American Journal of Bioinformatics Research 2019, 9(2): 45-55
DOI: 10.5923/j.bioinformatics.20190902.01

Bioinformatics Analysis and the Revelation of Thirteen Novel Mutations in Human LH-B Gene Related to PCOS

Nidal Essa[1,*], Abdelrahman H. Abdelmoneiom[2], Razan M. Badawi[2], Afnan Mohamed Mohamed Khair[2], Rania Anwer[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 Luteinizing hormone beta subunit (LH-B) (protein ID P01229) is gonadotropin hormone secreted from the anterior pituitary belongs to the glycoprotein family, mapped on chr19p13.3 and consists of three exons, three transcript variants with two phenotypes encoded for one protein known as Lutropin subunit beta (P01229) It has a central role in promoting spermatogenesis and ovulation by stimulating the testes and ovaries for steroidogenesis. PCOS is a common endocrinopathy affecting of women within the reproductive age, abnormal ovulation associated with a high level of LH as a result of gene polymorphisms lead to infertility problems. In this study, we used various computational approaches to identify nsSNPs which probably be deleterious to the structure and/or function of LH-B protein that might be associated with polycystic ovary syndrome. The data on human LH-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 LH-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. After retrieval of SNPs from the NCBI database, 140SNPs were classified as missense SNPs. From functional analysis software, 39 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. Hence, application of these 13 novel mutations through genetic studies will contribute towards our understanding of the pathophysiology of PCOS.

Keywords PCOS, In Silico analysis, LH-B, Folliculogenesis, Computational analysis, SNPs, Mutation

1. Introduction

Luteinizing hormone beta subunit (LH-B, OMIM 152780) as hypothalamus-pituitary-gonadal axis (*HPO*) gene, responsible for the production of steroid hormones and ovulatory process, mainly secreted from the anterior pituitary gland, and it belongs to glycoprotein family which has two subunits; an identical alpha subunit and a hormone-specific beta subunit [1,2]. LH-B also known as Lutropin beta chain, mapped onchr19p13.3 (1111bp, linear DNA) and consists of three exons, three transcript variants with two phenotypes encoded for one protein known as Lutropin subunit beta (P01229)which consists of 141 amino acids and 15,345(Da). It has a central role in promoting spermatogenesis and ovulation by stimulating the testes and ovaries for steroidogenesis [3], during ovulation, which is mainly LH dependent process, requires surge of LH-B levels to initiate the steroidogenesis which is essential for follicular development and corpus luteum formation through binding of LH to its high affinity receptor, LH/choriogonadotropin receptor (LHCGR) in specific theca and granulosa cells for androgen production mainly progesterone and estrogen, which maintain the endometrium for conception [4,5]. If conception or implantation fails to occur, the endometrium is shed and the corpus luteum atrophies to an atretic follicle. Therefore, anovulation due to abnormal LH signaling leads to a liberal role in amplifying ovarian androgen production that might be associated with polycystic ovary syndrome (PCOS) [5-7].

The appearance of PCOS as Stein & Leventhal syndrome [8] in 1935 as first report of the disease also described it as a syndrome of contradictory, because there is heterogenecity of both ovarian morphology and clinical findings such as hirsutism, obesity, infertility and polycystic ovary, also this heterogenecity extended to its genetic background with polygenic mode of inheritance. In addition to other genetics studies revealed that genes responsible from signal

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

transduction pathways which control many of them such as steroidogenesis, steroid hormones action, gonadotropin action and regulation, insulin action and secretion, energy homeostasis and chronic inflammation show abnormal expressions among PCOS population [9-12].

Hyperandrogenaemia as well as anovulation is the central hallmarks of PCOS, the majority of females with PCOS suffering from abnormal menstrual cycles and anovulatory problem. Many studies detected abnormal elevation of LHB due to gene polymorphisms which resulted in bioactivity changes in structure and even through the functional part of the hormone associated with anovulation among PCOS women. But until now, the relationship between LH signaling pathway and PCOS has not been defined clearly. Therefore a few previous studies concerning the role of LHb and LHCGR polymorphisms in PCOS have provided conflicting results such as [13-15].

Moreover, LHCGR polymorphisms may also lead to biological activity changes that support the worse effect of abnormal LHB action because of its transmembrane integral protein acts as a G-protein-coupled receptor superfamily controlling the LHB level in the theca cell membrane [5-18].

It has been reported that PCOS women of different ethnicity display different clinical manifestations. Amal Salah et al revealed that LHβ G1052A and LHCGR G935A genes polymorphisms are associated with increased risk of PCOS in Egyptian women, especially in obese cases because obesity aggravates menstrual irregularity and increases serum total testosterone level. Thus, Hyperandrogenism can affect the follicle growth and metabolic process, playing a key role in the occurrence and the development of PCOS in the obese women [19,20]. Additionally, two missense point mutations were discovered among PCOS women in LHβ gene (Trp 8 Arg and Ile 15 Thr) with obesity association only in Japanese population, while obese PCOS women from north European countries showed lower frequency [5]. Also, another variation that occurs due to ethnicity factor reported that LHβ gene variant G102A in exon 3 (Gly102Ser) was found to be higher in Singapore Chinese women who had menstrual disorders in contradiction to Korean research that found no difference in elevated blood glucose associated with insulin resistance as a metabolic feature of PCOS also associated with LHb 1052A allele [24-26]. High level of blood glucose associated with insulin resistance as a metabolic feature of PCOS also associated with LHB 1052A allele. While LHB 1052A allele carriers showed lower LH level and higher total and free testosterone levels due to SNPs effect on gene bioactivity.

The aim of computational biology is to permit researchers to perform experiments in silico on digital virtual models of biological molecules such as cells, organs and organisms. These virtual models hold great promise for enhancing the pace and covering the scope of biomedical research by freeing scientists from biological experimental barriers associated with ethical restrictions [27-30].

This leading and unique study benefits from the in silico analysis tools to build up knowledge about PCOS and to determine the influence of various SNPs in LH-B gene using in silico prediction software and assessing their effects on the structure/function of LH protein that may have an important role in disease susceptibility. This role will be examined through wet laboratory work as a part of PhD project that will enhance in understanding the genetic mechanisms of the disease through these useful tools. We followed the same protocol that was used for FSH-B gene analysis [31].

2. Material and Methods

2.1. Data retrieval: The data on human LH-B gene was collected from the National Center for Biological Information (NCBI) web site. The SNP information (protein accession number and SNP ID) of the LH gene was retrieved from the NCBI dbSNP (http://www.ncbi.nlm.nih.gov/snp/) [1] and the protein sequence was collected from Uni Prot databases [3].

2.2. SIFT (Sorting Intolerant from Tolerant): https://sift.bii.a-star.edu.sg/: Is an online tool that predicts if an amino acid substitution affects protein function or not by using sequence homology. It performs analysis based on different algorithms and it interprets the homologous sequences using the Swiss-Prot (version 51.3) and TrEMBL (version 34.3). It gives scores to each amino acid residue ranging from zero to one. The threshold intolerance score for SNPs is 0.05 or less [32].

2.3. Polyphen-2 (Polymorphism Phenotypingv2), http://genetics.bwh.harvard.edu/pph2/: Is used to predict the possible impact of an amino acid substitution on both structure and function of protein by analysis of multiple sequence **alignment** and protein 3D structure. It estimates the position specific independent count score (as probably damaging, possibly damaging or benign according to the score ranging from (0–1). [33].

PSIC) for every variant and then determines the difference between them, the higher the PSI, the higher the functional impact of the amino acid on the protein function may be. Prediction outcomes could be classified

2.4. Provean (Protein Variation Effect Analysis) (http://provean.jcvi.org/index.php): This 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 [34].

2.5. SNAP2 (Screening of Non-Acceptable Polymorphism 2; https://rostlab.org/services/snap2web/): 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 a neutral, non- neutral effect [35].

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, which 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 [36].

2.7. SNPs &GO (Single nucleotide polymorphism & Gene Ontology, PHD-SNP (http://snps.biofold.org/snps-and-goinformation

evolutionary information and function as encoded in the Gene Ontology terms and other outperforms with available predictive methods [37].

2.8. P-mut (http://mmb2.pcb.ub.es:8080/PMut): Is 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 [38].

2.9. I–Mutant 3.0 (http://gpcr2.biocomp.unibo.it/cgi/**predictors/I-Mutant3.0/ I-Mutant3.0.cgi):** It is a neural network based tool, predicts the change in the stability of the protein upon mutation The output is obtained in the form of protein stability change upon mutation and Gibbs-free energy change (DDG) either increased or decreased stability [39].

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 [40].

2.11. RaptorX (http://raptorx.uchicago.edu/): 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 the tertiary structure [41].

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 3un-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 [42].

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 [43].

3. Results

Using dbSNP database for gene data set, the human *LH-B* gene, gene of our interest, contained a total of 39SNPs missense nsSNPs, 13 nsSNPs of them were predicted to have the most damaging effects on LH-B protein's structure and function.

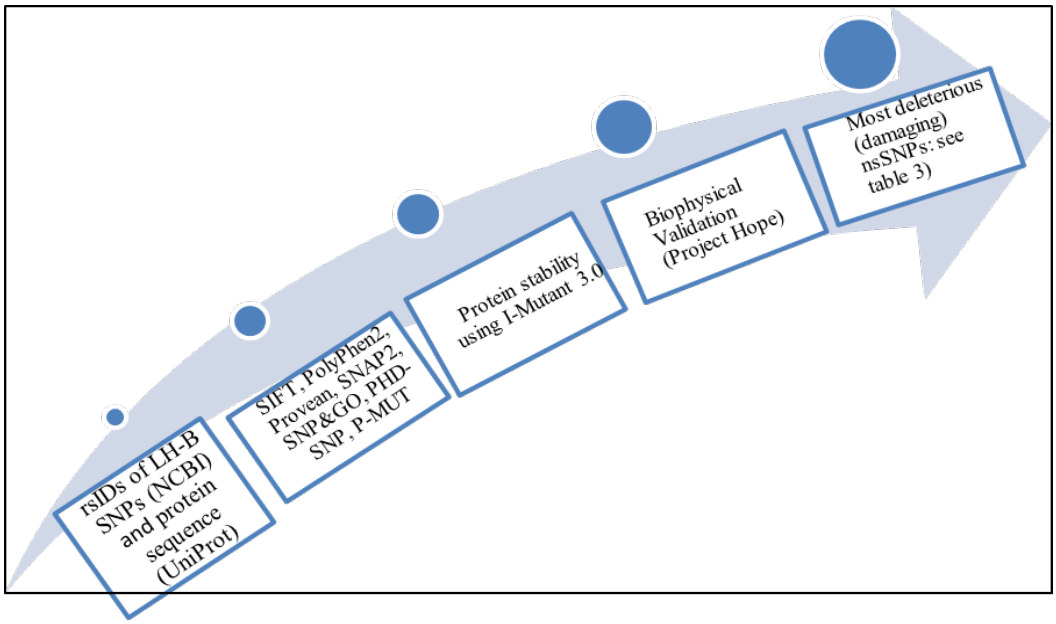

Figure 1. Diagrammatic representation of LH-B gene in silico work flow

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

dbSNP rs#	Substitution	SIFT PREDICTION	SIFT SCORE	POLPYHEN 2 PREDICTION	PLYPHEN2 SCORE	SNAP2 PREDICTION	SNAP 2 SCORE	Prediction (cutoff= -2.5)	PROVEAN score
rs749399282	R26W	affect	0.01	probably damaging	1	effect	84	Deleterious	-3.174
rs1337220278	P27L	affect	0.01	probably damaging	1	effect	42	Deleterious	-2.812
rs1330471939	C29Y	affect	0	probably damaging	1	effect	88	Deleterious	-9.723
rs370066297	N33S	affect	0.01	probably damaging	0.972	effect	54	Deleterious	-4.466
rs1007910415	E39K	affect	0	probably damaging	1	effect	76	Deleterious	-3.805
rs1468963976	K40Q	affect	0.01	probably damaging	1	effect	60	Deleterious	-3.325
rs1267880224	E41A	affect	0.01	probably damaging	0.989	effect	35	Deleterious	-4.76
rs375879463	I47N	affect	0.02	probably damaging	0.997	effect	67	Deleterious	-3.878
rs1236916690	T51N	affect	0	probably damaging	0.994	effect	88	Deleterious	-4.777
rs375748166	A55V	affect	0.01	probably damaging	1	effect	76	Deleterious	-3.036
rs121912517	G56D	affect	0	probably damaging	1	effect	94	Deleterious	-6.734
	G56A	affect	0	probably damaging	1	effect	84	Deleterious	-5.755
rs143118489	G56S	affect	0	probably damaging	1	effect	90	Deleterious	-5.755
rs1454408984	C58S	affect	0	probably damaging	0.998	effect	86	Deleterious	-9.637
rs1447389231	T60I	affect	0	probably damaging	1	effect	76	Deleterious	-5.19
rs746121371	L69P	affect	0	probably damaging	0.999	effect	65	Deleterious	-4.869
rs1395465019	L72R	affect	0.01	probably damaging	0.979	effect	15	Deleterious	-3.369
rs5030773	Q74R	affect	0	probably damaging	0.998	effect	86	Deleterious	-3.965
rs1303660947	V76L	affect	0.01	probably damaging	0.974	effect	63	Deleterious	-2.782
rs750110473	R80G	affect	0.01	possibly damaging	0.882	effect	21	Deleterious	-4.049
	R80C	affect	0	probably damaging	0.999	effect	46	Deleterious	-5.818
rs754210093	D81Y	affect	0	probably damaging	0.997	effect	76	Deleterious	-5.687
rs764734315	V82A	affect	0.01	probably damaging	0.994	effect	48	Deleterious	-2.513
rs776040203	R83G	affect	0.03	probably damaging	0.999	effect	78	Deleterious	-4.676
	R83C	affect	0.02	probably damaging	1	effect	63	Deleterious	-5.479
rs138049984	F84L	affect	0.01	probably damaging	0.99	effect	56	Deleterious	-4.326
rs780038003	S86C	affect	0	probably damaging	0.998	effect	38	Deleterious	-2.67
	S86F	affect	0	possibly damaging	0.927	effect	46	Deleterious	-3.567
rs756933264	G91A	affect	0.03	probably damaging	1	effect	35	Deleterious	-3.009
rs766619478	G95D	affect	0	probably damaging	1	effect	53	Deleterious	-5.667
rs149003040	V96M	affect	0	probably damaging	1	effect	61	Deleterious	-2.562
rs760453863	P98H	affect	0	probably damaging	1	effect	74	Deleterious	-7.914
	P103S	affect	0.01	possibly damaging	0.456	effect	71	Deleterious	-7.344
rs771999985	V104A	affect	0.01	probably damaging	0.999	effect	53	Deleterious	-3.68
rs778529225	A105P	affect	0	probably damaging	1	effect	92	Deleterious	-4.764

rs1187747882/ rs749003379	S107R	affect	0	probably damaging	0.999	effect	33	Deleterious	-4.7
rs528776938	R114C	affect	0.01	probably damaging	0.998	effect	29	Deleterious	-3.509
rs759571421	C120W	affect	0	probably damaging	1	effect	64	Deleterious	-10.339
rs1420843452	C120R	affect	0	probably damaging	1	effect	81	Deleterious	-11.415

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

dbSNP rs#	Substitution	Snp and go Prediction	RI	Probability	PHD Prediction	RI	Probability	PMUT Prediction	PMUT Probability	I mutant prediction	I mutant probability
rs1330471939	C29Y	Disease	8	0.878	Disease	8	0.909	Disease	0.73 (87%)	decrease	-0.27
rs375879463	I47N	Disease	3	0.651	Disease	4	0.694	Disease	0.66 (85%)	decrease	-1.47
rs121912517	G56D	Disease	6	0.789	Disease	8	0.897	Disease	0.62 (83%)	decrease	-0.99
	G56A	Disease	4	0.699	Disease	5	0.765	Disease	0.62 (83%)	decrease	-0.77
rs143118489	G56S	Disease	6	0.8	Disease	7	0.865	Disease	0.62 (83%)	decrease	-1.08
rs1454408984	C58S	Disease	6	0.822	Disease	8	0.896	Disease	0.62 (83%)	decrease	-0.81
rs5030773	Q74R	Disease	6	0.824	Disease	6	0.802	Disease	0.67 (85%)	decrease	-0.2
	R80C	Disease	1	0.551	Disease	4	0.689	Disease	0.63 (84%)	decrease	-1.22
rs760453863	P98H	Disease	1	0.548	Disease	4	0.685	Disease	0.57 (81%)	decrease	-1.73
rs771999985	V104A	Disease	4	0.721	Disease	4	0.69	Disease	0.52 (79%)	decrease	-1.64
rs778529225	A105P	Disease	4	0.718	Disease	7	0.833	Disease	0.62 (83%)	increase	0.05
rs759571421	C120W	Disease	4	0.696	Disease	4	0.707	Disease	0.60 (83%)	decrease	-0.11
rs1420843452	C120R	Disease	5	0.729	Disease	4	0.72	Disease	0.60 (83%)	decrease	-0.34

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

Location	dbSNP ID	Variant type	Wobble base pair	Ancestral Allele	Allele	miR ID	Conservation	miRSite	Function Class	Exp Support	context+ score change
49519269	rs189656791	SNP	Y	G	A	hsa-miR-342-5p	2	CACCCCAatcctc	C	N	-0.172
						hsa-miR-4651	2	CACCCCAatcctc	C	N	-0.161
						hsa-miR-4664-5p	2	CACCCCAatcctc	C	N	-0.185
						hsa-miR-608	2	CACCCCAatcctc	C	N	-0.165
						hsa-miR-6747-5p	2	CACCCCAatcctc	C	N	-0.218
49519271	rs141729250	SNP	N	C	C	hsa-miR-6747-5p	2	gACACCCCgatcc	D	N	-0.181
49519293	rs200045996	SNP	N	C	C	hsa-miR-4785	7	ccatcCCGACTCc	D	N	-0.222
					T	hsa-miR-2392	2	CCATCCTgactcc	C	N	-0.167
						hsa-miR-5702	7	ccatcCTGACTCc	C	N	-0.116

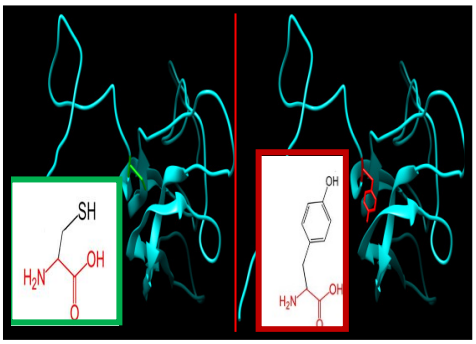

Figure 2. SNP ID: rs1330471939, (C29Y) The amino acid Cysteine in the native protein (green box) changed to Tyrosine in the mutant protein (red box) at position 29

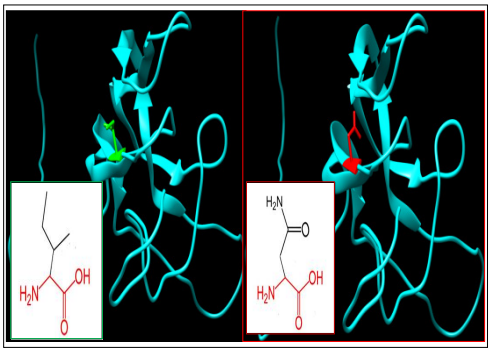

Figure 3. SNP ID: rs375879463, (I47N) The amino acid Isoleucine in the native protein (green box) changed to Asparagine in the mutant protein (red box) at position 47

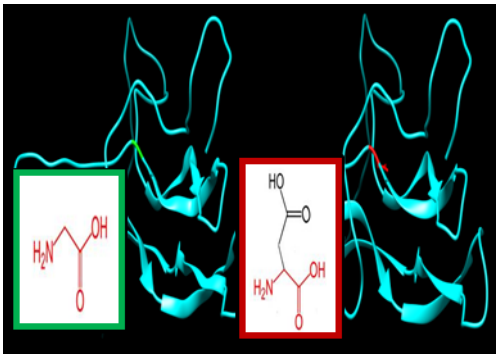

Figure 4. SNP ID: rs121912517, (G56D) The amino acid Glycine in the native protein (green box) changed to Aspartic Acid in the mutant protein (red box) at position 56

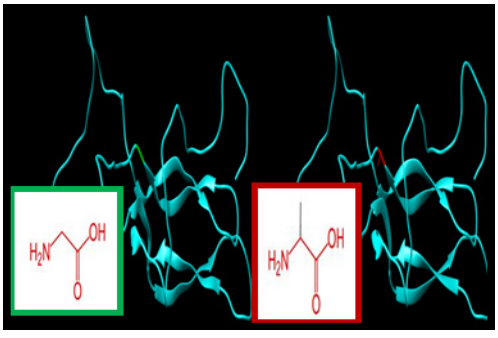

Figure 5. SNP ID: rs121912517, (G56A) The amino acid Glycine in the native protein (green box) changed to Alanine in the mutant protein (red box) at position 56

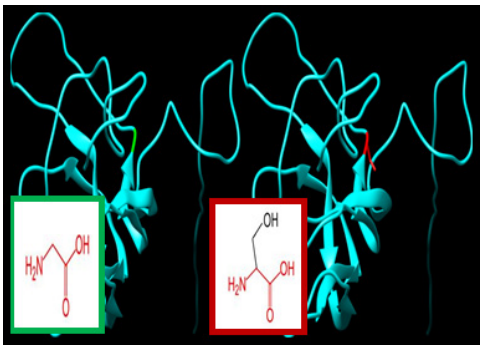

Figure 6. SNP ID: rs143118489, (G56S) The amino acid Glycine in the native protein (green box) changed to Serine in the mutant protein (red box) at position 56

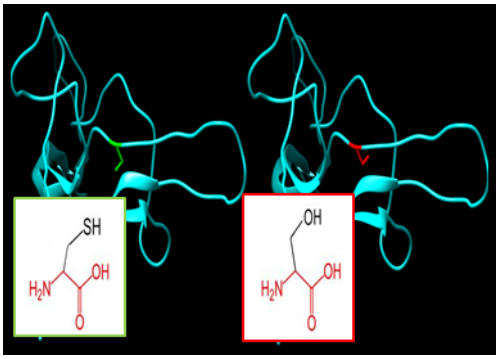

Figure 7. SNP ID: rs1454408984, (C58S) The amino acid Cysteine in the native protein (green box) changed to Serine in the mutant protein (red box) at position 58

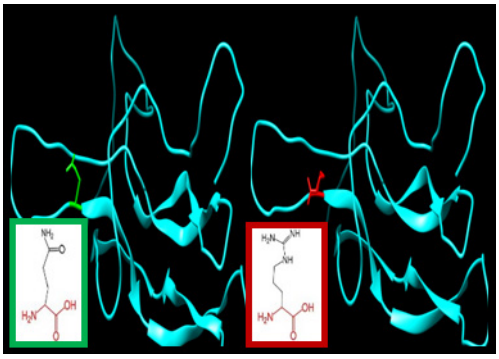

Figure 8. SNP ID: rs5030773, (Q74R) The amino acid Glutamine in the native protein (green box) changed to Arginine in the mutant protein (red box) at position 74

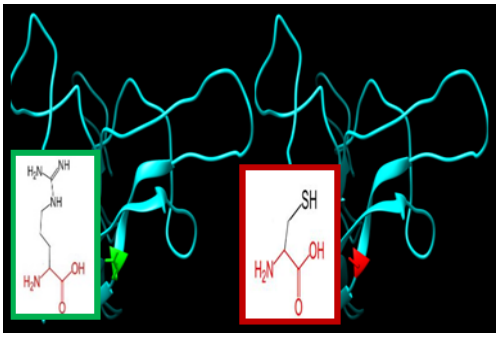

Figure 9. SNP ID: rs5030773, (R80C) The amino acid Glutamine in the native protein (green box) changed to Arginine in the mutant protein (red box) at position 80

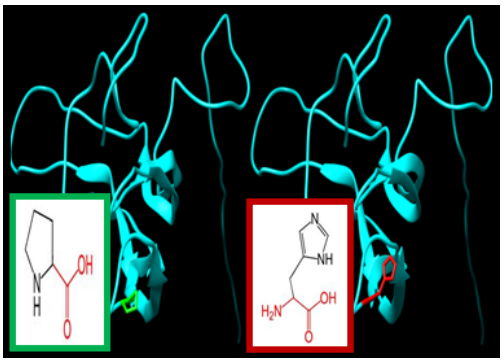

Figure 10. SNP ID: rs760453863, (P98H) The amino acid Proline in the native protein (green box) changed to Histidine in the mutant protein (red box) at position 98

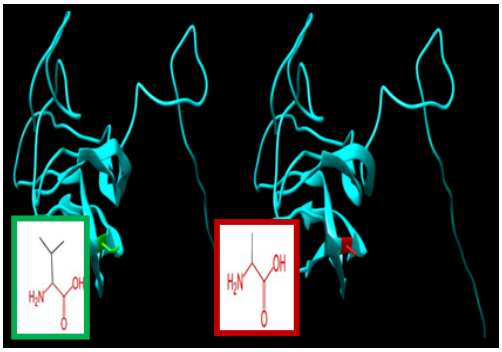

Figure 11. SNP ID: rs771999985, (V104A) The amino acid Valine in the native protein (green box) changed to Alanine in the mutant protein (red box) at position 104

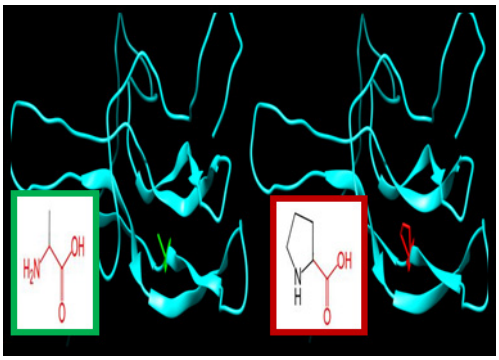

Figure 12. SNP ID: rs778529225, (A105P) The amino acid Alanine in the native protein (green box) changed to Proline in the mutant protein (red box) at position 105

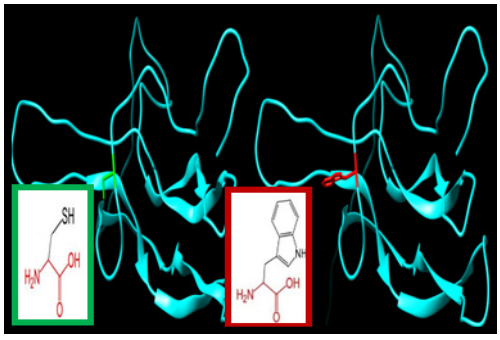

Figure 13. SNP ID: rs759571421, (C120W) The amino acid Cysteine in the native protein (green box) changed to Tryptophan in the mutant protein (red box) at position 120

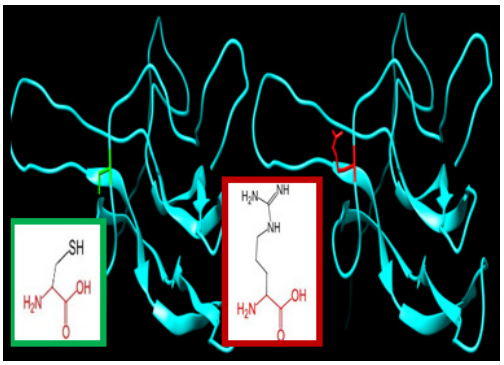

Figure 14. SNP ID: rs1420843452, (C120R) The amino acid Cysteine in the native protein (green box) changed to Arginine in the mutant protein (red box) at position 120

4. Discussion

LHB gene was investigated in dbSNP National Centre of Biotechnology Information (NCBI public database). This gene contains a total of 404 SNPs in the coding region; of which 140 are missense, 67 synonymous, five nonsense, seven frame shift and 444 are in the non-coding region, of which 300 in the 3'un-translated region (3' UTR) and 18 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 2).

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 affect protein stability; only one SNP increases protein stability (rs778529225, (A105P)) while the rest decrease its stability. Alteration of one amino acid into another mutated residue results in increasing the size of the protein in different locations such as (rs759571421, (C120W), (rs1420843452, (C120A) and rs760453863, (P98H)) and only two mutations lead to declining the size of it (rs5030773, (R80C) and rs771999985, (V104A)).

rs1330471939, (C29Y); In this location, Tyrosine (mutant residue) is bigger than Cysteine the (wild-type residue) which is probably will not fit for cysteine bridge formation which is important for protein stability and this consequently lead to loss of the hydrophobic interactions in the core of the protein. OR This mutation lead to decrease protein stability due to the absence of cysteine bridge, which is important for the stability of the protein and the mutant residue Tyrosine is bigger in size than Cysteine which also results in loss of hydrophobic interactions in the core of the protein.

rs375879463 (I47N): alteration of Isoleucine to Asparagine in this location which is nearby to a residue that makes a cysteine bond will affect the binding site of other

molecules and loss of hydrophobic interactions in the core of the protein.

In position 56; three different mutations occur (rs121912517 (G56D and (G56A) and rs143118489 (G56S)): and affect the physiochemical properties of the protein; all mutated residues lead to decrease stability and increase the size of the protein. Alteration of Glycine to Aspartic Acid with a negative charge which causes repulsion between the mutant residue and neighbouring residues and can disturb interactions with other molecules or other parts of the protein. The torsion angles for this residue are unusual due to the inflexibility of the mutant residue that results in an incorrect conformation and will disturb the local structure of the protein. Moreover, conversion of Glycine to Alanine leads to disturbing interactions with other molecules or other parts of the protein and probably damaging the protein. Also in the same location with other rs143118489 (G56S) Glycine is converted to Serine and results in the formation of unusual torsion angles which probably results into an incorrect conformation and will disturb the local structure of the protein.

rs1454408984 (C58S): When Cysteine (wild-type residue) is mutated to Serine (mutant residue) it decreases protein stability due to absences of cysteine bridge and loss of hydrophobic interactions with other molecules on the surface of the protein.

rs760453863(P98H): The Histidine (mutant residue) is bigger than Proline the (wild-type residue) but with less hydrophobicity and more flexibility which disturbs the special conformation of the wild-type residue and might cause loss of hydrophobic interactions in the core of the protein.

rs771999985(V104A): In this location, the mutation might causes an empty space in the core of the protein because Alanine (mutant residue) is smaller than Valine the (wild-type residue) and cause slightly destabilization of the protein but not totally damage the protein structure.

rs778529225(A105P): Only in this location the mutation leads to increase protein stability with bigger size of the mutant residue Proline and probably lead to damage of the protein because it is located near a highly conserved position in the core of the protein.

Two mutations occur in position 120 with different consequence on the stability and the neutrality of the protein; one when Cysteine (wild-type residue) is mutated to Tryptophan (mutant residue), rs759571421 (C120W), it decreases protein stability due to absences of cysteine bridge and loss of hydrophobic interactions with other molecules on the surface of the protein. And the other exchange (rs1420843452 (C120A)); when it is mutated to Arginine with a positive charge and bigger size than Cysteine, which causes repulsion between the mutant residue and neighbouring residues because it is located on the surface of the protein and disturbs interactions with other molecules or other parts of the protein.

Historically, women with PCOS characterized by a high level of serum LH, high LH/FSH ratio with normal/low FSH level but less is known about the genetic background that leads to overexpression of LH-B gene and its association with PCOS [15-16]. Regarding our findings, two of our most damaging SNPs were reported among the male population suffering from hypogonadism problems associated with low LH level which is uncommon among PCOS population associated with anovulatory and infertility problems [44].

Regarding substitution of Glutamine to mutated residue Arginine (rs5030773, Q74R) results in unbinding of luteinizing hormone to its receptors, which consequently, results in failure of puberty to develop spontaneously, with infertility in the homozygous state [45]. In male heterozygotes, the effect of the mutation is more subtle, causing impaired steroidogenesis and a high incidence of infertility, despite the development of normal secondary sex characteristics while a female with the heterogeneous pattern has normal sexual development and fertile which is not agree with the evidence that PCOS associated with high LH level [46-47].

Most reported mutations such as LHβ G1052A and LHβ gene (Trp 8 Arg and Ile 15 Thr) among different populations are with contradictory due to ethnicity and geographical distributions failed to success when analysed by our protocol and did not have an effect on protein structure [19-25]. Therefore, consideration should be taken to these thirteen novel mutations when we are carrying out genetic studies through human samples as the real benefit of computational biology that will contribute in understanding the pathophysiology of the disease that may be useful in discovering new diagnostic and therapeutic markers. Moreover, mutations occur in the 3'UTR region associated with abnormal expression of the gene among target organ explain the abnormal ovulation process which occurs in collaboration 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.

Taking these findings into consideration with the hormonal profiles of women with PCOS will increase our knowledge about the consequences of these mutations that result in overexpression of LH level associated with anovulation problems.

5. Conclusions

13 nsSNPs with different positions were predicted to be the most damaging mutations for beta subunit of LH 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 nine SNPs at the 3UTR were predicted to disrupt miRNA binding sites and hence affect the gene expression function that might be associated with anovulation problems.

Uses of in silico analysis tools that allow 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] LHB luteinizing hormone beta polypeptide [*Homo sapiens* (human)] Gene ID: 3972, updated on 21-Apr-2019 Retrieved from https://www.ncbi.nlm.nih.gov/gene/3972.

[2] LUTEINIZING HORMONE, BETA POLYPEPTIDE; LHB. Retrieved from https://omim.org/entry/152780.

[3] Gene: LHB ENSG00000104826 Retrieved from http://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000104826;r=19:49015980-49017091.

[4] UniproktB - P01229 (LSHB_HUMAN) Retrieved from https://www.uniprot.org/uniprot/P01229.

[5] Mariani Carla Prudente Batista, Eliane de Fatima Duarte, Michele Delarmelina dos Reis Borba, Emilie Zingler, João Mangussi-Gomes, Beatriz Taynara Araújo dos Santos, Olivia Laquis de Moraes, Sylvia Asaka Yamashita Hayashida, Edmund C. Baracat, Francisco de Assis da Rocha Neves, Gustavo Arantes Rosa Maciel, Tania Aparecida Sartori Sanchez Bachega, Gustavo Barcelos Barra, Adriana Lofrano-Porto, Trp28Arg/Ile35Thr LHB gene variants are associated with elevated testosterone levels in women with polycystic ovary syndrome, Gene, Volume 550, Issue 1, 2014, Pages 68-73.

[6] Tsilchorozidou, T., Overton, C. and Conway, G. (2004). The pathophysiology of polycystic ovary syndrome. Clinical Endocrinology, 60: 1-17. doi:10.1046/j.1365-2265.2003.01842.x.

[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] Stein I, Leventhal M (1935). Amenorrhoea associated with bilateral polycystic ovaries. Am J Obstet Gynecol 29: 181–185.

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

[10] 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.

[11] 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 syndromeHuman Reproduction, Vol.24, No.8 pp. 2014–2022.

[12] 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).

[13] 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.

[14] FJ Broekmans, EAH Knauff, et al. a PCOS according to the Rotterdam consensus criteria: change in prevalence among WHO-II anovulation and association with metabolic factors BJOG 2006; 113:1210–1217.

[15] 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.

[16] The Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). Hum Reprod 2004; 19:41–7.

[17] Azziz R. Diagnostic criteria for polycystic ovary syndrome: a reappraisal. Fertil Steril 2005;83:1343–6.

[18] Ha L, Shi Y, Zhao J, Li T, Chen Z-J (2015). Association Study between Polycystic Ovarian Syndrome and the Susceptibility Genes Polymorphisms in Hui Chinese Women. PLoS ONE 10(5): e0126505. doi:10.1371/journal.pone.0126505.

[19] El‐Shal, A. S., Zidan, H. E., Rashad, N. M., Abdelaziz, A. M. and Harira, M. M. (2016). Association between genes encoding components of the Leutinizing hormone/ Luteinizing hormone–choriogonadotrophin receptor pathway and polycystic ovary syndrome in Egyptian women. IUBMB

Life, 68: 23-36.

[20] Bassiouny YA1, Rabie WA, Hassan AA, Darwish RK. Association of the luteinizing hormone/choriogonadotropin receptor gene polymorphism with polycystic ovary syndrome. Gynecol Endocrinol. 2014 Jun; 30(6): 428-30.

[21] Shilpi Dasgupta, P.V.S. Sirisha, K. Neelaveni, K. Anuradha, G. Sudhakar, B. Mohan Reddy, Role of luteinizing hormone β-subunit gene variants among South Indian women with polycystic ovary syndrome, Gene,Volume 494, Issue 1, 2012, Pages 51-56.

[22] Thathapudi, S., Kodati, V., Erukkambattu, J., Addepally, U., & Qurratulain, H. (2015). Association of luteinizing hormone chorionic gonadotropin receptor gene polymorphism (rs2293275) with polycystic ovarian syndrome. Genetic testing and molecular biomarkers, 19(3), 128–132. doi:10.1089/gtmb.2014.0249.

[23] Ramanujam, L. N., Liao, W. X., Roy, A. C., Loganath, A., Goh, H. H. and Ng, S. C. (1999). Association of molecular variants of luteinizing hormone with menstrual disorders. Clinical Endocrinology, 51: 243-246.

[24] L. Skrgatic, D. Pavicic Baldani, J.Z. Cerne, P. Ferk, K. Gersak, CAG repeat polymorphism in androgen receptor gene is not directly associated with polycystic ovary syndrome but influences serum testosterone levels.The Journal of Steroid Biochemistry and Molecular Biology, Volume 128, Issues 3–5, 2012.Pages 107-112.

[25] Liu N1, Ma Y, Wang S, Zhang X, Zhang Q, Zhang X, Fu L, Qiao J. Association of the genetic variants of luteinizing hormone, luteinizing hormone receptor and polycystic ovary syndrome. Reprod Biol Endocrinol. 2012 Apr 30;10:36.

[26] Alviggi, C., Pettersson, K., Longobardi, S., Andersen, C. Y., Conforti, A., De Rosa, P., … Humaidan, P. (2013). A common polymorphic allele of the LH beta-subunit gene is associated with higher exogenous FSH consumption during controlled ovarian stimulation for assisted reproductive technology. *Reproductive biology and endocrinology*: RB&E, 11, 51. doi:10.1186/1477-7827-11-51.

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

[28] 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.

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

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

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

[32] 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.

[33] 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, Unit 7.20.

[34] 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.

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

[36] Wu, J., & Jiang, R. (2013). Prediction of deleterious nonsynonymous single-nucleotide polymorphism for human diseases. TheScientificWorldJournal, 2013, 675851.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] LLOYD AXELROD, ROBERT M NEER, BERNARD KLIMAN, Hypogonadism in a Male with Immunologically Active, Biologically Inactive Luteinizing Hormone: An Exception to a Venerable Rule, *The Journal of Clinical Endocrinology & Metabolism,* Volume 48, Issue 2, 1 February 1979, Pages 279–287, https://doi.org/10.1210/jcem-48-2-279.

[45] Weiss, Jeffrey, Axelrod, Lloyd, Whitcomb, Randall W. Harris, Philip E. Crowley, William F. Jameson, J. Larry. (1992). Hypogonadism Caused by a Single Amino Acid Substitution in the β Subunit of Luteinizing Hormone. New England Journal of Medicine VI - 326.

[46] Hernan Valdes-Socin, Roberto Salvi, Albert Thiry, Adrian F. Daly, François P. Pralong, Rolf Gaillard, Albert Beckers, Testicular Effects of Isolated Luteinizing Hormone Deficiency and Reversal by Long-Term Human Chorionic Gonadotropin Treatment, *The Journal of Clinical*

Endocrinology & Metabolism, Volume 94, Issue 1, 1 January 2009, Pages 3–4.

[47] Valdes-Socin, Hernán Salvi, Roberto Daly, Adrian Gaillard, Rolf C.- Quatresooz, Pascale Tebeu, Pierre-Marie Pralong, François P. Beckers, Albert (2004). Hypogonadism in a Patient with a Mutation in the Luteinizing Hormone Beta-Subunit Gene. New England Journal of Medicine VI - 351 2619-2625.