International Journal of Multidisciplinary and Current Research ISSN: 2321-3124
Research Article Available at: http://ijmcr.com

In silico Analysis of Single Nucleotide Polymorphisms (SNPS) in Human Factor VIII Gene

Arwa Abdelhalim Mohammed[1], Soada Ahmed Osman[2], Rayan Sidig Adam[3], Ezdihar Elhadi Elamin[4] and Mohamed Ahmad Salih[4]

[1]Department of Pharmacy, Sudan Medical Council, Khartoum-Sudan
[2,4]Department of Biotechnology, Africa city of Technology, Sudan
[3]Department of Cytopathology and Histopathology, University of Gezira, Sudan

Accepted 02 Jan 2017, Available online 06 Jan 2017, **Vol.5 (Jan/Feb 2017 issue)**

Abstract

F8 gene is extremely large (186 kb) and consists of 26 exons. F8 is located on the long arm of the Xq28 region of the X chromosome. Mutations in this gene can occur at diverse sites in a variety of types, such as structural variation and sequence variation. Single Nucleotide Polymorphisms (SNPs) are the most abundant sequence variations encountered in a genome playing a major role in understanding of the genetic basis of many complex human diseases, the nonsynonymous SNPs (nsSNPs) are known to be deleterious or disease-causing variations because they alter protein sequence, structure, and function. F8 gene plays a major role in hemophilia A. A computational-based (In Silico) analysis has been done to evaluate the phenotypic effect of nsSNPs in human Out of total of 6021 SNPs in the F8 gene, 1276 were nsSNPs, 626 were missenes mutation and 650 duo to nonsense mutation, 54 occurred in the 3'utr and 24 occurred in 5'utr Among the predicted nsSNPs, (rs371422922), (rs370369511), (rs1800288), (rs368808810) and (rs373079141), were identified as deleterious and highly damaging by the SIFT and PolyPhen programs. The protein structural analysis of these amino acid variants was performed using (I mutant and PHD), where Modeling of these amino acid substitutions prepared by Project hope. From a comparison of the stabilizing residues of the native and mutant proteins, we propose that these variants could be an important candidate for hemophilia A, caused by the F8 gene and other pathological condition.

Keywords: *Single Nucleotide Polymorphisms (SNPs), F8 gene, In silico analysis, nonsynonymous SNPs (nsSNPs), 3'utr, 5'utr, Polyphen and SIFT, hemophilia A (HA)*

1. Introduction

F8 gene provides instructions for making a protein called coagulation factor VIII. Coagulation factors are a group of related proteins that are essential for the formation of blood clots. After an injury, clots protect the body by sealing off damaged blood vessels and preventing further blood loss. [1,2] F8 gene is extremely large (186 kb) and consists of 26 exons. [3] The cytogenetic location of *F8* gene is Xq28, which is the long (q) arm of the X chromosome at position 28 where the molecular location of this gene starts from base pairs 154,835,788 to 155,022,723 on the X chromosome (Homo sapiens Annotation Release 108, GRCh38.p7) [4] Coagulation factor VIII is made chiefly by cells in the liver. This protein circulates in the bloodstream in an inactive form, bound to another molecule called von Willebrand factor, until an injury that damages blood vessels occurs. In response to injury, coagulation factor VIII is activated and separates from von Willebrand factor. The active protein (sometimes written as coagulation factor VIIIa) interacts with another coagulation factor called factor IX. This interaction sets off a chain of additional chemical reactions that form a blood clot. [1] Defects in this gene results in hemophilia A, a common recessive X-linked coagulation disorder. [5] The X-linked bleeding disorder of hemophilia A (HA) is characterized by coagulation factor VIII (FVIII) deficiency. [6] Furthermore Nonsevere hemophilia A is generally caused by *F8* missense mutations. [7] Despite information on large numbers of *F8* mutations associated with nonsevere hemophilia A that is collected in international databases. [8,9] Missenes mutation and nonsense mutation are both part of nonsynonymous mutations. Nonsynonymous mutations have a much greater effect on an individual than a synonymous mutation. In a nonsynonymous mutation, there is either an insertion or deletion of a single nucleotide in the sequence during transcription when the messenger RNA is copying the DNA. This single missing or added nucleotide causes a frame shift mutation, which throws off the entire reading frame of the amino acid sequence and mixes up the codons. This affects the amino acids that are coded for and change the resulting protein that is expressed. The severity of this

kind of mutation depends on how early in the amino acid sequence it happens. If it happens near the beginning and the entire protein is changed, this could become a lethal mutation. [10] Currently, HA is treated by administration of plasma-derived or recombinant FVIII, [11] but this strategy is complicated by the development of inhibitory antibodies in 30–40% of patients affected by the severe form of the disease. [12] Curative gene and cell therapies are, therefore, of interest for HA. It would be useful for such therapies to delineate the cell types capable of producing FVIII in necessary amounts. [13] It is not possible to calculate the inhibitor risk for specific *F8* mutations, as data on exposure days to therapeutic factor VIII concentrates are lacking. Patients with hemophilia -A who are treated with factor VIII concentrates are at risk of developing factor VIII neutralizing alloantibodies (inhibitors). [14,15] Inhibitor development is one of the most challenging complications in the treatment of hemophilia A, as it increases the bleeding tendency while it renders treatment with therapeutic factor VIII concentrates ineffective. Treatment of patients with inhibitors can be very complicated and costly. [16] One of the most important predictors of the risk of inhibitor development in severe hemophilia A is the *F8* gene mutation type. [17, 18] Reported absolute and relative risks of inhibitor development according to the different *F8* mutation types vary markedly between studies. Consequently in silico -analysis can yield more particular estimates of the risk for different types of *F8* mutations. Currently, there are no data on a gene linked to *F8* that also causes inhibitors, but of course this is theoretically possible. [19] In this study we used different computational methods to identify the *F8* gene SNPs and the effects of predicted mutation at the proteomic level. The identification of *F8* mutations and it`s effects is important for a precise diagnosis of HA, understanding of genotype-phenotype correlation, carrier detection, prenatal diagnosis, and predicting inhibitor development. We considered this study a distinctive one, because there is no researches deal with matter in silico studies.

2. Material and Methods

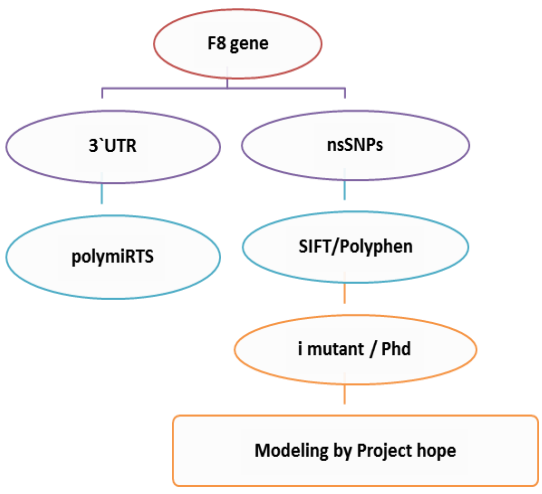

Figure 1: Workflow Software's used in SNPs analysis

2.1 Data mining

The data on human F8 gene was collected from National Center for Biological Information (NCBI) web site. The SNP information (protein accession number and SNP ID) of the F8 gene was retrieved from the NCBI dbSNP (http://www.ncbi.nlm.nih.gov/snp/) and Swiss Prot databases (http://expasy.org/). [20]

2.2 Evaluation of the functional impact of coding nsSNPs using a sequence homology tool (SIFT)

SIFT Available at (http://sift.bii.a-star.edu.sg/). 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. Considers the position at which the change occurred and the type of amino acid change. Given a protein sequence, SIFT chooses related proteins and obtains an alignment of these proteins with the query. Based on the amino acids appearing at each position in the alignment, SIFT calculates the probability that an amino acid at a position is tolerated conditional on the most frequent amino acid being tolerated. If this normalized value is less than a cutoff, the substitution is predicted to be deleterious. SIFT scores <0.05 are predicted by the algorithm to be intolerant or deleterious amino acid substitutions, whereas scores >0.05 are considered tolerant. [21]

2.3 Predicting Functional Effect of Human Missense Mutations by PolyPhen-2

The PolyPhen-2Web interface can be reached at http://genetics.bwh.harvard.edu/pph2/. The input form at this URL allows querying for a single individual amino acid substitution or a coding, non-synonymous SNP annotated in the dbSNP database. It calculates position-specific independent count (PSIC) scores for each of the two variants and computes the difference of the PSIC scores of the two variants. The higher a PSIC score difference, the higher functional impact a particular amino acid substitution is likely to have. PolyPhen scores were designated as ''probably damaging'' (0.95–1), ''possibly damaging'' (0.7–0.95), and ''benign'' (0.00–0.31). [22]

2.4 Predicting protein folding, misfolding and diseases by I-Mutant and PHD_SNP

I-Mutant 3.0 suite (http://gpcr2. biocomp. unibo.it/cgi/predictors/ I-Mutant3.0/I-Mutant3.0. cgi) and PHD_SNP (http://snps.biofold.org/phd-snp/phd-snp.html) Is a neural network based tool for the routine analysis of protein stability and alterations by taking into account the single-site mutations. The FASTA sequence of protein retrieved from UniProt is used as an input to predict the mutational effect on protein stability. I-Mutant also

provides the scores for free energy alterations due to single point protein mutations by adopting a hypothesis of thermodynamic reversibility of the existing experimental data. [23]

2.5 Modeling nsSNP locations on protein structure

Mapping the deleterious nsSNPs into protein 3D structure information was performed by project hope to know the impact of SNPs on structural level (http://www.cmbi.ru.nl/hope/input). [24]

2.6 Identification of DNA polymorphisms in miRNAs and miRNA target sites by polymirts

(http://compbio.uthsc.edu/miRSNP/) we submitted SNPs located within the 3'-UTRs then the polymirts checked if the SNP variants could alter putative miRNA target sites focusing on mutations that alter sequence complementarity to miRNA seed regions possibly leading to hemophilia A. [25]

2.7 Prediction for gene prioritization and predicting gene function by geneMANIA

GeneMANIA (http://www.genemania.org/) We submitted genes and selected from a list of data sets that they wish to query. GeneMANIA approach to know protein function prediction integrate multiple genomics and proteomics data sources to make inferences about the function of unknown proteins. [26]

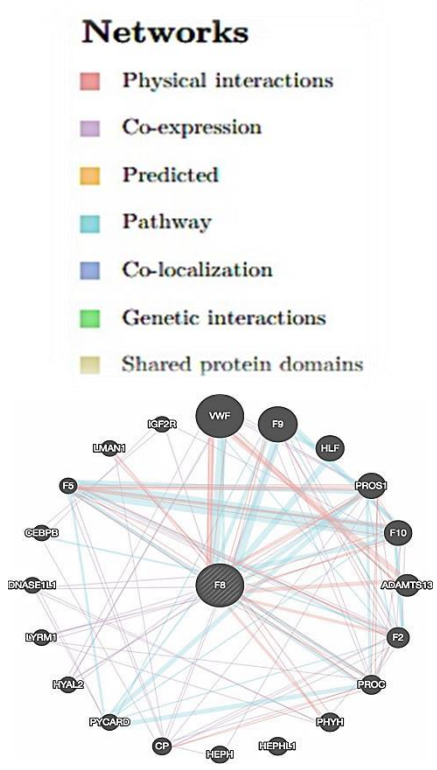

Figure 2: interaction between *F8* and its related genes

3. Results and Discussion

3.1 F8 gene function and activites illustrated by GENEMANIA

F8 Protein has many vital functions. It has many activities such as: peptidyl-glutamic acid carboxylation, protein carboxylation, blood coagulation, intrinsic pathway, peptidyl-glutamic acid modification, fibrin clot formation, protein activation cascade, Golgi lumen, platelet activation, post-translational protein modification, platelet alpha granule lumen, platelet alpha granule, secretory granule lumen, cytoplasmic membrane-bounded vesicle lumen, vesicle lumen, platelet degranulation, endoplasmic reticulum lumen, secretory granule, exocytosis, serine-type endopeptidase activity and leukocyte migration. The genes co-expressed with, share similar protein domain, or participate to achieve similar function are illustrated by using GENEMANIA and shown in Figure (2), Table (1) and (2) below.

Table 1: The *F8* gene functions and its appearance in network and genom

Function	FDR	Genes in network	Genes in genome
peptidyl-glutamic acid carboxylation	4.03E-09	5	11
protein carboxylation	4.03E-09	5	11
blood coagulation, intrinsic pathway	4.96E-08	5	18
peptidyl-glutamic acid modification	5.05E-08	5	19
blood coagulation, fibrin clot formation	7.06E-08	5	21
protein activation cascade	1.30E-07	6	61
Golgi lumen	7.32123E-05	5	84
platelet activation	0.000160522	6	211
post-translational protein modification	0.000160522	6	209
platelet alpha granule lumen	0.000375801	4	48
platelet alpha granule	0.00088901	4	61
secretory granule lumen	0.00088901	4	62
cytoplasmic membrane-bounded vesicle lumen	0.001735315	4	76
vesicle lumen	0.001735315	4	76
platelet degranulation	0.002198693	4	82
endoplasmic reticulum lumen	0.017341899	4	140
secretory granule	0.022565552	4	152
exocytosis	0.049373587	4	191
serine-type endopeptidase activity	0.049373587	3	66
leukocyte migration	0.096466105	4	230

***FDR:** False discovery rate is greater than or equal to the probability that this is a false positive

Table 2: The genes co-expressed, share domain and Interaction with *F8* gene

Entity 1	Entity 2	Weight	Network group
HYAL2	F8	0.019201733	Co-expression
IGF2R	F8	0.022348957	Co-expression
LYRM1	F8	0.014900776	Co-expression
DNASE1L1	F8	0.02426847	Co-expression
LYRM1	F8	0.01839787	Co-expression
DNASE1L1	F8	0.026349438	Co-expression
LYRM1	F8	0.016701337	Co-expression

LYRM1	F8	0.006861785	Co-expression
HYAL2	F8	0.013577192	Co-expression
F10	F8	0.013215183	Co-localization
IGF2R	F8	0.013168437	Co-localization
HYAL2	F8	0.019501498	Co-localization
VWF	F8	0.06012475	Pathway
F9	F8	0.09405152	Pathway
HLF	F8	0.18598104	Pathway
PROS1	F8	0.12179842	Pathway
F10	F8	0.13173525	Pathway
F2	F8	0.031619057	Pathway
PROC	F8	0.081246145	Pathway
CEBPB	F8	0.03429785	Pathway
VWF	F8	0.29483756	Pathway
F9	F8	0.28152388	Pathway
PYCARD	F8	0.0399082	Pathway
VWF	F8	0.39948943	Physical interactions
ADAMTS13	F8	0.39948943	Physical interactions
F2	F8	0.25483292	Physical interactions
VWF	F8	0.17619766	Physical interactions
PROS1	F8	0.25030422	Physical interactions
F10	F8	0.13846181	Physical interactions
PHYH	F8	0.2940384	Physical interactions
LMAN1	F8	0.07140928	Physical interactions
VWF	F8	0.08163424	Physical interactions
F9	F8	0.15136333	Physical interactions
PROS1	F8	0.08478585	Physical interactions
F10	F8	0.040783245	Physical interactions
PROC	F8	0.0664233	Physical interactions
PHYH	F8	0.09007206	Physical interactions
LMAN1	F8	0.20435846	Physical interactions
VWF	F8	0.14242032	Physical interactions
F9	F8	0.19860737	Physical interactions
PROS1	F8	0.12405749	Physical interactions
F10	F8	0.06901379	Physical interactions
PROC	F8	0.08327525	Physical interactions
PHYH	F8	0.21720189	Physical interactions
LMAN1	F8	0.13916731	Physical interactions
HEPHL1	F8	0.074891426	Shared protein domains
HEPH	F8	0.07124354	Shared protein domains
CP	F8	0.07124354	Shared protein domains
F5	F8	0.040583298	Shared protein domains
HEPHL1	F8	0.17869668	Shared protein domains
HEPH	F8	0.15368262	Shared protein domains
CP	F8	0.15368262	Shared protein domains
F5	F8	0.0812287	Shared protein domains

3.2 Deleterious or damaging nsSNPs predicted by SIFT and PolyPhen

These SNPs were found on the coding region, 3'UTR and 5'UTR SNPs were selected for computational analysis. SNPs lied on coding region (in X chromosome) were Predicted by SIFT and Polyphen-2: Predictions of deleterious nsSNPs were performed by SIFT and Polyphen software; from 1276 nsSNPs, 17 mutations were predicted to be deleterious/damaging by both servers. First, we submitted batch nsSNPs (rs SNPs) to SIFT server; then the resultant damaging nsSNPs were submitted to Polyphen as query sequences in FASTA Format, it traced 13 probably damaging predictions with score (0.976-1), the other 4 predictions were scored as possibly damaging with score (0.651-0.974). The result are reported in (table - 3)

Table 3: Double positive result of Polyphen and SIFT Table

Amino Acid	Protin ID	SIFT Prediction	SIFT score	Polyphen Prediction	PSIC SD
T2310I	ENSP00000353393	DELETERIOUS	**0.006**	Probably damaging	**0.998**
T175I	ENSP00000327895	DELETERIOUS	**0.016**	Probably damaging	**0.998**
Q2208H	ENSP00000353393	DELETERIOUS	**0.045**	Probably damaging	**0.976**
S2113Y	ENSP00000353393	DELETERIOUS	**0.008**	Probably damaging	**1**
C1922W	ENSP00000353393	DELETERIOUS	**0.004**	Probably damaging	**0.999**
L1778Q	ENSP00000353393	DELETERIOUS	**0**	Probably damaging	**1**
P1765T	ENSP00000353393	DELETERIOUS	**0.013**	Possibly damaging	**0.947**
V1752A	ENSP00000353393	DELETERIOUS	**0**	Probably damaging	**1**
R1664C	ENSP00000353393	DELETERIOUS	**0.025**	Probably damaging	**0.995**
E1094Q	ENSP00000353393	DELETERIOUS	**0.043**	Possibly damaging	**0.651**
T1056I	ENSP00000353393	DELETERIOUS	**0.007**	Possibly damaging	**0.883**
L606F	ENSP00000353393	DELETERIOUS	**0.016**	Probably damaging	**0.986**
C573Y	ENSP00000353393	DELETERIOUS	**0**	Probably damaging	**1**
R503H	ENSP00000353393	DELETERIOUS	**0.038**	Probably damaging	**0.992**
Y365C	ENSP00000353393	DELETERIOUS	**0.002**	Probably damaging	**0.997**
D364N	ENSP00000353393	DELETERIOUS	**0.049**	Possibly damaging	**0.651**
D75V	ENSP00000353393	DELETERIOUS	**0**	Probably damaging	**1**

***PolyPhen-2 result**: POROBABLY DAMAGING (more confident prediction) / POSSIBLY DAMAGING (less confident prediction), **PSIC SD:** Position-Specific Independent Counts software if the score is ≥ 0.5,**Tolerance Index**: Ranges from 0 to 1. The amino acid substitution is predicted damaging if the score is ≤ 0.05, and tolerated if the score is > 0.05.

3.3 Prediction of change in stability due to mutation used I-Mutant 3.0 and PHD-SNP servers

I mutant 3.0-server output demonstrated that protein stability with relate free energy has changed due to mutation. All mutations were detected by SIFT/Polyphen servers gave the results as following: all eleven mutations (Q→H, C→W, L→Q, P→T, V→A, R→C, E→Q, T→ I, L →F, R→H, and D→V) predicted a dramatic decrease of the protein stability, while six mutations (T→I, S→Y, C →Y, Y→C and D→N) predicted increase of stability of F8 protein. While in PHD-SNP output demonstrated whether an nsSNP is disease- related (disease) or neutral polymorphism, All 17 mutations detected by SIFT/Polypen-2 servers represented the following results: eight mutations (T→I, Q→H, S→Y, L →Q, R→C, C →Y, Y→C and D →V) predicted disease related polymorphism and eight mutations (C→W, P → T, V → A , E→Q , T→I , L→F , R→H and D→N) predicted neutral polymorphism, Table-3.

3.4 Modeling of amino acid substitution effects due to nsSNPs on protein structure

According to Sift /Polyphen results in which TOLERANCE INDEX ≤0.005and PSIC SD equals 1(double positive), we found that 15 SNPs in 18 mutations achieved these scores and these mutation occurs in following SNPs rs370369511, rs368808810, rs1800288 and rs371422922 with TI= 0 and PSIC SD equal 1. While rs373079141 with TI= **0.006**, TI= **0.016** and PSIC SD equal **0.998** these 5 SNPs were selected to be submitted to the Project Hope software to reveal the 3D structure for the truncated proteins with its new candidates; in addition, it described the reaction and physiochemical properties of these candidates. 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. All native and mutant structure of F8 protein showed in Figure (3,4,5,6 and 7) respectively, the wild type is displayed by green color while the mutated type displayed by red one.

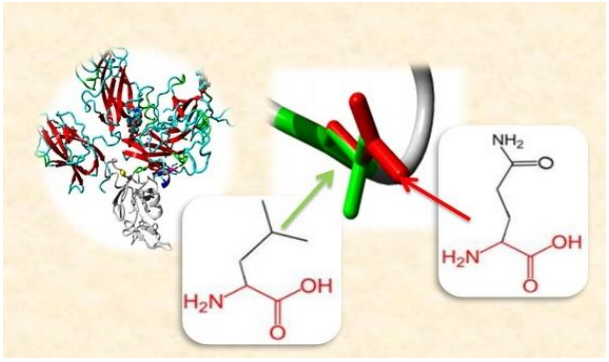

Figure 3: shoes 1st SNP that occur at the position 1778 L / Q.

1st SNP (rs371422922) that occur at the position 1778 caused conversion of leucine to a glutamine

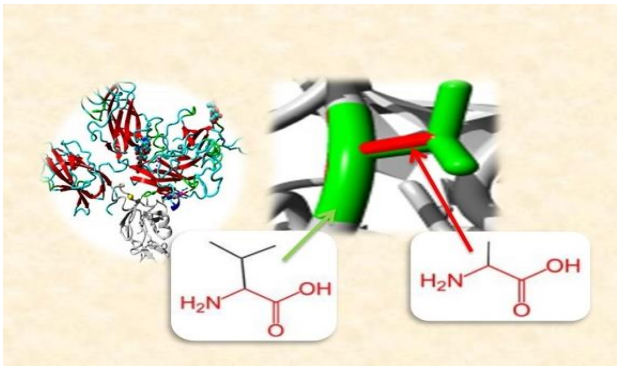

Figure 4: 2nd SNP at the position 1752 V / A.

2nd SNP (rs370369511) at the position 1752 caused conversion of valine to alanine

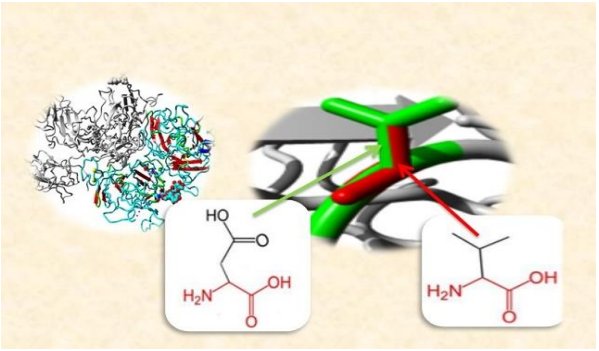

Figure 5: 3rd SNP at the position 75 D / V.

3rd SNP (rs1800288) at the position 75, caused conversion of aspartic acid into valine

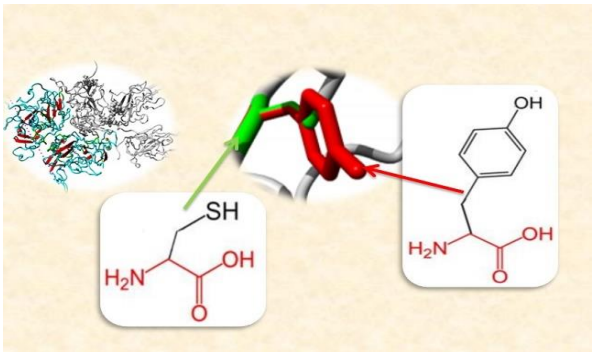

Figure 6: 4th SNP at the position 573 C / Y.

4th SNP (rs368808810) at the position 573 caused conversion of cysteine into a tyrosine.

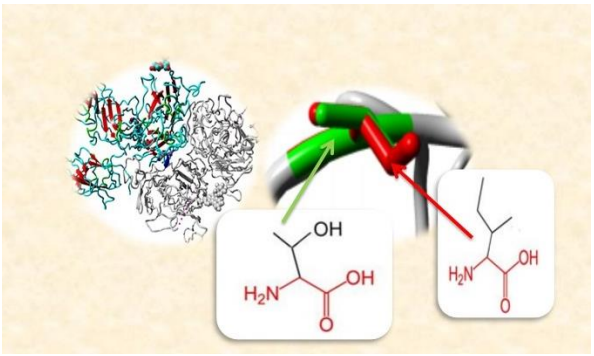

Figure 7: 5th SNP at the position 175 T / I

5th SNP (rs373079141) at the position 175 caused conversion of threonine into isoleucine

The present study was designed to study the properties of the protein that result from the single nucleotide polymorphism in the coding regions of Factor VIII. Five interesting SNPs were chosen according to the double positive result of Polyphen and SIFT, 1st SNP (rs371422922) that occur at the position 1778 was changed from Leucine to a Glutamine this mutant residue is located near a highly conserved position.

Table 4: SNPs and INDELs in miRNA target sites in *F8* gene

Location	dbSNP ID	miR ID	Conservation	miR Site	Function Class	context+ score change
154064257	rs182784273	hsa-miR-4655-5p	3	atagCCCGGTAga	D	-0.293
		hsa-miR-1911-3p	3	ataGCCTGGTAga	C	-0.491
		hsa-miR-6753-5p	3	atagCCTGGTAga	C	-0.184
154064344	rs1803603	hsa-miR-6804-3p	2	atgatgCAGGTGC	D	-0.259
154064397	rs34683807	hsa-miR-583	4	tgaCCTCTTTgga	D	-0.093
		hsa-miR-3974	5	TGACCTTtttgga	C	-0.218
154064456	rs186338743	hsa-miR-562	2	agaatgGCTACTT	D	-0.188
		hsa-miR-411-5p	2	agaatGTCTACTt	C	-0.228
154064482	rs202040251	hsa-miR-450b-3p	2	tgGATCCCAAtct	D	-0.378
		hsa-miR-4713-3p	2	tGGATCCCAatct	D	-0.459
		hsa-miR-5089-5p	4	tggATCCCAAtct	D	-0.101
		hsa-miR-638	2	tgGATCCCAatct	D	-0.194
		hsa-miR-6513-5p	4	tggATCCCAAtct	D	-0.095
		hsa-miR-6728-5p	4	tggATCCCAAtct	D	-0.113
		hsa-miR-769-3p	2	tgGATCCCAAtct	D	-0.368
		hsa-miR-640	2	TGGATCAcaatct	C	-0.219
		hsa-miR-8070	4	tggATCACAAtct	C	-0.076
154064493	rs36101366	hsa-miR-153-5p	3	tgaaagAAAAATG	D	0.054
		hsa-miR-1250-3p	3	tgaaagGAAAATG	C	-0.012
154064658	rs34700571	hsa-miR-6771-3p	4	tgtgaaGGGTTTA	D	-0.034
154064687	rs4487960	hsa-miR-31-5p	2	cccctTCTTGCCc	D	-0.086
154064725	rs199597237	hsa-miR-122-5p	2	actggtgaaaacagaaaaaACACTCCAgtctgccatatcaccacaca	O	0.087
		hsa-miR-3659	2	actggtgaaaacagaaaaAACACTCcagtctgccatatcaccacaca	O	-0.005
		hsa-miR-504-3p	3	actggtgaaaacagaaaaaaCACTCCAgtctgccatatcaccacaca	O	0.096
154064812	rs192048301	hsa-miR-128-3p	2	atatACTGTGAct	C	-0.052
		hsa-miR-216a-3p	2	atatACTGTGAct	C	-0.042
		hsa-miR-27a-3p	2	atatACTGTGAct	C	-0.141
		hsa-miR-27b-3p	2	atatACTGTGAct	C	-0.141
		hsa-miR-3681-3p	2	atatACTGTGAct	C	-0.059
154064871	rs183289212	hsa-miR-4419a	3	ctCTCCCTCtact	D	-0.191
		hsa-miR-4510	3	ctCTCCCTCtact	D	-0.181
		hsa-miR-6127	3	ctCTCCCTCtact	D	-0.172
		hsa-miR-6129	3	ctCTCCCTCtact	D	-0.181
		hsa-miR-6130	3	ctCTCCCTCtact	D	-0.181
		hsa-miR-6133	3	ctCTCCCTCtact	D	-0.172
		hsa-miR-6731-5p	2	CTCTCCCtctact	D	-0.18
		hsa-miR-6760-5p	2	cTCTCCCTctact	D	-0.123
		hsa-miR-6873-5p	2	ctcTCCCTCTAct	D	-0.246
		hsa-miR-8085	2	CTCTCCCtctact	D	-0.198
		hsa-miR-3154	2	ctctCCTTCTAct	C	-0.032
		hsa-miR-4428	2	cTCTCCTTctact	C	-0.084
154065169	rs5904396	hsa-miR-5683	3	aggCATCTGTAt	O	-0.151
154065194	rs189078164	hsa-miR-3162-5p	2	TCCCTAAgtcccc	D	-0.088
154065851	rs201147256	hsa-miR-1207-5p	2	gcaCCTGCCActg	D	-0.133
		hsa-miR-3665	3	gCACCTGCcactg	D	-0.149
		hsa-miR-4736	2	gcACCTGCCActg	D	-0.328
		hsa-miR-4763-3p	2	gcaCCTGCCActg	D	-0.133
		hsa-miR-7150	2	gcaCCTGCCActg	D	-0.171
		hsa-miR-764	2	GCACCTGccactg	D	-0.16

D: The derived allele disrupts a conserved miRNA site (ancestral allele with support > = 2).
C: The derived allele creates a new miRNA site
O: The ancestral allele cannot be determined

The mutation is present within a domain, annotated in UniProt as Plastocyanin-like 5, the domain type is A3 (multicopper oxidase domain), that introduces an amino acid with different properties. The residue of interest is located in a disordered region, but with partially disordered (70%) which annul its function. The mutant residue is bigger than the wild residue makes the new residue not buried in the core of the protein as wild type, which lead to loss of hydrophobic interactions at that position.

2nd SNP (rs370369511) at the position 1752 altered from Valine to Alanine this mutant residue is located near a highly conserved region, like the 1st SNP this mutation occur within the same domain and also different properties and percentage disorder except that the residue of interest is not placed in a disordered region. The wild - type and mutant amino acids differ in size (The mutant residue is smaller than the wild-type residue). That difference will cause an empty space in the core of the protein above all, this may not damage it.

3rd SNP (rs1800288) at the position 75, changed from Aspartic acid into Valine the mutant and wild type located in a conserved area. The mutation is located within a domain, annotated in UniProt as: Plastocyanin-like 1 the domain type is F5/8 A 1(multicopper oxidase domain), That introduces an amino acid with different properties and the residue of interest is located in a disordered region, but with partially disordered (70%) which abolish

its function. It is somewhat surprising that the mutation is possibly not damaging to the protein but according to its location on the surface of a domain, this will cause a possible loss of external interactions. 4th SNP (rs368808810) at the position 573 caused conversion of Cysteine into a Tyrosine. The mutation is located within a domain, annotated in UniProt as: Plastocyanin-like 3 the domain type is F5/8 A2 (multicopper oxidase domain), this domain can be disturb in function by this mutation also. As a result of that it is probably damaging to the protein. The observed difference between mutant and wild residue in size (The mutant residue is bigger than the wild-type residue) , will cause loss of hydrophobic interactions in the core of the protein. So the mutant and wild-type residue are not very similar.

5th SNP (rs373079141) at the position 175 caused change of threonine into isoleucine. The mutation is placed within a domain, annotated in UniProt as: F5/8 type C 2 (promote membrane binding), That introduces an amino acid with different properties and the residue of interest is located in a disordered region, but with partially disordered (55%) which annul its function. The wild-type residue forms a hydrogen bond with: aspartic acid at position 172. The size difference between wild-type and mutant residue (The mutant residue is bigger than the wild-type residue) makes that the new residue is not in the correct position to make the same hydrogen bond as the initial wild-type residue did. This difference in hydrophobicity will affect hydrogen bond formation. As 4th SNP The residue is located on the surface of the protein, mutation of this residue can disturb interactions with other molecules or other parts of the protein.

3.5 SNPs at the 3'UTR region

SNPs in 3'UTR of *f8* gene were submitted as batch to PolymiRTS server. The output showed that among 33 SNPs in 3'UTR region of *f8* gene, about 14 SNPs were predicted, namely rs182784273, rs1803603, rs34683807, rs186338743, rs202040251, rs36101366, rs34700571, rs4487960, rs199597237, rs192048301, rs183289212, rs5904396, rs189078164 and rs201147256. Among these 14 SNPs, 31 alleles disrupted a conserved miRNA site and 9 derived alleles created a new site of miRNA. As an example rs202040251 SNP contained (D) allele had 7 miRSite as target binding site can be disrupts a conserved miRNA and (C) alleles had 2 miRSite disrupts a conserved miRNA site, while 4 ancestral alleles (O) cannot be determined in the all predicted 14 SNPs. Table (4) below demonstrates the SNPs predicted by Polymirt to induce disruption or formation of mirRNA binding site.

Conclusion

The *F8* gene was investigated in this work by evaluating the influence of functional SNPs through computation methods. Out of total of 6021 SNPs were found on the *F8* gene of homo sapiens; of which 626 were missense, 650 were nonsense mutation, 1276 non-synonymous SNPs (nsSNPs), while 54 occurred in the 3'un-translated region and 24 occurred in 5'un-translated region and the rest were other types of SNPs. From 1276 nsSNPs, 17 SNPs were predicted to be deleterious/damaging by both SIFT and polyphen server. Further, we evaluated the protein stability based upon mutations caused by all 17 nsSNPs by using two distinct servers (IMUTANT, PHD). Consequently eleven mutations predicted a dramatic decrease of the protein stability and six mutations predicted disease related polymorphism. Structural analysis results showed that the amino acid residue substitutions which had the greatest impact on the stability of the F8 protein were mutations (rs371422922), (rs370369511), (rs1800288), (rs368808810) and (rs373079141). Based on our results we conclude that these SNPs should be considered important candidates in causing diseases related to *F8* gene mutation. These variants are reagents for further protein function and molecular epidemiology studies of hemophilia A susceptibility also can be used for pharmacogenomics and pharmacokinetic studies. Finally some appreciations of wet lab techniques are suggested to support our computational analysis results.

Acknowledgments

The authors wish to acknowledge the enthusiastic cooperation of Africa City of Technology, Sudan and the department of Pharmacy, Sudan Medical Council, Khartoum-Sudan.

Competing interests

The authors declare that they have no competing interests.

References

[1]. < https://ghr.nlm.nih.gov/gene/F8#resources >
[2]. <https://www.msdmanuals.com/home/blood-disorders/blood-clotting-process/how-blood-clots >
[3]. Graw, J., Brackmann, H.H., Oldenburg, J., Schneppenheim, R., Spannagl, M., & Schwaab, R.(2005), Haemophilia A: from mutation analysis to new therapies, Nat Rev Genet,Vol. 6, No. 6, pp 488-501, ISSN 1471-0056 (Print).
[4]. < https://www.ncbi.nlm.nih.gov/gene/2157 >
[5]. < http://www.ncbi.nlm.nih.gov/refseq/2008>
[6]. Graw J, Brackmann HH, Oldenburg J, Schneppenheim R, Spannagl M, Schwaab R. Haemophilia A: from mutation analysis to new therapies. Nat Rev Genet. 2005;6(6): 488–501. [PubMed]
[7]. d'Oiron R, Pipe SW, Jacquemin M. Mild/moderate haemophilia A: new insights into molecular mechanisms and inhibitor development. Haemophilia 2008;14(Suppl 3):138-146.
[8]. Kemball-Cook G. The Haemophilia A Mutation, Structure, Test and Resource Site.http://www.hadb.org.uk/. Accessed April 15, 2013.
[9]. Centers for Disease Control and Prevention. CHAMP: CDC Hemophilia A Mutation Project.

http://www.cdc.gov/ncbddd/hemophilia/champs.html. Accessed April 15, 2013.

[10]. <http://evolution.about.com/od/Overview/a/Synonymous-Vs-Nonsynonymous-Mutations.htm>

[11]. Den Uijl IE, Mauser Bunschoten EP, Roosendaal G, et al. Clinical severity of haemophilia A: does the classification of the 1950s still stand? Haemophilia. 2011; 17(6): 849–853. [PubMed]

[12]. Dimichele D. Inhibitors: resolving diagnostic and therapeutic dilemmas. Haemophilia. 2002;8(3): 280–287. [PubMed]

[13]. Terraube V, O'Donnell JS, Jenkins PV. Factor VIII and von Willebrand factor interaction: biological, clinical and therapeutic importance. Haemophilia. 2010; 16(1): 3–13. [PubMed]

[14]. Wight J, Paisley S.The epidemiology of inhibitors in haemophilia A: a systematic review.Haemophilia 2003;9(4):418-435.

[15]. Fijnvandraat K, Cnossen MH, Leebeek FW, Peters M. Diagnosis and management of haemophilia.BMJ 2012;344:e2707.

[16]. Di Minno MN, Di Minno G, Di Capua M, Cerbone AM, Coppola A. Cost of care of haemophilia with inhibitors. Haemophilia 2010; 16(1): e190-e201.

[17]. Schwaab R, Brackmann HH, Meyer C, et al. Haemophilia A: Mutation type determines risk of inhibitor formation. Thromb Haemost 1995; 74(6): 1402-1406.

[18]. Ter Avest PC, Fischer K, Mancuso ME, et al. Risk stratification for inhibitor development at first treatment for severe haemophilia A: a tool for clinical practice. J Thromb Haemost 2008; 6(12): 2048-2054.

[19]. http://www.bloodjournal.org/content/122/11/1954.full?sso-checked=true

[20]. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski M, Sirotkin K, dbSNP: the NCBI database of genetic variation.Nucleic Acids Research, 2001; 29:308–311.

[21]. Ng PC, Henikoff S, (2003) SIFT: predicting amino acid changes that affect protein function, Nucleic Acids Research, 31, (13). 3812–3814.

[22]. Adzhubei I, Jordan D M, Sunyaev SR, (2013) Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2.Human Genetics.76: 7.20.1-7.20.41.

[23]. Capriati E., Calabrese R. and Casadio R. (2006). Predicting the insurgence of human genetic disease associated to single point protein mutations with Support Vector Machines and evolutionary information. Bioiformatics 22(22): 2729-2734.

[24]. Venselaar H, Beek T A, Kuipers R KP, Hekkelman M L, Vriend G (2010) Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics.11 (1): 548.

[25]. Anindya Bhattacharya, Jesse D. Ziebarth, Yan Cui (2014). PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic Acids Research.42: 86–91.

[26]. Farley D W, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C , Kazi F, Lopes C T, Maitland A , Mostafavi S , Monto J, Shao Q, Wright G , Bader G D , Morris Q, (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function Nucleic Acids Researcher. 38 (2): 214-220.