EUROPEAN JOURNAL OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES
http://www.ejbps.com

ISSN 2349-8870
Volume: 7
Issue: 1
124-136
Year: 2020

COMPUTATIONAL ANALYSIS OF SINGLE NUCLEOTIDE POLYMORPHISM OF HUMAN *CFH* GENE

Reem A. Osman[1,2], Afra M. Al Bkrye[1], Hind Abdelaziz Elnasri[1] and Mona Abdelrahman Mohamed Khaier*[1]

[1]Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan.
[2]Africa Specialized Hospital for Ear, Nose and Throat, Khartoum, Sudan.

***Corresponding Author: Mona Abdelrahman Mohamed Khaier**
Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan.

Article Received on 13/12/2019 **Article Revised on 15/12/2019** **Article Accepted on 17/12/2019**

ABSTRACT
Introduction: Factor H is a plasma protein that controls activation of the complement system in the fluid phase, the mutations in *CFH* gene causes impaired regulation of the alternative pathway (C3bBb) production of abnormal protein. This protein had been associated with a typical haemolytic-uremic syndrome. This study will focuses on predicting the effects of single nucleotide polymorphisms, nsSNPs that have been reported in *CFH* gene using computational analysis. Materials and Methods: Data about *CFH* gene was retrieved from NCBI database and were analyzed by computational software. Coding region of SNPs that are non-synonymous (nsSNPs) detected by (GeneMANIA, SIFT, Polyphen, PROVEAN, SNPs&GO, PHD-SNPs, I-mutant, MUpro and Project Hope) software. SNPs at un-translated region at3 ends (3UTR) and SNPs at un-translated region at 5ends (5UTR) were analyzed by SNPs function prediction software. **Results:** 357are the total SNPs in *CFH* gene reported in *Homo sapiens*, 309 SNPs in the coding region, 18 in the 3UTR and 30 in the 5UTR. Only SNPs present in the coding region, 3UTR and 5UTR regions were selected for analysis. **Conclusion:** seventeen nsSNPs were highly damaging and affect the function and stability of the protein, the score with PSIC SD range (1-0.99) and TOLERANCE INDEX equal(0-), Only one nsSNPs (rs121913051) from these seventeen nsSNPs was reported to be disease related by previous studies.

INTRODUCTION
A Typical Haemolytic-Uremic Syndrome (THUS) is a condition that causes abnormal blood clots (thrombi) to be formed in small blood vessels in the kidneys, these clots can cause serious medical problems if they restrict or block blood flow, a typical haemolytic-uremic syndrome is characterized by three major features related to abnormal clotting: hemolytic anemia, thrombocytopenia, and kidney failure.[1]

Factor H is a plasma protein (molecular weight155, 000 Ad) that controls activation of the complement system in the fluid phase and in cellular surfaces. Factor H binds to C3b, accelerates the decay of the alternative-pathway C3 converts, and acts as a cofactor in the factor I–mediated proteolysis inactivation of C3b.[2] It can also interact with polyanionicmolecules (sialicacid sorglycosaminoglycans) on certain cellular surfaces, conferring to them resistance to damage as a consequence of complement activation through the alternative pathway.[3]

More than 100 mutations in the complement factor H(*CFH)* gene have been reported with this disease, Mutations in this gene increase the risk of a severe form of the disorder that usually appears early in life.[4]

Most *CFH* gene mutations associated with a typical hemolytic-uremic syndrome affect a region of the complement factor H protein known as the C-terminal domain.[5] These mutations in genes or gene coding for the *CFH* cause impaired regulation of the alternative pathway convertase C3bBb[6] resulting in the production of an abnormal or nonfunctional version of the protein. The resulting shortage of complement factor H can lead to uncontrolled activation of the complement system on the surface of cells, this gene provides instructions for making a protein called complement factor H. This protein helps in regulating part of the body's immune response (innate immune defense system) known as the complement system. The complement system is glycoprotein component produced by liver, blood monocyte, tissue macrophage found in the circulation in an inactive form called (pro_ proteins) this proteins work together to clear microbes and damaged cells from an organism, promotes inflammation and attacks the pathogens cell membrane.

Factor H is composed of 20 repetitive units of 60 amino acids, and these units are named "short consensus repeats" (SCRs pathway)[7] and has three binding sites for C3b, in SCR1-4, SCR6-10, and SCR16-20. Patients with a HUS has resulted in the identification of a remarkable clustering of missense mutation in the SCR16-20 region of factor H.[8] This region is involved in the binding of factor H to C3b deposited on cellular surfaces.[9] Deletion of any of the C3b binding sites significantly decreases factor H binding to C3b deposited on cellular surfaces.[10]

Cytogenetic Location: 1q31.3, which is the long (q) arm of chromosome 1 at position 31.3.[11]

Single nucleotide polymorphisms (SNPs) are the most common type of genetic Variation among people. When SNPs occur within a gene or in a regulatory region near a gene, they may play a more direct role in disease by affecting the gene's function.[12]

OBJECTIVE

This study aimed to detect highly damaging SNPs of complement factor H gene by using computational analysis.

MATERIALS AND METHODS

Sequence Datasets and Polymorphism Identification

All information about *CFH* gene was retrieved from NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/snp). A total of 357 SNPs in *CFH* gene were reported, 309 in the coding sequence, 18 in 3UTR and 30 in 5UTR, were filtered for investigation. Detected. Predictions of deleterious rsSNPs was performed by SIFT and Polyphen-2 and PROVEN software. I-Mutant 3.0 and Mue-pro performed prediction of change in stability due to mutation while the functional impact of the deleterious SNPS was analyze by project hope. Project hope software was used to highlight the changes occurred because of the deleterious SNPs at the molecular level of the protein 3D structure.

GeneMANIA

Gene MANIA (http://www.genemania.org) is a web interface that helps predicting the function of genes and gene sets. Gene MANIA finds other genes that are related to a set of input genes, using a very large set of functional association data. Association data include protein and genetic interactions, pathways, co-expression, co-localization and protein domain similarity. Gene MANIA can be used to detect new members of a pathway or complex, find additional genes you may have missed in your screen or find new genes with a specific function, such as protein kinases. Your question is defined by the set of genes you input.[13]

Sorting intolerant from tolerant (SIFT)

SIFT (http://siftdna.org/www/SIFT_dbSNP.html) prophesy the tolerated and deleterious SNPs and identify amino acid substitutions in protein-coding regions. Each substitution has the potential to affect protein function. This a program that predicts whether an amino acid substitution affects protein function so that users can prioritize substitutions for further study. We have shown that SIFT can distinguish between functionally neutral and deleterious amino acid changes in mutagenesis studies and on human polymorphisms. Scores ≥ 0.05 predicted as tolerated where scores ≤ 0.05 predicted as deleterious.[14]

PolyPhen-2

Polyphen- 2 (http://genetics.bwh.harvard.edu/pph2/) is an online bioinformatics program to automatically foretell the consequence of an amino acid change on the structure and function of a protein. This prediction is based on a number of features comprising the sequence, phylogenetic and structural information characterizing the substitution. Basically, this program searches for 3D protein structures, multiple alignments of homologous sequences and amino acid contact information in several protein structure databases, then calculates position-specific independent count scores (PSIC) for each of the two variants, and then computes the PSIC scores difference between two variants The higher a PSIC score difference, the higher the functional impact a particular amino acid substitution is likely to have. Prediction outcomes could be classified as. ≥ 1.0: probably damaging ≥ 0.75: possibly damaging where ≤ 0.5 benign nsSNPs that predicted to be intolerant by Sift has been submitted to Polyphen as protein sequence in FASTA format that obtained from UniprotKB/ExPASY after submitting the relevant ensemble protein (ESNP) there, and then we entered position of mutation, native amino acid and the new substituent for both structural and functional predictions.[15]

PROVEAN: (Protein Variation Effect Analyzer), PROVEAN was developed to predict whether a protein sequence variation affects protein function or not. PROVEAN is able to provide predictions for any type of protein sequence variations including single or multiple amino acid substitutions, insertions or deletions. The cut off is -2.282 < -2.282 considered neutral > -2.282 considered deleterious.[16]

Single Nucleotide Polymorphism &Gene Ontology (SNPs & Go)

Classifies the SNPs as disease related or neutral. Use information derived from (protein sequence database, protein 3D structure, protein sequence profile, protein function and Gene Ontology (GO) annotation. Score > 0.5 indicates that the mutation is disease related.[17]

Predictor of Human Deleterious Single Nucleotide Polymorphisms (PHD-SNP)

PHD- SNP is a web-based tool available at (http: //snps. biofold.org/phd-snp/phd-snp.html). It predicts whether the new phenotype derived from a nsSNP is a disease related or not (neutral). Protein sequence from

UniprotKB is submitted to the program after providing position and the new amino acid residue.
Score > 0.5 indicates that the mutation is disease related.[18]

I-Mutant 3.0 suite
I-Mutant version 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) was used to predict the protein stability changes upon single-site mutations. I-Mutant basically can evaluate the stability change of a single site mutation starting from the protein structure or from the protein sequences. The cut-off scores of this software >0 (+) increase the protein stability where < 0 (-) decrease the protein stability.[19]

MuPro
The software uses both sequence and structural information to predict whether a mutation will increase or decrease the stability of the protein structure.[20]

Project HOPE
Project **HOPE** (HOPE; http: //www. cmbi.ru.nl/hope/home) is an automatic mutation analysis server to study the insight structural features of native protein and the variant models. HOPE provides the 3D structural visualization of mutated proteins, and gives the results by using UniProt and DAS prediction servers. Input method of Project HOPE carries the protein sequence and selection of Mutant variants. HOPE server predicts the output in the form of structural variation between mutant and wild type residues. The sequence was submitted and only damaging mutation by both SIFTS and Polyphen were selects to be analyzed by the servers.[21]

RESULTS

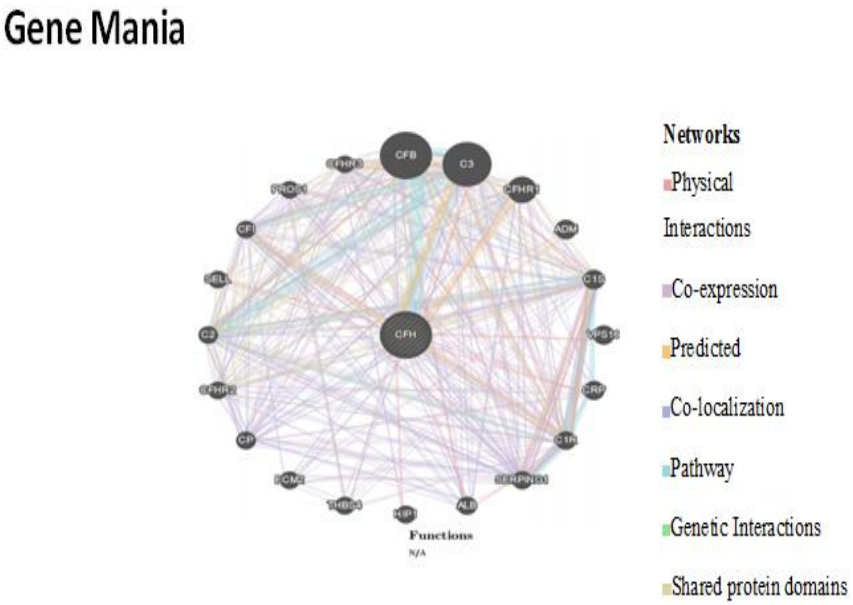

Figure 1: Showing Genes Co- Expression with *Cfh* Gene.

Table 1: Shows The Genes Co-Expressed And Shred Domain With *Cfh*.

Gene Name	Description
CFH	Complement factor H
CFB	Complement factor B
C3	Complement component 3
CFHR1	Complement factor H related 1
ADM	Adrenomedullin
C1S	Complement component 1, s subcomponent

SIFT

Table 2: Show SNPs of Human CFH Protein Predicted by SIFT.

	Amino Acid Change	Protein id	Sift Prediction	Sift Score
rs11539862	N997T	ENSP00000356399	DELETERIOUS	0.04
rs35292876	I1059T	ENSP00000356399	DELETERIOUS	0.013
rs56116390	V1089M	ENSP00000356399	DELETERIOUS	0.002
rs61819913	R830W	ENSP00000356399	DELETERIOUS	0.016
rs112088345	K145E	ENSP00000356399	DELETERIOUS	0
rs115722139	R1215G	ENSP00000356399	DELETERIOUS	0.001
rs121913051	C536R	ENSP00000356399	DELETERIOUS	0.018
rs121913052	C959Y	ENSP00000356399	DELETERIOUS	0.001
rs121913056	C431S	ENSP00000356399	DELETERIOUS	0.003
rs121913058	R127L	ENSP00000356399	DELETERIOUS	0
rs121913058	R1210C	ENSP00000356399	DELETERIOUS	0.015
rs121913063	P1114L	ENSP00000356399	DELETERIOUS	0.002
rs138609319	V149A	ENSP00000356399	DELETERIOUS	0
rs138890387	L3V	ENSP00000356399	DELETERIOUS	0.045
rs139254423	R175P	ENSP00000356399	DELETERIOUS	0.009
rs141021080	S58A	ENSP00000356399	DELETERIOUS	0.017
rs141336681	W978G	ENSP00000356399	DELETERIOUS	0.016
rs142201753	R2T	ENSP00000356399	DELETERIOUS	0.032
rs142266551	T724K	ENSP00000356399	DELETERIOUS	0.038
rs143237092	V144M	ENSP00000356399	DELETERIOUS	0.005
rs146147358	E201G	ENSP00000356399	DELETERIOUS	0.011
rs147549495	E201K	ENSP00000356399	DELETERIOUS	0.029
rs147864267	N1056K	ENSP00000356399	DELETERIOUS	0.006
rs149364107	Q950H	ENSP00000356399	DELETERIOUS	0.019
rs149938052	E902A	ENSP00000356399	DELETERIOUS	0.016
rs200203654	I330T	ENSP00000356399	DELETERIOUS	0.002
rs200306092	P213A	ENSP00000356399	DELETERIOUS	0.042
rs200638179	S437Y	ENSP00000356399	DELETERIOUS	0.003
rs200921441	L532S	ENSP00000356399	DELETERIOUS	0.001
rs201427724	G1110A	ENSP00000356399	DELETERIOUS	0.005
rs201778232	G1002R	ENSP00000356399	DELETERIOUS	0.035
rs201816520	L697F	ENSP00000356399	DELETERIOUS	0.032
rs267598267	G1194S	ENSP00000356399	DELETERIOUS	0.024
rs367687415	Y534C	ENSP00000356399	DELETERIOUS	0.001
rs368437292	P300L	ENSP00000356399	DELETERIOUS	0.044
rs369371759	T447R	ENSP00000356399	DELETERIOUS	0.041
rs370384708	S159N	ENSP00000356399	DELETERIOUS	0.007
rs370640334	Q840H	ENSP00000356399	DELETERIOUS	0.023
rs371462079	N217D	ENSP00000356399	DELETERIOUS	0.012
rs373339058	G133R	ENSP00000356399	DELETERIOUS	0
rs373941930	A421T	ENSP00000356399	DELETERIOUS	0.008
rs374704701	P1051L	ENSP00000356399	DELETERIOUS	0.004
rs375439160	G967E	ENSP00000356399	DELETERIOUS	0.002
rs376742584	N434H	ENSP00000356399	DELETERIOUS	0.008
rs121913056	C367S	ENSP00000352658	DELETERIOUS	0
rs121913058	R127L	ENSP00000352658	DELETERIOUS	0
rs141336681	S58A	ENSP00000352658	DELETERIOUS	0.026
rs200203654	I266T	ENSP00000352658	DELETERIOUS	0.002
rs200306092	P149A	ENSP00000352658	DELETERIOUS	0.029
rs200644601	S373Y	ENSP00000352658	DELETERIOUS	0.021
rs371642012	N153D	ENSP00000352658	DELETERIOUS	0.049
rs373339058	G133R	ENSP00000352658	DELETERIOUS	0
rs376742584	N370H	ENSP00000352658	DELETERIOUS	0.008

Comment: Using this software 53 SNPs were detected as deleterious affecting the function of protein.

Polyphen_2

Table 3: SNPs results of Human CFH Protein Predicted by Polyphen-2.

SNPs	Amino Acid Change	Protein ID	polyphen-2 prediction	Polyphen-2 Score
rs11539862	N997T	ENSP00000356399	Possibly Damaging	0.751
rs35292876	I1059T	ENSP00000356399	Benign	0.025
rs56116390	V1089M	ENSP00000356399	Probably Damaging	1
rs61819913	R830W	ENSP00000356399	Probably Damaging	0.986
rs112088345	K145E	ENSP00000356399	Probably Damaging	1
rs115722139	R1215G	ENSP00000356399	Probably Damaging	1
rs121913051	C536R	ENSP00000356399	Probably Damaging	1
rs121913052	C959Y	ENSP00000356399	Probably Damaging	0,977
rs121913056	C431S	ENSP00000356399	Probably Damaging	1
rs121913058	R127L	ENSP00000356399	Probably Damaging	1
rs121913058	R1210C	ENSP00000356399	Benign	0.024
rs121913063	P1114L	ENSP00000356399	Probably Damaging	0.981
rs138609319	V149A	ENSP00000356399	Probably Damaging	0.975
rs138890387	L3V	ENSP00000356399	Possibly Damaging	0.919
rs139254423	R175P	ENSP00000356399	Benign	0.383
rs141021080	S58A	ENSP00000356399	Benign	0.065
rs141336681	W978G	ENSP00000356399	Probably Damaging	1
rs142201753	R2T	ENSP00000356399	Possibly Damaging	0.903
rs142266551	T724K	ENSP00000356399	Possibly Damaging	0.648
rs143237092	V144M	ENSP00000356399	Probably Damaging	0.996
rs146147358	E201G	ENSP00000356399	Benign	0.002
rs147549495	E201K	ENSP00000356399	Benign	0.024
rs147864267	N1056K	ENSP00000356399	Probably Damaging	0.994
rs149364107	Q950H	ENSP00000356399	Possibly Damaging	0.809
rs149938052	E902A	ENSP00000356399	Probably Damaging	0.983
rs200203654	I330T	ENSP00000356399	Possibly Damaging	0.63
rs200306092	P213A	ENSP00000356399	Probably Damaging	0.999
rs200638179	S437Y	ENSP00000356399	Possibly Damaging	0.692
rs200921441	L532S	ENSP00000356399	Probably Damaging	1
rs201427724	G1110A	ENSP00000356399	Probably Damaging	1
rs201778232	G1002R	ENSP00000356399	Probably Damaging	0.985
rs201816520	L697F	ENSP00000356399	Probably Damaging	1
rs267598267	G1194S	ENSP00000356399	Probably Damaging	0.996
rs367687415	Y534C	ENSP00000356399	Probably Damaging	1
rs368437292	P300L	ENSP00000356399	Probably Damaging	0.999
rs369371759	T447R	ENSP00000356399	Probably Damaging	0.997
rs370384708	S159N	ENSP00000356399	Possibly Damaging	0.848
rs370640334	Q840H	ENSP00000356399	Probably Damaging	1
rs371462079	N217D	ENSP00000356399	Probably Damaging	0.991
rs373339058	G133R	ENSP00000356399	Probably Damaging	1
rs373941930	A421T	ENSP00000356399	Benign	0.001
rs374704701	P1051L	ENSP00000356399	Probably Damaging	1
rs375439160	G967E	ENSP00000356399	Probably Damaging	1
rs376742584	N434H	ENSP00000356399	Probably Damaging	0.964
rs121913056	C367S	ENSP00000352658	Probably Damaging	1
rs121913058	R127L	ENSP00000352658	Probably Damaging	1
rs141336681	S58A	ENSP00000352658	Benign	0.009
rs200203654	I266T	ENSP00000352658	Probably Damaging	0.999
rs200306092	P149A	ENSP00000352658	Probably Damaging	0.999
rs200644601	S373Y	ENSP00000352658	Probably Damaging	1
rs371642012	N153D	ENSP00000352658	Possibly Damaging	0.865
rs373339058	G133R	ENSP00000352658	Probably Damaging	1
rs376742584	N370H	ENSP00000352658	Possibly Damaging	0.765

Comment: Probably damaging SNPs were 35, possibly damaging SNPs were 10 SNPs while 8 SNPs were benign.

Provean

Table (4): SNPs of human CFH protein predicted by Provean.

SNPs	Amino Acid Change	Protein ID	Provean Prediction	Provean Score
rs11539862	N997T	ENSP00000356399	Deleterious	-2.59
rs35292876	I1059T	ENSP00000356399	Deleterious	-2.92
rs56116390	V1089M	ENSP00000356399	Neutral	-2.08
rs61819913	R830W	ENSP00000356399	Neutral	-1.31
rs112088345	K145E	ENSP00000356399	Neutral	-2.33
rs115722139	R1215G	ENSP00000356399	Deleterious	-5.98
rs121913051	C536R	ENSP00000356399	Deleterious	-7.9
rs121913052	C959Y	ENSP00000356399	Deleterious	-7.47
rs121913056	C431S	ENSP00000356399	Deleterious	-7.82
rs121913058	R127L	ENSP00000356399	Deleterious	-4.31
rs121913058	R1210C	ENSP00000356399	Neutral	-0.75
rs121913063	P1114L	ENSP00000356399	Deleterious	-7.22
rs138609319	V149A	ENSP00000356399	Neutral	-2.16
rs138890387	L3V	ENSP00000356399	Neutral	-0.85
rs139254423	R175P	ENSP00000356399	Neutral	-2.23
rs141021080	S58A	ENSP00000356399	Neutral	-0.86
rs141336681	W978G	ENSP00000356399	Deleterious	-8.72
rs142201753	R2T	ENSP00000356399	Neutral	-1.78
rs142266551	T724K	ENSP00000356399	Deleterious	-3.72
rs143237092	V144M	ENSP00000356399	Neutral	-1.55
rs146147358	E201G	ENSP00000356399	Deleterious	-3.1
rs147549495	E201K	ENSP00000356399	Neutral	-1.7
rs147864267	N1056K	ENSP00000356399	Deleterious	-3.7
rs149364107	Q950H	ENSP00000356399	Deleterious	-2.98
rs149938052	E902A	ENSP00000356399	Deleterious	-3.83
rs200203654	I330T	ENSP00000356399	Deleterious	-3.49
rs200306092	P213A	ENSP00000356399	Deleterious	-4.25
rs200638179	S437Y	ENSP00000356399	Deleterious	-3.56
rs200921441	L532S	ENSP00000356399	Deleterious	-3.7
rs201427724	G1110A	ENSP00000356399	Deleterious	-4.73
rs201778232	G1002R	ENSP00000356399	Deleterious	-2.83
rs201816520	L697F	ENSP00000356399	Deleterious	-3.01
rs267598267	G1194S	ENSP00000356399	Deleterious	-4.13
rs367687415	Y534C	ENSP00000356399	Deleterious	-5.81
rs368437292	P300L	ENSP00000356399	Deleterious	-4.76
rs369371759	T447R	ENSP00000356399	Deleterious	-3.13
rs370384708	S159N	ENSP00000356399	Neutral	-1.38
rs370640334	Q840H	ENSP00000356399	Deleterious	-3.19
rs371462079	N217D	ENSP00000356399	Deleterious	-2.89
rs373339058	G133R	ENSP00000356399	Deleterious	-4.93
rs373941930	A421T	ENSP00000356399	Neutral	-0.26
rs374704701	P1051L	ENSP00000356399	Deleterious	-6.18
rs375439160	G967E	ENSP00000356399	Deleterious	-5.22
rs376742584	N434H	ENSP00000356399	Deleterious	-3.2
rs121913056	C367S	ENSP00000352658	Deleterious	-9.12
rs121913058	R127L	ENSP00000352658	Deleterious	-4.22
rs141336681	S58A	ENSP00000352658	Neutral	-0.89
rs200203654	I266T	ENSP00000352658	Deleterious	-3.96
rs200306092	P149A	ENSP00000352658	Deleterious	-4.5
rs200644601	S373Y	ENSP00000352658	Deleterious	-3.78
rs371642012	N153D	ENSP00000352658	Deleterious	-2.88
rs373339058	G133R	ENSP00000352658	Deleterious	-4.96
rs376742584	N370H	ENSP00000352658	Deleterious	-3.46

Comment: this software detected 39 SNPs as Deleterious and 14 SNP as Neutral.

SNP&GO

Table (5): Show SNPs of human CFH protein predicted by SNPs&GO.

SNPs	Amino Acid Change	Protein ID	SNP PR	SNP RI	SNP PRO
rs11539862	N997T	ENSP00000356399	Neutral	6	0.218
rs35292876	I1059T	ENSP00000356399	Neutral	6	0.178
rs56116390	V1089M	ENSP00000356399	Neutral	6	0.182
rs61819913	R830W	ENSP00000356399	Neutral	4	0.302
rs112088345	K145E	ENSP00000356399	Neutral	5	0.267
rs115722139	R1215G	ENSP00000356399	Neutral	5	0.258
rs121913051	C536R	ENSP00000356399	Disease	7	0.848
rs121913052	C959Y	ENSP00000356399	Disease	9	0.961
rs121913056	C431S	ENSP00000356399	Disease	5	0.731
rs121913058	R127L	ENSP00000356399	Disease	8	0.894
rs121913058	R1210C	ENSP00000356399	Neutral	4	0.304
rs121913063	P1114L	ENSP00000356399	Neutral	8	0.098
rs138609319	V149A	ENSP00000356399	Neutral	4	0.283
rs138890387	L3V	ENSP00000356399	Neutral	9	0.071
rs139254423	R175P	ENSP00000356399	Disease	7	0.855
rs141021080	S58A	ENSP00000356399	Neutral	8	0.108
rs141336681	W978G	ENSP00000356399	Disease	2	0.575
rs142201753	R2T	ENSP00000356399	Neutral	7	0.126
rs142266551	T724K	ENSP00000356399	Neutral	3	0.345
rs143237092	V144M	ENSP00000356399	Neutral	5	0.257
rs146147358	E201G	ENSP00000356399	Neutral	7	0.13
rs147549495	E201K	ENSP00000356399	Neutral	7	0.149
rs147864267	N1056K	ENSP00000356399	Disease	2	0.587
rs149364107	Q950H	ENSP00000356399	Neutral	7	0.153
rs149938052	E902A	ENSP00000356399	Neutral	7	0.128
rs200203654	I330T	ENSP00000356399	Disease	0	0.515
rs200306092	P213A	ENSP00000356399	Neutral	5	0.253
rs200638179	S437Y	ENSP00000356399	Disease	4	0.693
rs200921441	L532S	ENSP00000356399	Disease	2	0.612
rs201427724	G1110A	ENSP00000356399	Neutral	9	0.072
rs201778232	G1002R	ENSP00000356399	Disease	1	0.532
rs201816520	L697F	ENSP00000356399	Neutral	6	0.185
rs267598267	G1194S	ENSP00000356399	Disease	2	0.582
rs367687415	Y534C	ENSP00000356399	Disease	6	0.807
rs368437292	P300L	ENSP00000356399	Neutral	9	0.045
rs369371759	T447R	ENSP00000356399	Neutral	4	0.282
rs370384708	S159N	ENSP00000356399	Neutral	6	0.198
rs370640334	Q840H	ENSP00000356399	Neutral	7	0.133
rs371462079	N217D	ENSP00000356399	Disease	0	0.509
rs373339058	G133R	ENSP00000356399	Disease	5	0.768
rs373941930	A421T	ENSP00000356399	Neutral	9	0.05
rs374704701	P1051L	ENSP00000356399	Neutral	4	0.281
rs375439160	G967E	ENSP00000356399	Disease	8	0.877
rs376742584	N434H	ENSP00000356399	Neutral	2	0.419
rs121913056	C367S	ENSP00000352658	Disease	4	0.704
rs121913058	R127L	ENSP00000352658	Disease	8	0.884
rs141336681	S58A	ENSP00000352658	Neutral	7	0.134
rs200203654	I266T	ENSP00000352658	Disease	1	0.566
rs200306092	P149A	ENSP00000352658	Neutral	4	0.29
rs200644601	S373Y	ENSP00000352658	Disease	4	0.686
rs371642012	N153D	ENSP00000352658	Neutral	1	0.466
rs373339058	G133R	ENSP00000352658	Disease	5	0.765
rs376742584	N370H	ENSP00000352658	Neutral	2	0.412

Comment: this software detected 21 SNPs causing Disease and 32SNPs were Neutral.

PHD&SNP

Table 6: SNPs of human CFH protein predicted by PHD&SNP.

SNPs	Amino Acid Change	Protein ID	PHD PR	PHD RI	PHD PRO
rs11539862	N997T	ENSP00000356399	Disease	4	0.7
rs35292876	I1059T	ENSP00000356399	Disease	1	0.539
rs56116390	V1089M	ENSP00000356399	Neutral	0	0.486
rs61819913	R830W	ENSP00000356399	Disease	3	0.64
rs112088345	K145E	ENSP00000356399	Neutral	3	0.34
rs115722139	R1215G	ENSP00000356399	Disease	1	0.545
rs121913051	C536R	ENSP00000356399	Disease	9	0.946
rs121913052	C959Y	ENSP00000356399	Disease	10	0.979
rs121913056	C431S	ENSP00000356399	Disease	7	0.831
rs121913058	R127L	ENSP00000356399	Disease	6	0.793
rs121913058	R1210C	ENSP00000356399	Disease	1	0.553
rs121913063	P1114L	ENSP00000356399	Neutral	5	0.247
rs138609319	V149A	ENSP00000356399	Neutral	3	0.328
rs138890387	L3V	ENSP00000356399	Neutral	8	0.099
rs139254423	R175P	ENSP00000356399	Disease	8	0.892
rs141021080	S58A	ENSP00000356399	Neutral	9	0.071
rs141336681	W978G	ENSP00000356399	Disease	8	0.911
rs142201753	R2T	ENSP00000356399	Neutral	7	0.134
rs142266551	T724K	ENSP00000356399	Neutral	0	0.481
rs143237092	V144M	ENSP00000356399	Disease	0	0.517
rs146147358	E201G	ENSP00000356399	Neutral	5	0.251
rs147549495	E201K	ENSP00000356399	Neutral	4	0.323
rs147864267	N1056K	ENSP00000356399	Disease	6	0.805
rs149364107	Q950H	ENSP00000356399	Neutral	4	0.314
rs149938052	E902A	ENSP00000356399	Neutral	4	0.301
rs200203654	I330T	ENSP00000356399	Disease	4	0.686
rs200306092	P213A	ENSP00000356399	Disease	2	0.604
rs200638179	S437Y	ENSP00000356399	Disease	6	0.814
rs200921441	L532S	ENSP00000356399	Disease	3	0.636
rs201427724	G1110A	ENSP00000356399	Neutral	4	0.316
rs201778232	G1002R	ENSP00000356399	Disease	6	0.82
rs201816520	L697F	ENSP00000356399	Neutral	2	0.385
rs267598267	G1194S	ENSP00000356399	Disease	5	0.772
rs367687415	Y534C	ENSP00000356399	Disease	9	0.939
rs368437292	P300L	ENSP00000356399	Neutral	7	0.141
rs369371759	T447R	ENSP00000356399	Neutral	5	0.268
rs370384708	S159N	ENSP00000356399	Neutral	5	0.241
rs370640334	Q840H	ENSP00000356399	Neutral	4	0.287
rs371462079	N217D	ENSP00000356399	Disease	1	0.548
rs373339058	G133R	ENSP00000356399	Disease	6	0.806
rs373941930	A421T	ENSP00000356399	Neutral	7	0.172
rs374704701	P1051L	ENSP00000356399	Disease	4	0.69
rs375439160	G967E	ENSP00000356399	Disease	9	0.944
rs376742584	N434H	ENSP00000356399	Neutral	2	0.389
rs121913056	C367S	ENSP00000352658	Disease	6	0.798
rs121913058	R127L	ENSP00000352658	Disease	7	0.834
rs141336681	S58A	ENSP00000352658	Disease	1	0.527
rs200203654	I266T	ENSP00000352658	Disease	6	0.788
rs200306092	P149A	ENSP00000352658	Disease	3	0.651
rs200644601	S373Y	ENSP00000352658	Disease	6	0.789
rs371642012	N153D	ENSP00000352658	Disease	1	0.542
rs373339058	G133R	ENSP00000352658	Disease	7	0.829
rs376742584	N370H	ENSP00000352658	Neutral	1	0.432

Comment: this software detected 32 SNPs cause disease and 21 SNPs is neutral.

I- MUTANT

Table (7): SNPs of Human Cfh Protein Predicted by I-Mutant.

SNPs	Amino Acid Change	Protein ID	I-MUTANT Prediction	I-MUTANT Score
rs11539862	N997T	ENSP00000356399	Increase	1
rs35292876	I1059T	ENSP00000356399	Decrease	9
rs56116390	V1089M	ENSP00000356399	Decrease	7
rs61819913	R830W	ENSP00000356399	Decrease	5
rs112088345	K145E	ENSP00000356399	Decrease	7
rs115722139	R1215G	ENSP00000356399	Decrease	5
rs121913051	C536R	ENSP00000356399	Decrease	2
rs121913052	C959Y	ENSP00000356399	Increase	1
rs121913056	C431S	ENSP00000356399	Decrease	5
rs121913058	R127L	ENSP00000356399	Decrease	8
rs121913058	R1210C	ENSP00000356399	Decrease	3
rs121913063	P1114L	ENSP00000356399	Decrease	7
rs138609319	V149A	ENSP00000356399	Decrease	10
rs138890387	L3V	ENSP00000356399	Decrease	8
rs139254423	R175P	ENSP00000356399	Decrease	6
rs141021080	S58A	ENSP00000356399	Decrease	8
rs141336681	W978G	ENSP00000356399	Decrease	9
rs142201753	R2T	ENSP00000356399	Decrease	7
rs142266551	T724K	ENSP00000356399	Decrease	4
rs143237092	V144M	ENSP00000356399	Decrease	9
rs146147358	E201G	ENSP00000356399	Decrease	6
rs147549495	E201K	ENSP00000356399	Decrease	8
rs147864267	N1056K	ENSP00000356399	Decrease	6
rs149364107	Q950H	ENSP00000356399	Decrease	5
rs149938052	E902A	ENSP00000356399	Decrease	9
rs200203654	I330T	ENSP00000356399	Decrease	7
rs200306092	P213A	ENSP00000356399	Decrease	9
rs200638179	S437Y	ENSP00000356399	Decrease	1
rs200921441	L532S	ENSP00000356399	Decrease	8
rs201427724	G1110A	ENSP00000356399	Decrease	8
rs201778232	G1002R	ENSP00000356399	Decrease	7
rs201816520	L697F	ENSP00000356399	Decrease	8
rs267598267	G1194S	ENSP00000356399	Decrease	8
rs367687415	Y534C	ENSP00000356399	Decrease	5
rs368437292	P300L	ENSP00000356399	Decrease	6
rs369371759	T447R	ENSP00000356399	Decrease	6
rs370384708	S159N	ENSP00000356399	Decrease	0
rs370640334	Q840H	ENSP00000356399	Decrease	8
rs371462079	N217D	ENSP00000356399	Decrease	2
rs373339058	G133R	ENSP00000356399	Decrease	9
rs373941930	A421T	ENSP00000356399	Decrease	5
rs374704701	P1051L	ENSP00000356399	Decrease	7
rs375439160	G967E	ENSP00000356399	Decrease	4
rs376742584	N434H	ENSP00000356399	Decrease	6
rs121913056	C367S	ENSP00000352658	Decrease	5
rs121913058	R127L	ENSP00000352658	Decrease	8
rs141336681	S58A	ENSP00000352658	Decrease	8
rs200203654	I266T	ENSP00000352658	Decrease	7
rs200306092	P149A	ENSP00000352658	Decrease	8
rs200644601	S373Y	ENSP00000352658	Decrease	1
rs371642012	N153D	ENSP00000352658	Decrease	2
rs373339058	G133R	ENSP00000352658	Decrease	9
rs376742584	N370H	ENSP00000352658	Decrease	6

Comment: 32 SNPs is Decrease stability of protein can lead to disease and 21 SNPs is increase stability of protein that means the function of protein is normal.

Project-Hope results

This software determines the effect of mutation on amino acid structure, charge and hydrophobicity. The schematic structures of the original amino acid and the mutant amino acid. The backbone is same for the two amino acids.

1- (rs267598267), the mutation of a Glycine into a Serine at position 1194 as shown in figure (1). Each amino acid has its own specific size, charge, and hydrophobicity-value. The original wild-type residue and newly introduced mutant residue often differ in these properties. The mutant residue is bigger than the wild-type residue.

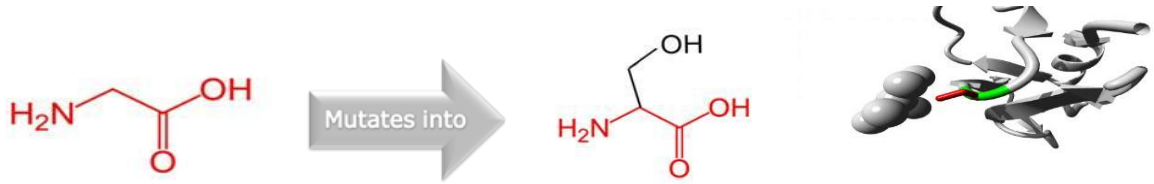

Figure 1: Amino acid change from Glycine to a Serine the protein (grey color) residues changed from wild type (green color) to mutant (red color).

2- (rs121913056) mutation of a Cysteine into a Serine at position 367 as shown in figure (2). Each amino acid has its own specific size, charge, and hydrophobicity-value. The original wild-type residue and newly introduced mutant residue often differ in these properties the wild-type residue is more hydrophobic than the mutant residue.

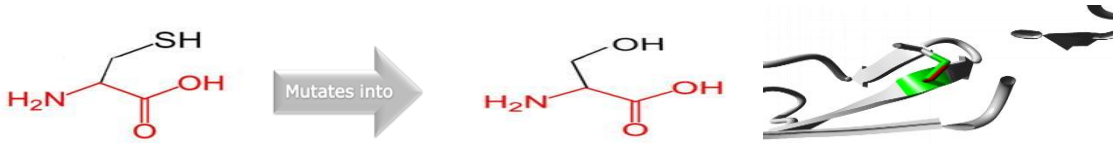

Figure (2): Amino acid change a Cysteine to a Serine, the protein (grey color) residues changed from wild type(green color) to mutant (red color).

(3) (rs200203654) the mutation of a Isoleucine into a Threonine at position 266 as shown in figure (3). Each amino acid has its own specific size, charge, and hydrophobicity-value the original wild-type residue and newly introduced mutant residue often differ in these properties; the mutant residue is smaller than the wild-type residue, the wild-type residue is more hydrophobic than the mutant residue.

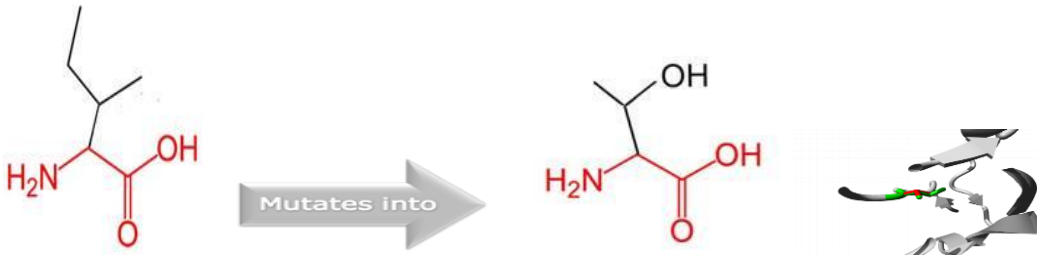

Figure (3): Amino acid change a Isoleucine to a Threonine, the protein (grey color) residues changed from wild type (green color) to mutant (red color).

(4) (rs200644601) the mutation of a Serine into a Tyrosine at position 373 as shown in figure(4).Each amino acid has its own specific size, charge, and hydrophobicity-value the original wild-type residue and newly introduced mutant residue often differ in these properties. The mutant residue is bigger than the wild-type residue.

Figure (4): Amino acid change a Serine to a Tyrosine, the protein (grey color) residues changed from wild type (green color) to mutant (red color).

(5) (rs147864267) the mutation of a Asparagine into a Lysine at position 1056 as shown in figure(5). Each amino acid has its own specific size, charge, and hydrophobicity-value. The original wild-type residue and newly introduced mutant residue often differ in these properties. The mutant residue is bigger than the wild-type residue. The wild-type residue charge was NEUTRAL; the mutant residue charge is POSITIVE.

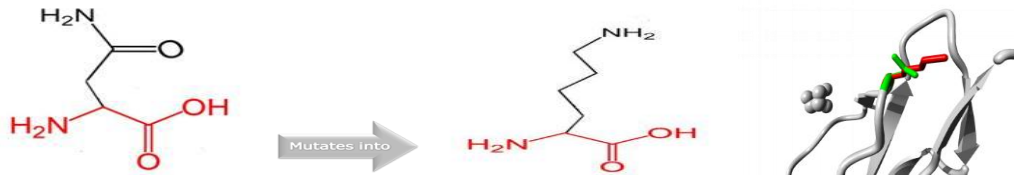

Figure (5): Amino Acid Change A Asparagine To A Lysine, The Protein (Grey Color) Residues Changed From Wild Type (Green Color) To Mutant (Red Color).

(6) (rs200921441) the mutation of a Leucine into a Serine at position 532 as shown in figure(6) Each amino acid has its own specific size, charge, and hydrophobicity-value the original wild-type residue and newly introduced mutant residue often differ in these properties. The mutant residue is smaller than the wild-type residue. The wild-type residue is more hydrophobic than the mutant residue.

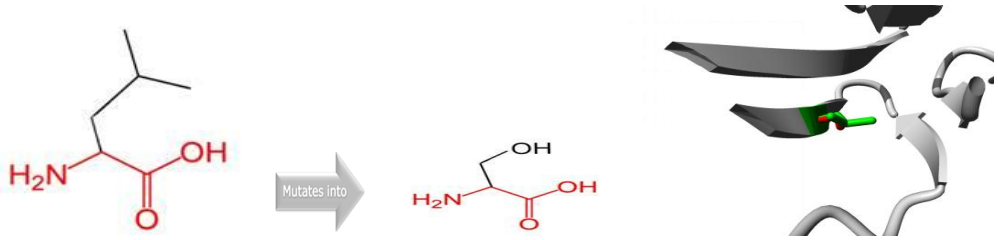

Figure (6): Amino acid change a Leucine into a Serine, the protein (grey color) residues changed from wild type (green color) to mutant (red color).

(7) (rs121913051) the mutation of a Cysteine into a Arginine at position 536 as shown in figure(7) Each amino acid has its own specific size, charge, and hydrophobicity-value the original wild-type residue and newly introduced mutant residue often differ in these properties. The mutant residue is bigger than the wild-type residue. The wild-type residue charge was NEUTRAL; the mutant residue charge is POSITIVE. The wild-type residue is more hydrophobic than the mutant residue.

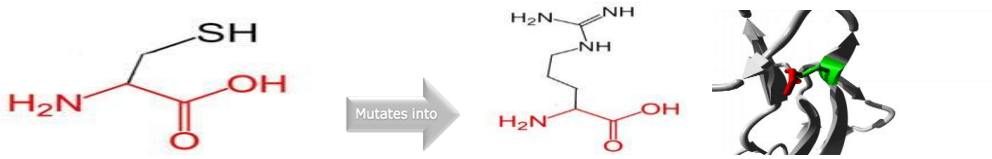

Figure (7): Amino acid change a Cysteine to a Arginine, the **protein (grey color) residues changed from wild type (green color) to mutant (red color).**

(8) rs367687415 the mutation of a Tyrosine into a Cysteine at position 534 as shown in figure(8) Each amino acid has its own specific size, charge, and hydrophobicity-value the original wild-type residue and newly introduced mutant residue often differ in these properties. The mutant residue is smaller than the wild-type residue. The mutant residue is more hydrophobic than the wild-type residue.

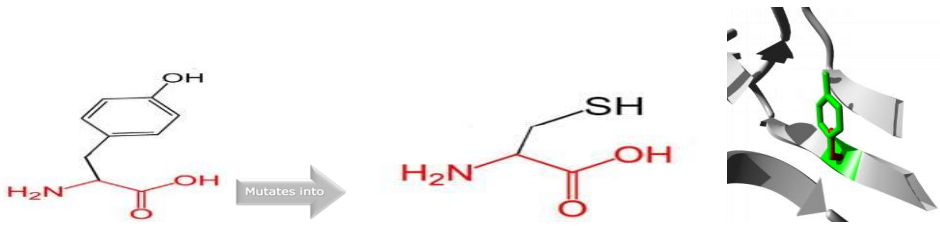

Figure (8): Amino Acid Change A Tyrosine To A Cysteine, The **Protein (Grey Color) Residues Changed From Wild Type (Green Color) To Mutant (Red Color).**

DISCUSSION

CFH gene was investigated in NCBI database (http://www.ncbi.nlm.nih.gov/). A total of 357 SNPs in *CFH* gene were reported on Homo sapiens, 309 in the coding region, 18 in 3UTR and 30 in 5UTR. Non synonymous SNPs were analyzed by SIFT, proven software, All SNPs predicted to be deleterious, 53 deleterious detected by sift, 39 deleterious and14 neutral by proven.

These deleterious SNPs were analyzed using PolyPhen-2 software to predict the damaging SNPs, we found that 35 SNPs were predicted to be deleterious & probably damaging in both software's and 10 possibly damaging while only 8 SNPS predicted as benign. The I-*mutant* software used to detect the stability of protein we found that 48 SNPs were scored decreased in the stability of the protein, While only5 SNPs had increased the stability of the protein, by MUE PRO software 51 SNPs are decreased in stability and 2 SNPs with increased in stability , it was predicted to be disease related by SNP & Go software that only 21 SNPs was reported as a disease, while the 32SNPs were reported as a neutral and PHD-SNP software detect 32 SNPs were scored as a disease related, while 21 scored as neutral.

The rs121913051 result in substitution of a Cysteine into bigger, less hydrophobic and was positively charged Arginine at position 536. The rs200921441 caused conversion of a Leucine amino acid with smaller and less hydrophobic a Serine at position 532. The rs367687415 the mutation of a Tyrosine into smaller, more hydrophobic to a Cysteine at position 534. According to project hope software this detects the effect of the mutation on the protein structure and function.

This study came out with that CFH protein structure and function was affected by seventeen nsSNPs, One nsSNPs (rs121913051) from these seventeen nsSNPs were reported to be damaging and disease related by another study, conducted by Warwicker, *et al*.,[22] three other SNPs (rs460897 - rs121913059 - rs121913059) were reported to cause disease in another study

CONCLUSIONS AND RECOMMENDATIONS

The available data from the NCBI dbSNP database for *CFH*gene has been analyzed by computational analysis to predicted deleterious and damaging SNPs by both Polyphen server and SIFT program, and detect protein function and stability by I-mutant, MUpro software. Through investigations and wet lab experimentation are needed to explore the effects of these polymorphisms.*

More studies may help in exploring the effects of these not reported SNPs.

REFERENCES

1. Genetic home references (https://ghr.nlm.nih.gov/)
2. Pangburn MK, Schreiber RD, Mu¨ller-Eberhard HJ Human complement C3b in activator: isolation, characterization, and demonstration of an absolute requirement for the serum protein b1H for cleavage of C3b and C4b in solution. *J Exp. Med*., 1977; 146: 257–270.
3. Meri S, Pangburn, M.K. Discrimination between activators and non-activators of the alternative pathway of complement Regulation via a sialic acid/polyanionic binding site on factor H. Proc *Natl Acad Sci* USA, 1990; 87: 3982–3986.
4. Genetic home references (https://ghr.nlm.nih.gov/)
5. Genetic home references (https://ghr.nlm.nih.gov/)
6. Miha´ ly Jo´ zsi, 1 Christoph Licht, 2 Stefanie Strobel, 1 Svante L. H. Zipfel, 1 Heiko Richter, 1 Stefan Heinen, 1 Peter F. Zipfel, 1, 3 and Christine Skerka1.
7. Ripoche J, Day AJ, Harris TJR, Sim RB the complete amino acid sequence of human complement factor H. *Biochem J*., 1988; 249: 593–602.
8. Pe´rez-Caballero D, Gonza´lez-Rubio C, Gallardo ME, Vera M, Lo´ pez-Trascasa M, Rodrı´guez de Co´ rdoba S, Sa´nchez-CorralP Clustering of missense mutations in the C-terminal Region of factor H in atypical hemolytic uremic syndrome. *Am J Hum Genet*, 2001; 68: 478–484.
9. Pangburn MK, Pangburn KLW, Koistinen V, Meri S, Sharma AK Molecular mechanisms of target recognition in an innate immune system: interactions among factor H, C3b, and target in the alternative pathway of human complement. *J Immunol*, 2000; 164: 4742–4751.
10. Sharma AK, Pangburn MK Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. Proc *Natl Acad Sci* USA, 1996; 93: 10996–11001.
11. Genetic home references (https://ghr.nlm.nih.gov/)
12. Genetic home references (https://ghr.nlm.nih.gov/)
13. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *J. Nucleic Acids Res*., 2010.
14. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*., 2003; 31: 3812-3814.
15. Ivan Adzhubei, Daniel M. Jordan, and Shamil R. Sunyaev. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr Protoc Hum Genet*, 2013; 0 7: Unit7.20.
16. J. Cheng A. Randall, P. Baldi. Prediction Of protein Stability Changes for Single -Site Mutations Using Support Vector Machines. *Proteins*, 2005; 62(4): 1125-1132.
17. Bava K. A, Gromiha M. M, Uedaira H, Kitajima K, Sarai A. ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res.*, 2004; 32: 120-121.

18. Gonzalez Perez A, López-Bigas N. Improving the assessment of the outcome of no synonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet*, 2011; 88: 440-449.
19. J. Cheng, A. Randall, and P. Baldi. Prediction of Protein Stability Changes for Single-Site Mutations Using Support Vector Machines. *Proteins*, 2006; 62(4): 1125-1132.
20. Capriotti E, Fariselli P, Calabrese R, Casadio R. Predicting protein stability changes from sequences using support vector machines. *Bioinformatic,* 2005; 21: 54-58.
21. P. Warwicker, T. H. J. Goodship, R. L. Donne,…… *Kidney International*, 1998; 53(4): 836–844.