scientific reports

OPEN

Clinical-exome sequencing unveils the genetic landscape of polycystic ovarian syndrome (PCOS) focusing on lean and obese phenotypes: implications for cost-effective diagnosis and personalized treatment

Shrinjana Dhar & Pritha Bhattacharjee

Polycystic ovarian syndrome (PCOS) is one of the most common endocrinopathies among reproductive women worldwide, contributing greatly on the incidence of female infertility and gynecological cancers. It is a complex health condition combining of multiple symptoms like androgen excess, uncontrolled weight gain, alopecia, hirsutism, etc. Conventionally PCOS was associated with obesity while it is often found among lean women nowadays, making the disease more critical to diagnose as well treatment. The disorder has an impact on several signal transduction pathways, including steroidogenesis, steroid hormone activity, gonadotrophin regulation, insulin secretion, energy balance, and chronic inflammation. Understanding the aetiology and pathophysiology of PCOS is difficult due to its multiple causes, which include environmental factors, intricate genetic predisposition, and epigenetic modifications. Despite research supporting the role of familial aggregations in PCOS outcomes, the inheritance pattern remains unknown. Henceforth, to reduce the burden of PCOS, it is inevitably important to diagnose at early ages as well as intervene through personalized medicine. With this brief background, it was imperative to elucidate the genetic architecture of PCOS considering BMI as an controlling factor. This study aims to investigate the genetic basis behind obesity-mediated PCOS, focusing on both obese and lean individuals. It uses a comprehensive bioinformatics methodology to depict pathways and functionality enrichment, allowing for cost-effective risk prediction and management. In the present research, the representative study participants (N = 2) were chosen from a cross-sectional epidemiological survey, based on their anthropometric parameters and confirmation of PCOS. Upon voluntary participation and written consent, biological fluids (whole blood and buccal swab) were taken from where DNA was extracted. The clinical-exome sequencing was performed by the Next-generation Illumina platform using the Twist Human Comprehensive Exome Kit. A comprehensive bioinformatics methodology was employed to identify the most important, unique, and common genes. A total of 26,550 variants were identified in clinically important exomes from two samples, with 5170 common and 2232 and 2322 unique among PCOS lean and obese phenotypes, respectively. Only 262 and 94 variants were PCOS-specific in lean and obese PCOS. Three filters were applied to shortlist the most potent variants, with 4 unique variants in lean PCOS, 2 unique variants in obese PCOS, and 5 common variants in both. The study found that leptin signalling impairment and insulin resistance, as well as mutations in *CYP1A1*, *CYP19A1*, *ESR1*, *AR*, *AMH*, *AdipoR1*, *NAMPT*, *NPY*, *PTEN*, *EGFR*, and *Akt*, all play significant roles in PCOS in the studied group. Young women in West Bengal, India, are more likely to have co-occurring PCOS, which includes estrogen resistance, leptin receptor insufficiency, folate deficiency, T2DM, and acanthosis nigricans, with obesity being a common phenotypic expression.

sequencing (CES), Comorbidity analysis, Mutation-phenotype interactions, Molecular screening

Environmental Epigenomics Laboratory, Department of Environmental Science, University of Calcutta, 35, Ballygunge Circular Road, Kolkata, West Bengal 700019, India. ✉email: pbenvs@caluniv.ac.in; 777.pritha@gmail.com

Polycystic ovarian syndrome (PCOS), one of the most common heterogeneous, severe endocrinopathies that affect 15–20% of reproductive women globally, has been linked to infertility and even some gynaecological cancers[1–3]. The condition is classified as a syndrome rather than a disease due to its diverse phenotypic symptoms, such as ultrasonographic polycystic ovarian morphology (PCOM), clinical or biochemical hyperandrogenism, hirsutism, acne, acanthosis nigricans, obesity, infertility, and hair loss[4,5]. Understanding the aetiology and pathophysiology of PCOS has been a difficult task because of its plentiful causes, including environmental variables, genetic predisposition, and epigenetic changes[2,3]. On the other hand, despite having sufficient evidence to confirm the importance of familial aggregations in PCOS outcomes[6–10], the inheritance pattern remains unclear. PCOS affects multiple signal transduction pathways, including steroidogenesis, steroid hormone activity, gonadotrophin control, insulin secretion, energy balance, and chronic inflammation[2–4]. According to the database[11,12], 533 potential genes and 145 SNPs have been linked to PCOS, and countless studies are being conducted throughout the world to comprehend the molecular signals and their intricate underlying mechanisms.

It is terrible that women are still not completely aware of their reproductive health. In our previous survey, we discovered that the majority of the population is unaware of their menstrual health while experiencing a few symptoms such as irregular menstruation, dysmenorrhea, and so on[13]. The condition stays the same in both developing and developed nations, and this could be related to traditional and cultural views[14–16]. Over the last 2 decades, the global burden of PCOS-related infertility has grown dramatically[17]. A study showed the same observation as that of Dhar et al.[13], at 30 years of age and those with a prior diagnosis of PCOS were 4 times more likely to have reported difficulties becoming pregnant than those undiagnosed and regularly sought medical treatment[18]. Contradictory to this, an earlier study reported that POM was present among 14% of healthy women and it was more prevalent among women in the age group < 35 years[19]. In addition, irregular menstrual bleeding is less trustworthy until it is prolonged, but moderate hair growth and chronic anovulation are normal in late puberty and early adolescence, which can lead to a delayed diagnosis[20] and failure in PCOS management. So, accumulating shreds of evidence reinforces the importance of detection of PCOS early in youth or before puberty to ensure timely diagnosis and adequate health care (Fig. 1).

In the era of medical advent, the introduction of next-generation sequencing (NGS) has caused a paradigm shift in genomics research, providing unequalled capabilities for analysing DNA and RNA molecules in a high-throughput manner. It provides insights into genome structure, genetic variants, gene expression levels, and epigenetic alterations, broadening research into uncommon genetic illnesses, cancer genomics, and precision medicine[21]. Clinical exome sequencing (CES) is a cutting-edge molecular diagnostic technology that rapidly discovers disease-causing genetic abnormalities in any human gene and is gaining popularity in clinical settings. Its high diagnostic yield makes it cost-effective and suitable for routine diagnosis in patients with a variety of phenotypes[22]. Since the last few years, a plethora of studies have been conducted to elucidate the genetic architecture of PCOS, with the majority of the studies using whole-exome sequencing methodology[8,9,23–26].

The current advances in genomics research of PCOS have primarily focused on family cohorts to better understand the disease's inheritance pattern, whereas the genetic mechanism behind obesity-mediated PCOS remains unknown. With this background, the present study has attempted to explore the clinically relevant mutational landscape of PCOS emphasizing the obese and lean category subject. In addition to this, this study has tried to elaborately represent the pathways enrichment, and functionality enrichment using a holistic bioinformatics approach so that further risk can be predicted and managed cost-effectively.

Materials and methods
Selection of study subjects
This study was conducted in urban and peri-urban areas of West Bengal, India, based on the epidemiological survey[13]. For subject selection, Rotterdam criteria were employed for confirmation of PCOS and two representative probands (N = 2) were included in this study. The reason behind the selection of these subjects due to their opposite anthropometric features.

To start with, all demographic and lifestyle details have been collected. Pathological screening like blood parameters and pelvic ultrasonography have been done along with other phenotypic assessments such as hirsutism, acne, alopecia,acanthosis nigricans, skin tags/moles. Pedigree data was also taken to understand familial aggregation. Subjects provided written informed consent for the genetic research studies, which were performed following the study protocol approved by the Institutional ethical committee of the University of Calcutta (***Ref. No. 07/ET/20-21/1777***) and in concordance with the Helsinki Declaration of 1975, as revised in 2008.

Biological sample collection and preparation
3 mL of whole blood was collected from each subject and genomic DNA was extracted from peripheral blood mononuclear cells (PBMCs) through density gradient centrifugation using HiSep™ (Himedia, India, LS001), followed by TriZol (Invitrogen™, United States) method. The concentration and purity of extracted DNA were assessed using a NanoDrop instrument (NanoDrop Technologies, Wilmington, DE, USA).
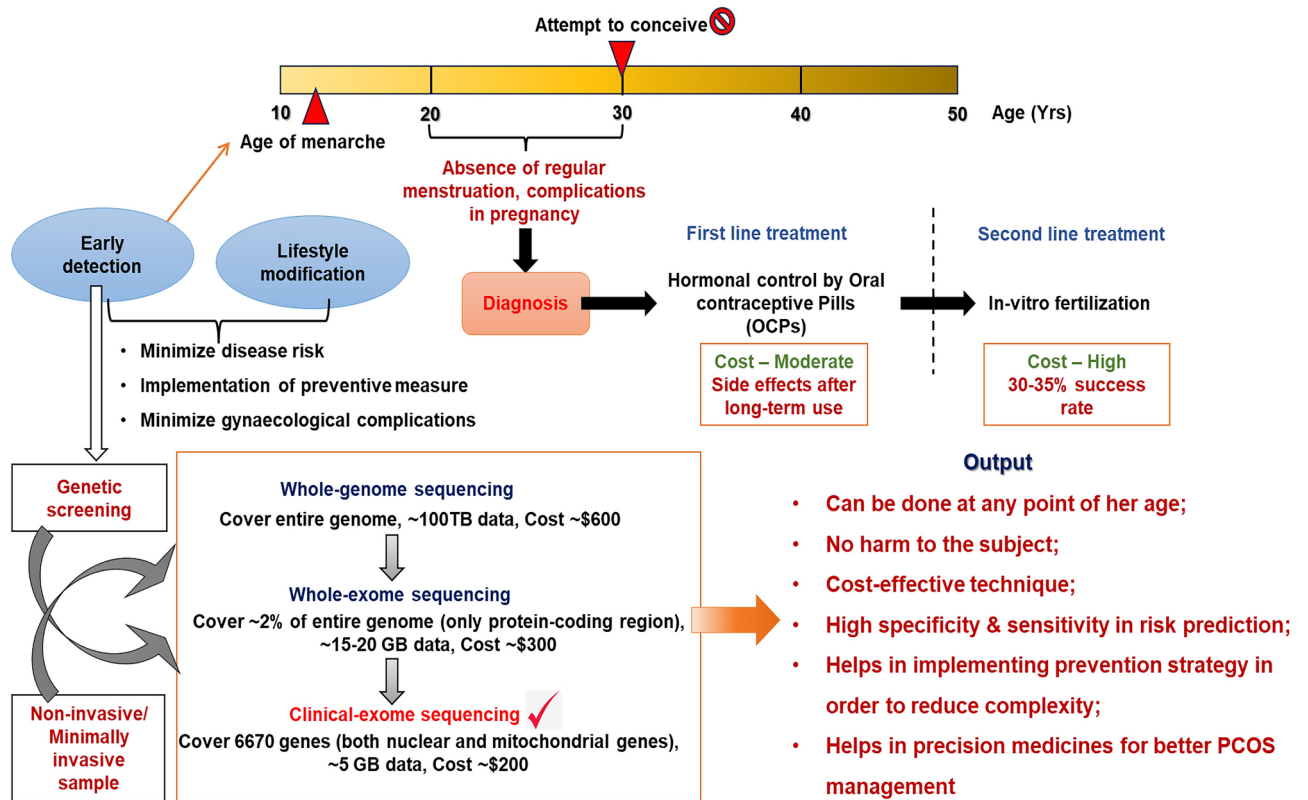
Fig. 1. Gaps & scope of the present research hypothesis.

Clinical exome sequencing

For quality control, genomic DNA was further quantified using a Qubit assay (Invitrogen, USA). Exome regions were captured using the Twist Human Comprehensive Exome Kit (Twist Bioscience, San Francisco, CA, USA). Paired-end sequencing (2×150 bp) was performed using an Illumina HiSeqX / NovaSeq next-generation sequencing platform (San Diego, CA, USA), yielding a minimum read depth of 80-100X. The sequences were aligned with the human reference assembly (hg38). The standard pipeline using the GATK tools for exome sequencing data analysis which has been shown to yield results of higher quality was used for analysis.

Variants screening using bioinformatics approach

Ultra-filtration was done in different approaches using VEP, Variation origin, OMIM report and Variation class. The functional enrichment analysis in terms of both pathways and ontology enrichment was done using the EnrichR bioinformatic tool[27–29] and visualization was done using Appyters (https://appyters.maayanlab.cloud/Enrichment_Analysis_Visualizer) tool. For pathway enrichment analysis, 6 websites viz. Reactome 2022, KEGG 2021, Elsevier Pathway Collection, WikiPathway 2023, Biocarta 2016 and Panther 2016 has been used. For ontology analysis, Biological process, Cellular components and Molecular functions has been taken into consideration. Protein–protein interaction (PPI) network was created using another web-based tool viz. NetworkAnalyst v3.0 (https://www.networkanalyst.ca/NetworkAnalyst/).

Results

Description of the study participants

This study encompasses two Indian Hindu Bengali families with 2 PCOS probands, one PCOS lean phenotype and another PCOS obese phenotype. Demography and anthropometric features have been defined in Table 1.

Proband-1 (PCOS lean phenotype): At the time of this study, the age of the subject was 28 years and the age of onset was 22 years. The pelvic ultrasonography clearly showed the bilateral polycystic ovaries; the right ovary was enlarged (16.8 cc vol.) and the left ovary was normal in size (11.6 cc vol.) with peripheral pea-sized multiple follicles. The level of Thyroid stimulating hormone (TSH), Prolactin, Leutinizing hormone (LH) Follicle stimulating hormone (FSH), Anti-mullerian hormone (AMH) and fasting insulin were 1.3 µIU/mL, 8.9 ng/mL, 1.9 mIU/mL, 2.7 mIU/mL, 7.29 ng/mL and 9.3 U/mL respectively. Other phenotypic characteristics including acne, acanthosis nigricans, and hirsutisms were not seen, but a slight alopecia was observed. Therefore, the subject belongs to Phenotype-D, i.e., Non-hyperandrogenic PCOS category. The family pedigree is depicted in Fig. 2a.

Proband-2 (PCOS obese phenotype): The age of the subject (at the time of this study) was 29 years and the age of onset was 16 years. The hormonal profile and fasting glucose data were 4.76 µIU/mL, 26.70 ng/mL, 10.46 mIU/mL, 4.74 mIU/mL, 2.85 ng/mL and 110.5 mg/dL respectively. The pelvic features showed bilateral

Features	PCOS-Lean	PCOS-Obese
Age (years)	28	29
Residential status	Urban	Peri-urban
Economic status	Higher-middle class	
Literacy	Yes	
Food habit	Non-vegetarian	
Any medications	Yes (OCP + Metformin + Folate supplement)	Yes (only OCP)
Physical activity	High	Low to moderate
Family history	No	Yes
Blood pressure	136/89	118/78
Pulse rate	94	101
Body weight (kg)	45	79.4
Total fat content (%)	17.7	37.6
Visceral fat content (%)	0.5	9
BMI (kg/m^2)	19.5	29.5
Resting metabolism	919	1532
Body age (years)	18	46
Subcutaneous mass content (%)		
Whole body	14	34.5
Trunk	9.7	30.4
Arm	28.6	52.2
Leg	24.4	50.2
Skeletal muscle mass content (%)		
Whole body	29.4	23.4
Trunk	25.8	17.3
Arm	37.1	20.5
Leg	39.6	37.1

Table 1. Demographic and anthropometric details of the Subjects participate in the present study.

polycystic ovaries with both enlarged ovaries; the right ovary was 13 cc vol. and the left ovary was 14 cc vol. with peripheral pea-sized multiple follicles. Subject exhibited extreme hirsutism and slight alopecia whereas acne and acanthosis nigricans were not observed. Therefore, the subject belongs to Phenotype-A, i.e., Full-blown PCOS category. The family history is depicted in Fig. 2b.

Variants summary

A total of 26,550 variants were identified in the clinically important exomes from two samples, among which 93.56% were SNPs and the rest were InDels. After filtering and eliminating the synonymous variations, we identified 92.68% missense variants and the rest contained nonsense, startloss, stoploss, frameshift-InDels and inframe-InDels.

After passing through variant filtering, 5170 variations were found to be common in all samples whereas 2232 and 2322 variations were unique among PCOS lean phenotype and PCOS obese phenotype respectively. Among 5170 variants, only 262 variants from 131 genes were PCOS-specific, according to the PCOSKb database (https://pcoskb.bicnirrh.res.in/index.php). Similarly, 120 variants from 70 genes were PCOS-specific in lean PCOS and 94 variants from 70 genes in obese PCOS.

Classification of variants and functional enrichment analysis

Figure 3 depicts the distribution of various types of mutations. To increase the chance of detecting biological processes related to the above genetic variants, all variants, regardless of the expected effect, were included in functional and pathway enrichment analysis. Functional enrichment analysis on 131 genes to which the above 262 variants (Supplementary Fig. 1 and 2) were assigned showed predominant biological processes like *positive regulation of protein phosphorylation*, *cytokine production*, *protein kinase B signalling*, *regulation of MAPK cascade*, etc. which involved in molecular functions like *steroid hydroxylase activity*, *heme binding*, *IGF-binding*, *transcription regulatory region nucleic acid binding*, *Wnt receptor activity*, etc. However, in the lean-PCOS subject, analysis of 70 genes (Supplementary Fig. 3) showed the most important biological processes like *cellular response to peptide hormone stimulus*, *response to insulin*, *regulation of protein phosphorylation*, *regulation of glucose import*, etc. thereby participating in molecular functions such as *transmembrane receptor protein tyrosine kinase activity*, *sequence-specific dsDNA binding*, *steroid hydroxylase activity*, *ATP binding*, *transcription cis-regulatory region binding*, etc. (Fig. 4). Similarly, in obese-PCOS subject, analysis of 70 genes (Supplementary Fig. 4) showed important biological processes like *regulation of glucose import*, *negative regulation of cell differentiation*, *cellular response to insulin stimulus*, *regulation of fat cell differentiation*, etc. thereby participating in molecular
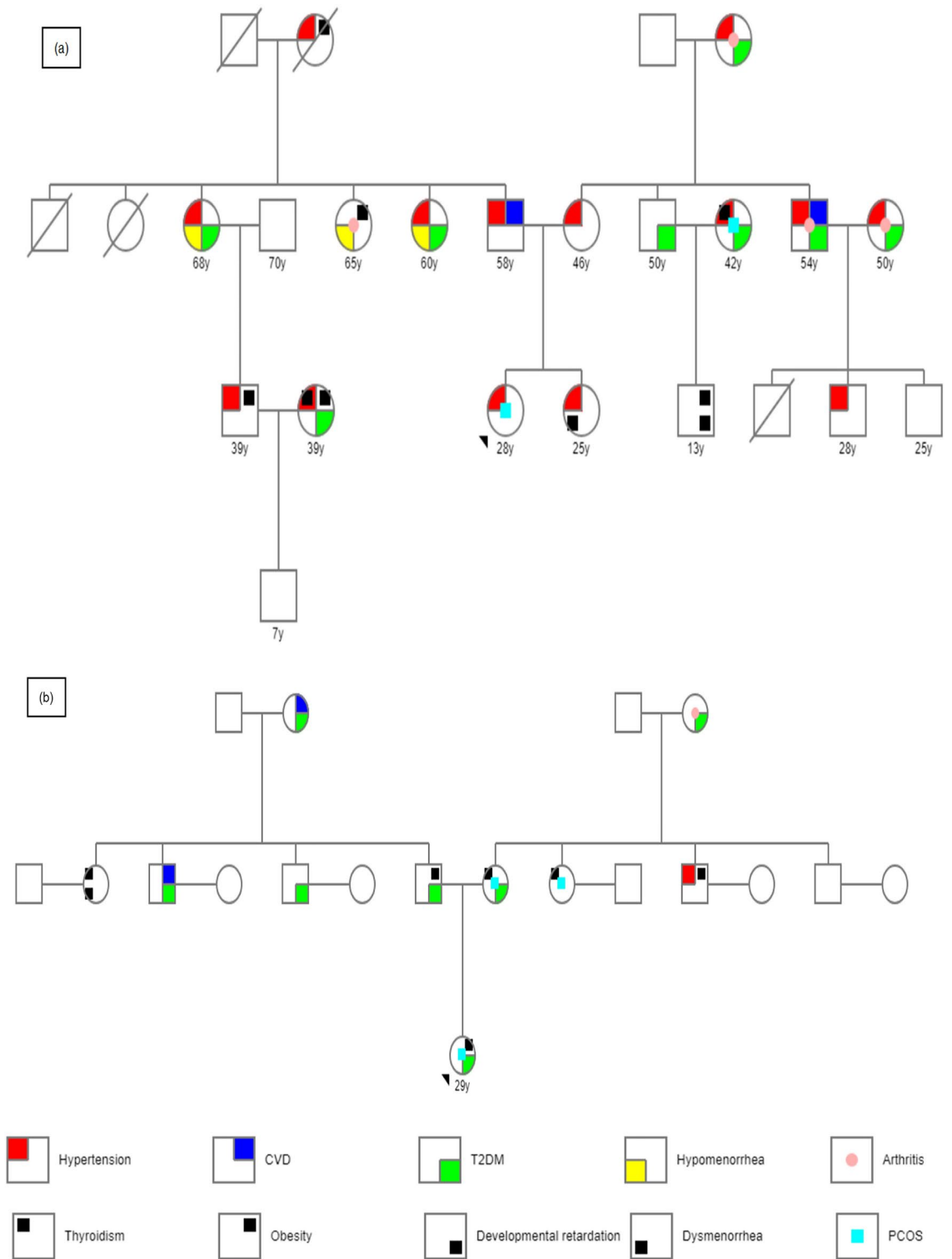
Fig. 2. Familial aggregation chart. (**a**) Lean PCOS subject, (**b**) Obese PCOS subject.

functions such as *LDL particle binding, phosphatidylinositol phospholipase activity, IGF binding, Ca²⁺ binding, protein homodimerization activity*, etc. (Fig. 5).

The pathway enrichment analysis of 131 common genes was depicted in Supplementary Fig. 5 and 6. In lean-PCOS (Supplementary Fig. 7), *HIF-1 signalling, PI3K-Akt signalling, AMPK signalling, complement & coagulation cascading*, etc. play the major role in PCOS manifestation (Fig. 6) whereas in obese-PCOS (Supplementary Fig. 8), *longevity regulating pathway, pathways in cancer, FoxO signalling, GPCR downstream signalling* plays a crucial role (Fig. 7).
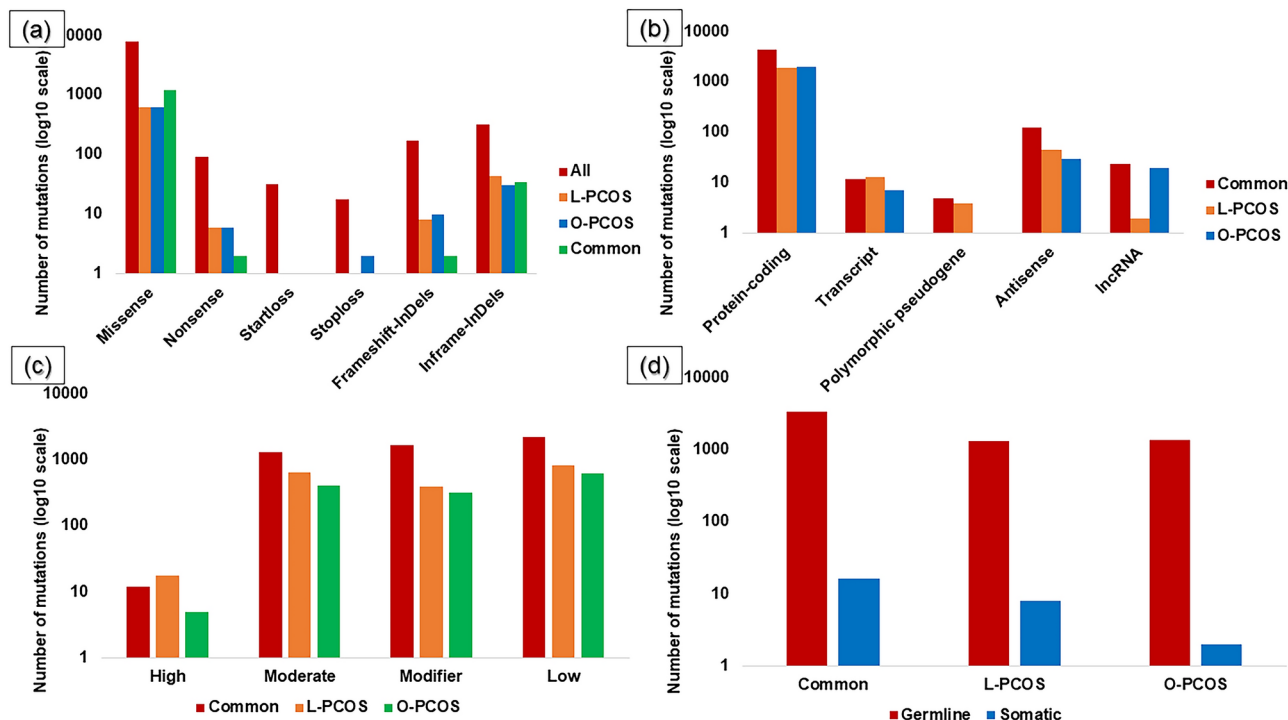
Fig. 3. Distribution of different types of mutations found in the study participants.

The PPI network for both subjects has been demonstrated in Fig. 8. Genes with higher scores in terms of node degree and betweenness were considered 'hub' genes. The hub genes include *EGFR*, *NFKB1*, *AR*, *NCOR1*, *ERBB2*, *PIK3R1*, *INSR*, *VEGFA*, *RUNX2* and *AKT3* in the case of PCOS subject with lean phenotype while *AKT1*, *YAP1*, *PTEN*, *PIK3CA*, *BRCA1*, *TLR2*, *INSR* and *TLR2* in the case of PCOS subject with obese phenotype.

Variants prioritization

To shortlist the most potent clinically relevant variants, 3 filters have been applied to the clinical-exome sequencing data, i.e. only germline mutations with 'High' VEP and OMIM ID have been selected. Therefore, 4 unique variants were found in the lean-PCOS subject, 2 unique variants were found in the obese-PCOS subject and 5 common variants were found in both (Table 2). Among them, *VDR* (rs2228570), *CYP21A2* (rs7755898), *KISS1* (rs71745629) and *XDH* (rs773456900) were already found as PCOS-associated genes.

To find out candidate mutations in mitochondrial DNA, 7 unique variants were found in the lean-PCOS subject, 3 unique variants were found in the obese-PCOS subject and 5 common variants were found in both (Table 3).

Comment
Principle findings

PCOS has become a global concern due to its significant role in female infertility, as well as ovarian and/or endometrial cancer. It is one of the most complex transgenerationally inherited endocrine-related metabolic disorders and is also difficult to detect at an early stage of life due to its intricate phenotypic nature. Hitherto, a plethora of studies have already revealed the genetic association of PCOS and the familial aggregation of PCOS, but the genetic architecture of the disease is still not clear. The present study highlights the important genetic variations in PCOS in the Bengali population from West Bengal, India and attempts to understand the molecular crosstalk of PCOS while stratified by BMI. The results may imply the genetic difference between a lean PCOS group and an obese PCOS group, which in turn, helps in precision medicine invention.

Results in the context of what is known

Substantial reports have been found the increasing trends of PCOS prevalence worldwide. In 2016, Bozdag et al. revealed that the PCOS prevalence according to NIH, Rotterdam and AE-PCOS Society was 6%, 10% and 10% respectively[27] while these proportions have become 5.5%, 11% and 7.1% respectively in 2024[28]. Overall, global prevalence of hirsutism, hyper-androgenism, PCOM and oligo-anovulation were 13%, 11%, 28% and 15% respectively[27]. Nonetheless, studies have also shown that the proportion of PCOS-A, PCOS-B, PCOS-C and PCOS-D phenotype were 44.8%, 14.9%, 16.2% and 19.5% respectively in USA and Europe[29] whereas that of 23.9%, 46.3%, 21.6% and 8.2% respectively in Middle-east countries[30] and PCOS-D was more prevalent in East-Asian countries[31]. In the last few years, studies have shown that the prevalence of PCOS phenotype-A has been decreased from 67.7% to 22.72% while that of PCOS phenotype-D has been sharply increased from 3.6% to 49.53% in India[32,33]. However, Dadachanji and his group reported that prevalence of PCOS phenotype-D was
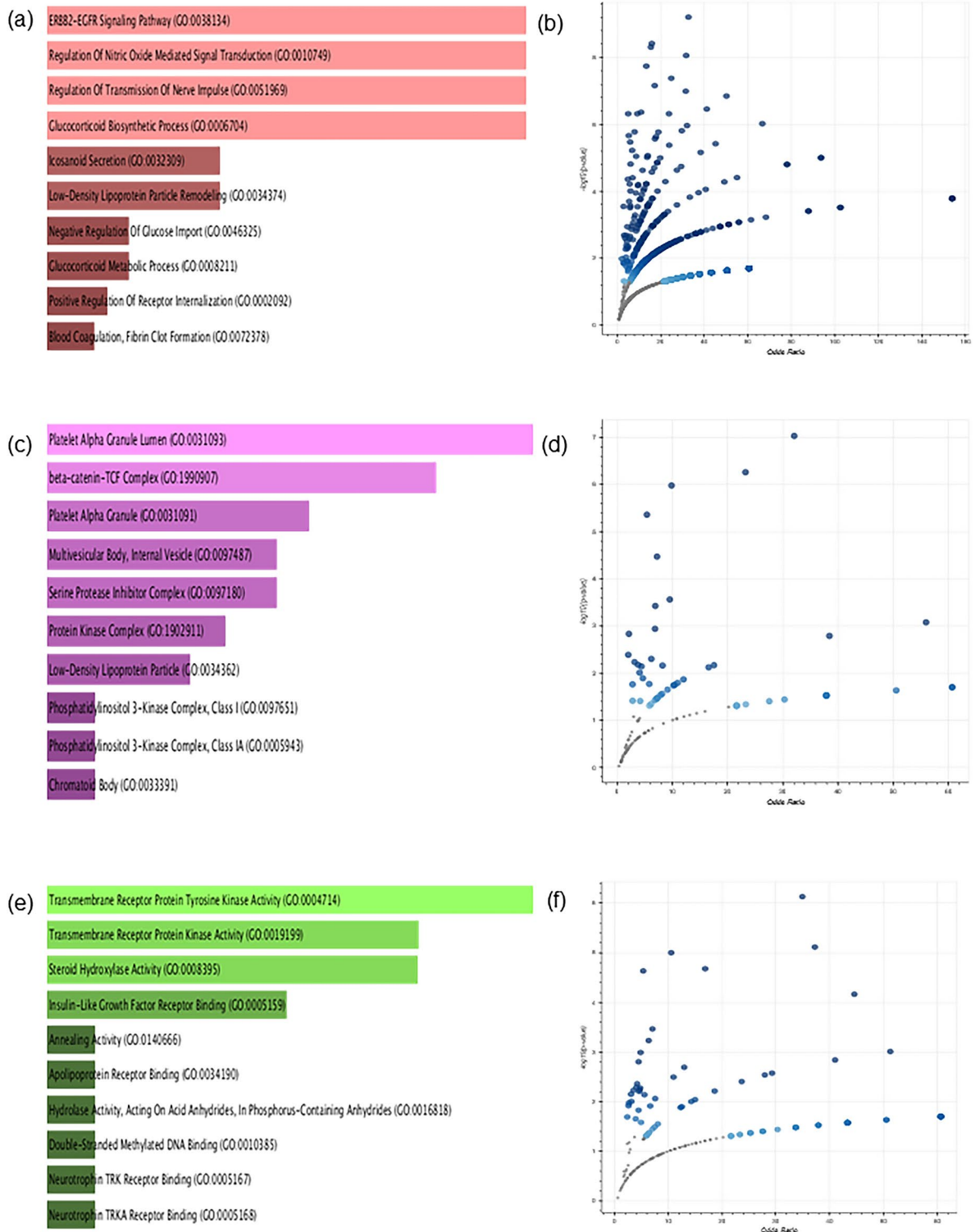
Fig. 4. Functional enrichment analysis in Lean PCOS. (**a,c,e**) Bar charts of top 10 terms from the GO-2023 database gene set library, based on a combined score of p-value and q-value; (**b,d,f**) Volcano plots of terms from the GO-2023 database gene set; each point represents a single term, plotted by the corresponding odds ratio in X-axis and -log$_{10}$ p-value in Y-axis from the enrichment results of the input query gene set; larger and darker-coloured the point, the more significantly enriched the input gene set for the term.
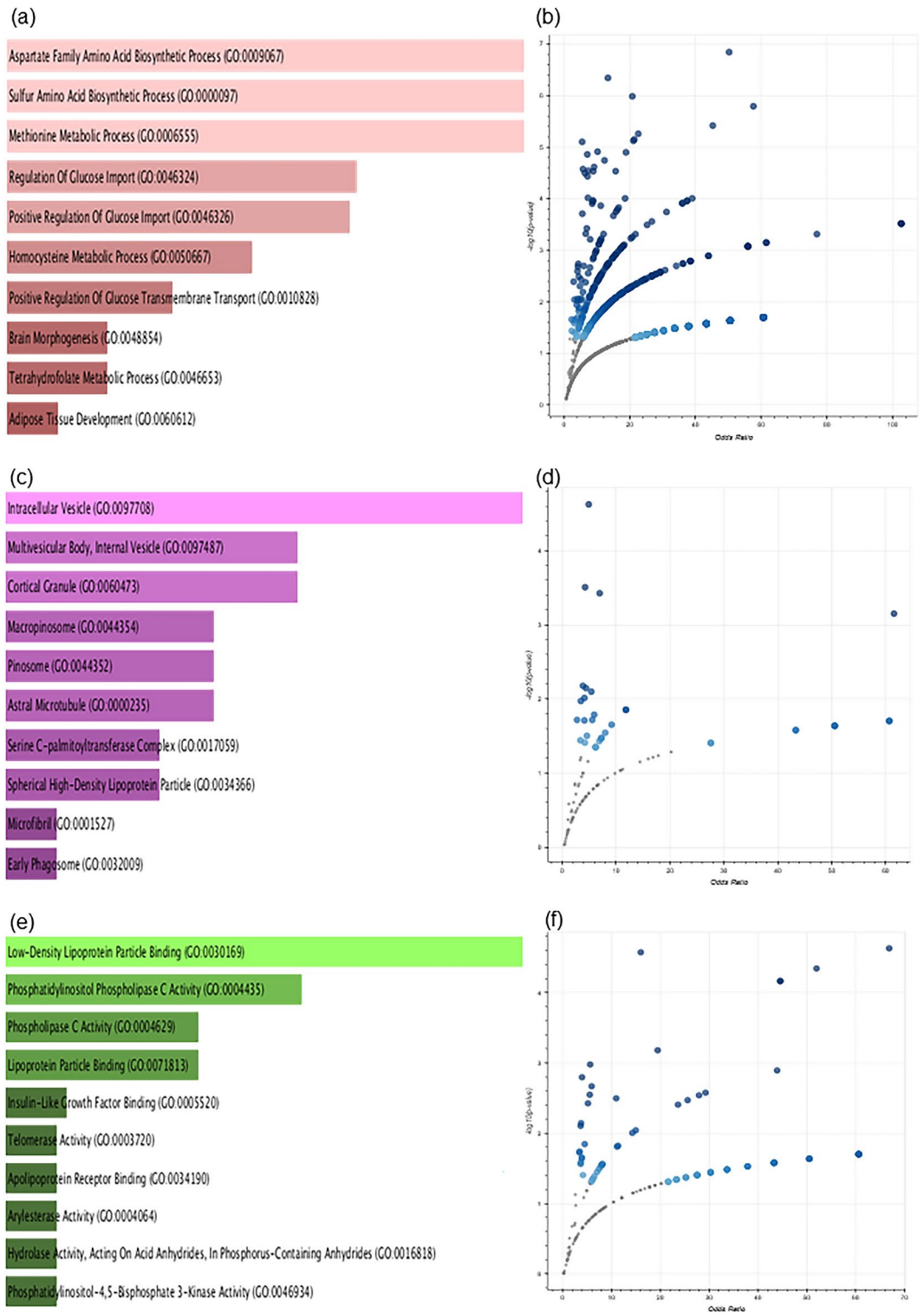
Fig. 5. Functional enrichment analysis in Obese PCOS. (**a,c,e**) Bar charts of top 10 terms from the GO-2023 database gene set library, based on a combined score of p-value and q-value; (**b,d,f**) Volcano plots of terms from the GO-2023 database gene set; each point represents a single term, plotted by the corresponding odds ratio in X-axis and -$\log_{10}$ p-value in Y-axis from the enrichment results of the input query gene set; larger and darker-coloured the point, the more significantly enriched the input gene set for the term.
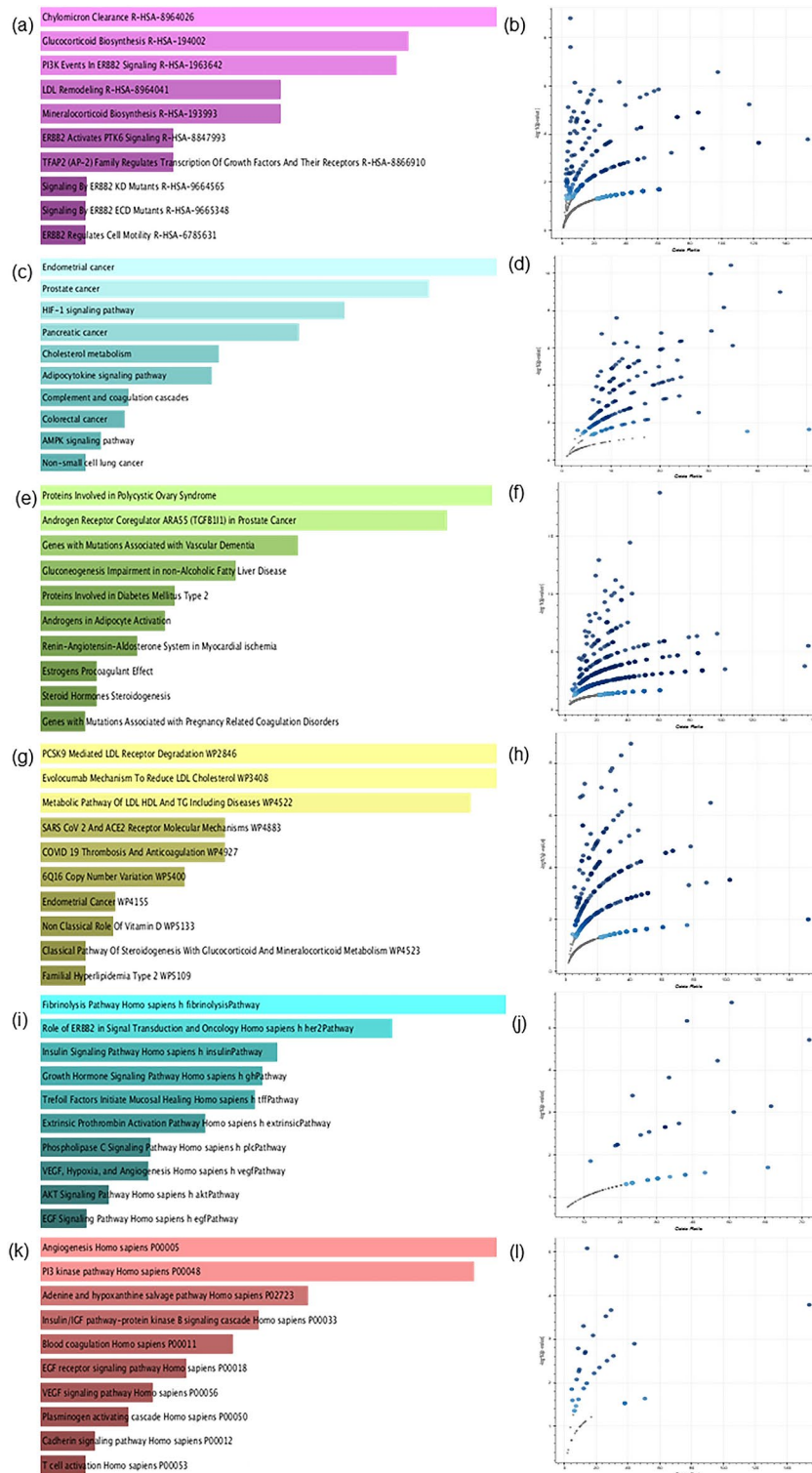
Fig. 6. Pathway enrichment analysis in Lean PCOS. (**a,c,e,g,i,k**) Bar charts of top 10 terms from pathway analysis database gene set library, based on a combined score of p-value and q-value; (**b,d,f,h,j,l**) Volcano plots of terms from the pathway analysis database gene set; each point represents a single term, plotted by the corresponding odds ratio in X-axis and -log$_{10}$ p-value in Y-axis from the enrichment results of the input query gene set; larger and darker-coloured the point, the more significantly enriched the input gene set for the term.
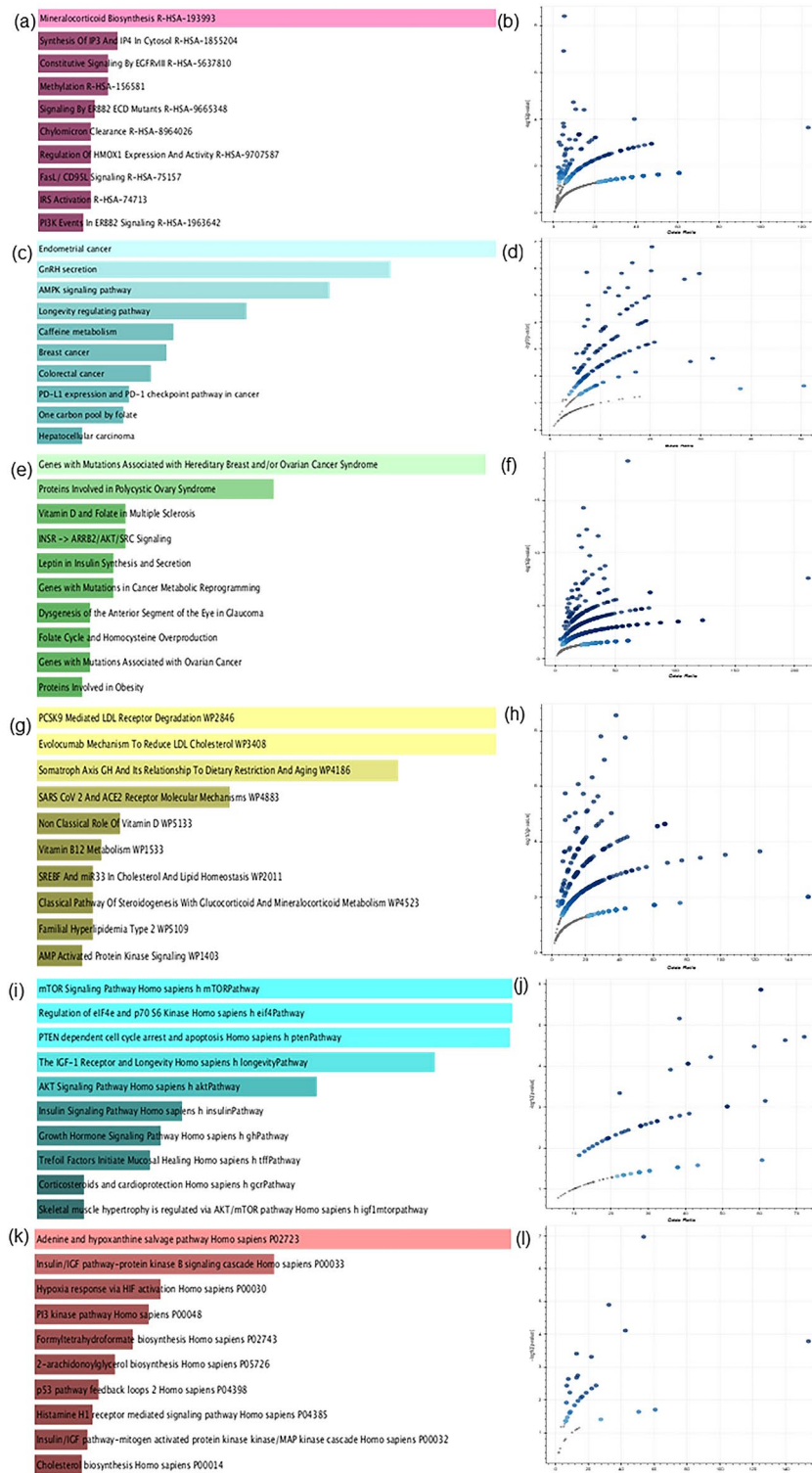
Fig. 7. Pathway enrichment analysis in Obese PCOS. (**a,c,e,g,i,k**) Bar charts of top 10 terms from pathway analysis database gene set library, based on a combined score of p-value and q-value; (**b,d,f,h,j,l**) Volcano plots of terms from the pathway analysis database gene set; each point represents a single term, plotted by the corresponding odds ratio in X-axis and -log$_{10}$ p-value in Y-axis from the enrichment results of the input query gene set; larger and darker-coloured the point, the more significantly enriched the input gene set for the term.
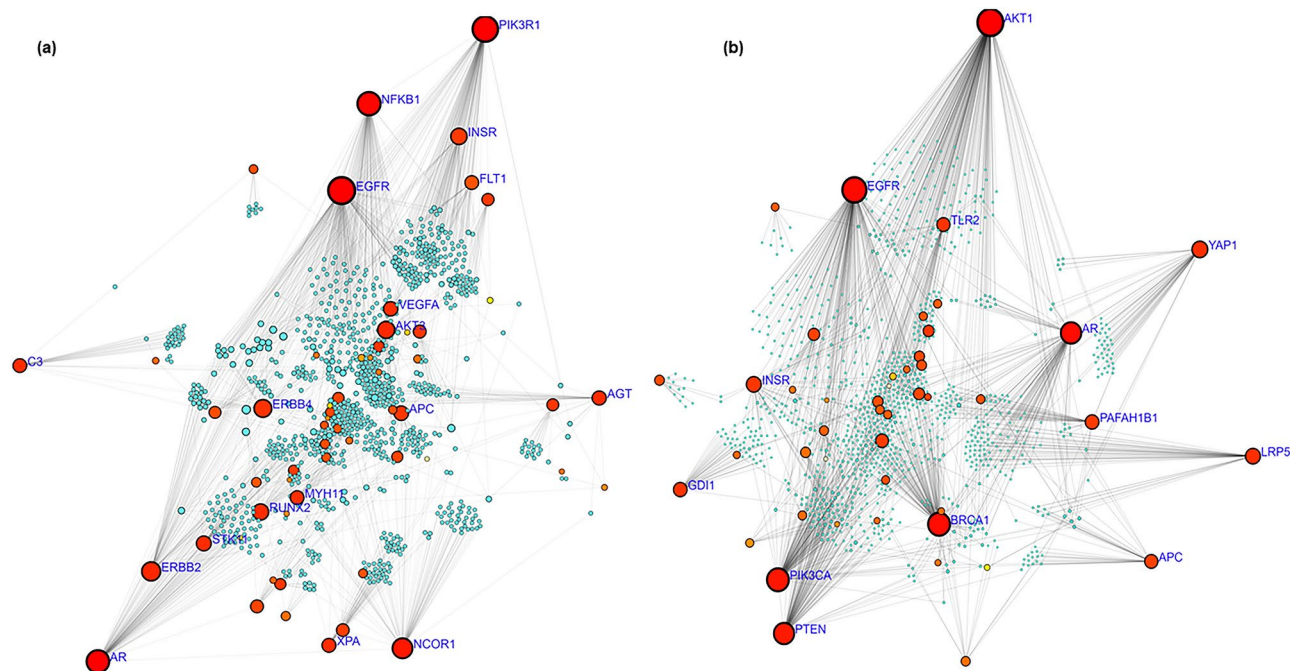
Fig. 8. Protein–protein interaction network. (**a**) Lean PCOS, (**b**) Obese PCOS.

Subject	Gene (rsID)	Variation class	Inheritance	Clinical significance	Zygosity	Associated phenotypes	1000G	gnomAD SAS
Both	*TMEM216* (rs10897158)	Intronic splice acceptor	Autosomal recessive	Benign	Homo.[1]	Joubert syndrome, Meckel syndrome	0.719649	0.819658
Both	***VDR* (rs2228570)**	**Start-loss**	**Autosomal recessive**	**Likely pathogenic**	**Hetero.[2]**	**Vitamin-D deficiency with or without alopecia**	**0.671526**	**0.738688**
Both	*CTU2* (rs11278302)	Intronic splice donor	Autosomal recessive	Benign	Homo	Microcephaly, facial dysmorphism, renal agenesis, and ambiguous genitalia syndrome	0.773762	0.844626
Both	*SON* (rs34373121)	Frameshift insertion	Autosomal dominant	Benign	Homo	ZTTK syndrome	1	0.999793
Both	*SON* (rs34377180)	Frameshift deletion						
L-PCOS	*CHIT1* (rs150192398)	Nonsense	Autosomal recessive	Benign	Hetero	Chitotriosidase deficiency	0.289137	0.427436
L-PCOS	*FUT6* (rs145035679)	Nonsense	NA	Uncertain	Hetero	Fucosyltransferase-6 deficiency	0.0289537	0.0532974
L-PCOS	***Cyp21A2* (rs7755898)**	**Nonsense**	**Autosomal recessive**	**Pathogenic**	**Hetero**	**Hyperandrogenism, congenital adrenal hyperplasia due to 21-hydroxylase deficiency**	**NA[3]**	**0.0128205**
L-PCOS	*Fam20C* (rs774848096)	Nonsense	Autosomal recessive	Benign	Hetero	Raine syndrome	NA	0.317136
Ob-PCOS	***KISS1* (rs71745629)**	**Stop-loss**	**Autosomal recessive**	**Benign**	**Hetero**	**Congenital idiopathic hypogonadotropic hypogonadism**	**0.221845**	**0.194549**
Ob-PCOS	***XDH* (rs773456900)**	**Frameshift insertion**	**Autosomal recessive**	**Pathogenic**	**Hetero**	**Xanthinuria**	**NA**	**0.0008285**

Table 2. Summary of the important candidate variants from clinical-exome sequencing of PCOS subjects. [1]Homozygous [2]Heterozygous [3]Not found Bold represents the PCOS-specific genes according to PCOSkb database.

highest (63.1%) followed by phenotype-A (30%) among lean subjects whereas the prevalence of phenotype-A (46.8%) and D (42.2%) were relatively closure among obese group[34]. In general, PCOS-A possesses insulin resistance, excess LH, T, 17αOHPG and ovarian estradiol by involving ovarian core pathways whereas PCOS-D possesses excess progesterone by involving adrenal regulatory pathways[35]. In the present study, the Proband-1 exhibit LH : FSH ratio ≤ 1 and increased AMH level suggesting the possibility of functional hypothalamic amenorrhea (FHA) which cause ovulation dysfunction and ovarian cyst formation[36]. Contrasting to this, high fasting glucose level (insulin resistance) might disrupt the ovarian steroidogenesis mechanism and thereby cause PCOS by initiating HPO feedback loop[35].

Subject	Gene (rsID)	Variation class	Variation impact	Gene type	Zygosity	Allele frequency	Ontology
Both	*CYB* (rs527236041)	Missense (Thr7Ile)	Moderate	Protein	Homo	NA	Oxidoreductase and ubiquinol-cytochrome-c reductase activity
	CYB (rs2853508)	Missense (Thr194Ala)	Moderate	Protein	Homo	NA	
	RNR1 (rs2001030)	Exonic-NC	Modifier	rRNA	Homo	NA	NA
	RNR1 (rs2853518)	Exonic-NC	Modifier	rRNA	Homo	NA	
	ATP6 (rs2001031)	Missense (Thr112Ala)	Moderate	Protein	Homo	NA	ATP hydrolysis activity and proton transmembrane transporter activity
L-PCOS	*ATP6* (rs879150284)	Missense (Thr45Ala)	Moderate	Protein	Homo	NA	
	ATP6 (**novel**)	Missense (Thr13Ala)	Moderate	Protein	Homo	NA	
	TL2 (rs2853498)	Exonic-NC	Modifier	tRNA	Homo	NA	NA
	ND1 (rs200180511)	Missense (Thr263Ala)	Moderate	Protein	Homo	NA	NADH dehydrogenase (ubiquinone) activity
	ND2 (rs878939965)	Missense (Trp239Cys)	Moderate	Protein	Homo	NA	
	ND5 (rs28359178)	Missense (Ala458Thr)	Moderate	Protein	Homo	NA	
	RNR2 (**novel**)	Exonic-NC	Modifier	rRNA	Homo	NA	NA
O-PCOS	*RNR2* (rs200040509)	Exonic-NC	Modifier	rRNA	Hetero	NA	
	TC (rs9659239)	Exonic-NC	Modifier	tRNA	Homo	NA	NA
	CO1 (**novel**)	Missense (Ile416Thr)	Moderate	Protein	Homo	NA	Fe-ion binding and electron transfer activity

Table 3. Summary of shortlisted variations in mitochondrial DNA of PCOS women.

In previous studies, *VDR* polymorphism (rs2228570) *aka* Fok1 was found to be associated with systemic lupus erythematosus[37], chronic spontaneous urticaria[38], T2DM[39], Parkinson's disease[40], autoimmune thyroiditis[41], Vit-D deficiency[42], and many other disorders[43–45]. In PCOS individuals, rs2228570 showed significant association with phenotype C, i.e., hyperandrogenemia and PCOM[46] and homozygous mutant genotype showed significant association with traits like hirsutism, infertility, alopecia and acne[47]. Lurie et al.[45] reported that each copy of the mutant allele was associated with a modest 9% increased risk of invasive ovarian carcinoma. In our study, a heterozygous genotype of *Fok1* has been found in both categories of PCOS women and thus can be correlated with the phenotypic features. Polymorphism in *Cyp21A2* (rs7755898) was found to be associated with Congenital Adrenal Hyperplasia due to 21-hydroxylase deficiency[48]. However, this rare variant (~1% of the population) has been found in Lean PCOS individuals for the first time in our study. The *KISS1* gene or its polymorphisms have not been reported to be associated directly with PCOS yet. However, follicular development is controlled by both estrogen production and feedback, with low estrogen levels fine-tuning GnRH/LH pulses via negative feedback. Estrogen production rises with follicular development, resulting in a surge in GnRH/LH and ovulation via positive feedback. Kisspeptin neurons, which are connected to GnRH neurons, play an important role in LH preovulatory surge and ovulation defoliation. Kisspeptin's positive feedback and raised LH levels play important roles in PCOS pathogenesis[49]. In this study, *KISS1* (rs71745629) has been found in Obese PCOS subject as a unique variant. It is noteworthy to mention that this stop-loss mutation was earlier found to be associated with metastatic cancer patients via *VEGF*-dependent angiogenesis signalling[50]. Besides this, another polymorphism in the *XDH* gene (rs773456900) has been found in our study which is a very rare variant and has not been reported to be associated with any disease. Apart from these PCOS-associated genes, *TMEM216* (rs10897158), *CTU2* (rs11278302), *SON* (rs34373121 & rs34377180) were found in both PCOS subjects whereas *CHIT1* (rs150192398), *FUT6* (rs145035679), *Fam20C* (rs774848096) were found in lean PCOS.

In this study, the family pedigree tree showed that T2DM is the most predominant metabolic disorder among both probands' families indicating a genetic pattern of shared architecture with T2DM. In addition, hypertension has been frequently found in the case of the Lean-PCOS family, which denotes the cardiometabolic risk in the subject. However, PCOS was not detected in any of the family members of lean-PCOS although hypomenorrhea on the paternal side, but it was detected in the mother of obese PCOS, thereby confirming the non-mendelian inheritance pattern of PCOS. Dhar and her colleagues[51] predicted mutations in 11 loci, i.e. *LepR*, *CAPN10*, *INSR*, *IRS1*, *CYP1A1*, *CYP11A1*, *CYP17A1*, *CYP19A1*, *ESR1*, *FSHR*, and *HSD11B1* which are functionally co-expressed with 5 loci (*Lep*, *INS*, *FDX1*, *PTPN1* and *PTPN11*). In another study, mutations in 6 loci, i.e. *ERBB4*, *GATA4*, *INSR*, *LHCGR*, *SUOX* and *YAP1* were predicted to have a potential effect on the manifestation of PCOS[52]. In this study, mutations in *LepR*, *CAPN10*, *INSR*, *GATA4*, *SUOX*, *CYP1A1*, *CYP19A1* and *ESR1* were observed in both category subjects while mutations in *Lep*, *ERBB4* were found in lean PCOS and mutations in *YAP1* was found in obese PCOS. Combining results suggest that leptin signalling impairment and insulin resistance thus play a major role in PCOS of the study population. Besides this, mutations in *CYP1A1*, *CYP19A1* and *ESR1* genes disrupt ovarian steroidogenesis and initiate a negative feedback loop, thereby leading to estrogen sensitivity and aromatase deficiency[53,54]. Additionally, mutations in *AR* (affecting androgen signalling), *AMH* (affecting the secretion of AMH), *AdipoR1* (affecting adiponectin signalling), *NAMPT*, *NPY* (affecting energy homeostasis), *PTEN*, *EGFR*, *Akt* (cancer signalling) were observed.

Clinical implications

Comorbidity analysis using the data from the present study showed in Fig. 9. Based on genetic network of the study participants, it was found that the risk for estrogen resistance, leptin receptor deficiency, MTHFR/folate deficiency, T2DM and acanthosis to co-occur with PCOS was much higher while shared-gene based
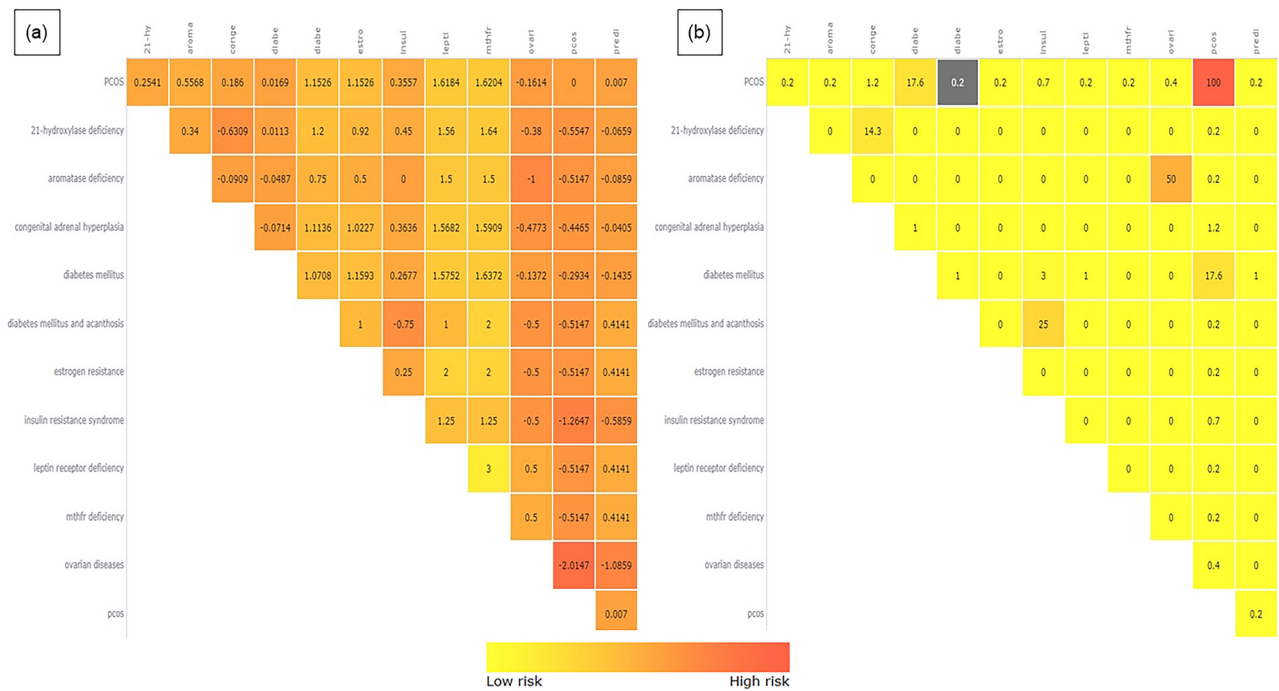
Fig. 9. Comorbidity analysis of PCOS individuals based on (**a**) network-based separation of shared genes, (**b**) shared genes.

analysis showed very high risk for T2DM to co-occur with PCOS. Overall, this study may give a comprehensive overview of PCOS among young women from West Bengal, India emphasizing obesity as the major phenotypic expression. Henceforth, folate supplementation, insulin sensitizers can effectively act as a therapeutic drug upon the associated genetic risk factors.

Research implications

Previously, a whole-genome association study on women from Asian and European populations revealed that *DENND1A*, *THADA*, *FSHR*, *LHCGR*, *AMH*, *AMHR2*, *ADIPOQ*, *FTO*, *HNF1A*, *CYP19*, *YAP1*, *HMGA2*, *RAB5B*, *SUOX*, *INSR*, and *TOX3* were the common genes associated with PCOS[55]. In Table 4, the comprehensive summary of genomics study on PCOS subjects. With this background, our study, for the first time, attempted to identify and differentiate the genetic outlook of PCOS between Obese and lean subjects. Our findings support future exploration with a statistically significant sample size to validate the high-throughput data. On the other hand, the identified genetic variations in this study must be extrapolated using a sanger-sequencing technique in a large population size in order to establish the variant as an early indicator of PCOS. Nonetheless, protein–protein interaction network and comorbidity analysis, when combined, show the primary signalling pathways and their ultimate fate, implying the extent of risk detection at the pre-pubertal stage and protection from future or long-term issues.

Strength and limitations

The present study, so far, is the first attempt to unravel the genetic mapping of PCOS women stratified by BMI. The study has been conducted with both invasive (whole-blood) and non-invasive (buccal swab) biological method, resulting same sensitivity and specificity towards high-throughput mutation screening technique, thereby open a door for hassle-free diagnostic strategy at any point of age group. On the other hand, genetic screening by clinical-exome sequencing can significantly reduce the ambiguity of data as well as the diagnostic costs when compared to whole-exome or whole-genome sequencing, which in turn, may have a positive impact on nations' economy. Like any other research work, this study has also certain drawbacks. First of all, the sample size is very small considering the clinical-exome analysis of sporadic cases, which may partially subsidize to the poor validation of sequencing-derived classifiers in an independent group of patients.

Conclusion

PCOS is a complicated and multidimensional endocrinopathy that affects reproductive women. It is typically diagnosed in adolescents or those under 30 years old, and can lead to pregnancy problems. A considerable amount of study has previously been conducted to investigate the role of genetic predisposition in the manifestation of PCOS, despite the fact that it varies depending on phenotypic expression. Previously, it was believed that PCOS only affected obese or overweight people, however extensive research has shown that the illness can also affect lean people. Our study infers that the underlying genetic association and mechanisms behind obese and lean PCOS women are significantly different. Along with the identification of key genes, gene network analysis and

Study population	Study design	Methodology	Sample size	Observation	References
China	Cross-sectional	WGS	PCOS = 19	Associated genes: *ZFYVE28*, *COL19A1*, *SIK3*, *ANKK1*, *PPIG*, and *REPIN1* Associated variants: *CELSR1* rs11914038, *ZBTB* rs651748	56
India	Observational	WES	PCOS = 52	Rare variants: *UCP1* (Thr227Ile), *UCP2* (Arg88Cys), *IRS1* (Ser892Gly) and *GHRL* (Leu72Met)	25
Pakistan	Family cohort	WES	PCOS = 2; Non-PCOS = 2	Rare variations: *MAML3*, *PCGF6*, *TFAP2C*, *NBPF16*, *ACTL9*, *CACNA1A*, *DOCK4*, *CPEB2*, *AMDHD2*, *OR2T35*, *LRP1*	9
China	Observational	WES	POI = 24	Rare variations: *DNAH6* (Q803K, V2894M), *BNC1* (D575V, D568V, L525P), *HFM1* (G1034S), *EIF2B4* (R466Q), *EIF2B2* (G26S, V308M), *EIF2B3* (M130T), *FOXL2* (A226T, G118C), *MCM9* (A830T), *FANCA* (H780Q), *ATM* (N789K), *GHR* (W94X), *LRPP* (A3S) and *RC* (R989C)	57
Brazil	Cohort	Targeted sequencing	PCOS = 45	*LMNA* Arg249Gln, *PIK3R1* Leu26Phe, *DLK1* Gly199Alafs*11, INSR Tyr1190, *LMNA* Gly638Arg & Arg644Cys, *NR5A1* Pro129Leu, *GATA4* Arg265Cys, *BMP15* Arg68Trp, *FSHR* Arg283Trp	26
India	Cohort	WES	PCOS = 43	Novel variants: *FSHR*, *SCARB1*, *INSR*, *DFFB*, *ACTG1*, *GPX4*, *CYC1* and *ALDOA*	8

Table 4. Comprehensive summary of identified genetic variations among different study populations.

comorbities suggest different mode of cellular and molecular signalling leading to more complicated disease. However, additional research involving a large number of people are required to validate our results, which can be used as a detection biomarker in a focused panel. Since the oral contraceptive pills (OCPs) are a generalized mode of therapy, these might cause certain toxicity to the patients as well. Herein, our study can improvise the generalized treatment into a more personalized treatment strategy by using cost-effective individual genetic profile. Moreover, our study implies the significance of early disease prediction in order to minimize disease severity and later stage complications by treating with precision medicine and specific cost-effective diagnosis.

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

References
1. Kulkarni, S. et al. Polycystic ovary syndrome: Current scenario and future insights. *Drug Discov. Today.* **28**(12), 103821. https://doi.org/10.1016/j.drudis.2023.103821 (2023).
2. Ajmal, N., Khan, S. Z. & Shaikh, R. Polycystic ovary syndrome (PCOS) and genetic predisposition: A review article. *Eur. J. Obstet. Gynecol. Reprod. Biol. X.* **3**, 100060. https://doi.org/10.1016/j.eurox.2019.100060 (2019).
3. Prapas, N. et al. Genetics of polycystic ovary syndrome. *Hippokratia.* **13**(4), 216–223 (2009).
4. Escobar-Morreale, H. F. Polycystic ovary syndrome: Definition, aetiology, diagnosis and treatment. *Nat. Rev. Endocrinol.* **14**(5), 270–284. https://doi.org/10.1038/nrendo.2018.24 (2018).
5. Azziz, R. et al. The androgen excess and PCOS society criteria for the polycystic ovary syndrome: The complete task force report. *Fertil. Steril.* https://doi.org/10.1016/j.fertnstert.2008.06.035 (2009).
6. Govind, A., Obhrai, M. S. & Clayton, R. N. Polycystic ovaries are inherited as an autosomal dominant trait: Analysis of 29 polycystic ovary syndrome and 10 control families. *J. Clin. Endocrinol. Metab.* **84**(1), 38–43. https://doi.org/10.1210/jcem.84.1.5382 (1999).
7. Franks, S. et al. The genetic basis of polycystic ovary syndrome. *Hum. Reprod.* **12**(12), 2641–2648. https://doi.org/10.1093/humrep/12.12.2641 (1997).
8. Janani, D. M. et al. Whole exome and targeted sequencing reveal novel mutations associated with inherited PCOS condition in an Indian cohort. *J. Hum. Genet.* **68**(1), 39–46. https://doi.org/10.1038/s10038-022-01093-2 (2023).
9. Khan, M. J., Nazli, R., Ahmed, J. & Basit, S. Whole genome sequencing instead of whole exome sequencing is required to identify the genetic causes of polycystic ovary syndrome in Pakistani families. *Pak. J. Med. Sci.* **34**(3), 540–545. https://doi.org/10.12669/pjms.343.14644 (2018).
10. Diamanti-Kandarakis, E., Kandarakis, H. & Legro, R. S. The role of genes and environment in the etiology of PCOS. *Endocrine.* **30**(1), 19–26. https://doi.org/10.1385/ENDO:30:1:19 (2006).
11. Joseph, S., Barai, R. S., Bhujbalrao, R. & Idicula-Thomas, S. PCOSKB: A KnowledgeBase on genes, diseases, ontology terms and biochemical pathways associated with PolyCystic Ovary Syndrome. *Nucleic Acids Res.* **44**(D1), D1032–D1035. https://doi.org/10.1093/nar/gkv1146 (2016).
12. Sharma, M. et al. PCOSKBR2: A database of genes, diseases, pathways, and networks associated with polycystic ovary syndrome. *Sci. Rep.* **10**(1), 14738. https://doi.org/10.1038/s41598-020-71418-8 (2020).
13. Dhar, S., Mondal, K. K. & Bhattacharjee, P. Influence of lifestyle factors with the outcome of menstrual disorders among adolescents and young women in West Bengal, India. *Sci. Rep.* **13**(1), 12476. https://doi.org/10.1038/s41598-023-35858-2 (2023).
14. Kamran, H. S. & Saeed, T. Situation analysis, cultural beliefs, lifestyle, and the psychological impact of polycystic ovary syndrome in Europe. In *Polycystic Ovary Syndrome* (eds Kamran, H. S. & Saeed, T.) 216–219 (Elsevier, 2024). https://doi.org/10.1016/B978-0-323-87932-3.00037-2.
15. Khan, R., Rehman, R. & Alam, F. Situation analysis of polycystic ovary syndrome in Western Asia. In *Polycystic Ovary Syndrome* (eds Khan, R. et al.) 207–215 (Elsevier, 2024). https://doi.org/10.1016/B978-0-323-87932-3.00036-0.
16. Mohamed Rashid Sokwala, S. B. & Dodia, R. Global approach to polycystic ovary syndrome in Africa. In *Polycystic Ovary Syndrome* (eds Mohamed Rashid Sokwala, S. B. & Dodia, R.) 220–228 (Elsevier, 2024). https://doi.org/10.1016/B978-0-323-87932-3.00038-4.
17. Liu, X., Zhang, J. & Wang, S. Global, regional, and national burden of infertility attributable to PCOS, 1990–2019. *Hum. Reprod.* **39**(1), 108–118. https://doi.org/10.1093/humrep/dead241 (2024).
18. Fernandez, R. C. et al. Diagnosis delayed: health profile differences between women with undiagnosed polycystic ovary syndrome and those with a clinical diagnosis by age 35 years. *Hum. Reprod.* **36**(8), 2275–2284. https://doi.org/10.1093/humrep/deab101 (2021).

19. Koivunen, R. et al. The prevalence of polycystic ovaries in healthy women. *Acta Obstet. Gynecol. Scand.* **78**(2), 137–141. https://doi.org/10.1080/j.1600-0412.1999.780212.x (1999).
20. Jeffrey Chang, R. & Coffler, M. S. Polycystic ovary syndrome: Early detection in the adolescent. *Clin. Obstet. Gynecol.* **50**(1), 178–187. https://doi.org/10.1097/GRF.0b013e31802f50fc (2007).
21. Satam, H. et al. Next-generation sequencing technology: Current trends and advancements. *Biology (Basel).* **12**(7), 997. https://doi.org/10.3390/biology12070997 (2023).
22. Fogel, B. L., Satya-Murti, S. & Cohen, B. H. Clinical exome sequencing in neurologic disease. *Neurol. Clin. Pract.* **6**(2), 164–176. https://doi.org/10.1212/CPJ.0000000000000239 (2016).
23. Zeber-Lubecka, N. et al. Exome sequencing to explore the possibility of predicting genetic susceptibility to the joint occurrence of polycystic ovary syndrome and Hashimoto's thyroiditis. *Front. Immunol.* **14**(July), 1–17. https://doi.org/10.3389/fimmu.2023.1193293 (2023).
24. Gao, M. et al. Association between single nucleotide polymorphisms, TGF-β1 promoter methylation, and polycystic ovary syndrome. *BMC Pregnancy Childbirth.* **24**(1), 1–15. https://doi.org/10.1186/s12884-023-06210-3 (2024).
25. Sharma, P., Halder, A., Jain, M. & Tripathi, M. Whole exome sequencing reveals rare variants in genes associated with metabolic disorders in women with PCOS. *J. Hum. Reprod. Sci.* **16**(4), 307–316. https://doi.org/10.4103/jhrs.jhrs_13_23 (2023).
26. Crespo, R. P. et al. High-throughput sequencing to identify monogenic etiologies in a preselected polycystic ovary syndrome cohort. *J. Endocr. Soc.* **6**(9), 1–10. https://doi.org/10.1210/jendso/bvac106 (2022).
27. Bozdag, G., Mumusoglu, S., Zengin, D., Karabulut, E. & Yildiz, B. O. The prevalence and phenotypic features of polycystic ovary syndrome: A systematic review and meta-analysis. *Hum. Reprod.* **31**(12), 2841–2855. https://doi.org/10.1093/humrep/dew218 (2016).
28. Salari, N. et al. Global prevalence of polycystic ovary syndrome in women worldwide: A comprehensive systematic review and meta-analysis. *Arch. Gynecol. Obstet.* https://doi.org/10.1007/s00404-024-07607-x (2024).
29. Chiaffarino, F. et al. Prevalence of polycystic ovary syndrome in European countries and USA: A systematic review and meta-analysis. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **279**, 159–170. https://doi.org/10.1016/j.ejogrb.2022.10.020 (2022).
30. Farhadi-Azar, M. et al. The prevalence of polycystic ovary syndrome, its phenotypes and cardio-metabolic features in a community sample of Iranian population: Tehran lipid and glucose study. *Front. Endocrinol. (Lausanne).* https://doi.org/10.3389/fendo.2022.825528 (2022).
31. Myers, S. H., Montanino Oliva, M., Nordio, M. & Unfer, V. PCOS phenotype focus: Phenotype D under the magnifying glass. *Arch. Gynecol. Obstet.* **309**(6), 2307–2313. https://doi.org/10.1007/s00404-024-07408-2 (2024).
32. Patel, S. et al. Evaluation of hormonal profile and ovarian morphology among adolescent girls with menstrual irregularities in a tertiary care centre at central India. Talbott EO, Ed. *Sci. World J.* **2022**, 1–8. https://doi.org/10.1155/2022/3047526 (2022).
33. Sachdeva, G., Gainder, S., Suri, V., Sachdeva, N. & Chopra, S. Comparison of the different PCOS phenotypes based on clinical metabolic, and hormonal profile, and their response to clomiphene. *Indian J. Endocrinol. Metab.* **23**(3), 326. https://doi.org/10.4103/ijem.IJEM_30_19 (2019).
34. Dadachanji, R., Patil, A., Joshi, B. & Mukherjee, S. Elucidating the impact of obesity on hormonal and metabolic perturbations in polycystic ovary syndrome phenotypes in Indian women. Qu F, ed. *PLoS One.* **16**(2), e0246862. https://doi.org/10.1371/journal.pone.0246862 (2021).
35. Elsayed, A. M. et al. Clinical phenotypes of PCOS: A cross-sectional study. *Reprod. Sci.* **30**(11), 3261–3272. https://doi.org/10.1007/s43032-023-01262-4 (2023).
36. Boegl, M. et al. The LH:FSH ratio in functional hypothalamic amenorrhea: An observational study. *J. Clin. Med.* **13**(5), 1201. https://doi.org/10.3390/jcm13051201 (2024).
37. Meza-Meza, M. R. et al. Vitamin D Receptor (VDR) genetic variants: Relationship of FokI genotypes with VDR expression and clinical disease activity in systemic lupus erythematosus patients. *Genes (Basel).* **13**(11), 2016. https://doi.org/10.3390/genes13112016 (2022).
38. Egea, E. et al. Significant association between TaqI and FokI VDR gene polymorphisms and chronic spontaneous urticaria in a Colombian Caribbean population. *Allergol. Immunopathol. (Madr.).* **50**(SP2), 8–14. https://doi.org/10.15586/aei.v50iSP2.696 (2022).
39. Mohammed, A. A. et al. Allelic discrimination of vitamin D receptor polymorphisms and risk of type 2 diabetes mellitus: A case-controlled study. *Healthcare.* **11**(4), 485. https://doi.org/10.3390/healthcare11040485 (2023).
40. Agliardi, C. et al. The VDR FokI (rs2228570) polymorphism is involved in Parkinson's disease. *J. Neurol. Sci.* **428**, 117606. https://doi.org/10.1016/j.jns.2021.117606 (2021).
41. Glonti, S. et al. The study of VDR FOKI RS2228570 SNP in autoimmune thyroiditis. *Georgian Med. News.* **337**, 99–103 (2023).
42. Tuncel, G., Temel, S. G. & Ergoren, M. C. Strong association between VDR FokI (rs2228570) gene variant and serum vitamin D levels in Turkish Cypriots. *Mol. Biol. Rep.* **46**(3), 3349–3355. https://doi.org/10.1007/s11033-019-04796-6 (2019).
43. Gnagnarella, P. et al. Vitamin D receptor polymorphism FokI and cancer risk: A comprehensive meta-analysis. *Carcinogenesis.* **35**(9), 1913–1919. https://doi.org/10.1093/carcin/bgu150 (2014).
44. Das, B. et al. Frequency of fokI and taqI polymorphism of vitamin D receptor gene in Indian population and its association with 25-hydroxyvitamin D levels. *Indian J. Hum. Genet.* **15**(3), 108. https://doi.org/10.4103/0971-6866.60186 (2009).
45. Lurie, G. et al. Vitamin D receptor rs2228570 polymorphism and invasive ovarian carcinoma risk: Pooled analysis in five studies within the Ovarian Cancer Association Consortium. *Int. J. Cancer.* **128**(4), 936–943. https://doi.org/10.1002/ijc.25403 (2011).
46. Albahlol, I. A. et al. Vitamin D receptor gene polymorphism and polycystic ovary syndrome susceptibility. *BMC Med. Genomics.* **16**(1), 108. https://doi.org/10.1186/s12920-023-01541-8 (2023).
47. Dasgupta, S., Dutta, J., Annamaneni, S., Kudugunti, N. & Battini, M. Association of vitamin D receptor gene polymorphisms with polycystic ovary syndrome among Indian women. *Indian J. Med. Res.* **142**(3), 276. https://doi.org/10.4103/0971-5916.166587 (2015).
48. Doleschall, M. et al. A unique haplotype of RCCX copy number variation: From the clinics of congenital adrenal hyperplasia to evolutionary genetics. *Eur. J. Hum. Genet.* **25**(6), 702–710. https://doi.org/10.1038/ejhg.2017.38 (2017).
49. Uenoyama, Y., Inoue, N., Nakamura, S. & Tsukamura, H. Kisspeptin neurons and estrogen-estrogen receptor α signaling: unraveling the mystery of steroid feedback system regulating mammalian reproduction. *Int. J. Mol. Sci.* **22**(17), 9229. https://doi.org/10.3390/ijms22179229 (2021).
50. Paré-Brunet, L. et al. Genetic variations in the VEGF pathway as prognostic factors in metastatic colorectal cancer patients treated with oxaliplatin-based chemotherapy. *Pharmacogenomics J.* **15**(5), 397–404. https://doi.org/10.1038/tpj.2015.1 (2015).
51. Dhar, S., Mridha, S. & Bhattacharjee, P. Mutational landscape screening through comprehensive in silico analysis for polycystic ovarian syndrome-related genes. *Reprod. Sci.* **29**(2), 480–496. https://doi.org/10.1007/s43032-021-00752-7 (2022).
52. Prabhu, B. N. et al. Conceptualization of functional single nucleotide polymorphisms of polycystic ovarian syndrome genes: an in silico approach. *J. Endocrinol. Invest.* https://doi.org/10.1007/s40618-021-01498-4 (2021).
53. Liu, F. et al. Identification of a novel ESR1 mutation in a Chinese PCOS woman with estrogen insensitivity in IVF treatment. *Reprod. Biol. Endocrinol.* **20**(1), 157. https://doi.org/10.1186/s12958-022-01029-7 (2022).
54. Heidarzadehpilehrood, R. et al. A review on CYP11A1, CYP17A1, and CYP19A1 polymorphism studies: Candidate susceptibility genes for polycystic ovary syndrome (PCOS) and infertility. *Genes (Basel).* **13**(2), 302. https://doi.org/10.3390/genes13020302 (2022).

55. Castillo-Higuera, T. et al. A comprehensive overview of common polymorphic variants in genes related to polycystic ovary syndrome. *Reprod. Sci.* **28**(9), 2399–2412. https://doi.org/10.1007/s43032-020-00375-4 (2021).
56. Lai, W. et al. Identifying genetic factors of polycystic ovary syndrome in women with epilepsy: A whole-genome sequencing study. *Neuroendocrinology.* **114**(3), 223–233. https://doi.org/10.1159/000534531 (2024).
57. Liu, H. et al. Whole-exome sequencing in patients with premature ovarian insufficiency: Early detection and early intervention. *J. Ovarian Res.* **13**(1), 1–8. https://doi.org/10.1186/s13048-020-00716-6 (2020).

Acknowledgements

Author contributions

Both S.D. and P.B. conceptualized and visualized the research. S.D. carried out the entire experiments, data analysis and representation, and the original draft of manuscript preparation; P.B. supervised the overall research including resource management, fund acquisition, manuscript reviewing & editing.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-75719-0.

Correspondence and requests for materials should be addressed to P.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.