

Data Mining Analysis of *ESCO2* Gene Single Nucleotide Polymorphisms Associated with Roberts's Syndrome

Sahar Mohamed Ali Mohamed Babiker, Afra Mohamed Al Bkrye, Hind Abdelaziz Elnasri, Mona Abdelrahman Mohamed Khaier*

Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bari, Khartoum, Sudan

Email address:

munakhaier@gmail.com (M. A. M. Khaier)

*Corresponding author

To cite this article:

Sahar Mohamed Ali Mohamed Babiker, Afra Mohamed Al Bkrye, Hind Abdelaziz Elnasri, Mona Abdelrahman Mohamed Khaier. Data Mining Analysis of *ESCO2* Gene Single Nucleotide Polymorphisms Associated with Roberts's Syndrome. *International Journal of Genetics and Genomics*. Vol. 8, No. 1, 2020, pp. 41-47. doi: 10.11648/j.ijgg.20200801.15

Received: December 20, 2019; Accepted: January 13, 2020; Published: January 31, 2020

Abstract: Roberts's syndrome is a genetic disorder characterized by limb and facial abnormalities. Affected individuals also grow slowly before and after birth. This syndrome is associated with *ESCO2* (Establishment of Sister Chromatid cohesion N-acetyltransferase 2) gene mutations. SNPs in the coding region (exonal SNPs) that are non-synonymous (nsSNPs), the SNPs and related ensembles protein (ESNP) were obtained from the SNPs database (dbSNP) for computational analysis. Bioinformatics analysis of *ESCO2* exonal non-synonymous SNPs initiated by GeneMANIA, SIFT, Polyphen-2, PHD, SNP&GO, Provean and ProjtHope. There were 85 nsSNPs, they had been submitted to SIFT software to predict the tolerant and intolerant SNPs, they had been sorted to 65 Tolerated SNPs and 20 Deleterious SNPs. SIFT deleterious SNPs had been tested by polyphen-2 software and the result was 3 benign SNPs, 3 possibly damaging and 14 probably damaging SNPs. The same 20 SNPs were tested using SNP&GO software and gave the same result for PHD and SNP&GO (4 diseased and 16 neutral) and the result obtained when using Provean software was 12 SNPs were neutral while only 8 SNPs were deleterious. The total nsSNPs affecting the structure, function and causing disease in the tested software were 4 nsSNPs (rs80359868, rs146312522, rs200548692, rs373708669) Protein structural analysis was done using all of CPH server, Raptor X, Project HOPE and chimera for the 4 pathological SNPs (W539, C392Y, R427C and D403V) resulted in all function prediction software. and, these results are at use for further researches and studies on this gene and it's mutations.

Keywords: *ESCO2* Gene, Roberts's Syndrome, SIFT, Polyphen-2, PHD, SNP&GO and Provean

1. Introduction

Roberts's syndrome (RBS), (SC phocomelia syndrome or Roberts's tetraphocomelia syndrome) is a genetic disorder characterized by limb and facial abnormalities. Affected individuals also grow slowly before and after birth [1].

Roberts's syndrome is known by limbs abnormalities, hypomelia; particularly in forearms and lower legs. In severe cases, phocomelia occurs. People with Roberts syndrome may also have abnormal or missing fingers and toes, and joint deformities commonly occur at the elbows and knees [2]. Individuals with Roberts's syndrome typically have numerous facial abnormalities, including cleft lip with or without cleft palate, micrognathia, ear abnormalities,

hypertelorism, small nostrils, they may have microcephalic head and in severe cases affected individuals have encephalocele. Infants with severe form of Roberts's syndrome are often stillborn or die shortly after birth. Mildly affected individuals may live up to adulthood [1].

ESCO2 (Establishment of Sister Chromatid cohesion N-acetyltransferase 2); this gene is a member of a conserved protein family that provides instructions for making a protein that is important for proper chromosome separation during S phase of cell division phases [2]. The *ESCO2* protein plays an important role in establishing the glue that attaches the sister chromatids together until the chromosomes are ready to divide [1].

At least 26 mutations have been found to cause Roberts syndrome. All of these mutations prevent the cell from producing any functional *ESCO2* protein, the absence of

functional ESCO2 protein causes some of the glue between sister chromatids to be missing at the chromosome's centromere [1].

ESCO2 gene is located in chromosome 8 [1]. Molecular analysis detected that all *ESCO2* mutations were found in

exon 3, and they were homozygous mutations in affected people with heterozygous mutation in their parents [3].

In this study the aim was to analyze *ESCO2* single nucleotide polymorphisms computationally to predict their association with Roberts's syndrome using in silico methods.

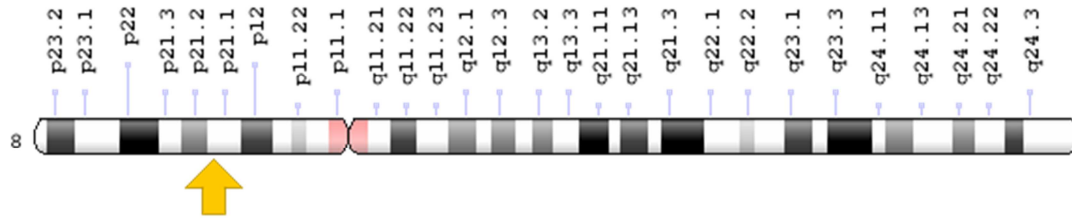


Figure 1. Gene *ESCO2* location on chromosome 8 [3].

2. Material and Methods

The critical step in this work was to select SNPs for analysis by computational software; In this selection the priority was given to SNPs in the coding region (exonal SNPs) that are non-synonymous (ns SNPs), the SNPs and related ensembles protein (ESNP) were obtained from the SNPs database (dbSNP) for computational analysis from <http://www.ncbi.nlm.nih.gov/snp/> and UniprotKB database.

2.1. Function Prediction Software

2.1.1. SIFT (Sorting Intolerant From Tolerant)

Predict whether an amino acid substitution affects protein function based on sequence homology and physical properties of amino acids

(<http://siftdata.org/www/SIFT-dbSNP.html>) [4].

2.1.2. Polyphen-2 (Polymorphism phenotyping v2)

Is a tool which predicts possible impact of amino acid substitution on the structure and function of a human protein using straight forward

Physical and comparative consideration (<http://genetics.bwh.harvard.edu/pph2/>) [5].

2.1.3. SNP&GO (Single Nucleotide Polymorphism & Gene Ontology) Predicting Disease Associated Variations Using GO Terms

SNP&GO is a SVM classifier consisting of a single SVM that takes in input protein sequence profile and functional information (<http://snps.biofold.org/snps-and-go/snps-and-go.html>) [6].

2.1.4. Provean (Protein Variation Effect Analyzer)

Provean is a software tool which predicts whether an amino acid substitution has an impact on the biological function of a protein. (<http://provean.jcvi.org/index.php>) [7].

2.2. Stability Software

2.2.1. I-Mutant Suite -3 (Predictor of the Effects of a Single Point Mutation)

Used to predict protein stability changes upon single point mutation using protein structure or sequence, and can be used to

predict single point mutation disease association from protein sequence (<http://gpcr2.biocomp.unibo.it/cgi/predictors/I-mutant3.0/I-Mutant3.0cgi>) [8]

2.2.2. MUpro (Prediction of Protein Stability Changes for Single Site Mutations from Sequences)

It is a set of machine learning programs to predict how single-site amino acid mutation affects protein stability.

(<http://mupro.proteomics.ics.uci.edu>) [9].

2.3. Function and Co-expression Software: GeneMANIA

Is a web interface that helps predict the function of genes and gene sets. GeneMANIA finds other genes that are related to a set of input gene, using a very large set of functional association data. Associated data include protein and genetic interactions, pathways, co-expressions, co-localizations and protein domains similarity [10]. (<http://www.genemania.org>)

2.4. Structure Prediction Software

2.4.1. Project HOPE (an Online Web Service Where the User Can Submit a Sequence and Mutation and HOPE Will Show the Effect of That m1. Utation Based on Protein Structure Change)

HOPE collects structural information from a series of sources, including calculations on 3D protein structure, sequence annotations in UniprotKB and prediction from Reproof software HOPE combines this information to analyze the effect of a certain mutation on the protein structure.

(HOPE; <http://www.cmbi.ru.nl/hope/html>) [11].

2.4.2. Raptor-X (a Web Portal for Protein Structure and Function Prediction)

This web portal for protein structure and function prediction excelling at secondary, tertiary and contact prediction for protein sequences without close homology in the Protein Data Bank, PDB (<http://raptorX-uchicago.edu/>) [12].

2.4.3. Chimera 1.10.2 (an Extensible Program for Interactive Visualization and Analysis of Molecular Structures and Related Data)

UCSF Chimera is a highly extensible program for

interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles. High-quality images and animations can be generated. (<https://www.cgl.ucsf.edu>) [13].

3. Results and Discussion

3.1. Functional Analysis Results

ESCO2 gene single nucleotide polymorphisms and ensembles protein were obtained from the SNPs database (dbSNP) for computational analysis (<http://www.ncbi.nlm.nih.gov/snp/>) and UniprotKB database are 308 SNPs, which had been filtered to 241 SNPs for only *ESCO2* gene in *Homo sapiens*, then filtered to 114 for the isoform ENSP00000306999, which filtered for

only 85 SNPs representing exonal non-synonymous single nucleotide polymorphism (nsSNPs).

These 85 nsSNPs had been submitted to SIFT software to predict the tolerant and intolerant SNPs, they had been sorted to 65 Tolerated SNPs and 20 Deleterious SNPs. SIFT deleterious SNPs had been tested by polyphen-2 software and the result was 3 benign SNPs, 3 possibly damaging and 14 probably damaging SNPs. The same 20 SNPs were tested using SNP&GO software and gave the same result for PHD and SNP&GO (4 diseased and 16 neutral) and the result obtained when using Provean software was 12 SNPs were neutral while only 8 SNPs were deleterious. The total nsSNPs affecting the structure, function and causing disease in the tested software were 4 nsSNPs (rs80359868, rs146312522, rs200548692, rs373708669) (Table 1).

Table 1. The results of SIFT, Polyphen-2, PHD, SNP&GO and Provean.

SNP	Amino Acid Change	SIFT Prediction	Polyphen-2 prediction	PHD Prediction	SNP&GO prediction	Provean Prediction
rs1380676	P321H	Deleterious	probably damaging	Neutral	Neutral	Neutral
rs80359868	W539G	Deleterious	probably damaging	Disease	Disease	Deleterious
rs115144373	P579T	Deleterious	probably damaging	Neutral	Neutral	Deleterious
rs140062105	I376M	Deleterious	probably damaging	Neutral	Neutral	Neutral
rs143406229	L81F	Deleterious	Benign	Neutral	Neutral	Neutral
rs143530690	K498T	Deleterious	Benign	Neutral	Neutral	Deleterious
rs145248677	R9K	Deleterious	Benign	Neutral	Neutral	Neutral
rs146312522	C392Y	Deleterious	probably damaging	Disease	Disease	Deleterious
rs191457728	D373Y	Deleterious	probably damaging	Neutral	Neutral	Deleterious
rs199665460	K221I	Deleterious	possibly damaging	Neutral	Neutral	Deleterious
rs200548692	R427C	Deleterious	possibly damaging	Disease	Disease	Deleterious
rs202084183	K369E	Deleterious	possibly damaging	Neutral	Neutral	Neutral
rs369056136	S391C	Deleterious	probably damaging	Neutral	Neutral	Neutral
rs373708669	D403V	Deleterious	probably damaging	Disease	Disease	Deleterious

3.2. Stability Results

The stability for substitution proteins predicted using I-Mutant suite3.0 and MUpro software, the results were shown in Table 2.

Table 2. Results of I-Mutant and MUpro.

SNP	Amino Acid Change	I-Mutant stability	I-Mutant RI	DDG Value Kcal/mol	MUpro Prediction
rs1380676	P321H	Decrease	7	-1.35	Decrease
rs80359868	W539G	Decrease	9	-2.14	Decrease
rs115144373	P579T	Decrease	9	-2.14	Decrease
rs140062105	I376M	Increase	4	0.02	Decrease
rs143406229	L81F	Decrease	7	-0.65	Decrease
rs143530690	K498T	Increase	2	-0.09	Decrease
rs145248677	R9K	Decrease	8	-0.75	Decrease
rs146312522	C392Y	Decrease	6	-0.56	Decrease
rs191457728	D373Y	Decrease	2	-0.08	Decrease
rs199665460	K221I	Decrease	4	-0.25	Increase
rs200548692	R427C	Decrease	0	-1.01	Decrease
rs202084183	K369E	Decrease	6	-0.99	Decrease
rs369056136	S391C	Decrease	5	-1.35	Decrease
rs373708669	D403V	Decrease	2	-0.66	Decrease

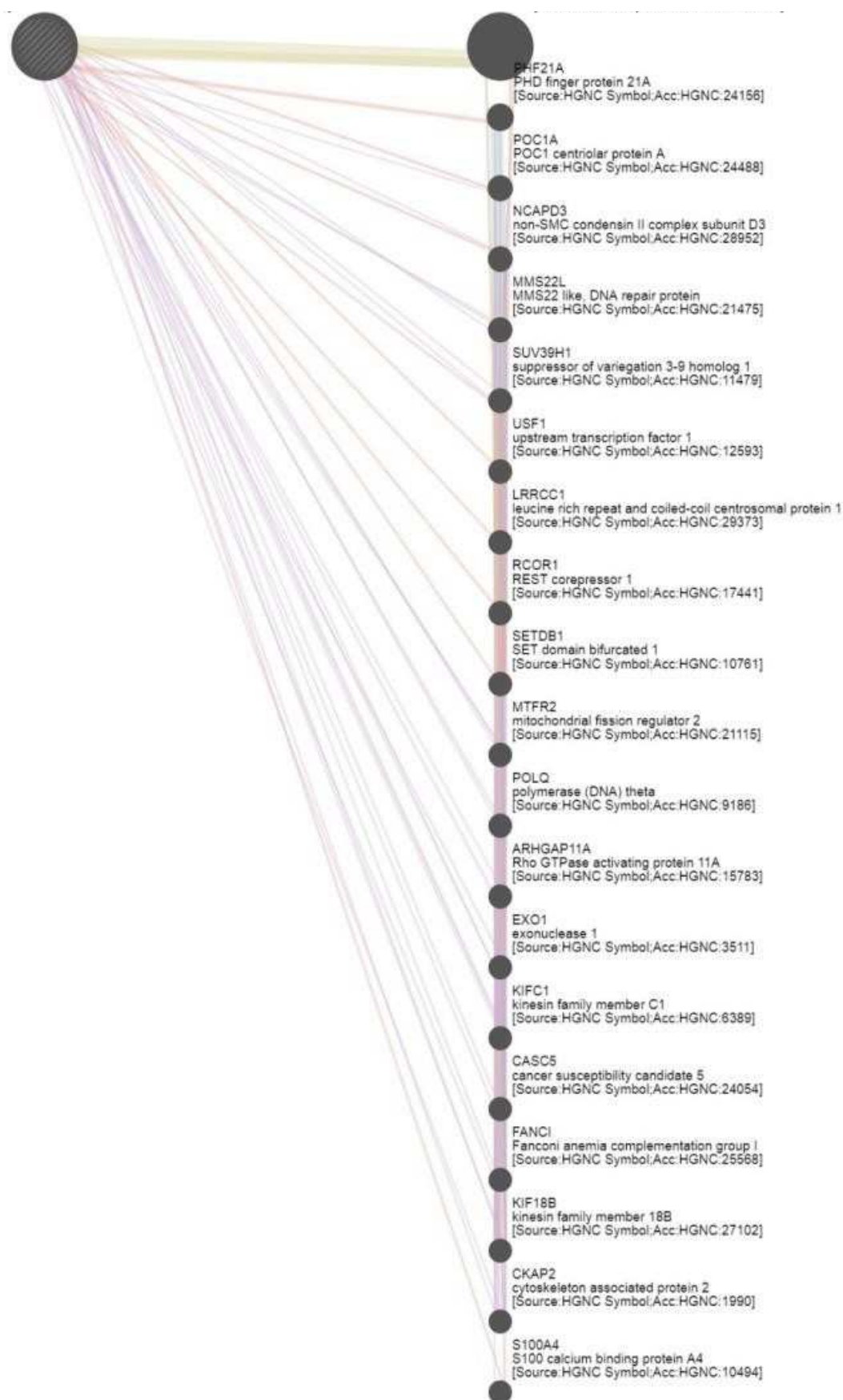
3.3. GeneMANIA Results

1. *ESCO2* shared domain with *ESCO1*
2. *ESCO2* has co-expression with (*POC1A*, *MMS22L*, *MTFR2*, *ARHGAP11A*, *EXO1*, *KIFC1*, *FANCI*, *KIF18B*

and *CKAP2*).

3. *ESCO2* has physical interaction with (*PHF21A*, *POC1A*, *NCAPD3*, *SUV39H1*, *USF1*, *LRRCC1*, *RCOR1*, *SETDB1*, *POLQ*, *CASC5* and *S100A4*).

Shown in figure 2.



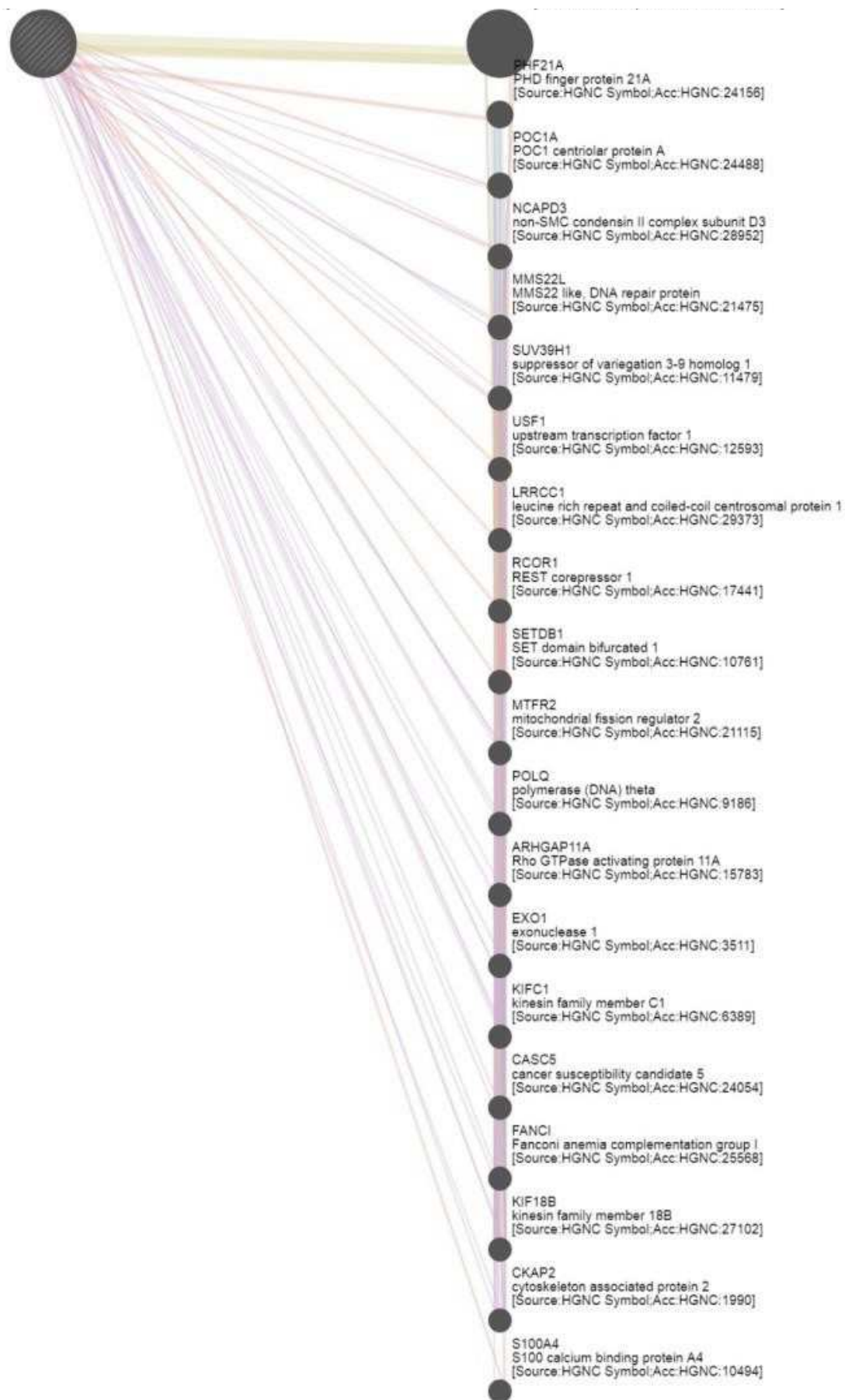


Figure 2. GeneMANIA results.

3.4. Structural Analysis

ESCO2 Protein

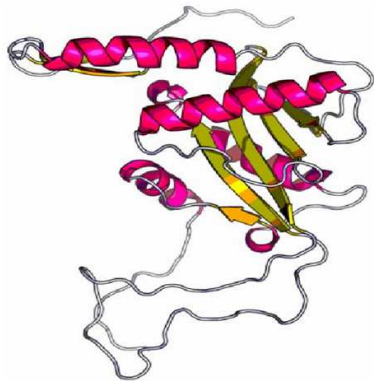


Figure 3. ESCO2 protein by Chimera.

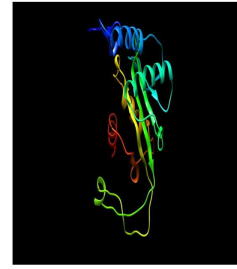


Figure 4. ESCO2 protein by Raptor-X.

To analyze substitution proteins structure the 4 positive mutations had been selected that were associated with disease in all function prediction software, and submitted into project HOPE to get their effect on protein structure and the result for all of them contained no structure image because of structure information lacking but according to the amino acid change the structural change predicted were as in the table 3:

Table 3. Project HOPE structural change result Further analysis for the 4 mutations had been done using chimera after getting the protein secondary structure using Raptor X, the 3D structure.

Mutation	Project hope wild type	Project hope mutant type	Chimera wild type	Chimera mutant type
C 392Y				No image found
D403 V				No image found
R 427 C				
W 539 G				

Table 4. Chimera results.

Amino acid change	SNP	Structural change
C239Y	rs146312522	Loss of hydrophobic interactions bumps, disturb a zinc finger domain
D403V	rs373708669	Disturb a zinc finger domain
R427C	rs200548692	Disturb correct folding (lack of structural information)
W539G	rs80359868	Disturb the rigidity of the protein

When the structure using Raptor-X was not obtained, the CPH server was used

(Protein homology modeling server) and the query for position 392 and position 403 was not obtained, the query was started from position 424 [14]. The goal of this study was to identify the *ESCO2* gene ns-SNPs and predict their effect and association with Roberts's syndrome.

It was found that a total of 85 exonal ns-SNPs, there are 65 neutral and 20 deleterious SNPs according to SIFT prediction

result; from the 20 there are 14 SNPs predicted to be probably damaging according to polyphen-2. Testing the same 20 SNPs using PHD, SNP&GO and Provean we have got the same result (4 diseased SNPs and 16 neutral SNPs).

Upon these results; there are 4 SNPs were positive for disease association prediction (W539G, R427C, D403V, C392Y) and their stability according to I-mutant-3 and MUpro is decreasing for all 4 SNPs.

ESCO2 is a human orthologous of yeast Eco1 and is

mutated in Roberts syndrome, knowing that ESCO2 is essential for cohesion establishment and double strand break repair through binding with zinc finger domain then any disturbance in zinc finger domain like in the mutation of a cysteine in to a tyrosine at position 392 and the mutation of aspartic acid in to a valine at position 403, this disturbance lead to inappropriate DNA binding. These two mutations (C392Y and D403V) can be considered novel. For the mutation of arginine in to cysteine at position 427, the wild type had a positive charge when the mutant is neutral which can lead to loss of interactions with other molecules and increase the hydrophobicity in this position which can result in loss of hydrogen bonds and disturb correct folding, and for the mutation of tryptophan in to glycine at position 539, this mutation is located in a region with known splice variants described Roberts syndrome (MIM: 268300), and knowing that glycine is very flexible amino acid then this mutation can disturb the required rigidity of the protein. this mutation matches a described variants annotated with severity disease [15].

4. Conclusion

Among all analyzed nsSNPs there were only 4 nsSNPs that are deleterious and probably highly associated with Roberts's syndrome, especially the mutation of tryptophan into glycine at position 539 (W539G). The others 3 (R427C, D403V and C392Y) nsSNPs are novel and can be proposed as diagnostic nsSNPs for the disease.

5. Recommendation

We recommend further analysis for other types of SNPs and mutations and forward protein modeling and analysis (in silico and in vivo).

References

- [1] Miriam Gordillo, Hugo Vega, Ethylin. Wang Jabs Roberts Syndrome. Gene Reviews. April 14, 2009; <http://www.ncbi.nlm.nih.gov/books/nbk1153/>. Accessed 2/11/20.
- [2] Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics. 2010 Nov 8; 11 (1): 548. DOI: 10.1186/1471-2105-11-548. PubMed: 21059217.
- [3] Genetic Home Reference (<http://ghr.nlm.nih.gov/>).
- [4] Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003; 31: 3812-3814.
- [5] 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.
- [6] Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. (2009). Functional annotations improve the predictive score of human disease-related mutations in proteins. Human Mutation. 30; 1237-1244.
- [7] Choi, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the functional effect of amino acid substitutions and Indels. PLoS ONE 7: e46688.
- [8] Capriotti E, Fariselli P, Casadio R. (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res., 33 (Web Server issue): W306-W310.
- [9] J. Cheng, A. Randall, and P. Baldi. Prediction of Protein Stability Changes for Single Site Mutations Using Support Vector Machines. Proteins: Structure, Function, Bioinformatics. In press, 2005.
- [10] Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010 Jul 1; 38 Suppl: W214-20
- [11] Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics. 2010 Nov 8; 11: 548.
- [12] Sheng Wang, Wei Li, Shiwang Liu, and Jinbo Xu. RaptorX-property a web server for protein structure property prediction. Nucleic acids Research, 2016.
- [13] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004; 25: 1605-12.
- [14] Nielsen M., Lundegaard C., Lund O., Petersen TN. CPHmodels- 3.0--remote homology modeling using structure-guided sequence profiles. Nucleic Acids Research, 2010, Vol. 38.
- [15] Van der Lelij P, Godthelp BC, Van Zon W, Van Gostaliga D, Oostra AB, et al, (2009) The cellular phenotype of Roberts syndrome fibroblasts as revealed by ectopic expression of ESCO2, PLoS ONE 4 (9): e 6936. doi: 10.1371/ Journal. pone. 0006936.