

Genomic analyses implicate hormonal and metabolic dysregulation in polycystic ovary syndrome

Received: 18 April 2024

Accepted: 11 February 2026

Published online: 23 April 2026

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Polycystic ovary syndrome (PCOS) and its underlying features remain poorly understood. In this genetic study ($n = 544,513$), we expand the number of genetic loci from 16 to 29, and additionally identify 31 associated plasma proteins. Many risk-increasing loci were associated with later age at menopause, underscoring the reproductive longevity related to an increased oocyte number and/or availability across the lifespan. Hormonal regulation in the etiology of this condition, through metabolic and reproductive features, was emphasized. The proteomic analysis highlighted metabolic biology known to be related to PCOS. A polygenic risk score (PRS) was associated with adverse cardiometabolic outcomes, with differing relevance of testosterone and body mass index in women and men. Finally, while oligo-anovulation and anovulatory infertility are features of PCOS, we observed no impact of PCOS susceptibility on childlessness. We suggest that PCOS susceptibility confers balanced pleiotropic influences on fertility in women, and life-long adverse metabolic consequences in both sexes.

Polycystic ovary syndrome (PCOS) is the most common reproductive endocrinopathy¹, with impacts across the lifespan. The diagnostic criteria require two of three features—hyperandrogenism (HA), oligo-anovulation and/or polycystic ovarian morphology (PCOM)². PCOS is the most common cause of anovulatory infertility and is associated with insulin resistance, conferring an increased risk of metabolic outcomes such as type 2 diabetes (T2D)^{3,4}. Previous large-scale genetic studies demonstrated that PCOS is a complex polygenic disorder encompassing interactions among brain, metabolic and gonadal function⁵. High body mass index (BMI) and fasting insulin levels were identified as causal risk factors for PCOS⁵. Association at the *FSHB* locus highlighted the pituitary as a driver for PCOS^{5,6}. The genetic susceptibility for later age at menopause was identified as causal for PCOS, linking PCOS etiology to the DNA damage response^{6,7}. However, the small number of identified loci has limited further exploration⁸. There are no adequately powered prospective studies of women with PCOS beyond their reproductive years. Therefore, long-term health outcomes remain unknown. In addition, understanding of the genetic risk factors for PCOS on other health outcomes in women and men is incomplete.

To address these limitations, we conducted a meta-analysis in genome-wide association study (GWAS), including data from 20,818 cases and 523,695 controls, which doubled the number of women with PCOS compared to previous GWAS⁶ (Supplementary Table 1). We assessed the identified signals with a range of phenotypes encompassing three relevant mechanisms—metabolic pathways, hormonal regulation (including the hypothalamic–pituitary–gonadal axis) and the oocyte/follicle complement. We also conducted a complementary proteomic-based analysis to further identify the biology of this condition.

Previous studies used genetic instruments to explore the causal links from a range of phenotypes to PCOS, but not the downstream impacts of PCOS⁹. Many of these are likely to be a feature of ‘common soil’ effects, where several conditions stem from the same source; in this case, the adverse metabolic or hormonal background may be common to both women and men. However, there may be conditions where PCOS has a specific, additional, adverse effect. In particular, previous work implicated shared genetic influences on male-pattern balding and PCOS⁵, but whether PCOS risk variants impact disease risk for cardiometabolic and other health conditions in men has only

✉ e-mail: cwelt@genetics.utah.edu; tugce.karaderi@sund.ku.dk; felix.day@mrc-epid.cam.ac.uk

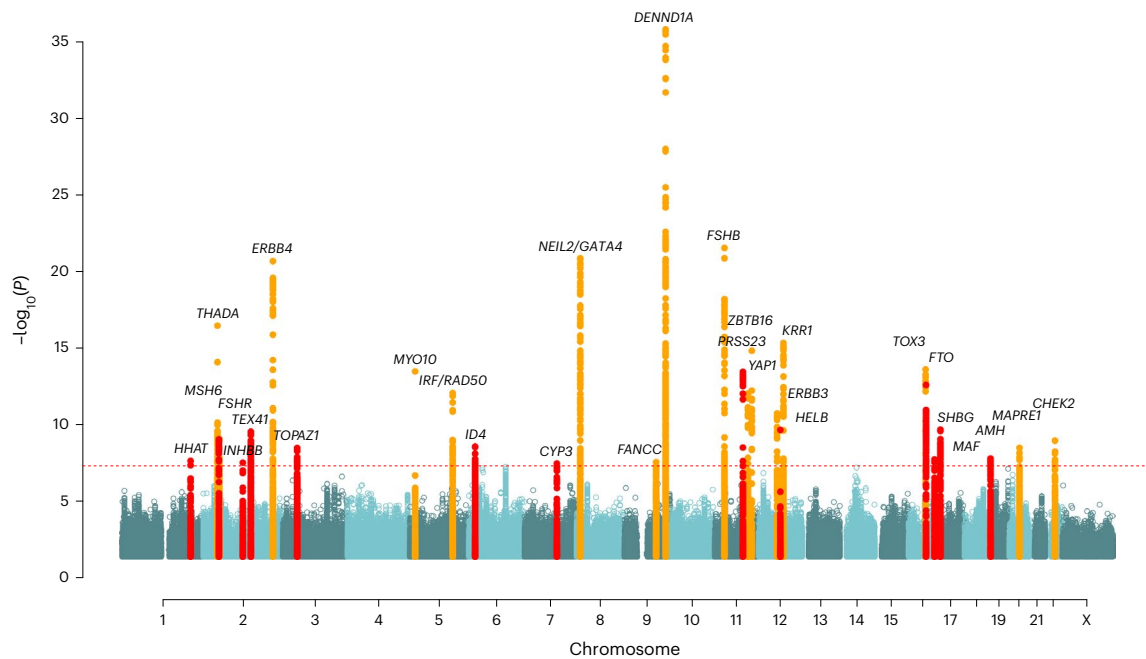


Fig. 1 | Manhattan plot showing the 29 genomic loci associated with PCOS. Two-tailed P values were generated by meta-analysis using an inverse variance weighted approach. Variants within 300 kb on either side of a genome-wide

significant signal identified in this study are highlighted in red, those previously identified are in yellow. The dotted line indicates the genome-wide significance level of $P = 5 \times 10^{-8}$. Gene names indicate the consensus PCOS gene at each locus.

been addressed to a limited extent¹⁰. Here the impact of PCOS on cardiometabolic disease, disorders of reproductive organs and mental health is examined in men and women.

Results

Genome-wide discovery for PCOS signals

We identified 29 independent loci associated with PCOS ($P < 5 \times 10^{-8}$) in the all-ancestries meta-analysis, of which 13 had not previously been reported^{5,6,11–13} (Fig. 1 and Supplementary Table 2). The majority of the cohort is from European ancestry (93%), while the rest have African American (5%), East Asian (1%) or Hispanic ancestry (1%; Supplementary Fig. 1). These GWAS signals include a variant at *FTO* (rs8047587), confirming earlier findings¹⁴, and confirms the reported effect of increasing BMI on risk of PCOS^{6,15}. Other PCOS signals have relevance to reproductive hormone pathways—*AMH* (rs732310), *INHBB* (rs6712151) and *SHBG* (rs1641518). Alongside the known European signal at *FSHB* (rs11031005), we report a signal at *FSHR* (rs13004711), replicating the association previously observed in Han Chinese women¹¹. Three of the 29 loci (*INHBB*, *NEIL2*, *DENND1A*) had evidence of secondary signals (within 500 kb of the lead signal; Supplementary Table 3).

We also performed a BMI-adjusted model in a subset of the cohorts (Supplementary Table 4 and Supplementary Fig. 2). In this analysis, only the *FTO* locus was substantially attenuated ($P = 0.019$ after adjustment; Supplementary Fig. 3). To explore whether our findings were affected by differences in the criteria used to diagnose PCOS, we stratified studies based on the case definition. There were no differences in the effect sizes of the 29 PCOS signals by case definition (Supplementary Fig. 4), and they were comparable across individual studies (Supplementary Fig. 5). We also assessed the ten PCOS signals previously reported in East Asian women¹¹ (Supplementary Table 5). Of these, all but three (variants near *C9orf3*, *INSR* and *SUMO1P1*) had statistically significant associations in our study ($P < 0.005$; for more details, see Supplementary Note). Finally, fine-mapping of the identified loci was performed, resolving the credible window for the associations (Supplementary Tables 6 and 7 and Supplementary Note).

Identifying genes of interest

We used two approaches to identify PCOS risk genes, which we call 'consensus genes'. First, we performed a literature review of all genes within 500 kb of the signals, prioritizing those with a reported link to one of the following four preselected processes: (1) reproductive function, (2) steroid metabolism and sex-hormone levels, (3) metabolic syndrome and (4) DNA damage repair. Evidence for genes linked to at least one of these processes is described in detail in Supplementary Table 8. Second, we used the GWAS-to-gene bioinformatic approach that leverages data on expression quantitative trait loci (eQTLs), protein QTLs (pQTLs), predicted deleterious variants and variant-based scoring methods to rank genes based on their causal likelihood (Supplementary Table 9)¹⁶. Findings from these two approaches were then harmonized for each PCOS signal (Supplementary Table 8).

In 16 cases, both approaches prioritized the same gene. In the other cases, there was a strong rationale for prioritizing the literature-based gene instead of the bioinformatics-identified gene. For example, the gene *SHBG* was prioritized over *ATPIB2* at rs1641518, because the variant was also associated with circulating sex-hormone binding globulin (SHBG) levels (Fig. 2). The rs732310 variant was assigned to the *AMH* gene, based on the known functions of anti-Müllerian hormone (AMH) in inhibiting recruitment of ovarian follicles from the primordial follicle pool, inhibiting follicle-stimulating hormone (FSH) sensitivity of growing follicles and regulating gonadotropin-releasing hormone (GnRH)-dependent luteinizing hormone (LH) pulsatility^{17,18}. Although rs732310 shows no association with AMH levels (Fig. 2), the GWAS data used for the AMH analysis were derived from normo-ovulatory women¹⁹, and may not reflect variations in AMH levels in women with PCOS, in whom the expression pattern of AMH differs²⁰.

Relationship of the identified loci with other phenotypes

Several of the 29 PCOS variants had previously been associated in GWASs for age at menopause (14 variants), age at menarche (6), female testosterone levels (7), BMI (8) and male-pattern baldness (2; Supplementary Table 10). We annotated our signals using publicly available GWAS results, focusing on the relationships between these

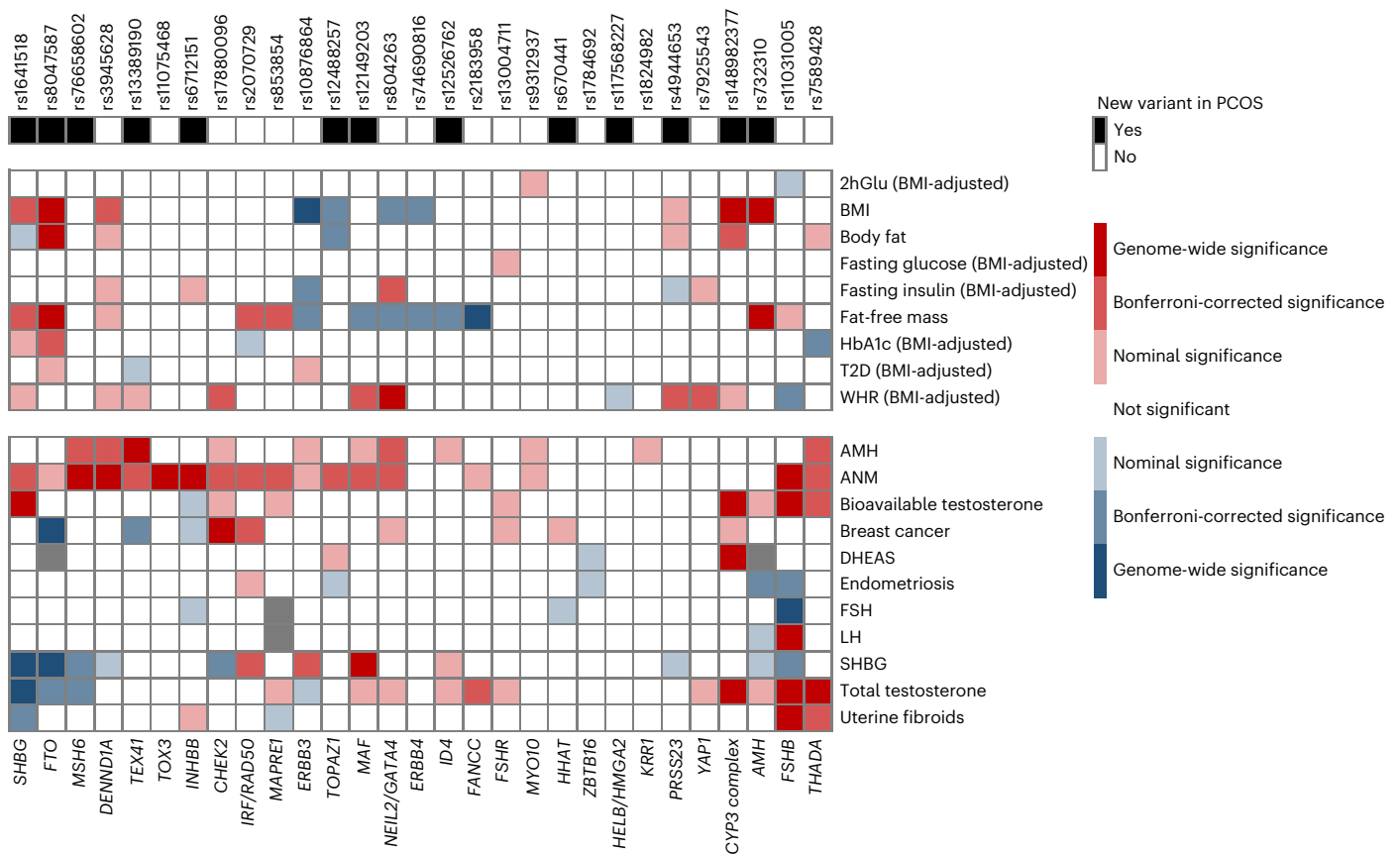


Fig. 2 | Heatmap of GWAS associations for the 29 PCOS loci with other relevant traits. Direction and the strength of association between the 29 PCOS risk-increasing alleles (top) with 20 other relevant traits and diseases with available GWAS summary statistics (right). Color coding indicates strength and direction (z scores) of associations—positive (red) and negative (blue). In the upper row of the heatmap, new PCOS loci identified in this study (black boxes)

and previously reported loci (white boxes) are shown with corresponding lead variants. The top and bottom of the heatmap show metabolic and reproductive phenotypes, respectively. Gray boxes indicate missing variant-trait association data. Genes are presented in the lower x axis as ‘consensus gene’. DHEAS, dehydroepiandrosterone sulfate.

variants and a range of metabolic, reproductive and hormonal phenotypes. All 29 PCOS signals showed at least nominal significance with one or more metabolic, reproductive and/or hormonal trait(s), providing evidence to support the multifactorial etiology and comorbidities (Fig. 2 and Supplementary Tables 11 and 12).

In validation of the substantial overlap between signals for PCOS and age at natural menopause (ANM), eight signals showed evidence of colocalization (posterior probability ≥ 0.75). PCOS signals at *FSHB*, *DENND1A*, *TOX3*, *RAD50* and *MAF/MAFTRR* were associated with ANM (Supplementary Table 13); conversely, the reported ANM signals at *FSHB*, *DENND1A*, *CASC22*, *BMP4*, *PPARG* and *MAF/MAFTRR* were associated with PCOS (Supplementary Table 14). At all eight colocalized signals, the PCOS risk-increasing allele conferred later ANM. The *FSHB* signal has been well described to affect other reproductive phenotypes, including age at menarche²¹ and dizygous twinning²². Other shared loci are also related to breast cancer pathways—*TOX3*, *CASC22* and *RAD50*. Interestingly, the shared *PPARG* locus suggests an effect of metabolic pathways independent of BMI-related pathways.

A number of PCOS-associated loci also showed strong effects on hormone levels, particularly SHBG levels (including rs1641518 near *SHBG*). Approximately 30% (8/29) PCOS risk-increasing alleles had nominal association with lower SHBG levels. There were two loci in which PCOS risk alleles were associated with lower SHBG and higher total testosterone, suggesting a relationship driven by higher androgens; whereas three other loci, including *SHBG* and *FTO*, had lower SHBG and lower total testosterone, suggesting an

SHBG-mediated effect (Fig. 2). Three of the signals had evidence of colocalization with *SHBG* levels, including the *CYP3* complex, *FTO* and *FSHB* (Supplementary Table 15). The signal at *FTO* is likely due to the established links between increasing BMI and decreasing *SHBG*²³. The *CYP3* complex metabolizes oestradiol and testosterone²⁴, with mouse knockouts showing substantially increased free testosterone levels²⁵. The *FSHB* locus is associated with increased LH, which stimulates androstenedione, and, therefore, testosterone LH production²⁶, which would lower *SHBG* levels.

There was additional evidence for the role of PCOS-associated loci impacting the regulation of gonadotropins and the functional ovarian reserve, that is, the pool of recruitable oocytes/follicles across a lifespan. The PCOS risk-increasing allele at the *FSHB* locus was associated with lower levels of FSH, and *INHBB* was nominally associated with lower FSH. Although the *FSHR* variant was not associated with FSH levels in 2,913 Europeans, the rs2268361-T variant found in Han Chinese women, which shows some evidence of linkage disequilibrium (LD; $R^2 = 0.3$ in Europeans) with the variant reported here, was associated with lower FSH levels^{27,28} (Fig. 2 and Supplementary Table 11). *FSHR* has also been linked to twinning rates and FSH levels using gene-based approaches²². At five loci, PCOS risk alleles were associated with higher AMH levels ($P < 0.0017$). All five loci overlap with the loci related to ANM and are associated with higher AMH. AMH is measured clinically to indicate the number of growing follicles, as a proxy of the functional ovarian reserve, and its concentrations are strongly related to age at menopause¹⁸. This suggests their involvement in the establishment

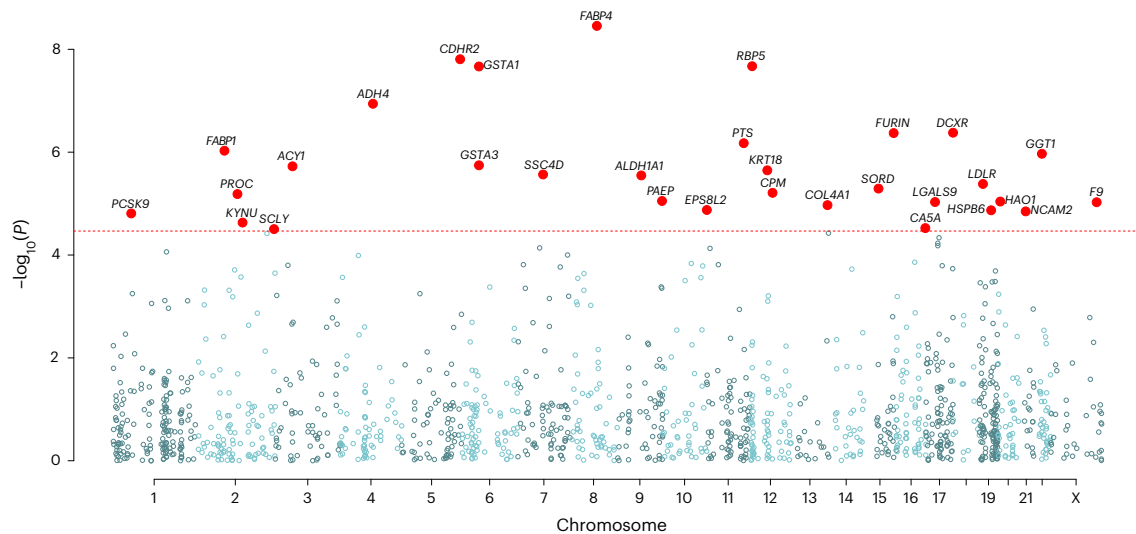


Fig. 3 | Manhattan plot showing the genomic positions of the genes for the 31 plasma proteins associated with ovarian dysfunction. Ovarian dysfunction was defined as the ICD-10 code E28, the supracategory that includes PCOS. Gene names and locations for proteins reaching significance are shown at the sentinel red dots. *P* values are two-tailed and come from linear regression analyses. The dotted line indicates a Bonferroni-corrected level of significance of $P = 3.4 \times 10^{-5}$. Significant proteins are (gene–protein) as follows: *ACY1*–aminoacylase 1; *ADH4*–alcohol dehydrogenase 4 (class II), pi polypeptide; *ALDH1A1*–aldehyde dehydrogenase 1 family member A1; *CASA*–carbonic anhydrase 5A; *CDHR2*–cadherin-related family member 2; *COL4A1*–collagen type IV α 1 chain; *CPM*–carboxypeptidase M; *DCXR*–dicarbonyl and L-xylulose reductase; *EPS8L2*–EPS8 signaling adaptor L2; *F9*–coagulation factor IX; *FABP1*–fatty acid-binding

protein 1; *FABP4*–fatty acid-binding protein 4; *FURIN*–furin, paired basic amino acid cleaving enzyme; *GGT1*– γ -glutamyltransferase 1; *GSTA1*–glutathione S-transferase α 1; *GSTA3*–glutathione S-transferase α 3; *HA01*–hydroxyacid oxidase 1; *HSPB6*–heat shock protein family B (small) member 6; *KRT18*–keratin 18; *KYNU*–kynureninase; *LDLR*–low-density lipoprotein receptor; *LGALS9*–galectin 9; *NCAM2*–neural cell adhesion molecule 2; *PAEP*–progesterone associated endometrial protein; *PCSK9*–proprotein convertase subtilisin/kexin type 9; *PROC*–protein C, inactivator of coagulation factors Va and VIIIa; *PTS*–6-pyruvoyltetrahydropterin synthase; *RBP5*–retinol binding protein 5; *SCLY*–selenocysteine lyase; *SORD*–sorbitol dehydrogenase; *SSC4D*–scavenger receptor cysteine rich family member with four domains.

and preservation of the functional ovarian reserve as a fundamental element of PCOS.

Protein-based analysis

The levels of 31 plasma proteins were phenotypically associated with the International Classification of Diseases (ICD) 10 code E28, ovarian dysfunction, which includes PCOS (all $P < 3.4 \times 10^{-5}$; Fig. 3 and Supplementary Table 16). These included recognized metabolic disease-associated proteins such as PCSK9, LDLR, FURIN, FABP1 and FABP4. Other associated proteins metabolize hydroxysteroids, retinol and lipids, including ALDH1A1 and ADH4, that may regulate the metabolic response to a high-fat diet^{29,30}. Other proteins are potential contributors to diabetes and metabolic disease, such as GGT1 (ref. 31), or their complications, SORD³². Finally, there were enzymes important for the biosynthesis of progesterone or testosterone, GSTA1 and GSTA3³³, and fertilization and implantation, PAEP³⁴. We used GProfiler to perform a combined pathway-based analysis using proteins drawn from either the protein-based approach or the GWAS-associated loci (Supplementary Fig. 6 and Supplementary Tables 17 and 18)³⁵. The two discovery methods highlighted different biology, with only one pathway, ‘benzaldehyde dehydrogenase activity’, driven by both proteomic and GWAS findings (which included *FANCC*, a known DNA damage repair gene). The results show evidence of enrichment for pathways representing androgen binding, ovarian follicle development, neuregulin receptor activity and the PCSK9-LDLR complex consistent with the lipid abnormalities of PCOS^{36,37}.

To build potentially causal pathways, we identified plasma proteins whose levels were associated with each of our PCOS GWAS variants, resulting in 299 proteins ($P < 3.4 \times 10^{-5}$; Supplementary Table 19). The PCOS signals at *ERBB3*, *ERBB4* and *ZBTB16* were associated with plasma levels of their encoded proteins, providing support for these consensus genes. The PCOS signal at *SHBG* was associated with levels of TNSF12 and TNSF13, encoded by genes in the same region, and related

to apoptosis and regulation of steroidogenesis³⁸. One PCOS signal at *RAD50/IRF1* accounted for 199 protein associations. This signal lies in a region on chromosome 2 that contains many immune-related genes—*IRF1*, *IL4*, *IL5* and *IL13*—and likely has a widespread impact on the plasma proteome. Another signal, overlapping the established obesity signal at *NEIL2/GATA4*, was associated with 29 plasma proteins. Other associated proteins included leptin, which is higher in obesity; PPY, a regulator of food intake; and several fatty acid-binding proteins, which regulate fatty acid uptake in adipose cells³⁹. We resolved these to nine variant–protein pairs, and compared the variance explained in PCOS and in the respective protein levels (Supplementary Note and Supplementary Table 20), and for proteins linked to the signal at *FTO* BMI, similar to the logic of Steiger filtering⁴⁰. All five *FTO*-associated proteins, NCAM2 associated with *ERBB3* and IGFBP2 associated with *MAF* appeared to be upstream (that is, more likely to be determinants, or common soil) of PCOS. The two other proteins associated with *MAF*, *CDHR2* and *CPM* may have their levels altered as a consequence of PCOS (Supplementary Fig. 7).

Inferring causal impacts of PCOS on other comorbidities

The impact of PCOS on fertility, metabolic disease and mental health is well known, but few studies have used a genetic approach to uncover additional comorbidities⁴¹. Previous genetic causal modeling, using Mendelian randomization (MR) approaches, has shown that aspects of metabolic syndrome traits are risk factors for PCOS⁶. However, those MR studies did not determine whether PCOS had an effect on broader health status. Therefore, we calculated a PRS for PCOS comprising -1.1 million genetic variants to explore the likely causal effect of PCOS on a number of other outcomes.

To identify phenotypes that may share genetic influences with PCOS and to determine potential differences between women and men, we performed PRS-based analyses in the UK Biobank study (independent of the discovery dataset). We used the PRS-CS software, a Bayesian

Table 1 | PCOS PRS associations with traits and diseases of interest

Trait	Women or men	n_{total} or $n_{\text{cases/controls}}$	Age-adjusted PRS		BMI-adjusted PRS		Heterogeneity	
			β (s.e.) or OR (95% CI)	P	β (s.e.) or OR (95% CI)	P	I^2	P for heterogeneity between sexes for age-adjusted PRS
Cardiometabolic								
BMI	Women	206,214	0.296 (0.012)	4×10^{-143}	0.037 (0.012)	0.0017	91%	0.0007
	Men	175,708	0.243 (0.010)	1×10^{-121}	0.027 (0.010)	0.009		
WHR (adjusted BMI)	Women	206,444	0.002 (0.0001)	2×10^{-42}	0.002 (0.0001)	1×10^{-37}	99%	<0.0001
	Men	176,000	0.0008 (0.0001)	4×10^{-11}	0.0007 (0.0001)	5×10^{-8}		
Obesity	Women	47,124/159,733	1.12 (1.10, 1.13)	1×10^{-89}	1.02 (1.01, 1.03)	3×10^{-4}	0%	0.6
	Men	44,499/131,861	1.12 (1.11, 1.13)	1×10^{-87}	1.02 (1.01, 1.03)	3×10^{-4}		
T2D	Women	6,713/117,099	1.13 (1.10, 1.16)	9×10^{-21}	1.06 (1.03, 1.09)	2×10^{-5}	17%	0.3
	Men	11,730/97,146	1.11 (1.09, 1.14)	2×10^{-24}	1.05 (1.03, 1.07)	5×10^{-6}		
CAD	Women	4,716/202,141	1.09 (1.05, 1.12)	7×10^{-8}	1.06 (1.03, 1.10)	5×10^{-5}	82%	0.02
	Men	15,857/160,503	1.04 (1.02, 1.06)	2×10^{-6}	1.01 (0.99, 1.03)	0.3		
HbA1c (mmol mol ⁻¹)	Women	197,172	0.123 (0.013)	1×10^{-21}	0.069 (0.013)	1×10^{-7}	88%	0.004
	Men	168,088	0.187 (0.018)	2×10^{-25}	0.076 (0.018)	3×10^{-5}		
HDL (mmol l ⁻¹)	Women	179,166	-0.013 (0.001)	2×10^{-45}	-0.005 (0.001)	4×10^{-8}	92%	0.0004
	Men	155,355	-0.008 (0.001)	2×10^{-23}	-0.002 (0.001)	0.01		
Triglycerides (mmol l ⁻¹)	Women	197,011	0.025 (0.002)	1×10^{-38}	0.016 (0.002)	2×10^{-16}	19%	0.3
	Men	168,122	0.029 (0.003)	2×10^{-24}	0.019 (0.003)	2×10^{-11}		
Hormonal/reproductive								
PCOS	Women	1,003/205,849	1.42 (1.33, 1.52)	9×10^{-27}	1.27 (1.19, 1.36)	2×10^{-13}	-	-
SHBG (nmol l ⁻¹)	Women	177,167	-1.593 (0.075)	6×10^{-99}	-1.102 (0.076)	2×10^{-47}	99%	<0.0001
	Men	154,245	-0.747 (0.042)	3×10^{-69}	-0.614 (0.043)	8×10^{-47}		
Testosterone (nmol l ⁻¹)	Women	164,553	0.014 (0.002)	4×10^{-19}	0.012 (0.002)	5×10^{-14}	100%	<0.0001
	Men	166,673	-0.144 (0.009)	2×10^{-53}	-0.097 (0.009)	5×10^{-25}		
FAI	Women	148,849	0.097 (0.005)	4×10^{-71}	0.068 (0.006)	2×10^{-34}	97%	<0.0001
	Men	153,352	0.296 (0.036)	1×10^{-16}	0.297 (0.036)	1×10^{-16}		
Breast cancer	Women	11,035/195,822	1.03 (1.01, 1.05)	0.008	1.03 (1.01, 1.05)	0.006	-	-
Age at menopause	Women	118,309	0.167 (0.015)	5×10^{-30}	0.157 (0.015)	5×10^{-26}	-	-
Childlessness ^a	Women	39,248/167,346	0.98 (0.97, 1.00)	0.009	0.99 (0.98, 1.00)	0.04	0%	0.98
	Men	37,567/137,139	0.98 (0.97, 1.00)	0.01	0.99 (0.98, 1.01)	0.4		
Other								
Depression ^a	Women	15,711/32,761	1.04 (1.02, 1.07)	1×10^{-5}	1.03 (1.01, 1.05)	0.002	0%	0.5
	Men	8,802/34,069	1.03 (1.01, 1.06)	0.008	1.02 (1.00, 1.05)	0.1		
Asthma ^{a,b}	Women	29,117/135,476	1.02 (1.01, 1.03)	0.003	1.00 (0.99, 1.01)	0.9	33%	0.2
	Men	21,684/123,136	1.01 (0.99, 1.02)	0.3	0.99 (0.98, 1.01)	0.3		

^aControlled for Townsend deprivation index in addition to standard covariates. ^bControlled for smoking status (yes/no) in addition to standard covariates. We calculated the PCOS PRS in women and men in the UK Biobank. Then, we tested for association between PCOS PRS and phenotypes of interest. We applied Bonferroni correction for multiple testing for associations with 17 phenotypes; hence, associations with $P < 0.003$ are considered statistically significant. Statistically significant heterogeneity between the sexes was considered when $I^2 > 80\%$ and Cochran's Q P value for heterogeneity < 0.004 (0.05/14 phenotypes analyzed). HbA1c, hemoglobin A1c; HDL, high-density lipoprotein.

regression framework that weights the effect size of each variant by the strength of its association (P value) in the GWAS meta-analysis⁴² to calculate a PRS. This was standardized; effect sizes (β or odds ratio (OR)) in Table 1 are reported per 1 s.d. increase in the PRS. To validate the score, we confirmed that it was associated with PCOS in women in the UK Biobank ($P = 9 \times 10^{-27}$) and that the odds of PCOS increased across increasing quintiles of the PRS (Supplementary Fig. 8). As expected, there was also a strong association between a higher PCOS PRS and an increased BMI in both women and men (Table 1).

Because increased BMI is a risk factor for both PCOS and many of the expected metabolic comorbidities, we assessed the

BMI-independent causal relationship between PCOS and metabolic outcomes in the following two additional analyses: (1) we included measured BMI as a covariate and (2) we generated and tested a second BMI-adjusted PRS. The association of the BMI-adjusted PRS with baseline BMI in the UK Biobank was substantially attenuated in women and men (Table 1 and Supplementary Table 21). The PRS analyses were replicated using an additional PRS tool LDpred2, which provided consistent results. In both women and men, a higher PCOS PRS was associated with increased risk of coronary artery disease (CAD), T2D and obesity (Table 1). It was also associated with cardiometabolic risk factors—higher waist-to-hip (WHR) ratio adjusted for BMI (WHR_{adjBMI}),

higher HbA1c, higher triglycerides and lower high-density lipoprotein cholesterol. Regarding hormonal/reproductive risk factors, a higher PRS was associated with lower SHBG levels and higher free androgen index (FAI) in both sexes (Table 1). These associations were also seen in the BMI-adjusted model analyses.

We tested for the presence of sex differences in the effects of the PRS on (1) cardiometabolic outcomes, where different effects of sex were observed for BMI, WHR_{adjBMI} , HbA1c and high-density lipoprotein cholesterol; and (2) hormonal/reproductive outcomes, where differential sex-related effects were observed for SHBG, total testosterone and FAI (Table 1). There has been increasing interest in associations between reproductive diseases and mental health^{43,44}, but we found only a nominal association between the PRS and depression in women after BMI adjustment.

We used MR to provide additional evidence of causality (Supplementary Tables 22 and 23). While MR provides more robust causal inference, there are caveats—outcome data are often in men and women combined, and with only 29 variants, this analysis is likely underpowered. Most of the associations with nonreproductive phenotypes were not significant, supporting the idea of an underpowered analysis. Interestingly, the MR analysis showed associations with two noncardiometabolic outcomes—depression and asthma—with no attenuation when controlling for BMI.

Pleiotropic effects of PCOS on reproductive outcome

Consistent with the overlap across signals for PCOS and ANM, PCOS susceptibility was associated with later ANM in both PRS (Table 1) and MR analyses (Supplementary Table 23). We also observed an apparent effect of susceptibility to later ANM on higher PCOS risk (Supplementary Table 23), indicating a bidirectional relationship. Given the importance of DNA repair as a mechanism that regulates ANM, we tested whether the same pathways also contributed to PCOS. We performed MR analyses for PCOS with reported variants for ANM stratified by a previous annotation of a DNA-repair pathway⁷. Both strata of menopause variants indicated an effect of later ANM on higher risk of PCOS (Supplementary Table 23). However, the estimated effect of ANM was larger using non-DNA-repair ANM variants than when using DNA-repair variants (for the difference between estimates, $P = 1.5 \times 10^{-6}$; Supplementary Fig. 9), suggesting a role for sex-hormone-related pathways common to PCOS and ANM. This is highlighted by the shared signal at *FSHB* associated with lower FSH levels, higher PCOS risk and later age at menopause.

The subset of DNA-repair-related ANM variants that are associated with PCOS has selective effects. For example, ANM variants at *BRCA1* and *BRCA2*, genes involved in homologous recombination repair of double-strand DNA breaks, are robustly associated with an earlier ANM but not with PCOS ($P = 0.94$ and $P = 0.11$, respectively). Conversely, *CHEK2*, known to maintain DNA integrity through checkpoint control⁴⁵, is associated with a later ANM, greater risk of PCOS and higher serum AMH levels in PCOS^{7,18}.

Given that PCOS is one of the most common conditions resulting in reduced fertility, the PCOS PRS showed unexpected nominal associations with lower risk of childlessness, although not confirmed in the MR analysis. We therefore examined the impact of PCOS susceptibility on eight infertility phenotypes in the Copenhagen Hospital Biobank (CHB)⁴⁶, applying Bonferroni correction for multiple testing (0.05/8 phenotypes analyzed = $P < 0.006$; Supplementary Table 24). A higher PCOS PRS was associated with increased risk of infertility in women (OR = 1.03, $P = 0.02$ after age adjustment; OR = 1.04, $P = 0.00011$ after age and BMI adjustment). The stronger association after adjustment for BMI suggests that, in addition to affecting fertility through BMI, PCOS also affects fertility through BMI-independent mechanisms, such as HA.

Interestingly, the PCOS PRS was associated with an increased number of oocytes aspirated during in vitro fertilization treatment ($\beta = 0.025$, $P = 1 \times 10^{-4}$ after age adjustment; $\beta = 0.027$, $P = 2 \times 10^{-3}$ after

age and BMI adjustment). In a separate dataset of 812 women, two of our loci, *ZBTB16* and *SHBG*, were associated with larger ovarian volume, which correlates with oocyte/follicle number and is a symptomatic presentation of PCOS⁴⁷. These findings suggest a greater available oocyte pool in PCOS.

There was also support for the hypothesis that some genetic PCOS susceptibility might exhibit a balancing pleiotropy effect on reproductive success. We assessed the links from PCOS to age at first birth⁴⁸, age at last birth⁴⁸, childlessness⁴⁹ and number of children⁴⁹ using publicly available GWAS datasets (Supplementary Table 23). There were no apparent associations with childlessness or number of children. The latter result was confirmed in the CHB data, in which there was no association between the PCOS PRS and completed family size ($P = 0.07$) or rates of pregnancy ($P = 0.25$; Supplementary Table 24). However, there was a nominally significant association with later age at last birth when data were controlled for age and BMI ($P = 0.01$). We further assessed the association between PCOS and age at last live birth in the UK Biobank in 1,003 women with PCOS and 205,849 controls. PCOS was associated with later age at last birth ($\beta = 0.46$ years, $P = 0.04$ after age adjustment; $\beta = 0.81$ years, $P = 0.0003$ after age and BMI adjustment). These results could be explained by a longer reproductive window or by a shifting of the window, and the lack of association with childlessness indicates that some compensatory mechanisms exist.

Discussion

This study expands the number of PCOS genome-wide significant loci from 16 to 29. The locus at *FTO*, well established for obesity, highlights the link among PCOS, metabolic syndrome and obesity. Other signals at *SHBG*, *FSHR* (associated in a European GWAS) and the *CYP3* complex highlight hormonal regulation in the etiology of PCOS. Alongside these results, we also present proteins that are associated with ovarian dysfunction. The protein associations confirm some candidate genes at GWAS loci (*ERBB3*, *ERBB4* and *ZBTB16*).

The identified loci underscore the sex hormonal origins of PCOS. While our literature-based method might be biased toward particular pathways, there was clear evidence of important biology. We identified signals at *FSHB*⁵⁰ and *FSHR*¹¹, highlighting the role of pituitary gonadotrophs (LH and FSH) in ovarian stimulation. Other consensus genes *SHBG*, *INHBB*, *AMH* and *TEX41*—this last robustly associated with AMH levels—also point to hormones related to ovarian folliculogenesis, and feedback on the hypothalamic–pituitary–gonadal axis. Inhibin B is secreted by granulosa cells of small to large antral follicles and inhibits FSH release, ensuring the growth of one dominant follicle^{51,52}, and the *INHBB* variant was nominally, inversely associated with FSH levels. Circulating AMH levels reflects the number of growing small antral follicles and AMH reduces FSH sensitivity of growing follicles¹⁸. Furthermore, AMH inhibits aromatase activity at the level of a growing follicle and increases LH-dependent GnRH pulsatility at the hypothalamus⁵³. Finally, SHBG is a binding protein for androgens, thereby regulating free and bioavailable androgen levels³⁴. In men, there is a corresponding increase in FAI that explains the connection between PCOS variants and male-pattern balding⁶. In summary, PCOS risk is affected by a number of classical and well-established sex hormonal pathways.

Many of the identified PCOS loci overlap with those associated with age at menopause with PCOS risk alleles conferring later ANM. Two possible mechanisms could explain, perhaps in tandem, the links between these two phenotypes. First, several overlapping variants are related to genes linked to DNA-repair mechanisms such as *MSH6*, *CHEK2* and *RAD50*. The partitioned MR analysis suggested that both DNA damage repair and non-DNA damage repair (predominantly hormonal pathways) were causal for PCOS, but the latter had a stronger influence. While ANM is thought to be impacted by a range of pathways linked to DNA repair, the signals shared with PCOS might be related to more specific mechanisms, such as *CHEK2*, where the effect is to have DNA-damaged oocytes persist for longer⁷. Thus, there may be follicles

with DNA-damaged oocytes that remain in the ovary because the DNA checkpoint removal mechanism failed. In PCOS, this may reduce oocyte atresia, leading to continuous AMH expression and thereby stronger inhibition of primordial follicle recruitment (associated with later ANM) and reduced FSH sensitivity (contributing to the polycystic ovary morphology seen in PCOS)^{17,18}. In addition, PCOS was not associated with *BRCA1* or *BRCA2* variants, which appear to influence earlier ANM based on less functional DNA-repair mechanisms and potential loss of damaged oocytes⁷.

Second, changes in hormonal levels may increase follicle numbers as demonstrated for increased androgen levels⁵⁵ and activin, the product of two inhibin β subunits⁵⁶. It is also possible that hormone levels reduce depletion of the primordial follicle pool, causing a later end to the reproductive window, as has been demonstrated for the variants causing lower FSH levels⁵⁷. Moreover, the variant in *FSHB* was linked to less follicle selection across the reproductive lifespan, potentially leading to a greater follicle pool consistent with PCOS. Similarly, increased serum AMH levels, seen in PCOS patients, reduce the rate of primordial follicle recruitment and may thereby slow follicle pool depletion, leading to later menopause^{17,18}. Observational studies suggest that women with PCOS or PCOS symptoms have children as often as asymptomatic women, with reproductive success in the long term^{58–61}. In both our PCOS PRS and epidemiological analysis of women with PCOS in the UK Biobank, there was a suggestion of later age at last birth and this effect was also seen in the CHB data. This is consistent with a previous finding that a longer or shifted window of reproduction was required for the same cumulative family size in women with PCOS⁶¹. The main cause of infertility in PCOS is irregular ovulation. The relative infertility at younger ages may be balanced by improved ovulations at later ages^{62,63}.

In summary, the new loci contain risk genes expected to increase the follicle complement in PCOS^{64,65}. This finding supports the Rotterdam diagnostic criteria for PCOS, highlighting PCOM (number of growing small antral follicles), HA (hormone regulation), and related irregular ovulation and menses as primary etiologic features of PCOS².

The score-based analyses stressed the link to metabolic diseases, with a number of strong associations between PCOS and clinical endpoints, mirroring observed associations⁶⁶, and previous studies^{10,40}. A higher PCOS PRS was associated with higher BMI in both women and men; thus, much of the effect on cardiometabolic diseases seen in the BMI-unadjusted models is through the ‘common soil’ impact of BMI. However, many of the associations remained significant in women (although substantially attenuated), and not in men, after controlling for BMI. These female-specific effects, unrelated to BMI, suggest a shared causal factor between PCOS and metabolic disease. A plausible mechanism is that higher androgen levels in PCOS are a risk factor for CAD^{67–69}. Although testosterone and other androgens decrease to the same level as in controls after menopause, the continued lower SHBG in women with PCOS after menopause and the lasting impact of androgens during reproductive age on physiology may result in long-term increased CAD risk^{70,71}.

The proteins associated with reproductive dysfunction stress the links between reproductive phenotypes and the metabolic syndrome. Associations were seen with classical adiposity and metabolic proteins, including leptin and furin. Other associated proteins are vital to cholesterol metabolism, such as PCSK9 and the low-density lipoprotein receptor, which are important for both cardiovascular risk and steroidogenesis. There were also proteins that contribute to the metabolic response to a high-fat diet. It is important to consider that most of the women in whom the protein-based analysis was done were assessed after the end of their reproductive window. The data again implicate lower SHBG and higher free androgen levels in PCOS after menopause⁷⁰, and potentially sustained effects of HA even after the reproductive years. Thus, these results and the score-based analyses together suggest that there is an ongoing, adverse pattern of cardio-metabolic health in women with a genetic risk for PCOS.

Conclusions

Here we identify genetic regions and proteins associated with PCOS. The genomic loci appear to primarily implicate hormonal pathways as the causal factors for PCOS, while the proteins stress the common factors that influence PCOS and metabolic disease, particularly pathways related to increased BMI. Our findings highlight important links between PCOS and T2D and CAD, through mechanisms that are related to and also independent of adiposity. We also expand our understanding of the factors affecting the ovarian follicle complement on the condition, including both hormonal influences and specific DNA-repair mechanisms, and their role in PCOS. We also demonstrate some evidence of balanced pleiotropy conferred by PCOS genetic susceptibility that maintains the high prevalence of PCOS in the population.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-026-02543-9>.

References

1. Knochenhauer, E. S. et al. Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States: a prospective study. *J. Clin. Endocrinol. Metab.* **83**, 3078–3082 (1998).
2. Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum. Reprod.* **19**, 41–47 (2004).
3. Ehrmann, D. A., Barnes, R. B., Rosenfield, R. L., Cavaghan, M. K. & Imperial, J. Prevalence of impaired glucose tolerance and diabetes in women with polycystic ovary syndrome. *Diabetes Care* **22**, 141–146 (1999).
4. Legro, R. S., Kunesman, A. R., Dodson, W. C. & Dunaif, A. Prevalence and predictors of risk for type 2 diabetes mellitus and impaired glucose tolerance in polycystic ovary syndrome: a prospective, controlled study in 254 affected women. *J. Clin. Endocrinol. Metab.* **84**, 165–169 (1999).
5. Day, F. R. et al. Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nat. Commun.* **6**, 8464 (2015).
6. Day, F. et al. Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PLoS Genet.* **14**, e1007813 (2018).
7. Ruth, K. S. et al. Genetic insights into biological mechanisms governing human ovarian ageing. *Nature* **596**, 393–397 (2021).
8. Vink, J. M., Sadrzadeh, S., Lambalk, C. B. & Boomsma, D. I. Heritability of polycystic ovary syndrome in a Dutch twin-family study. *J. Clin. Endocrinol. Metab.* **91**, 2100–2104 (2006).
9. Zhu, T. & Goodarzi, M. O. Causes and consequences of polycystic ovary syndrome: insights from Mendelian randomization. *J. Clin. Endocrinol. Metab.* **107**, e899–e911 (2022).
10. Zhu, J. et al. Evidence from men for ovary-independent effects of genetic risk factors for polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.* **107**, e1577–e1587 (2022).
11. Shi, Y. et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat. Genet.* **44**, 1020–1025 (2012).
12. Hayes, M. G. et al. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. *Nat. Commun.* **6**, 7502 (2015).
13. Tyrmi, J. S. et al. Leveraging Northern European population history: novel low-frequency variants for polycystic ovary syndrome. *Hum. Reprod.* **37**, 352–365 (2022).

14. Barber, T. M. et al. Association of variants in the fat mass and obesity associated (FTO) gene with polycystic ovary syndrome. *Diabetologia* **51**, 1153–1158 (2008).
15. Brower, M. A. et al. Bidirectional Mendelian randomization to explore the causal relationships between body mass index and polycystic ovary syndrome. *Hum. Reprod.* **34**, 127–136 (2019).
16. Kentistou, K. A. et al. Understanding the genetic complexity of puberty timing across the allele frequency spectrum. *Nat. Genet.* **56**, 1397–1411 (2024).
17. Silva, M. S. B. & Giacobini, P. New insights into anti-Müllerian hormone role in the hypothalamic-pituitary-gonadal axis and neuroendocrine development. *Cell. Mol. Life Sci.* **78**, 1–16 (2021).
18. Visser, J. A. & Themmen, A. P. Anti-Müllerian hormone and folliculogenesis. *Mol. Cell. Endocrinol.* **234**, 81–86 (2005).
19. Verdiesen, R. M. G. et al. Genome-wide association study meta-analysis identifies three novel loci for circulating anti-Müllerian hormone levels in women. *Hum. Reprod.* **37**, 1069–1082 (2022).
20. Kristensen, S. G. et al. Quantitative differences in TGF- β family members measured in small antral follicle fluids from women with or without PCO. *J. Clin. Endocrinol. Metab.* **104**, 6371–6384 (2019).
21. Day, F. R. et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat. Genet.* **49**, 834–841 (2017).
22. Mbarek, H. et al. Identification of common genetic variants influencing spontaneous dizygotic twinning and female fertility. *Am. J. Hum. Genet.* **98**, 898–908 (2016).
23. Cooper, L. A., Page, S. T., Amory, J. K., Anawalt, B. D. & Matsumoto, A. M. The association of obesity with sex hormone-binding globulin is stronger than the association with ageing—implications for the interpretation of total testosterone measurements. *Clin. Endocrinol. (Oxf.)* **83**, 828–833 (2015).
24. Aoyama, T. et al. Cytochrome P-450 hPCN3, a novel cytochrome P-450 IIIA gene product that is differentially expressed in adult human liver. cDNA and deduced amino acid sequence and distinct specificities of cDNA-expressed hPCN1 and hPCN3 for the metabolism of steroid hormones and cyclosporine. *J. Biol. Chem.* **264**, 10388–10395 (1989).
25. Hashimoto, M. et al. Cyp3a deficiency enhances androgen receptor activity and cholesterol synthesis in the mouse prostate. *J. Steroid Biochem. Mol. Biol.* **163**, 121–128 (2016).
26. Magoffin, D. A. & Weitsman, S. R. Differentiation of ovarian theca-interstitial cells in vitro: regulation of 17 α -hydroxylase messenger ribonucleic acid expression by luteinizing hormone and insulin-like growth factor-I. *Endocrinology* **132**, 1945–1951 (1993).
27. Ruth, K. S. et al. Genome-wide association study with 1000 genomes imputation identifies signals for nine sex hormone-related phenotypes. *Eur. J. Hum. Genet.* **24**, 284–290 (2015).
28. Saxena, R. et al. Han Chinese polycystic ovary syndrome risk variants in women of European ancestry: relationship to FSH levels and glucose tolerance. *Hum. Reprod.* **30**, 1454–1459 (2015).
29. Koppaka, V. et al. Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application. *Pharmacol. Rev.* **64**, 520–539 (2012).
30. Pares, X., Farres, J., Kedishvili, N. & Duester, G. Medium- and short-chain dehydrogenase/reductase gene and protein families: medium-chain and short-chain dehydrogenases/reductases in retinoid metabolism. *Cell. Mol. Life Sci.* **65**, 3936–3949 (2008).
31. Kunutsor, S. K., Abbasi, A. & Adler, A. I. γ -Glutamyl transferase and risk of type II diabetes: an updated systematic review and dose-response meta-analysis. *Ann. Epidemiol.* **24**, 809–816 (2014).
32. Cortese, A. et al. Biallelic mutations in SORD cause a common and potentially treatable hereditary neuropathy with implications for diabetes. *Nat. Genet.* **52**, 473–481 (2020).
33. Johansson, A. S. & Mannervik, B. Human glutathione transferase A3-3, a highly efficient catalyst of double-bond isomerization in the biosynthetic pathway of steroid hormones. *J. Biol. Chem.* **276**, 33061–33065 (2001).
34. Sawyer, L. β -lactoglobulin and glycodelin: two sides of the same coin? *Front. Physiol.* **12**, 678080 (2021).
35. Kolberg, L. et al. g:Profiler— intraoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* **51**, W207–W212 (2023).
36. Wild, R. A., Rizzo, M., Clifton, S. & Carmina, E. Lipid levels in polycystic ovary syndrome: systematic review and meta-analysis. *Fertil. Steril.* **95**, 1073–1079 (2011).
37. Guo, F. et al. The lipid profiles in different characteristics of women with PCOS and the interaction between dyslipidemia and metabolic disorder states: a retrospective study in Chinese population. *Front. Endocrinol. (Lausanne)* **13**, 892125 (2022).
38. Mikhaylova, I. V., Kuulasmaa, T., Jaaskelainen, J. & Voutilainen, R. Tumor necrosis factor- α regulates steroidogenesis, apoptosis, and cell viability in the human adrenocortical cell line NCI-H295R. *Endocrinology* **148**, 386–392 (2007).
39. Mallick, R. et al. Fatty acids and their proteins in adipose tissue inflammation. *Cell Biochem. Biophys.* **82**, 35–51 (2024).
40. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
41. Joo, Y. Y. et al. A polygenic and phenotypic risk prediction for polycystic ovary syndrome evaluated by phenome-wide association studies. *J. Clin. Endocrinol. Metab.* **105**, 1918–1936 (2020).
42. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
43. Dokras, A. et al. Androgen excess-polycystic ovary syndrome society: position statement on depression, anxiety, quality of life, and eating disorders in polycystic ovary syndrome. *Fertil. Steril.* **109**, 888–899 (2018).
44. Teede, H. J. et al. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum. Reprod.* **33**, 1602–1618 (2018).
45. Bolcun-Filas, E., Rinaldi, V. D., White, M. E. & Schimenti, J. C. Reversal of female infertility by Chk2 ablation reveals the oocyte DNA damage checkpoint pathway. *Science* **343**, 533–536 (2014).
46. Sorensen, E. et al. Data resource profile: the Copenhagen Hospital Biobank (CHB). *Int. J. Epidemiol.* **50**, 719–720 (2021).
47. Tidwell, A., Zhu, J., Battiola, T. & Welt, C. K. Phenotypes associated with polycystic ovary syndrome risk variants. *J. Endocr. Soc.* **9**, bvae219 (2024).
48. Mills, M. C. et al. Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour. *Nat. Hum. Behav.* **5**, 1717–1730 (2021).
49. Mathieson, I. et al. Genome-wide analysis identifies genetic effects on reproductive success and ongoing natural selection at the FADS locus. *Nat. Hum. Behav.* **7**, 790–801 (2023).
50. Censin, J. C., Bovijn, J., Holmes, M. V. & Lindgren, C. M. Colocalization analysis of polycystic ovary syndrome to identify potential disease-mediating genes and proteins. *Eur. J. Hum. Genet.* **29**, 1446–1454 (2021).
51. Welt, C. K., Pagan, Y. L., Smith, P. C., Rado, K. B. & Hall, J. E. Control of follicle-stimulating hormone by estradiol and the inhibins: critical role of estradiol at the hypothalamus during the luteal-follicular transition. *J. Clin. Endocrinol. Metab.* **88**, 1766–1771 (2003).

52. Yding Andersen, C. Inhibin-B secretion and FSH isoform distribution may play an integral part of follicular selection in the natural menstrual cycle. *Mol. Hum. Reprod.* **23**, 16–24 (2017).
53. Dewailly, D., Barbotin, A. L., Dumont, A., Catteau-Jonard, S. & Robin, G. Role of anti-Müllerian hormone in the pathogenesis of polycystic ovary syndrome. *Front. Endocrinol. (Lausanne)* **11**, 641 (2020).
54. Laurent, M. R. et al. Sex hormone-binding globulin regulation of androgen bioactivity in vivo: validation of the free hormone hypothesis. *Sci. Rep.* **6**, 35539 (2016).
55. Drummond, A. E. The role of steroids in follicular growth. *Reprod. Biol. Endocrinol.* **4**, 16 (2006).
56. Bristol-Gould, S. K. et al. Postnatal regulation of germ cells by activin: the establishment of the initial follicle pool. *Dev. Biol.* **298**, 132–148 (2006).
57. Ruth, K. S. et al. Genetic evidence that lower circulating FSH levels lengthen menstrual cycle, increase age at menopause and impact female reproductive health. *Hum. Reprod.* **31**, 473–481 (2016).
58. Hudecova, M., Holte, J., Olovsson, M. & Sundstrom Poromaa, I. Long-term follow-up of patients with polycystic ovary syndrome: reproductive outcome and ovarian reserve. *Hum. Reprod.* **24**, 1176–1183 (2009).
59. Koivunen, R. et al. Fecundability and spontaneous abortions in women with self-reported oligo-amenorrhea and/or hirsutism: Northern Finland Birth Cohort 1966 Study. *Hum. Reprod.* **23**, 2134–2139 (2008).
60. West, S. et al. The impact of self-reported oligo-amenorrhea and hirsutism on fertility and lifetime reproductive success: results from the Northern Finland Birth Cohort 1966. *Hum. Reprod.* **29**, 628–633 (2014).
61. Persson, S. et al. Fecundity among women with polycystic ovary syndrome (PCOS)—a population-based study. *Hum. Reprod.* **34**, 2052–2060 (2019).
62. Elting, M. W., Korsen, T. J., Rekers-Mombarg, L. T. & Schoemaker, J. Women with polycystic ovary syndrome gain regular menstrual cycles when ageing. *Hum. Reprod.* **15**, 24–28 (2000).
63. Carmina, E., Campagna, A. M. & Lobo, R. A. A 20-year follow-up of young women with polycystic ovary syndrome. *Obstet. Gynecol.* **119**, 263–269 (2012).
64. Pigny, P. et al. Elevated serum level of anti-müllerian hormone in patients with polycystic ovary syndrome: relationship to the ovarian follicle excess and to the follicular arrest. *J. Clin. Endocrinol. Metab.* **88**, 5957–5962 (2003).
65. Webber, L. J. et al. Formation and early development of follicles in the polycystic ovary. *Lancet* **362**, 1017–1021 (2003).
66. Emerging Risk Factors Collaboration et al. Lipoprotein(a) concentration and the risk of coronary heart disease, stroke, and nonvascular mortality. *JAMA* **302**, 412–423 (2009).
67. Ruth, K. S. et al. Using human genetics to understand the disease impacts of testosterone in men and women. *Nat. Med.* **26**, 252–258 (2020).
68. Bots, S. H., Peters, S. A. E. & Woodward, M. Sex differences in coronary heart disease and stroke mortality: a global assessment of the effect of ageing between 1980 and 2010. *BMJ Glob. Health* **2**, e000298 (2017).
69. Vitale, C., Fini, M., Speziale, G. & Chierchia, S. Gender differences in the cardiovascular effects of sex hormones. *Fundam. Clin. Pharmacol.* **24**, 675–685 (2010).
70. Schmidt, J., Brannstrom, M., Landin-Wilhelmsen, K. & Dahlgren, E. Reproductive hormone levels and anthropometry in postmenopausal women with polycystic ovary syndrome (PCOS): a 21-year follow-up study of women diagnosed with PCOS around 50 years ago and their age-matched controls. *J. Clin. Endocrinol. Metab.* **96**, 2178–2185 (2011).
71. Puurunen, J. et al. Unfavorable hormonal, metabolic, and inflammatory alterations persist after menopause in women with PCOS. *J. Clin. Endocrinol. Metab.* **96**, 1827–1834 (2011).

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026

Loes M. E. Moolhuijsen^{1,100}, Jia Zhu^{2,3,4,100}, Benjamin H. Mullin^{5,6,100}, Natàlia Pujol-Gualdo^{7,8}, Ky'era V. Actkins^{9,10}, Jasmine A. Mack^{11,12,13}, Hridya Rao^{14,15}, Bhavi Trivedi¹⁶, Katherine A. Kentistou¹⁷, Yajie Zhao¹⁷, David Westergaard^{18,19}, Jaakko S. Tyrmi^{20,21}, Gudmar Thorleifsson²², Yanfei Zhang^{14,23}, Laura Wittmans²⁴, Amber DeVries²⁵, Kelly Brewer²⁶, Ryan Sisk²⁷, Rebecca Danning²⁸, Michael H. Preuss²⁹, Michelle R. Jones²⁵, Katherine S. Ruth³⁰, Marianne Andersen³¹, Ricardo Azziz³², Karina Banasik^{18,19}, Michael Boehnke³³, Linda Broer¹, Søren Brunak^{18,34}, Yee-Ming Chan^{2,3,4}, Daniel I. Chasman^{28,35}, Mark Daly^{36,37,38}, David A. Ehrmann³⁹, Bart C. Fauser⁴⁰, Lars G. Fritsche³³, M. Geoffrey Hayes^{41,42,43}, Chunyan He^{35,44}, Hongyan Huang⁴⁵, Irina Kowalska⁴⁶, Peter Kraft⁴⁵, Richard S. Legro⁴⁷, Nan Lin^{35,44}, Ruth J. Loos^{29,48,49,50}, Yvonne V. Louwers⁵¹, Reedik Magi⁷, Mark I. McCarthy^{52,53}, Laure Morin-Papunen⁵⁴, Jean V. Morrison⁵⁵, Cynthia Morton^{4,35,56,57,58}, Girish N. Nadkarni⁵⁹, Benjamin M. Neale^{60,61}, Henriette Svarre Nielsen^{19,62,63}, Mette Nyegaard⁶⁴, Sisse R. Ostrowski^{62,63,65}, Ole B. V. Pedersen^{65,66}, Erik Sørensen⁶⁵, Christina Mikkelsen^{65,67}, Christian Erikstrup^{68,69}, Kathrine A. Kaspersen^{68,70}, Mie T. Bruun^{71,72}, Bitten Aagaard⁷³, Henrik Ullum⁷⁴, Barbara Obermayer-Pietsch⁷⁵, Aarno Palotie^{37,60,61}, Mary P. Reeve³⁷, Andres Salumets^{76,77,78}, Richa Saxena³⁶, Timothy D. Spector⁷⁹, Bronwyn G. A. Stuckey^{5,80,81}, Unnur Thorsteinsdottir^{22,82}, André G. Uitterlinden^{1,83}, Margrit Urbanek^{41,43}, Sebastian Zöllner⁸⁴, Genes and Health Research Team*, DBDS Genomic Consortium*, 23andMe Research Team*, David A. van Heel¹⁶, Joel N. Hirschhorn^{2,3,85,86}, Kari Stefansson^{22,82}, John R. B. Perry^{17,87}, Unnur Styrkarsdottir²²

Scott G. Wilson^{1,5,6,80}, **Terhi Piltonen**⁸, **Triin Laisk**⁷, **Marjo-Riitta Jarvelin**^{1,8,21,88,89}, **Kharis Burns**^{81,90}, **Anne E. Justice**¹⁴, **Hannele Laivuori**^{1,20,37,91,92}, **Ken K. Ong**^{17,93}, **Mark O. Goodarzi**⁹⁴, **Lea K. Davis**^{9,10}, **Andrea Dunaif**²⁶, **Cecilia M. Lindgren**^{1,24,36,95}, **Joop S. E. Laven**⁵¹, **Stephen Franks**⁹⁶, **Jenny A. Visser**¹, **Corrine K. Welt**^{97,100} ✉, **Tugce Karaderi**^{18,98,100} ✉ & **Felix R. Day**^{17,100} ✉

¹Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. ²Division of Endocrinology, Boston Children's Hospital, Boston, MA, USA. ³Department of Pediatrics, Harvard Medical School, Boston, MA, USA. ⁴Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁵Department of Endocrinology & Diabetes, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia. ⁶School of Biomedical Sciences, University of Western Australia, Crawley, Western Australia, Australia. ⁷Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia. ⁸Department of Obstetrics and Gynecology, Research Unit of Clinical Medicine, Medical Research Centre, Oulu University Hospital, University of Oulu, Oulu, Finland. ⁹Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. ¹⁰Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. ¹¹Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. ¹²Department of Obstetrics and Gynaecology, University of Cambridge, Cambridge, UK. ¹³Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA. ¹⁴Department of Population Health Sciences, Geisinger, Danville, PA, USA. ¹⁵Department of Biobehavioral Health, Pennsylvania State University, University Park, PA, USA. ¹⁶Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. ¹⁷MRC Epidemiology Unit, Institute of Metabolic Science, School of Clinical Medicine, University of Cambridge, Cambridge, UK. ¹⁸Translational Disease Systems Biology, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ¹⁹Department Obstetrics and Gynecology, Copenhagen University Hospital Hvidovre, Hvidovre, Denmark. ²⁰Center for Child, Adolescent and Maternal Health Research, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. ²¹Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland. ²²deCODE genetics/Amgen, Reykjavik, Iceland. ²³Population Analytics and Insights, Data Science and Digital Health, Janssen R&D, Springhouse, PA, USA. ²⁴The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. ²⁵Center for Bioinformatics & Functional Genomics, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ²⁶Division of Endocrinology, Diabetes and Bone Disease, Icahn School of Medicine at Mount Sinai, New York City, NY, USA. ²⁷Division of Cardiology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. ²⁸Brigham and Women's Hospital, Boston, MA, USA. ²⁹The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York City, NY, USA. ³⁰Department of Clinical and Biomedical Science, University of Exeter, Exeter, UK. ³¹Odense University Hospital, University of Southern Denmark, Odense, Denmark. ³²Obstetrics & Gynecology, Medicine, and Healthcare Organization & Policy, Schools of Medicine and Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. ³³Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. ³⁴Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ³⁵Harvard Medical School, Boston, MA, USA. ³⁶Broad Institute of MIT and Harvard and Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. ³⁷Institute for Molecular Medicine Finland (FIMM), Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland. ³⁸Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. ³⁹Department of Medicine, Section of Adult and Paediatric Endocrinology, Diabetes, and Metabolism, The University of Chicago, Chicago, IL, USA. ⁴⁰Department of Reproductive Medicine and Gynaecology, University of Utrecht & University Medical Center, Utrecht, the Netherlands. ⁴¹Division of Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. ⁴²Department of Anthropology, Northwestern University, Evanston, IL, USA. ⁴³Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. ⁴⁴Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, MA, USA. ⁴⁵Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ⁴⁶Department of Internal Medicine and Metabolic Diseases, Medical University of Białystok, Białystok, Poland. ⁴⁷Department of Obstetrics and Gynecology and Public Health Sciences, Penn State University College of Medicine, Hershey, PA, USA. ⁴⁸The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York City, NY, USA. ⁴⁹The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York City, NY, USA. ⁵⁰The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medicine, University of Copenhagen, Copenhagen, Denmark. ⁵¹Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynaecology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands. ⁵²Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. ⁵³Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK. ⁵⁴Research Unit of Clinical Medicine, University of Oulu, Oulu, Finland. ⁵⁵Department of Human Genetics, University of Chicago, Chicago, IL, USA. ⁵⁶The Developmental Genome Anatomy Project, Brigham and Women's Hospital, Boston, MA, USA. ⁵⁷Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Boston, MA, USA. ⁵⁸Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA. ⁵⁹Icahn School of Medicine at Mount Sinai, New York City, NY, USA. ⁶⁰Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁶¹Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. ⁶²Institute of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. ⁶³Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ⁶⁴Department of Health Science and Technology, Aalborg University, Gistrup, Denmark. ⁶⁵Department of Clinical Immunology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark. ⁶⁶Department of Clinical Immunology, Zealand University Hospital, Køge, Denmark. ⁶⁷Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Science, Copenhagen University, Copenhagen, Denmark. ⁶⁸Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark. ⁶⁹Department of Clinical Medicine, Aarhus University, Aarhus, Denmark. ⁷⁰Danish Big Data Centre for Environment and Health (BERTHA), Aarhus University, Aarhus, Denmark. ⁷¹Clinical Immunology Research Unit, Department of Clinical Immunology, Odense University Hospital, Odense, Denmark. ⁷²Department of Clinical Research, University of Southern Denmark, Odense, Denmark. ⁷³Department of Clinical Immunology, Aalborg University Hospital, Aalborg, Denmark. ⁷⁴Statens Serum Institut, Copenhagen, Denmark. ⁷⁵Division of Endocrinology and Diabetology, Department of Internal Medicine, Medical University of Graz, Graz, Austria. ⁷⁶Competence Centre on Health Technologies, Tartu, Estonia. ⁷⁷Department of Obstetrics and Gynecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia. ⁷⁸Division of Obstetrics and Gynecology, Department of Clinical Science, Intervention and Technology, Karolinska Institutet and Karolinska University Hospital, Huddinge, Stockholm, Sweden. ⁷⁹Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. ⁸⁰Medical School, University of Western Australia, Crawley, Western Australia, Australia. ⁸¹Keogh Institute for Medical Research, Nedlands, Western Australia, Australia. ⁸²Faculty of Medicine, University of Iceland, Reykjavik, Iceland. ⁸³Department of Epidemiology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands. ⁸⁴Department of Biostatistics and Department of

Psychiatry, University of Michigan, Ann Arbor, MI, USA. ⁸⁵Programs in Metabolism and Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁸⁶Department of Genetics, Harvard Medical School, Boston, MA, USA. ⁸⁷Metabolic Research Laboratory, Wellcome-MRC Institute of Metabolic Science, School of Clinical Medicine, University of Cambridge, Cambridge, UK. ⁸⁸MRC Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. ⁸⁹Research Unit of Population Health, Faculty of Medicine, University of Oulu, Oulu, Finland. ⁹⁰Department of Endocrinology and Diabetes, Royal Perth Hospital, Perth, Western Australia, Australia. ⁹¹Department of Obstetrics and Gynecology, Tampere University Hospital, The Wellbeing Services County of Pirkanmaa, Tampere, Finland. ⁹²Medical and Clinical Genetics, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. ⁹³Department of Paediatrics, University of Cambridge, Cambridge, UK. ⁹⁴Division of Endocrinology, Diabetes, and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ⁹⁵Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ⁹⁶Institute of Reproductive & Developmental Biology, Department of Metabolism, Digestion & Reproduction, Imperial College London, London, UK. ⁹⁷Division of Endocrinology, Metabolism and Diabetes, University of Utah, Salt Lake City, UT, USA. ⁹⁸Center for Health Data Science, Faculty of Medical and Health Sciences, University of Copenhagen, Copenhagen, Denmark. ¹⁰⁰These authors contributed equally: Loes M. E. Moolhuijsen, Jia Zhu, Benjamin H. Mullin, Corrine K. Welt, Tugce Karaderi, Felix R. Day. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: cwelt@genetics.utah.edu; tugce.karaderi@sund.ku.dk; felix.day@mrc-epid.cam.ac.uk

Genes and Health Research Team

Bhavi Trivedi¹⁶ & David A. van Heel¹⁶

DBDS Genomic Consortium

Bitten Aagaard⁷³, Karina Banasik^{18,19}, Søren Brunak^{18,34}, Mie T. Bruun^{71,72}, Christian Erikstrup^{68,69}, Kathrine A. Kaspersen^{68,70}, Christina Mikkelsen^{65,67}, Mette Nyegaard⁶⁴, Sisse R. Ostrowski^{62,63,65}, Ole B. V. Pedersen^{65,66}, Erik Sørensen⁶⁵, Kari Stefansson^{22,82}, Unnur Thorsteinsdottir^{22,82}, Henrik Ullum⁷⁵ & David Westergaard^{18,19}

23andMe Research Team

Adam Auton⁹⁹, Alan Kwong⁹⁹, Anjali J. Shastri⁹⁹, Barry Hicks⁹⁹, Catherine H. Weldon⁹⁹, David A. Hinds⁹⁹, Emily DelloRusso⁹⁹, Emily M. Rios⁹⁹, Joyce Y. Tung⁹⁹, Kahsaia de Brito⁹⁹, Katelyn Kukar Bond⁹⁹, Keng-Han Lin⁹⁹, Matthew H. McIntyre⁹⁹, Matthew J. Kmieciak⁹⁹, Qiaojuan Jane Su⁹⁹, Robert K. Bell⁹⁹, Sayantan Das⁹⁹, Shubham Saini⁹⁹, Stella Aslibekyan⁹⁹, Vinh Tran⁹⁹, Wanwan Xu⁹⁹, Alisa P. Lehman⁹⁹, Noura S. Abul-Husn⁹⁹, R. RYanne Wu⁹⁹, Rebecca M. K. Berns⁹⁹, Ruth I. Tennen⁹⁹, Stacey B. Detweiler⁹⁹, Aditya Ambati⁹⁹, Anna Guan⁹⁹, Bertram L. Koelsch⁹⁹, Chris German⁹⁹, Éadaoin Harney⁹⁹, Ethan M. Jewett⁹⁹, G. David Poznik⁹⁹, James R. Ashenurst⁹⁹, Jingran Wen⁹⁹, Peter R. Wilton⁹⁹, Steven J. Micheletti⁹⁹ & William A. Freyman⁹⁹

⁹⁹23andMe Research Institute, Palo Alto, CA, USA. Full lists of members and their affiliations appear in the Supplementary Information.

Methods

Ethics

This study complies with all relevant ethical regulations—details of the specific ethical approval for the studies contributing to the meta-analysis, or other analyses can be found in the Supplementary Note. All identifiable participants in all cohorts provided written informed consent. Anonymized data were studied under a waiver of informed consent.

Study characteristics

This study included summary statistics of our previous meta-analysis, including 3,981 cases and 17,322 controls of European ancestry after removing overlapping samples (Supplementary Table 1; described in detail in ref. 6). An additional 16 studies from 13 cohorts were included for the current GWAS meta-analysis, resulting in 20,818 cases and 523,695 controls (Supplementary Table 1). The newly added cases included all ancestries, of which 87% cases were from European origin, 10% from East-Asian origin and the remaining were either from Hispanic or African-American ancestry. Included women with PCOS were diagnosed using three different PCOS criteria, namely (1) the clinical diagnostic criteria (either the Rotterdam or the National Institutes of Health (NIH) criteria), (2) women diagnosed using the electronic health record (EHR) ICD-8, ICD-9 or ICD-10 codes, or (3) self-reported cases (see more details in Supplementary Note and Supplementary Table 25). The Rotterdam criteria diagnosis is based on the presence of at least two of the three criteria—ovulatory dysfunction, HA and PCOM. The NIH criteria only require the presence of ovulatory dysfunction and HA. Detailed description of each cohort can be found in Supplementary Table 1. In Supplementary Fig. 1, we have visualized the distribution of diagnostic criteria and ethnic groups of the newly added cases and previous meta-analysis results⁶ combined.

Data collection and quality control

Summary results of GWAS using a case-control setting were provided by the studies contributing to the meta-analysis. At the study level, the analyses were adjusted for age, principal components and BMI (only for BMI-adjusted analyses). Central quality control was performed by two independent analysts using the EasyQC pipeline⁷². Variant exclusion filters used included—(1) minor allele frequency < 1%, (2) imputation quality (R^2) < 0.3 or info of < 0.4 for MACH and IMPUTE2, respectively⁷².

Meta-analysis

A fixed-effect, inverse-weighted-variance meta-analysis approach was used with the collected summary statistics from the individual studies. Either GWAMA⁷³ or METAL⁷⁴ was used as a meta-analysis tool. We performed meta-analyses for all ancestries combined and only for European ancestry combined. These meta-analyses were carried out using two models—age-adjusted and age and BMI-adjusted—given the association between obesity and PCOS⁵. Variants present in at least three strata were reported and used in further analyses.

These meta-analysis results were then combined with the previously published genome-wide meta-analysis summary statistics⁶ to increase the statistical power and discover further associations with PCOS status. We called this analysis ‘the two-strata meta-analysis’. The resulting sample size was 456,570 (15,634 cases and 440,936 controls; Supplementary Table 1). As previous research had found no substantial heterogeneity in variant discovery as a function of different diagnostic criteria⁶, studies with any method of case ascertainment were combined. Variants present in all strata were reported and used in the follow-up analyses. Identified variants were annotated and investigated further with regard to their biological function using FUMA⁷⁵. Forest plots for comparing the effect sizes across the strata in the meta-analysis were made using the ggplot2 package in R.

Furthermore, we compared the effect sizes across different phenotype definitions used; PCOS definitions based on EHRs, clinical

diagnosis and self-reports were included in this comparison. EHRs were based on de ICD-8/ICD-9/ICD-10 codes for PCOS (E28.2/256.4/256.9, or Hirsutism L68.0/704.1/704 and Irregular Menses N91.X, N92.5, N92.6/626.X/626). Clinical diagnosis was based on either the NIH criteria or the Rotterdam criteria. Additional meta-analysis was performed to statistically test for heterogeneity across these three PCOS definitions. In addition to visually inspecting forest plots for the meta-analysis, Cochran’s Q P value and I^2 were used for assessing heterogeneity. A P value below 0.05 was considered statistically significant.

The summary statistics from the age-adjusted meta-analyses were further combined with the previously published summary statistics from 23andMe to increase the statistical power⁶. We called this analysis ‘the three-strata meta-analysis’. The resulting sample size was 544,513 (20,818 cases and 523,695 controls; Supplementary Table 1).

To assess the effects of the lead 29 GWAS variants ($P < 5 \times 10^{-8}$) in the BMI-adjusted model, we performed an individual variant look-up in the summary statistics of the BMI-adjusted model for the three-strata (Supplementary Table 2) and two-strata (Supplementary Table 4) analyses. For this single variant association study, associations were considered statistically significant if they passed the Bonferroni correction for 29 variants ($P < 0.05/29 = 0.0017$).

We assessed the lead GWAS variants ($P < 5 \times 10^{-8}$) by examining their relationship with 20 related metabolic, hormonal and reproductive phenotypes with available GWAS results data. Except for LH and FSH all other traits were publicly available (Supplementary Table 10). The heatmap was drawn using the ‘pheatmap’ library in R (v3.6.1). Finally, colocalization analysis was performed on the three-strata data using the R package ‘coloc’⁷⁶, based on a window 500 kb either side of the sentinel signals.

Fine-mapping

To identify a credible set of variants containing the most likely causal variant underlying our association signals, we conducted fine-mapping using the shotgun stochastic search method as performed in FINEMAP⁷⁷. We used summary statistics from our two-stage summary GWAS meta-analysis results without the data from 23andMe, and considered all variants within 1 Mb \pm from our tag variants. We used two different contributing studies as LD references to perform fine-mapping. First, we used unrelated (up to second degree), European ancestry women from the MyCode Community Health Initiative Study (DiscovEHR; $n = 47,061$) with genetic data imputed to the 1000 Genomes Phase III global reference panel. European ancestry was inferred using genetic data as described elsewhere⁷⁸. Second, we used an unrelated dataset of European-ancestry females ($n = 36,890$) in the EHR-linked biobank at Vanderbilt University Medical Center (BioVU). Genetic data were imputed to the Haplotype Reference Consortium and European ancestry was defined by principal components⁷⁹. We assumed a single causal variant for all loci, and for four loci with evidence of a secondary signal, we also performed fine-mapping assuming two causal variants.

Functional mapping and annotation of GWAS

Functional mapping and annotation of GWAS were performed with FUMA, and further annotation of the association results with PhenoScanner (date accessed on 25 March 2022)^{80,81}. FUMA analyses were performed using the summary statistics for (1) the top 29 PCOS-associated variants in the three-strata meta-analysis and (2) the genome-wide two-strata meta-analysis. Unless specified otherwise, the default settings were used in the FUMA analyses for both SNP2GENE and GENE2FUNC⁷⁵.

Proteomic analysis

Proteomic analysis using logistic regression for the association of normalized plasma protein levels with disease was run in the ~22,000 women with data from the Olink panel of plasma proteins, aged 56.5 ± 8.1 years⁸². Here the outcome was the first occurrence data,

and to maximize sample size, we used a diagnosis of any of the ICD-10 code E28, the supergroup that includes PCOS. Proteins were considered significantly associated if they passed a Bonferroni-corrected P -value threshold of 3.4×10^{-5} . The generation of the protein data is described elsewhere⁸³.

Separately, we performed a protein PheWAS for each variant identified in the GWAS meta-analysis, using the total sample of ~44,000 (men and women) and Olink data to identify proteins linked to our PCOS signals. Again, we used a Bonferroni-corrected P -value threshold of 3.4×10^{-5} . Once this panel of proteins had been identified, we prioritized proteins significantly associated with the ICD-10 code E28 diagnosis, as defined above. Finally, once we had established our variant–protein pairs, we attempted to establish the position in the causal pathways by considering the relative R^2 , with those variants that had a higher R^2 with PCOS suggesting that the protein was downstream of PCOS, and vice versa.

Annotating genes of interest

First, we performed a literature review of all genes within 500 kb of the 29 signals to identify genes with a reported link to one of the following four processes: (1) reproductive function, (2) steroid metabolism and sex-hormone levels, (3) pathways related to metabolic syndrome and (4) DNA damage repair, selected due to reported links between PCOS and age at menopause for which DNA damage repair is the dominant process⁷. The literature search was conducted using PubMed and the Cochrane databases. For each gene, the most important findings regarding the four preselected processes are summarized in Supplementary Table 8.

Second, we used the established GWAS-to-genes pipeline, which integrates genomic and functional data through multiple gene annotation methods to highlight likely causal genes at each of the identified signals, as described elsewhere¹⁶. Briefly, tissue enrichment for GWAS associations was performed using LD score regression to identify key tissues for annotations with tissue-specific datasets. Then, a gene score is generated from the following panels of annotations: (1) the closest gene to the signal scored 1.5 points. (2) eQTL⁸⁴ colocalization from both SMR-HEIDI and the ‘coloc’ package⁷⁶ was scored 1.5 points, or 1 if only from one of these. An additional point was given to genes with eQTLs at secondary signals. (3) Colocalization with pQTL derived from plasma^{82,85} scored the same as for eQTLs. (4) Coding variants, with variants of deleterious or damaging predicted consequence in LD with GWAS PCOS signals, were scored 1.0 point, or only 0.5 points if the coding variants were predicted to be benign or tolerated. (5) Genes targeted by enhancers that overlapped with or were correlated with GWAS PCOS signals were scored 1.0 point. (6) Polygenic priority score prioritized genes at each locus were scored 1.5 points⁸⁶.

Gene-set enrichment analysis

To perform gene-set enrichment analysis that leveraged information across both the proteomics and the implicated genes, we used GProfiler, selecting either the consensus gene or the associated protein from the proteomics analysis. Clustering of the pathways was done using an index of dissimilarity based on the shared genes across the enriched intersections of each pathway¹⁶.

PRS analyses in the UK Biobank

We used the PRS-CS software to calculate a PRS for PCOS, which is a Bayesian regression framework that applies continuous shrinkage parameters to estimate posterior effect sizes⁴². This work was performed in the UK Biobank, a population-based cohort of ~500,000 individuals in the United Kingdom⁸⁷, which was independent of the discovery GWAS sample. The tuning or global shrinkage parameter $\phi = 1 \times 10^{-4}$ that optimized the association of the PRS for PCOS in the UK Biobank, as previously reported, was used¹⁰. Using this method, our PCOS PRS included 1,119,009 genetic variants. In the same UK Biobank

sample, we replicated these analyses using another PRS tool, LDpred2, that uses a Bayesian shrinkage model⁸⁸.

To identify women with PCOS in the UK Biobank study, data from self-report, primary-care clinical events and/or ICD-9 and ICD-10, as previously reported,¹⁰ were used. We binned women with or without a diagnosis of PCOS by their quintile of PRS and used logistic regression to determine the odds of PCOS for each quintile using the lowest quintile as a reference. Women with a PCOS PRS in the highest quintile had an increased odds of PCOS (OR = 2.41, 95% CI = 1.96–2.98; $P = 2 \times 10^{-16}$). Thus, our PCOS PRS is able to represent the genetic risk for PCOS in women in the UK Biobank.

Ascertainment of cardiometabolic and androgenic phenotypes has been previously reported¹⁰. All other phenotypes, including measures of fertility and longevity, asthma and mental health disorders, were based on a composite of self-reported measures, diagnosis codes from hospitalization records and age at diagnosis (Supplementary Note). We used linear and logistic regressions to analyze the associations between continuous and dichotomous phenotypes and the PCOS PRS, respectively. We adjusted all analyses for age, age squared, genotyping array, the UK Biobank assessment center and the first ten genetic principal components; for asthma and psychological outcomes, we additionally controlled for the Townsend deprivation index, and, for asthma, we further controlled for smoking status. Adjustment for BMI was performed in the following two ways: (1) a measured BMI was included in the model as a covariate, and (2) we constructed a score based on the GWAS meta-analysis where the genetic associations were adjusted for BMI.

PRS analysis in CHB based on the Danish Registries

PRSs for PCOS were calculated using LDpred2 (ref. 89). These genome-wide scores were calculated using the meta-analysis excluding data from 23andMe. Autosomal genotype data from 138,669 individuals in the CHB were filtered to only include variants present in a set of 1,054,330 reference variants recommended by LDpred2 developers. Missing genotype information was imputed to be the reference allele for the affected locus. GWAS summary statistics were preprocessed with MungeSumStats.

The completed family size was determined by counting the number of live births from the Medical Birth Registry⁹⁰. This study was initiated in 1973, and data are considered complete. Only women born in the years 1957–1973 were included in this analysis, the youngest would be 45 years old, and 61 years old when data collection ended (31 December 2018). Data were treated as count data, and we tested to determine whether there was equidispersion, underdispersion or overdispersion using the AER R package (v.1.2.10). We found significant underdispersion (dispersion = 0.60, $P < 2.2 \times 10^{-16}$). Consequently, data were analyzed using a Conway–Maxwell–Poisson distribution.

From the Medical Birth Registry, we also identified the age at first birth and last birth (expressed in days). Data were analyzed using a linear regression, and model fit was inspected from residuals. There were no signs of deviation from a Gaussian error.

The Danish IVF registry was initiated in 1994 and contains data on all treatments and procedures related to medically assisted reproduction. Reporting is mandatory for both private and public clinics. Furthermore, there is information on any procedure-related treatments and their duration. Female infertility was defined using the 628 (ICD-8) and N97 (ICD-10; excluding N97.4) in the National Patient Registry (public hospitals only⁹¹) and ‘female cause’ in the IVF Register. Male infertility was defined using the 606 (ICD-8) and N46 (ICD-10) in the National Patient Registry and ‘male cause’ (excluding male infertility due to sterilization) in the IVF Register. For the number of oocytes, we extracted all aspirations performed between 18 January 1994 and 31 December 2018. The mandatory reported data were changed in 2005, and, thus, we analyzed the two time periods (1994–2005 and 2006–2018) separately and meta-analyzed. We additionally extracted

information on treatment (Klomifen, HMG-FSH, GnRH-A, Estrogen, Progesterone, HCG) and the number of treatment days before the aspiration. Both time periods were overdispersed and were analyzed using a negative binomial distribution. To take into account multiple aspirations for a single woman, we included an individual random term.

Finally, we investigated the number of cycles before a woman got pregnant or ceased treatment. Each woman was only included until the first pregnancy. All transfer or insemination attempts were summarized, and analyzed using a Conway–Maxwell–Poisson distribution, as data were underdispersed. Furthermore, a term for zero inflation was also tested, as only 81% of the population became pregnant. A model that included a zero-inflation term was found to fit significantly better ($P < 2.2 \times 10^{-16}$, likelihood-ratio test). This was also substantiated by lower Akaike information criterion and Bayesian information criterion scores.

All models were fit using glmmTMB (/services/tools/R/4.0.0/R_PACKAGES.txt:glmmTMB '1.1.5'), and no rate models, except the number of cycles until pregnant, had indications of zero inflation.

GWAS catalog accessions for calculated bioavailable testosterone, total testosterone and SHBG are GCST90012102, GCST90012106 and GCST90012112, respectively.

MR analysis

MR analysis was performed using two-sample inverse-weighted methods⁹². In addition, the intercept from the MR-EGGER⁹³ was calculated to provide a test of directional pleiotropy and the I^2 metric to assess general heterogeneity of the variants. Data for the outcomes were taken from the most recent genome-wide study for each outcome (thus, in most cases, these data were not sex-specific). To correct for any impact of the role of BMI on analyses multivariate inverse-weighted method⁹⁴ was implemented. The β s for these variants to BMI were taken from the most recent GIANT consortium study that combined GWAS meta-analysis data with that from UK Biobank⁹⁵. For the analysis of association between menopause variants and PCOS split by evidence for a DNA damage effect, the variants were classified based on the proximity to a known DNA damage repair gene as discussed in ref. 7.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Genome-wide data from the two-strata model (that is, the meta-analysis without the data from 23andMe) will be available from the GWAS catalog, accessions GCST90570579 and GCST90570580. Data from 23andMe can be made available to qualified investigators who enter into an agreement with 23andMe that protects participant confidentiality; details on the access procedures for 23andMe, including links to data access request forms and details of time frames for requests, can be found at <https://research.23andme.com/dataset-access/>.

Code availability

No custom code was used in this project, details of the reference for code used can be found in the methods.

References

72. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
73. Magi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
74. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
75. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
76. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
77. Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
78. Dewey, F. E. et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* **354**, aaf6814 (2016).
79. Dennis, J. K. et al. Clinical laboratory test-wide association scan of polygenic scores identifies biomarkers of complex disease. *Genome Med.* **13**, 6 (2021).
80. Staley, J. R. et al. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* **32**, 3207–3209 (2016).
81. Kamat, M. A. et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* **35**, 4851–4853 (2019).
82. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
83. Sun, B. B. et al. Genetic associations of protein-coding variants in human disease. *Nature* **603**, 95–102 (2022).
84. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
85. Pietzner, M. et al. Mapping the proteo-genomic convergence of human diseases. *Science* **374**, eabj1541 (2021).
86. Weeks, E. M. et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat. Genet.* **55**, 1267–1276 (2023).
87. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
88. Choi, S. W., Mak, T. S. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**, 2759–2772 (2020).
89. Prive, F., Arbel, J. & Vilhjalmsón, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2020).
90. Bliddal, M., Broe, A., Pottgard, A., Olsen, J. & Langhoff-Roos, J. The Danish Medical Birth Register. *Eur. J. Epidemiol.* **33**, 27–36 (2018).
91. Schmidt, M. et al. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin. Epidemiol.* **7**, 449–490 (2015).
92. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
93. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32**, 377–389 (2017).
94. Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* **181**, 251–260 (2015).
95. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700,000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).

Acknowledgements

This study was conducted using the data from the UK Biobank (applications 9905 and 31832). The authors would like to thank the research participants and employees of 23andMe for making this study possible. This work was supported by the Medical Research

Council (unit programs—MC_UU_12015/2, MC_UU_00006/2, MC_UU_12015/1 and MC_UU_00006/1) and the NIHR Cambridge Biomedical Research Centre. T.K. is supported by the Novo Nordisk Foundation Data Science Investigator (grant NNF20OC0062294). J.Z. was supported by the National Institute of Child Health and Human Development (NICHD; grant K08HD110723), the American Heart Association (grant 23CDA1054471), a Clinical Scholar Award from the Pediatric Endocrine Society and a Career Development Fellowship from the Boston Children's Hospital Office of Faculty Development. C.K.W. was supported by the NICHD (grant R01HD065029). B.H.M. was supported by the Department of Health, Western Australia (merit award 1186046) and the Australian National Health and Medical Research Council (ideas grant 2003629). Further Acknowledgments details are provided in the Supplementary Note.

Author contributions

L.M.E.M., J.Z., B.H.M., N.P.G., K.K.O., M.O.G., L.K.D., A.D., C.M.L., J.S.E.L., S.F., J.A.V., C.K.W., T.K. and F.R.D. wrote the first draft of the manuscript. L.M.E.M., J.Z., B.H.M., N.P.G., K.V.A., J.A.M., H.R., B.T., K.A.K., Y.Z., D.W., J.S.T., G.T., Y.Z., L.W., A.D.e.V., K.B., R.S., R.D., M.H.P., M.R.J., K.B., A.E.J., T.K. and F.R.D. performed the analysis. L.M.E.M., J.Z., B.H.M., N.P.G., J.A.M., H.R., B.T., K.A. Kentistou, Y.Z., D.W., J.S.T., G.T., Y.Z., L.W., A.D.e.V., M.-R.J., A.E.J., H.L., K.K.O., M.O.G., L.K.D., K.B., R.S., A.D., C.M.L., J.S.E.L., S.F., J.A.V., C.K.W., T.K. and F.R.D. provided interpretation. R.D., M.H.P., M.R.J., K.S.R., M.A., R.A., K.B., M.B., L.B., S.B., Y.-M.C., D.I.C., M.D., D.A.E., B.C.F., L.G.F., M.G.H., C.H., H.H., I.K., P.K., R.S.L., N.L., R.J.L., Y.V.L., R.M., M.I.M., L.M.-P., J.V.M., C.M., G.N.N., B.M.N., H.S.N., M.N., S.R.O., O.B.V.P., E.S., C.M., C.E., K.A. Kaspersen, M.T.B., B.A., H.U., B.O.-P., A.P., M.P.R., A.S., R.S., T.D.S., B.G.A.S., U.T., A.G.U., M.U., S.Z., D.A.V.H., J.N.H., K.S., J.R.B.P., U.S., S.G.W., T.P., T.L., M.-R.J., K.B., A.E.J., H.L., M.O.G., L.K.D., A.D., C.M.L., J.S.E.L., S.F. and C.K.W. provided data, and contributed to the reading, commentary and critical review of the manuscript. U.S., J.R.B.P., F.R.D., A.D., J.S.E.L., S.F., C.M.L. and C.K.W. form the steering committee of the International PCOS Consortium. All authors participated in reading, revising and approving the final submission of the manuscript.

Competing interests

The 23andMe Research team members are employees of and hold stock or stock options in the company. S.B. reports ownerships in Intomics A/S, Hoba Therapeutics Aps, Novo Nordisk A/S, Lundbeck

A/S and ALK abello A/S. G.T., U.T., K.S. and U.S. are employees of deCODE genetics/Amgen. M.I.M. serves on advisory panels for Pfizer and Novo Nordisk; has received honoraria from Pfizer, Novo Nordisk and Eli Lilly; has received research funding from Pfizer, Novo Nordisk, Eli Lilly, AstraZeneca, Sanofi Aventis, Boehringer Ingelheim, Merck, Roche, Janssen, Takeda and Servier; and is now an employee of Genentech and a holder of Roche stock. J.L. has received consultancy fees from Danone, Metagenics, Titus Healthcare, Roche and Euroscreen. R.A. has received consultancy fees from Spruce Biosciences, Fortress Biotech, Rani Therapeutics, Core Access Surgical Technologies, Proera and May Health, and has equity in Arora Forge Advisors and Martin Imaging. M.O.G. has served on an advisory board for Nestle Health Science. B.C.F. has received fees and/or grant support during the last 4 years from the following organizations: Bain Capital, Controversies in Obstetrics & Gynecology & Infertility, Dutch Heart Foundation (Nederlandse Hartstichting), Elsevier, European Society of Human Reproduction and Embryology, Ferring, International Federation of Fertility Societies, London Women's Clinic, Myovant, the Netherlands Organisation for Health Research and Development (ZonMW), Pantharei Bioscience, Partners Group, PregLem/Gideon Richter, Reproductive Biomedicine Online, Shieldler and UpToDate. T.D.S. is a cofounder and shareholder of Zoe. J.N.H. has equity in Camp4 Therapeutics. K.A.K. and J.R.B.P. are employees and shareholders of Insmed. J.R.B.P. receives research funding from GSK. Y.Z. was a UK University worker at GSK. J.A.V. has received royalties from AMH assays, paid to the institute/laboratory, with no personal financial gain. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-026-02543-9>.

Correspondence and requests for materials should be addressed to Corrine K. Welt, Tugce Karaderi or Felix R. Day.

Peer review information *Nature Genetics* thanks Jenny Fung and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used.
Data analysis	<p>Central quality control (QC) was performed using the EasyQC pipeline (version 9.2). A fixed-effect, inverse-weighted-variance meta-analysis approach was used with the collected summary statistics from the individual studies, using either GWAMA version is 1_2_6) or METAL (version released on 2011-03-25). Identified variants were annotated using FUMA (v1.5.2). Forest plots for comparing the effect sizes across the strata in the meta-analysis were made using the ggplot2 package in R (version 4.3.1).</p> <p>Fine-mapping was performed using the shotgun stochastic search method as performed in FINEMAP. Functional mapping and annotation of GWAS was performed with FUMA (v1.5.2), and further annotation of the association results with PhenoScanner (date accessed 25 March 2022).</p> <p>The GWAS-to-Genes pipeline, incorporating eQTL colocalisation from both SMR-HEIDI (Version 0.68) and the “coloc” package, was used to highlight likely causal genes at each of the identified signals. We used the PRS-CS software to calculate a polygenic risk score (PRS) for PCOS.</p> <p>In the same UK Biobank sample, we replicated these analyses using another PRS tool, LDpred-2, which employs Bayesian shrinkage model. PRS for PCOS were also calculated in the Copenhagen Hospital Biobank using LDpred2. GWAS summary statistics were pre-processed with MungeSumStats. The AER R package (v1.2.10) was used to determine whether there was equi-, under-, or over-dispersion. Lastly, using the Danish IVF registry, data models were fit using glmTMB (/services/tools/R/4.0.0/R_PACKAGES.txt:glmTMB “1.1.5”).</p>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Cohorts should be contacted individually for access to their raw data. UK Biobank data are available on application (<https://ams.ukbiobank.ac.uk/ams/>). Access to the two strata summary statistics can be found at the DOI in the data access section of the paper.

We used the NCBI RefSeq gene map for GRCh37 which is available via <http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/>. GTEx eQTL data was used (V7) and is available via <https://gtexportal.org>.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	All analysis are disaggregated by sex. All included studies had ethics approval and collected informed written consent.
Reporting on race, ethnicity, or other socially relevant groupings	Study samples were defined within each study that contributed to the meta-analysis, usually by position based on genetic principle components. As these were done at a study level, we have used the geographical based ancestry terminology.
Population characteristics	Described in Supplementary Table 1 "Cohort descriptives"
Recruitment	This study is a meta-analysis, combining data from different sources across the world. Recruitment varied across the different included studies.
Ethics oversight	<p>UK Biobank data has approval from the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB). 23andMe research participants provided informed consent and volunteered to participate in the research online under a protocol approved by the external AAHRPP-accredited IRB, Ethical & Independent (E&I) Review Services. For each of the other individual studies that contributed data, ethical approval was given by the relevant boards.</p> <p>The Boston cohort was approved by the Partners IRB (# 2002P001924 and 2012P002417) and the University of Utah IRB (IRB_00076659). The deCODE cohort was approved by the National Bioethics Committee of Iceland (VSN 03–007), which was conducted in agreement with conditions issued by the Data Protection Authority of Iceland. Personal identities of the participants' data and biological samples were encrypted by a third-party system (Identity Protection System), approved and monitored by the Data Protection Authority.</p> <p>The UK cohort was approved by the Parkside Health Authority (Now—NHS Health Research Authority, NRES Committee—West London & GTAC, UK, London, UK) under EC2359 "The Molecular Genetics of Polycystic Ovaries."</p> <p>The Rotterdam PCOS cohort, was approved by institutional review board (Medical Ethics Committee) of the Erasmus Medical Center (04-263). Controls from the Lifelines Cohort Study have been approved by the UMCG Medical ethical committee under number 2007/152.</p> <p>The Chicago PCOS cohort was approved by the Northwestern IRB (#STU00008096). The control subjects from the NUGene study were approved by the Northwestern IRB (# STU00010003).</p> <p>The Estonia cohort was approved by the Research Ethics Committee of the University of Tartu approved the study (198T-18).</p> <p>The Western Australian PCOS study was approved by the SCGOPHCG Human Research Ethics Committee (RGS0000001467) and controls by HRA North West – Liverpool East Research Ethics Committee (19/NW/0187; TwinsUK).</p> <p>The Nurses' Health Study (NHS I and II) was approved by the Partners Human Research Committee (#1999-P-011114).</p> <p>Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, older research cohorts, collected prior the start of FinnGen (in August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by the National Supervisory Authority for Welfare and Health, Fimea. Recruitment procedures followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) approved the FinnGen study protocol (Nr HUS/990/2017). The FinnGen study was approved by Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019, THL/1524/5.05.00/2020, and THL/2364/14.02/2020); Digital and population data service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3); the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020); and Statistics Finland (permit numbers: TK-53- 1041-17 and TK-53-90-20). The Biobank access decisions for</p>

FinnGen samples and data utilized in the FinnGen Data Freeze 6 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, Auria Biobank AB17-5154, Biobank Borealis of Northern Finland_2017_1013, Biobank of Eastern Finland 1186/2018, Finnish Clinical Biobank Tampere MH0004, Central Finland Biobank 1-2017, and Terveystalo Biobank STB 2018001.

Analyses in the EstBB were carried out under ethical approval 1.1-12/624 from the Estonian Committee on Bioethics and Human Research and data release N05 from the EstBB.

Cedars Sinai -The study was approved by the institutional review boards of the recruiting centers and Cedars-Sinai Medical Center (CSMC). Written informed consent was obtained from all participants.

BioVU – approved by the Institutional Review Board at Vanderbilt University (#160279).

At enrolment, Women's Genome Health Study (WGHS) participants consented to ongoing analyses linking blood-derived observations with clinical measures collected at baseline and throughout observation. All analysis in the WGHS had been approved by the institutional review board of Brigham and Women's Hospital, Boston, MA.

The Michigan Genomics Initiative (MGI) adhered to the principles of the Declaration of Helsinki. Consent forms and study protocols for MGI participants received approval from the Institutional Review Board of the University of Michigan Medical School (IRB IDs: HUM00071298, HUM00099197, HUM00151834, HUM00156162, HUM00164162, HUM00143789, HUM00167679, HUM00041845, HUM00044507, HUM00176147, HUM00148297, HUM00141565, and HUM00097962).

The Genes & Health study was approved by the London South East NRES Committee of the Health Research Authority (14/LO/1240).

BioMe was approved by the Icahn School of Medicine at Mount Sinai's Institutional Review Board (23-00583). All participants provided written informed consent.

The Danish Blood Donor Study (DBDS): Ethics committee approval NVK-1700407, data protection agency P-2019-99. Repro: Ethics committee approval NVK-1805807, data protection agency P-2019-49.

All participants provided informed consent to participate in the original MyCode Community Health Initiative as approved by the Geisinger Institutional Review Board. The project described in this paper was reviewed and determined to be not human subjects research by the Geisinger Institutional Review Board (IRB#: #2017-158).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This study included summary statistics of our previous meta-analysis including 10,074 cases and 103,164 controls of European ancestry (described in detail in Day et al. 2019). An additional 13 cohorts were included for the current GWAS meta-analysis, resulting in 21,570 cases and 523,971 controls. The newly added cases included all-ancestries, of which 87% of the cases were from European origin, 10% from East-Asian origin and the remaining being either from Hispanic or African-American ancestry.
Data exclusions	Individuals failing standard genotyping quality control parameters defined in the individual studies or missing genotype, phenotype, or covariate data were excluded from analysis. This decision was made prior to performing any downstream analysis.
Replication	We conducted our analysis using a sufficiently powered dataset. Additionally, no suitable dataset was available for replication. However, we did replicate our PRS in the Biobank, testing for the presence of PCOS.
Randomization	Participants were not allocated into experimental groups. The principle exposure in this study is naturally occurring genetic variants, meaning that we were unable to randomize the individuals in the study.
Blinding	Blinding is by design not possible in this study, as it is a genome wide association study of common and rare genetic variants and not a randomized study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Plants

Seed stocks	<input type="text" value="N/A"/>
Novel plant genotypes	<input type="text" value="N/A"/>
Authentication	<input type="text" value="N/A"/>