EAS Journal of Biotechnology and Genetics
Abbreviated Key Title: EAS J Biotechnol Genet
ISSN: 2663-189X (Print) & ISSN: 2663-7286 (Online)
Published By East African Scholars Publisher, Kenya

OPEN ACCESS

Volume-2 | Issue-3|May-June-2020 | DOI: 10.36349/easjbg.2020.v02i03.002

Research Article

Computational Analysis of Non-Synonymous Single Nucleotide Polymorphism (nsSNPs) in Human *CYCS* Gene

Mohmmed Elmoataz Widaa [1], Afra M. Al Bkrye[1], Hind Abdelaziz Elnasri[1] & Mona A. M. Khaier*[1]

[1] Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan

Article History
Received: 04.04.2020
Accepted: 08.05.2020
Published: 13.05.2020

Journal homepage:
https://www.easpublisher.com/easjbg

Quick Response Code

Abstract: ***Background*****:** The cytochrome c, somatic (*CYCS*) gene encodes a small haeme protein that functions as a central component of the electron transport chain in mitochondria. Most important health condition related to mutations in the coding region of this gene was Thrombocytopenia 4 and affect apoptosis. ***Aims*****:** This study aimed to investigate the effect of non-synonymous SNPs (ns SNPs) of *CYCS* gene in protein function and structure using different computational software. ***Material and Methods*****:** Different nsSNPs and protein related sequences were obtained from NCBI and ExPASY database (2020). Deleterious and damaging effect of SNPs were analyzed using SIFT, Provean, Polyphen-2 and SNPs & GO software. Protein stability was investigated using I-Mutant and MUpro software. The interaction of *CYCS* with other genes was studied using GeneMANIA software. The structural and functional impact of point mutations was predicted using Project Hope software. ***Results***: A total of 100 Single Nucleotide Polymorphisms were retrieved from National Center of Biotechnology Information (NCBI). From these 68 were in the 3`UTR, 28 in the 5`UTR and 14 in the coding region (nsSNP). Only these 14 were selected for further investigations. Six nsSNPs were found to be deleterious while 8 were tolerated by SIFT. Using Provean 11 nsSNPs were found to be deleterious while 3 were Neutral. For PolyPhen-2, 5 nsSNPs were observed to be damaging while 9 were benign. Using PHD and SNPs&GO 9 and 6 nsSNPs were found to be disease related for the two software respectively. The change in the chemical nature of amino acid and how it affects the protein structure was analyzed using Project Hope. ***Conclusion*****:** Six highly deleterious, damaging and disease related nsSNPs (rs17851278, rs11548796, rs121918552, rs11548820, rs11548818 and rs11548778) were detected at *CYCS* gene. These nsSNPs can be considered as candidate nsSNPs for diagnosis of Thrombocytopenia 4 and normal apoptosis.

Keywords: *CYCS* gene, Computational analysis, Thrombocytopenia and nsSNPs.

Copyright © 2020 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)** which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

INTRODUCTION

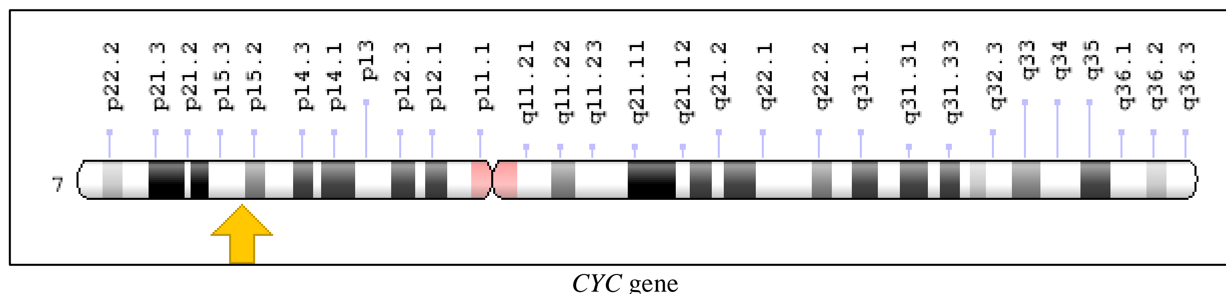
The cytochrome c, somatic *CYCS* gene encodes a small haeme protein that functions as a central component of the electron transport chain in mitochondria. The encoded protein associates with the inner membrane of the mitochondrion where it accepts electrons from cytochrome b and transfers them to the cytochrome oxidase complex (ghr.nlm.nih.gov/gene/*CYCS*). It is an electron carrier protein. The oxidized form of the cytochrome c haeme group can accept an electron from the haeme group of the cytochrome c1 subunit of cytochrome reductase. Cytochrome c then transfers this electron to the cytochrome oxidase complex, the final protein carrier in the mitochondrial electron-transport chain (www.uniprot.org).

Mutations in this gene are associated with autosomal dominant non-syndromic thrombocytopenia (ghr.nlm.nih.gov/gene/*CYCS*). Numerous processed pseudogenes of this gene are found throughout the human genome. (Kim, D. P. *et al.*, 2012). Thrombocytopenia 4 (THC4) is defined by a decrease in the number of platelets in circulating blood, resulting in the potential for increased bleeding and decreased ability for clotting (OMIM Entry).

Also *CYCS* gene plays a role in apoptosis. Suppression of the anti-apoptotic members or activation of the pro-apoptotic members of the Bcl-2 family leads to altered mitochondrial membrane permeability resulting in release of cytochrome c into the cytosol. Binding of cytochrome c to Apaf-1 triggers the activation of caspase-9, which then accelerates apoptosis by activating other caspases (www.uniprot.org).

Cytogenetic Location:

The gene is located on chromosome 7p15.3, which is the short (p) arm at position 15.3, Figure (1).



CYC gene

Figure (1) Cytogenetic Location of *CYC* gene in chromosome 7 (ghr.nlm.nih.gov/gene/*CYCS*).

This study aimed to investigate the effect of non-synonymous SNPs (ns SNPs) of *CYCS* gene in protein function and structure using different computational software.

MATERIAL AND METHODS

SNPs were obtained from the National Center for Biological Information (NCBI) SNPs database (2020). Different software for *CYCS* gene analysis were used for prediction of the effect of the nsSNPs on the structure, function and stability of the protein. These included GeneMANIA, SIFT, Provean, Polyphen-2, I-Mutant, MUpro, SNPs & Go and Project HOPE.

GeneMANIA (http://www.genemania.org)

This is a web interface that helps predicting the function of genes and gene sets. GeneMANIA 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. GeneMANIA can be used to find new members of a pathway or complex, finds additional genes which might have been missed in the screen (Warde-Farley, D. *et al*., 2010).

Sorting intolerant from tolerant (SIFT)
(http://sift.jcvi.org/www/SIFT_chr_coords_submit.html)

SIFT predicts the tolerated and deleterious SNPs and identifies the impact of amino acid substitution on protein function and phenotype alterations, so that users can prioritize substitutions for further studies. The main principle of this program is that it generates alignments with a large number of homologous sequences and assigns scores to each residue ranging from zero to one. The threshold in tolerance score for SNPs is 0.05 or less. The input for this software were nsSNPs obtained from NCBI (Ng, P. C., & Henikoff, S. 2003).

Provean
(http://provean.jcvi.org/protein_batch_submit.php?species=humn)

This tool provides PROVEAN prediction for all human proteins and variants. It also shows SIFT predictions when precomputed scores are available. The input for this software were nsSNPs obtained from SIFT and the output was a prediction whether these nsSNPs were deleterious or neutral software (González-Pérez, A., & López-Bigas, N. 2011).

Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/)

Polyphen-2, is an online bioinformatics program which automatically predicts 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 substitution. This program basically 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 the 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 benign, possibly damaging or probably damaging, according to the posterior probability intervals (0, 0.2), (0.2, 0.85) and (0.85, 1), respectively, nsSNPs that were predicted to be intolerant by SIFT has been submitted to Polyphen as protein sequence in FASTA format obtained from UniprotKB/ExPASY after submitting the relevant ensemble protein (ESNP) there. The position of mutation should be entered together with the native amino acid and the new substituent for both structural and functional predictions (Venselaar, H. *et al*., 2010).

I-Mutant (http://gpcr2.biocomp.unibo.it/cgi/)

I-Mutant version 3.0 was used to predict protein stability changes in 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 and the position and type of the mutation. The output is either increased or decreased stability (Bava, K. A. *et al*., 2004; Capriotti, E. *et al*., 2005).

Mu Pro (http://mupro.proteomics.ics.uci.edu/)

It is a set of machine learning programs to predict how single-site amino acid mutation affects protein stability. It is developed based on two machine learning methods: Support Vector Machines and Neural Networks (Cheng, J. *et al*., 2006). The output is either increased or decreased stability.

SNPs & Go (http://snps.biofold.org/snps-and-go//snps-and-go.html)

(SNPs &Gene Ontology), It is software that predicts the disease related mutations from protein FASTA sequence. Its output is prediction of results based on the discrimination among: disease related and neutral variations of protein sequence. The probability score higher than 0.5 reveals the disease related effect of mutation (Capriotti, E. *et al*., 2005).

Project HOPE (http://www.cmbi.ru.nl/hope/home)

Project HOPE is an easy-to-use web server that analyses the structural effects of intended mutation. 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 (Venselaar, H. *et al*., 2010).

RESULTS AND DISCUSSION

Using GeneMANIA software 20 genes were co-expressed with *CYCS* gene the results were shown in Figure (2).

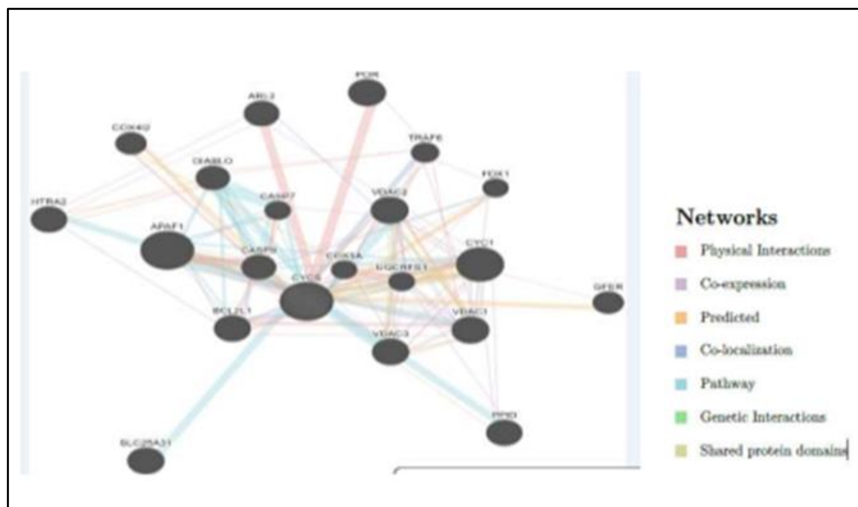


Figure 2, GeneMANIA results showing physical interaction, Co-expression, Predicted, Co-localization, Pathway, Genetic interaction and Shared protein domains

A total 100 Single Nucleotide Polymorphisms were retrieved from National Center of Biotechnology Information (NCBI). From these 68 were in the 3UTR, 28 in the 5UTR and 14 in the coding region (nsSNP). Only these 14 were selected for further investigations. Six nsSNPs were found to be deleterious while 8 were tolerated by SIFT. Using Provean 11 nsSNPs were found to be deleterious while 3 were Neutral. For PolyPhen-2, 5 nsSNPs were observed to be damaging while 9 were benign (Table 1).

Table 1, SIFT, Provean and Poolyphenn-2 results

SNPs	**Amino Acid Change**	**SIFT Score**	**SIFT Prediction**	**Provean P Prediction (cutoff=-2.5)** - 2.5)	**Provean Score**	Polyphen-2 **Prediction**	Polyphen-2 **Score**
rs11548772	A51P	0.111	Tolerated	Neutral	-0.343	Benign	0.000
rs11548778	M81T	0.02	Deleterious	Deleterious	-5.731	Benign	0.165
rs11548795	K56R	0.068	Tolerated	Deleterious	-2.768	Benign	0.053
rs11548796	P31S	0.01	Deleterious	Deleterious	-7.847	Possibly Damaging	0.575
rs11548798	K26M	0.048	Deleterious	Deleterious	-5.087	Possibly Damaging	0.939
rs11548799	T29A	0.199	Tolerated	Neutral	-2.476	Benign	0.000
rs11548807	A44T	0.209	Tolerated	Deleterious	-3.261	Benign	0.058
rs11548818	P77H	0	Deleterious	Deleterious	-7.691	Possibly Damaging	0.744
rs11548820	Y75C	0.079	Tolerated	Deleterious	-8.175	Benign	0.040

rs17851278	C18Y	0.017	Deleterious	Deleterious	-10.444	Probably Damaging	0.960
rs113106786	M13T	0.606	Tolerated	Neutral	-0.467	Benign	0.000
rs121918552	G42S	0.002	Deleterious	Deleterious	-5.904	Possibly Damaging	0.882
rs146569499	I12F	0.051	Tolerated	Deleterious	-3.157	Benign	0.000
rs367666281	T103A	0.072	Tolerated	Deleterious	-3.659	Benign	0.002

For I-Mutant, 12 nsSNPs were found to decrease the stability of the protein and 2 nsSNPs were found to increase the stability of the protein. Also for MUpro, the 14 nsSNPs were found to decrease the stability of the protein(Table 2)

Table 2, I-Mutant 0.3 & MUpro results

SNPs	**Amino Acid Change**	I-Mutant 0.3 **SVM2 Prediction**	**DDG Value Prediction Kcal/mol**	**RI**	MUpro **Prediction**	detal delta G
rs11548772	A51P	Decrease	**-0.59**	**2**	Decrease	-0.95451101
rs11548778	M81T	**Decrease**	**-1.18**	**8**	Decrease	-1.0269979
rs11548795	M 56R	**Decrease**	**-0.31**	1	Decrease	-1.1385467
rs11548796	P31S	**Decrease**	-1.66	7	Decrease	-0.95203726
rs11548798	K26M	Increase	**0.41**	5	Decrease	-0.048660445
rs11548799	T29A	**Decrease**	**-0.62**	6	Decrease	-0.7241127
rs11548807	A44T	Increase	**-0.13**	2	Decrease	-0.73097298
rs11548818	P77H	**Decrease**	**-0.66**	8	Decrease	-1.6784456
rs11548820	Y75C	**Decrease**	**-0.59**	4	Decrease	-0.60710436
rs17851278	C18Y	**Decrease**	**-0.98**	3	Decrease	-1.3336257
rs113106786	M13T	**Decrease**	**-0.15**	1	Decrease	-2.1701814
rs121918552	G42S	**Decrease**	**-1.04**	6	Decrease	-0.53414758
rs146569499	I12F	**Decrease**	**-0.24**	6	Decrease	-1.0970067
rs367666281	T103A	**Decrease**	**-1.74**	9	Decrease	-0.99810303

Using PHD and SNP&GO 9 and 6 nsSNPs were found to be disease related for the two software respectively and 5, 8 were neutral for the same software(Table 3).

Table 3, PHD-SNPs and SNP&GO results

Amino Acid Change	PhD-SNP **Prediction**	**RI**	**Probability**	**SNPs&GO Prediction**	**RI**	**Probability**
I12F	**Neutral**	**4**	0.318	**Neutral**	**8**	0.115
M13T	**Neutral**	8	0.116	**Neutral**	9	0.051
C18Y	**Disease**	10	0.978	**Disease**	9	0.927
K26M	**Neutral**	3	0.344	**Neutral**	8	0.112
T29A	**Neutral**	4	0.320	**Neutral**	7	0.146
P31S	**Disease**	7	0.871	**Disease**	6	0.790
G42S	**Disease**	7	0.863	**Disease**	5	0.760
A44T	**Disease**	0	0.508	**Neutral**	7	0.160
A51P	**Disease**	6	0.805	**Neutral**	3	0.359
K56R	**Disease**	1	0.544	**Neutral**	4	0.295

Y75C	**Disease**	7	0.861	**Disease**	5	0.748
P77H	**Disease**	8	0.897	**Disease**	4	0.700
M81T	**Disease**	8	0.911	**Disease**	7	0.841
T103A	**Neutral**	6	0.188	**Neutral**	8	0.084

The Effect of the Mutation on the Protein Using Project Hope Software

The rs17851278, C18Y, The mutant residue is bigger than the wild-type residue. The wild-type residue was buried in the core of the protein 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 (BMC Bioinformatics. 2010).

The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively. The mutated residue is located in a domain that is important for the activity of the protein and in contact with another domain that is also important for the activity. The interaction between these domains could be disturbed by the mutation, which might affect the function of the protein. The results were shown in Figure 3.

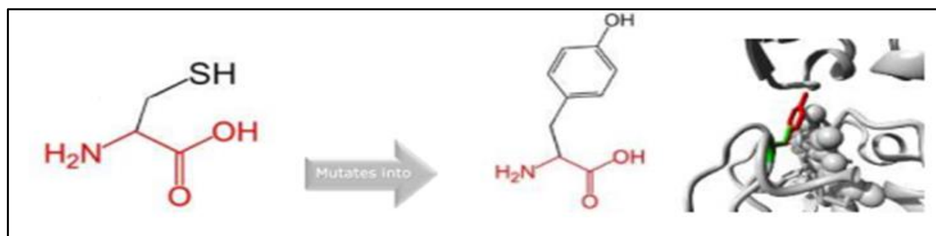


Figure 3, rs178512780, Cysteine into a Tyrosine at position 18.

For the rs11548796, P31S, The mutant residue is smaller than the wild-type residue. causing an empty space in the core of the protein. The hydrophobicity of the wild-type and mutant residue differs. The mutation will cause loss of hydrophobic interactions in the core of the protein. The wild-type residue is a proline. Prolines are known to be very rigid and therefore induce a special backbone conformation which might be required at this position. The mutation can disturb this special conformation. Based on conservation scores this mutation is probably damaging to the protein. The mutated residue is located in a domain that is important for the activity of the protein and in contact with another domain that is also important for the activity. The interaction between these domains could be disturbed by the mutation, which might affect the function of the protein (BMC Bioinformatics. 2010). The results were shown inn Figure 4.

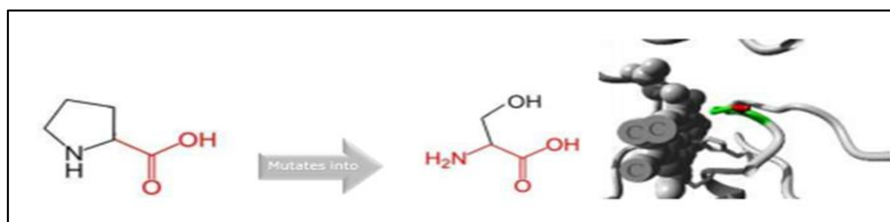


Figure 4, rs11548796, Proline into a Serine at position 31

The**,** rs121918552, G42S, The wild-type residue is a glycine, the most flexible and conserved of all residues. This flexibility might be necessary for the protein's function. Mutation of this glycine can abolish this function. 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. (BMC Bioinformatics. 2010). 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 results were shown in Figure 5.

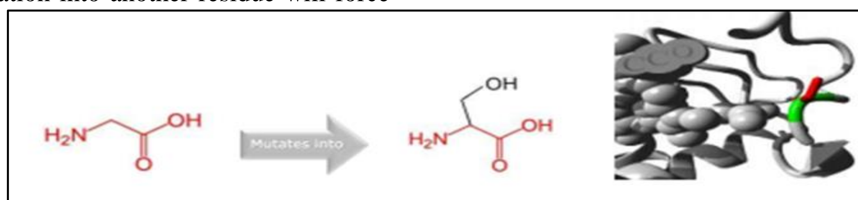


Figure 5, rs121918552, Glycine into a Serine at position 42.

For the rs11548820, Y75C, The mutant residue is smaller than the wild-type residue. The mutation will cause an empty space in the core of the protein. The hydrophobicity of the wild-type and mutant residue differs. The mutation will cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding. The wild-type residue is much conserved, but a few other residue types have been observed at this position too, the mutant residue was not among the other residue types observed at this position in other, homologous proteins. However, residues that have some properties in common with your mutated residue were observed. This means that in some rare cases your mutation might occur without damaging the protein. The mutation showed benign effect with Polyphen-2 software. (Table 1).Results were shown in Figure 6.

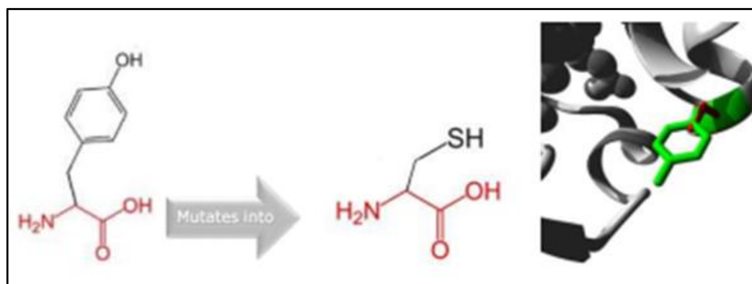


Figure 6, rs11548820, Tyrosine into a Cysteine at position 75.

For the rs11548818, P77H, the mutant residue is bigger than the wild-type residue. 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. The hydrophobicity of the wild-type and mutant residue differs. The mutation might cause loss of hydrophobic interactions with other molecules on the surface of the protein. The wild-type residue is a proline, conserved, and is known to be very rigid and therefore induce a special backbone conformation which might be required at this position. The mutation can disturb this special conformation (BMC Bioinformatics. 2010). Results were shown in Figure 7.

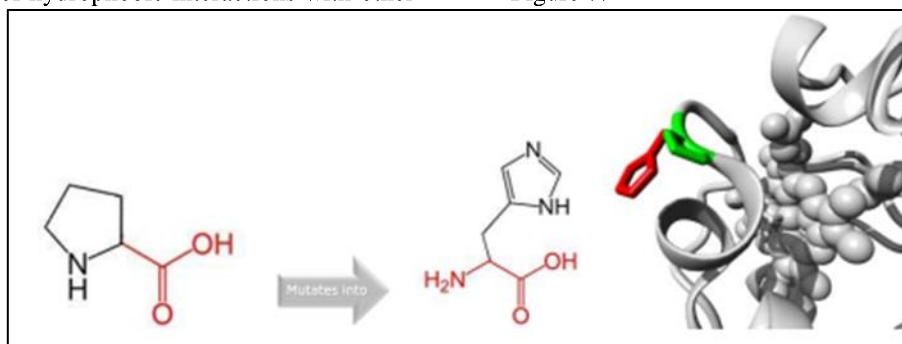


Figure 7, rs11548818, Proline into a Histidine at position 77

In case rs115487780, M81T, The mutation will cause an empty space in the core of the protein. The mutation will cause loss of hydrophobic interactions in the core of the protein. The wild-type was involved in a metal-ion contact. The size differences between the wild-type and mutant residue disturb the interaction with the metal-ion: "Iron (heme axial ligand)" (BMC Bioinformatics. 2010). The results were shown in Figure 8.

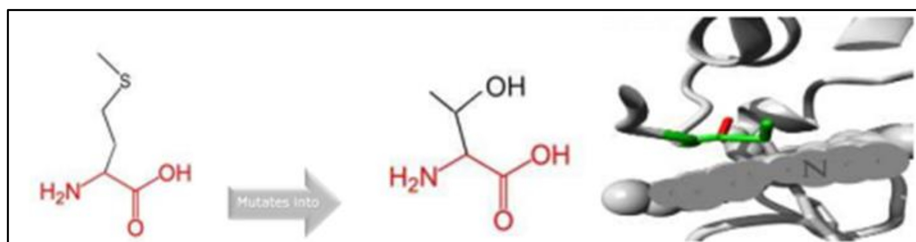


Figure 8, rs115487780 Methionine into a Threonine at position 81

The rs121918552, G42S was identified at 6-generation pedigree segregating autosomal dominant transmission of thrombocytopenia by Morison *et al*. (2008). All affected family members had a gly42-to-ser substitution in the *CYCS* gene (Morison, I. M. *et al*., 2008).

Also in 4 affected members of an Italian family with THC4, De Rocco *et al*. (2017) identified a heterozygous missense mutation in the *CYCS* gene (Y48H; 123970.0002). In vitro studies in yeast and murine knockout cells showed that both the Y48H and G42S mutations reduced the mitochondrial respiratory rate and increased apoptosis (De Rocco, D. *et al*., 2017).

CONCLUSION

In conclusion broad functional and structural analyses were carried out to predict possibly damaging and deleterious nsSNPs of *CYCS* gene using bioinformatics software and *in silico* methods. In the study, 6 high confidence damaging nsSNPs are identified from 14 nsSNPs. Although bioinformatics tools have their limitations, the results from this study may be useful in future for further population based research activities and towards development of medicines.

Acknowledgments

The authors would like to thank the staff of the Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan for facilitating this work. The authors did not receive any funding for this study.

Competing Interests

The authors declare that they have no competing interests.

REFERENCES

1. Genetic Home Reference 2017 https://ghr.nlm.nih.gov/gene/*CYCS*
2. www.uniprot.org › UniprotKB. Cytochrome c - Ricinus communis (Castor bean) – UniprotKB
3. Kim D. Pruitt, Tatiana Tatusova, Garth R. Brown, Donna R. Maglott. Nucleic Acids Res. 2012 ; 40(Database issue): D130–D135.
4. OMIM Entry - *# 612004* - THROMBOCYTOPENIA 4; THC4.
5. Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., ... & Maitland, A. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research*, *38*(suppl_2), W214-W220.
6. Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research*, *31*(13), 3812-3814.
7. González-Pérez, A., & López-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *The American Journal of Human Genetics*, *88*(4), 440-449.
8. Venselaar, H., te Beek, T. A., Kuipers, R. K., 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.
9. Bava, K. A., Gromiha, M. M., Uedaira, H., Kitajima, K., & Sarai, A. (2004). ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic acids research*, *32*(suppl_1), D120-D121.
10. Capriotti, E., Fariselli, P., Calabrese, R., & Casadio, R. (2005). Predicting protein stability changes from sequences using support vector machines. *Bioinformatics*, *21*(suppl_2), ii54-ii58.
11. Cheng, J., Randall, A., & Baldi, P. (2006). Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins: Structure, Function, and Bioinformatics*, *62*(4), 1125-1132.
12. Capriotti, E., Fariselli, P., Calabrese, R., & Casadio, R. (2005). Predicting protein stability changes from sequences using support vector machines. *Bioinformatics*, *21*(suppl_2), ii54-ii58.
13. Venselaar, H., te Beek, T. A., Kuipers, R. K., 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.
14. BMC Bioinformatics. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *8;11(1),*548. DOI: 10.1186/1471-2105-11-548. PubMed: 21059217.
15. Morison, I. M., Bordé, E. M. C., Cheesman, E. J., Cheong, P. L., Holyoake, A. J., Fichelson, S., ... & Fagerlund, R. D. (2008). A mutation of human cytochrome c enhances the intrinsic apoptotic pathway but causes only thrombocytopenia. *Nature genetics*, *40*(4), 387-389.
16. De Rocco, D., Melazzini, F., Marconi, C., Pecci, A., Bottega, R., Gnan, C., ... & Heller, P. G. (2017). Mutations of RUNX1 in families with inherited thrombocytopenia. *American journal of hematology*, *92*(6), E86.