



Received on 08 April, 2018; received in revised form, 11 July, 2018; accepted, 18 July, 2018; published 01 December, 2018

IN-SILICO ANALYSIS OF SNPs FROM cAMP-GEFII GENE ASSOCIATED WITH POLYCYSTIC OVARIAN SYNDROME

Jeyabaskar Suganya, Mahendran Radha^{*}, Sharanya Manoharan and Vasudevan Poornima

Department of Bioinformatics, School of Life Sciences, VISTAS, Pallavaram, Chennai - 600117, Tamil Nadu, India.

Keywords:

Polycystic ovary syndrome (PCOS), Candidate gene, Protein, SNP, SNP analysis tool

Correspondence to Author:

Mahendran Radha

Department of Bioinformatics,
School of Life Sciences, VISTAS,
Pallavaram, Chennai - 600117,
Tamil Nadu, India.

E-mail: suganyaj11@gmail.com

ABSTRACT: Polycystic ovary syndrome (PCOS) is one of the most common disorders that occur in women at any age due to the endocrine hormone imbalance. The cause for this disorder is still not identified, but on recent research provided that disorder may be caused by some genetic variation. Predicting and understanding the effects of genetic variation that occurred in the gene are becoming more important for single nucleotide polymorphism to understand the molecular basis of genetic disease. From the literature survey, the candidate gene which is responsible for causing genetic PCOS was identified. In this work using computational methods, this candidate gene was analyzed completely to find out the genetic variation which in charge for altering the expression and the functional of the gene. On analyzing the gene, it was predicted that the protein which was translated from the gene played a key role for causing the major alteration in the gene. Using SNP analysis tool, further investigation were carried out to the disease causing protein and were predicted that the particular mutation occurred in the protein altered the function and structure of the gene. By using bioinformatics tool, an attempt was made to stop the mutation by replacing the original amino acid to the structure and sequence of the protein, which was suggested by the tools. Some clinical studies can be carried out further to confirm that the protein which was responsible for gene alteration in PCOS will function normally after some necessary modification are made in the protein which was suggested by computational methods.

INTRODUCTION: Polycystic ovary syndrome (PCOS) is a mysterious reproductive disorder that causes irregular menstrual periods¹. An average of 5 to 10 percent of women was found to have this disorder in their endocrine glands². PCOS women's of overweight or obese would have higher risk of developing Type 2 diabetes³ and infertility than the normal women⁴.

The mode of inheritance and role of genetic factors in PCOS syndrome have not been fully investigated until now⁵. The genes responsible for altered expression in PCOS causes genetic abnormality which affects signal transduction ruling steroid genesis, steroid hormones action, gonadotropin action and regulation, insulin action and secretion, energy homeostasis, chronic inflammation and others^{6,7}.

Several large research groups are actively searching for a genetic cause of this syndrome and noted out the increased level of insulin resistance in PCOS affected woman⁸. A study was conducted to find out the main candidate gene which causes PCOS to women by collecting 50 candidate genes which are

<p>QUICK RESPONSE CODE</p> 	<p style="text-align: center;">DOI: 10.13040/IJPSR.0975-8232.9(12).5216-20</p> <hr/> <p style="text-align: center;">Article can be accessed online on: www.ijpsr.com</p>
<p>DOI link: http://dx.doi.org/10.13040/IJPSR.0975-8232.9(12).5216-20</p>	

linkage and association with PCOS. That study resulted the cAMP-GEFII were identified as new candidate genes that may be involved in the genetic etiology of PCOS⁹. The genes cAMP-GEFII, which is an intracellular signaling molecule that is activated by cAMP¹⁰ and it was the direct target of cAMP in regulated exocytosis¹¹.

Later genetic causes of this syndrome were identified that if Single nucleotide polymorphisms *i.e.* mutation occurs in the candidate gene would result in PCOS^{12, 13}. Single nucleotide polymorphisms are single base (A, T, G, C) variations between genomes within a species¹⁴. SNPs can also be used to track the inheritance of disease genes within families¹⁵. When SNPs occur within a gene or in a regulatory region, it affects the stability and functions of the protein¹⁶ due to deleterious amino acid changes occur in the protein sequences would play a very vital role in causing the disease by affecting the candidate gene's functions¹⁷. SNP analysis would help to identify the multiple genes associated with such complex diseases like cancer, diabetes, vascular disease and mental illness^{18, 19}. In this study, the candidate gene of PCOS was analysed using Bioinformatics and step were taken to stop the mutation occurred in the gene which result in normal gene function.

MATERIALS AND METHODS:

Sequence Analysis: The Gene ID, nucleotide sequences and protein sequences of Human cAMP-GEFII gene were retrieved using NCBI and SWISSPROT Database.

The physical and chemical properties of protein sequence can be analyzed using ProtParam. The properties include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).

The domain and family of protein were identified using Pfam to understand the function of protein.

SNPs Analysis: The list of SNP'S (rs ID) present in the gene was predicted by submitting gene ID using NCBI- dbSNP. The dbSNP database shows the disease-causing clinical mutations as well as neutral polymorphisms which is present in the gene.

The Gene ID of cAMP-GEFII were submitted to database and list of SNPs which observed in the gene were predicted (<http://www.ncbi.nlm.nih.gov/SNP/>).NCBI – db SNP database.

The SNP IDs of human cAMP-GEFII were submitted to identify where amino acid substitution occurs in gene and which affect the protein function. (http://siftdna.org/www/SIFT_dbSNP.html).

Sorting Intolerant from Tolerant (SIFT dbSNP) Predicted protein sequence and SNPs of human cAMP-GEFII were submitted to evaluate the effect of single amino acid substitution on protein function (<https://www.rostlab.org/services/snap/submit>).²⁰

Screening for Non-Acceptable Polymorphisms (SNAP):

The FASTA format of protein sequence and amino acid substitution were submitted to predict whether a nonsynonymous single nucleotide polymorphism (nsSNP) having a phenotypic effect or not. (<http://snpanalyzer.uthsc.edu/>).

NsSNP Analyzer: The FASTA format of protein sequence and amino acid substitution were submitted to predict the particular nonsynonymous (amino acid changing) SNP which cause a functional impact on the protein. (<http://www.pantherdb.org/tools/csnpscoreForm.jsp>).

PANTHER: Evolutionary analysis of coding SNPs The Gene ID of cAMP-GEFII was submitted to assign molecular functional effects of non-synonymous SNPs based on structure and sequence analysis (<http://www.snps3d.org/>).

SNPs3D: Pymol is a potent and interactive molecular visualization software tool and used to view the protein, ligand²¹. The three dimensional structure of protein and ligand were visualized using Pymol and saved in PDB file format for further analysis²².

Swiss pdb viewer: Swiss pdb viewer is an interactive molecular graphics program for visualization and analysis of protein structures. Using Mutation²³ and Energy minimization tool²⁴ for the protein, mutation has been altered and

Energy minimization was performed before and after mutation.

RESULTS AND DISCUSSION:

Sequence Analysis: The FASTA file format of the Nucleotide sequence >gi|568815596:172687774-173100516 Homo sapiens chromosome 2, GRCh38 Primary Assembly and protein sequence. >gi|155030204|ref|NP_008954.2| rap guanine nucleotide exchange factor 4 isoform a [Homo sapiens]. The Nucleotide and protein sequence was

retrieved from the NCBI & SWISSPROT database which contains 412225 base-pairs in the mRNA and 1011 amino acid residues

The physiochemical properties of the protein - molecular weight (115521.5), theoretical pI (6.37) and aliphatic index (88.80), Grand average of hydropathicity – GRAVY (0.334), molecular formula (C₅₁₄₀H₈₁₃₁N₁₄₁₉O₁₅₁₁S₄₈) and amino acid composition are shown in the **Table 1** were predicted by protparam.

TABLE 1: SHOWS THE AMINO ACID COMPOSITION OF cAMP-GEFII PROTEIN

Amino acid	In number	In %	Amino acid	In number	In %	Amino acid	In number	In %
Ala (A)	6.6	6.5%	Leu (L)	106	10.5%	Gly (G)	46	4.5%
Arg (R)	58	5.7%	Lys (K)	64	6.3%	His (H)	35	3.5%
Asn (N)	44	4.4%	Met (M)	30	3.0%	Ile (I)	53	5.2%
Asp (D)	56	5.5%	Phe (F)	42	4.2%	Trp (W)	10	1.0%
Cys (C)	18	1.8%	Pro (P)	44	4.4%	Tyr (Y)	29	2.9%
Gln (Q)	46	4.5%	Ser (S)	55	5.4%	Val (V)	73	7.2%
Glu (E)	77	7.6%	Thr (T)	59	5.8%	Sec (U)	0	0.0%

The five (3 different & 2 same) domain and 1 family of protein were identified and found that all the 5 domain were significant. In the domain, 2

clan were found (CL0542 & CL0072) were predicted using Pfam are shown in the **Table 2**.

TABLE 2: SHOWS THE FAMILY OF cAMP-GEFII PROTEIN

Family	Description	Entry Type	CLAN
cNMP binding	Cyclic nucleotide-binding domain	Domain	-
DEP	Domain found in Dishevelled	Domain	-
cNMP binding	Cyclic nucleotide-binding domain	Domain	-
RasGEF N	RasGEF N-terminal motif	Domain	CL0542
RA	Ras association domain	Domain	CL0072
RasGEF	RasGEF domain	Family	-

CL0542: Name of the clan - Ras guanyl-nucleotide exchange factor activity N-term. This is the more N-terminal domain of the RAS-GEF superfamily. This clan contains 2 families and the total number of domains in the clan is 2682.

CL0072: Name of the clan - Ubiquitin superfamily. This family includes proteins that share the ubiquitin fold. This clan contains 41 families and the total number of domains in the clan is 107825.

SNPs Analysis: From NCBI-Gene database the cAMP-GEFII gene ID 11069 were identified and then the gene ID were submitted to dbSNP database which predicted that there were 193 SNPs ID (rsID) for this gene, in that some of them are nonsynonymous (ns) and synonymous, which occur

in intron region and untranslated region. The non-synonymous (80) SNPs were taken for further analysis because the presence of non-synonymous SNPs can only affect the protein function and structure^{25, 26}. The list of non-synonymous SNPs was as shown in the **Table 3**.

TABLE 3: SHOWS THE LIST OF NON-SYNONYMOUS SNPs

rs375512263	rs200210661	rs79167858	rs369674227	rs18363265	rs376787687
rs57766910	rs146735016	rs201933844	rs182318587	rs200268023	rs369838179
rs373393896	rs141738165	rs376497880	rs370988008	rs374468976	rs376255045
rs199696709	rs202063054	rs200306531	rs375404385	rs146330323	rs17857212
rs376257745	rs369456567	rs376593491	rs139102560	rs368287979	rs369571635
rs34347501	rs2166589	rs201545225	rs368794857	rs373777968	rs17857211
rs112722051	rs372092370	rs185499414	rs374929947	rs370697748	rs372769775
rs201425570	rs183096630	rs199578267	rs200217066	rs372171158	rs150495482
rs190538163	rs17853965	rs368919451	rs373366834	rs376379807	rs182317769
rs370077088	rs17852173	rs368051118	rs201085724	rs368206145	rs372827002
rs61741755	rs200419099	rs17852174	rs181938855	rs368290386	rs369850284
rs368653870	rs199973105	rs373120144	rs201741962	rs61756296	rs372917537
rs369058413	rs201446102	rs201906810	rs200990002	rs373438925	rs17853967
	rs36815776			rs370128565	

Using Sift dbSNP database, nsSNP was analyzed to spot rsID (rs17852173) which was responsible for damaging the function and the structure of cAMP-GEFII were shown in the **Table 4**.

TABLE 4: SHOWS THE PREDICTION OF SNPs WHICH CAUSES DAMAGE TO cAMP-GEFII PROTEIN

SNP IDs	Amino acid change	Predicted
rs182318587	G54E (G)	DAMAGE
rs17857212	P171R (P)	DAMAGE
rs2166589	S240F (S)	DAMAGE
rs17857211	H358R (H)	DAMAGE
rs17852173	H679P (H)	DAMAGE
rs17853967	P749 (P)	DAMAGE

Predicted nsSNP: The SIFT database predicted 6 nsSNP (G54E, P171R, S240F, H358R, H679P,

P749) which causes damage to the gene. From SNAP, nsAnalyzer, Panther Database the exact SNP which affect the function of protein and also for the possibility of structural alteration were predicted. In SNAP database SNP of H679P were found to be Non-neutral with high Reliability index 5 when compared with other SNP which are neutral. From nsAnalyzer Database Phenotype of given SNP were analyzed H679P SNP were found to cause disease with siftscore of 0. Result predicted from Panther database show that SNP H679P does not show any multiple sequences position (MSA) when compare with other SNP were shown in the **Table 5**.

TABLE 5: SHOWS THE RESULT OF SNPs ANALYSIS OF cAMP-GEFII PROTEIN

SNP Variation	SNAP Prediction	nsSNP analyzer	PANTHER (MSA position)	SNP3D (Mol effect)
G54E	Neutral	Neutral	Found	No lose
P171R	Neutral	Neutral	Found	No lose
S240F	Neutral	Neutral	Found	No lose
H358R	Neutral	Neutral	Found	No lose
H679P	Non - Neutral	Disease	Not Found	Loses H2 bond
P749	Neutral	Neutral	Found	No lose
G54E	Neutral	Neutral	Found	No lose

From 3D SNP database, the 3D structure of cAMP-GEFII Protein was retrieved and viewed using PyMol. On analyzing the sequence and structure of the protein it was confirmed that the amino acid proline was present in the 679th position of amino acid with the loss of Hydrogen Bond in the structure resulting in the alteration of the gene function. By Swiss Pdb Viewer, the exact amino acid Histidine was placed in proline in the 679th position of protein structure shown in the **Fig. 1**, the altered 3D structure of the protein is saved in PDB file format. On analyzing the both structure of the protein, the amino acid histidine present in the 679th position showed the proper arrangement of atoms and thereby regulating the protein function when compared with the amino acid proline in the 679th position of protein structure.

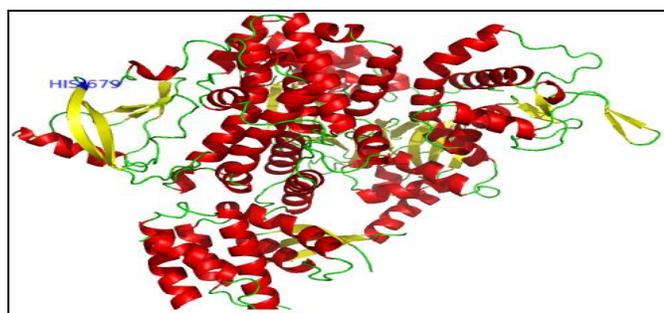


FIG. 1: 3D STRUCTURE OF PROTEIN OF cAMPGEFII WITH HISTIDINE IN 679th POSITION

CONCLUSION: Mutation occurs in the gene affects both the function and structure of the protein which lead to a disease in human. In the current study, the gene which was responsible for causing reproductive disorder of PCOS was identified. On further analyses of cAMP-GEFII gene using bioinformatics tools, it was predicted that mutation (SNP) present in the gene was in charge in causing disease.

From the above obtained result, we can conclude that non-synonymous SNP of rsID (rs17852173) in the cAMP-GEFII gene was not only affects the protein function and structure but also cause serious damage to gene. The investigation of rsID was carried out to identify the exact location of SNP occurred in the protein sequence and structure, resulted that H679P SNP presented in the protein played a vital role for the occurrence of PCOS in many women. The mutations (SNP) in the protein structure were dysfunction by replacing the amino acid Histidine (H) instead of proline (P) in 679th position of protein structure to make the protein function normally by Swiss PDB Viewer. Some clinical trials are needed in future to confirm that this disease can be treated well by altering the mutation in the gene.

ACKNOWLEDGEMENT: We acknowledge Vels Institute of Science, Technology and Advanced Studies (VISTAS) for providing us with required infrastructure and support system needed.

CONFLICT OF INTEREST: The authors declared no conflicts of interest.

REFERENCES:

- Azziz R, Woods KS, Reyna R, Key TJ, Knochenhauer ES, and Yildiz BO: The Prevalence and Features of the Polycystic Ovary Syndrome in an Unselected Population. *J Clin Endocrinol Metab.* 2004; 89(6): 2745-9.
- Wild RA, Painter PC, Coulson PB, Carruth KB and Ranney GB: Lipoprotein Lipid Concentrations and Cardiovascular Risk in Women with Polycystic Ovary Syndrome. *J Clin Endocrinol Metab.* 1985; 61(5): 946-51.
- Knochenhauer ES, Key TJ, Kahsar-Miller M, Waggoner W, Boots LR and Azziz R: Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States: a prospective study. *J Clin Endocrinol Metab.* 1998; 83: 3078-3082.
- Richard S Legro, Kunselman AR, Dodson WC and Dunaif A: Prevalence and predictors of risk for type 2 diabetes and impaired glucose tolerance in polycystic ovary syndrome: a prospective, controlled study in 254 affected women. *J Clin Endocrinol Metab.* 1999; 84: 165-169.
- Legro RS and Strauss JF: Molecular progress in infertility: polycystic ovary syndrome. *Fertil Steril* 2002; 78(3): 569-76.
- Prapas N, Karkanaki A, Prapas, Kalogiannidis, Katsikis, and Panidis D: Genetics of Polycystic Ovary Syndrome. *Hippokratia* 2009; 13(4): 216-223.
- Gharani N, Waterworth DM and Batty S: Association of the steroid synthesis gene CYP11a with polycystic ovary syndrome and hyperandrogenism. *Hum Mol Genet.* 1997; 6: 397-402
- Legro RS and Strauss JF: Molecular progress in infertility: polycystic ovary syndrome. *Fertil Steril.* 2002; 78: 569-576.
- Wood JR, Nelson VL, Ho C, Jansen E, Urbanek CYWM, McAllister JM, Mosselman S and Strauss JF: The Molecular Phenotype of Polycystic Ovary Syndrome (PCOS) Theca Cells and New Candidate PCOS Genes Defined by Microarray Analysis. *J Biol Chem.* 2003; 278(29): 26380-90.
- Kawasaki H, Springett GM, Mochizuki N, Toki S, Nakaya M, Matsuda M, Housman DE and Graybiel AM: A family of cAMP-binding proteins that directly activate Rap1. *Science* 1998; 282: 2275-2279.
- Ozaki N, Shibasaki T, Kashima Y, Miki T, Takahashi K, Ueno H, Sunaga Y, Yano H, Matsuura Y, Iwanaga T, Takai Y and Seino S: cAMP-GEFII is a direct target of cAMP in regulated exocytosis. *Nat Cell Biol.* 2000; 2(11): 805-11.
- Camp Kang G, Chepurny OG and Holz GG: cAMP-regulated guanine nucleotide exchange factor II (Epac2) mediates Ca²⁺-induced Ca²⁺ release in INS-1 pancreatic beta-cells. *J Physiol* 2001; 536: 375-385
- Carey AH, Waterworth D and Patel K: Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. *Hum Mol Genet* 1994; 3: 1873-1876.
- Chakravarti A: Single nucleotide polymorphisms: To a future of genetic medicine. *Nature* 2001; 409: 822-823.
- Teng S, Alexova ME and Alexov E: Approaches and Resources for Prediction of the Effects of Non-Synonymous Single Nucleotide Polymorphism on Protein Function and Interactions 2008; 9 (2): 123-133.
- Pauline CN and Henikoff S: Predicting Deleterious Amino Acid Substitutions. *Genome Res* 2001; 11(5): 863-74.
- Panneerselvam P, Sivakumari K, Jayaprakash P and Srikanth R: SNP analysis of follistatin gene associated with polycystic ovarian syndrome. *Adv Appl Bioinforma Chem* 2010; 3: 111-119.
- Carey AH, Chan KL, Short F, White D, Williamson R and Franks S: Evidence for a single gene effect causing polycystic ovaries and male pattern baldness. *Clin Endocrinol.* 1993; 38: 653-658.
- Legro RS, Driscoll D, Strauss JF, Fox J and Dunaif A: Evidence for a genetic basis for hyperandrogenemia in the polycystic ovary syndrome. *Proc Natl Acad Sci.* 1998; 95: 14956-14960.
- Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, Donnell CJ, Paul IW and de Bakker: SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008; 24(24): 2938-9.
- Itoh N and Ohta H: Roles of FGF20 in dopaminergic neurons and Parkinson's disease. *Front Mol Neurosci.* 2013; 6: 15.
- Volkow ND, Fowler JS, Logan J, Alexoff D, Zhu W, Telang F, Wang GJ, Jayne M, Hooker JM, Wong C, Hubbard B, Carter P, Warner D, King P, Shea C, Xu Y, Muench L and Apelskog-Torres L: Effects of Modafinil on Dopamine and Dopamine Transporters in the Male Human Brain: Clinical Implications *JAMA* 2009; 301(11): 1148-1154.
- Saggu H, Cooksey J, Dexter D, Wells FR, Lees A, Jenner P and Marsden CD: A selective increase in particulate superoxide dismutase activity in parkinsonian substantia nigra. *J Neurochem* 1989; 53(3): 692-7.
- Sharma RB and Chetia D: Docking studies on quinine analogs for plasmepsin-II of malaria parasite using bioinformatics tools. *Int J Pharm Sci,* 5(3): 68-685.
- Jordan DM, Ramensky VE and Sunyaev SR: Human allelic variation: perspective from protein function, structure and evolution. *Curr Opin Struct Biol.* 2010; 20(3): 342-350.
- Ramensky V, Bork P and Sunyaev S: Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* 2002; 30(17): 3894-3900.

How to cite this article:

Suganya J, Radha M, Manoharan S and Poornima V: *In-silico* analysis of SNPs from cAMP-GEFII gene associated with polycystic ovarian syndrome. *Int J Pharm Sci & Res* 2018; 9(12): 5216-20. doi: 10.13040/IJPSR.0975-8232.9(12).5216-20.

All © 2013 are reserved by International Journal of Pharmaceutical Sciences and Research. This Journal licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

This article can be downloaded to **ANDROID OS** based mobile. Scan QR Code using Code/Bar Scanner from your mobile. (Scanners are available on Google Playstore)