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

ISSN 2349-8870
Volume: 7
Issue: 3
27-32
Year: 2020

PREDICTION OF STRUCTURAL AND FUNCTIONAL EFFECTS OF SINGLE NUCLEOTIDE POLYMORPHISMS IN *PAX1* GENE ASSOCIATED WITH KLIPPEL–FEIL SYNDROME

Lana Salah Mohamed Salih[1,2], Afra M. AlBkrye[1], Hind Abdelaziz Elnasri[1] and Mona A. M. Khaier*[1]

[1]Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan.
[2]Royal Care Hospital, Ministry of Health, Khartoum, Sudan.

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

Article Received on 29/12/2019 **Article Revised on 19/01/2020** **Article Accepted on 09/02/2020**

ABSTRACT
Background: The *PAX1* gene is a member of paired Pax family that develops control genes that which encodes transcription factor. This gene plays critical roles during fetal development and during embryogenesis and essential for development of the vertebral column. It is located in chromosome 20 at position 11.22 causing Klippel–Feil Syndrome. **Materials and methods**: The nsSNPs of *PAX1* gene were obtained from NCBI dbSNP database and were analyzed using computational bioinformatics tools. They were analyzed by (SIFT, Polyphen-2, Provean, I- mutant, SNPs& GO, PHD, GeneMANIA and Project Hope software). **Results**: The total number of SNPs collected from NCBI database were 4148, 164 SNPs were in the coding region, 38 in the 3´UTRs, 3 in the 5´UTRs. Only SNPs that were found in the coding region were analyzed. Four nsSNPs were found to be deleterious and having high score in all software used (rs143731938, rs147752664, rs199692693, rs372580256), they were also predicted to change protein stability.

INTRODUCTION
Klippel–Feil syndrome (KFS) is a condition characterized by failed segmentation of the cervical vertebrae with the clinical of a short, immobile neck and a low posterior hair line.[1] Vertebral fusions may also occur elsewhere along the spine and other vertebral anomalies such as hemi vertebrae may be present.[2] Other features include Sprengel's shoulder, renal, cardiac and neurological abnormalities.[3] KFS appears to be an etiologically heterogeneous condition with sporadic occurrence, autosomal dominant and autosomal recessive modes of inheritance that have been reported.[4]

*PAX*1 gene is a member of the paired (PAX) family of transcription factor. These genes play critical roles during fetal development and also play a role in pattern formation during embryogenesis and may be essential for development of the vertebral column.[5] This gene is silenced by methylation in ovarian and cervical cancers and may be a tumor suppressor gene.[6] Mutations in this gene are also associated with vertebral malformations.[7]

The most conserved functional motif in all PAX proteins is the 128 amino-acid paired domain, which exhibits DNA-binding activity. Several defects occur due to deletion and insertion of functional motifs in this gene.[8]

The homology between mouse and human mutants of *PAX* gene has been a useful tool in defining conditions in man. However, the human homologue of the first mouse *PAX1* mutant *undulated* (*un*) remains elusive.[9] (*Un)* is caused by a missense mutation in the paired box of *PAX1,* which decreases the DNA-binding affinity of the protein and alters its DNA-binding specificity.[10]

This study aimed to use computational and bioinformatics tools to identify and evaluate nsSNPs in human *PAX1* gene that might affect the function and stability of PAX protein.

Cytogenetic Location: 20p11.22, which is the short (p) arm of chromosome 20 at position 11.22.

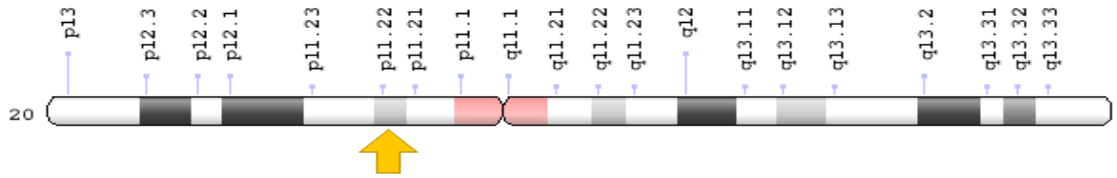


Figure (1): Cytogenetic Location of *PAX1* gene.[11]

MATERIAL AND METHODS

Computational methods were used to detect nsSNPs in *PAX1* gene, eight software were used to analyze SNP from coding region.

Sorting Intolerant From Tolerant (SIFT) (http://sift.jcvi.org/www/SIFT_dbSNP.html

It is a multistep algorithm that predicts whether an amino acid substitution would affect protein function or not, by filtering out the mutations based on tolerance score. SNPs data were retrieved from dbSNPs as an rsIDs, then submitted to the SIFT server for analysis. The SIFT server reports the results as prediction scores between 0 and 1. A score within a range of 0–0.05 is considered to be deleterious or pathogenic, whereas scores above 0.05 to 1 are considered to be neutral or non-pathogenic.[12]

Polyphen-2(Polymorphic Phenotyping-2)
(http://genetics.bwh.harvard.edu/pph2/dbsearch.shtml)

It predicts the possible impacts of amino acid substitution on stability and function of human protein using structure and comparative evolutionary consideration.

Prediction is based on number of sequence, phylogenetic and structural features.

Polyphen-2 uses the UniprotKB database as reference source for all protein sequence. The queries were submitted to the polyphen2 server in the form of dbSNP IDs or sequence. The output levels would be appraised qualitatively as benign, possibly damaging (less confident prediction) and probably damaging (more confident prediction), If the Polyphen-2 score is greater than or equals 0.5, it can be classified as deleterious, and if the score is less than 0.5 it can be regarded as benign or tolerated.[13]

(**Pro**tein **V**ariation **E**ffect **An**alyzer)**PROVEAN**

It is a software tool which predicts whether an amino acid substitution or indel has an impact on the biological function of a protein. Single or multiple amino acid deletions cut off is -2.282 considered neutral l>2.282considered deleterious.[15]

(**S**ingle **N**ucleotide **P**olymorphism & **G**ene **O**ntology)**SNPs&GO**

Classifies the nsSNP as disease related or neutral. It uses information derived from protein sequence database, protein 3D structure, protein sequence profile, protein function and gene ontology annotation. Score>0.5 indicate that the mutation is disease related.[16]

(**P**rediction of Human **D**eleterious **s**ingle **N**ucleotide **P**olymorphism)**PHD-SNP**
(http://snps.biofold.org/phd-snp/phd-snp.html).

Classifies nsSNP according to the ratio between the frequencies of the wild type and substituted amino acid. For both software probability>0.5 indicates that the mutation is disease related.[17]

SNPs & GO

This software classified the nsSNPs as disease related or neutral according to the information derived from (protein sequence, 3D structure, protein function).

I-MUTANT
(http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi)

Predicting stability changes upon mutation from the protein sequence or structure.

It is a neutral network based web server for automatic prediction of protein stability change upon single site mutation.[18]

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

This software helps to predict the function of the desired gene and gene sets, and to construct gene –gene function, interaction network from gene list. The result includes most related genes to the original list, and functional annotation from gene ontology.[18]

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

Hope is an online web and easy to use, analyze the structural effects of point mutation in a protein sequence. The input is the protein sequence and the mutation. HOPE will collect and combine available information from a series of web services and database and will produce a report.[19]

RESULTS

Different bioinformatics tools were utilized to predict the effects of nsSNPs. The total number of nsSNPs were 164. Using SIFT software10 deleterious nsSNPs were predicted and the rest were tolerated while with Polyphen-2, 6 nsSNPS were predicted as possibly damaging with high score and 2 were benign. PROVEAN software was used also and resulted in 6 deleterious nsSNPs and 4 Neutral.

Table (1): Result of SIFT, POLYPHEN-2, PROVEN software.

SNP	Amino Acid Change	Protein ID	SIFT Score	SIFT Prediction	Polyphen- 2	Polyphen-2 score	Proven prediction	Provean Score
rs17861059	P453L	ENSP00000381499	0.029	Deleterious	Benign	0.005	0.39	Neutral
rs113116112	P471Q	ENSP00000381499	0.009	Deleterious	Benign	0.006	0.27	Neutral
rs142565607	K224R	ENSP00000381499	0.022	Deleterious	possibly damaging	0.49	-2.89	Deleterious
rs143731938	G377C	ENSP00000381499	0.042	Deleterious	possibly damaging	1	-7.98	Deleterious
rs147752664	G297D	ENSP00000381499	0.001	Deleterious	possibly damaging	1	-6.72	Deleterious
rs199692693	R222C	ENSP00000381499	0	Deleterious	possibly damaging	1	-7.98	Deleterious
rs201542749	P392R	ENSP00000381499	0.017	Deleterious	possibly damaging	1	-1.48	Neutral
rs367842641	Y154F	ENSP00000381499	0.022	Deleterious	possibly damaging	1	-3.86	Deleterious
rs368447637	T445M	ENSP00000381499	0.017	Deleterious	possibly damaging	0.89	-0.25	Neutral
rs372580256	S291L	ENSP00000381499	0.022	Deleterious	possibly damaging	1	-5.06	Deleterious

I-MUTANT

The results here were classified into decrease or increase stability of protein with prediction and scoring. Only 10nsSNPs were found as decreasing the stability.

Table (2): Results of I-MUTANT Software.

SNP	AMINO ACID CHANGE	PROTEIN ID	I-Mutant prediction	I-Mutant score
rs17861059	P453L	ENSP00000381499	Decrease	5
rs113116112	P471Q	ENSP00000381499	Decrease	5
rs142565607	K224R	ENSP00000381499	Decrease	7
rs143731938	G377C	ENSP00000381499	Decrease	4
rs147752664	G297D	ENSP00000381499	Decrease	6
rs199692693	R222C	ENSP00000381499	Decrease	4
rs201542749	P392R	ENSP00000381499	Decrease	3
rs367842641	Y154F	ENSP00000381499	Decrease	7
rs368447637	T445M	ENSP00000381499	Decrease	2
rs372580256	S291L	ENSP00000381499	Decrease	1

SNPs & GO

This software classified the nsSNPs as disease related or neutral according to the information derived from (protein sequence, 3D structure, protein function). So according to this, 5 nsSNPs out of 10 were found as disease related.

Table (3): SNP and GO Software results.

SNP	Amino Acid Change	SNP&GO Prediction	SNP&GO RI	SNP&GO Probability	PHD prediction	PHD RI	PHD probability
rs17861059	P453L	Neutral	10	0.014	Neutral	8	0.95
rs113116112	P471Q	Neutral	6	0.194	Disease	3	0.639
rs142565607	K224R	Neutral	6	0.208	Neutral	2	0.375
rs143731938	G377C	Disease	2	0.578	Disease	8	0.913
rs147752664	G297D	Disease	3	0.649	Disease	8	0.899
rs199692693	R222C	Disease	6	0.817	Disease	8	0.899
rs201542749	P392R	Neutral	8	0.082	Neutral	1	0.46
rs367842641	Y154F	Neutral	4	0.306	Disease	5	0.764
rs368447637	T445M	Neutral	8	0.081	Neutral	1	0.456
rs372580256	S291L	Disease	4	0.709	Disease	7	0.826

GeneMANIA

Using this software 20 genes were co expressed.

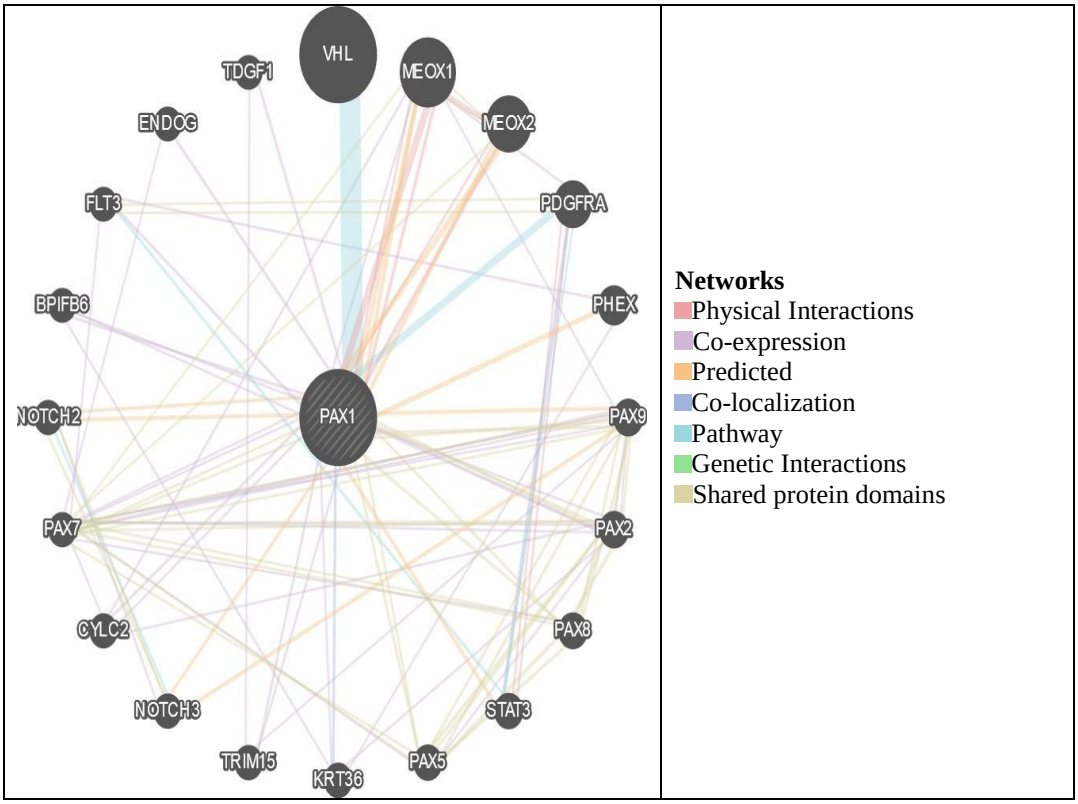
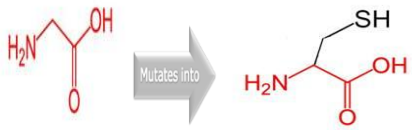
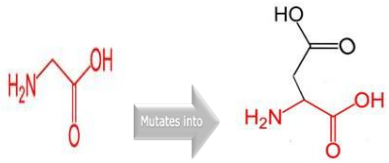
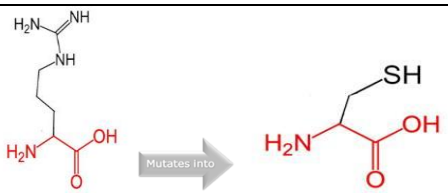


Figure (1): Related genes with *PAX1* gene by GeneMANIA software.

Project Hope

This software analyzes the structure that affect point mutation in protein sequence and function.

SNPs	Amino acid change	Amino acid structure of wild and mutant type
rs143731938	G377C	 Change in the amino acid glycine into cysteine.
rs147752664	G297D	 Change in amino acid glycine into aspartic acid
rs199692693	R222C	 Change in amino acid arginine into cysteine

rs372580256	S291L	 Change in amino acid Serine into leucine.

Figure (2): project hope results in *pax1* gene show position of mutation.

DISCUSSION

All nsSNPs were subjected to eight different prediction algorithms to investigate whether these SNPs have any effect on the structure, function and stability of *PAX1*gene. Non synonymous SNPs were analyzed by SIFT, 10 were deleterious and 154 were tolerated. For Proven software, 6nsSNPs were deleterious and 4 were neutrals. These deleterious nsSNPs were analyzed using PolyPhen-2 software to predict the damaging SNPs, we found that 8nsSNPs were predicted to be possibly damaging and only 2 nsSNPS were predicted as benign. The, I-*mutant* software used to detect the stability of protein we found that 10 nsSNPs were scored decreased in the stability of the protein. For the prediction to be disease related by SNP & Go software, only 4 SNPs were reported as disease related, PHD- software scored 6nsSNPsas disease related. Finally 4 nsSNPs out of 164 nsSNPs met the criteria and were classified as highly damaging. They were taken as an input for Project Hoped software to check for the size, hydrophobicity, charge and the 3D structure, the were shown in Figure 2. The rs143731938, the change in amino acid glycine into cysteine (at postion377), the wild type glycine is a most flexible of all residues these flexibilities necessary for the protein's function so the mutant type might disrupt this function, the size of mutant type is bigger which lead to bumps. The mutation into another residue make incorrect conformation and will disturb the local structure as in rs147752664 which the glycine converted into Aspartic acid (at position 297) and also the mutant type introduce charge this can cause repulsion of ligands. The rs199692693 convert arginine into cysteine, the mutant is smaller, neutral and more hydrophobic which lead to incorrect folding or loss of hydrogen bounds. In the last rs372580256 the substitution serine into leucine (at position 291) the mutant is bigger and more hydrophobic. The mutant type located in a domain that important for many activities so the mutation might disturb this function. There are many SNPs that affected *PAX1* gene on structure and function that will be damaging and most of them were not.

CONCLUSIONS

Computational biology tools are very powerful especially when provided with good data and used by experts. This study concluded that only 4 nsSNPs out of 164 nsSNPs met the criteria of the 8 software used and were classified as deleterious and highly damaging. These nsSNPS can be used for the diagnosis.

Recommendations

More lab experimentation was needed to prove the effect of nsSNPs on structure and function of the protein. To draw the full picture for the effect of the SNPs on the disease the 3UTRs, 5UTRs SNPs must be included in the study.

REFERENCES

1. Nancy McLaughlin, Alexander G. Weil, Jacques Demers1, Daniel Shedid. Klippel–Feil syndrome associated with a craniocervico-thoracic dermoid cyst. *Surgical Neurology International.,* 2013; 4: s61-66.
2. Alessio Pirino, Maria Alessandra Sotgiu, Andrea Montella, Erich Cosmi's,. Association of Klippel-Feil syndrome, Dandy-Walker malformation, spina bifida: A case report. *Clinical Orthopaedics and Related Research*., 2019; 14(3): 415-418.
3. J M McGaughran, A Oates, D Donnai, A P Read, M Tassabehji,. Mutations in *PAX1* may be associated with Klippel–Feil syndrome. *European Journal of Human Genetics.*, 2003; 11: 468–474.
4. Karaca, Ender, Yuregir, Ozge O, Bozdogan, Sevcan T, Aslan, Huseyin, Aslan, Huseyin. 2015. Rare variants in the notch signaling pathway describe a novel type of autosomal recessive Klippel–Feil syndrome. *AMRICAN JOURNAL OF MEDICAL GENETICS., 167(11): 2795-2799.*
5. Genetics Home Reference. From NCBI GENE. 2012.
6. Jun Xu, [1]Ling Xu, [1]Baohua Yang, [1]Lifeng Wang, [1]Xiao Lin, [1]Hong Tu[2] Assessing methylation status of PAX1 in cervical scrapings, as a novel diagnostic and predictive biomarker, was closely related to screen cervical cancer. *International journal of clinical and experimental pathology*, 2015; 8(2): 1674–1681.
7. PF Giampietroa, CL Raggiob, CEReynoldsc, SK Shuklac, EMcPhersona, N Ghebraniousd, FSJacobsene, V Kumarf, TFaciszewskig, RM Paulih, KRasmussena, JK Burmesteri, CZaleskia, S Merchanta,*, D Davidi, JL Weberi, I Glurichiand RD Blan. An analysis of PAX1in the development of vertebral malformations. *Medical Genetic Services*, 2005: 68: 448–453.
8. https://www.ncbi.nlm.nih.gov/
9. JM McGaughran1, A Oates2, D Donnai2, AP Read2and M Tassabehji. Mutations in PAX1 may be associated with Klippel–Feil syndrome. *European Journal of Human Genetics.*, 2003; *11: 468 – 474.*
10. JM McGaughran1, A Oates2, D Donnai2, AP Read2 and M Tassabehji. Mutations in PAX1 may be associated with Klippel–Feil syndrome. *European Journal of Human Genetics.*, 2003; *11: 468 – 474.*
11. Genetic Home Reference.

12. Kumar P., Henikoff S., and Ng P.C., Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols., 2009; 4(7): 1073–1081. pmid:19561590
13. Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A. ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res.*, 2004; 32: 120-121.
14. Capriotti Emidio, Fariselli Piero, and Casadio Rita." I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure." Nucleic acids research, 2005; 33(2): W306–W310.
15. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE, 2012; 7(10): e46688.
16. Thusberg, J., Olatubosun, A. and Vihinen, M. Performance of mutation pathogenicity prediction methods on missense variants. *Human Mutation.*, 2011; 32: 358-368.
17. Capriotti E, Fariselli P, Calabrese R, Casadio R. Predicting protein stability changes from sequences using support vector machines. Bioinformatic., 2005; 21: 54-58.
18. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P,. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *J. Nucleic Acids Res.*, 2010.
19. Adzhubei I. A., Schmidt S., Peshkin L., Ramensky V. E., Gerasimova A., Bork P., A method and server for predicting damaging missense mutations. *Nature methods.*, 2010; 7(4): 248–249. pmid: 20354512.