




Mendelian Randomization Analysis Identified Potential Genes Pleiotropically Associated with Polycystic Ovary Syndrome

Qian Sun^{1,2} · Yuan Gao¹ · Jingyun Yang^{3,4} · Jiayi Lu⁵ · Wen Feng¹ · Wen Yang^{1,2} 

Received: 17 July 2021 / Accepted: 14 October 2021 / Published online: 26 October 2021
© Society for Reproductive Investigation 2021

Abstract

Polycystic ovary syndrome (PCOS) is a common endocrine disorder with unclear etiology. Some genes may be pleiotropically or potentially causally associated with PCOS. In the present study, the summary data-based Mendelian randomization (SMR) method integrating genome-wide association study (GWAS) for PCOS and expression quantitative trait loci (eQTL) data was applied to identify genes that were pleiotropically associated with PCOS. Separate SMR analysis was performed using eQTL data in the ovary and whole blood. Although no genes showed significant pleiotropic association with PCOS after correction for multiple testing, some of the genes exhibited suggestive significance. *RPS26* showed the strongest suggestive pleiotropic association with PCOS in both SMR analyses (β [SE]=0.10[0.03], $P_{SMR}=1.72\times 10^{-4}$ for ovary; β [SE]=0.11[0.03], $P_{SMR}=1.40\times 10^{-4}$ for whole blood). *PM20D1* showed the second strongest suggestive pleiotropic association with PCOS in the SMR analysis using eQTL data for the whole blood and was also among the top ten hit genes in the SMR analysis using eQTL data for the ovary. Two other genes, including *CTC-457L16.2* and *NEIL2*, were among the top ten hit genes in both SMR analyses. In conclusion, this study revealed multiple genes that were potentially involved in the pathogenesis of PCOS.

Keywords Polycystic ovary syndrome · Pleiotropic association · Expression quantitative trait loci · Summary Mendelian randomization

Abbreviations

BPA bisphenol A
eQTL expression quantitative trait loci
GWAS genome-wide association study
HEIDI heterogeneity in dependent instruments
mQTL DNA methylation quantitative trait loci
IR insulin resistance

MR Mendelian randomization
PCOS polycystic ovary syndrome
QoL quality of life
SMR summary data-based Mendelian randomization
T2D type 2 diabetes

Introduction

Polycystic ovary syndrome (PCOS) is a common endocrine disorder that often occurs in women of childbearing age, often accompanied by insulin resistance (IR) and obesity [1]. According to the revised Rotterdam 2003 criteria [2], PCOS is diagnosed when two of the following three criteria are present: (1) oligo- or an-ovulation, (2) clinical and/or biochemical signs of hyperandrogenism, and (3) polycystic ovaries on ultrasound. As a highly prevalent endocrine disorder, PCOS affects 5–13% of reproductive-aged women and more than 10% of adolescent girls [3–6]. PCOS is associated with a variety of consequences, such as reproductive issues, metabolic abnormalities, and psychological disorders [7]. Furthermore, more than half of the PCOS patients suffer from infertility [8], and infertile women and adolescent

Qian Sun and Yuan Gao contributed equally to this work.

✉ Wen Yang
1010393828@qq.com

- ¹ Department of Gynecology, The First Affiliated Hospital of Kangda College of Nanjing Medical University, 6 Zhenhua Road, Lianyungang 222061, Jiangsu, China
- ² Department of Gynecology, The First People's Hospital of Lianyungang, Lianyungang, Jiangsu, China
- ³ Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA
- ⁴ Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA
- ⁵ Department of Finance, School of Economics, Shanghai University, Shanghai, China

girls with PCOS were reported to have reduced quality of life (QoL) [9–11].

PCOS is a multifaceted disease, and previous studies have found its association with genetic factors [12], environmental factors such as exposure to Bisphenol A (BPA) [13], and life style factors such as smoking [14]. However, the exact etiology of PCOS remains to be elucidated. More studies are needed to further explore the pathological mechanisms underlying PCOS to facilitate diagnosis and treatment of this common and jeopardizing disease.

Genetics play an important role in the pathogenesis of PCOS. Previous twin studies demonstrated that genetics explained more than 70% of PCOS pathogenesis [15]. Multiple genome-wide association studies (GWAS) conducted on Han Chinese, Korean, and European populations have identified a few genetic variants in association with PCOS, such as single-nucleotide polymorphisms (SNPs) in/near *FSHR*, *THADA* [16], *YAP1*, *HMGA2* [17], *KHDRBS3* [18], *FSHB*, *GATA4/NEIL2* [19], *ERBB4/HER4*, and *RAD50* [20]. However, biological interpretation of the identified genetic variants in the etiology of PCOS remains largely unclear. It is likely that the genetic variants identified in GWAS could exert their effects on diseases/disorders via gene expression because many of them are located in non-coding regions [21]. Therefore, exploring the relationship between genetic variation and gene expression can help better understand the regulatory pathways underlying the pathogenesis of PCOS.

Mendelian randomization (MR) is a method for exploring potentially causal association between an exposure and an outcome by using genetic variants as the instrumental variables (IVs) for exposure [22–24]. Compared with traditional statistical methods used in the association studies, MR can reduce confounding and reverse causation and is becoming increasingly popular in the exploration of etiological mechanisms [25, 26]. A novel analytical framework

through a summary data-based MR (SMR) approach integrating cis- expression quantitative trait loci (cis-eQTL, i.e., genetic variants near a gene that explain the variance of the expression level of the gene [27]) or cis- DNA methylation QTL (cis-mQTL, i.e., CpG sites near a gene that explain the variance of the methylation level of the CpG site [28]) and GWAS data was recently proposed [29]. This method has been employed in identifying gene expressions or DNA methylation loci that are pleiotropically or potentially causally associated with various phenotypes, such as severity of COVID-19, major depression, and Alzheimer’s disease [30–32], implying that it is a promising tool to explore genes that are pleiotropically associated with complex traits.

In this paper, the SMR method integrating summarized GWAS data for PCOS and cis- eQTL data were applied to prioritize genes that are pleiotropically/potentially causally associated with PCOS.

Materials and Methods

This study utilized SMR method integrating GWAS summary results for PCOS and eQTL data in ovary and blood. All the data were publicly available. The analytic process of the present study is illustrated in Fig. 1.

Data Sources

eQTL Data

In the SMR analysis, cis-eQTL genetic variants were used as the instrumental variables (IVs) for gene expression. SMR analysis were performed using eQTL data in ovary because ovary is directly involved in PCOS. Analysis using eQTL

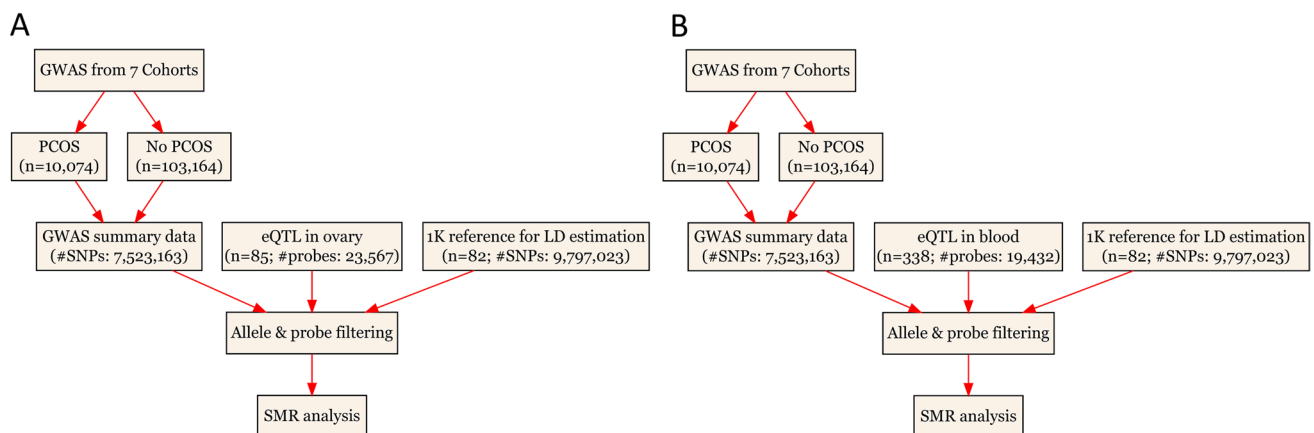


Fig. 1 Flow chart for the SMR analysis. **A** SMR analysis using eQTL data from ovary and **B** SMR analysis using eQTL data from blood. eQTL, expression quantitative trait loci; GWAS, genome-wide asso-

ciation studies; LD, linkage disequilibrium; PCOS, polycystic ovary syndrome; SMR, summary data-based Mendelian randomization; SNP, single nucleotide polymorphisms

data from blood were also conducted as blood might reflect hormonal or metabolic traits associated with PCOS. The eQTL data were obtained from the V7 release of the GTEx summarized data. Detailed information regarding sample acquisition and treatment were provided elsewhere [33]. The summarized data included 85 participants for ovary and 338 participants for blood [34]. The eQTL data can be downloaded at <https://cnsgenomics.com/data/SMR/#eQTLsummarydata>.

GWAS Data for PCOS

The GWAS summarized data for CCT were provided by a recent genome-wide association meta-analysis of PCOS [35]. The results were based on meta-analyses of seven cohorts of European descent, including a total of 10,074 PCOS cases and 103,164 controls. Details of the study participants and design of the study can be found in the paper published previously [35]. Diagnosis of PCOS was based on the NIH criteria [36], which was established in 1990 and requires hyperandrogenism (HA) and ovulatory dysfunction (OD), or the Rotterdam Criteria [2] which was established in 2003, includes the presence of polycystic ovarian morphology (PCOM) and requires the presence of at least two of three traits as specified above. The self-report-based data from 23andMe, Inc. (Mountain View, CA, USA) were excluded. The GWAS summarized data can be downloaded at <https://www.repository.cam.ac.uk/handle/1810/283491>.

SMR Analysis

In the SMR analyses, cis-eQTL was used as the instrumental variable, gene expression was the exposure, and PCOS was the outcome. The analysis was done using the method as implemented in the software SMR. SMR applies the principles of MR to jointly analyze GWAS and eQTL summary statistics to test for pleiotropic association between gene expression and a trait due to a shared and potentially causal variant at a locus. Detailed information regarding the SMR method was reported in a previous publication [29]. The heterogeneity in dependent instruments (HEIDI) test [29] was conducted to evaluate the existence of linkage in the observed association. Rejection of the null hypothesis (i.e., $P_{\text{HEIDI}} < 0.05$) indicates that the observed association could be due to two distinct genetic variants in high linkage disequilibrium with each other. The default settings in SMR were adopted (e.g., $P_{\text{eQTL}} < 5 \times 10^{-8}$, minor allele frequency [MAF] > 0.01 , removing SNPs in very strong linkage disequilibrium [LD, $r^2 > 0.9$] with the top associated eQTL, and removing SNPs in low LD or not in LD [$r^2 < 0.05$] with the top associated eQTL), and false discovery rate (FDR) was used to adjust for multiple testing.

Data cleaning and statistical/bioinformatical analysis were performed using R version 4.0.4 (<https://www.r-project.org/>), PLINK 1.9 (<https://www.cog-genomics.org/plink/1.9/>), and SMR (<https://cnsgenomics.com/software/smr/>).

Results

Basic Information of the Summarized Data

The GWAS summarized data were based on GWAS meta-analysis of 113,238 subjects (10,074 PCOS cases and 103,164 controls) [35]. After checking of allele frequencies among the datasets and LD pruning, there were about 6.4 million eligible SNPs included in the final SMR analysis. The number of participants for GETx eQTL data in ovary is smaller ($n=85$), compared with that in whole blood ($n=385$), so is the number of eligible probes (1530 vs. 4490). The detailed information is shown in Table 1.

Pleiotropic Association with PCOS

Information of the top ten probes using eQTL data for the ovary and whole blood was presented in Table 2. Although no genes showed significant pleiotropic association with PCOS after correction for multiple testing, some of genes exhibited suggestive significance. Specifically, *RPS26* (ENSG00000197728.5) showed the strongest suggestive pleiotropic association with PCOS in both SMR analyses ($\beta[\text{SE}] = 0.10[0.03]$, $P_{\text{SMR}} = 1.72 \times 10^{-4}$ for ovary; $\beta[\text{SE}] = 0.11[0.03]$, $P_{\text{SMR}} = 1.40 \times 10^{-4}$ for whole blood; Fig. 2). *PM20D1* (ENSG00000162877.8) showed the second strongest suggestive pleiotropic association with PCOS in the SMR analysis using eQTL data for the whole blood, and it was also among the top ten hit genes in the SMR analysis using eQTL data for the ovary ($\beta[\text{SE}] = 0.10[0.03]$, $P_{\text{SMR}} = 2.94 \times 10^{-3}$ for ovary; $\beta[\text{SE}] = 0.14[0.04]$, $P_{\text{SMR}} = 8.38 \times 10^{-4}$ for blood; Fig. 3). Two other genes, including *CTC-457L16.2* (ENSG00000262319.1; $\beta[\text{SE}] = -0.2[0.06]$, $P_{\text{SMR}} = 1.55 \times 10^{-3}$ for ovary; $\beta[\text{SE}] = -0.31[0.09]$, $P_{\text{SMR}} = 1.07 \times 10^{-3}$ for blood; Fig. 4) and

Table 1 Basic information of the GWAS and eQTL data.

Data source	Total number of participants	Number of eligible genetic variants or probes
eQTL data		
Ovary	85	1530
Whole blood	338	4490
GWAS data	113,238	6,400,755

GWAS, genome-wide association studies; QTL, quantitative trait loci

Table 2 The top ten probes identified in the SMR analysis.

eQTL data	Probe ID	Gene	CHR	Top SNP	P_{eQTL}	P_{GWAS}	Beta	SE	P_{SMR}	P_{HEIDI}	N_{SNP}
Ovary	ENSG00000197728.5	<i>RPS26</i>	12	rs1131017	3.89×10^{-55}	1.60×10^{-4}	0.10	0.03	1.72×10^{-4}	1.87×10^{-1}	20
	ENSG00000261390.1	<i>RPI1-345M22.2</i>	16	rs7196490	1.02×10^{-11}	1.20×10^{-5}	-0.28	0.08	3.18×10^{-4}	8.51×10^{-1}	20
	ENSG00000154328.11	<i>NEIL2</i>	8	rs804271	1.71×10^{-14}	3.70×10^{-4}	-0.20	0.06	1.02×10^{-3}	1.87×10^{-1}	20
	ENSG00000225706.1	<i>RPI1-75C9.1</i>	9	rs77301379	7.45×10^{-20}	5.30×10^{-4}	0.13	0.04	1.41×10^{-3}	8.71×10^{-1}	20
	ENSG00000262319.1	<i>CTC-457L16.2</i>	17	rs5885404	3.59×10^{-9}	1.80×10^{-4}	-0.20	0.06	1.55×10^{-3}	9.58×10^{-1}	12
	ENSG00000172000.3	<i>ZNF556</i>	19	rs4807349	2.34×10^{-10}	5.70×10^{-4}	-0.24	0.08	2.24×10^{-3}	1.84×10^{-1}	20
	ENSG00000162877.8	<i>PM20D1</i>	1	rs823080	3.53×10^{-22}	8.50×10^{-4}	0.10	0.03	2.94×10^{-3}	9.31×10^{-1}	20
	ENSG00000186300.7	<i>ZNF555</i>	19	rs917653	3.31×10^{-12}	8.30×10^{-4}	-0.15	0.05	3.18×10^{-3}	8.66×10^{-1}	17
	ENSG00000165623.5	<i>UCMA</i>	10	rs78415540	1.12×10^{-20}	3.80×10^{-3}	-0.09	0.03	3.69×10^{-3}	1.46×10^{-1}	20
	ENSG00000100211.6	<i>CBY1</i>	22	rs62229988	3.09×10^{-13}	2.50×10^{-3}	0.21	0.07	3.90×10^{-3}	6.00×10^{-1}	20
Whole blood	ENSG00000197728.5	<i>RPS26</i>	12	rs1131017	5.82×10^{-101}	1.60×10^{-4}	0.11	0.03	1.40×10^{-4}	3.60×10^{-1}	20
	ENSG00000162877.8	<i>PM20D1</i>	1	rs1772143	3.17×10^{-45}	5.20×10^{-4}	0.14	0.04	8.38×10^{-4}	8.78×10^{-1}	20
	ENSG00000262319.1	<i>CTC-457L16.2</i>	17	rs5885404	2.28×10^{-11}	1.80×10^{-4}	-0.31	0.09	1.07×10^{-3}	9.96×10^{-1}	16
	ENSG00000100201.14	<i>DDX17</i>	22	rs2267390	5.13×10^{-18}	4.90×10^{-4}	-0.77	0.24	1.08×10^{-3}	3.23×10^{-2}	20
	ENSG00000127483.13	<i>HPIBP3</i>	1	rs9426742	2.54×10^{-16}	6.90×10^{-4}	0.60	0.19	1.53×10^{-3}	9.17×10^{-1}	20
	ENSG00000242550.1	<i>SERPINB10</i>	18	rs9955526	4.76×10^{-37}	1.00×10^{-3}	0.22	0.07	1.64×10^{-3}	1.11×10^{-1}	20
	ENSG00000128482.11	<i>RNF112</i>	17	rs7502682	2.34×10^{-20}	9.90×10^{-4}	0.23	0.07	1.71×10^{-3}	2.58×10^{-1}	20
	ENSG00000254415.3	<i>SIGLEC14</i>	19	rs872629	2.70×10^{-30}	1.20×10^{-3}	0.20	0.07	1.97×10^{-3}	2.71×10^{-1}	20
	ENSG00000154328.11	<i>NEIL2</i>	8	rs804270	2.11×10^{-14}	1.00×10^{-3}	-0.30	0.10	2.25×10^{-3}	1.33×10^{-1}	20
	ENSG00000105501.7	<i>SIGLEC5</i>	19	rs872629	7.55×10^{-14}	1.20×10^{-3}	0.52	0.18	3.15×10^{-3}	1.83×10^{-1}	18

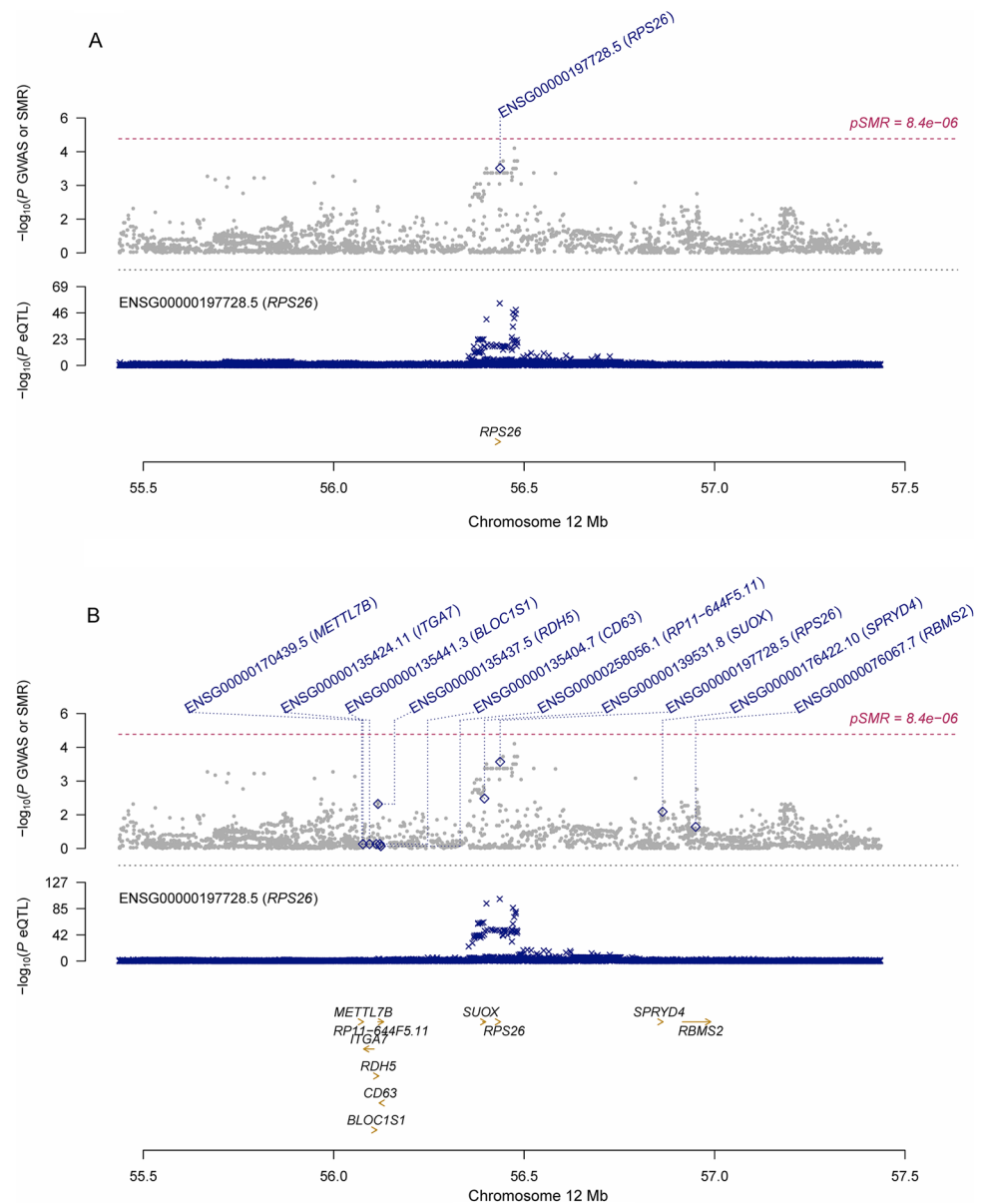
The GWAS summarized data were provided by the study of Shungin et al. and can be downloaded at <https://www.repositorio.cam.ac.uk/handle/1810/283491>. The GTEx eQTL data can be downloaded at <https://cnsgenomics.com/data/SMR/#eQTLsummarydata>

P_{eQTL} is the P -value of the top associated cis-eQTL in the eQTL analysis, and P_{GWAS} is the P -value for the top associated cis-eQTL in the GWAS analysis. Beta is the estimated effect size in SMR analysis, SE is the corresponding standard error, P_{SMR} is the P -value for SMR analysis, P_{HEIDI} is the P -value for the HEIDI test, and N_{SNP} is the number of SNPs involved in the HEIDI test. FDR was calculated at $P=10^{-3}$ threshold.

Bold font means statistical significance after correction for multiple testing using FDR.

CHR, chromosome; HEIDI, heterogeneity in dependent instruments; SMP, single-nucleotide polymorphism; SMR, summary data-based Mendelian randomization; QTL, quantitative trait loci; FDR, false discovery rate; GWAS, genome-wide association studies

Fig. 2 Pleiotropic association of *RPS26* with PCOS. **A** SMR analysis results using eQTL data for ovary. **B** SMR analysis results using eQTL data for whole blood. Top plot, grey dots represent the $-\log_{10}(P)$ values for SNPs from the GWAS of PCOS, with solid rhombuses indicating that the probes pass HEIDI test. Middle plot, eQTL results. Bottom plot, location of genes tagged by the probes. GWAS, genome-wide association studies; SMR, summary data-based Mendelian randomization; HEIDI, heterogeneity in dependent instruments; eQTL, expression quantitative trait loci; PCOS, polycystic ovary syndrome



NEIL2 (ENSG00000154328.11; $\beta[\text{SE}] = -0.2[0.06]$, $P_{\text{SMR}} = 1.02 \times 10^{-3}$ for ovary; $\beta[\text{SE}] = -0.30[0.10]$, $P_{\text{SMR}} = 2.25 \times 10^{-3}$ for blood; Fig. 5), were among the top ten hit genes in both SMR analyses.

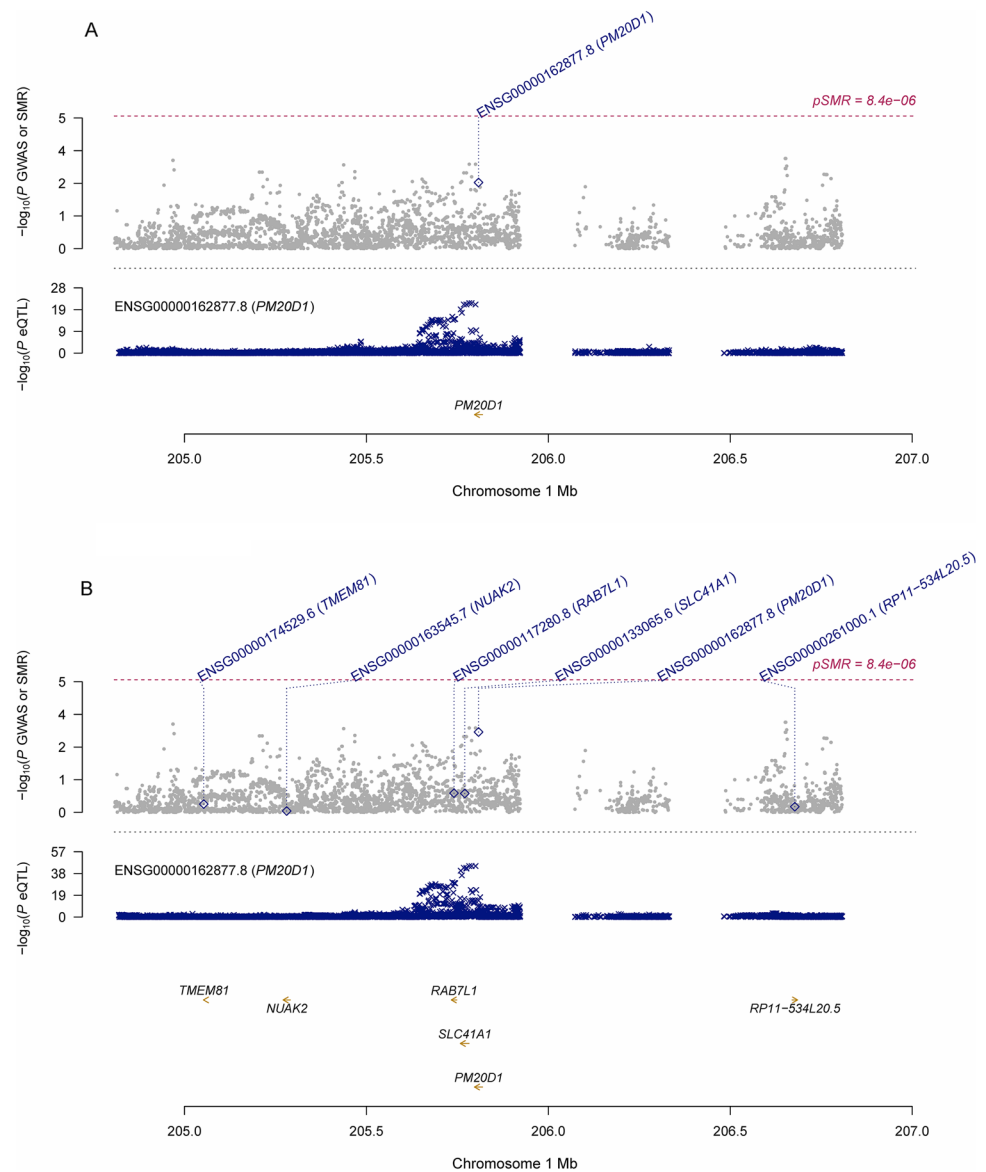
Discussion

This study integrated GWAS summarized data for PCOS and eQTL data in the SMR analysis to explore genes that were pleiotropically or potentially causal associated with PCOS. It was found that multiple genes were potentially involved in the pathogenesis of PCOS. To the best knowledge of the authors, this is the first study to explore genes

in pleiotropic association PCOS through a Mendelian randomization approach.

In the present study, *RPS26* (Ribosomal Protein S26) showed the strongest suggestive pleiotropic association with PCOS in the SMR analyses using both ovary and blood eQTL data (Table 2). *RPS26*, located on 12q13.2, is a gene encoding a ribosomal protein which is a component of the 40S subunit and belongs to the S26E family of ribosomal proteins [37]. Recent research found that *RPS26* critically regulated T-cell survival in a p53-dependent manner [38]. Knockout of *RPS26* in mouse oocytes led to retarded follicle development from pre-antral follicles to antral follicles and arrested chromatin configurations of the oocytes [39]. An earlier study using human ovary cDNA library found that *RPS26* was downregulated in PCOS ovary, compared

Fig. 3 Pleiotropic association of *PM20D1* with PCOS. **A** SMR analysis results using eQTL data for ovary. **B** SMR analysis results using eQTL data for whole blood. Top plot, grey dots represent the $-\log_{10}(P)$ values for SNPs from the GWAS of PCOS, with solid rhombuses indicating that the probes pass HEIDI test. Middle plot, eQTL results. Bottom plot, location of genes tagged by the probes. GWAS, genome-wide association studies; SMR, summary data-based Mendelian randomization; HEIDI, heterogeneity in dependent instruments; eQTL, expression quantitative trait loci; PCOS, polycystic ovary syndrome

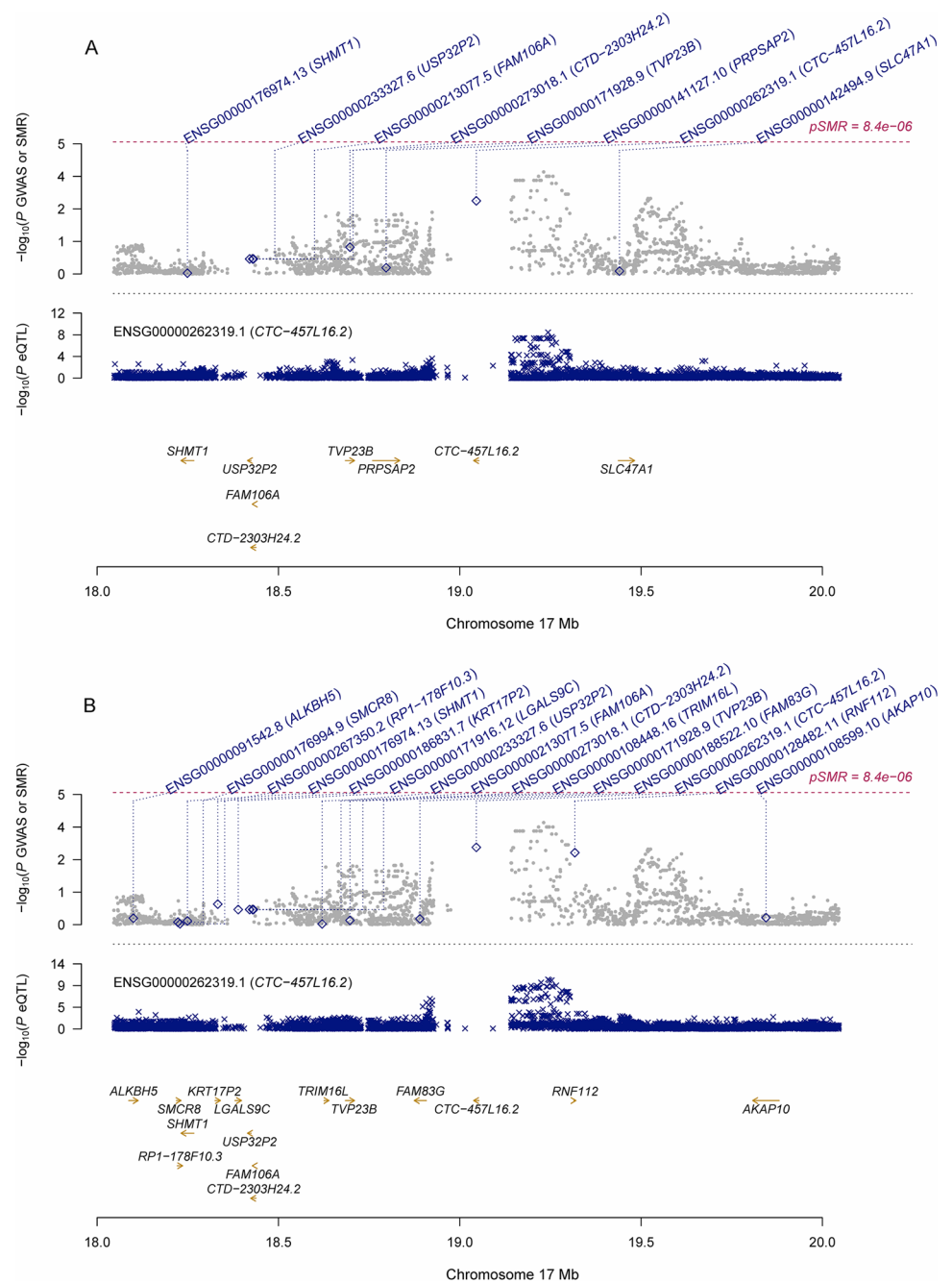


with the normal human ovary [40]. A recent study examined genes in PCOS-associated regions using a Bayesian colocalization approach (Coloc) and found seven genes, including *RPS26*, that harbored potential causal variants accounting for approximately 30% of known PCOS signals [41]. These findings, together with findings from the present study, suggested that *RPS26* might play a critical role in the etiology of PCOS and highlighted the potential of this gene as a promising target for the prevention and treatment of PCOS.

PM20D1 (Peptidase M20 Domain Containing 1) showed the second strongest suggestive pleiotropic association with PCOS in the SMR analysis using blood eQTL data and was also among the top hit genes in the SMR analysis using ovary eQTL data (Table 2). *PM20D1*, located on 1q32.1, is a gene encoding a bidirectional enzyme capable of catalyzing both the condensation of fatty acids and amino acids to

generate N-acyl amino acids and the reverse hydrolytic reaction, thereby regulating energy homeostasis [42]. Mice with increased circulating *PM20D1* had increased N-acyl amino acids in blood and improved glucose homeostasis, and *PM20D1* knockout in mice resulted in impaired glucose tolerance and insulin resistance (IR) [43]. In human adipocytes, *PM20D1* is one of the most highly upregulated genes by the antidiabetic thiazolidinedione drug rosiglitazone, suggesting a potential role of this enzyme and/or N-fatty acyl amino acids in obesity and diabetes [44]. In addition, *PM20D1* has been found to be associated with various diseases such as Parkinson's disease [45] and Alzheimer's disease [46]. To date, research is scarce on the association of *PM20D1* with PCOS. Given that PCOS is a metabolic disorder that is closely related with obesity, IR, type 2 diabetes (T2D), and metabolic syndrome [47], it is highly likely that *PM20D1*

Fig. 4 Pleiotropic association of *CTC-457L16.2* with PCOS. **A** SMR analysis results using eQTL data for ovary. **B** SMR analysis results using eQTL data for whole blood. Top plot, grey dots represent the $-\log_{10}(P$ values) for SNPs from the GWAS of PCOS, with solid rhombuses indicating that the probes pass HEIDI test. Middle plot, eQTL results. Bottom plot, location of genes tagged by the probes. GWAS, genome-wide association studies; SMR, summary data-based Mendelian randomization; HEIDI, heterogeneity in dependent instruments; eQTL, expression quantitative trait loci; PCOS, polycystic ovary syndrome

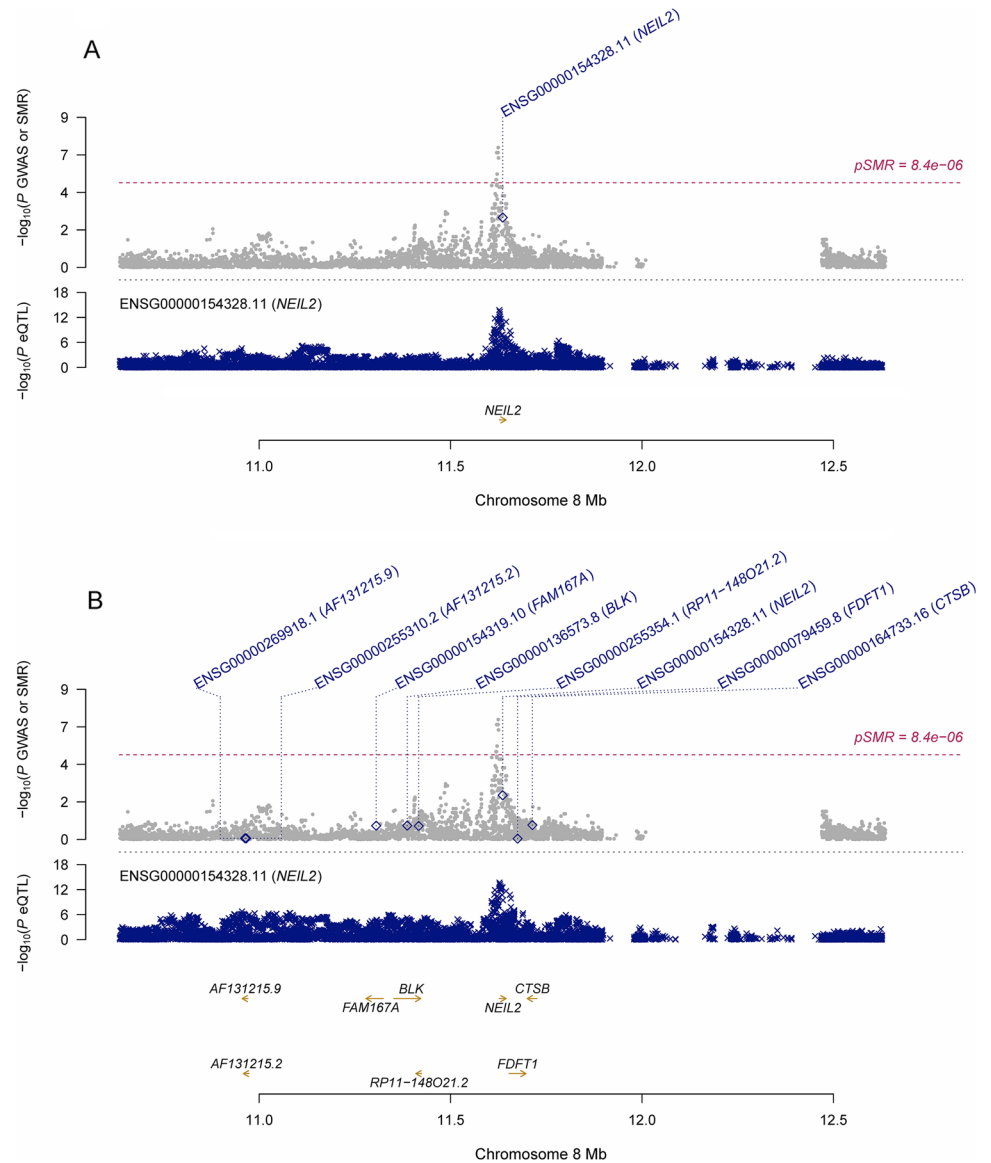


could be involved in the etiology of PCOS, and more studies are needed to explore the exact roles that *PM20D1* plays in the pathogenesis of PCOS.

NEIL2 (Nei Like DNA Glycosylase 2) showed the third strongest suggestive pleiotropic association with PCOS in the SMR analysis using ovary eQTL data and was also among the top hit genes in the SMR analysis using whole blood eQTL data (Table 2). *NEIL2*, located on 8p23.1, is a gene encoding a member of the Fpg/Nei family of DNA glycosylases. The encoded enzyme is primarily involved in DNA repair by cleaving oxidatively damaged bases and

introducing a DNA strand break via its abasic site lyase activity [48, 49]. The recent meta-analysis of GWAS on PCOS, on which the SMR analysis of the present study was based, found that the genetic variant rs804279 in *GATA4/NEIL2* showed significant association with PCOS (OR=1.14, 95% CI: 1.10–1.18; $P=3.76 \times 10^{-12}$); however, significant heterogeneity was observed across the different studies [35]. This genetic variant also showed significant association with polycystic ovarian morphology and ovulatory dysfunction [35]. Another genetic variant rs8191514 in *NEIL2* was predicted to generate a binding site for twenty

Fig. 5 Pleiotropic association of *NEIL2* with PCOS. **A** SMR analysis results using eQTL data for ovary. **B** SMR analysis results using eQTL data for whole blood. Top plot, grey dots represent the $-\log_{10}(P)$ values for SNPs from the GWAS of PCOS, with solid rhombuses indicating that the probes pass HEIDI test. Middle plot, eQTL results. Bottom plot, location of genes tagged by the probes. GWAS, genome-wide association studies; SMR, summary data-based Mendelian randomization; HEIDI, heterogeneity in dependent instruments; eQTL, expression quantitative trait loci; PCOS, polycystic ovary syndrome



transcription factors, and it was in linkage disequilibrium with the PCOS-identified genetic variant rs804279 ($R^2=0.4$, $D'=0.97$; $P<0.0001$) [50]. Moreover, 8p23.1, where the *NEIL2* is located, is also the region of *GATA4* (GATA Binding Protein 4), and knock-out of *GATA4* led to abnormal responses to exogenous gonadotropins and impaired fertility in mice [51]. It also encompasses the promoter region of *FDFT1* (Farnesyl-Diphosphate Farnesyltransferase 1), a gene encoding farnesyl-diphosphate farnesyl transferase that is involved in cholesterol biosynthesis pathway [52], thereby influencing testosterone biosynthesis. These findings, together with findings from the present study, highlight the importance of this region in association with PCOS.

This study has some limitations. The number of probes used in the SMR analyses was limited, and the sample size in the eQTL analysis was limited, especially for the eQTL

data in ovary, which may lead to reduced power in the eQTL analysis. Consequently, some important genes implicated in PCOS may have been missed. Future SMR studies with larger samples for the eQTL analysis are warranted to identify additional genes underlying the pathogenesis of PCOS. The SMR analyses were performed only in participants of European ethnicity, and future studies are needed to explore whether the findings can be generalized to other ethnicities. Moreover, this study only explored gene expression probes in association with PCOS, and it is possible that genetic variants exert their effect on PCOS through other epigenetic mechanisms, such as DNA methylation. More studies integrating multi-omics data are needed to more systematically explore the complex mechanisms underpinning PCOS.

In summary, the present SMR study integrating GWAS of PCOS and eQTL data revealed that multiple genes were

potentially involved in the pathogenesis of PCOS. Some of the genes were reported to be involved in the regulation of T-cell survival, energy homeostasis, and DNA repair. More studies are needed to examine the exact functions of these genes in the etiology of PCOS and to explore additional genes implicated in the pathogenesis of PCOS.

Author Contribution QS, JY, WF, and WY designed the study. JL and JY analyzed data and performed data interpretation. QS, YG, and JY wrote the initial draft and JY and WY contributed to subsequent revisions of the manuscript. All authors reviewed the study findings and read and approved the final version before submission.

Funding This work was supported by the Lianyungang Health and Family Planning Science and Technology Project (No. 201706). Dr. Jingyun Yang's research was supported by NIH/NIA grants (P30AG10161, R01AG15819, R01AG17917, R01AG36042, U01AG61356, and 1RF1AG064312-01).

Data Availability All data generated or analyzed during this study are publicly available as specified in the methods section of this paper. The eQTL data can be downloaded at <https://cnsgenomics.com/data/SMR/#eQTLsummarydata>, and the GWAS summarized data can be downloaded at <https://www.repository.cam.ac.uk/handle/1810/283491>.

Code Availability Not applicable.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors declare no competing interests.

References

- Bellver J, Rodriguez-Tabernero L, Robles A, Munoz E, Martinez F, Landeras J, et al. Polycystic ovary syndrome throughout a woman's life. *J Assist Reprod Genet.* 2018;35(1):25–39.
- Rotterdam EA-SPCWG. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril.* 2004;81(1):19–25.
- Teede H, Deeks A, Moran L. Polycystic ovary syndrome: a complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan. *BMC Med.* 2010;8:41.
- Azziz R, Carmina E, Chen Z, Dunaif A, Laven JS, Legro RS, et al. Polycystic ovary syndrome. *Nat Rev Dis Primers.* 2016;2:16057.
- Naz MSG, Tehrani FR, Majd HA, Ahmadi F, Ozgoli G, Fakari FR, et al. The prevalence of polycystic ovary syndrome in adolescents: a systematic review and meta-analysis. *Int J Reprod Biomed.* 2019;17(8):533–42.
- Rotterdam EA-SP. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod.* 2004;19(1):41–7.
- Legro RS, Arslanian SA, Ehrmann DA, Hoeger KM, Murad MH, Pasquali R, et al. Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab.* 2013;98(12):4565–92.
- Joham AE, Teede HJ, Ranasingha S, Zoungas S, Boyle J. Prevalence of infertility and use of fertility treatment in women with polycystic ovary syndrome: data from a large community-based cohort study. *J Womens Health (Larchmt).* 2015;24(4):299–307.
- Barnard L, Ferriday D, Guenther N, Strauss B, Balen AH, Dye L. Quality of life and psychological well being in polycystic ovary syndrome. *Hum Reprod.* 2007;22(8):2279–86.
- Wilson NA, Pena AS. Quality of life in adolescent girls with polycystic ovary syndrome. *J Paediatr Child Health.* 2020;56(9):1351–7.
- Naumova I, Castelo-Branco C, Kasterina I, Casals G. Quality of life in infertile women with polycystic ovary syndrome: a comparative study. *Reprod Sci.* 2020.
- Zhao H, Lv Y, Li L, Chen ZJ. Genetic studies on polycystic ovary syndrome. *Best Pract Res Clin Obstet Gynaecol.* 2016;37:56–65.
- Kandaraki E, Chatzigeorgiou A, Livadas S, Palioura E, Economou F, Koutsilieris M, et al. Endocrine disruptors and polycystic ovary syndrome (PCOS): elevated serum levels of bisphenol A in women with PCOS. *J Clin Endocrinol Metab.* 2011;96(3):E480–4.
- Zhang B, Zhou W, Shi Y, Zhang J, Cui L, Chen ZJ. Lifestyle and environmental contributions to ovulatory dysfunction in women of polycystic ovary syndrome. *BMC Endocr Disord.* 2020;20(1):19.
- Vink JM, Sadrzadeh S, Lambalk CB, Boomsma DI. Heritability of polycystic ovary syndrome in a Dutch twin-family study. *J Clin Endocrinol Metab.* 2006;91(6):2100–4.
- Chen ZJ, Zhao H, He L, Shi Y, Qin Y, Shi Y, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat Genet.* 2011;43(1):55–9.
- Shi Y, Zhao H, Shi Y, Cao Y, Yang D, Li Z, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet.* 2012;44(9):1020–5.
- Lee H, Oh JY, Sung YA, Chung H, Kim HL, Kim GS, et al. Genome-wide association study identified new susceptibility loci for polycystic ovary syndrome. *Hum Reprod.* 2015;30(3):723–31.
- Hayes MG, Urbanek M, Ehrmann DA, Armstrong LL, Lee JY, Sisk R, et al. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. *Nat Commun.* 2015;6:7502.
- Day FR, Hinds DA, Tung JY, Stolk L, Styrkarsdottir U, Saxena R, et al. Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nat Commun.* 2015;6:8464.
- Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet.* 2012;90(1):7–24.
- Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* 2014;23(R1):R89–98.
- Burgess S, Foley CN, Allara E, Staley JR, Howson JMM. A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nat Commun.* 2020;11(1):376.
- Wang K, Han S. Effect of selection bias on two sample summary data based Mendelian randomization. *Sci Rep.* 2021;11(1):7585.
- Thanassoulis G, O'Donnell CJ. Mendelian randomization: nature's randomized trial in the post-genome era. *JAMA.* 2009;301(22):2386–8.
- Burgess S, Timpson NJ, Ebrahim S, Davey SG. Mendelian randomization: where are we now and where are we going? *Int J Epidemiol.* 2015;44(2):379–88.
- Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci.* 2013;368(1620):20120362.

28. Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ, et al. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genomics*. 2014;15:145.
29. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48(5):481–7.
30. Liu D, Yang J, Feng B, Lu W, Zhao C, Li L. Mendelian randomization analysis identified genes pleiotropically associated with the risk and prognosis of COVID-19. *medRxiv*. 2020.
31. Yang H, Liu D, Zhao C, Feng B, Lu W, Yang X, et al. Mendelian randomization integrating GWAS and eQTL data revealed genes pleiotropically associated with major depressive disorder. *Transl Psychiatry*. 2021;11(1):225.
32. Pavlides JM, Zhu Z, Gratten J, McRae AF, Wray NR, Yang J. Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome Med*. 2016;8(1):84.
33. Carithers LJ, Moore HM. The Genotype-Tissue Expression (GTEx) Project. *Biopreserv Biobank*. 2015;13(5):307–8.
34. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204–13.
35. Day F, Karaderi T, Jones MR, Meun C, He C, Drong A, et al. Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PLoS Genet*. 2018;14(12):e1007813.
36. Zawadzki JK, Dunaif A. *Polycystic ovary syndrome*. Cambridge: Blackwell Scientific Publications; 1992.
37. Filipenko ML, Vinichenko NA, Karpova GG, Mertvetsov NP, Amaldi F. Isolation, structural analysis and mapping of the functional gene of human ribosomal protein S26. *Gene*. 1998;211(2):287–92.
38. Chen C, Peng J, Ma S, Ding Y, Huang T, Zhao S, et al. Ribosomal protein S26 serves as a checkpoint of T-cell survival and homeostasis in a p53-dependent manner. *Cell Mol Immunol* 2021.
39. Liu XM, Yan MQ, Ji SY, Sha QQ, Huang T, Zhao H, et al. Loss of oocyte Rps26 in mice arrests oocyte growth and causes premature ovarian failure. *Cell Death Dis*. 2018;9(12):1144.
40. Diao FY, Xu M, Hu Y, Li J, Xu Z, Lin M, et al. The molecular characteristics of polycystic ovary syndrome (PCOS) ovary defined by human ovary cDNA microarray. *J Mol Endocrinol*. 2004;33(1):59–72.
41. Censin JC, Bovijn J, Holmes MV, Lindgren CM. Colocalization analysis of polycystic ovary syndrome to identify potential disease-mediating genes and proteins. *Eur J Hum Genet*. 2021.
42. Long JZ, Svensson KJ, Bateman LA, Lin H, Kamenecka T, Lokurkar IA, et al. The Secreted Enzyme PM20D1 Regulates Lipidated Amino Acid Uncouplers of Mitochondria. *Cell*. 2016;166(2):424–35.
43. Long JZ, Roche AM, Berdan CA, Louie SM, Roberts AJ, Svensson KJ, et al. Ablation of PM20D1 reveals N-acyl amino acid control of metabolism and nociception. *Proc Natl Acad Sci USA*. 2018;115(29):E6937–45.
44. Benson KK, Hu W, Weller AH, Bennett AH, Chen ER, Khetarpal SA, et al. Natural human genetic variation determines basal and inducible expression of PM20D1, an obesity-associated gene. *Proc Natl Acad Sci USA*. 2019;116(46):23232–42.
45. Pihlstrom L, Rengmark A, Bjornara KA, Dizdar N, Fardell C, Forsgren L, et al. Fine mapping and resequencing of the PARK16 locus in Parkinson's disease. *J Hum Genet*. 2015;60(7):357–62.
46. Sanchez-Mut JV, Heyn H, Silva BA, Dixsaut L, Garcia-Esparcia P, Vidal E, et al. PM20D1 is a quantitative trait locus associated with Alzheimer's disease. *Nat Med*. 2018;24(5):598–603.
47. Condorelli RA, Calogero AE, Di Mauro M, La Vignera S. PCOS and diabetes mellitus: from insulin resistance to altered beta pancreatic function, a link in evolution. *Gynecol Endocrinol*. 2017;33(9):665–7.
48. Dou H, Mitra S, Hazra TK. Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *J Biol Chem*. 2003;278(50):49679–84.
49. Hazra TK, Kow YW, Hatahet Z, Imhoff B, Boldogh I, Mokkapati SK, et al. Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *J Biol Chem*. 2002;277(34):30417–20.
50. Prabhu BN, Kanchamreddy SH, Sharma AR, Bhat SK, Bhat PV, Kabekkodu SP, et al. Conceptualization of functional single nucleotide polymorphisms of polycystic ovarian syndrome genes: an in silico approach. *J Endocrinol Invest*. 2021.
51. LaVoie HA. The GATA-keepers of ovarian development and folliculogenesis. *Biol Reprod*. 2014;91(2):38.
52. Ha NT, Lee CH. Roles of farnesyl-diphosphate farnesyltransferase 1 in tumour and tumour microenvironments. *Cells*. 2020;9(11).

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