



# Comprehensive Analysis of rsSNPs Associated with Hypertension Using In-Silico Bioinformatics Tools

Alsadig Gassoum<sup>1,2</sup>, Nahla E. Abdelraheem<sup>1</sup>, Nehad Elsadig<sup>1</sup>

<sup>1</sup>National Center of Neurological Sciences, Khartoum, Sudan

<sup>2</sup>Nahda College, Khartoum, Sudan

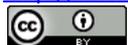
Email: sadig9@gmail.com

Received 20 June 2016; accepted 26 July 2016; published 29 July 2016

Copyright © 2016 by authors and OALib.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Genetic epidemiological studies have suggested that several genetic variants increase the risk for hypertension. It is likely that a number of genes rather than a single gene account for the heritability of this complex disorder. However, the genetic analysis of hypertension produced complex, inconsistent and nonreproducible results, which makes it difficult to draw conclusions about the association between specific genes and hypertension. Material and methods: In this study, we aimed to analyze SNPs that had been investigated in hypertension. These SNPs were collected from text-mind hypertension, obesity and diabetic (T-HOD) data base program, during the period of 31 may 2016. SNPs lists which were reported with hypertension were collected in excel file sheet and processed for analysis using different types of bioinformatics tools and programs. Results: SNPs were evaluated for their deleterious effect on the protein function and stability, in the present study, 7 SNPs were predicted deleterious (A288S, M731T, R172C, R50Q, G460W, K197N, G75V). Mutation3D server showed 3 of mutations (STEAA4, PLD2, AZIN2, rs28933400, rs2286672, rs16835244 genes and corresponding rsSNPs respectively) were found to increase risk to hypertension.

## Keywords

Hypertension, SNPs, In-Silico

Subject Areas: Bioinformatics

## 1. Introduction

Hypertension (elevated blood pressure levels exceeding 140/90 mmHg according to WHO criteria) is a common

**How to cite this paper:** Gassoum, A., Abdelraheem, N.E. and Elsadig, N. (2016) Comprehensive Analysis of rsSNPs Associated with Hypertension Using In-Silico Bioinformatics Tools. *Open Access Library Journal*, **3**: e2839.

<http://dx.doi.org/10.4236/oalib.1102839>

complex disorder, which affects 15% - 20% of adult population in Western societies [1]. It is classified as primary (essential) or secondary hypertension. The former type is used to describe hypertension without a known pathology.

### 1.1. Genetics of Hypertension

Genetic epidemiological studies have suggested that several genetic variants increase the risk for hypertension [2]. It is likely that a number of genes rather than a single gene account for the heritability of this complex disorder. However, the genetic analysis of hypertension produced complex, inconsistent and nonreproducible results, which makes it difficult to draw conclusions about the association between specific genes and hypertension [3].

### 1.2. T-HOD Data Base

Text-mined Hypertension, Obesity and Diabetes candidate gene database (T-HOD), employed the state-of-art text-mining technologies, including a gene identification (GI) system [4] [5], a disease term recognition system and the disease-gene relation extraction system—HypertenGene [6]. Because gene names vary a great deal, different genes may contain the same name. Moreover, gene names may be ambiguous and easily confused with terms employed in other research fields. The employed GI system was designed to alleviate the above problems, which was used to recognize gene terms and link them to their corresponding Entrez Gene IDs using a collective entity linking approach [6]. For extracting hypertension-related genes, we formulated the task as a binary classification problem in HypertenGene: for each recognized disease-gene pair from sentences in an abstract, determine whether it is a key relation. HypertenGene applies a maximum entropy model with a set of features, such as *n*-gram, chunk, parse tree and template features. We then rank all extracted genes according to their probability as calculated by the model. We extended and optimized the above systems to extract HOD genes in our T-HOD.

### 1.3. SNP

Single Nucleotide Polymorphism, causes the most common genetic mutation in human. Around 93% of human genes represent SNPs [7]. They are generating the majority of biological variations among individuals. SNPs can fall within the coding regions (coding SNPs) or non-coding regions of genes (noncoding SNPs), or in the intergenic region between two genes [8] [9]. While non-coding SNPs and intergenic may have a subtle impact [10], nonsynonymous coding SNPs have the major impact on individual by changing the protein sequence. The alterations that caused by these mutations on proteins sequences may alter protein function and structure [11]. Moreover these SNPs can affect the binding site for many transcriptional factors [10].

Normally, two different alleles, and also triallelic SNPs in which three different base variations may coexist within a population [11]. SNPs are been associated with many hereditary diseases like sickle cell anaemia, beta thalathemia and lung fibrosis [12]-[14]. Severity of disease or responds to treatment may be associated with existence of some types of SNPs [15]. Association studies can find relationship between different types of SNPs and diseases related to target populations [16]. The distribution of SNPs within human genome is not homogeneous, but the most SNPs are found within the non-coding region of the DNA, fewer are falling in the coding region, this is due to the natural selection mechanism adaptation [10] [15].

### 1.4. Biomedical Research

SNPs' has a greatest importance in biomedical research is for comparing regions of the genome between cohorts (such as with matched cohorts with and without a disease) in genome-wide association studies. SNPs have been used in genome-wide association studies as high-resolution markers in gene mapping related to diseases or normal traits. SNPs without an observable impact on the phenotype (so called silent mutations) are still useful as genetic markers in genome-wide association studies, because of their quantity and the stable inheritance over generations [17].

### 1.5. Disease

A single SNP may cause a Mendelian disease, though for complex diseases, SNPs do not usually function indi-

vidually, rather, they work in coordination with other SNPs to manifest a disease condition as has been seen in Osteoporosis [18]. All types of SNPs can have an observable phenotype or can result in disease.

SNPs in non-coding regions can manifest in a higher risk of cancer [19], and may affect mRNA structure and disease susceptibility [20].

### 1.6. SNPs in Coding Regions

Synonymous Substitutions by definition do not result in a change of amino acid in the protein, but still can affect its function in other ways. An example would be a seemingly silent mutation in the multidrug resistance gene 1 (MDR1), which codes for a cellular membrane pump that expels drugs from the cell, can slow down translation and allow the peptide chain to fold into an unusual conformation, causing the mutant pump to be less functional [21].

### 1.7. Non-Synonymous Substitutions

Missense-single change in the base results in change in amino acid of protein and its malfunction which leads to disease (e.g. c.1580G > T SNP in LMNA gene-position 1580 (nt) in the DNA sequence (CGT codon) causing the guanine to be replaced with the thymine, yielding CTT codon in the DNA sequence, results at the protein level in the replacement of the arginine by the leucine in the position 527 [22], at the phenotype level this manifests in overlapping mandibuloacral dysplasia and progeria syndrome)

Nonsense-point mutation in a sequence of DNA that results in a premature stop codon, or a *nonsense codon* in the transcribed mRNA, and in a truncated, incomplete, and usually nonfunctional protein product (e.g. Cystic fibrosis caused by the G542X mutation in the cystic fibrosis transmembrane conductance regulator gene) [23].

## 2. Material and Methods

### 2.1. Data Collection

In this study we aimed to analyze SNPs that had been investigated in hypertension. These SNPs were collected from text-mind hypertension, obesity and diabetic (T-HOD) data base program, during the period of 31 may 2016. The reported SNPs with hypertension were collected in excel file sheet and processed for analysis using different types of bioinformatics tools and programs.

### 2.2. SNPs Analysis

Functional effects of nsSNPs were predicted using different types of bioinformatics tools and programs, these program included SIFT (<http://sift.jcvi.org/>, <http://provean.jcvi.org/index.php>), PhD-SNP (<http://snps.biofold.org/phd-snp/phd-snp.html>), SNPs & GO (<http://snps-and-go.biocomp.unibo.it/snps-and-go/>), and MutPred (<http://mutpred.mutdb.org/>), furthermore polyphen was used to confirm PROVEAN results.

### 2.3. Prediction of Functional SNPs by SIFT and PROVEAN

SIFT is a sequence homology-based tool that sorts intolerant from tolerant amino acid substitutions and predicts whether an amino acid substitution in a protein will have a phenotypic effect. SIFT is based on the premise that protein evolution is correlated with protein function. Positions important for function should be conserved in an alignment of the protein family, whereas unimportant positions should appear diverse in an alignment [24].

Substitution of amino acid effects was predicted in protein function based on the conservation degree of the amino acid in the protein sequence, SIFT score of <0.05 is predicted by the algorithm to be damaged and >0.05 is considered to be tolerated [25].

### 2.4. PROVEAN (Protein Variation Effect Analyzer)

Is a soft ware that predict the amino acid substitution has any impact on the biological function of the protein, the assessment is based on PROVEAN score, where score of <-2.5 indicated that the protein variants is predicted have a deleterious effects, while the score of >-2.5 the variant is predicted to have a “neutral” effect. [26].

## 2.5. Detection of Deleterious nsSNPs by PANTHER

The PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System was designed to classify proteins (and their genes) in order to facilitate high throughput analysis. In this study the amino acid sequences were analyzed using PANTHER program to classify proteins.

## 2.6. PolyPhen-2 (<http://Genetics.bwh.harvard.edu/pph/data/>)

**PolyPhen-2 (Polymorphism Phenotyping v2)** is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations. In this study **PolyPhen-2** program was used to classify proteins into deleterious and benign.

## 2.7. SNP & GO

It is a support vector machine (SVM), based on the method to predict accurately where the mutation is related to disease from the protein sequence. Protein sequence was prepared in FASTA format and processed for analysis, the output results was obtained as neutral or disease related variation, the RI (reliability index ) with value  $> 5$  indicate disease related effect on function caused by the mutation on the protein [27].

## 2.8. Prediction of Harmful Mutations by MutPred

The MutPred server (<http://mutpred.mutdb.org/>), used to classify amino acid substitution (aas) as disease associated or neutral, also it predict disease/deleterious amino acids. The output of MutPred contains a general score (G), the probability that the amino acid substitution is deleterious/disease-associated and top 5 property scores (p).

## 2.9. Analysis of the Effects of nsSNPs on the Protein Stability by I-Mutant 2.0 and MUpro

I-Mutant 2.0 is a SVM based tools, support vector machine based tool that leads to automatic protein stability change prediction which is caused by single point mutation. Positive  $\Delta\Delta G$  value indicated that the mutated protein is of higher stability [28].

## 2.10. MUpro

Is a support vector machine-based tool for the prediction of protein stability changes upon non-synonymous SNPs. A score  $< 0$  means the variant decreases the protein stability; while, a score  $> 0$  means the variant increases the protein stability.

## 2.11. Prediction of the Stability Effects upon Mutation in Both Domain Cores and Domain-Domain Interfaces

ELASPIC is a novel ensemble machine learning approach that predicts the effects of mutations on protein folding and protein-protein interactions. The web server can be used to evaluate the effect of mutations on any protein in the Uniprot database, and allows all predicted results, including modeled wild-type and mutated structures, to be managed and viewed online and downloaded if needed.

## 2.12. Structural Analysis

The detection of nsSNPs Location in Protein Structure uses Mutation3D. Mutation3D (<http://mutation3d.org>) is a functional prediction tool for studying the spatial arrangement of amino acid substitutions on protein models and structures. This tool was used to analyse proteins structure for selected SNPs from hypertension data according to T-HOD data base.

## 2.13. Modeling Amino Acid Substitution, H-Bonding and Clash

UCSF Chimera is a highly extensible program for interactive visualization and analysis of molecular structures and related data. Chimera (version 1.8) software was used to scan the 3D (three-dimensional) structure of spe-

cific protein [29], Chimera (version 1.8) currently available within the Chimera package and available from the Chimera web site <http://www.cgl.ucsf.edu/chimera/>.

## 2.14. Modelling Amino Acid Sequence Using ModWeb

ModWeb: A Server for Protein Structure Modeling: was used to analyse and remodelling of protein sequences of Q687X5, P35611 and Q9UBU3-2 proteins.

## 2.15. M4T Server ver. 3.0

Comparative Modelling uses a combination of multiple templates and iterative optimization of alternative alignments.

## 2.16. Project HOPE

Project HOPE is an easy-to-use web-server that analyses the structural effects of mutation of interest. The server was used to analyse protein sequences in this study. Project HOPE collects and combine available information from a series of web-servers and databases and produced a mutation report complete with results, figures and animations. Where available Project HOPE will use the 3D structure of the protein but the server can also build a homology model if necessary. Other information sources include the Uniprot database and a series of DAS prediction servers [30].

## 3. Results

The T-HOD data-base server was used to retrieve hypertension SNPs, a total of 282 rsSNPs, were analyzed using **variants effects predictor**, results showed that intron-variants 30%, down\_stream\_gene\_variants 23%, non-coding transcript variants 15%, upstream gene variants 11%, missense variants 4%, regulatory region variants 3%, 3 prime UTR variants 3%, 2%. The coding consequences represent missense variants 69%, synonymous variants 28%, stop-lost 2%, and coding sequence variants 1%. Only missense non-synonymous coding SNPs were chosen for further analysis. nsSNPs, and mutations position were displayed in **Table 1**.

### 3.1. Prediction of Tolerated and Deleterious nsSNPs

The all SNPs were submitted to SIFT program to predict their effects on protein, out of 282 rsSNPs screened 27 rsSNPs were tolerated, 7 rsSNPs were damaging from which 4 were deleterious and 3 were neutral, SIFT couldn't find 248 SNPs (**Table 2**). From the 7 damaging SNPs 3 were reported with essential hypertension (rs28933400, rs4961, and rs1981529).

### 3.2. Deleterious nsSNPs by PROVEAN Server and PANTHER

The above mentioned damaging 7 rsSNPs, were submitted to provean server, 4 of them were deleterious while 3 were neutral. Significant correlation was found between SIFT and PROVEAN results, that the results of SIFT

**Table 1.** This table shows rsSNPs, protein ID, mutation and codons.

rsSNP	Protein ID	Amino acid change	Polymorphism
rs16835244	NP_443724	A288S	Gca/Tca
rs28933400	NP_000693	M731T	aTg/aCg
rs2286672	NP_002654	R172C	Cgt/Tgt
rs34911341	NP_001128413	R50Q	cGa/cAa
rs4961	NP_001110	G460W	Ggg/Tgg
rs5370	NP_001161791	K197N	aaG/aaT
rs1981529	NP_078912	G75V	gGc/gTc

**Table 2.** Prediction effects on protein using SIFT, Polyphen and Panther.

rsSNPs	SIFT	Score	PROVEAN	Score	Polyphen	Panther
rs16835244	damaging	0	deleterios	-2.792	benign	probably damaging
rs28933400	damaging	0	deleterios	-4.465	possibly damaging	probably damaging
rs2286672	damaging	0	deleterios	-2.511	benign	probably benign
rs34911341	damaging	0	deleterios	-2.593	probably damaging	probably damaging
rs4961	damaging	0.03	neutral	-2.31	probably damaging	probably benign
rs5370	damaging	0.01	neutral	-0.928	possibly damaging	probably benign
rs1981529	damaging	0.03	neutral	-1.937	benign	

showed 7 of the SNPs were damaging while PROVEAN detected 4 of the 7 SNPs were deleterious, SIFT and PROVEAN prediction may suggest protein disruption and function. Panther server was also used, out of the 7 SNPs, 3 were probably damaging and the rest were benign (**Table 2**).

### 3.3. Damaging nsSNPs Found by SNPs & GO, PHD-SNP

SNPs & GO results showed that out of 7 SNPs 3 were predicted to have disease causing ability while the rest were neutral by PHD-SNP, by SNP & GO 2 of the SNPs were predicted to have causing disease ability (**Table 3**).

### 3.4. Identification of Functional nsSNP

Changes in protein stability were examined by I-mutant 2 and MUpro software programs. The results of I-mutant 2 showed that (A288S, M731T, R171C, R50Q, G460W, K197N, G75V) were predicted decreasing of the free energy of proteins except G460W was predicted to increase of the free energy of protein. MUpro results predict increase stability of protein in all of the variants (**Table 4**).

### 3.5. Prediction of Functional Effects of nsSNP Using MutPred

MutPred analysis was done to determine the degree of tolerance for each amino acid substitution on the basis of physio-chemical properties. **Table 5** shows the results of MutPred.

### 3.6. Prediction of Stability Effects upon Mutation in Both Domain Cores and Domain-Domain Interfaces by ELASPIC

SNPs were classified according to their structural location, into core or interface, in the present study 3 variants were core structural location while the rest were not classified, detailed results of ELASPIC was displayed in **Table 6**.

### 3.7. Distributions of nsSNPs by Mutation3D Server

Results of Mutation3D indicated that 3 of mutations (**STEA4**, **PLD2**, **AZIN2**, **rs28933400**, **rs2286672**, **rs16835244** genes and corresponding **rsSNPs** respectively) were found to be with a high risk to hypertension, they located in the protein domain, detailed results were displayed in **Figures 1-6**.

### 3.8. Homology Modeling of New and Wild Amino Acids of Deleterious nsSNPs

3D of protein structure is very important to verify the deleterious mutations and possible effects on the structure and function of protein, in this study 4 proteins were modelled by Chimera UCSF program 1.8, and H bonding inter-actions and clashes were calculated using Chimera 1.8 program. Modeller server [31]-[34] was used to create 3D structure protein for 3 proteins (**Figures 7-34**).

## 4. Discussion

In the present study we aimed to investigate SNPs which were reported with hypertension, and as we mentioned

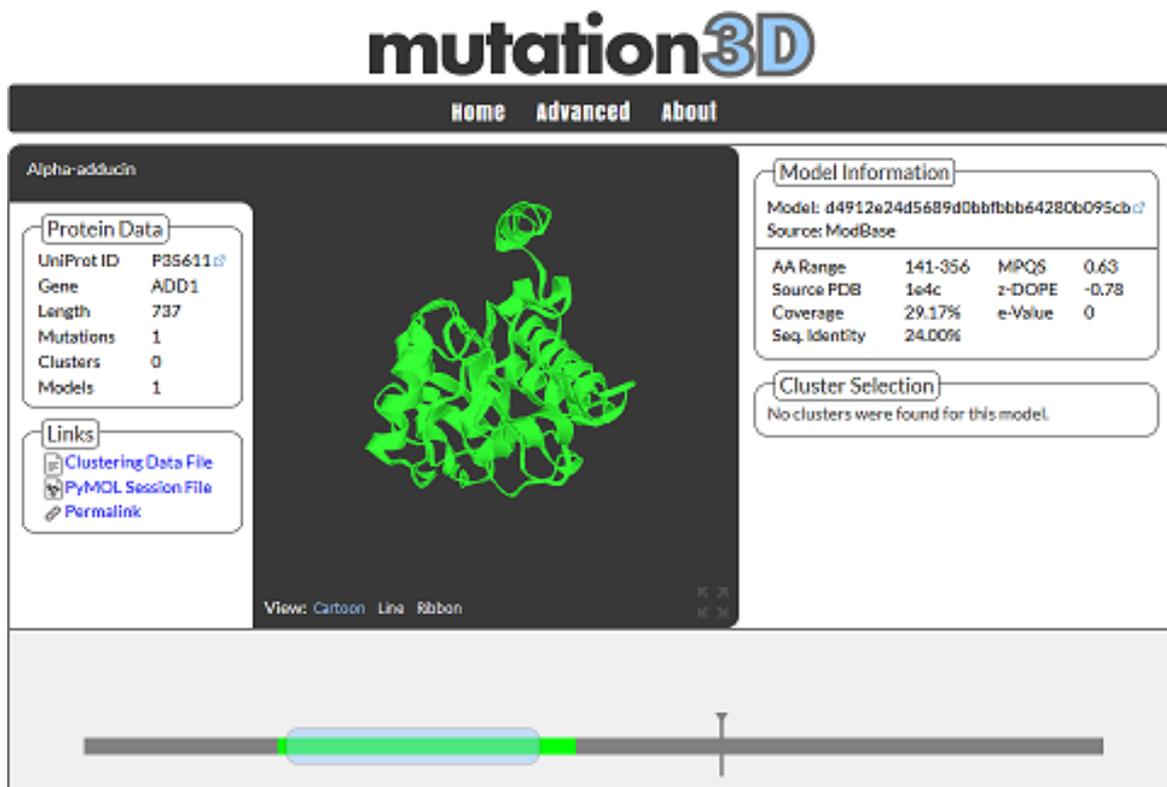


Figure 1. rs4961 (G460W) SNP structure: mutation outside the core of protein.



Figure 2. rs5370 (K197N) SNP structure: mutation not within the core of protein.



Figure 3. rs1981529 (G75V) SNP structure: mutation in the core of protein.



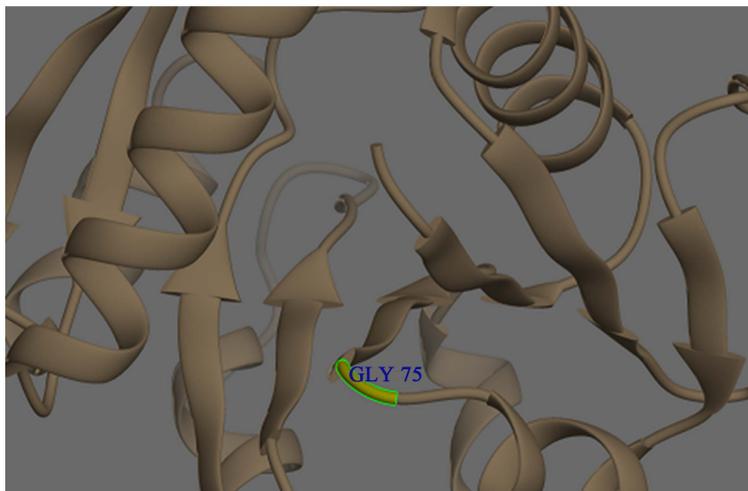
Figure 4. rs2286672 (R172C) SNP structure: mutation in the core of protein.



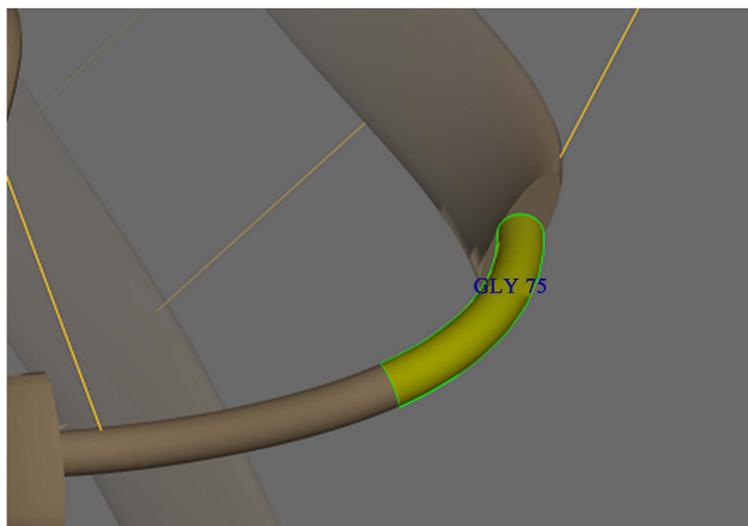
Figure 5. rs16835244 (A288S) SNP structure: mutation in the core of protein.



Figure 6. rs28933400 (M731T) SNP structure: mutation not within the core of protein.



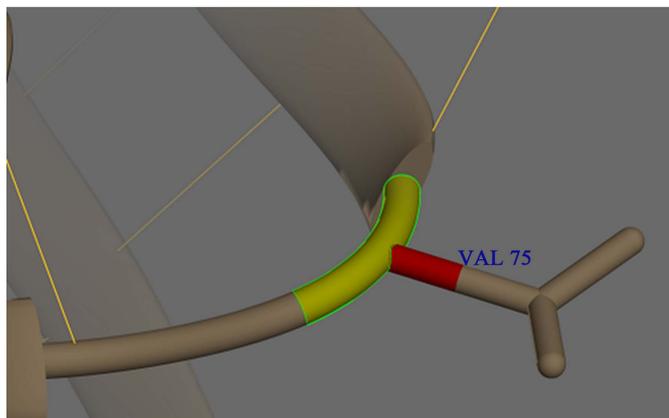
**Figure 7.** rs1981529 (G75V) SNP structure: wild type in green (ribbon yellow).



**Figure 8.** rs1981529 (G75V) SNP structure: focused on wild type residue which present in green colour (ribbon yellow).



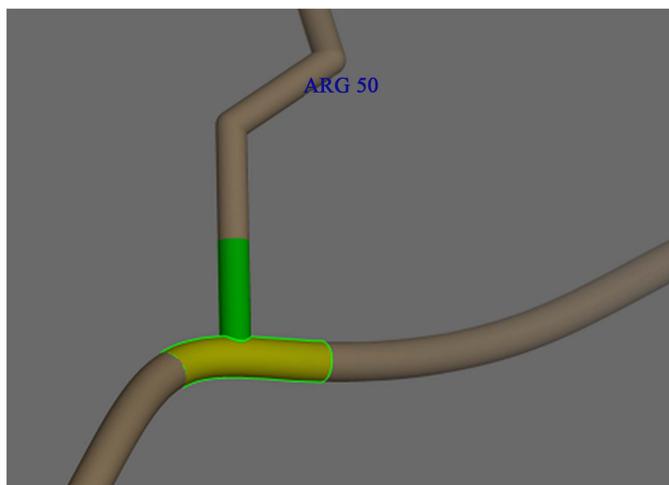
**Figure 9.** rs1981529 (G75V) SNP structure: mutant type in red (ribbon yellow).



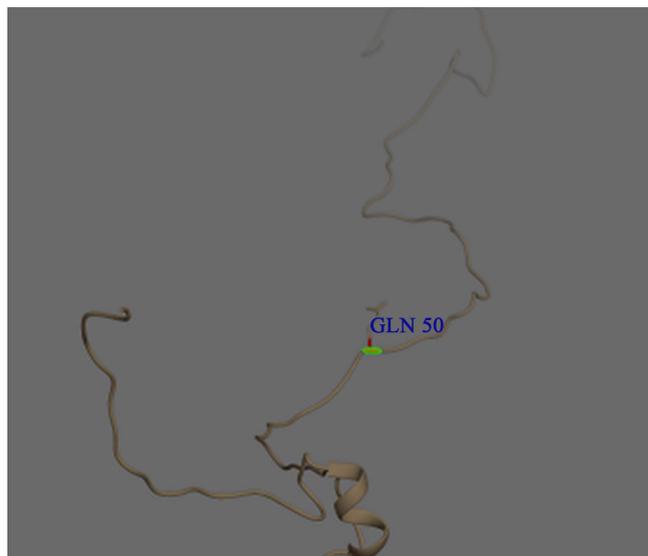
**Figure 10.** rs1981529 (G75V) SNP structure: showed magnified view of mutant type in red (ribbon yellow).



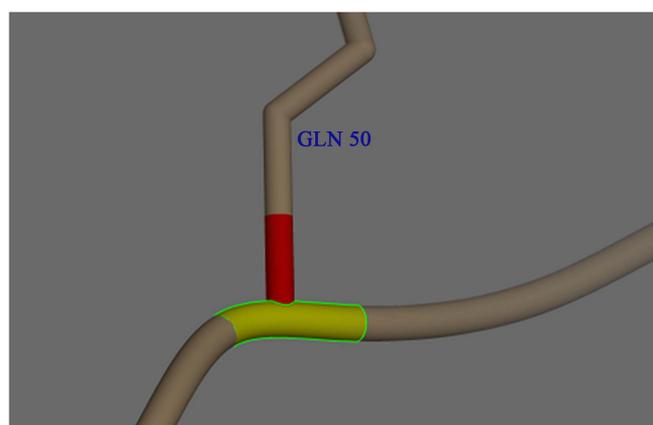
**Figure 11.** rs34911341 (A50Q) SNP structure: wild type green.



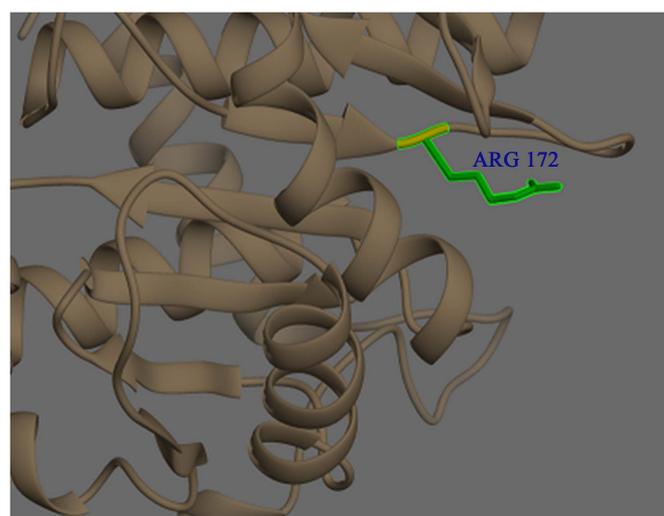
**Figure 12.** rs34911341 (A50Q) SNP structure: this figure focused on the wild type Arg residue in green.



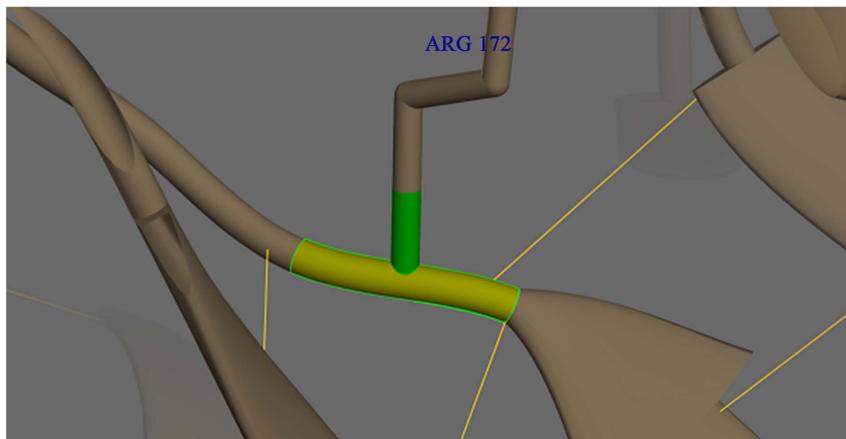
**Figure 13.** rs34911341 (A50Q) SNP structure: mutant red.



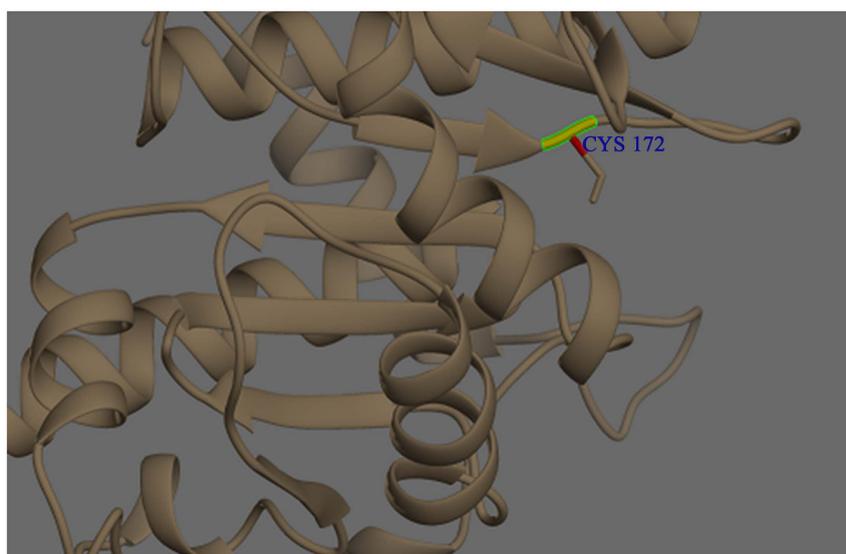
**Figure 14.** rs34911341 (A50Q) SNP structure: mutant residue was magnified in red.



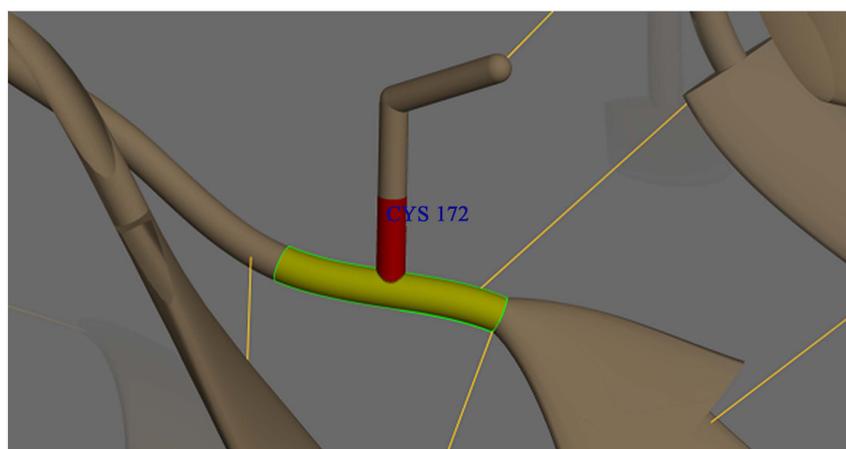
**Figure 15.** rs2286672 (A172C) SNP structure: wild type green colour.



**Figure 16.** rs2286672 (A172C) SNP structure: wild type green colour, 2Hbonding interactions were observed.



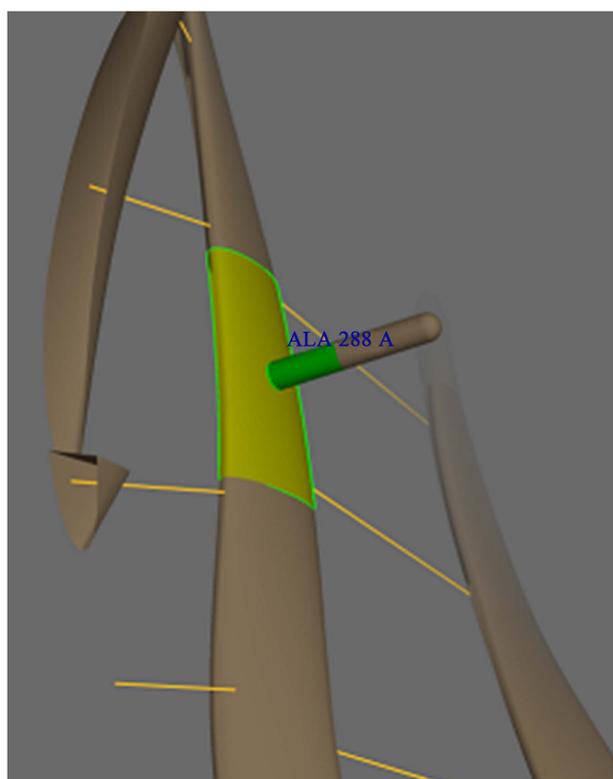
**Figure 17.** rs2286672 (A172C) SNP structure: mutant type red colour.



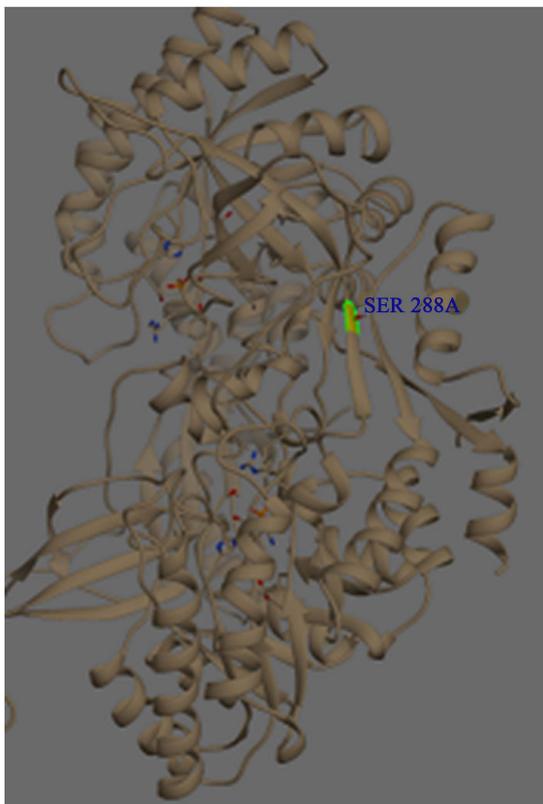
**Figure 18.** rs2286672, shows mutant type in red colour. In this figure 2H bonding interaction in mutant residue.



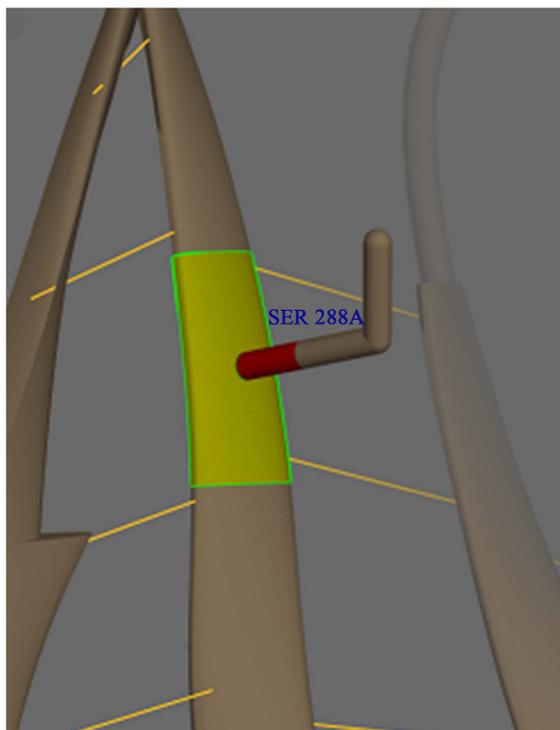
**Figure 19.** rs16835244 (A288S) SNP structure: wild type green.



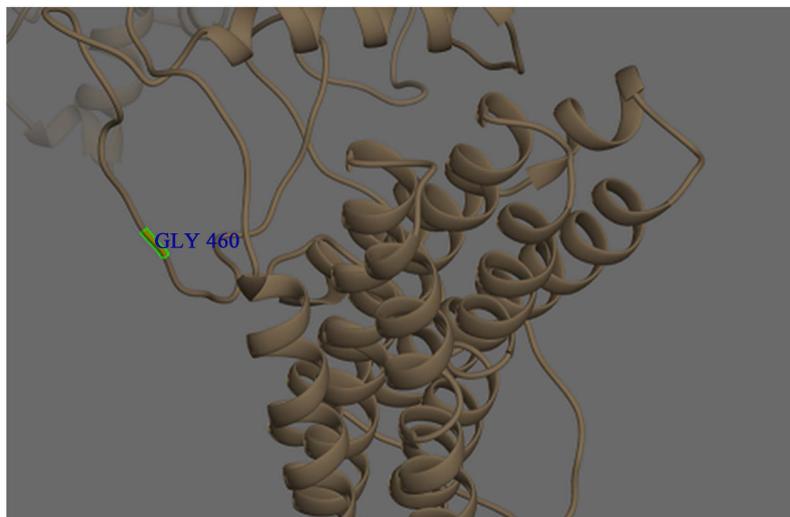
**Figure 20.** rs16835244 (A288S) SNP structure: the wild type was present in green colour, and 2 H-bonding interactions were observed.



**Figure 21.** rs16835244 (A288S) SNP structure: the mutant type SER was indicated in red colour.



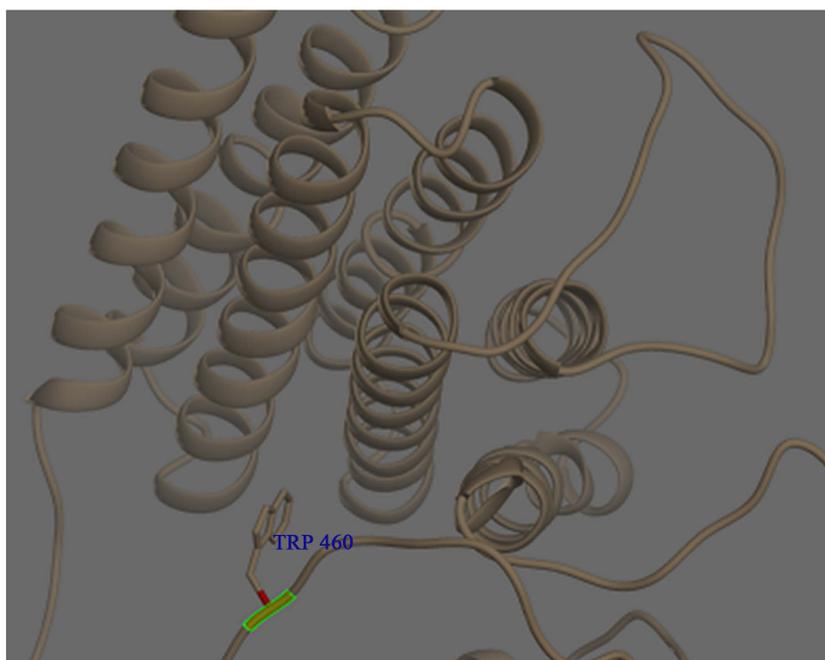
**Figure 22.** rs16835244 (A288S) SNP structure: magnified mutant SER residue was shown in red with 2 H-bonding interactions.



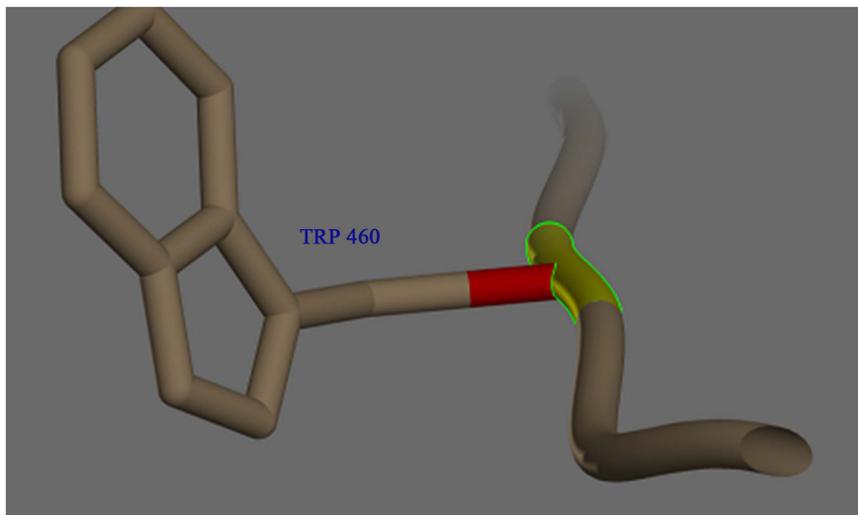
**Figure 23.** rs4961 (G460T) SNP structure: wild type green (ribbon yellow).



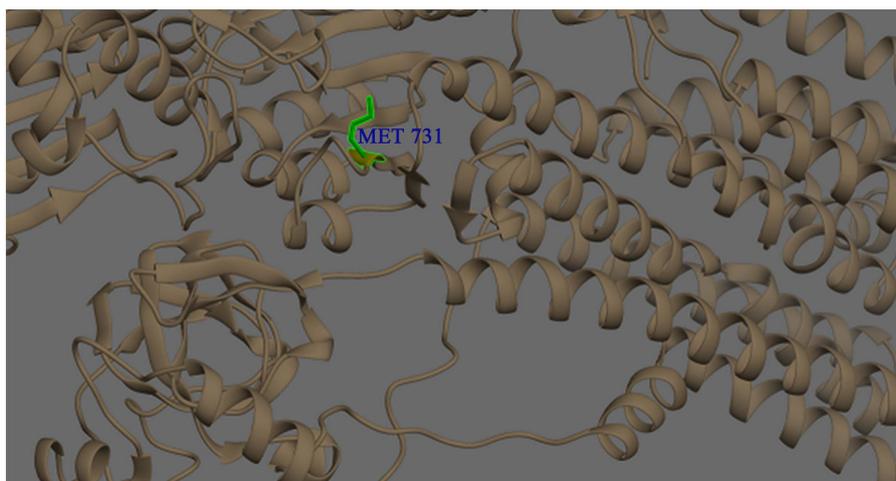
**Figure 24.** rs4961 (G460T) SNP structure: wild type green (ribbon yellow).



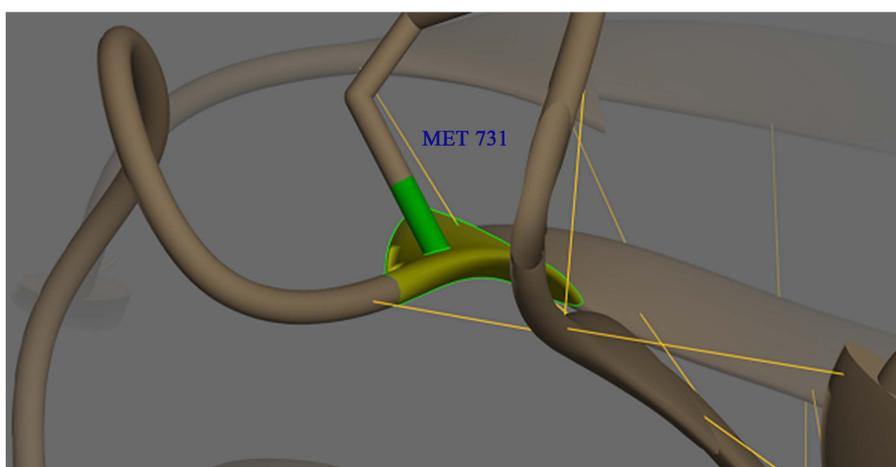
**Figure 25.** rs4961 (G460T) SNP structure: TRP mutant residue was present in red (ribbon yellow).



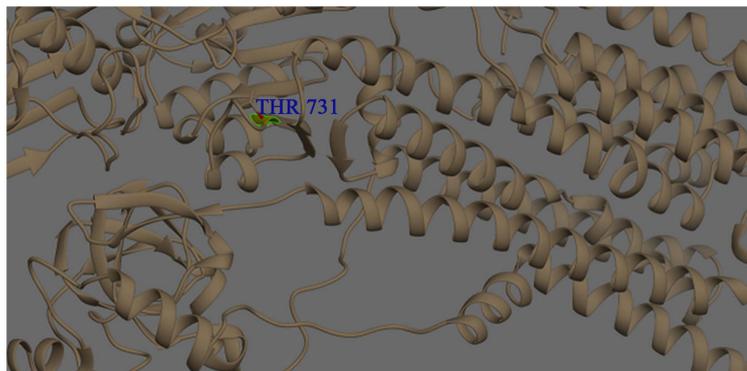
**Figure 26.** rs4961 (G460T) SNP structure: magnification of the mutated residue was shown in red (ribbon yellow).



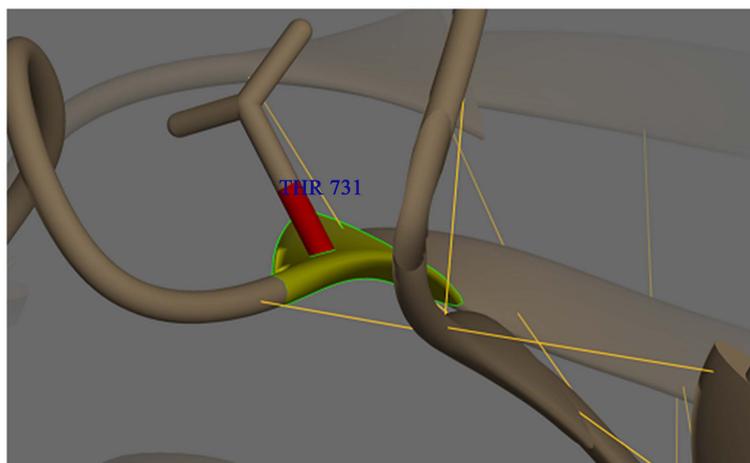
**Figure 27.** rs28933400 (M731T) SNP structure: wild type green ribbon yellow.



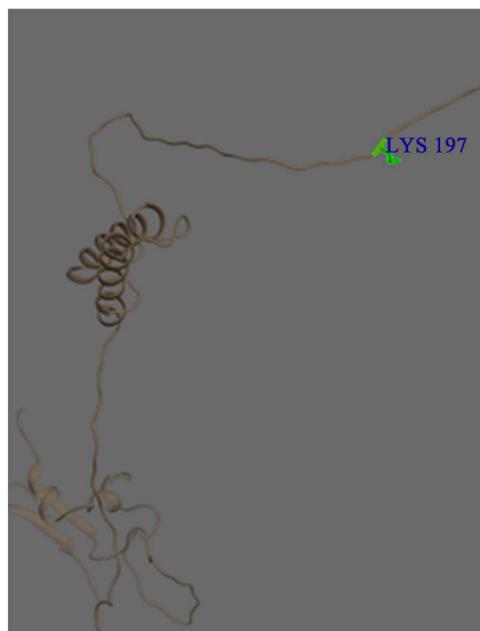
**Figure 28.** rs28933400 (M731T) SNP structure: wild type green and ribbon yellow, 2 H-bonding interaction.



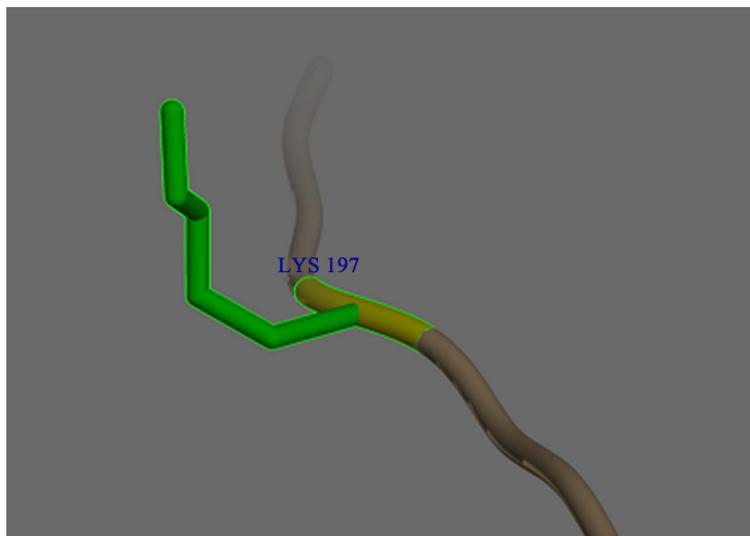
**Figure 29.** rs28933400 (M731T) SNP structure: mutant type red, ribbon yellow.



**Figure 30.** rs28933400 (M731T) SNP structure: mutant type shown in red colour and ribbon in yellow, 2 H-bonding interaction was observed.



**Figure 31.** rs5370 (K197N) SNP structure: wild type was present in green colour.



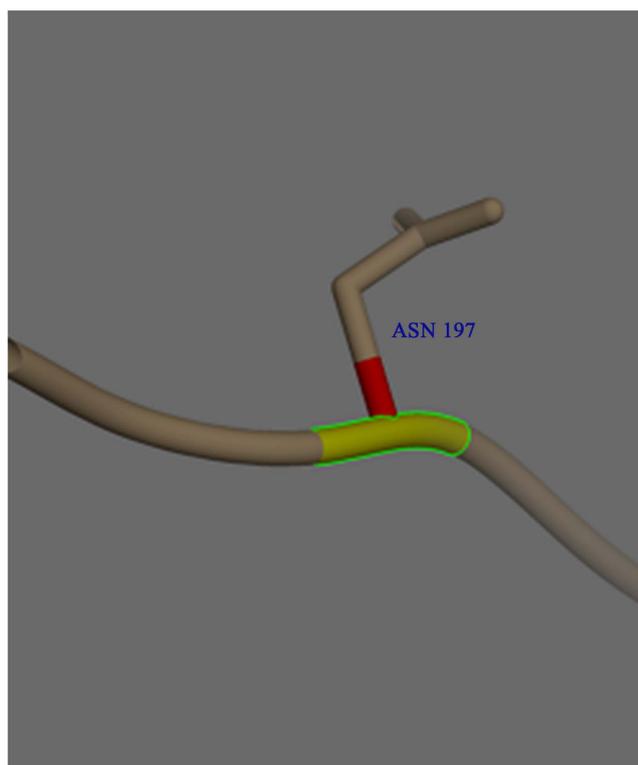
**Figure 32.** rs5370 (K197N) SNP structure: focused on the wild type which present in green colour.



**Figure 33.** rs5370 (K197N) SNP structure: mutant type red.

**Table 3.** Prediction of disease related mutation by PHD-SNP and SNP & GO.

rsSNPs	PHD_SNP	RI	SNP & GO	RI
rs16835244	disease	5	disease	0
rs28933400	disease	3	disease	10
rs2286672	disease	3	neutral	1
rs34911341	neutral	1	unclassified	
rs4961	disease	0	neutral	4
rs5370	neutral	3	unclassified	
rs1981529	disease	2	neutral	3



**Figure 34.** rs5370 (K197N) SNP structure: magnified mutant residue in red colour.

**Table 4.** This table shows protein stability prediction by I-Mutant and MuPro.

rsSNPs	I-Mutant SVM2	DDG value	MuPro	Confidence score
rs16835244	decrease	<b>-1.24</b>	increase stability	-0.805949883
rs28933400	decrease	<b>-0.43</b>	increase stability	-0.793908856
rs2286672	decrease	<b>-1.18</b>	increase stability	0.613637436
rs34911341	decrease	<b>-0.6</b>	increase stability	0.707156682
rs4961	increase	<b>0.35</b>	increase stability	-0.791147359
rs5370	decrease	<b>-2.27</b>	increase stability	-0.928539952
rs1981529	decrease	<b>-1.54</b>	increase stability	0.955986275

**Table 5.** Summarizes MutPred results.

Mutation	Prob. deleterious	Loss of sheet	Gain of helix	Loss of loop	Glycos. S283	Gain of MoRF binding
A288S	0.253	( $P = 0.0228$ )	( $P = 0.0893$ )	( $P = 0.2897$ )	( $P = 0.4302$ )	( $P = 0.4656$ )
M731T	0.937	( $P = 0.1358$ )	( $P = 0.1466$ )	( $P = 0.1872$ )	( $P = 0.2205$ )	( $P = 0.2897$ )
R50Q	0.581	( $P = 0.0115$ )	( $P = 0.0921$ )	( $P = 0.132$ )	( $P = 0.1688$ )	( $P = 0.1894$ )
K197N	0.075	( $P = 0.02$ )	( $P = 0.0668$ )	( $P = 0.0997$ )	( $P = 0.1299$ )	( $P = 0.1579$ )
G460W	0.078	( $P = 0.0549$ )	( $P = 0.1312$ )	( $P = 0.1551$ )	( $P = 0.1736$ )	( $P = 0.1907$ )
M731T	0.937	( $P = 0.1358$ )	( $P = 0.1466$ )	( $P = 0.1872$ )	( $P = 0.2205$ )	( $P = 0.2897$ )
G75V	0.388	( $P = 0.0359$ )	( $P = 0.0477$ )	( $P = 0.0556$ )	( $P = 0.0609$ )	( $P = 0.1131$ )

**Table 6.** Predictions of mutation effects in protein function by using ELASPIC.

rsSNPs	ELASPIC	$\Delta G_{wt}$	$\Delta G_{mut}$	$\Delta\Delta G$
rs16835244	core	195.996	196.748	-0.96341
rs28933400	core	327.034	331.157	0.575316
rs2286672	core	276.518	276.932	0.34204
rs34911341	unclassified	unclassified	unclassified	unclassified
rs4961	unclassified	unclassified	unclassified	unclassified
rs5370	unclassified	unclassified	unclassified	unclassified
rs1981529	core	14.7475	13.7847	-0.87053

before these SNPs were retrieved from T-HOD database web site. The methods used to assess nsSNPs (mutations) in this study were based on different types of bioinformatics tools, describing pathogenicity and providing some clue on molecular level about the effect of mutations. It is very difficult to use one method or bioinformatics tool to predict pathogenic effect of SNPs, so in the present study we used 12 different in silico prediction algorithms (SIFT, PROVEAN, PHD-SNP, Panther, MUpro, MutPred, I-Mutation2, polyphen, SNP & GO, Project-Hope, Chimera and modeller to sort tolerant and diseased SNPs.

The findings of this study showed that 7 SNPs were damaged by using SIFT (A288S, M731T, R172C, R50Q, G460W, K197N, G75V) and 4 (A288S, M731T, R172C, R50Q) out of the seven SNPs were deleterious by PROVEAN, while 4 SNPs were found to be disease caused by PHD-SNP (A288S, R172C) and 2 SNPs (A288S, M731T, R172C, G75V and G460W) by using SNPS & GO. Polyphen results showed that 4 SNPs (M731T, R50Q, G460W, K197N) were probably and possibly damaging, moreover Panther results indicated A288S, M731T and G460W were probably damaging. I-Mutant Suite results showed that 6 mutations were decreasing protein stability (A288S, M731T, R172C, R50Q, K197N, G75V) while G460W showed increased stability of protein. By comparing output of the 6 above mentioned in-silico bioinformatics tools, A288S, M731T, R172C, G75V, G460W, R50Q and K197N mutations were found functionally significant. Using MutPred to determine the degree of tolerance of each amino acid substitution on the bases of physico-chemical properties, results of this study showed that, A288S, R50Q, K197N and G460W were harmful with loss of sheet  $P = 0.0228, 0.0115, 0.02$  and  $0.0549$  respectively. Furthermore, these 7 SNPs were analysed by structurally and functionally by using 5 bioinformatics tools; Chimera, Mutation 3D, PDB, modeller and ELASPIC. In the present study the “core” residues were found predominant within the mutations, this residues are defined as residues which are exposed in the monomeric protein but buried in the protein complex. Core residues are typically hydrophobic with a composition strongly divergent from the composition of the remainder of the protein surface [35] Core residues supply the bulk of the energy driving association by hydrophobic interactions [35] The hydrophobic interactions within the complex cause the core region to become tightly packed upon complex association with little room for conformational variability. For these reasons, the core residues are strongly conserved during evolution [36] and mutations in this region are usually more strongly unfavorable when compared to mutations at the periphery of the interface. Results of Mutation3D server showed 3 of mutations (STEA4, PLD2, AZIN2, **rs28933400**, **rs2286672**, **rs16835244** genes and corresponding rsSNPs respectively) were found to be with a high risk to hypertension. Hydrogen bonding and clashes of the mutations A288S, M731T, and R172C showed different numbers of hydrogen bonding between mutant residue and wild type, the differences of H-bonding between the wild and mutant residues may indicated a significant effect on protein stability, these results were obtained by using Chimera program 1.8.

#### **rs28933400**

The mutant residue is smaller than the wild-type residue. The wild-type residue is more hydrophobic than the mutant residue. The mutated residue is located in a domain that is important for binding of other molecules. The mutated residue is in contact with residues in another domain. It is possible that the mutation disturbs these contacts. 3D of protein of this mutation showed that the mutation was located in the core of protein, and this may increase the risk of hypertension. Moreover this mutation showed differences in H-bonding between the wild type and mutant type residues, and these differences may affect protein stability.

**rs2286672**

The mutant residue is smaller than the wild-type residue. The wild-type residue was positively charged, the mutant residue is neutral. The mutant residue is more hydrophobic than the wild-type residue. The mutation is located within a domain. The mutation introduces an amino acid with different properties, which can disturb this domain and abolish its function. There is a difference in charge between the wild-type and mutant amino acid.

The charge of the wild-type residue will be lost, and this can cause loss of interactions with other molecules or residues. The wild-type and mutant amino acids differ in size. The mutant residue is smaller, and this might lead to loss of interactions. The hydrophobicity of the wild-type and mutant residue differs. The mutation introduces a more hydrophobic residue at this position. This can result in loss of hydrogen bonds and/or disturb correct folding. 3D of protein of this mutation showed that the mutation was located in the core of protein, and this may increase the risk of hypertension. Moreover this mutation showed differences in H-bonding between the wild type and mutant type residues, and these differences may affect protein stability.

**rs16835244**

The wild-type and mutant amino acids differ in size. The mutant residue is bigger than the wild-type residue. The wild-type residue was buried in the core of the protein. The mutant residue is bigger and probably will not fit. The hydrophobicity of the wild-type and mutant residue differs. The mutation will cause loss of hydrophobic interactions in the core of the protein. 3D of protein of this mutation showed that the mutation was located in the core of protein, and this may increase the risk of hypertension. Moreover this mutation showed differences in H-bonding between the wild type and mutant type residues, and these differences may affect protein stability.

**rs34911341**

There is a difference in charge between the wild-type and mutant amino acid. The charge of the wild-type residue will be lost, and this can cause loss of interactions with other molecules or residues. The wild-type and mutant amino acids differ in size. The mutant residue is smaller, and this might lead to loss of interactions.

**rs4961**

The wild-type and mutant amino acids differ in size. The mutant residue is bigger, this might lead to bumps. The torsion angles for this residue are unusual. Only Glycine is flexible enough to make these torsion angles, mutation into another residue will force the local backbone into an incorrect conformation and will disturb the local structure.

**rs5370**

The mutation is located within the signal peptide. This sequence of this peptide is important because it is recognized by other proteins and often cleaved of to generate the mature protein.

The new residue that is introduced in the signal peptide differs in its properties from the original one. It is possible that this mutation disturbs recognition of the signal peptide.

There is a difference in charge between the wild-type and mutant amino acid. The charge of the wild-type residue will be lost, and this can cause loss of interactions with other molecules or residues. The wild-type and mutant amino acids differ in size. The mutant residue is smaller, and this might lead to loss of interactions.

**rs1981529**

The wild-type and mutant amino acids differ in size; in addition to that the mutant residue is bigger than the wild-type residue. Moreover the mutation is located on the surface of the protein; mutation of this residue can disturb interactions with other molecules or other parts of the protein. The torsion angles for this residue are unusual. Only glycine is flexible enough to make these torsion angles, mutation into another residue will force the local backbone into an incorrect conformation and will disturb the local structure.

## 5. Conclusion

The available hypertension rsSNPs from T-HOD data base were retrieved, and then analyzed using different types of bioinformatics tools, and the predicted deleterious SNPs were evaluated for their deleterious effect on the protein function and stability. In the present study, 7 SNPs were predicted deleterious (A288S, M731T, R172C, R50Q, G460W, K197N, G75V). Mutation3D server showed that 3 of mutations (STEA4, PLD2, AZIN2, rs28933400, rs2286672, rs16835244 genes and corresponding rsSNPs respectively) were found to increase risk to hypertension.

## References

- [1] Lifton, R.P. (1996) Molecular Genetics of Human Blood Pressure Variation. *Science*, **272**, 676-680.

- <http://dx.doi.org/10.1126/science.272.5262.676>
- [2] Izawa, H., Yamada, Y., Okada, T., Tanaka, M., Hirayama, H. and Yokota, M. (2003) Prediction of Genetic Risk for Hypertension. *Hypertension*, **41**, 1035-1040. <http://dx.doi.org/10.1161/01.HYP.0000065618.56368.24>
  - [3] Williams, S.M., et al. (2000) Combinations of Variations in Multiple Genes Are Associated with Hypertension. *Hypertension*, **36**, 2-6. <http://dx.doi.org/10.1161/01.HYP.36.1.2>
  - [4] Dai, H.-J., Chang, Y.-C., Tsai, R.T.-H. and Hsu, W.-L. (2011) Integration of Gene Normalization Stages and Co-Reference Resolution Using a Markov Logic Network. *Bioinformatics*, **27**, 2586-2594. <http://dx.doi.org/10.1093/bioinformatics/btr358>
  - [5] Dai, H.-J., Lai, P.-T. and Tsai, R.T.-H. (2010) Multistage Gene Normalization and SVM-Based Ranking for Protein Interactor Extraction in Full-Text Articles. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **7**, 412-420.
  - [6] Tsai, R.T.-H., Lai, P.-T., Dai, H.-J., et al. (2009) HypertenGene: Extracting Key Hypertension Genes from Biomedical Literature with Position and Automatically-Generated Template Features. *BMC Bioinformatics*, **10**, S9. <http://dx.doi.org/10.1186/1471-2105-10-S15-S9>
  - [7] Single Nucleotide Polymorphism/SNP (2015) Learn Science at Scitable. <http://www.nature.com>
  - [8] Nachman, M.W. (2001) Single Nucleotide Polymorphisms and Recombination Rate in Humans. *Trends in Genetics*, **17**, 481-485. [http://dx.doi.org/10.1016/S0168-9525\(01\)02409-X](http://dx.doi.org/10.1016/S0168-9525(01)02409-X)
  - [9] Varela, M.A. and Amos, W. (2010) Heterogeneous Distribution of SNPs in the Human Genome: Microsatellites as Predictors of Nucleotide Diversity and Divergence. *Genomics*, **95**, 151-159. <http://dx.doi.org/10.1016/j.ygeno.2009.12.003>
  - [10] Barreiro, L.B., Laval, G., Quach, H., Patin, E. and Quintana-Murci, L. (2008) Natural Selection has Driven Population Differentiation in Modern Humans. *Nature Genetics*, **40**, 340-345. <http://dx.doi.org/10.1038/ng.78>
  - [11] Hodgkinson, A. and Eyre-Walker, A. (2009) Human Triallelic Sites: Evidence for a New Mutational Mechanism? *Genetics*, **1**.
  - [12] Ingram, V.M. (1956) A Specific Chemical Difference between the Globins of Normal Human and Sickle-Cell Anaemia Haemoglobin. *Nature*, **178**, 792-794. <http://dx.doi.org/10.1038/178792a0>
  - [13] Chang, J.C. and Kan, Y.W. (1979) Beta 0 Thalassemia, a Nonsense Mutation in Man. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 2886-2889. <http://dx.doi.org/10.1073/pnas.76.6.2886>
  - [14] Hamosh, A., King, T.M., Rosenstein, B.J., Corey, M., Levison, H., Durie, P., Tsui, L.C., McIntosh, I., Keston, M., Brock, D.J., Macek, M., Zemková, D., Krásničanová, H., Vávrová, V., Macek, M., Golder, N., Schwarz, M.J., Super, M., Watson, E.K., Williams, C., Bush, A., O'Mahoney, S.M., Humphries, P., Dearce, M.A., Reis, A., Bürger, J., Stuhmann, M., Schmidtke, J., Wulbrand, U. and Dörk, T. (1992) Cystic Fibrosis Patients Bearing Both the Common Missense Mutation Gly-Asp at Codon 551 and the Delta F508 Mutation Are Clinically Indistinguishable from Delta F508 Homozygotes, Except for Decreased Risk of Meconiumileus. *American Journal of Human Genetics*, **51**, 245-250.
  - [15] Wolf, A.B., Caselli, R.J., Reiman, E.M. and Valla, J. (2012) APOE and Neuroenergetics: An Emerging Paradigm in Alzheimer's Disease. *Neurobiology of Aging*, **34**, 1007-1017. <http://dx.doi.org/10.1016/j.neurobiolaging.2012.10.011>
  - [16] Abecasis, G.R. and Cookson, W.O. (2000) GOLD—Graphical Overview of Linkage Disequilibrium. *Bioinformatics*, **16**, 182-183. <http://dx.doi.org/10.1093/bioinformatics/16.2.182>
  - [17] Thomas, P.E., Klinger, R., Furlong, L.I., Hofmann-Apitius, M. and Friedrich, C.M. (2011) Challenges in the Association of Human Single Nucleotide Polymorphism Mentions with Unique Database Identifiers. *BMC Bioinformatics*, **12**, S4. <http://dx.doi.org/10.1186/1471-2105-12-S4-S4>
  - [18] Singh, M., Singh, P., Juneja, P.K., Singh, S. and Kaur, T. (2010) SNP-SNP Interactions within APOE Gene Influence Plasma Lipids in Postmenopausal Osteoporosis. *Rheumatology International*, **31**, 421-423. <http://dx.doi.org/10.1007/s00296-010-1449-7>
  - [19] Li, G., Pan, T., Guo, D. and Li, L.C. (2014) Regulatory Variants and Disease: The E-Cadherin-160C/A SNP as an Example. *Molecular Biology International*, **2014**, Article ID: 967565. <http://dx.doi.org/10.1155/2014/967565>
  - [20] Lu, Y.-F., Mauger, D.M., Goldstein, D.B., Urban, T.J., Weeks, K.M. and Bradrick, S.S. (2015) IFNL3 mRNA Structure Is Remodeled by a Functional Non-Coding Polymorphism Associated with Hepatitis C Virus Clearance. *Scientific Reports*, **5**, Article No. 16037. <http://dx.doi.org/10.1038/srep16037>
  - [21] Kimchi-Sarfaty, C., Oh, J.M., Kim, I.W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V. and Gottesman, M.M. (2007) A "Silent" Polymorphism in the *MDR1* Gene Changes Substrate Specificity. *Science*, **315**, 525-528. <http://dx.doi.org/10.1126/science.1135308>
  - [22] Al-Haggag, M., Madej-Pilarczyk, A., Kozłowski, L., Bujnicki, J.M., Yahia, S., Abdel-Hadi, D., Shams, A., Ahmad, N.,

- Hamed, S. and Puzianowska-Kuznicka, M. (2012) A Novel Homozygous p.Arg527Leu *LMNA* Mutation in Two Unrelated Egyptian Families Causes Overlapping Mandibuloacral Dysplasia and Progeria Syndrome. *European Journal of Human Genetics*, **20**, 1134-1140. <http://dx.doi.org/10.1038/ejhg.2012.77>
- [23] Cordovado, S.K., Hendrix, M., Greene, C.N., Mochal, S., Earley, M.C., Farrell, P.M., Kharrazi, M., Hannon, W.H. and Mueller, P.W. (2012) CFTR Mutation Analysis and Haplotype Associations in CF Patients. *Molecular Genetics and Metabolism*, **105**, 249-254.
- [24] Talavera, D., Robertson, D.L. and Lovell, S.C. (2011) Characterization of Protein-Protein Interaction Interfaces from a Single Species. *PLoS ONE*, **6**, e21053. <http://dx.doi.org/10.1371/journal.pone.0021053>
- [25] Andreani, J., Faure, G. and Guerois, R. (2012) Versatility and in Variance in the Evolution of Homologous Heteromeric Interfaces. *PLoS Computational Biology*, **8**, e1002677. <http://dx.doi.org/10.1371/journal.pcbi.1002677>
- [26] Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P.L. and Casadio, R. (2009) Functional Annotations Improve the Predictive Score of Human Disease-Related Mutations in Proteins. *Human Mutation*, **30**, 1237-1244. <http://dx.doi.org/10.1002/humu.21047>
- [27] Capriotti, E., Fariselli, P. and Casadio, R. (2005) I-Mutant2.0: Predicting Stability Changes upon Mutation from the Protein Sequence or Structure. *Nucleic Acids Research*, **33**, 306-310. <http://dx.doi.org/10.1093/nar/gki375>
- [28] Sofia, B.M. and Mohamed, M.H. (2016) Insilico Validation of Babesia Bovis Merozoite Surface Antigen-1, Merozoite Surface Antigen-2b and Merozoite Surface Antigen-2c Proteins for Vaccine and Drug Development. *International Journal of Bioinformatics and Biomedical Engineering*, **2**, 30-39.
- [29] Guharoy, M. and Chakrabarti, P. (2005) Conservation and Relative Importance of Residues across Protein-Protein Interfaces. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 15447-15452.
- [30] Venselaar, H., te Beek, T.A.H., Kuipers, R.K.P., Hekkelman, M.L. and Vriend, G. (2010) Protein Structure Analysis of Mutations Causing Inheritable Diseases. An e-Science Approach with Life Scientist Friendly Interfaces. *BMC Bioinformatics*, **11**, 548. <http://dx.doi.org/10.1186/1471-2105-11-548>
- [31] Källberg, M., Wang, H.P., Wang, S., Peng, J., Wang, Z.Y., Lu, H. and Xu, J.B. (2012) Template-Based Protein Structure Modeling Using the RaptorX Web Server. *Nature Protocols*, **7**, 1511-1522. <http://dx.doi.org/10.1038/nprot.2012.085>
- [32] Ma, J.Z., Wang, S., Zhao, F. and Xu, J.B. (2013) Protein Threading Using Context-Specific Alignment Potential. *Bioinformatics*, **29**, i257-i265. <http://dx.doi.org/10.1093/bioinformatics/btt210>
- [33] Peng, J. and Xu, J.B. (2011) A Multiple-Template Approach to Protein Threading. *Proteins*, **79**, 1930-1939. <http://dx.doi.org/10.1002/prot.23016>
- [34] Peng, J. and Xu, J.B. (2011) RaptorX: Exploiting Structure Information for Protein Alignment by Statistical Inference. *Proteins*, **79**, 161-171. <http://dx.doi.org/10.1002/prot.23175>
- [35] Kumar, P., Henikoff, S. and Ng, P.C. (2009) Predicting the Effects of Coding Non-Synonymous Variants on Protein Function Using the SIFT Algorithm. *Nature Protocols*, **4**, 1073-1081. <http://dx.doi.org/10.1038/nprot.2009.86>
- [36] Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. and Chan, A.P. (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*, **7**, e46688. <http://dx.doi.org/10.1371/journal.pone.0046688>

## Abbreviations

**SIFT**: is a sequence homology-based tool that **Sorts Intolerant From Tolerant** amino acid substitutions and predicts whether an amino acid substitution in a protein will have a phenotypic effect

**PROVEAN**: (**Protein Variation Effect Analyzer**)

**ELASPIC**: **Ensemble Learning Approach for Stability Prediction of Interface and Core mutation**

**SNP**: Single Nucleotide polymorphism



**Submit or recommend next manuscript to OALib Journal and we will provide best service for you:**

- Publication frequency: Monthly
- 9 [subject areas](#) of science, technology and medicine
- Fair and rigorous peer-review system
- Fast publication process
- Article promotion in various social networking sites (LinkedIn, Facebook, Twitter, etc.)
- Maximum dissemination of your research work

Submit Your Paper Online: [Click Here to Submit](#)

Contact Us: [service@oalib.com](mailto:service@oalib.com)