Indian Journal of Biochemistry & Biophysics
Vol. 57, October 2020, pp. 584-593

In silico analysis of functional non-synonymous and intronic variants found in a polycystic ovarian syndrome (PCOS) candidate gene: *DENND1A*

Dakshina Moorthy Janani & Balasundaram Usha*

Department of Genetic Engineering, School of Bioengineering, SRM Institute of Science and Technology, Kattankulathur-603 203, Tamil Nadu, India

Received 04 March 2020; *revised 31 July 2020*

The major thrust of our study confers to the identification of non-synonymous and intronic variants of the *DENND1A* gene *via in silico* methods and determination of its effect on the structural integrity of the protein. The outcome identifies potential disease -causing SNPs. The pathogenic variants of *DENND1A* were deduced *via in silico* analysis using various tools that include SIFT, PolyPhen-2, PROVEAN, SNP & GO, and PANTHER. The intronic variants were analysed using RegulomeDB. The 3D protein structure was obtained using the SWISS PDB modeler and validated by Ramachandran plots and QMEAN server. The effect on the stability of the protein structure caused by the SNPs was evaluated on the PYMOL and SWISS model platform. The functional changes caused by the SNPs were analysed *in silico* with I Mutant and Mutation Taster. The post-translational modifications were also predicted. STRING database was used for screening the protein interaction network. The SNPs rs2479106 and rs10986105 on the splice sites were found to be pathogenic for PCOS. The amino acid changes V179G and P331L were found to be disease-causing but the disease association with PCOS is yet to be validated.

Keywords: *DENND1A*, *In silico* analysis, Polycystic ovarian syndrome (PCOS), Protein structure, Protein stability, SNPs

A major purpose of human genetics is to comprehend the role of major genetic variants pathogenic to common diseases. This involves characterizing the gene variation found in human populations, arranging the extensive list of single-nucleotide polymorphisms [SNPs] of candidate genes together, and carrying out association studies for particular diseases. Before proceeding to any variant analysis in a population it is essential to characterize the variants using *in silico* tools.

The tools provide information of the SNPs that alter the amino acid sequence of a gene product [non-synonymous SNPs] and to discover the relationship between genetic and phenotypic variation; it is fundamental to determine the structural impact of the respective non-synonymous mutations on protein structure. Non – synonymous SNPs [nsSNPs] are sequence variants present as point mutations that lead to an alteration in the sequence of the encoded protein[1]. Usually, these variants are neutral, with little or no evident consequence on protein function, but some nsSNPs are known to be deleterious *i.e.* they are associated with either Mendelian or complex genetic diseases. Most of the SNPs that affect biological function occur outside the coding regions of genes namely the regulatory region and very few occur within the protein – coding sequence. Roughly there are around 13000 exonic SNPs per person, of which around 58% are non-synonymous[2]. These SNP effects upon gene-environment interaction develop various complex disorders in humans who are predisposed.

The Poly Cystic Ovarian Syndrome (PCOS) is one such complex disorder with a gynecological endocrinopathy and affects 5-12% of women in their reproductive age and is associated with various metabolic and reproductive complications[3]. This syndrome is characterized by increased androgen secretion, anovulation, and polycystic ovary morphology. The long-term health risks may include infertility[4], endometrial cancer and cardiovascular disorders. It is likely to be polygenic or oligogenic in origin[5,6].

GWAS (Genome-Wide Association Studies) studies for PCOS by Chen *et al.* and its replication studies in European cohorts have reported the *DENND1A* [differentially expressed in normal and neoplastic cells containing domain 1A] gene as a potential candidate marker[7,8]. It is being considered

*Correspondence:
Phone: 044-27452270
E-mail: sundaram.usha@gmail.com

that overexpression of *DENND1A* can increase the possibility of a woman's susceptibility to PCOS. More precisely, it is found that polymorphisms in the *DENND1A*.variant 2 [*DENND1A*.V2] lead to an increased expression of PCOS ovarian theca cells and adrenal zona reticularis, both of which are responsible for androgen secretion[9].

Also, recently, *DENND1A* has been under the spotlight for finding the association between its mutants and disorders. *DENND1A*, a member of the connecdenn family, and it is found to show its major functions as guanine nucleotide exchange factor [GEFs] for the early endosomal small GTPase RAB35 and also binds to the clathrin adaptor protein-2. This gene which consists of the clathrin domain and is considered to facilitate endocytosis and receptor-mediated turnover[10].

DENND1A is found on the chromosomal region 9q33.2 and consists of 22 exons that extend over 500000 bases and encodes the protein DENN domain-containing 1A [*DENND1A*], which contains an N-terminal DENN [differentially expressed in normal and neoplastic cells] domain, a common and evolutionarily ancient protein module. The primary structure of *DENND1A* protein is predicted to have DENN domains, a clathrin-binding motif, a proline-rich domain, and an adapter protein binding site at its C – terminal end (Fig. 1)[11,12]. Connecdenn1 is thought to be expressed primarily in brain and testis, although it can be detected in multiple additional tissues and cell lines[13]. A minimum of two splice variants is predicted from databases: Variant 1 and Variant 2[14].

We have tried to exclusively predict the effect of *DENND1A* variants on the functionality of the *DENND1A* protein and also deduce its role in PCOS pathophysiology *via in silico* analyses. Hence, the variants identified may give insight into the pathophysiology of PCOS thus helping the clinician identify a diagnosis module or treatment method after evaluating the risk ratio of the SNP in the population.

Materials and Methods

SNPs datasets

The Online Mendelian Inheritance in Man [OMIM] and Entrez Gene on National Centre for Biological Information [NCBI] websites were used to retrieve the specifics for the *DENND1A* gene. The SNP information such as rs ID, functional class, variation class, clinical significance, and its allele frequency was obtained from the database of NCBI for SNPs [dbSNP] found at [https://www.ncbi.nlm.nih.gov/snp]. Entrez Gene [http://grch37.ensembl.org/index.html] gave the gene's details on chromosome location and coordinates in base pairs, consequence type, and amino acid coordinates. The list of rs IDs for further downstream analysis was obtained from dbSNP and also NCBI's Variation Viewer [https://www.ncbi.nlm.nih.gov/variation/view/] since this table gives varied options for classification. The SNP list thus retrieved was combined and duplicates were removed for being used as input for the various *in silico* tools[15].

The SNPs identified by GWAS for PCOS were retrieved from PCOS knowledgebase [PCOSKB-http://pcoskb.bicnirrh.res.in/] and the chromosome coordinates were given as input to RegulomeDB and Mutation Taster.

The particulars on the *DENND1A* protein were obtained from UniProt Knowledgebase accessed from [http://www.uniprot.org/]. The mRNA accession number [NM_024820], the genomic DNA accession number [NC_000009.12] and the protein accession number [Q8TEH3] of the canonical structure which was used here for the native 3D structure and the FASTA sequence for mutation modelling was obtained from Protein Data Bank [PDB] [http://www.rcsb.org/pdb/home/home.do].

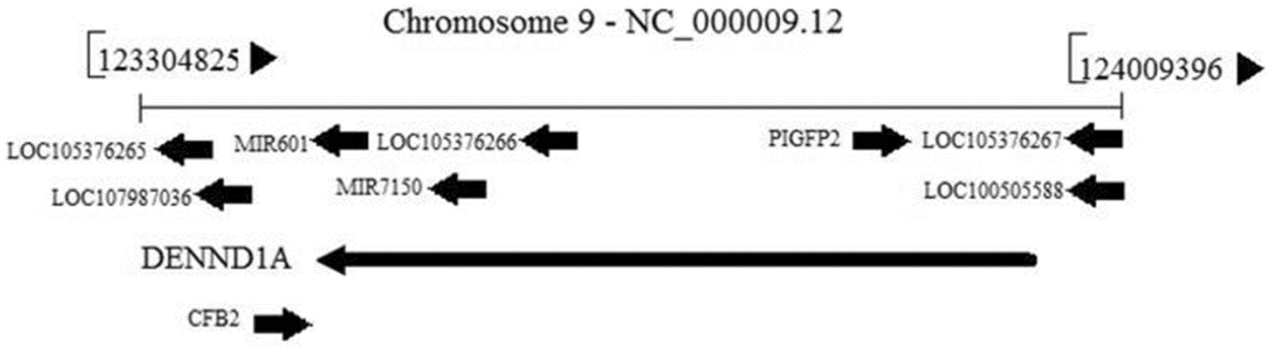

Fig. 1 — A diagrammatic representation of *DENND1A* gene in the chromosome 9_NC_000009.12 extracted from www.ncbi.nih.gov. Gene ID: 57706

Functional analysis prediction of non-synonymous SNPs

The tools SIFT, PolyPhen-2, PROVEAN, SNP & GO and PANTHER were used for characterizing the damaging effects of SNP. SIFT [http://sift.jcvi.org/] was used to identify the tolerance of an amino acid substitution on the protein's function. SIFT predicts the deleterious effect of an nsSNP by using algorithms that align closely related sequences using PSI-BLAST and predict the damage using the degree of conservation of sequences. PolyPhen-2 [http://genetics.bwh.harward.edu/pp2] abbreviated as Polymorphism Phenotyping version 2. PolyPhen helps to understand the probable impact of amino acid substitution on the structural and functional properties of the protein. It first characterizes using sequence-based methods followed by mapping the polymorphisms against the structure for generating scores. PROVEAN [http://provean.jcvi.org] tool enables predicting if an amino acid substitution has an effect on protein's biological function and further filtering sequence variants to find out non-synonymous variants. PROVEAN was used to gather pair-wise sequence alignment scores to generate pre-computed predictions. SNP & GO is an accurate method that takes in to account the protein sequence and predicts whether the polymorphisms are disease-related variations or not by exploiting the functional annotation of the corresponding protein [http://snps.biofold.org/snps-and-go/]. PANTHER [www.pantherdb.org/tools/csnpScoreForm.jsp] was used to predict the damaging effect of nsSNPs by calculating the substitution position-specific evolutionary conservation [sub-PSEC] score and thus to predict the amino acid substitution that will cause any functional effect using HMM[Hidden Markov Model][16].

Prediction of *DENND1A* protein structure

Protein BLAST[17] was used to find the proteins that are similar to *DENND1A*. Identification of highly-homologous sequences was done using a BLASTP search against every macromolecular structure deposited in protein data bank [http://www.rcsb.org/pdb/home]. SWISS-MODEL online protein modeling software was used to predict the structure of *DENND1A* protein since the 3D structure of *DENND1A* protein is not available in protein data bank[18]. We used SWISS-MODEL [http://swissmodel.expasy.org/] to generate a 3D structural model for *DENND1A* protein. SWISS-MODEL is software that uses homology modelling based structure prediction for automated protein structure modeling. Assigning the fold junctions, aligning the target according to the template sequences, building a model based on the alignment, and then validating the structure are the four stages involved in the modelling methodology[19].

Validation and evaluation of the predicted protein structure

Stereochemical characterization of *DENND1A* protein was performed using RAMPAGE. Ramachandran plot is a scattered two-dimensional plot of Φ and Ψ pairs comparing them to a predicted distribution. [http://modred.bioc.cam.ac.uk/~rapper/rampage.php] RAMPAGE Ramachandran plot uses database statistics for validation of modeled protein structure[20].

Evaluation of protein structure was carried out using the QMEAN server. The QMEAN server [http://swissmodel.expasy.org/qmean] provides access to both composite scoring function QMEAN and clustering method QMEAN clust. QMEAN score is a calculation of 6 various terms [1] C_β interaction energy [2] all-atom pairwise energy [3] solvation energy [4] torsional angle energy [5] solvent accessibility agreement and [6] total QMEAN-score.

The post-translational modification sites on *DENND1A* protein

Glycation sites of ε amino groups of lysine residues were predicted using a NetGlycate 1.0 server [http://www.cbs.dtu.dk/services/NetGlycate/]. In NetGlycate a score of >0.5 was considered to be a site for glycation. Phosphorylation sites were predicted using a NetPhos2.0 sever [http://www.cbs.dtu.dk/services/NetPhos/]. In NetPhos2.0, serine, threonine, and tyrosine residues with a score of >0.5 were considered to be phosphorylated[21,22]. Ubiquitylation sites were predicted using UbPred [www.ubpred.org]. In Ubpred, lysine residues with a score of ≥ 0.62 were considered ubiquitylated. Sumoylation is important for the protein to attain its correct conformation and thus these sites were predicted using SUMOplot [http://www.abgent.com/sumoplot]. For SUMOplot high probability motifs having a score of 0.5 were considered sumoylated.

***DENND1A* interacting protein partners**

Protein-protein interaction networks are eminent to uncover and annotate all functional interfaces among cell proteins. To find out the interaction of *DENND1A* protein with other cell proteins, the online database resource "Search Tool for the Retrieval of Interacting proteins" [STRING; http://string-db.org/] was used. This imparted the knowledge relating to the unique relations and convenient access to both experimental

and theoretical interaction evidence of - DENND1A. The input options for the STRING database include protein name, protein sequence, and multiple sequences[23].

Predicting the effect of SNPs on protein stability

I-Mutant 2.0 [http://folding.biofold.org/i-mutant/i-mutant2.0.html], a neural network-based tool, predicts the variation in the stability of the protein upon mutation. This tool automatically predicts protein stability changes upon single-site mutations. Prediction can be done using a protein structure or sequence from databases. The FASTA sequence of the *DENND1A* protein with the accession number Q8TEH3 was retrieved from UNIPROT to predict the mutational effect on protein stability. Output illustrates mutations with respect to the change in stability and Gibbs-free energy change [G]. The energy change due to the incorporation of the mutation was calculated using the SWISS PDB VIEWER software. The mutations were incorporated in the predicted structure using the PyMol software and the mutated proteins were used for predicting the energy change using the tools in SWISS PDB VIEWER.

Results

The dbSNP database contains both legitimated and not legitimated polymorphisms. Despite these shortcomings, dbSNP was used because it is the most elaborate SNP database. *DENND1A* gene contains 1,16,762 SNPs compiled from NCBI and the dbSNP database. This list was then manually curated to validate the mutations. When these SNPs were filtered to have their multiple allele frequencies in the range of 0.05 to 0.49, only 1984 SNPs were displayed. This list was furthered filtered using a heterozygosity score of 40 – 50 to obtain only 577 SNPs to be validated using dbSNP – Q and those SNP which have been withdrawn were deleted. Thus a list of 215 SNPs remains, which was merged with the SNP list from the variation viewer platform to give a list of 890 SNPs to be used as input for further downstream analyses.

Deleterious nsSNPs of *DENND1A* gene

The above-curated list of 890 SNPs was fed as input for SIFT program that screens for intolerable SNP mutations (Table 1). SIFT reported 4 prominent SNPs out of which 3 reported being deleterious. We used PROVEAN and PANTHER to characterize the functional substitution effect through evolutionary relationship classification in *DENND1A* protein. PROVEAN is advantageous for proteins like *DENND1A* where only the primary structure is known and it has predicted only 2 SNPs as deleterious whereas PANTHER predicts 3 SNPs as majorly deleterious as they occur in a highly conserved region. PANTHER predicts the deleterious SNPs based on the PSEP [Position Specific Evolutionary Preservation] score when the PSEP score is ≥ 200my [million years] and those with a score of < 200my is said to be benign. SNP & GO gave positive disease-causing damaging effects for 2 SNPs; V179G and P331L, respectively. All the results are briefed in (Table 1). These SNPs were used for further structural analysis.

***DENND1A* SNPs identified by GWAS for PCOS**

The SNPs listed by GWAS and the RegulomeDB output is summarized in (Table 2). The intronic variants identified by GWAS were further scrutinized for their disease-causing ability using Mutation Taster. It was found that two of the reported SNPs

Table 1 — Effect of nsSNP in *DENND1A* gene predicted using different *in silico* tools

Snp	Amino Acid Change	Sift		Provean		Panther		Snp & Go		Polyphen2	
		Score	Prediction	Score	Prediction	Preservation Time (My)	Prediction	Score	Prediction	Score	Prediction
rs78862236	V179G	0	Damaging	–6.575	Damaging	842	Damaging	0.881	Damaging	1	Damaging
rs112018705	T519M	0	Damaging	–1.519	Neutral	361	Damaging	---	---	0.99	Damaging
rs115363971	P331L	0	Damaging	–8.444	Damaging	750	Damaging	0.925	Damaging	---	---
rs116352541	P827S	0.09	Tolerant	–0.406	Neutral	----	-----	---	---	---	---

Table 2 — REGULOMEdb score for intronic polymorphisms in *DENND1A* gene

Chromosome	Coordinates	Coordinates	rs ID	Regulome Score
chr9	126446777	126446778	rs10818854	4
chr9	126525211	126525212	rs2479106	5
chr9	126549954	126549955	rs10986105	5

protein features	start (aa)	end (aa)	feature	details	
	24	91	DOMAIN	UDENN.	might get lost (downstream of altered splice site)
	92	273	DOMAIN	DENN.	might get lost (downstream of altered splice site)
	304	371	DOMAIN	dDENN.	might get lost (downstream of altered splice site)
	394	394	MOD_RES	Phosphotyrosine.	might get lost (downstream of altered splice site)
	456	456	CONFLICT	E -> G (in Ref. 1; BAB15002).	might get lost (downstream of altered splice site)
	519	519	MOD_RES	Phosphothreonine.	might get lost (downstream of altered splice site)
	520	520	MOD_RES	Phosphoserine.	might get lost (downstream of altered splice site)
	523	523	MOD_RES	Phosphoserine.	might get lost (downstream of altered splice site)
	536	536	MOD_RES	Phosphoserine.	might get lost (downstream of altered splice site)
	538	538	MOD_RES	Phosphoserine.	might get lost (downstream of altered splice site)
	546	546	MOD_RES	Phosphoserine.	might get lost (downstream of altered splice site)
	592	592	MOD_RES	Phosphoserine.	might get lost (downstream of altered splice site)
	649	995	COMPBIAS	Pro-rich.	might get lost (downstream of altered splice site)

Fig. 2 — Mutation Taster results for SNPs rs 2479106 and rs 10986105 which were identified *via* GWAS

[rs2479106 & rs10986105] were found to alter a splicing region in the upstream and thus affects almost all protein features of *DENND1A* (Fig. 2).

Interestingly, the analysis also predicted another SNP [rs112018705] in the protein-coding region which is highly damaging and decreases the protein stability.

Modeling and validating structural stability of *DENND1A* protein

The prediction of the structure of *DENND1A* is essential for understanding the structural variation possibly caused due to a mutation. SWISS-MODEL was used as a web-based platform to generate a 3D structure model for wild-type *DENND1A*. The input template required for modeling was obtained by running a BLASTP algorithm against the PDB database. A template of DENND1B -Chain A, GEF domain complex with Rab GTPase Rab35, showed 69% identity with *DENND1A* and was chosen for homology modeling. Once the template sequence and the target sequence were aligned, the 3D model was constructed automatically using an auto-model class. The homologous model generated using a template [SMTL id 3tw8.2] covered a total of 376 residues starting from met1 to asn376 position out of 1009 amino acid residues of *DENND1A* protein (Fig. 3). This region is found to be similar in both the variants as the variant 2 has the first 559 amino acids similar to the variant 1. The variant 2 lacks the proline-rich domain and has some alterations in the C-terminal domain.

Ramachandran Plot was used to validate the protein model obtained from SWISS-MODEL workspace, according to which 378 residues were obtained in the

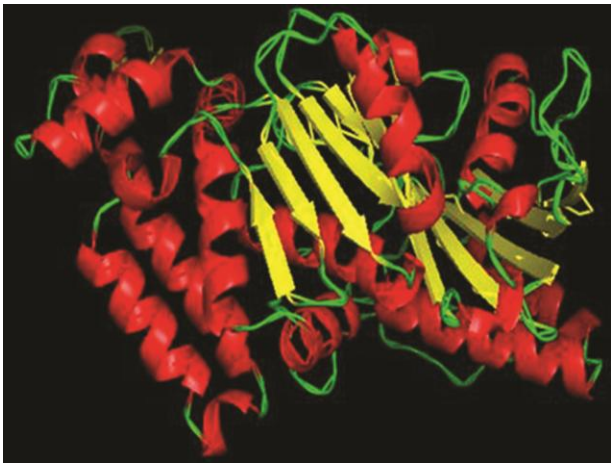

Fig. 3 — A pictorial representation of *DENND1A* protein structure that was predicted partially from 1 to 376 using SWISS MODELER platform and validated using the Ramachandran plot. This representation was visualized using PyMol software

final *DENND1A* model (Fig. 4). Out of all the amino acids, 360 [96.3%] were in the most favourable region, 11 [2.9%] in allowed regions, and 3 [0.8%] in disallowed regions. Hence, the stability of SWISS-MODEL is acceptable. The QMEAN score consists of a linear combination of 6 terms. The pseudo energies of the contributing terms are the total Q mean score which is relatable with its z-score -1.86, which comes from under an estimated model reliability value between 0 and 1.

Post-translation modification and ligand binding sites on *DENND1A* protein

We used various *in silico* tools to study how nsSNPs influence post-translational modification of

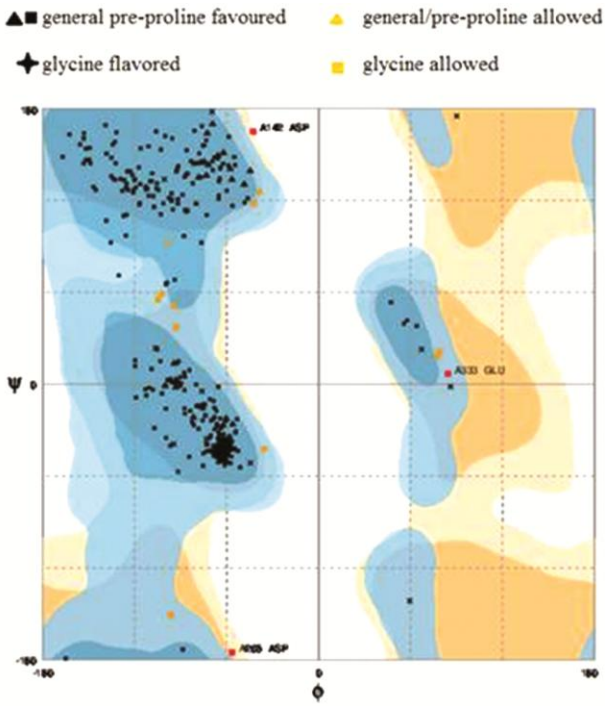

Fig. 4 — Ramachandran plot validation output graph from the SWISS MODELER predicted protein. The plot shows that the structure can be considered valid and that 96.3% of the predicted residues portion fall under favorable region 2.9% in the allowable regions and 0.8% in the disallowed region

DENND1A protein. NetGlycan predicted that 12 residues undergo glycation. According to NEtPhos, 70 threonine, 105 serine, and 10 tyrosine residues undergo phosphorylation. Protein sequence with a mutational position and amino acid variant associated with nsSNPs were submitted as input in UbPred, 17 residue positions have a confidence level that is capable of variations (Table 3). Similarly, with SUMOplot we predicted that there are 8 different positions that have the chances for sumoylation (Table 4).

Protein-Protein Interactions

A 3D molecular structure becomes meaningful only when its context with other proteins is identified through its assembly, regulation, and signalling interactions. The STRING provides such a platform to understand in detail about the protein interaction, which is a report based on critical assessment of curated data from experiments and databases. The Figure 5 shows the first-line interacting proteins with *DENND1A*. Three of the major interacting proteins with DENDD1A as reported by STRING are (i) Rab35, an important protein responsible for membrane protein trafficking inside a cell; (ii) CLTB and CLCT, Clathrin light chains and heavy chains, respectively, which coat the pits in vesicles for endocytosis; (iii) GAK, essential for uncoating of Clathrin-coated pits. Surprisingly NEK6 and NEK7, which are responsible for regulating meiotic spindle formation, are also key interactors of *DENND1A* thus showing the indirect association of *DENND1A* in cell cycle regulation. Other important first line interacting proteins are CRB2, IFT27, and CENPJ.

Table 3 — UbPred Scoring that predicts the possible ubiquitination in the *DENND1A* protein structure

Residue	Score	Ubiquitination
6	0.68	Yes Low confidence
452	0.67	Yes Low confidence
469	0.82	Yes Medium confidence
479	0.71	Yes Medium confidence
482	0.89	Yes High confidence
553	0.92	Yes High confidence
591	0.76	Yes Medium confidence
600	0.69	Yes Medium confidence
637	0.62	Yes Low confidence
651	0.71	Yes Medium confidence
659	0.87	Yes High confidence
686	0.87	Yes High confidence
723	0.76	Yes Medium confidence
937	0.79	Yes Medium confidence
980	0.88	Yes High confidence
982	0.92	Yes High confidence
1003	0.63	Yes Low confidence

Table 4 — Sumoylation predicted by Sumoplot that gives a prediction on possible sumoylation that can occur in *DENND1A* protein variant 2

Number	Position	Group	Score
1	K427	SYRNALKIEPEEPI	0.91
2	K608	PRPHVVKRPKSNIA	0.82
3	K540	FAKDHAKMGIKEVK	0.62
4	K703	EDLRAPKDLREQP	0.61
5	K791	ILNPSDKEEVPTPT	0.5
6	K465	Q LQLFKQFIDGRL	0.5
7	K233	LLETLHKLPIPDPG	0.41
8	K811	PRPQGRKTPELGIV	0.34

Effect of mutations on the stability of the protein

DENND1A mutants were modeled against their wild-type using PyMOL (Fig. 6). A comparative and superimposed observation showed no difference between the mutants and the wild-type. They were hence extensively studied. The total energy after minimization and electrostatic constraint of native protein were −21245.529 kJ/mol and −12995.61 kJ/mol

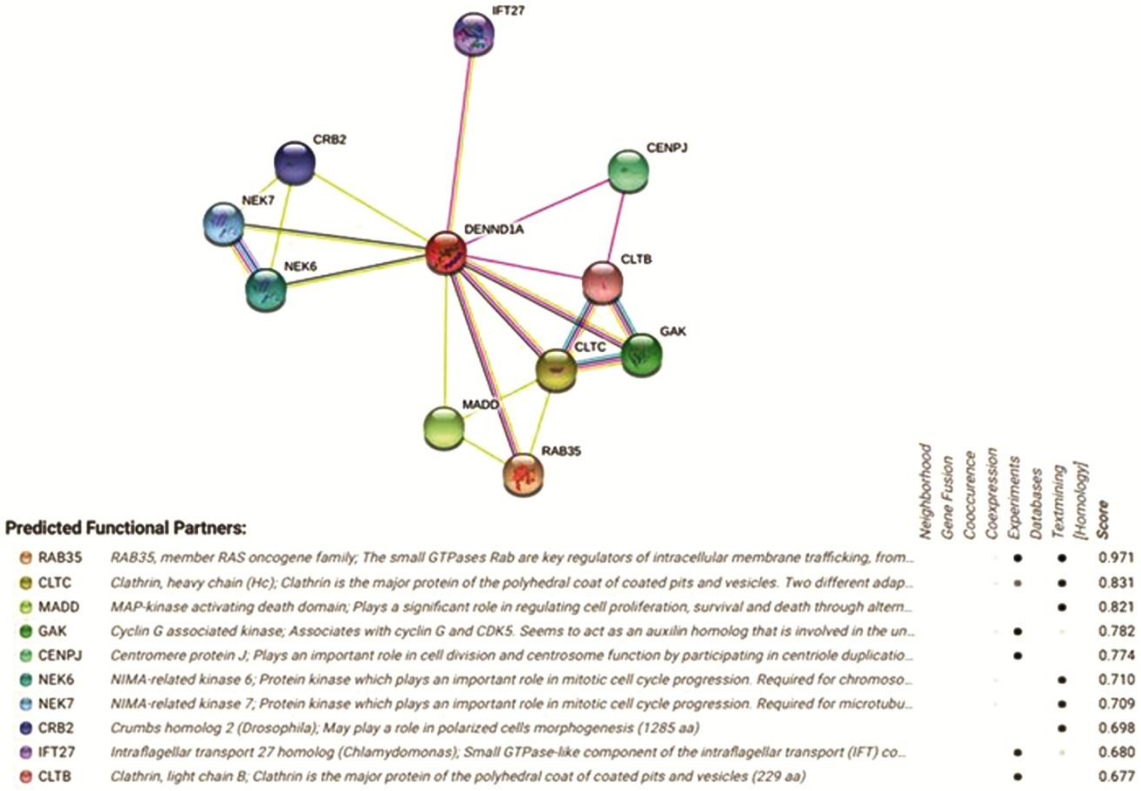

Fig. 5 — A pictorial representation of the first line interacting proteins shows that *DENND1A* have major interactions with Rab35, CLTB, CLCT, GAK, NEK6 and NEK7, CRB2, IFT27 and CENPJ

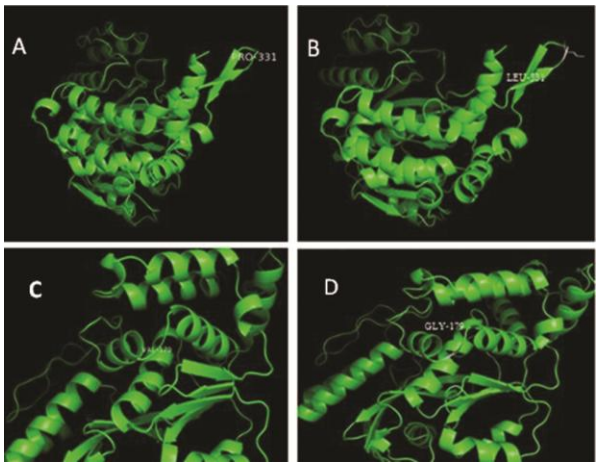

Fig. 6 — *DENND1A* protein structure predicted using SWISS MODELER and mutation incorporated structure using PyMol protein visualization software to understand the possibility of structural isolation due to the incorporation of mutation. (A) Native structure with Proline at position 331; (B) Mutation of Leucine incorporated at position 331; (C) Native structure with Valine at position 179; and (D) Mutation of Glycine incorporated at position 179

whereas, that of the Mut179 were −21138.232 kJ/mol and −12352.56 and Mut331 were −21275.221 kJ/mol and −12420.47. The energy of the mutant structures showed thermodynamically favourable changes in comparison with the wild-type. These genomic variants have the probability of further decreasing the stability of the *DENND1A* protein.

Table 5 — I-Mutant table that gives the change in the stability of protein with an introduction of polymorphism

Position	WT	New	Stability	RI	pH	T
179	D	G	Decrease	7	7.0	25
519	G	M	Decrease	5	7.0	25
331	V	L	Decrease	8	7.0	25
827	Q	S	Decrease	5	7.0	25

Stability prediction [ΔG] and reliability index [RI] upon mutation of the protein was reported by I – MUTANT (Table 5). Those SNPs which are reported to be damaging or deleterious by SIFT, PolyPhen, PROVEAN, and PANTHER were given as input. There are higher chances that protein stability might decrease due to mutations at positions179, 519, 331, and 827. Thus, the stability will be more affected at position 331 and 179 when compared to the rest of the residues.

Discussion

Trait variation among the world population is contributed by Single Nucleotide Polymorphisms

[SNPs]. SNP is a position at which two of the alleles occur at an appreciable frequency [0.1%] in the human population and are stable because of the low rate of recurrent mutation making them a probable biomarker[24]. Such inherited differences in the DNA sequence contribute majorly to the phenotypic variation thus affecting an individual's anthropometric characteristics; response to the environment and possibility of a disease. Thus, SNPs may serve as genetic markers for screening disease-causing genes through family analysis for linkage studies, investigating the linkage disequilibrium among the isolated population, association analysis of patients and controls *etc*[25].

Polycystic Ovarian Syndrome [PCOS] is one such complex disorder where the disease manifests due to the altered response of the variants to the environment[26]. Though each interaction and response needs to be studied in detail to unravel the pathophysiology, initial *in silico* analysis of the candidate genes and its variants may give indications for translational analysis.

For PCOS, six Genome-Wide Association Studies [GWAS] have been carried out until now and has listed out more than 20 candidate genes[9,10,27-31]. *DENND1A* SNPs were identified as potentially pathogenic by three of the GWAS and many literatures give the association of *DENND1A* with PCOS[32,33]. The *in silico* analysis of this gene carried out in this study, has revealed a few pathogenic mutations leading to alterations in the protein structure and consequently its function.

Usually, the nsSNPs present on the shell of a protein is possibly tolerated but those identified by GWAS are positioned at functionally important sites like the splicing region and thus alter the protein structure, protein-protein interface and ultimately impair the protein's stability[34].

Apart from the 3 *DENND1A* SNPs identified by GWAS studies for PCOS, our *in silico* analysis has identified 4 other SNPs. These SNPs are present in the coding region and are found to be highly pathogenic. But the disease associations of these variants are yet to be studied. The identified pathogenicity can be authenticated by the fact that almost all the disease-causing mutations are present on the sites which are highly conserved and thus are likely to be structurally and functionally important[35-38].

It is also known that the splicing mechanism of *DENDD1A* gives rise to two variants – *DENND1A* Variant 1 and *DENND1A* Variant 2. The longer of these transcripts, variant1, encodes 1009 amino-acid protein which consists of a C-terminal proline-rich domain. The variant 2 encodes a truncated 559 amino-acid protein that contains the DENN domain and the clathrin-binding domain, but it does not contribute the proline-rich domain and it also contains a C-terminal 33 amino-acid sequence which is not found in the larger connec denn 1 variant[39].

The *DENND1A* Variant 2 is more closely associated with PCOS by activating the signal transduction cascades and regulating it *via* gonadotropins. It also increases the expression of steroidogenic enzymes by regulating the intranuclear transport of signal transducers in conjunction with *DENND1A* Variant 1[40]. The *DENND1A* Variant 2 expressed on the ovarian theca cells is responsible for androgen biosynthesis and a PCOS phenotypic ovary has increased expression of *DENND1A* Variant 2[9].

Thus the DENN domains and Clathrin Binding motifs are conserved in both the variants and thus the SNPs found in these regions may play a role in causing PCOS. By using *Caenorhabditis elegans* as a model organism it was shown that the DENN domain which binds to Rab-35 is responsible for the recruitment of this protein to Clathrin-Coated Vesicles [CCVs] for subsequent transport to endosomes where it controls yolk receptor recycling[41]. Thus the [GWAS] SNPs which are found to be altering the protein function of the DENN Domain may play a major role in PCOS.

Post Translational Modification [PTM] are important for protein localization, enzyme activity, degradation and interaction with other proteins to give signals. Some important PTMs which we have screened are for phosphorylation, glycosylation, ubiquitination, and sumoylation. The DENN domain has a sumolytion region while the proline-rich region has the most PTM sites. It has 4 sumolytion sites, 7 ubiquitination sites, 18 serine phosphorylation sites, and 14 threonine sites. It was found that the rs2479106 (Table 2) lies on a conservatory motif – Hmbox1, which is a Transcription Factor Binding site.

Conclusion

Through this study, *in silico* analysis, we have listed the most potential variants of *DENND1A* associated to Polycystic Ovarian Syndrome. The intronic variants previously reported in the Chinese, and European populations were found to affect the transcription process. Similarly, the predicted damaging nsSNPs reported in our study, affect the stability of the protein and thus alter the bio functionality of *DENND1A*. The instability of the DENND1A might affect the function

of the protein and its downstream processes and thus contribute to the pathology of PCOS.

The *DENND1A* seems to play a major role in androgen biosynthesis and malfunction or overexpression of a particular type of *DENND1A* variant [*DENND1A* Variant 2] leads to an increased steroid hormone biosynthesis. This information hints us on the fact that the SNPs on *DENDD1A* may lead to a hyperandrogenic phenotype of PCOS and thus may help in precise diagnosis and treatment. Thus to verify the identified nsSNP's association with this disease state, population screening needs to be done for the respective ethnic population.

Acknowledgement

The author, Dakshina Moorthy Janani, is supported by funds from Women Scientist – A Scheme (SR/ WOS-A/LS-253/2017), Department of Science and Technology, Government of India. We would also extend our thanks to the PCOS Awareness Association of India, The PCOS Foundation, Chennai (PCOSSAI) in creating awareness on the challenges faced by the researchers in improvising the diagnosis and treatment methods for PCOS.

Conflict of Interest

All authors declare no conflict of interest.

References

1 Ramensky V, Bork P & Sunyaev S, Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*, 30 (2002) 3894.

2 Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G & Kang HM, Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*, 337 (2012) 64.

3 Shinde G, Khavale S, Dadachanji R, D'Souza S, Nandedkar T & Mukherjee S, Estrogenized mouse model of polycystic ovary highlights mitochondrial pathway of apoptosis. *Indian J Exp Biol*, 56 (2018) 293.

4 Xi W, Gong F & Lu G, Correlation of serum Anti-Müllerian hormone concentrations on day 3 of the *in vitro* fertilization stimulation cycle with assisted reproduction outcome in polycystic ovary syndrome patients. *J Assist Reprod Genet*, 29 (2012) 397.

5 Shah D & Rasool S, Polycystic ovary syndrome and metabolic syndrome: the worrisome twosome?. *Climacteric*, 19 (2016) 7.

6 Bharathi RV, Swetha S, Neerajaa J, Madhavica JV, Janani DM, Rekha SN, Ramya S & Usha B, An epidemiological survey: Effect of predisposing factors for PCOS in Indian urban and rural population. *Middle East Fertility Soc J*, 22 (2017) 313.

7 Marat AL, Dokainish H & McPherson PS, DENN domain proteins: regulators of Rab GTPases. *J Biol Chem*, 286 (2011) 13791.

8 Chen ZJ, Zhao H, He L, Shi Y, Qin Y, Shi Y, Li Z, You L, Zhao J, Liu J & Liang X, Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16. 3, 2p21 and 9q33. 3. *Nature Genet*, 43 (2011) 55.

9 Yoshimura SI, Gerondopoulos A, Linford A, Rigden DJ & Barr FA, Family-wide characterization of the DENN domain Rab GDP-GTP exchange factors. *J Cell Biol*, 191 (2010) 367.

10 Allaire PD, Ritter B, Thomas S, Burman JL, Denisov AY, Legendre-Guillemin V, Harper SQ, Davidson BL, Gehring K & McPherson PS, Connecdenn, a novel DENN domain-containing protein of neuronal clathrin-coated vesicles functioning in synaptic vesicle endocytosis. *J Neurosci*, 26 (2006) 13202.

11 McAllister JM, Modi B, Miller BA, Biegler J, Bruggeman R, Legro RS & Strauss JF, Overexpression of a *DENND1A* isoform produces a polycystic ovary syndrome theca phenotype. *Proc Natl Acad Sci USA*, 111 (2014) E1519.

12 Marat AL & McPherson PS, The connecdenn family, Rab35 guanine nucleotide exchange factors interfacing with the clathrin machinery. *J Biol Chem*, 285 (2010) 10627.

13 Chen ZJ, Zhao H, He L, Shi Y, Qin Y, Shi Y, Li Z, You L, Zhao J, Liu J & Liang X, Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16. 3, 2p21 and 9q33. 3. *Nature Genet*, 43 (2011) 55.

14 Shi Y, Zhao H, Shi Y, Cao Y, Yang D, Li Z, Zhang B, Liang X, Li T, Chen J & Shen J, Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nature Genet*, 44 (2012) 1020.

15 Jaiswal A & Srivastava JP, Changes in reactive oxygen scavenging systems and protein profiles in maize roots in response to nitric oxide under waterlogging stress. *Indian J Biochem Biophys*, 55 (2018) 26.

16 Hepp D, Goncalves GL & De Freitas OTR, Prediction of the damage-associated non-synonymous single nucleotide polymorphisms in the human MC1R gene. *PLoS One*, 10 (2015) 1.

17 Srivastava S & Pandey A, Computational screening of anticancer drugs targeting miRNA155 synthesis in breast cancer. *Indian J Biochem Biophys*, 57 (2020) 389.

18 Schwede T, Kopp J, Guex N & Peitsch MC, SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res*, 31 (2003) 3381.

19 Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L & Schwede T, SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*, 42 (2014) 1.

20 Lovell SC, Davis IW, Adrendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS & Richardson DC, Structure validation by C alpha geometry: phi,psi and C beta deviation. *Proteins-Structure Funct Genet*, 50 (2003) 437.

21 Chattopadhyay S, Bagchi P, Dutta D, Mukherjee A, Kobayashi N & Chawlasarkar M, Computational identification of posttranslational modification sites and functional families reveal possible moonlighting role of rotaviral proteins. *Bioinformation*, 4 (2010) 448.

22 Blom N, Gammeltoft S & Brunak S, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol*, 294 (1999) 1351.

23 Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP & Kuhn M, STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*, 43 (2014) D447.

24 Syvänen AC, Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet.*, 2 (2001) 930.

25 Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL & Hunt SE, A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409 (2001) 928.

26 De Leo V, Musacchio MC, Cappelli V, Massaro MG, Morgante G & Petraglia F, Genetic, hormonal and metabolic aspects of PCOS: an update. *Reprod Biol Endocrinol*, 14 (2016) 38.

27 Hwang JY, Lee EJ, Jin GM, Sung YA, Lee HJ, Heon Kwak S, Jang HC, Soo PK, Lee HJ, Byul JH & Song J, Genome-wide association study identifies GYS2 as a novel genetic factor for polycystic ovary syndrome through obesity-related condition. *J Hum Genet*, 57 (2012) 660.

28 Lee H, Oh JY, Sung YA, Chung H, Kim HL, Kim GS, Cho YS & Kim JT, Genome-wide association study identified new susceptibility loci for polycystic ovary syndrome. *Hum Reprod*, 30 (2015) 723.

29 Day FR, Hinds DA, Tung JY, Stolk L, Styrkarsdottir U, Saxena R, Bjonnes A, Broer L, Dunger DB, Halldorsson BV & Lawlor DA, Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nat Commun*, 6 (2015) 9464.

30 Day F, Karaderi T, Jones MR, Meun C, He C, Drong A, Kraft P, Lin N, Huang H, Broer L & Magi R, Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PLoS Genet*, 14(12) (2018) e1007813.

31 Hiam D, Moreno-Asso A, Teede HJ, Laven JS, Stepto NK, Moran LJ & Gibson-Helm M, The Genetics of Polycystic Ovary Syndrome: An Overview of Candidate Gene Systematic Reviews and Genome-Wide Association Studies. *J clin Med*, 10 (2019) 1606.

32 Eriksen MB, Nielsen MF, Brusgaard K, Tan Q, Andersen MS, Glintborg D & Gaster M, Genetic alterations within the *DENND1A* gene in patients with polycystic ovary syndrome [PCOS]. *PLoS One*, 8 (2013) 7186.

33 Bao S, Cai JH, Yang SY, Ren Y, Feng T, Jin T & Li ZR, Association of *DENND1A* gene polymorphisms with polycystic ovary syndrome: a meta-analysis. *J Clin Res Pediatr Endocrinol*, 8 (2016)135.

34 David A, Razali R, Wass MN & Sternberg MJE, Protein–protein interaction sites are hot spots for disease-associated non-synonymous SNPs. *Hum Mutat*, 33 (2012) 359.

35 Paul JJ, Surendran A & Thatheyus AJ, Efficacy of orange peel in the decolourization of the commercial auramine yellow dye used in textile industry. *Indian J Biochem Biophys*, 57 (2020) 481.

36 Anithasri A, Arulvijayavani S, Gopinath P & Varuni AK, Association of Lipid biomarkers with drinking patterns and severity in alcoholic liver disease - A hospital- based cross-sectional study. *Indian J Biochem Biophys*, 57 (2020) 395.

37 Burke DF, Worth CL, Priego EM, Cheng T, Smink LJ, Todd JA & Blundell TL, Genome bioinformatic analysis of non-synonymous SNPs. *BMC Bioinformatics*, 8 (2007) 301.

38 Yates CM & Sternberg MJ, The effects of non-synonymous single nucleotide polymorphisms [nsSNPs] on protein–protein interactions. *J Mol Biol*, 425 (2013) 3949.

39 McAllister JM, Legro RS, Modi BP & Strauss JF, Functional genomics of PCOS: from GWAS to molecular mechanisms. *Trends Endocrin Met*, 26 (2015) 118.

40 McAllister JM, Modi B, Miller BA, Biegler J, Bruggeman R, Legro RS & Strauss JF, Overexpression of a *DENND1A* isoform produces a polycystic ovary syndrome theca phenotype. *Proc Natl Acad Sci U S A*, 111 (2014) E1519.

41 Sato M, Sato K, Liou W, Pant S, Harada A & Grant BD, Regulation of endocytic recycling by *C. elegans* Rab35 and its regulator RME-4, a coated pit protein. *EMBO J*, 27 (2008) 1183.