Journal of Endocrinological Investigation (2021) 44:1783–1793
https://doi.org/10.1007/s40618-021-01498-4

ORIGINAL ARTICLE

Conceptualization of functional single nucleotide polymorphisms of polycystic ovarian syndrome genes: an in silico approach

B. N. Prabhu[1] · S. H. Kanchamreddy[1] · A. R. Sharma[1] · S. K. Bhat[2] · P. V. Bhat[2] · S. P. Kabekkodu[3] · K. Satyamoorthy[3] · P. S. Rai[1]

Received: 23 June 2020 / Accepted: 2 January 2021 / Published online: 27 January 2021

Abstract

Purpose Polycystic ovarian syndrome (PCOS) is a multi-faceted endocrinopathy frequently observed in reproductive-aged females, causing infertility. Cumulative evidence revealed that genetic and epigenetic variations, along with environmental factors, were linked with PCOS. Deciphering the molecular pathways of PCOS is quite complicated due to the availability of limited molecular information. Hence, to explore the influence of genetic variations in PCOS, we mapped the GWAS genes and performed a computational analysis to identify the SNPs and their impact on the coding and non-coding sequences.

Methods The causative genes of PCOS were searched using the GWAS catalog, and pathway analysis was performed using ClueGO. SNPs were extracted using an Ensembl genome browser, and missense variants were shortlisted. Further, the native and mutant forms of the deleterious SNPs were modeled using I-TASSER, Swiss-PdbViewer, and PyMOL. MirSNP, PolymiRTS, miRNASNP3, and SNP2TFBS, SNPInspector databases were used to find SNPs in the miRNA binding site and transcription factor binding site (TFBS), respectively. EnhancerDB and HaploReg were used to characterize enhancer SNPs. Linkage Disequilibrium (LD) analysis was performed using LDlink.

Results 25 PCOS genes showed interaction with 18 pathways. 7 SNPs were predicted to be deleterious using different pathogenicity predictions. 4 SNPs were found in the miRNA target site, TFBS, and enhancer sites and were in LD with reported PCOS GWAS SNPs.

Conclusion Computational analysis of SNPs residing in PCOS genes may provide insight into complex molecular interactions among genes involved in PCOS pathophysiology. It may also aid in determining the causal variants and consequently contributing to predicting disease strategies.

Keywords Polycystic ovarian syndrome · Single nucleotide polymorphisms · miRNAs · Transcription factors · Enhancers

Introduction

Polycystic ovarian syndrome (PCOS) is a multifactorial endocrine disorder with uncertain etiologies among reproductive-aged females and is a frequent cause of infertility in women [1]. It is manifested by several endocrine disturbances such as chronic anovulation, hyperandrogenism characterized by frontal alopecia, acne and hirsutism, presence of multiple cysts in ovaries, and metabolic consequences including a high risk of obesity, insulin resistance, type 2 diabetes mellitus (T2DM) and cardiovascular diseases [2, 3] and psychological complications such as increased distress and depression [4]. Although not understood completely, this complex disorder is considered to be caused due to intricate interplay between various factors such as genetic and epigenetic predisposition, ethnicity, environmental influences, and lifestyle [5]. It was also conferred

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40618-021-01498-4.

✉ P. S. Rai
padmalatha.rai@manipal.edu

[1] Department of Biotechnology, Manipal School of Life Sciences, MAHE, Manipal, Karnataka, India

[2] Department of Obstetrics and Gynaecology, Dr. T.M.A Pai Hospital, MMMC, MAHE, Manipal, Karnataka, India

[3] Department of Cell and Molecular Biology, Manipal School of Life Sciences, MAHE, Manipal, Karnataka, India

as an evolutionary paradox for impairing fertility in women without diminishing in disease prevalence. Earlier reports on evolutionary dynamics in PCOS encompass only females and not the male's role in the genotype/phenotype distinction. As this disease is known to affect only females, yet males might be the carrier of PCOS linked features such as hyperandrogenism and may contribute to conserving the genetics predisposing to PCOS [6, 7]. Further, these factors can significantly influence the phenotypic complexity of the syndrome.

The pathophysiology of PCOS is relatively challenging due to the involvement of numerous pathways such as insulin signaling pathway, androgen synthesis, altered gonadotropin ratios, glucose, and lipid metabolism [8]. Despite the challenge of the multifaceted nature of PCOS, the heritable factors, including genes and their interaction, gene-environment relation, epigenetic modifications, alteration in proteins, and metabolites, have been reported through different approaches such as genomics, transcriptomics, proteomics, and metabolomics to delineate the molecular pathomechanisms of PCOS [9]. Since the significant information in this complex endocrinopathy is inadequate; there is a prerequisite to integrate the data from Genome-Wide Association Study (GWAS) with in silico analysis.

A gene and its products are controlled by numerous mechanisms that comprise interaction between various genes, pathways, and factors [10]. The most predominant form of genomic variation is Single-nucleotide polymorphisms (SNPs), where two substitute bases exist at a noticeable frequency in humans [11]. Researchers were accustomed to focusing on the SNPs in the coding region of the genome, particularly non-synonymous SNPs (nsSNPs), as they are expected to significantly change the function of encoded proteins [12]. Besides, the unpredicted discovery of the GWAS revealed that > 90% of disease-linked SNPs reside in the non-coding sequence, which is also responsible for contributing to complex diseases [11], and confirms that SNPs can serve as a valuable biomarker to investigate the heritability that influences individuals to specific phenotype including diseases [10]. In the present study, we intended to determine the impact of SNPs in the selected GWAS genes using bioinformatics tools and evaluate their detrimental effects on the structure and function of a protein, miRNA controllers, transcription factor binding elements, and enhancers, which may play a critical role in PCOS susceptibility and assist in delineating the precise pathomechanisms of PCOS.

Methods

Identification of genes involved in the pathogenesis of PCOS

A comprehensive literature screening was conducted using the GWAS catalog (https://www.ebi.ac.uk/gwas/). A manual curation procedure was implemented using the search key term "polycystic ovary syndrome" to identify the causative genes at genome-wide significance ($P < 5 \times 10E^{-8}$) involved in PCOS pathogenesis.

Pathway interaction among PCOS genes

The identified PCOS GWAS genes were imported to the Cytoscape tool, and a plug-in named ClueGO v2.5.7 [13] was used for biological and functional interpretation of a large number of genes to constitute the networks. Molecular function, cellular components, biological process, KEGG, and reactome pathways were the different ontologies used in the framework. Kappa statistics were used to connect the terms, and the network was visualized in the circular layout.

Data retrieval and SNPs characterization

The identified genes and their symbols were subjected to SNP search in the Ensembl genome browser (m.ensembl.org) using the option variant table. The list of SNPs identified was further categorized into 5′-UTR SNPs, synonymous SNPs, intronic SNPs, missense SNPs, 3′-UTR SNPs, splice region SNPs, splice donor SNPs, splice acceptor SNPs, stop retained SNPs, stop-gained SNPs, stop-lost SNPs, and non-coding transcript exon SNPs. Among these SNPs, non-synonymous SNPs (nsSNPs) were subsequently used for downstream analysis.

Prediction of nsSNP functional impacts by in silico analysis

The retrieved nsSNP were analyzed using six different tools with mutation score available in the Ensembl genome browser, namely PolyPhen-2 (Polymorphism Phenotyping), SIFT (Sorting Intolerant from Tolerant), CADD (Combined Annotation-Dependent Depletion), Revel (Rare exome variant ensemble learner), MetaLR, and Mutation assessor. Finally, the SNPs categorized as "deleterious" in all 6 tools were selected and analyzed to influence the protein structure and stability.

Protein modeling and impact of the mutation on protein structure

The native and mutant forms of deleterious SNPs were modeled to predict the mutation's effect on protein structure and function. We tabulated the hydropathy index proposed by Jack Kyte and Russell F Doolittle [14], which revealed the modification in hydrophilicity or hydrophobicity due to amino acid change in the protein. The proteins structures were computed using Iterative Threading ASSEmbly Refinement (I-TASSER) [15] using an amino acid template from

the Uniprot database. Further mutation analysis and energy calculations were performed on the Swiss-Pdb viewer. PyMOL software's align function was used to calculate the root-mean-square deviation (RMSD) value of mutant type from native protein.

Functional microRNA target SNPs prediction

The identified genes involved in PCOS pathogenesis were subjected to functional microRNA binding SNP prediction using the miRNA-related SNPs (MirSNP) database [16], the PolymiRTS database [17], and the microRNA related Single Nucleotide Polymorphisms v3 (miRNASNP3) database [18]. The gene symbols of the shortlisted genes were used in the MirSNP database to search the miRNA binding SNP sites and their effects on the target site. In the PolymiRTS database, the search options containing gene symbol was used to retrieve the SNPs and their associated miRNAs at ancestral and mutant allele. The miRNASNP3 database was used to retrieve microRNA related SNPs with their impact on the target gain/loss in the 3′-UTR region.

SNPs at transcription factor binding site

The identified PCOS genes were utilized to find the SNPs in transcription factor binding sites using SNP2TFBS [19]. The annotated variant option was used to retrieve the SNPs present in the 5′-UTR and upstream regions. The SNPInspector (trail access version) in Genomatix Software Suite (https://www.genomatix.de/) was used to predict whether SNPs in TFBS create or disrupt the transcription factor binding sites.

SNPs in enhancers

The identified GWAS genes at genome-wide significance in PCOS were used to examine the impact of SNPs in enhancers using EnhancerDB [20] and HaploReg v4.1, which is developed by ENCODE laboratories [21]. The search option containing gene was used in the EnhancerDB database to search the SNPs located in the enhancers of the respective genes, and the regulatory motifs that were altered of those SNPs were reported using HaploReg.

Linkage disequilibrium analysis of functional SNPs

The identified SNPs that may be functional, obtained by analysing SNPs in coding region, 3′-UTR, 5′-UTR, upstream region and introns of selected GWAS genes in PCOS were further evaluated by performing Linkage disequilibrium (LD) analysis. These SNPs were further correlated with reported PCOS GWAS SNPs using LDlink [22] to examine their impact on disease progression.

Results

Identification of genes associated with the pathogenesis of PCOS

We shortlisted 25 GWAS genes linked with PCOS pathogenesis. The details of the genome-wide significant SNPs used to identify the in/nearest genes associated with PCOS were tabulated from the reported studies (Online Resource 1, 2). The shortlisted genes were mapped them using Idiographica. The representation showed the distribution of genes across 9 autosomes including chromosome 2, 5, 8, 9, 11, 12, 16, 19, 20 all over the genome (Fig. 1). The schematic representation of in silico workflow is depicted in the Fig. 2 (Fig. 2).

Pathway interaction among PCOS genes

The association between PCOS genes using the molecular function, cellular components, biological process, KEGG, and reactome pathways displayed a network showing the interaction of 9 out of 25 shortlisted genes and their pathways after performing enrichment/depletion (Two-sided hypergeometric test) (Fig. 3). The framework also showed 4 Kappa score groups such as hormone ligand-binding receptors, peptide hormone metabolism, cardiac muscle tissue regeneration, and positive regulation of phosphatidylinositol 3-kinase signaling (Fig. 3). It was found that *ERBB4, GATA4* and, *YAP1* genes contributed 60 percent in cardiac muscle tissue regeneration (Fig. 3).

Characterization of SNPs

A total of 16,71,896 SNPs were retrieved by a search using the Ensembl genome browser (GRCh38.p13). As 1000 Genomes Project was recognized with ample account of genetic variations in humans, these SNPs were filtered for the 1000 Genomes Project lead to the identification of 1,04,034 SNPs. Further, these SNPs were categorized based on their function. 260 SNPs were present in the 5′-UTR region, 436 were synonymous SNPs, 1,00,494 were intronic SNPs, 1702 were 3′-UTR SNPs, 86 were splice variants (splice region, splice donor, splice acceptor), 1 stop retained SNP, 16 stop-gained SNPs, 1 stop-lost SNP, 77 were non-coding transcript exon SNPs, and 961 were missense variants of the genes involved in the PCOS (Figs. 4, 5).

Selection of deleterious nsSNPs

Among 961 missense variants, 285 (29.65%) were reported as "deleterious" by SIFT, while the frequency of mutation was reduced to 159 (16.54%) as "probably damaging" by

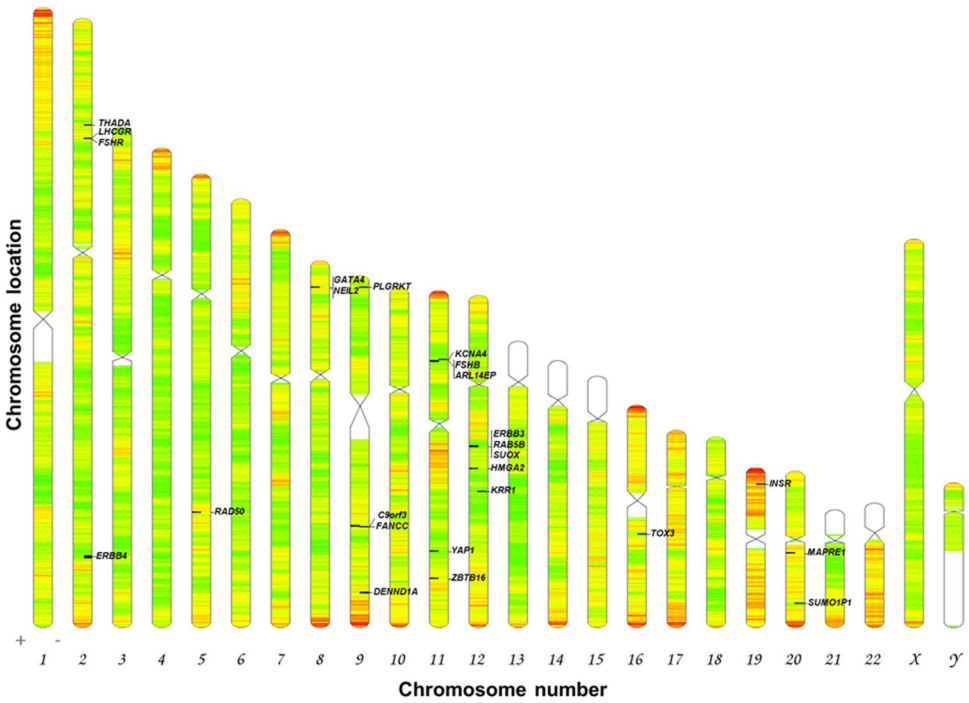

Fig. 1 Chromosome-wide distribution of PCOS GWAS genes

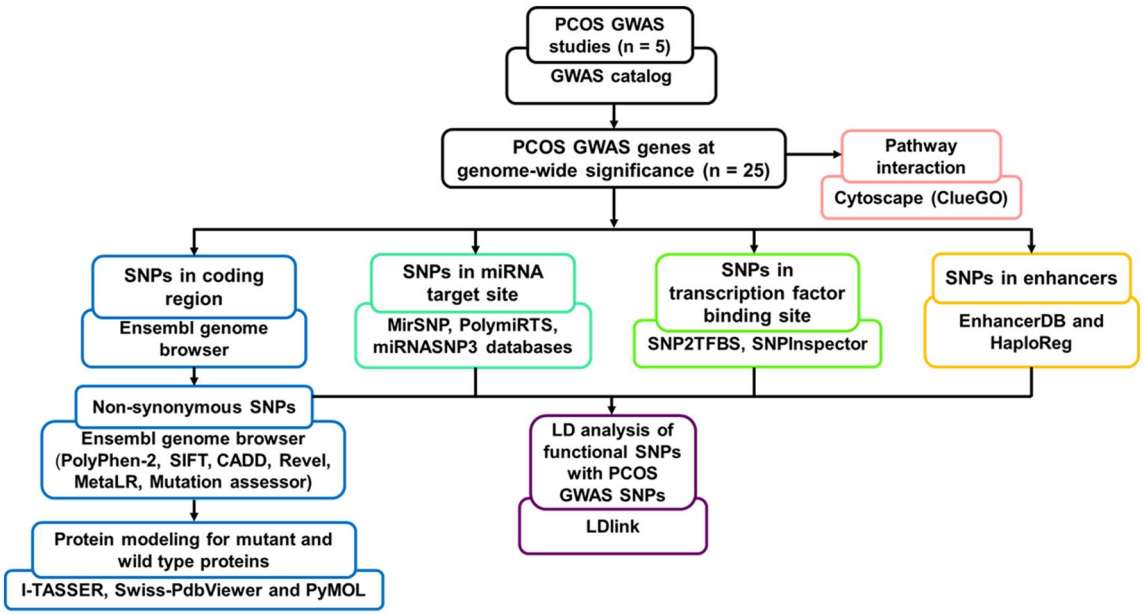

Fig. 2 Schematic representation of in silico workflow

PolyPhen-2, 21 (2.18%) as "likely deleterious" were analysed by CADD, and 123 (12.79%) as "likely disease-causing" by Revel, 150 (15.60%) as "damaging" by Meta LR and 21 (2.18%) as "high" by Mutation Assessor (Fig. 6). Six different bioinformatic tools (SIFT, PolyPhen-2, CADD, Revel, Meta LR, Mutation Assessor) collectively highlighted 7 deleterious nsSNPs (Fig. 7) which included *ERBB4* rs192066345 and rs528780505, *GATA4* rs180765750, *INSR* rs79312957, *LHCGR* rs121912525, *SUOX* rs575660698, and *YAP1* rs199505545 (Online Resource 3).

Protein modeling and impact of the mutation on protein structure

The structures of the proteins were modelled using I-TASSER (Fig. 8). Out of 7 nsSNPs identified, change in

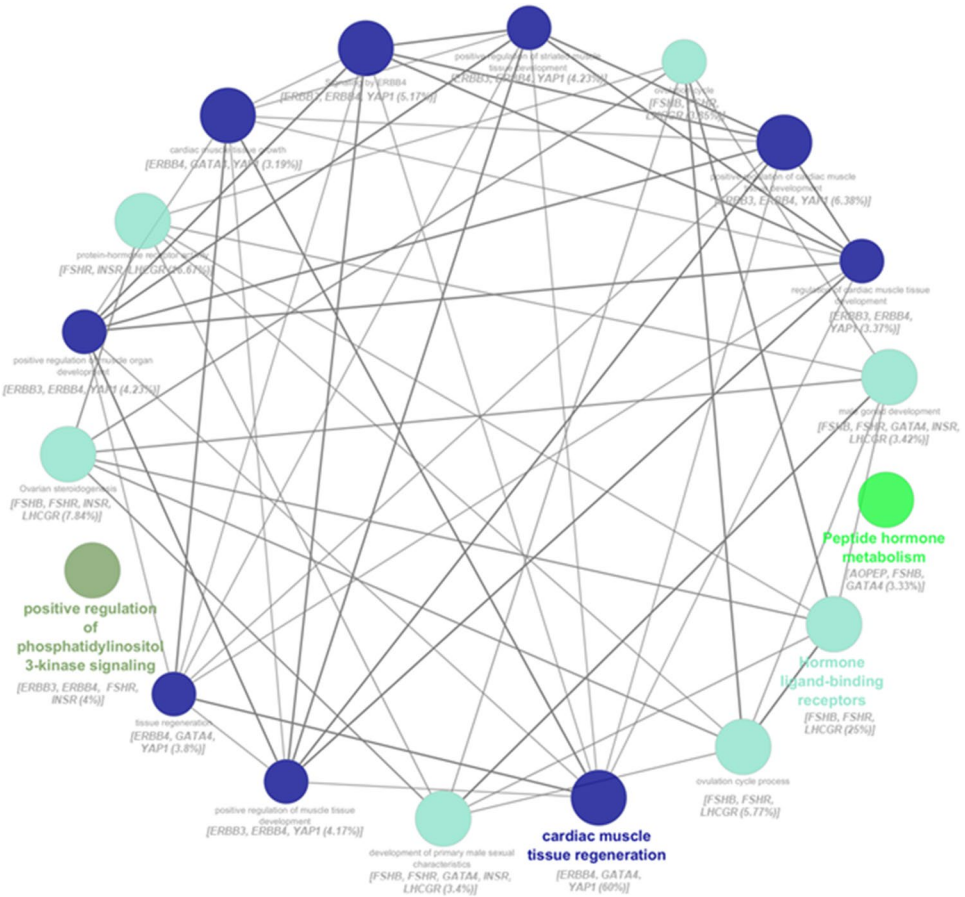

Fig. 3 Pathway interaction of PCOS genes

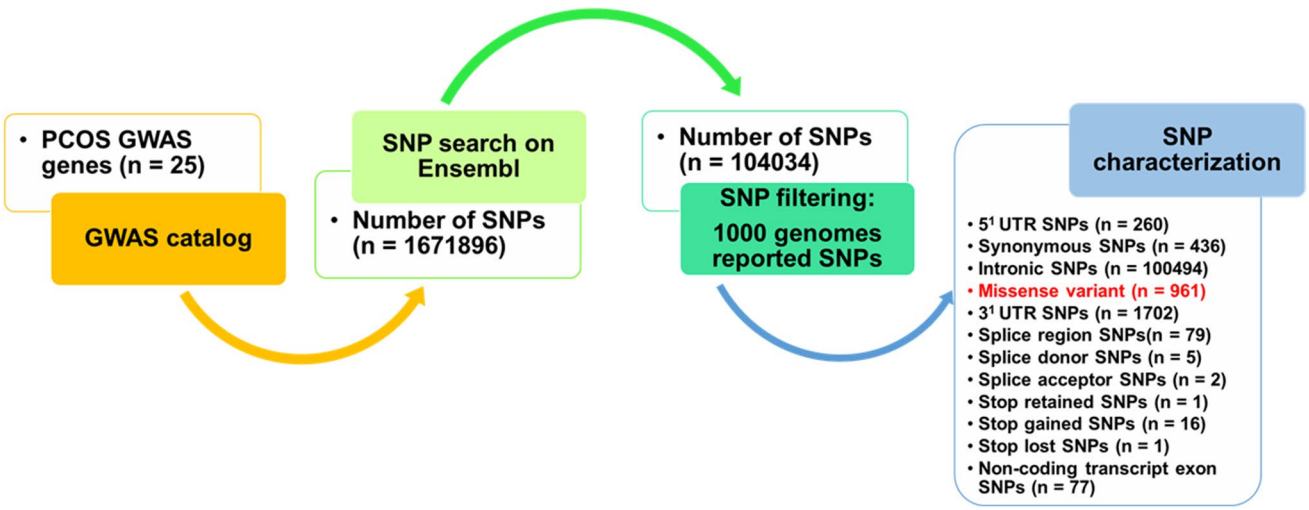

Fig. 4 Schematic representation of in silico SNP search and characterization

amino acid in ERBB4 (rs528780505) suggested a change in polarity and hydrophobicity/hydrophilicity (Online Resource 4). The polarity and hydropathy index for all the polymorphisms are listed in Online Resource 4. The rs528780505 showed altered amino acid from isoleucine to asparagine at 362nd position, which resulted in a change in polarity from non-polar to polar and the hydropathy index from 4.5 to −3.5. There was an observed difference in the total free energy of the wild type (−33,905.453 kJ/mol) and mutant type (−34,064 kJ/mol) protein (Online Resource 5). The

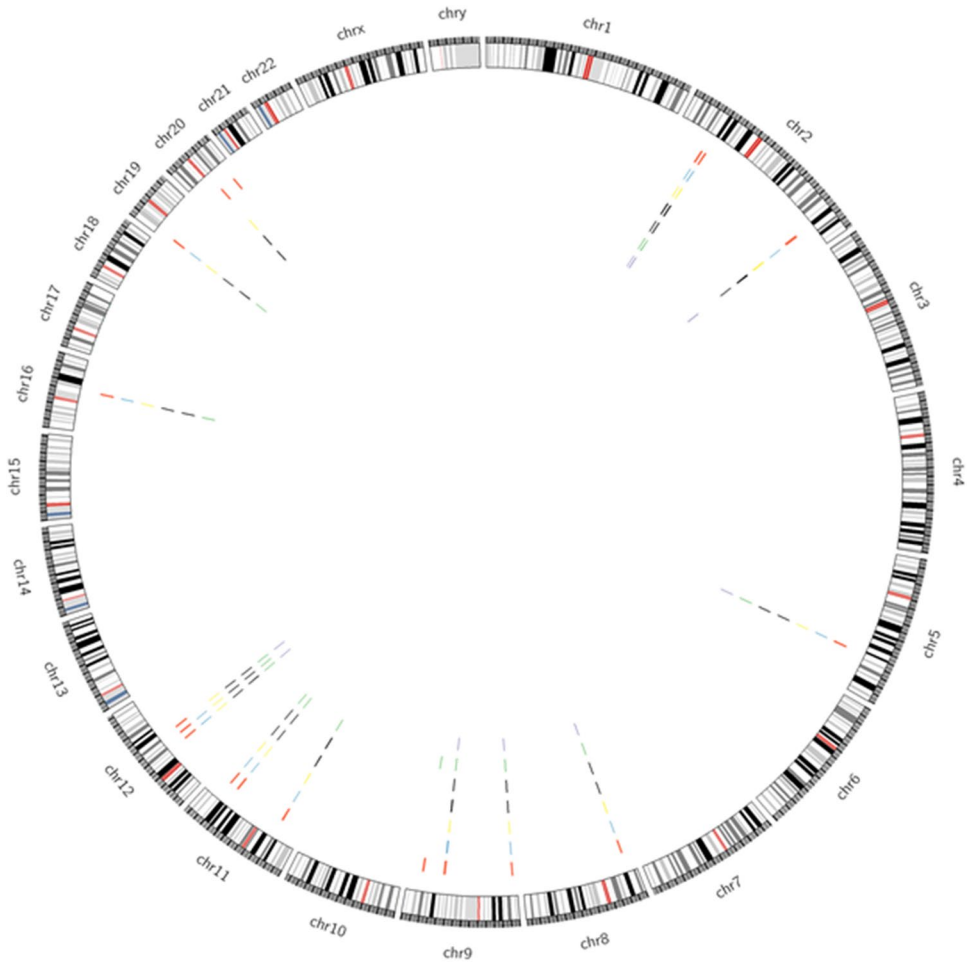

Fig. 5 Circos plot representing SNP distribution across 25 genes involved in PCOS pathogenesis showing (outer ring) all the chromosomes, 25 genes (from outer ring inwards), 5′-UTR SNPs, synonymous SNPs, missense variants, 3′-UTR SNPs, splice variants (splice region, splice donor, splice acceptor), inner most ring constitutes stop retained, stop-gained and stop-lost SNPs

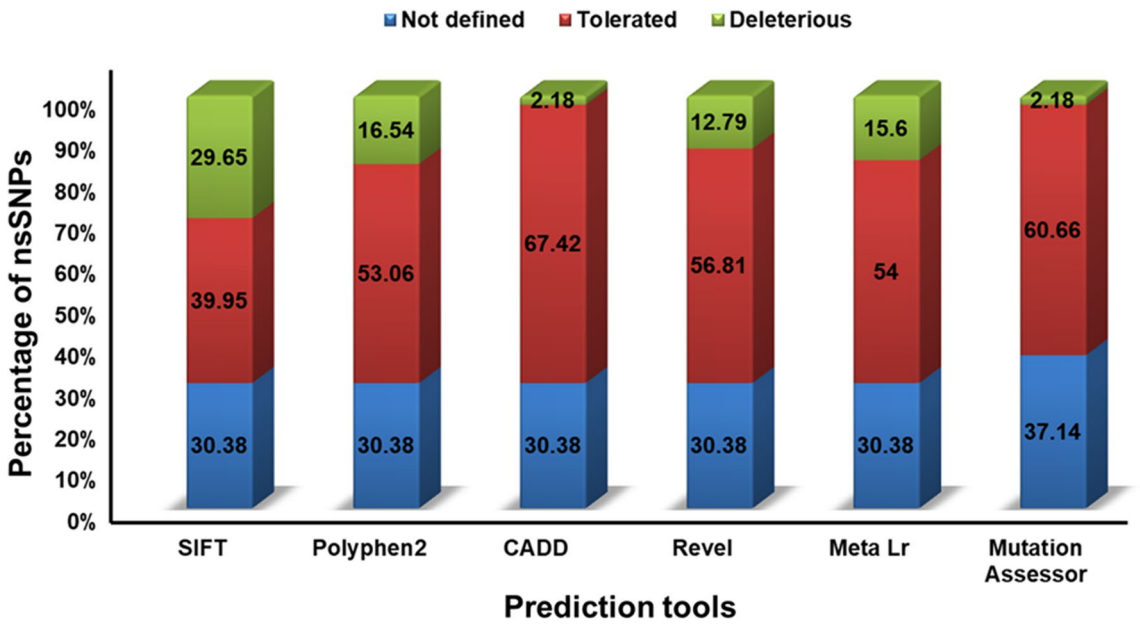

Fig. 6 Functional characterization of SNPs in PCOS genes

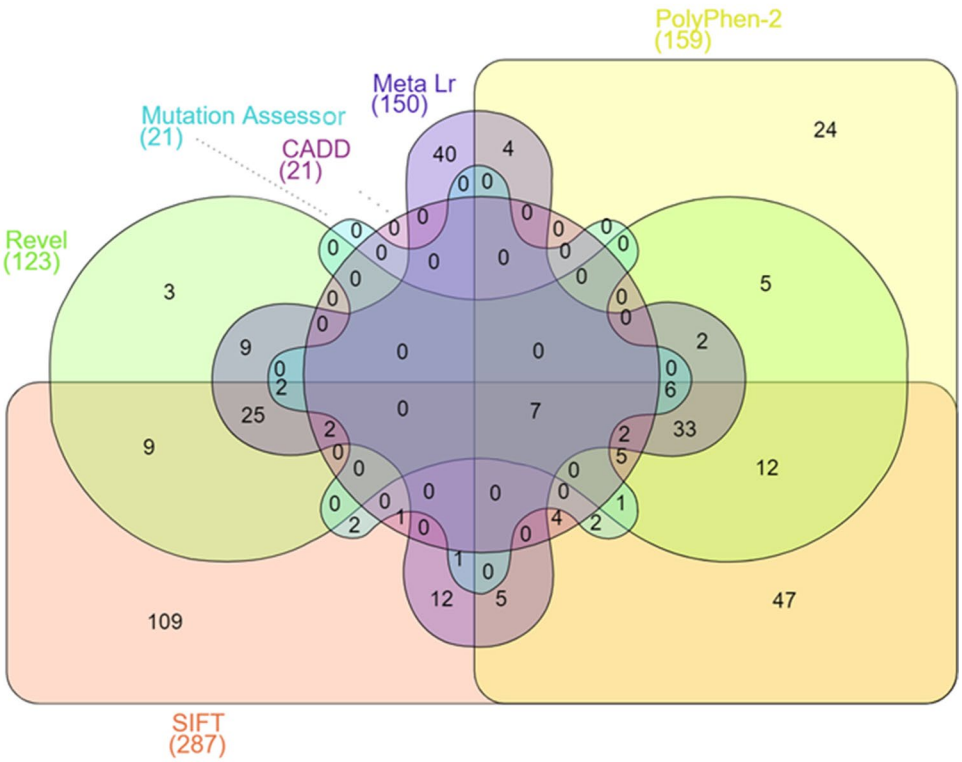

Fig. 7 Functional prediction of common non-synonymous SNPs by six pathogenicity predictions

root-mean-square deviation calculated between the wild types and mutants was 0.001 Å for ERBB4 rs528780505. The RMSD value of all the proteins are tabulated (Online Resource 5).

Prediction of functional microRNA target SNPs

In the study, we used 3 different tools (MirSNP, PolymiRTS, miRNASNP3) which concordantly showed 3 SNPs (Online Resource 6) in the microRNA target binding sites, namely, rs1042725, rs7312910 in the *HMGA2* gene, and rs242538 in the *MAPRE1* gene with the minor allele frequency (MAF) > 0.1. The table also showed whether miRNAs associated with SNPs within the target site would create or break or decrease or enhance a miRNA-mRNA binding site (Online Resource 6).

SNPs at transcription factor binding siteSsec2

Using SNP2TFBS, a total of 10 SNPs with MAF > 0.1 were identified in TFBS, out of which 9 SNPs are present in the upstream and 1 SNP in the 5′-UTR region. Among these, SNPInspector predicted that rs8191514 in the *NEIL2* generated a binding site for twenty transcription factors, and rs62579216 in the *DENND1A* gene deleted the binding site for nine transcription factors. The impact of 10 SNPs at TFBS reported whether SNPs would generate or delete the sites for the binding of transcription factors (Online Resource 7).

SNPs in enhancers

In the present study, we used 2 databases (EnhancerDB and HaploReg), which collectively reported 8 intronic SNPs in the enhancers with MAF > 0.1. Among these, rs11670022 in the *INSR* gene showed 5 altered regulatory motifs which included E2A, HEN1, Lmo2, Myf, ZEB1 followed by rs73488786 in the *INSR* gene had shown 4 altered regulatory motifs namely, AP-1, BDP1, CTCF, SMC3 and rs56394135 in the *RAD50* gene showing 4 altered regulatory motifs namely, Dbx2, Maf, Pou2f2, THAP1. The details of enhancer SNPs and their altered regulatory motifs are tabulated (Online Resource 8).

Linkage disequilibrium analysis of functional SNPs

Using LDlink, a total of 28 SNPs that may be functional were further examined to correlate with reported PCOS GWAS SNPs. Out of which 4 SNPs were in LD, namely, rs8191514 in the *NEIL2* gene is correlated with rs804279. rs242538 in the *MAPRE1* gene is correlated with rs853854. rs12237685 in the *DENND1A* gene is correlated with rs9696009 and rs2479106. rs3846732 in the *RAD50* gene is correlated with rs13164856. R^2, D′, and *p* value of the

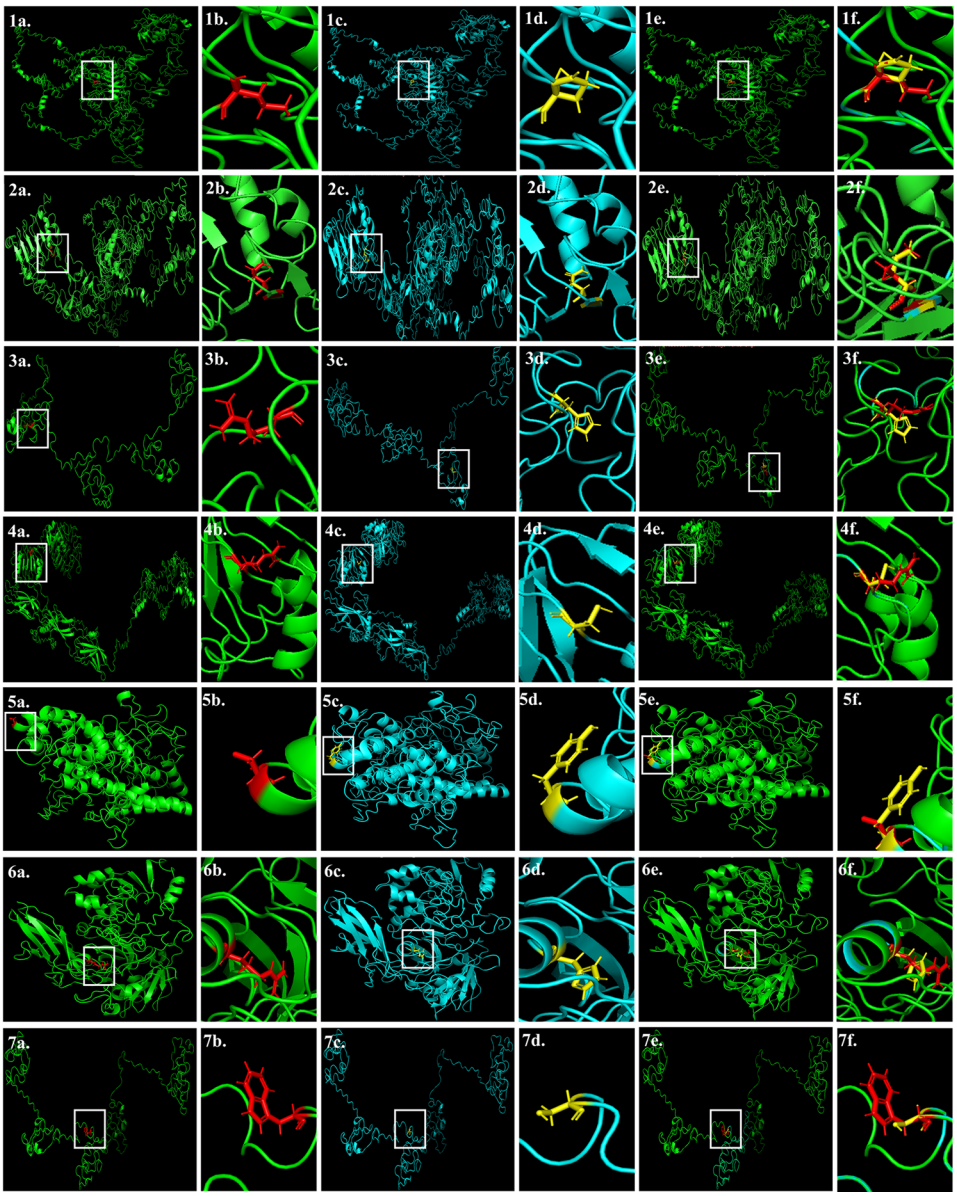

Fig. 8 Native, mutant, and superimposition of native and mutant modeled structures of the *ERBB4* (1) rs192066345 and (2) rs528780505, (3) *GATA4* rs180765750, (4) *INSR* rs79312957, (5) *LHCGR* rs121912525, (6) *SUOX* rs575660698, and (7) *YAP1* rs199505545. **a** Structure of native protein. **b** Enlarged structure of native protein (**c**) Structure of mutant protein. **d** Enlarged structure of mutant protein (**e**) Superimposed model of native and mutant protein structures. **f** Enlarged superimposed model of native and mutant protein structures

selected SNPs with reported PCOS GWAS SNPs were calculated and cataloged (Table 1).

Discussion

Exertions intended to interpret the molecular mechanisms of multifaceted diseases like PCOS are supported by high-throughput approaches to identify genetic variations resulting in the generation of large amounts of data [10]. To manage these vast amounts of data and to provide insight into PCOS development, researchers have used a variety of in silico prediction tools [23]. In the present study, after reviewing publications from the GWAS catalog, the potential causal genes at genome-wide significance were shortlisted and subsequently examined to identify and predict the deleterious SNPs and their impact on disease progression. Prediction of SNPs was made using six different tools, namely, SIFT, PolyPhen-2, CADD, Revel, Meta Lr, and Mutation assessor. The interpretation of these data should be evaluated accurately to address the significance of gene and should be verified whether the genetic variants are deleterious and impact protein structure or not [24]. Hence evaluation of these genetic variations is carefully performed with the use of different SNP prediction tools by selecting the overlapping predictions to mitigate the false-positive interpretation [10].

Our computational approach has identified 7 deleterious nsSNPs from 6 SNP prediction tools. These genetic variations reside in different genes such as *ERBB4, GATA4,*

Table 1 Linkage disequilibrium analysis showing correlation of functional SNPs with reported PCOS GWAS SNPs

Sl no	Gene	rsID	MAF	Correlation with reported PCOS GWAS SNPs	R^2	D′	*p* value
1	*NEIL2*	rs8191514	0.13	rs804279	0.4	0.97	< 0.0001
2	*MAPRE1*	rs242538	0.19	rs853854	0.69	1	< 0.0001
3	*DENND1A*	rs12237685	0.36	rs9696009	0.18	0.68	< 0.0001
				rs2479106	0.61	0.87	< 0.0001
4	*RAD50*	rs3846732	0.4	rs13164856	0.79	0.97	< 0.0001

NEIL2 nei like dna glycosylase 2, *MAPRE1* microtubule associated protein rp/eb family member 1, *DENND1A* DENN domain containing 1A, *RAD50* RAD50 double strand break repair protein

INSR, LHCGR, SUOX, and *YAP1*. So far, minimal investigations have been carried out to predict the effect of nsSNPs. Despite, few studies have been reported the role of *INSR* rs79312957, *LHCGR* rs121912525, in complex traits. An in silico study conducted by Mahmud et al. 2016 identified that mutation in *INSR* (rs79312957) caused type A insulin resistance, which is a prominent feature observed in PCOS females [25]. During adolescence, the type A insulin resistance in PCOS females shows higher insulin levels in the bloodstream which interacts with the different hormones and induce aberrations in menstruation, presence of multiple cysts in the ovaries, and other related features of the syndrome [26]. Interestingly, the mutation in LH receptor (rs121912525) has a higher chance of causing partial ovarian failure manifested by defects in ovarian folliculogenesis, anovular menstruation, luteal phase defects, imperfect feminization at adolescence, amenorrhoea and, infertility in females [27], which are again the characterized features of PCOS.

The effect of nsSNP, rs79312957 in *INSR*, can cause numerous insulin-resistant diseases. An earlier computational study by Mahmud et al., 2016 showed the structural modification between the native and mutant forms of protein INSR rs79312957, based on the value of Gibbs free energy [25]. The variation in free energy, when it deviates from native to mutant type, the variation in free energy indicates protein stability [10]. The authors also provided computational evidence for the destabilizing effect of nsSNP rs79312957 on the insulin receptor which is considered to impact protein structure and function [25]. Hence, we used a structural-based method to determine the influence of 7 deleterious nsSNPs on its protein structure. We have assessed changes in polarity, hydrophobicity/hydrophilicity, and hydropathy index in the present study. Besides, we have also calculated change in energy from native to mutant protein type and RMSD value for all the 7 nsSNPs, which might contribute strength to assess the protein function. Our study also confirms the expected effects of INSR rs79312957 by depicting the deviation of RMSD value from native to the mutant form of protein.

Research on miRNAs has shown that miRNAs binding at the 3′-UTR region silences the genes and is involved in gene regulation at a posttranscriptional level. Also, alterations in the miRNA binding sites can induce impaired binding of the miRNAs affecting its function [10]. The outcome of the GWAS has resulted in the discovery of a massive number of SNPs. Although the impact of SNPs in the non-coding site of the gene is scant, we focussed on 3′-UTR SNPs in the present study. Thus, we retrieved the SNP data of the genes responsible for PCOS pathology to decipher the miRNA sites using MirSNP, PolymiRTS, miRNASNP3 databases and further investigated whether miRNAs associated with SNPs within the target site would create or break or decrease a miRNA-mRNA binding site. In the current approach, LD analysis was performed between selected SNPs that may be functional and PCOS GWAS SNPs to examine their impact on PCOS pathogenesis. LD analysis revealed that *MAPRE1* rs242538 was correlated with the reported GWAS SNP rs853854 (*MAPRE1*) in PCOS (R^2: 0.6, D′: 1, *p* value < 0.0001).

Similarly, the effect of SNPs in TFBS and enhancers were also taken into consideration. SNPs at TFBS possibly affect gene regulation by changing the binding ability of the corresponding TF created by SNP alleles [28]. Our study collectively showed 10 SNPs in the 5′-UTR and upstream region, which controls the expression of genes involved in PCOS. Out of 10 SNPs, rs8191514 in the *NEIL2* gene generated a binding site for twenty transcription factors and was found to be in LD with the reported GWAS SNP rs804279 (*NEIL2*) in PCOS (R^2: 0.4, D′: 0.97, *p* value < 0.0001). Studies have revealed that disease or trait linked non-coding SNPs modify the functions of regulatory motifs, such as enhancers that classically control gene expression [29]. A sum of 8 SNPs in the enhancers with their altered regulatory motifs were identified. Out of which, 2 SNPs were found to be LD with the reported GWAS SNPs in PCOS namely, *DENND1A* rs12237685, *RAD50* rs3846732. Henceforth in the current study, a total of 4 SNPs that were correlated with PCOS GWAS SNPs which implies these linked SNPs would be more likely pathogenic in PCOS than functional SNPs not

so linked, thus that are discussed above should be crucially taken into account for delineating the precise pathomechanisms of PCOS.

Conclusion

In the present in silico analysis, efforts were taken to unveil the remarkable findings to report the genetic markers that regulate the expression of genes to portray the pathomechanisms of PCOS. The use of computational gene mining tactics assists primarily in identifying the causal genes and their interaction in PCOS pathway and aid in evaluating the impact of SNPs in different regions of the gene. The data constitutes a structural foundation to figure out complex molecular connections among genes involved in PCOS pathophysiology and consequently contributes to predicting disease strategies. However, when an SNP is likely linked with a trait or disease, it is commonly assumed that the SNP functions through nearby genes. Hence, it is evident that the current approach may miss some relevant genes. In addition, as we focused on genes, this study will not have identified intronic or intergenic SNPs that contribute to the pathophysiology of PCOS.

Acknowledgements This study was supported by Manipal Academy of Higher Education.

Author contributions Work and concept were initiated by PSR, SKB, PVB and KS; literature search and data interpretation were performed by NPB, SHK and ARS. The manuscript was written by NPB. PSR, SPK, SKB and KS critically reviewed the manuscript.

Funding Open access funding provided by Manipal Academy of Higher Education, Manipal. This paper received no financial assistance from any funding body.

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflict of interest.

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Unluturk U, Harmanci A, Kocaefe C, Yildiz BO (2007) The genetic basis of the polycystic ovary syndrome: a literature review including discussion of PPAR-γ. PPAR Res. https://doi.org/10.1155/2007/49109
2. Kosova G, Urbanek M (2013) Genetics of the polycystic ovary syndrome. Mol Cell Endocrinol 373:29–38. https://doi.org/10.1016/j.mce.2012.10.009
3. Baptiste CG, Battista MC, Trottier A, Baillargeon JP (2010) Insulin and hyperandrogenism in women with polycystic ovary syndrome. J Steroid Biochem Mol Biol 122:42–52. https://doi.org/10.1016/j.jsbmb.2009.12.010
4. Ko H, Teede H, Moran L (2016) Analysis of the barriers and enablers to implementing lifestyle management practices for women with PCOS in Singapore. BMC Res Notes 9:1–11. https://doi.org/10.1186/s13104-016-2107-2
5. Pereira-Eshraghi CF, Chiuzan C, Zhang Y et al (2020) Obesity and insulin resistance, not polycystic ovary syndrome, are independent predictors of bone mineral density in adolescents and young women. Horm Res Paediatr. https://doi.org/10.1159/000507079
6. Casarini L, Simoni M, Brigante G (2016) Is polycystic ovary syndrome a sexual conflict? A review. Reprod Biomed Online 32:350–361. https://doi.org/10.1016/j.rbmo.2016.01.011
7. Casarini L, Brigante G (2014) The polycystic ovary syndrome evolutionary paradox: a genome-wide association studies-based, in silico, evolutionary explanation. J Clin Endocrinol Metab 99:E2412–E2420. https://doi.org/10.1210/jc.2014-2703
8. Panda PK, Rane R, Ravichandran R et al (2016) Genetics of PCOS: A systematic bioinformatics approach to unveil the proteins responsible for PCOS. Genomics Data 8:52–60. https://doi.org/10.1016/j.gdata.2016.03.008
9. Afiqah-Aleng N, Mohamed-Hussein Z-A (2020) Computational systems analysis on polycystic ovarian syndrome (PCOS). Polycystic Ovarian Syndr. https://doi.org/10.5772/intechopen.89490
10. Vohra M, Sharma AR, Paul B et al (2018) In silico characterization of functional single nucleotide polymorphisms of folate pathway genes. Ann Hum Genet 82:186–199. https://doi.org/10.1111/ahg.12231
11. Madelaine R, Notwell JH, Skariah G et al (2018) A screen for deeply conserved non-coding GWAS SNPs uncovers a MIR-9-2 functional mutation associated to retinal vasculature defects in human. Nucleic Acids Res 46:3517–3531. https://doi.org/10.1093/nar/gky166
12. Nishizaki SS, Ng N, Dong S et al (2020) Predicting the effects of SNPs on transcription factor binding affinity. Bioinformatics 36:364–372. https://doi.org/10.1093/bioinformatics/btz612
13. Bindea G, Mlecnik B, Hackl H et al (2009) ClueGO: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25:1091–1093. https://doi.org/10.1093/bioinformatics/btp101
14. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132. https://doi.org/10.1016/0022-2836(82)90515-0
15. Yang J, Zhang Y (2015) Protein structure and function prediction using I-TASSER. Curr Protoc Bioinforma 52:5.8.1-5.8.15. https://doi.org/10.1002/0471250953.bi0508s52

16. Liu C, Zhang F, Li T et al (2012) MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. BMC Genomics. https://doi.org/10.1186/1471-2164-13-661
17. Bhattacharya A, Ziebarth JD, Cui Y (2014) PolymiRTS Database 3.0: Linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic Acids Res 42:86–91. https://doi.org/10.1093/nar/gkt1028
18. Gong J, Liu C, Liu W et al (2015) An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. Database 2015:1–8. https://doi.org/10.1093/database/bav029
19. Kumar S, Ambrosini G, Bucher P (2017) SNP2TFBS-a database of regulatory SNPs affecting predicted transcription factor binding site affinity. Nucleic Acids Res 45:D139–D144. https://doi.org/10.1093/nar/gkw1064
20. Kang R, Zhang Y, Huang Q et al (2019) EnhancerDB: a resource of transcriptional regulation in the context of enhancers. Database 2019:1–8. https://doi.org/10.1093/database/bay141
21. Ward LD, Kellis M (2012) HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 40:930–934. https://doi.org/10.1093/nar/gkr917
22. Machiela MJ, Chanock SJ (2015) LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31:3555–3557. https://doi.org/10.1093/bioinformatics/btv402
23. Wooley JC, Lin HS, National Research Council (2005) Computational modeling and simulation as enablers for biological discovery. In: Catalyzing inquiry at the interface of computing and biology. National Academies Press, US
24. Vihinen M (2012) How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. BMC Genomics 13(Suppl 4):S2. https://doi.org/10.1186/1471-2164-13-S4-S2
25. Mahmud Z, Malik SUF, Ahmed J, Azad AK (2016) Computational analysis of damaging single-nucleotide polymorphisms and their structural and functional impact on the insulin receptor. Biomed Res Int. https://doi.org/10.1155/2016/2023803
26. Musso C, Cochran E, Moran SA et al (2004) Clinical course of genetic diseases of the insulin receptor (type A and Rabson-Mendenhall syndromes): a 30 year prospective. Medicine (Baltimore) 83:209–222. https://doi.org/10.1097/01.md.0000133625.73570.54
27. Latronico AC, Anasti J, Arnhold IJP et al (1996) Brief report: testicular and ovarian resistance to luteinizing hormone caused by inactivating mutations of the luteinizing hormone-receptor gene. N Engl J Med 334:507–512. https://doi.org/10.1056/NEJM199602223340805
28. Buroker NE (2017) SNPs, transcriptional factor binding sites and disease. Biomed Genet Genomics 2:1–9. https://doi.org/10.15761/bgg.1000132
29. Kikuchi M, Hara N, Hasegawa M et al (2019) Enhancer variants associated with Alzheimer's disease affect gene expression via chromatin looping. BMC Med Genomics 12:1–16. https://doi.org/10.1186/s12920-019-0574-8
30. Shi Y, Zhao H, Shi Y et al (2012) Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. Nat Genet 44:1020–1025. https://doi.org/10.1038/ng.2384
31. Chen ZJ, Zhao H, He L et al (2011) Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. Nat Genet 43:55–59. https://doi.org/10.1038/ng.732
32. Hayes MG, Urbanek M, Ehrmann DA et al (2015) Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. Nat Commun 6:1–12. https://doi.org/10.1038/ncomms8502
33. Day F, Karaderi T, Jones MR et al (2018) Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. PLoS Genet 14:1–20. https://doi.org/10.1371/journal.pgen.1007813
34. Day FR, Hinds DA, Tung JY et al (2015) Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. Nat Commun 6:1–7. https://doi.org/10.1038/ncomms9464