Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 7, 2021, Pages. 1688 - 1695
Received 05 May 2021; Accepted 01 June 2021.

Androgen Receptor Gene Polymorphism and Polycystic Ovarian Syndrome: A Detailed *In-Silico* Study

Nasreen sultana[1], Sk. Jasmin[2], Md. Azeeza Siddiqa[3], K. kasturi[*]

ABSTRACT

Background: Polycystic ovary syndrome (PCOS) is characterized by adulatory dysfunction and hyper androgenism. Its etiopathology is not well understood but genetic factors seem to have a role. Polymorphism of the androgen receptor (AR) gene has been associated with different androgen pattern diseases.

Objective: To analyze functional and structural consequences of the nsSNPs of AR gene polymorphisms associated with PCOS.

Methodology: We have utilized numerous computational tools like SIFT, POLYPHEN, PROVEAN, I-MUTANT, PANTHER, PHD-SNP, SNP&GO, PMUT, and HOPE for the exploration the AR gene polymorphisms associated with PCOS.

Results: We filtered the most pathological mutations by combining the scores of the aforementioned servers and found five SNPs (R775C, H875Y, Y764C, L708R, C580F) as deleterious and disease causing. The findings implicate that this nsSNPs would possible association with the protein deteriorating and disease causal potentialities.

KEYWORDS: Computational tools, Polycystic Ovary Syndrome, Non-Synonymous Single Nucleotide Polymorphism, Androgen receptor, PMUT, polymorphisms.

INTRODUCTION:

Polycystic ovary syndrome (PCOS) is one of the most common endocrinopathies in young women[1]. Polycystic ovary syndrome impacts millions of women worldwide, with a prevalence of 6–10% in premenopausal women[2]. It is a heterogeneous disorder of unclear etiopathogenesis but there is evidence of the participation of a genetic component[3]. Hyper-androgenism is the hallmark of PCOS. Androgen gene is an X-linked gene at position Xq11-12 of the long arm (q) on the chromosome with eight exons and introns varying in size between 0.7 and 2.6 kb, encodes a protein called Androgen, which contains 919 amino acids[4]. It regulates the transcription of specific genes[5]. AR gene is a family of Steroid hormone and it contains three active domains includes: N-terminal domain (NTD), DNA binding domain (DBD) (Hold two zinc molecules) Ligand-Binding domain (LBD) (Provides Androgen regulation, Transcriptional activity). In central, between the DBD and LBD Hinge region is present[6] (Figure-1)

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 7, 2021, Pages. 1688 - 1695
Received 05 May 2021; Accepted 01 June 2021.

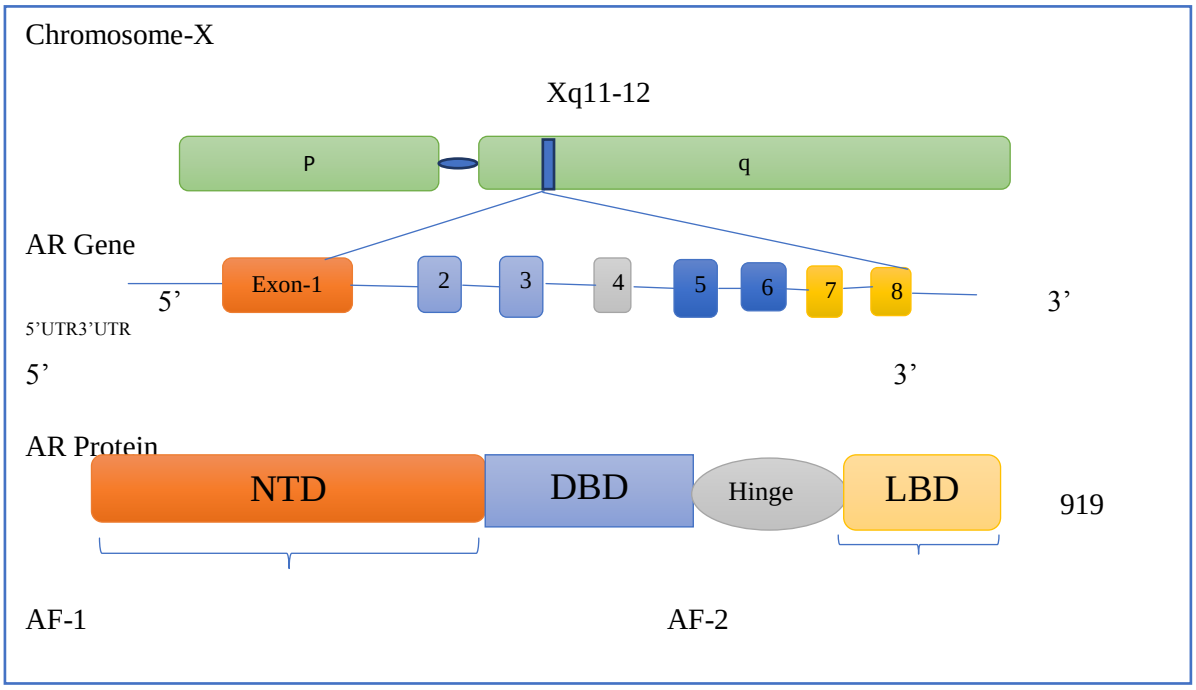


Figure-1 Depicts chromosomal location of Human AR gene and Protein

High serum androgen levels are associated with the inhibition of follicular development, an-ovulation, irregular menstruation, appearance of small cysts in the ovary[7] and follicular apoptosis[8]. Insulin[9] and LH levels are also elevated in many PCOS patients Women with polycystic ovaries arise because of excessive androgen action.

Single nucleotide polymorphism (SNP) plays a crucial role to identify common genetic variants and a potential biomarker for investigating the deleterious and neutral effects on protein function. The extensive study of the SNPs on the biological system is not easy, because SNPs have functional consequences like causing amino acid changes, changes to mRNA transcript stability, and changes to transcription factor binding affinity[10]. *In silico* studies are cost effective, less time consuming, easy to carry out, and more reliable as compared to experimental procedures[11]. The present study is undertaken to explore the nsSNPs of *AR* gene responsible for PCOs in women by using computational tools to predict the deleterious nature of the mutation and structural analysis.

MATERIALS AND METHODS:

The raw data of AR gene was retrieved from NCBI. Reference Id of the AR gene and protein sequence was retrieved from Ensemble and Uniprot tools. We have utilized nine computational tools (mentioned in figure-2) to examine the damaging, deleterious and disease associated SNPs.

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 7, 2021, Pages. 1688 - 1695
Received 05 May 2021; Accepted 01 June 2021.

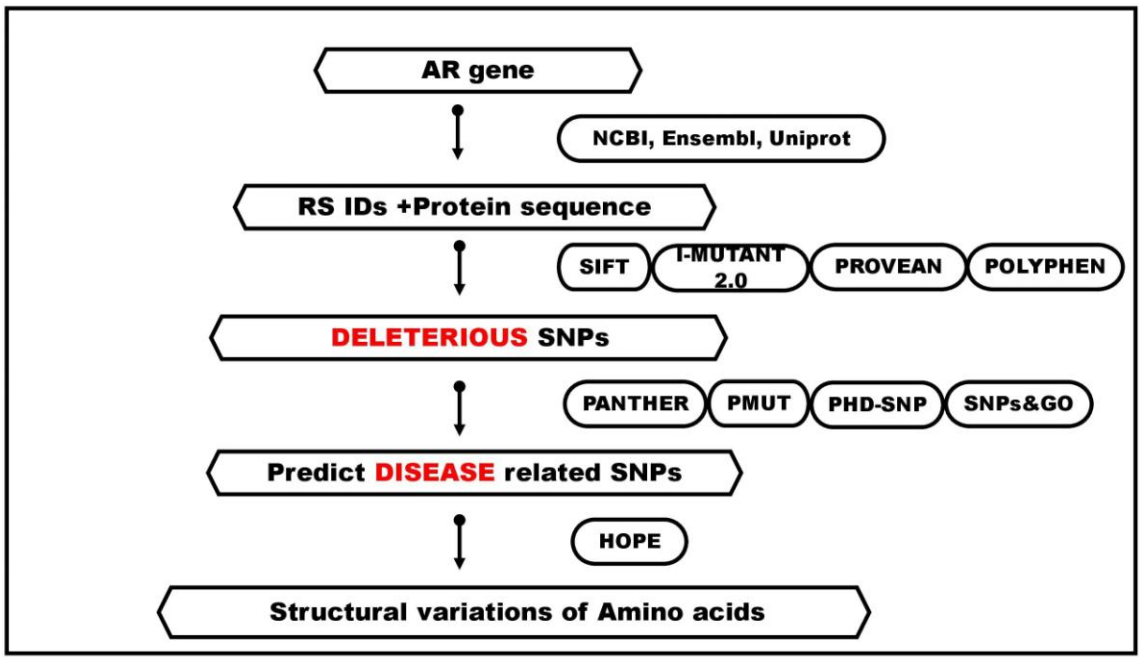


Figure-2 Diagrammatic representation of AR gene *In-silico* flow work

Prediction of Damaging and deleterious SNPs:

SIFT: It determines the amino acid replacement will contribute to protein function and mutations may be phenotype which is predicted on sequence homology.[10]SIFT grades categorized as Deleterious (≤0.05) and tolerated (>0.05).[12]

POLYPHEN-2: It winds up the possible sequence of single amino acid which deposits in both structure and function of protein activity, the grades categorized as probably damaging, possibly damaging (≤0.5) and benign (≥0.51).[13]

PROVEAN: It impacts based upon alternation protein sequence on performance of protein. If the grade is limit of 2.5, is assessed as deleterious, even if the grade is above 2.5, is assessed as neutral.[14]

I-MUTANT is a support vector machine (SVM) based tool, for the automatic prediction of mutations in the stability of protein caused by single point mutation. It predicts that SNPs are increase and decrease. Increased stability of protein having DDG>0 rate, whereas decreased stability of protein having DDG<0 rate with association with Gibbs free energy.[15]

Prediction of Disease related SNPs:

PANTHER is a web server predicted on prevalence occurrence of amino acid at a specific position in protein sequence competently related on evolution. The threshold subpsec -3 has been designated, below which the predictions are assessed as damaging.[16]

PMUT is a web server predicts a transformation pathogenicity index 0 to 1. If the grade is >0.5 considered as pathological mutation, while, <0.5 recommends neutral.[17]

SNP&GO it works based on alternation of protein by using gene ontology terms. It is a support vector machine-based classifier; the grades distinguish between pathological and neutral variations. One key feature of this server is displaying the PHD-SNP and PANTHER algorithm grades. In the event that the probability range is elevated above 0.5 which is considered as pathological, even if the range is not higher than 0.5 suggests neutral.[18]

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 7, 2021, Pages. 1688 - 1695
Received 05 May 2021; Accepted 01 June 2021.

PHD-SNP predicts the amino acid substitutions are associated with disease or neutrality, with a degree of reliability.[19, 20]

HOPE is an automated algorithm that analyses all of the data collected along with known features of the wild-type and mutant amino acids, such as size, charge, and hydrophobicity, to forecast how the mutation would affect the protein's structure and function[21].

RESUITS AND DISCUSSION:

To determine the deleterious non-synonymous single nucleotide polymorphism (nsSNPs), which might be involved in inducing disease associated phenomenon, is now among the most important field of computational genomic research. The disease associated mutations can be identified with the help of genome sequencing and its analysis. The advanced method in computational biology has now enabled us to determine the deleterious nsSNPs in the target candidate genes. Computational methods were applied to study the protein structural and functional effect on point mutation at molecular level. In this investigation we implemented multiple computational methods to identify the most likely pathogenic mutations in AR gene. Our results also revealed that implementations of different algorithms often serve as powerful tools for prioritizing candidate functional nsSNPs. Here we used SIFT, POLYPHEN, PROVEAN, I-MUTANT 2.0, PANTHER, PHD-SNP, SNP&GO and PMUT tools to examine the most deleterious and disease association nsSNPs from the SNP dataset.

Screening of Deleterious nsSNPs:

In SIFT, out of 20 mutations 16 mutations were predicted to be deleterious with the score≤0.05 remaining 4 were predicted as tolerated. A total of 19 nsSNPs were predicted to be damaging and the remaining one SNP was categorized as benign with POLYPHEN 2.0. All the nsSNPs submitted to Polyphen-2.0 and SIFT were also submitted as input to the PROVEAN and I-MUTANT 2.0 server. The PROVEAN predicted 12 were as damaging and remaining 6 were predicted as neutral. I-MUTANT 2.0, 18 mutations were affecting the stability of protein structure. By combing the all four tools score 10 (**R775C, Y764C, M788V**, **R775H, R608Q, H875Y, L708R, C580F, P549S and L713F)** SNPs were predicted as damaging and deleterious and carried these SNPs for further screening analysis.

S.No	RS.Ids	A.A change	SIFT		POLYPHEN-2		PROVEAN		I MUTANT2.0	
			Prediction	Score	Prediction	Score	Prediction	Score	Prediction	Score
1	Rs104894742	E2K	Deleterious	0.006	Pro.damg	0.996	Neutral	-1.351	Decrease	-0.68
2	Rs111468555	E642K	Deleterious	0.006	Pro.damg	0.996	Neutral	-2.284	Decrease	-0.70
3	**Rs137852562**	**R775C**	**Deleterious**	**0.00**	**Pro.damg**	**1.000**	**Deleterious**	**-6.150**	**Decrease**	**-1.94**
4	Rs137852564	V867M	Tolerated	0.401	Pro.damg	0.995	Neutral	-0.469	Decrease	-0.64
5	**Rs137852567**	**Y764C**	**Deleterious**	**0.009**	**Pro.damg**	**1.000**	**Deleterious**	**-6.185**	**Decrease**	**-0.78**
6	Rs137852569	A597T	Deleterious	0.001	Benign	0.295	Deleterious	-3.003	Decrease	-0.39
7	**Rs137852570**	**M788V**	**Deleterious**	**0.003**	**Poss.damg**	**0.653**	**Deleterious**	**-3.021**	**Decrease**	**-0.21**
8	**Rs137852572**	**R775H**	**Deleterious**	**0.001**	**Pro.damg**	**1.000**	**Deleterious**	**-3.827**	**Decrease**	**-2.23**
9	**Rs137852573**	**R608Q**	**Deleterious**	**0.002**	**Pro.damg**	**1.000**	**Deleterious**	**-3.092**	**Decrease**	**-0.53**
10	Rs137852574	I870M	Deleterious	0.008	Pro.damg	0.991	Neutral	-1.408	Decrease	-0.78

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 7, 2021, Pages. 1688 - 1695
Received 05 May 2021; Accepted 01 June 2021.

11	Rs137852580	T878S	Tolerated	0.169	Pro.damg	0.987	Deleterious	-2.576	Decrease	-1.62
12	**Rs137852581**	**H875Y**	**Deleterious**	**0.008**	**Pro.damg**	**0.977**	**Deleterious**	**-4.098**	**Decrease**	**-0.51**
13	Rs137852582	Q903R	Tolerated	0.056	Pro.damg	1.000	Neutral	-1.730	Decrease	-1.52
14	Rs137852583	A722T	Deleterious	0.011	Pro.damg	0.997	Neutral	-2.009	Decrease	-2.46
15	Rs137852584	S648N	Tolerated	0.347	Benign	0.001	Neutral	-0.881	Decrease	-1.01
16	**Rs137852585**	**L708R**	**Deleterious**	**0.00**	**Pro.damg**	**1.000**	**Deleterious**	**-4.714**	**Decrease**	**-0.23**
17	**Rs137852586**	**C580F**	**Deleterious**	**0.00**	**Pro.damg**	**1.000**	**Deleterious**	**-8.246**	**Decrease**	**-0.08**
18	Rs137852587	F583Y	Deleterious	0.00	Pro.damg	0.998	Neutral	-2.282	Increase	0.03
19	**Rs137852588**	**P549S**	**Deleterious**	**0.008**	**Pro.damg**	**0.999**	**Deleterious**	**-5.391**	**Increase**	**-0.57**
20	**Rs137852595**	**L713F**	**Deleterious**	**0.002**	**Pro.damg**	**1.000**	**Deleterious**	**-3.076**	**Decrease**	**0.18**

Screened Damaging and deleterious SNPs by various computational tools

Pro.damg- probably damaging; Poss.damg- possible damaging;

Prediction of Disease related SNPs:

PANTHER, PMUT, PHD-SNP and SNP&GO were performed to validate the results obtained from four tools. PANTHER predicted 20 nsSNPs to be associated with damaging (Table-2) PMUT predicted 17 nsSNPs, PHD-SNP and SNP&GO both were predicted 10 disease associated nsSNPs. Finally out of 20 nsSNPs. We found five nsSNPs namely **R775C (rs137852562), H875Y (rs137852581), Y764C (rs137852567), L708R (rs137852585) and C580F (rs237852586)** are predicted as damaging, deleterious and disease associated by combing score of all the computational tools.

Disease associated SNPs predicted by four different computational tools

S.NO	RS Ids	A.A change	PANTHER	PMUT		SNP&GO	PHD-SNP
				Prediction	Score		
1	Rs104894742	E2K	Damaging	Disease	0.51	Neutral	Neutral
2	Rs111468555	E642K	Damaging	Disease	0.65	Neutral	Neutral
3	**Rs137852562**	**R775C**	**damaging**	**Disease**	**0.84**	**Disease**	**Disease**
4	Rs137852564	V867M	damaging	Disease	0.84	Neutral	Neutral
5	**Rs137852567**	**Y764C**	**damaging**	**Disease**	**0.81**	**Disease**	**Disease**
6	Rs137852569	A597T	damaging	Disease	0.77	Disease	Neutral
7	Rs137852570	M788V	damaging	Disease	0.79	Neutral	Disease
8	Rs137852572	R755H	damaging	Disease	0.88	Neutral	Neutral
9	Rs137852573	R608Q	damaging	Disease	0.89	Disease	Neutral
10	Rs137852574	I870M	damaging	Disease	0.68	Neutral	Neutral
11	Rs137852580	T878S	damaging	Disease	0.79	Neutral	Neutral
12	**Rs137852581**	**H875Y**	**damaging**	**Disease**	**0.81**	**Disease**	**Disease**
13	**Rs137852582**	**Q903R**	**damaging**	**Disease**	**0.58**	**Disease**	**Disease**
14	Rs137852583	A722T	damaging	Disease	0.85	Neutral	Neutral
15	Rs137852584	S648N	damaging	Neutral	0.99	Neutral	Neutral
16	**Rs137852585**	**L708R**	**damaging**	**Disease**	**0.86**	**Disease**	**Disease**
17	**Rs137852586**	**C508F**	**damaging**	**Disease**	**0.90**	**Disease**	**Disease**

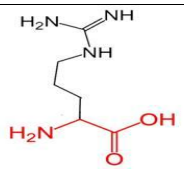
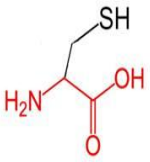
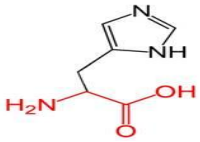
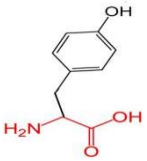
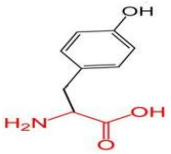
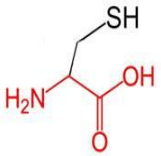
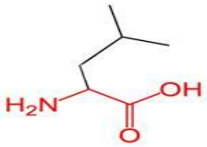
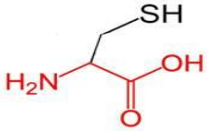
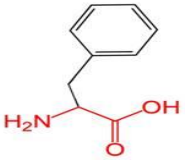
Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 7, 2021, Pages. 1688 - 1695
Received 05 May 2021; Accepted 01 June 2021.

18	**Rs137852587**	**F583Y**	**damaging**	**Disease**	**0.78**	**Disease**	**Disease**
19	Rs137852588	P549S	damaging	Neutral	0.21	Neutral	Disease
20	Rs137852595	L713F	damaging	Neutral	0.34	Disease	Disease

HOPE: The below tabular form shows the schematic representation of the original (left) and the mutant (right) amino acid. It shows an unexpected conclusion with amino acid properties in the mutations R775C, H875C, Y764C, L708R and C580F.The backbone, which is same for each amino acid, is coloured red. The side chain, unique for each amino acid, is coloured black. Each amino acid has its own specific size, and hydrophobicity value. The original wild-type residue and newly introduced mutant residue often differ in these properties.

- The mutant residue is smaller than the wild type residue.
- The wild type residue charge was POSITIVE; the mutant residue charge is NEUTRAL.
- The mutant residue is more hydrophobic than the wild-type residue.

Amino acid variations between wild and mutant protein at concern positions

RS IDs	Amino acid change	Wild type	Mutant type
Rs137852562	R775C		
Rs137852581	H875Y		
RS137852567	Y764C		
Rs137852585	L708R		
Rs137852586	C580F		

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 7, 2021, Pages. 1688 - 1695
Received 05 May 2021; Accepted 01 June 2021.

CONCLUSION:

Nowadays using *In-silico* tools is becoming an important approach for screening of disease related SNPs. In this study an extensive analysis of AR gene was carried out using different computational tools aiming to investigate the effect of nsSNPs on structure and function of the protein. Our study identified five pathogenic SNPs namely Rs137852562 (R775C), Rs137852581 (H875Y), RS137852567 (Y764C), Rs137852585 (L708R), Rs137852586 (C580F) in AR gene as disease associated and also revealed their probable malfunctioning mechanism via their structural destabilization. Compared to the wild type, all the selected mutations were altering the structural behaviour of the protein. Although using computational tools to investigate the effects of the SNPs may help in determining disease related SNPs, but nevertheless population genetics and clinical studies are important to confirm the outcomes of such study.

REFERENCES:

1. Knochenhauer, E. S., Key, T. J., Kahsar-Miller, M., Waggoner, W., Boots, L. R., & Azziz, R. (1998). Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States: a prospective study. *The Journal of Clinical Endocrinology & Metabolism*, *83*(9), 3078-3082.
2. Goodarzi, M. O., Dumesic, D. A., Chazenbalk, G., & Azziz, R. (2011). Polycystic ovary syndrome: etiology, pathogenesis and diagnosis. *Nature reviews endocrinology*, *7*(4), 219-231.
3. Azziz, R., Carmina, E., Dewailly, D., Diamanti-Kandarakis, E., Escobar-Morreale, H. F., Futterweit, W & Witchel, S. F. (2009). The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. *Fertility and sterility*, *91*(2), 456-488.
4. Tan, M. E., Li, J., Xu, H. E., Melcher, K., & Yong, E. L. (2015). Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacologica Sinica*, *36*(1), 3-23.
5. Gottlieb, B., Pinsky, L., Beitel, L. K., & Trifiro, M. (1999). Androgen insensitivity. *American journal of medical genetics*, *89*(4), 210-217.
6. Claessens, F., Denayer, S., Van Tilborgh, N., Kerkhofs, S., Helsen, C., & Haelens, A. (2008). Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. *Nuclear receptor signaling*, *6*(1), nrs-06008.
7. Adams, J., Polson, D. W., & Franks, S. (1986). Prevalence of polycystic ovaries in women with anovulation and idiopathic hirsutism. *Br Med J (Clin Res Ed)*, *293*(6543), 355-359.
8. Yuan, W., & Giudice, L. C. (1997). Programmed cell death in human ovary is a function of follicle and corpus luteum status. *The Journal of Clinical Endocrinology & Metabolism*, *82*(9), 3148-3155.
9. Dunaif, A., Scott, D., Finegood, D. I. A. N. E., Quintana, B., & Whitcomb, R. A. N. D. A. L. L. (1996). The insulin-sensitizing agent troglitazone improves metabolic and reproductive abnormalities in the polycystic ovary syndrome. *The Journal of Clinical Endocrinology & Metabolism*, *81*(9), 3299-3306.
10. Griffith, O. L., Montgomery, S. B., Bernier, B., Chu, B., Kasaian, K., Aerts, S & Open Regulatory Annotation Consortium. (2007). ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic acids research*, *36*(suppl_1), D107-D113.

11. Singh, P. K., & Mistry, K. N. (2016). A computational approach to determine susceptibility to cancer by evaluating the deleterious effect of nsSNP in XRCC1 gene on binding interaction of XRCC1 protein with ligase III. *Gene*, *576*(1), 141-149.
12. Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research*, *31*(13), 3812-3814.
13. Min, L., Nie, M., Zhang, A., Wen, J., Noel, S. D., Lee, V., ... & Kaiser, U. B. (2016). Computational analysis of missense variants of G protein-coupled receptors involved in the neuroendocrine regulation of reproduction. *Neuroendocrinology*, *103*(3-4), 230-239.
14. Min, L., Nie, M., Zhang, A., Wen, J., Noel, S. D., Lee, V., ... & Kaiser, U. B. (2016). Computational analysis of missense variants of G protein-coupled receptors involved in the neuroendocrine regulation of reproduction. *Neuroendocrinology*, *103*(3-4), 230-239.
15. Badgujar, N. V., Tarapara, B. V., & Shah, F. D. (2019). Computational analysis of high-risk SNPs in human CHK2 gene responsible for hereditary breast cancer: A functional and structural impact. *Plos one*, *14*(8), e0220711.
16. Badgujar, N. V., Tarapara, B. V., & Shah, F. D. (2019). Computational analysis of high-risk SNPs in human CHK2 gene responsible for hereditary breast cancer: A functional and structural impact. *Plos one*, *14*(8), e0220711.
17. Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., ... & Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome research*, *13*(9), 2129-2141.
18. Smith, P. J., Zhang, C., Wang, J., Chew, S. L., Zhang, M. Q., & Krainer, A. R. (2006). An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Human molecular genetics*, *15*(16), 2490-2508.
19. Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P. L., Altman, R. B., &Casadio, R. (2013). WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC genomics*, *14 Suppl 3*(Suppl 3), S6.
20. Capriotti, E., Calabrese, R., & Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*, *22*(22), 2729-2734.
21. Venselaar, H., Te Beek, T. A., Kuipers, R. K., Hekkelman, M. L., & Vriend, G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC bioinformatics*, *11*(1), 1-10.