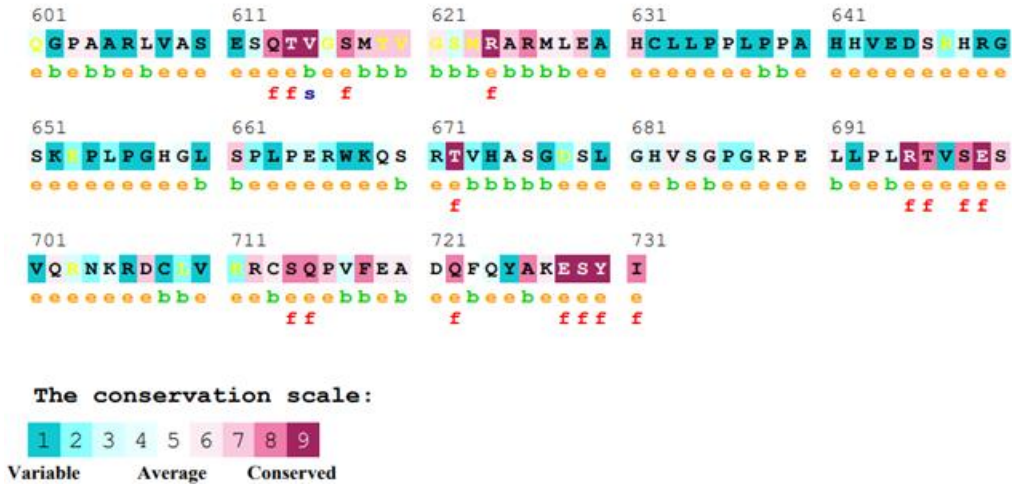

Figure 8: Pictorial Representation of ConSurf Result of TAGAP (b)

Analysis of nsSNP ID rs866898464 E136 has a conservation score of 9 indicating high exposure. The highly exposed protein G141, with nsSNP ID rs777042268 and conservation 8, was identified by ConSurf web analysis. The rs778438807 for V151 is well conserved, with a conservation score of 9 revealed. The N205 location (rs765146154) identified as highly exposed and conserved.

4.5. Project Hope Results

The HOPE results reveal that genetic polymorphism in the human genome is primarily based on SNPs, which are primarily composed of single base pair alternations in alleles. These alternations are the most common and significant type of variation in DNA sequences. The study indicates that non-synonymous SNPs cause mutations that lead to genetic diseases. Non-synonymous SNPs in the TAGAP gene are linked to severe illnesses due to their negative impact on protein structure. Results from SNP analysis are as follows. With project hope, protein SNP mutations have been explored. Various gene and protein mutations are associated to methionine synthase. The STRING database was used to examine these predictions. Further details on these mutations.

4.5.1. rs57752780 (V744L)

Mutation of amino acid to SNP V744L. Different angles have been used to see the residue where the mutation occurs. In the picture, H bonding is used for mutation. An angle of 744 on the left is employed. Mutations will occur at this angle, with SNPs present on the local bonding. All these interactions cause genetic mutation. Mutation changes the sequence of amino acids in DNA and SNPs, affecting their function.

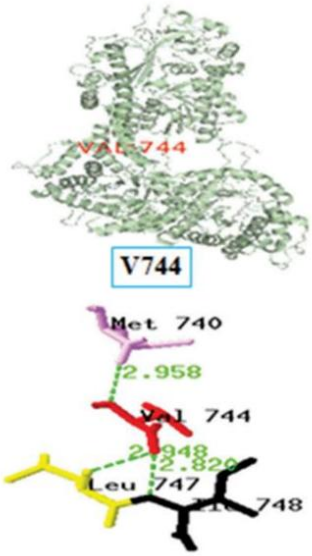

Figure 9: rs57752780 (V744L)

4.5.2. rs57752780 (L744L)

In this structure, gene mutations occur at angle 744 on the right, altering the genetic coding of DNA. Mutations can be harmful and affect the DNA sequence. These mutations also occur with DNA SNPs. This alters DNA sequence and creates a mutant gene.

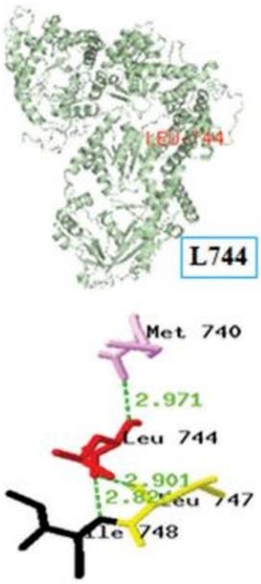

Figure 10: rs57752780 (L744L)

4.5.3. rs113914406 (G682D)

Mutations occur in amino acids rs113914406 at the angle of 682 on the left side of DNA due to hydrogen bonding. Hydrogen bonding is essential for SNP introduction and mutations. These mutations are harmful and cause genetic changes that impact human lives. Avoid this form of mutation to prevent potentially hazardous illnesses.

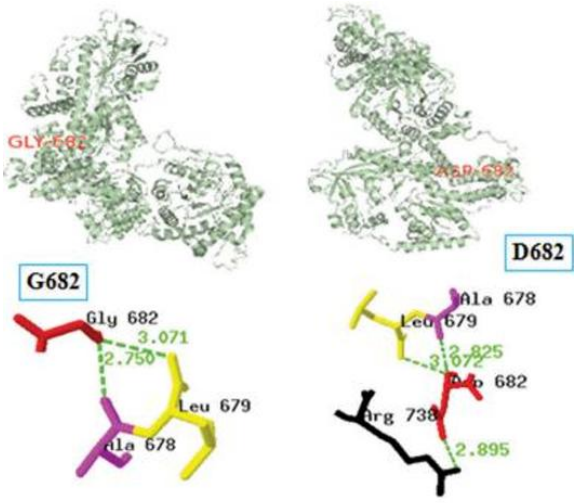

Figure 11: rs113914406 (G682D)

4.5.4. rs74710714 (V776E)

These SNP configurations enable mutation of amino acids rs74710714 on the left side of SPDBV at an angle of 766. Genetic mutation is promoted by gene interactions with GLY 828 and lle 826. Mutations can impact DNA function and the immune system. This form of mutation should be avoided. Mutations impair DNA repair. Mutations modify DNA structure, function, and ability by altering sequences via multiple mutations.

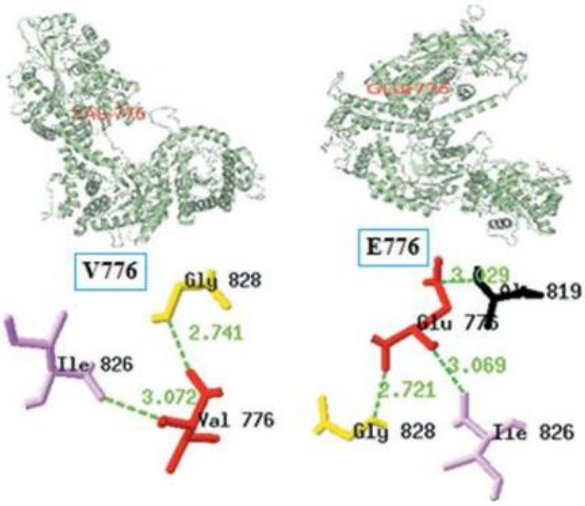

Figure 12: rs74710714 (V776E)

4.5.5. rs116836001 (R1027W)

A mutation at the right side of DNA at the angle of 1027 alters the mutation. This mutation occurs because to hydrogen bonding and SNP present. We know there are two SNP kinds. One form of SNP, synonymous, contributes to mutation. These mutations are harmful and increase the risk of cancer in humans. This sort of genetic mutation weakens the immune system, making the body more susceptible to many illnesses. Mutations can cause the death of many human cells, including white blood cells, red blood cells, and impulsive neurons in the central nervous system.

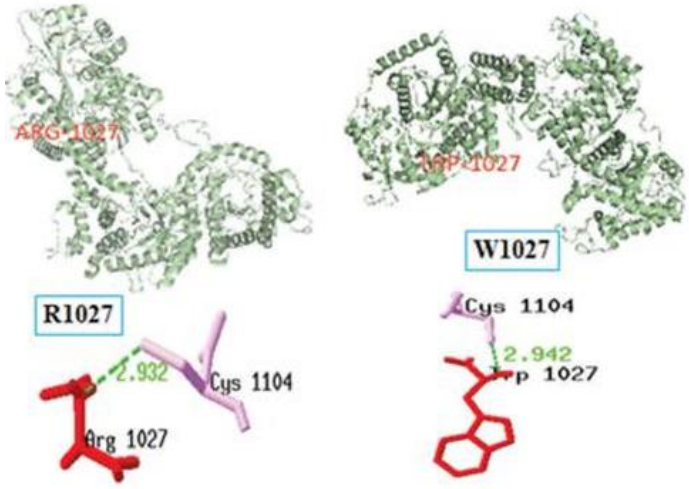

Figure 13: rs116836001 (R1027W)

5. Discussion:

The database analysis identified multiple SNPs, highlighting the challenges in determining which SNPs significantly influence protein structure and functionality. This study emphasizes the importance of distinguishing between coding and non-coding SNPs to better understand their effects on protein function, aligning with findings from Lim et al. (2021), who noted the critical role of SNP analysis in addressing genetic illnesses and improving cancer treatments. The TAGAP gene is responsible for the regulation of the immune system. Being closely related to various diseases like chronic myeloid leukemia (CML) linked to mutations, diabetes mellitus related to insulin regulation, it has been considered in studies.

From more than 60 significant SNPs analyzed through I-Mutant 3.0 and Consurf tools, nine missense SNPs have been identified that contribute the most deleterious impact on protein stability and functionality. These findings are in agreement with earlier research studies, such as Raghav & Sharma, 2013, that underscore the involvement of nsSNPs in protein malfunction and disease causation. The conservation analysis reinforces the importance of amino acid preservation in maintaining protein function, suggesting that mutations in conserved regions are more likely to disrupt gene stability and lead to diseases.

Potential Applications: The study has important implications for precision medicine. Identification of disease-causing SNPs in the TAGAP gene can be used to develop predictive models for autoimmune diseases, diabetes, and cancer, thus facilitating early diagnosis and personalized treatment approaches. In addition, knowledge about SNP-induced protein instability can help in drug discovery, especially when targeting structural weaknesses in proteins related to genetic diseases.

The study includes limits and biases but offers significant insights. Computational technologies like I-Mutant 3.0 and Consurf surface cannot fully reproduce biological system complexity, hence they may be inaccurate. Additionally, the NCBI SNP dataset may not fully cover varied populations, limiting generalizability. Missense SNPs are the main focus of the study, which does not examine synonymous or non-coding SNPs. The absence of experimental validation for computational discoveries makes it difficult to confirm the biological impact of detrimental SNPs on protein structure and function. To overcome these constraints, future study should use experimental methods and a more diversified genetic sample.

6. Conclusion:

The research report reveals that TAGAP, the most important protein in the human body, has numerous effective mechanisms. Over 50,000 single nucleotide polymorphisms (SNPs) were initially analysed to predict their behavioural impact on protein structural properties. Four tests, SIFT, PROVEAN, and

Polyphen-2, were performed to analyse the most affecting nsSNPs, which have the most adverse impact on the protein's functionality, leading to diseases like cancer, anxiety, and Diabetic Mellitus. The structural effect of SNPs on TAGAP is depicted through the Project Hope project. Over 30 nsSNPs were identified, impacting the protein's conservation and functionality. The physicochemical properties affected by mutations on the TAPAG gene were also assessed. Nine major nsSNPs were identified, showing extreme exposed and conserved surfaces. The study concludes that more research is needed to improve the protein's effectiveness in the human body, enhancing gene functionality and structural properties. The research contributes to the overall effectiveness of protein functionality in the human body.

References

[1] Li, Pei, Maozu Guo, Chunyu Wang, Xiaoyan Liu, and Quan Zou. "An overview of SNP interactions in genome-wide association studies." *Briefings in functional genomics* 14, no. 2 (2015): 143-155.

[2] Selvaraj, Suganya, and Shanmughavel Piramanayagam. "Impact of gene mutation in the development of Parkinson's disease." *Genes & diseases* 6, no. 2 (2019): 120-128.

[3] Ho, Daniel Sik Wai, William Schierding, Melissa Wake, Richard Saffery, and Justin O'Sullivan. "Machine learning SNP based prediction for precision medicine." *Frontiers in genetics* 10 (2019): 267.

[4] Ndungu, Anne, Anthony Payne, Jason M. Torres, Martijn van de Bunt, and Mark I. McCarthy. "A multi-tissue transcriptome analysis of human metabolites guides interpretability of associations based on multi-SNP models for gene expression." *The American Journal of Human Genetics* 106, no. 2 (2020): 188-201.

[5] Castro-Santos, Patricia, R. A. Verdugo, R. Alonso-Arias, M. A. Gutiérrez, J. Suazo, J. C. Aguillón, Jordi Olloquequi et al. "Association analysis in a Latin American population revealed ethnic differences in rheumatoid arthritis-associated SNPs in Caucasian and Asian populations." *Scientific Reports* 10, no. 1 (2020): 7879.

[6] Pehlivan, Melek, Tülay K. Ayna, Maşallah Baran, Mustafa Soyöz, Aslı Ö. Koçyiğit, Burcu Çerçi, and İbrahim Pirim. "Investigation of TAGAP gene polymorphism (rs1738074) in Turkish pediatric celiac patients." *Turkish Journal of Biochemistry* 46, no. 3 (2021): 293-298.

[7] Ali, Yasir, Mehran Akhtar, Kainat Khan, Nadia Farooqi, Shahla Gohar, Syed Ishfaq Ahmad, Madeeha Ayaz, Zia Ul Islam, Maria Arshad, and Fazal Jalil. "Screening for Deleterious non-synonymous SNPs in Human CCL21 Gene using in-silico analysis." *NUST Journal of Natural Sciences* 6, no. 2 (2021).

[8] Akhtar, Mehran, Tazkira Jamal, Hina Jamal, Jalal Ud Din, Muhsin Jamal, Muhammad Arif, Maria Arshad, and Fazal Jalil. "Identification of most damaging nsSNPs in human CCR6 gene: In silico analyses." *International journal of immunogenetics* 46, no. 6 (2019): 459-471.

[9] Czarny, Piotr, Paulina Wigner, Piotr Galecki, and Tomasz Sliwinski. "The interplay between inflammation, oxidative stress, DNA damage, DNA repair and mitochondrial dysfunction in depression." *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 80 (2018): 309-321.

[10] M. J. Islam, M. R. Parves, S. Mahmud, F. A. Tithi and M. A. Reza, "Assessment of structurally and functionally high-risk nsSNPs impacts on human bone morphogenetic protein receptor type IA (BMPR1A) by computational approach," Computational biology and chemistry, 80, pp. 31-45, 2019.

[11] R. Vaser, S. Adusumalli, S. N. Leng, M. Sikic and P. C. Ng, "SIFT missense predictions for genomes," Nature protocols, 11(1), pp. 1-9, 2016.

[12] Y. Choi and A. P. Chan, "PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels," Bioinformatics, 31(16), pp. 2745-2747, 2015. 55

[13] Y. S. L. B. B. C. M. J. Itan, J. G. Markle, R. Martinez-Barricarte and J. L. Casanova, "The mutation significance cutoff: gene-level thresholds for variant predictions," Nature methods, 13(2), pp. 109-110, 2016.

[14] R. S. E. Mohamed, "Early Detection of Parkinson's Diseases Using Bioinformatics and fMRI Image Processing," Doctoral dissertation, University of Gezira, 2018.

[15] M. A. Beg and L. S. Meena, "Mutational effects on structural stability of SRP pathway dependent co-translational protein ftsY of Mycobacterium tuberculosis H37Rv," Gene Reports, 15, p. 100395, 2019.

[16] M. Hecht, Y. Bromberg and B. Rost, "Better prediction of functional effects for sequence variants," BMC genomics, 16(8), pp. 1-12, 2015.

[17] E. Capriotti, P. L. Martelli, P. Fariselli and R. Casadio, "Blind prediction of deleterious amino acid variations with SNPs&GO," Human mutation, 38(9), pp. 1064-1071, 2017.

[18] R. H. Smith, Z. M. Khan, P. M. U. Ung, A. P. Scopton, L. Silber, S. M. Mack and A. C. Dar, "Type II binders targeting the “GLR-out” conformation of the pseudokinase STRADα," Biochemistry, 60(4), pp. 289-302, 2021.

[19] M. Nailwal and J. B. Chauhan, "Computational analysis of high risk missense variant in human UTY gene: a candidate gene of AZFa sub-region," Journal of Reproduction & Infertility, , p. 298, 2017.

[20] M. Nailwal and J. B. Chauhan, "Computational analysis of high-risk SNPs in human DBY gene responsible for male infertility: a functional and structural impact," Interdisciplinary Sciences: Computational Life Sciences, 11(3), pp. 412-427, 2019.

[21] M. Akhtar, T. Jamal, H. Jamal, J. U. Din, M. A. M. Jamal and F. Jalil, "Identification of most damaging nsSNPs in human CCR6 gene: In silico analyses," International journal of immunogenetics, 46(6), pp. 459-471, 2019. 56

[22] F. Alzahrani, F. Ahmed, M. Sharma, M. Rehan, M. Mahfuz, M. Baeshen and Y. Hawsawi, "Investigating the pathogenic SNPs in BLM helicase and their biological consequences by computational approach," Scientific reports, pp. 1-22, 2020.

[23] M. Basheir, A. Bakri, H. Elnasri and M. Khaier, "COMPUTATIONAL ANALYSIS OF FUNCTIONAL SINGLE NUCLEOTIDE POLYMORPHISM OF HUMAN EUKARYOTICS TRANSLATION INITIATION FACTOR2 B1 (EIF2B1) GENE," pp. 1 10, 2020.

[24] M. Desai and J. B. Chauhan, "Predicting the functional and structural consequences of nsSNPs in human methionine synthase gene using computational tools," Systems Biology in Reproductive Medicine, pp. 288-300, 2019.

[25] S. Olatunji, K. Bowen and C. Huang, "Structural basis of the membrane intramolecular transacylase reaction responsible for lyso-form lipoprotein synthesis," Nature communications, pp. 1-14, 2021.