American Journal of Bioinformatics Research 2019, 9(1): 11-21
DOI: 10.5923/j.bioinformatics.20190901.02

Comprehensive Computational Analysis Revealed Thirteen Novel Mutations in Human *FSH-B* gene Related to PCOS

Nidal Essa[1,*], Enas A. Osman[2], Hadeel M. Yousif[2], Kutuf A. Albushra[2], Amel Nasir Eltayeb Ali[2], Tebyan Ameer Abdelhameed Abbas[2], Mohamed A. Hassan[2,3]

[1]Faculty of Medical Laboratory Sciences, University of Medical Sciences and Technology, Khartoum, Sudan
[2]Department of Bioinformatics, Africa city of Technology, Khartoum, Sudan
[3]Department of Bioinformatics, DETAGEN Genetic Diagnostics Center, Kayseri, Turkey

Abstract Background: Follicular stimulating hormone beta subunit (***FSH-B***) as gonadotropin hormone secreted from the anterior pituitary belongs to the glycoprotein family located on 11p14.1 and consists of three exons. It is responsible for follicular growth and ovarian steroidogenesis in females and spermatogenesis in males. PCOS is a common endocrinopathy affecting 4-20% of women within the reproductive age the pathophysiological process is not fully understood and lowering of serum FSH level due to gene polymorphism lead to abnormal folliculogenesis and irregular menstrual cycle. In this study, we used various computational approaches to identify nsSNPs which probably be deleterious to the structure and/or function of FSH-B protein that might be associated with polycystic ovary syndrome. Methods: The data on human FSH-B gene was retrieved from dbSNP/NCBI. Eleven different bioinformatics prediction algorithms; SIFT, Polyphen, PROVEAN, SNAP2, Pmut, PhD-SNP, I-Mutant and Project Hope were used to analyze the effect of nsSNPs on functions and structure of the FSH-B protein, and RaptorX for protein modeling and Chimera for visualization of the model, in addition, we used PolymiRTS to detect SNPs on miRNA binding sites. Results: After retrieval of SNPs from the NCBI database, 164SNPs were classified as missense SNPs. From functional analysis softwares, 84 SNP were predicted to be deleterious then they analyzed by disease - related software 13 SNPs, when checked for protein stability, 12 of them decreased protein stability and one SNP increased its stability. Conclusion: Consideration should be taken to these 13 novel mutations when we are carrying out genetic studies through human samples.

Keywords PCOS, In Silico analysis, FSH-B, Folliculogenesis, Computational analysis

1. Introduction

The anterior pituitary glycoprotein hormone family consists of follicle-stimulating hormone, luteinizing hormone, chorionic gonadotropin, and thyroid-stimulating hormone [1]. All of these glycoproteins consist of an identical alpha subunit and a hormone-specific beta subunit. Follicular stimulating hormone beta subunit (FSH-B), also known as Follitropin subunit beta, located on 11p14.1 and consists of 3 exons, two transcript variants encoded for one protein with 129 amino acids which enables ovarian folliculogenesis to the antral follicle stage as well as Sertoli cell proliferation and maintenance of sperm quality in testis through binding to Follicular stimulating hormone receptor (FSHR) that expresses in the ovarian granulosa cells and Sertoli cells of the testis [1-3].

Polycystic ovary syndrome (PCOS), the disease of our interest, is a complex disorder, in which both environmental and genetic factors contribute to its pathogenesis [4]. Despite extensive researches, the precise pathophysiology of PCOS remains unknown. New theories were introduced, and an interest in the genetics of PCOS has increased considerably during recent years [5-6].

The etiology of PCOS is not fully understood, but the role of genetic factors has long been established by familial aggregation and genome-wide association studies revealing that altered expression of several genes affect signal transduction pathways controlling steroidogenesis, steroid hormones action, gonadotropin action and regulation, insulin action and secretion, energy homeostasis and chronic

* Corresponding author:
nidal.essa@umst-edu.sd (Nidal Essa)
Published online at http://journal.sapub.org/bioinformatics

inflammation [5-6]. Common phenotypic pictures of PCOS vary among different populations due to environmental, genetics and diagnostic criteria that define the appearance of the disease among women with PCOS such as irregular menstrual cycles, hyperandrogenism, polycystic ovarian morphology and acne [6-9].

In particular, Hypothalamus pituitary gonadal axis (HPO) genes as reproductive genes responsible for the production of steroid hormones and ovulatory process, any defect in the gene or its receptor due to mutations will result in structure and biological dysfunction which consequently results in anovulation and irregular menstrual cycle [10]. FSHB is thus likely to play an important role in the etiology of PCOS. Missense variants as one of these mutations lead to altering the translation of amino acid which consequently results in protein dysfunction either by their effect on protein solubility or by disrupting protein structure. Recent genome-wide association studies (GWAS) in Han Chinese women with PCOS demonstrate 11 genetic loci that associated with PCOS which are responsible for steroidogenesis, steroid hormones action, gonadotropin action and regulation, insulin action and secretion, energy homeostasis, chronic inflammation, and others. Replication studies have demonstrated that variants at several of these loci also confer risk for PCOS in women of European ethnicity. The strongest loci in Europeans contain genes for DENND1A and THADA, with additional associations in loci containing the LHCGR, FSHR, YAP1, and RAB5/SUOX. [4,7,11].

PCOS is a very complex syndrome, with a lot of undiscoverable genetic components that associate with its pathogenesis. Abnormal folliculogenesis that associated with PCOS is due to low serum FSH level but its mechanism of occurrence is unknown. Generally, these mutations affect its expression on the ovarian granulosa and result in abnormal serum level. Reported FSHB variants (rs11031010, rs11031006) showed significant association with abnormal serum levels of FSH and LH at genome-wide significance among both Chinese and European PCOS populations [12-16].

Exploitation of Bioinformatics algorithms with the construction of databases enables biomedical scientists to collect, retrieval, and analyze scientific data on a mass scale [17-18].

Combination of bioinformatics resources and computational "In-Silico" experiment will accelerate the pace and efficiency of scientific discovery [19-21].

This is the first study used the in silico analysis for building up knowledge about PCOS. Thus, we aimed to determine the influence of various SNPs in FSH-B gene using in silico prediction software and assessing their effects on the structure/function of FSH protein that may have an important role in disease susceptibility which will be examined through wet laboratory work and help in understanding the genetic mechanisms of the disease through these useful tools.

2. Material and Methods

2.1. Data retrieval: from the dbSNP (http://www.ncbi.nlm.nih.gov/SNP/), the SNPs information regarding *FSH-B* gene was obtained (SNP ID), the protein sequence and it is accession number was collected from Uni Prot database. http://www.uniprot.org.

2.2. SIFT (Sorting Intolerant from Tolerant): https://sift.bii.a-star.edu.sg/. As functional analysis tool based on the two concepts, sequence homology and conservation regions on protein sequence, the SFIT server predicts which amino acid substation lead to deleterious effect (SIFT scores <0.05) or tolerated effect (SIFT scores >0.05) on the protein structure [17].

2.3. Polyphen-2 (Polymorphism Phenotyping v2), http://genetics.bwh.harvard.edu/pph2/: Based on the position-specific independent count scores (PSIC) which calculated in dependence of the multiple sequence alignment analysis and protein 3D structure of the wild and the mutant protein. Its outcomes values ranged from (0-1) classified the predication product to benign, possibly damaging and probably damaging. MsSNPs that predicted to be intolerant by SIFT has been submitted to Polyphen as protein sequence in FASTA format that obtained from UniproktB /Expasy after submitting the relevant ensemble protein (ESNP) there, and then we entered the position of mutation, native amino acid and the new substituent for both structural and functional protein analysis [18].

2.4. Provean (Protein Variation Effect Analysis) (http://provean.jcvi.org/index.php): It is a software tool that predicts the impact of amino acid substitution on the biological function of the protein when we entered the protein FASTA sequence along with amino acid substitutions as input query for analysis. The outcome product according to the threshold score of −2.5 as tolerated or deleterious effect [19].

2.5. SNAP2 (Screening of Non-Acceptable Polymorphism 2; https://rostlab.org/services/snap2web/): It is a neural network dependent software tool that distinguishes between effect and neutral variants/non-synonymous SNPs by taking a variety of sequence and variant features into account. After entering the protein FASTA sequence as input query while the output predication as neutral, non- neutral effect [20].

2.6. PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms; http://snps.biofold.org/phd-snp/phd-snp.html): It is an online Support Vector Machine (SVM) based classifier, is optimized to predict if a given single point protein mutation can be classified as disease-related or as a neutral polymorphism, also we used the protein FASTA sequence as input query besides the residues change to obtain the output results [21].

2.7. SNPs &GO (Single nucleotide polymorphism & Gene Ontology, PHD-SNP (http://snps.biofold.org/snps-and-go). It's an accurate algorithm method that used the FASTA sequences of protein

to give unique framework information evolutionary information and function as encoded in the Gene Ontology terms, and other outperforms with available predictive methods [22].

2.8. P-mut (http://mmb2.pcb.ub.es:8080/PMut): a web-based tool for the annotation of pathological variants on proteins based on the uses of neural networks. The output display according to the pathogenicity index ranged from 0-1; more than 0.5 scored as single pathological mutations [23].

2.9. I–Mutant 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi): In regard to the neural network concept, it predicts the changes in protein stability in response to the occurrence of mutation. The output obtained is either increased or decreased stability of the protein [24].

2.10. Project Hope (http://www.cmbi.ru.nl/hope/): It is a new online web-server to search protein 3D structures (if available) by collecting structural information from a series of sources, including calculations on the 3D coordinates of the protein, sequence annotations from the Uni Prot database, and predictions by DAS services. Submission of protein sequence and the mutant variants as input information to the web server which displays the output as textual report supported by figures and animations [25].

2.11. RaptorX (http://raptorx.uchicago.edu/): It is a web server predicting structure property of a protein sequence without using any templates. It outperforms other servers, especially for proteins without close homologs in PDB or with very sparse sequence profile. The server predicts tertiary structure [26].

2.12. PolymiRTS Database 3.0: As well as it is a database of naturally occurring DNA variations in microRNAs (miRNA) seed region and miRNA target sites, it is used to predict 3 un-translated region polymorphism (3' UTR) effect upon microRNAs and their target sites which may influence miRNA-mRNA interaction, causing impact on miRNA-mediated gene repression. PolymiRTS database was made by examining 3UTRs of mRNAs in human and mouse for SNPs in miRNA target destinations. Then, the impact of polymorphism on gene expression and phenotypes are identified and then connected in the database. The PolymiRTS data base also includes polymorphism in target sites that have been supported by a variety of experimental methods and polymorphism in miRNA seed regions [27].

2.13. GeneMANIA: (http://www.genemania.org/). It is a web interface that predicts gene interaction with other genes in various pathways using genetic interactions, pathways, co-expression, co-localization and protein domain similarity [28].

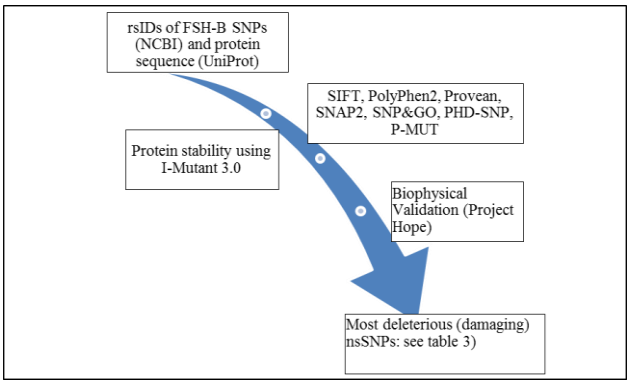

Figure 1. Diagrammatic representation of FSH-B gene in silico work flow

3. Results

Using dbSNP database for gene data set, the human FSH-B gene, gene of our interest, contained a total of 84 SNPs missense nsSNPs, 13 nsSNPs of them were predicted to have the most damaging effects on FSH-B protein's structure and function.

Table (1). Damaging or Deleterious or effect nsSNPs associated variations predicted by SIFT, Polyphen, PROVEAN, SNAP2 softwares

dbSNP rs#	Amino acid change	SIFT Prediction	SIFT score	polyphen2 prediction	poyphen2 score	provean prediction	provean score	SNAP2 prediction	SNAP2 Score
rs746553053	I28T	Deletrious	0.01	probably damaging	1	Deleterious	-4.24	effect	70
rs989734200	C35G	Deletrious	0	probably damaging	1	Deleterious	-11.3	effect	89
rs1166756851	C38R	Deletrious	0	probably damaging	1	Deleterious	-1.11	effect	84
rs770418547	C38Y	Deletrious	0	probably damaging	1	Deleterious	-10	effect	86
	C38F	Deletrious	0	probably damaging	1	Deleterious	-10.1	effect	87
rs773628212	I39T	Deletrious	0.01	possibly damaging	0.846	Deleterious	-3.19	effect	40
rs911993505	I41T	Deletrious	0	possibly damaging	0.513	Deleterious	-3.41	effect	76
rs944749696	N42S	Deletrious	0.03	probably damaging	0.999	Deleterious	-3.15	effect	20
rs760435888	C46Y	Deletrious	0	probably damaging	1	Deleterious	-10.7	effect	93

dbSNP rs#	Amino acid change	SIFT Prediction	SIFT score	polyphen2 prediction	poyphen2 score	provean prediction	provean score	SNAP2 prediction	SNAP2 Score
rs377437870	C50F	Deletrious	0	probably damaging	1	Deleterious	-10.7	effect	86
rs369839978	P60Q	Deletrious	0	probably damaging	0.979	Deleterious	-3.67	effect	49
	P60L	Deletrious	0	probably damaging	0.968	Deleterious	-3.92	effect	50
rs5030776	C69G	Deletrious	0	probably damaging	1	Deleterious	-11.9	effect	92
rs1202953613	T70I	Deletrious	0	possibly damaging	0.939	Deleterious	-4.39	effect	39
rs767219014	K72N	Deletrious	0.01	probably damaging	0.984	Deleterious	-2.52	effect	25
rs1048025235	Y76H	Deletrious	0	probably damaging	1	Deleterious	-4.446	effect	64
rs1242052340	E77V	Deletrious	0	possibly damaging	0.923	Deleterious	-5.535	effect	25
rs540759868	T78R	Deletrious	0	probably damaging	1	Deleterious	-5.07	effect	72
rs1171563230	V79A	Deletrious	0.04	possibly damaging	0.935	Deleterious	-3.003	effect	42
rs1044572003	P82H	Deletrious	0	probably damaging	1	Deleterious	-7.271	effect	12
rs757773706	G83S	Deletrious	0.02	possibly damaging	0.999	Deleterious	-4.317	effect	27
rs1366775720	G83V	Deletrious	0	probably damaging	1	Deleterious	-7.485	effect	71
rs1159738542	S90F	Deletrious	0	probably damaging	1	Deleterious	-3.81	effect	77
rs368414879	P95S	Deletrious	0	probably damaging	1	Deleterious	-7.868	effect	76
rs762866106	V96M	Deletrious	0	probably damaging	1	Deleterious	-2.943	effect	55
rs5030777	C100R	Deletrious	0	probably damaging	1	Deleterious	-11.74	effect	89
rs267602841	D108N	Deletrious	0.01	possibly damaging	0.673	Deleterious	-3.147	effect	47
rs1318841246	C122G	Deletrious	0	probably damaging	1	Deleterious	-9.25	effect	87

Table (2). Disease effect nsSNPs associated variations predicted by Pmut, SNP&GO and PHD softwares

dbSNP rs#	Amino acid change	Prediction SNP &GO	Probability SNP &GO	RI	Prediction PHD	Probability PHD	RI	Mutation P MUT	Prediction PMUT
rs989734200	C35G	Disease	0.849	7	Disease	0.886	8	C → G (Cys → Gly)	0.82 (90%) Disease
rs1166756851	C38R	Disease	0.864	7	Disease	0.882	8	C → R (Cys → Arg)	0.78 (88%) Disease
rs770418547	C38Y	Disease	0.86	7	Disease	0.912	8	C → Y (Cys → Tyr)	0.82 (90%) Disease
	C38F	Disease	0.868	7	Disease	0.895	8	C → F (Cys → Phe)	0.82 (90%) Disease
rs760435888	C46Y	Disease	0.904	8	Disease	0.954	9	C → Y (Cys → Tyr)	0.70 (86%) Disease
rs377437870	C50F	Disease	0.907	8	Disease	0.945	9	C → F (Cys → Phe)	0.70 (86%) Disease
rs5030776	C69G	Disease	0.805	6	Disease	0.852	7	C → G (Cys → Gly)	0.73 (87%) Disease
rs540759868	T78R	Disease	0.515	0	Disease	0.76	5	T → R (Thr → Arg)	0.59 (82%) Disease

dbSNP rs#	Amino acid change	Prediction SNP &GO	Probability SNP &GO	RI	Prediction PHD	Probability PHD	RI	Mutation P MUT	Prediction PMUT
rs757773706	G83S	Disease	0.605	2	Disease	0.768	5	G → S (Gly → Ser)	0.56 (81%) Disease
rs1366775720	G83V	Disease	0.785	6	Disease	0.856	7	G → V (Gly → Val)	0.69 (86%) Disease
rs368414879	P95S	Disease	0.804	6	Disease	0.801	6	P → S (Pro → Ser)	0.69 (85%) Disease
rs5030777	C100R	Disease	0.865	7	Disease	0.887	8	C → R (Cys → Arg)	0.66 (85%) Disease
rs1318841246	C122G	Disease	0.765	5	Disease	0.725	5	C → G (Cys → Gly)	0.70 (86%) Disease

Table (3). Stability analysis predicted by I-Mutant version 3.0 (also Show the novel mutations)

dbSNP rs#	Amino acid change	I mutant	I mutant
rs989734200	C35G	Decrease	-0.95
rs1166756851	C38R	Decrease	-0.13
rs770418547	C38Y	Increase	-0.08
	C38F	Decrease	0.16
rs760435888	C46Y	Decrease	-0.17
rs377437870	C50F	Decrease	-0.02
rs5030776	C69G	Decrease	-1.17
rs540759868	T78R	Decrease	-0.27
rs757773706	G83S	Decrease	-1.23
rs1366775720	G83V	Decrease	-0.43
rs368414879	P95S	Decrease	-1.42
rs5030777	C100R	Decrease	-0.28
rs1318841246	C122G	Decrease	-1.57

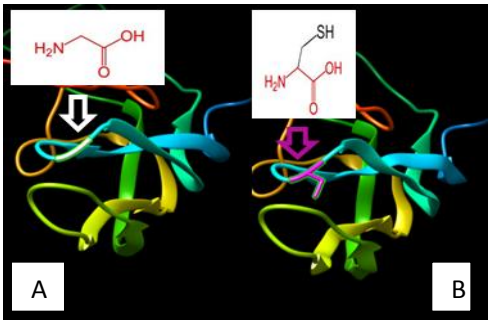

Figure 2. SNP ID: rs989734200, (C35G) The amino acid Cysteine in the native protein (A) changed to Glycine in the mutant protein (B) at position 35

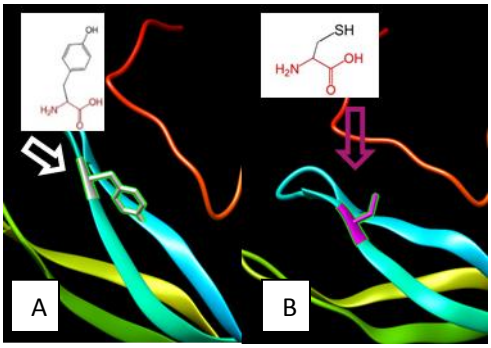

Figure 3. SNP ID: rs770418547, (C38F) The amino acid Cysteine in the native protein (A) changed to Phenylalanine in the mutant protein (B) at position 38

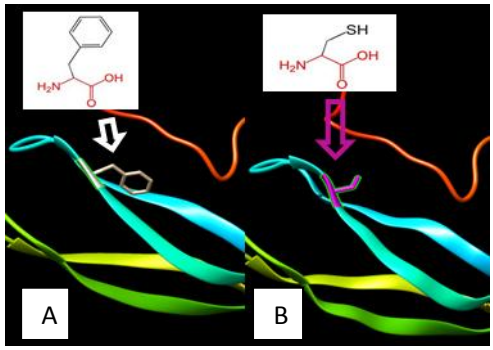

Figure 4. SNP ID: rs770418547, (C38R) The amino acid Cysteine in the native protein (A) changed to Tyrosine in the mutant protein (B) at position 38

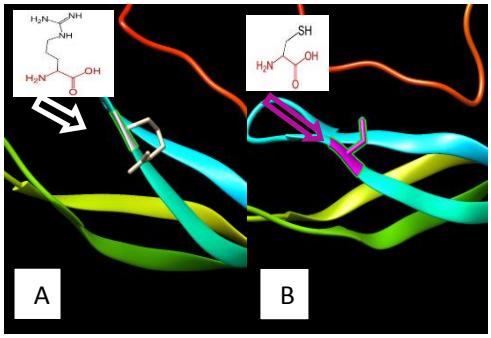

Figure 5. SNP ID: rs1166756851, (C38R) The amino acid Cysteine in the native protein (A) changed to Arginine in the mutant protein (B) at position 38

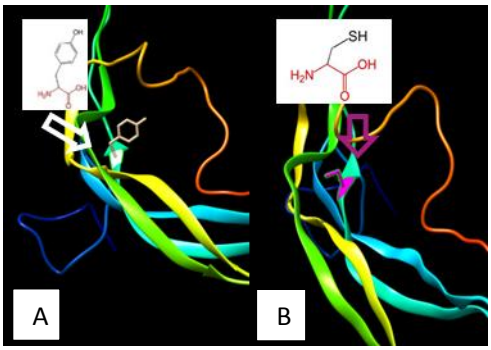

Figure 6. SNP ID: rs760435888, (C46Y) The amino acid Cysteine in the native protein (A) changed to Arginine in the mutant protein (B) at position 46

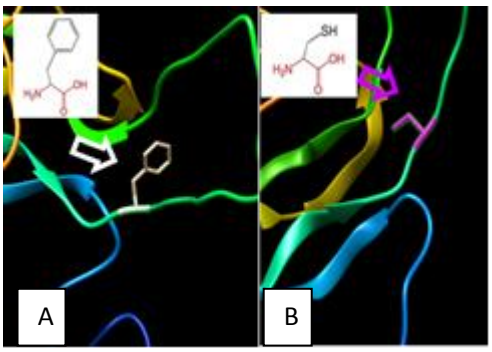

Figure 7. SNP ID: rs377437870, (C50F) The amino acid Cysteine in the native protein (A) changed to Phenylalanine in the mutant protein (B) at position 50

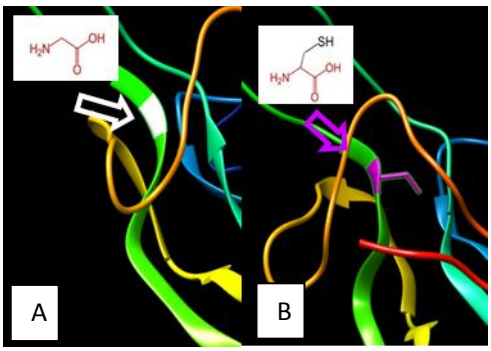

Figure 8. SNP ID: rs5030776, (C69G) The amino acid Cysteine in the native protein (A) changed to Glycine in the mutant protein (B) at position 69

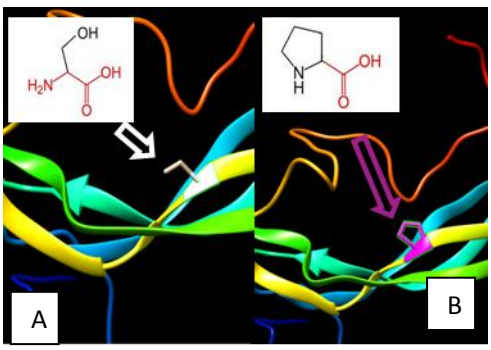

Figure 9. SNP ID: rs368414879, (P59S) The amino acid Proline in the native protein (A) changed to Serine in the mutant protein (B) at position 59

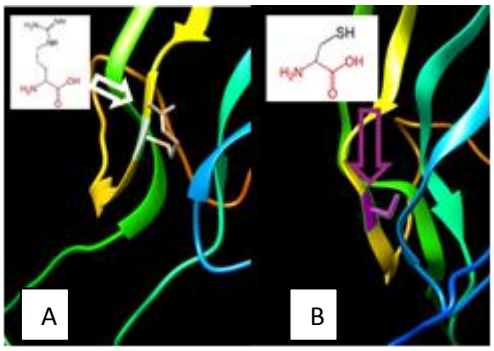

Figure 10. SNP ID: rs5030777, (C100R) The amino acid Cysteine in the native protein (A) changed to Arginine in the mutant protein (B) at position 100

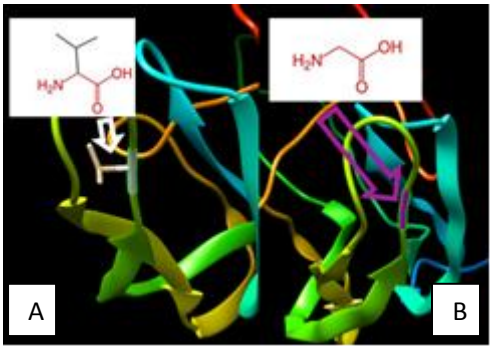

Figure 11. SNP ID: rs1366775720, (G83V) The amino acid Glycine in the native protein (A) changed to Valine in the mutant protein (B) at position 83

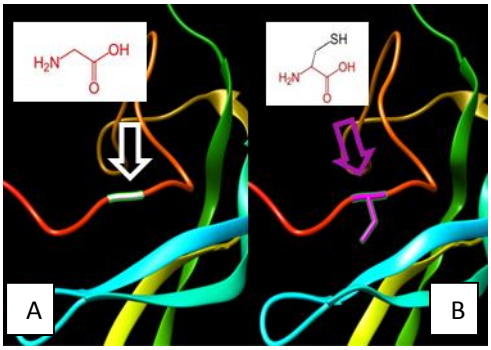

Figure 12. SNP ID: rs1318841246, (C122G) The amino acid Cysteine in the native protein (A) changed to Glycine in the mutant protein (B) at position 122

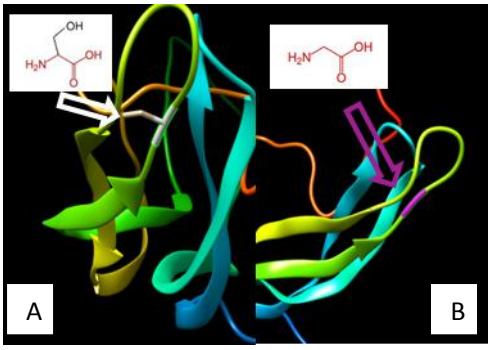

Figure 13. SNP ID: rs757773706, (G83S) The amino acid Glycine in the native protein (A) changed to Serine in the mutant protein (B) at position 83

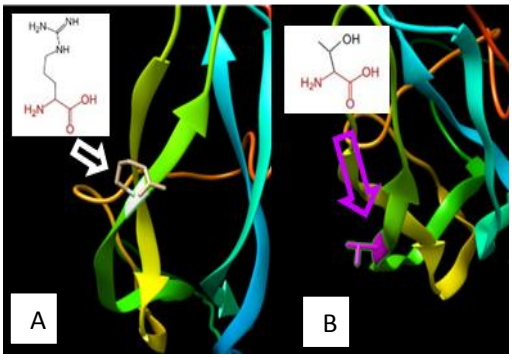

Figure 14. SNP ID: rs540759868, (T78R) The amino acid Threonine in the native protein (A) changed to Arginine in the mutant protein (B) at position 78

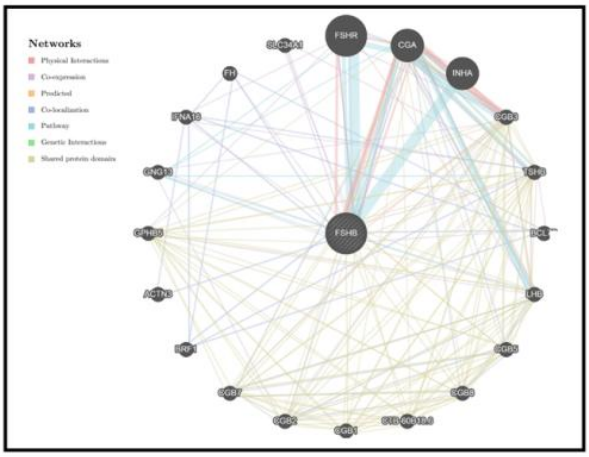

Figure 15. GeneMANIA result for *FSH-B* Gene

Table (4). Prediction of SNPs at the 3'UTR Region using PolymiRTS

Location	dbSNP ID	miR ID	miRSite	Function Class
30255406	rs182092464	hsa-miR-499b-5p	tcaaaAAGTCTGt	D
30255426	rs186768498	hsa-miR-1285-3p	caatgTGCCCAGg	D
		hsa-miR-3187-5p	caatgTGCCCAGg	D
		hsa-miR-5189-5p	caatgTGCCCAGg	D
		hsa-miR-612	caatgTGCCCAGg	D
		hsa-miR-6860	caatgTGCCCAGg	D
		hsa-miR-939-3p	caatgtGCCCAGG	D
		hsa-miR-1292-5p	caatGTTCCCAgg	C
		hsa-miR-204-3p	caatgTTCCCAGg	C
		hsa-miR-4471	caatGTTCCCAgg	C
		hsa-miR-4646-5p	caatgTTCCCAGg	C
		hsa-miR-8059	caatGTTCCCAgg	C
		hsa-miR-2115-3p	cTTCTGATAgact	C
		hsa-miR-361-5p	ctTCTGATAgact	C
30255797	rs201231187	hsa-miR-1910-3p	agcacTCTGCCTggaac	O
		hsa-miR-4514	agcactCTGCCTGgaac	O
		hsa-miR-4645-5p	agcactcTGCCTGGAac	O
		hsa-miR-4673	agcactcTGCCTGGAac	O
		hsa-miR-4692	agcactCTGCCTGgaac	O
		hsa-miR-6511a-5p	agcacTCTGCCTggaac	O
		hsa-miR-759	agCACTCTGcctggaac	O
		hsa-miR-194-5p	agcactCTGTTACctggaac	O
		hsa-miR-4797-5p	agcACTCTGTtacctggaac	O
30255887	rs184811057	hsa-miR-6072	agctGATGAGGct	D
		hsa-miR-6891-3p	agctGATGAGGct	D
30255980	rs190082386	hsa-miR-499a-3p	caTGATGTAtctt	D
		hsa-miR-499b-3p	caTGATGTAtctt	D
		hsa-miR-2681-3p	CATGATAtatctt	C
		hsa-miR-3074-3p	caTGATATAtctt	C
		hsa-miR-96-3p	CATGATAtatctt	C
30255982	rs676349	hsa-miR-499a-3p	TGATGTAtctttc	D
		hsa-miR-499b-3p	TGATGTAtctttc	D
		hsa-miR-642a-3p	tgATGTGTCttte	C
		hsa-miR-642b-3p	tgATGTGTCttte	C

Location	dbSNP ID	miR ID	miRSite	Function Class
30256033	rs140725941	hsa-miR-3978	gaTTTCCAAgcta	D
		hsa-miR-4453	gatttCCAAGCTA	D
		hsa-miR-4538	gatttCCAAGCTA	D
		hsa-miR-5093	gATTTCCAagcta	D
		hsa-miR-21-5p	gatttcTAAGCTA	C
		hsa-miR-5579-3p	gatttcTAAGCTA	C
		hsa-miR-5680	gATTTCTAagcta	C
		hsa-miR-590-5p	gatttcTAAGCTA	C
30256076	rs193044329	hsa-miR-578	ggCAAGAAAttgt	D
30256225	rs184744059	hsa-miR-4795-5p	CCACTTCtttgga	D
30256295	rs78946483			
		hsa-miR-6502-3p	ttaatTGGTCTAc	C
30256341	rs146658313	hsa-miR-5571-5p	aaAGAATTAgcca	D
		hsa-miR-138-2-3p	aaaGAAATAGcca	C
30256453	rs145887479	hsa-miR-1277-5p	ggttATATATTAa	D
		hsa-miR-410-3p	gGTTATATAttaa	D
		hsa-miR-5011-5p	ggtTATATATtaa	D
		hsa-miR-889-3p	ggttatATATTAA	D
		hsa-miR-374c-5p	ggttaTGTATTAA	C
		hsa-miR-655-3p	ggttaTGTATTAA	C
30256737	rs181864952	hsa-miR-3065-5p	atTTTGTTGgatt	D
		hsa-miR-1468-3p	ATTTTGCtggatt	C
		hsa-miR-338-3p	atttTGCTGGAtt	C
		hsa-miR-4530	atttTGCTGGAtt	C
		hsa-miR-545-3p	atTTTGCTGgatt	C
		hsa-miR-548p	aTTTTGCTggatt	C
		hsa-miR-5586-5p	attttGCTGGATt	C

4. Discussion

FSHB gene was investigated in dbSNP National Centre of Biotechnology Information (NCBI public database). This gene contains a total of 354 SNPs in coding region; of which 76 are missense, 4 synonymous, eight nonsense, five frame shift and 612 are in the non-coding region, of which 300 in the 3'un-translated region (3’ UTR) and 312 in 5’ un-translated region (5’UTR). We selected the missense coding SNPs and 3′UTR SNPs for our investigation.

13 novel mutations have been predicated which had an effect on protein stability and function using different bioinformatics algorisms such as SIFT, PolyPhen-2, Provean, SNAP2, SNP&GO, PHD-SNP, P-MUT and I-Mutant 3.0 (Figure 15).

Changes in physiochemical properties of the protein due to mutations will possibly result in critical changes in the protein function and consequently result in different phenotypic pictures of the disease.

Regarding protein stability, all of the predicted SNPs were predicted to deregulates the protein stability, (rs770418547, C38Y) which increased protein stability. While four of them, ((rs989734200, C35G), (rs5030776, C69G), (rs1318841246, C122G) and (rs368414879, P59S)), their substitutions result in lowering the protein size and lead to losing of hydrophobic interactions in the protein core, while the rest increase the molecule size which disturbs interactions with other molecules or other parts of the protein.

Additionally, three of them ((rs760435888, C46Y), (rs377437870, C50F) and (rs5030777, C100)) were unique in their position and lead to complete damaging of the protein which associated with protein folding problems.

In this mutation, (rs989734200, (C35G)); the mutant residue is smaller than the wild type residue which will cause an empty space in the core of the protein, loss of hydrophobic interactions property and decrease protein stability.

Substitution of Cysteine to Phenylalanine and Arginine rs770418547, (C38F and C38R) in the same location 38 due to mutation has many effects in protein structure appear in decreasing protein stability, increasing the size which can disturb interactions with other molecules or other parts of the protein. While in the same location when Cytosine is substituted to Arginine (C38R) the protein become positive in charge, loss of hydrophobic interaction and disturbs its interactions with other molecules or other parts of the protein.

Also mutation in this location rs1166756851, (C38R); results in increasing the size of the protein that changes the

physiochemical property of the protein to a positive charge while the wild-type residue charge is neutral which consequently might result in losing hydrophobic interactions with other molecules on the surface of the protein.

In this unique location rs760435888, (C46Y); Cytosine is also mutated to Tyrosine results in complete damaging of the protein, which might disturb the interactions with other molecules or other parts of the protein because this residue is located on the surface of the protein.

Furthermore, changing of Cytosine to Phenylalanine in location 50 (rs377437870, (C50F)) consequently damage the protein structure, increasing the size of it and can disturb interactions with other molecules or other parts of the protein.

One of the mutations that leads to decrease protein size occur when Cysteine is converted to Glycine (rs5030776, (C69G)) and cause a possible loss of external interactions beside loss of hydrophobic interactions with other molecules on the surface of the protein, As well as, when altering of Proline to Serine localized in 59 position (rs368414879, (P59S)), the mutant residue is smaller in size and possibly causes loss of external interactions and hydrophobic interactions with other molecules on the surface of the protein.

Conversion of Cysteine to Arginine in position 100 due to mutation (rs5030777, (C100R)) changes protein neutrality to positive charge and increases the size which consequently damages it due to protein folding problems with loss of hydrophobic interactions in the core of the protein. rs1366775720, (G83V); showed that mutant residue is more hydrophobic than the wild-type residue, bigger in size, and the torsion angles of 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 of the protein. rs1318841246, (C122G); the mutant residue is smaller than the wild-type residue. This will cause a possible loss of external interactions and loss of hydrophobic interactions with other molecules on the surface of the protein.

Moreover, conversion of Glycine to Serine in position 83 (rs757773706, (G83S)) associated with decrease protein stability, bigger size of the mutant residue and the unusual torsion angles will force the local backbone into an incorrect conformation and will disturb its local structure.

Mutated Arginine in position 78 (rs540759868, (T78R)) changes the protein neutrality to a positive charge, become bigger in sizeand might cause loss of hydrophobic interactions with other molecules on the surface of the protein.

Despite extensive genetic studies were done to study the association of FSH-B polymorphisms with PCOS, our computational analysis of the gene revealed that these thirteen novel SNPs success in the analysis by various softwares with no previous studies.

Only two of them (rs5030776, rs5030777) were reported by Nagirnaja, L [29], et al., Lindstedt, G, et al [30], and Layman, L.C, [31] et al., associated with increase serum FSH level and male infertility. FSH-B plays the same role in development and maturation of gametes in both genders and regard to female physiology; it plays a central role in the regulation of ovarian folliculogenesis, which is disordered in PCOS, under the influence of LH action.

FSH-B gene interacts with more than 20 genes to accomplish different biological functions. Therefore, damaging its protein due to one of these mutations will probably destroy these pathways.

Consideration should be taken to these thirteen novel mutations when we are carrying out genetic studies through human samples.

5. Conclusions

13 nsSNPs with different positions were predicted to be the most damaging mutations for FSH protein, altering its physiochemical properties such as size, charge, hydrophobicity, and stability, leading to loss or disturbance of the protein internal and external interactions and eventually loss of the protein's function as well as disease association. Furthermore, the 61 SNPs at the 3UTR were predicted to disrupt miRNA binding sites and hence affect the gene expression function

Currently, in silico analysis plays an essential role in narrowing the gap between the theoretical and applied parts of the medical research and by computational bioinformatics algorithms, we can increase our knowledge about disease pathogenic process which impacts in di covering newly preventive and treatment strategies tools.

ACKNOWLEDGEMENTS

The authors desire to acknowledge the excited cooperation of Africa City of Technology-Khartoum, Sudan and to Dr. Mounkaila Noma for language editing and proofreading of the manuscript, University of Medical Sciences and Technology, Khartoum, Sudan.

REFERENCES

[1] Kristiina Rull, Marina Grigorova, Aivar Ehrenberg, Pille Vaas, Aire Sekavin, Diana Nõmmemees, Mart Adler, Ele Hanson, Peeter Juhanson, and Maris Laan, 2018, FSHB −211 G>T is a major genetic modulator of reproductive physiology and health in childbearing age women. Human Reproduction. pp. 1–13.

[2] Grigorova, M., Margus Punab, Kristo Ausmees, and Maris Laan., 2008, FSHB promoter polymorphism within an evolutionary conserved element is associated with serum FSH level in men. Hum Reprod, 23(9): p. 2160-6.

[3] R. Azziz, D.A. Dumesic, and M.O. Goodarzi, 2011, Polycystic ovary syndrome: an ancient disorder? Fertil. Steril. 95 (5) 1544–1548.

[4] R. Saxena, N.A. Georgopoulos, T.J. Braaten, A.C. Bjonnes, V. Koika, D. Panidis, and C.K. Welt. 2015. Han Chinese polycystic ovary syndrome risk variants in women of European ancestry: relationship to FSH levels and glucose tolerance, Human Reproduction, Vol.30, No.6 pp. 1454–1459.

[5] Wassim Y. Almawi, Bayan Hubail, Dana Z. Arekat, Suhaila M. Al-Farsi, Shadha K. Al-Kindi, Mona R. Arekat, Naeema Mahmood and Samira Madan. 2015. Leutinizing hormone/choriogonadotropin receptor and follicle stimulating hormone receptor gene variants in polycystic ovary syndrome, J Assist Reprod Genet 32:607–614.

[6] Anderson Sanches de Melo, Sabrine Vilan Dias, Ricardo de Carvalho Cavalli, Viviane Cunha Cardoso, Heloisa Bettiol, Marco Antonio Barbieri, Rui Alberto Ferriani and Carolina Sales Vieira. (2015). Pathogenesis of polycystic ovary syndrome: multifactorial assessment from the foetal stage to menopause Reproduction.150 R11–R24.

[7] Hyejin Lee, Jee-Young Oh, Yeon-Ah Sung, Hyewon Chung, Hyung-Lae Kim, Gwang Sub Kim, Yoon Shin Cho, Jin Taek Kim; Genome-wide association study identified new susceptibility loci for polycystic ovary syndrome, Human Reproduction, Volume 30, Issue 3, 1 March 2015, Pages 723–731, https://doi.org/10.1093/humrep/deu352.

[8] Prapas N, Karkanaki A, Prapas I, Kalogiannidis I, Katsikis I, and Panidis D, (2009). Genetics.pathophysyiology. H21IP6POKRATIA, 13, 4: 216-223.

[9] Zi-Jiang Chen, Han Zhao, Lin He, Yuhua Shi, Yingying Qin, Yongyong Shi, Zhiqiang Li, Li You, Junli Zhao, Jiayin Liu, Xiaoyan Liang, Xiaoming Zhao, Junzhao Zhao, Yingpu Sun, Bo Zhang, Hong Jiang, Dongni Zhao, Yuehong Bian, Xuan Gao, Ling Geng, Yiran Li, Dongyi Zhu, Xiuqin Sun, Jin-e Xu, Cuifang Hao, Chun-e Ren, Yajie Zhang, Shiling Chen, Wei Zhang, Aijun Yang, Junhao Yan, Yuan Li, Jinlong Ma and Yueran Zhao., 2011, Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3., Nat Genet, 43(1): p. 55-9.

[10] O. Valkenburg A.G. Uitterlinden D. Piersma, A. Hofman, A.P.N. Themmen, F.H. de Jong, B.C.J.M. Fauser, and J.S.E. Laven, 2009, Genetic polymorphisms of GnRH and gonadotrophic hormone receptors affect the phenotype of polycystic ovary syndrome Human Reproduction, Vol.24, No.8 pp. 2014–2022.

[11] Yongyong Shi, Han Zhao, Yuhua Shi, Yunxia Cao, Dongzi Yang, Zhiqiang Li, Bo Zhang, Xiaoyan Liang, Tao Li, Jianhua Chen, Jiawei Shen, Junzhao Zhao, Li You, Xuan Gao, Dongyi Zhu, Xiaoming Zhao, Ying Yan, Yingying Qin, Wenjin Li, Junhao Yan, Qingzhong Wang, Junli Zhao, Ling , Jinlong Ma, Yueran Zhao, Guang He, Aiping Zhang, Shuhua Zou, Aijun Yang, Jiayin Liu, Weidong Li, Baojie Li, Chunling Wan, Ying Qin, Juanzi Shi, Jing Yang, Hong Jiang, Jin-e Xu, Xiujuan Qi, Yun Sun, Yajie Zhang, Cuifang Hao, Xiuqing Ju, Dongni Zhao, Chun-e Ren, Xiuqing Li, Wei Zhang, Yiwen Zhang, Jiangtao Zhang, Di Wu, Changming Zhang, Lin He and Zi-Jiang Chen., 2012, Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome., Nat Genet, 44:(9).

[12] Ye Tian, Han Zhao, Haitao Chen, Yingqian Peng, Linlin Cui, Yanzhi Du, Zhao Wang, Jianfeng Xu, and Zi-Jiang Chen, 2016, Variants in FSHB Are Associated With Polycystic Ovary Syndrome and Luteinizing Hormone Level in Han Chinese Women. J Clin Endocrinol Metab 101: 2178–2184.

[13] Priscilla Mutharasan, Eugene Galdones, Beatriz Pen alver Bernabe, Obed A. Garcia, Nadereh Jafari, Lonnie D. Shea, Teresa K. Woodruff, Richard S. Legro, Andrea Dunaif, and Margrit Urbanek., 2013, Evidence for Chromosome 2p16.3 Polycystic Ovary Syndrome Susceptibility Locus in Affected Women of European Ancestry J Clin Endocrinol Metab 98: E185–E190.

[14] Hayes, M.G., Margrit Urbanek, David A. Ehrmann, Loren L. Armstrong, Ji Young Lee, Ryan Sisk, Tugce Karaderi, Thomas M. Barber, Mark I. McCarthy, Stephen Franks, Cecilia M. Lindgren, Corrine K. Welt, Evanthia Diamanti-Kandarakis, Dimitrios Panidis, Mark O. Goodarzi, Ricardo Azziz, Yi ZhangRoland G. James, Michael Olivier, Ahmed H. Kissebah, Reproductive Medicine Network, Elisabet Stener-Victorin, Richard S. Legro, and , Andrea Dunaif. 2015, Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. Nat Commun, 6: p. 7502.

[15] Jan M. McAllister, Richard S. Legro, Bhavi P. Modi, and Jerome F. Strauss III,. 2015, Functional Genomics of PCOS: From GWAS to Molecular Mechanisms. Trends Endocrinol Metab, 26(3): 118–124.

[16] Welt, C.K. and J.M. Duran, 2014, Genetics of polycystic ovary syndrome. Semin Reprod Med. 32(3): p. 177-82.

[17] Altman RB (2012). Introduction to Translational Bioinformatics Collection. PLoS Comput Biol 8(12): e1002796. doi:10.1371/journal.pcbi.1002796.

[18] Kann, M. G. 2009, Advances in translational bioinformatics: computational approaches for the hunting of disease genes BRIEFINGS IN BIOINFORMATICS. VOL 11. NO 1. 96 -110.

[19] Sirintrapun, S.J., et al., Translational Bioinformatics and Clinical Research (Biomedical) Informatics. Clin Lab Med, 2016. 36(1): p. 153-81.

[20] Kennelly, P.J. and Rodwell W V., Bioinformatics & Computational Biology, 30th ed., McGraw-Hill Education. 2015.

[21] Jessica D. Tenenbaum, Translational Bioinformatics: Past, Present, and Future, Genomics, Proteomics & Bioinformatics, Volume 14, Issue 1, 2016.

[22] Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic acids research, 40(Web Server issue), W452-7.

[23] Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting the functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics, Chapter 7, Unit7.20.

[24] Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics (Oxford, England), 31(16), 2745-7.

[25] Bromberg, Y. and B. Rost, SNAP: predict the effect of non-synonymous polymorphisms on function. Nucleic Acids Res, 2007. 35(11): p. 3823-35.

[26] Wu, J., & Jiang, R. (2013). Prediction of deleterious nonsynonymous single-nucleotide polymorphism for human diseases. The Scientific World Journal, 2013, 675851.

[27] Desai, M. and J.B. Chauhan, Predicting the functional and structural consequences of nsSNPs in human methionine synthase gene using computational tools. 2019: p. 1-13.

[28] López-Ferrando, V., Gazzo, A., de la Cruz, X., Orozco, M., & Gelpí, J. L. (2017). PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. Nucleic acids research, 45(W1), W222-W228.

[29] Capriotti, E., Fariselli, P., & Casadio, R. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic acids research, 33(Web Server issue), W306-10.

[30] Venselaar, H., et al., Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics, 2010. 11: p. 548.

[31] Wang, S., Li, W., Liu, S., & Xu, J. (2016). RaptorX-Property: a web server for protein structure property prediction. Nucleic acids research, 44(W1), W430-5.

[32] Bhattacharya, A., Ziebarth, J. D., & Cui, Y. (2013). PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic acids research, 42 (Database issue), D86-91.

[33] Franz, M., Rodriguez, H., Lopes, C., Zuberi, K., Montojo, J., Bader, G. D., & Morris, Q. (2018). GeneMANIA update 2018. Nucleic acids research, 46(W1), W60-W64.

[34] Nagirnaja, L., et al., Genomics and genetics of gonadotropin beta-subunit genes: Unique FSHB and duplicated LHB/CGB loci. Mol Cell Endocrinol, 2010. 329(1-2): p. 4-16.

[35] Lindstedt, G., et al., Follitropin (FSH) deficiency in an infertile male due to FSHbeta gene mutation. A syndrome of normal puberty and virilization but underdeveloped testicles with azoospermia, low FSH but high lutropin and normal serum testosterone concentrations. Clin Chem Lab Med, 1998. 36(8): p. 663-5.

[36] Layman, L.C., et al., Delayed puberty and hypogonadism caused by mutations in the follicle-stimulating hormone beta-subunit gene. N Engl J Med, 1997. 337(9): p. 607-11.