

RESEARCH

Open Access



# Body mass index stratified meta-analysis of genome-wide association studies of polycystic ovary syndrome in women of European ancestry

Kharis Burns<sup>1,2\*</sup>, Benjamin H. Mullin<sup>3,4</sup>, Loes M. E. Moolhuijsen<sup>5</sup>, Triin Laisk<sup>6</sup>, Jaakko S. Tyrmi<sup>7,8</sup>, Jinrui Cui<sup>9</sup>, Ky'Era V. Actkins<sup>10,11,12</sup>, Yvonne V. Louwers<sup>13</sup>, Estonian Biobank Research Team, Lea K. Davis<sup>11,12</sup>, Frank Dudbridge<sup>14</sup>, Ricardo Azziz<sup>15</sup>, Mark O. Goodarzi<sup>9</sup>, Hannele Laivuori<sup>7,16,17,18</sup>, Reedik Mägi<sup>6</sup>, Jenny A. Visser<sup>5</sup>, Joop S. E. Laven<sup>13</sup>, Scott G. Wilson<sup>3,4,19</sup>, FinnGen, International PCOS Consortium, Felix R. Day<sup>20</sup> and Bronwyn G. A. Stuckey<sup>2,3,21</sup>

## Abstract

**Background** Polycystic ovary syndrome (PCOS) is a complex multifactorial disorder with a substantial genetic component. However, the clinical manifestations of PCOS are heterogeneous with notable differences between lean and obese women, implying a different pathophysiology manifesting in differential body mass index (BMI). We performed a meta-analysis of genome-wide association study (GWAS) data from six well-characterised cohorts, using a case–control study design stratified by BMI, aiming to identify genetic variants associated with lean and overweight/obese PCOS subtypes.

**Results** The study comprised 254,588 women (5,937 cases and 248,651 controls) from individual studies performed in Australia, Estonia, Finland, the Netherlands and United States of America, and separated according to three BMI stratifications (lean, overweight and obese). Genome-wide association analyses were performed for each stratification within each cohort, with the data for each BMI group meta-analysed using METAL software. Almost half of the total study population (47%,  $n = 119,584$ ) were of lean BMI ( $\leq 25$  kg/m<sup>2</sup>). Two genome-wide significant loci were identified for lean PCOS, led by rs12000707 within *DENND1A* ( $P = 1.55 \times 10^{-12}$ ) and rs2228260 within *XBP1* ( $P = 3.68 \times 10^{-8}$ ). One additional locus, *LINC02905*, was highlighted as significantly associated with lean PCOS through gene-based analyses ( $P = 1.76 \times 10^{-6}$ ). There were no significant loci observed for the overweight or obese sub-strata when analysed separately, however, when these strata were combined, an association signal led by rs569675099 within *DENND1A* reached genome-wide significance ( $P = 3.22 \times 10^{-9}$ ) and a gene-based association was identified with *ERBB4* ( $P = 1.59 \times 10^{-6}$ ). Nineteen of 28 signals identified in previous GWAS, were replicated with consistent allelic effect in the lean stratum.

\*Correspondence:

Kharis Burns

kharis.burns@research.uwa.edu.au

Full list of author information is available at the end of the article



© Crown 2024. **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 Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

There were less replicated signals in the overweight and obese groups, and only 4 SNPs were replicated in each of the three BMI strata.

**Conclusions** Genetic variation at the *XBP1*, *LINC02905* and *ERBB4* loci were associated with PCOS within unique BMI strata, while *DENND1A* demonstrated associations across multiple strata, providing evidence of both distinct and shared genetic features between lean and overweight/obese PCOS-affected women. This study demonstrated that PCOS-affected women with contrasting body weight are not only phenotypically distinct but also show variation in genetic architecture; lean PCOS women typically display elevated gonadotrophin ratios, lower insulin resistance, higher androgen levels, including adrenal androgens, and more favourable lipid profiles. Overall, these findings add to the growing body of evidence supporting a genetic basis for PCOS as well as differences in genetic patterns relevant to PCOS BMI-subtype.

**Keywords** Polycystic ovary syndrome, PCOS, GWAS, Meta-analysis, Lean, Obese, Body mass index, BMI

**Background**

Polycystic ovary syndrome (PCOS) is a common female endocrinopathy, affecting around 5–15% of women, though its aetiology remains to be fully explained [1]. Cardinal features include hyperandrogenism, oligoamenorrhoea, and often obesity and hyperinsulinaemia [1, 2]. Familial inheritance suggests a genetic basis and genome-wide association studies (GWAS) have identified numerous genetic loci significantly associated with this condition [3–7]. However, the relatively modest number of polymorphisms with robust association data identified to date do not appear to entirely explain the disease aetiology [4]. Most affected women are overweight or obese, with only 16–30% in the lean to normal BMI range [8–10]. Indeed, the clinical manifestations of PCOS are notably different between lean and obese women, potentially implying a different pathophysiology associated with differential body mass index (BMI) [11]. It seems possible that a difference in aetiology is attributable to distinct combinations of genotypes. Improved understanding of the genetic architecture of PCOS subtypes may assist in predicting comorbidity risk, facilitating earlier intervention and tailored patient management. Indeed, principal

component analysis has demonstrated clusters of risk factors explaining the variance in PCOS – involving women with i) high BMI, insulin resistance, low high-density lipoprotein and low sex hormone binding globulin, ii) hypertension, elevated low-density lipoprotein and hypertriglyceridemia and iii) a lean PCOS phenotype with elevated luteinizing hormone: follicle stimulating hormone ratio and total testosterone [11]. The lean PCOS phenotype therefore appears to be distinct, potentially necessitating different treatment paradigms, particularly with respect to traditional lifestyle and weight loss recommendations.

In this meta-analysis of PCOS case–control GWAS data we aimed to analyse genotype differences in women with the syndrome based on BMI stratification thus providing insight into the hypothesis that lean and overweight/obese PCOS phenotypes are genetically distinct.

**Results**

**Characteristics of the cohorts**

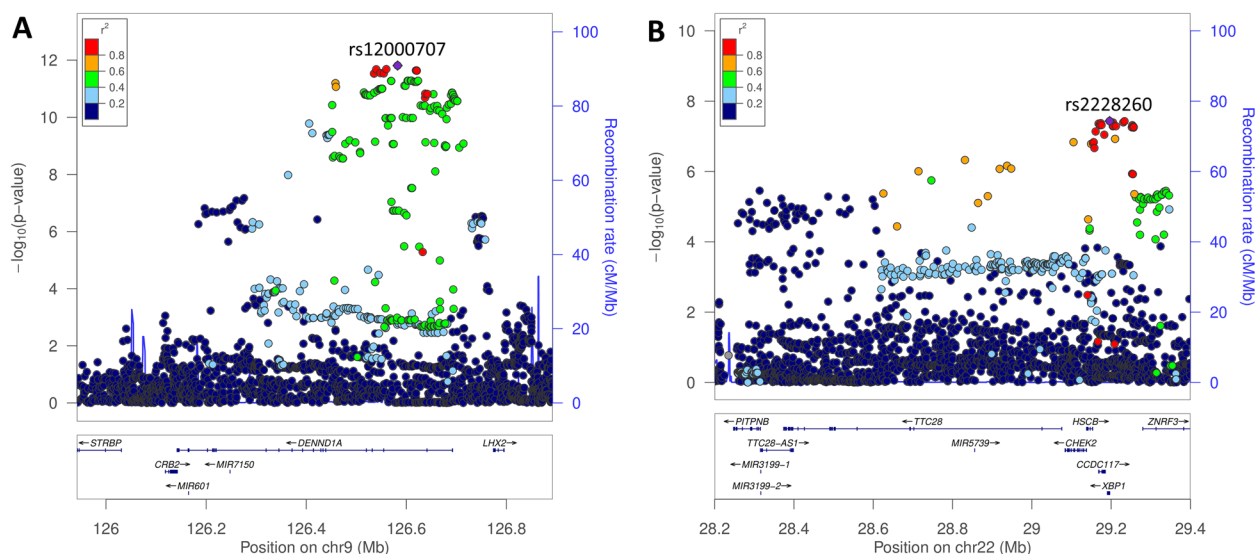
A total of 254,588 women were included in the meta-analysis, comprising 5,937 PCOS cases and 248,651

**Table 1** Details of the cohorts included in the BMI stratified PCOS meta-analysis

Group	Lean <sup>a</sup>		Overweight		Obese		Totals
	Case	Control	Case	Control	Case	Control	
WA PCOS	82	1,287	50	796	139	409	2,763
Estonian Biobank	2147	59,319	776	31,579	742	22,980	117,543
FinnGen	171	47,533	164	38,656	308	31,605	118,437
Cedars Sinai	101	138	66	69	192	60	626
Dutch cohort <sup>b</sup>	352	4,643	139	2,130	143	912	8,319
BioVU	66	3,745	73	1,388	226	1,402	6,900
<b>Totals</b>	<b>2,919</b>	<b>116,665</b>	<b>1,268</b>	<b>74,618</b>	<b>1,750</b>	<b>57,368</b>	<b>254,588</b>

<sup>a</sup> Lean = BMI ≤ 25 kg/m<sup>2</sup>, Overweight = 25 < BMI < 30 kg/m<sup>2</sup>, Obese = BMI ≥ 30 kg/m<sup>2</sup>

<sup>b</sup> Cases provided by the Rotterdam PCOS Cohort and controls provided by Lifelines Cohort Study



**Fig. 1** Regional association plots of genome-wide significant loci identified in the lean PCOS meta-analysis. **A** 9q33.3 (*DENND1A*) and **B** 22q12.1 (*XBPI*). Genetic variants are depicted by position (x-axis) together with their meta-analysis *P*-value ( $-\log_{10}$ ; y-axis). Variants are colour coded according to their LD ( $r^2$ ) with the lead variant. Mb = million bases

controls stratified into BMI subgroups i) BMI  $\leq 25$  kg/m<sup>2</sup> (lean), ii) BMI 25 to 30 kg/m<sup>2</sup> (overweight), and iii) BMI  $\geq 30$  kg/m<sup>2</sup> (obese) (Table 1). Almost half of the combined study subjects (47%,  $n = 119,584$ ) were of lean BMI. The majority of the cases and controls were from Estonian or Finnish datasets, with the remainder comprising American, Australian and Dutch Caucasian subjects (Table 1).

### Meta-analysis

QQ plots for both the SNP and gene-based analyses completed are presented in Supplementary Fig. 1.

### Single-variant based meta-analysis

For the purposes of this study, genetic loci are defined as regions of the genome containing association signals for PCOS. This study identified two genome-wide significant genetic loci ( $P < 5 \times 10^{-8}$ ) for lean PCOS ( $n = 2,919$  cases and 166,655 controls) on chromosome 9 in *DENND1A* (led by rs12000707;  $P = 1.55 \times 10^{-12}$ ) and on chromosome 22 in *XBPI* (led by rs2228260;  $P = 3.68 \times 10^{-8}$ ) (Supplementary Fig. 2 and Supplementary Fig. 3). There were no genome-wide significant loci identified for the overweight or obese sub-strata when analysed separately. When the overweight and obese groups were combined (i.e., non-lean subjects), one genome-wide significant locus was identified on chromosome 9, also in *DENND1A* (led by rs569675099;  $P = 3.22 \times 10^{-9}$ ) (Supplementary Fig. 3 and Supplementary Fig. 4). This variant is in moderate linkage disequilibrium (LD) ( $r^2 = 0.51$ ) with rs12000707 ( $P = 3.72 \times 10^{-8}$ ), which was the lead variant in the lean

strata meta-analysis. The variant rs569675099 did not meet GW significance in the lean group ( $P = 1.03 \times 10^{-5}$ ) though has previously been identified as associated with PCOS in women of European and Han Chinese ancestry [3, 6]. Co-localisation analysis [12] of the GWAS meta-analysis results for the *DENND1A* locus in the lean and non-lean groups generated a 95.5% posterior probability of co-localised association signals in the two datasets, indicating the presence of a shared causal variant.

The lead variant in the lean PCOS meta-analysis, rs12000707 (Fig. 1; Supplementary Table 1), is a non-coding intronic variant that has not previously been highlighted by GWAS but does have GTEx data supporting a role as an expression quantitative trait locus (eQTL) in subcutaneous adipose tissue (*DENND1A*;  $P = 7.0 \times 10^{-6}$ ) [13]. This locus contained a total of 124 genome-wide significant variants in the results for the lean PCOS meta-analysis. The lead single nucleotide polymorphism (SNP), rs12000707 is in complete LD with rs9696009 ( $r^2 = 1$ ), previously reported in a GWAS of PCOS conducted in European populations [4]. Data for this locus from FUMA [14] analysis illustrates the large size of the region and number of variants at this site in LD (Supplementary Fig. 5). Based on this data alone it is not possible to determine which variant(s) are the functional drivers within this LD block. The SNP rs12000707 also demonstrated nominally significant associations in the overweight and obese groups (Supplementary Table 1), with meta-analysis of the 3 BMI strata suggesting that there is no significant heterogeneity between the groups (het  $P = 0.56$ ).

The other signal highlighted in the meta-analysis of lean PCOS, on chromosome 22, led by rs2228260, a synonymous SNP in *XBPI* (Fig. 1; Supplementary Table 1), contained a total of 11 genome-wide significant variants. Examination of the LD between these variants suggested that they were all likely representative of the same signal (all  $r^2=1$ ). This association signal is part of a large LD block spanning multiple genes including *TTC28*, *CHEK2*, *HSCB*, *CCDC117*, *XBPI*, *ZNRF3* and *EMID1*. Any one of these genes could be driving the association signal, however publicly available eQTL data from GTEx shows that rs2228260 is an eQTL for *CHEK2* (adrenal gland;  $P=8.3\times 10^{-7}$ ) [13]. This particular variant has not been identified in any previous GWAS. *XBPI* has documented involvement in glucose and lipid metabolism, providing a potential biological link with PCOS, where metabolic disturbances including dyslipidaemia and insulin resistance are noted [11]. However, other genes within this LD block, including both *CHEK2* and *ZNRF3*, have previously been implicated in PCOS by data from the Finnish and Estonian cohorts analysed in isolation, populations included in this meta-analysis [15]. Tyrmi et al., highlighted two putative independent causal variants in the checkpoint kinase 2 (*CHEK2*) gene, which they proposed were the basis of the association [15]. The lead variant for the *CHEK2* locus identified in that study for the Estonian cohort, rs182075939, is not in strong linkage disequilibrium with rs2228260 ( $r^2\leq 0.2$ ) [16]. Considering this, we performed a conditional analysis for this locus in the lean PCOS meta-analysis using the COJO function of the GCTA package [17]. After conditioning on the *CHEK2* variant rs182075939, the association between rs2228260 and lean PCOS was no longer genome-wide significant ( $P_{\text{cond}}=1.9\times 10^{-5}$ ), and there was a reduction in effect size (conditioned beta=0.22, reduced from 0.28). Hence it is not possible to establish rs2228260 as an independent association signal. The lead variant for the Finnish cohort, rs145598156 [15], located closest to *ZNRF3*, and rs2228260 are in linkage equilibrium in Europeans ( $r^2=0.0$ ) [16]. However, it should be noted that rs145598156 is very rare in non-Finnish Europeans (MAF=0.003), making accurate estimation of LD difficult. The remaining genes within this LD block have no established biological link with PCOS. The variant rs145598156 was not analysed in this study due to its very low frequency.

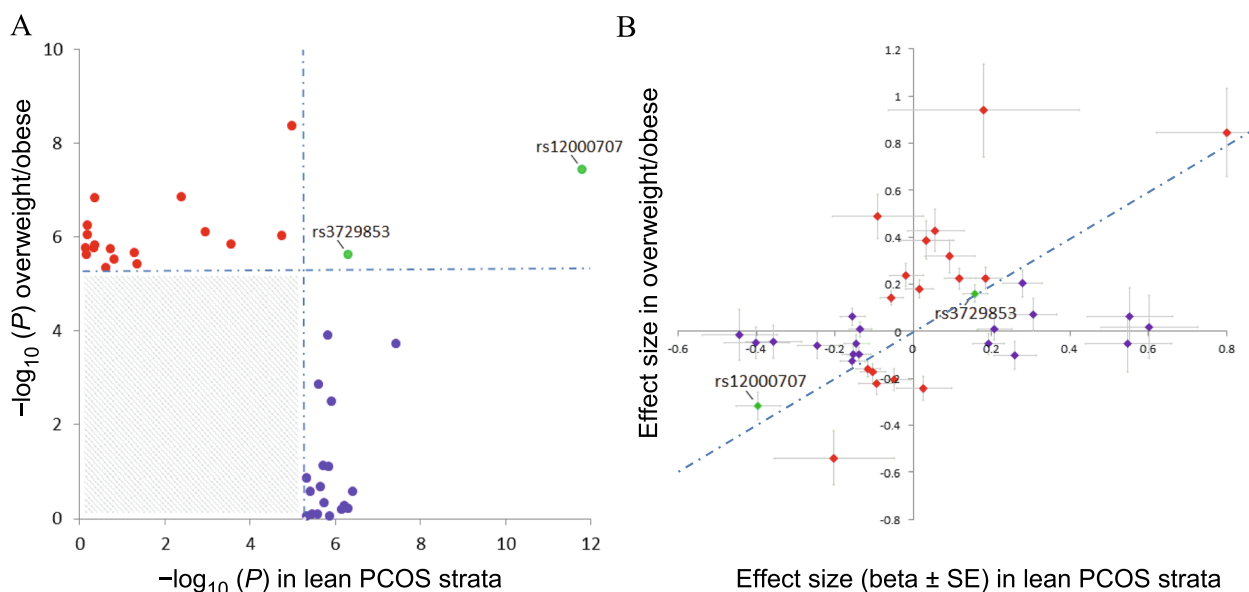
The significant associations identified in the lean PCOS meta-analysis were examined within each of the contributing cohorts (Supplementary Table 2). The Estonian Biobank demonstrated the strongest associations of the six cohorts for the two lead variants, which is not surprising considering that this cohort contributed the largest number of lean PCOS cases to the study, while the Western Australian PCOS research group (WA PCOS)

cohort demonstrated the greatest effect size for these two variants.

Genome-wide suggestive associations may represent true associations that have failed to reach the stringent genome-wide significance threshold for various reasons including statistical power, and which could be validated through further replication. Genetic variants meeting the criterion for genome-wide suggestive association with lean PCOS ( $P<5\times 10^{-6}$ ) are presented in Supplementary Table 1. A number of these signals have previously been identified in GWAS of PCOS affected women of both European and Chinese ancestry, specifically variants in *YAPI*, *KRR1*, *IRF1* and *BLK* [4–7]. However, there has been no previous research published specifically identifying these loci in lean PCOS affected subjects. Furthermore, from the analysis of the combined overweight/obese cohorts, (Supplementary Table 3), genome wide suggestive signals were identified for three previously reported PCOS loci. rs11031006 within *FSHB* was identified as genome-wide suggestive ( $P=1.42\times 10^{-7}$ ), which has been previously reported as a risk variant for PCOS [5, 6, 18]. The other signals in known PCOS loci include rs11453664 on chromosome 2 within *ERBB4* ( $P=7.85\times 10^{-7}$ ) and rs3729853 in *GATA4* on chromosome 8 ( $P=2.4\times 10^{-6}$ ). These specific SNPs have not previously been reported as risk variants for PCOS, though the signal reported in *GATA4*, rs3729853, is in modest LD with the previously published significant SNP rs804279 ( $r^2=0.36$ ) [4, 16]. Two lead variants were shared by the lean and combined overweight/obese PCOS strata, and with consistent allelic effects observed, when examining results that were of at least suggestive association ( $P<5\times 10^{-6}$ ; Fig. 2). These two variants were located in the *DENND1A* (rs12000707) and *GATA4* (rs3729853) loci. However, the other lead variants that were of at least suggestive association for the lean PCOS stratum and the combined overweight/obese PCOS strata were observed to be uniquely represented in only one stratum or the other (Fig. 2A), suggesting a potential difference in the genetic architecture of lean versus overweight/obese PCOS; the observed effect size of the risk alleles for those lead variants also followed a similar pattern of segregation (Fig. 2B).

#### Gene-based association testing

Gene-based association tests are commonly used following single-variant GWAS analysis to model the sum of the effects of all variants within a gene to determine if, despite individual variants not achieving significance, there is statistical evidence of a composite association signal. Gene-based association testing identified two significant associations in the lean PCOS group at the multiple testing corrected (Bonferroni) significance threshold ( $P<1.96\times 10^{-6}$ ) (Supplementary Fig. 6).



**Fig. 2** Graph of lead variants in the meta-analyses showing at least suggestive association from each locus. **A**  $P$ -values are shown from the meta-analyses of the lean PCOS strata (x-axis) or the combined overweight/obese PCOS strata (y-axis). Blue symbols show lead variants with at least suggestive association ( $P < 5 \times 10^{-6}$ ) in the lean PCOS strata, red symbols show variants with at least suggestive association in the combined overweight/obese PCOS strata and green variants are those with at least suggestive association in both lean PCOS and combined overweight/obese PCOS strata. Lead variants that did not achieve suggestive association fall in the grey region. Dashed lines show the threshold for genome-wide suggestive association ( $P < 5 \times 10^{-6}$ ). **B** Plot of the effect size (beta  $\pm$  SE) for the risk allele of lead variants showing at least suggestive association in the lean PCOS strata (blue symbols), combined overweight/obese PCOS strata (red) and variants that show at least suggestive association in both the lean PCOS and combined overweight/obese PCOS strata (green). Dashed line shows the diagonal

The leading signal was again *DENNDIA* on chromosome 9 ( $P = 4.04 \times 10^{-10}$ ) followed by *LINC02905* (also known as *C8orf49*) ( $P = 1.76 \times 10^{-6}$ ), a long intergenic non-protein coding RNA gene located between *GATA4* and *NEIL2* on chromosome 8. The *GATA4/NEIL2* locus has been identified as associated with PCOS in previous GWAS in European populations, with heterogeneous effects depending on diagnostic criteria applied [4]. The *GATA4/NEIL2* locus also has previous links with ovulatory dysfunction and polycystic ovary morphology [4]. Accordingly, *LINC02905* may be part of a PCOS-susceptibility gene cluster on chromosome 8.

There were no significant gene-based association signals identified for the overweight or obese groups when examined separately, though when combined, one significant signal was identified on chromosome 2 for *ERBB4* ( $P = 1.59 \times 10^{-6}$ ; Supplementary Fig. 7). This gene has previously been identified in GWAS as associated with PCOS, in women of both European and Chinese ancestry [4, 19, 20].

#### Replication of established PCOS loci

All loci associated with PCOS identified in the previously published PCOS GWAS [3–7] were investigated within

each of the three BMI strata (Supplementary Table 4). Each locus was examined and observation of  $P < 0.05$  in any strata of this meta-analysis was considered nominally significant in terms of replication. For any known loci demonstrating nominal association within the cohort, the beta value or odds ratio was checked for consistent allelic direction of effect with that previously reported. All SNPs known to be associated with PCOS through GWAS were analysed in this study with differing levels of significance across the three BMI strata.

Nineteen of these 28 signals were replicated in the lean cohort, demonstrating consistent allelic effect. There were fewer signals replicated in the overweight and obese groups, with 9 and 7 signals respectively with consistent allelic effect. Only four SNPs, rs9696009 within *DENNDIA*, rs2178575 within *ERBB4*, rs11031005 within *ARL14EP/FSHB* and rs1795379 within *KRR1* were replicated in all three BMI strata. The signal within *DENNDIA* was most significant in the lean group. The other previously identified SNP in *DENNDIA*, rs2479106, did not meet criteria for replication in any BMI tier. The signal within *KRR1* was also most significant in the lean group. The signals in *ERBB4* and *ARL14EP/FSHB* showed similar significance across all BMI strata.

## Discussion

This study aimed to identify differences in genetic architecture between lean, overweight and obese PCOS affected patients in order to provide further insight into potential differences in aetiology between these diverse phenotypes. Single-variant based analysis found evidence of two genome-wide significant genetic associations with the lean phenotype, and one significant association when the overweight and obese groups were combined. Gene-based testing confirmed two genes associated with the lean PCOS group, and one gene associated with the overweight and obese groups combined. Additional variants that demonstrated genome-wide suggestive association were observed in the strata, with lead variants for the lean and combined overweight/obese PCOS strata typically demonstrating greater effects in only one stratum (Figs. 2A, B), therefore, these data may suggest a difference in the genetic architecture underlying lean versus overweight/obese PCOS.

The strongest association signal in the lean analysis, led by rs12000707 on chromosome 9, is located in *DENND1A*, which has robust evidence for genetic involvement in PCOS [4]. This gene encodes the protein DENN/MADD domain containing 1A, which plays a role in endocytosis and receptor turnover and has been identified as associated with PCOS in a number of previous GWAS involving women of European and Han Chinese ancestry [3, 4]. Replication studies have further supported these findings, with certain SNPs identified as associated with increased PCOS risk, highlighting it as one of the most well recognised genes implicated in PCOS [3, 4, 20]. Variants within *DENND1A* have also been associated with hyperandrogenism and ovulatory dysfunction [4, 21]. Functional studies have shown the involvement of *DENND1A* in the pathophysiology of PCOS phenotypes, with laboratory studies demonstrating that ovarian thecal cells in PCOS affected women secrete higher androgen amounts than those from non-affected women, potentially related to upregulation of enzymatic activity in steroid pathways [22]. A *DENND1A* isoform, termed *DENND1A.V2*, has been implicated in the increased expression of genes *CYP17A1* and *CYP11A1*, which both play a role in the formation of key enzymes involved in androgen steroidogenesis [23]. This is thought to play a role in PCOS thecal cell androgen production, a feature of PCOS [23]. Furthermore, forced expression of this *DENND1A.V2* isoform in normal human theca cells has been shown to increase androgen and progesterone production, thus converting the normal theca cell into a PCOS phenotype [23]. Conversely, a knockdown model whereby *DENND1A.V2* expression was silenced in PCOS theca cells demonstrated a reduction in steroidogenesis [23]. A model creating transgenic h*DENND1A.V2* mice

lines has further demonstrated these concepts; elevation of both ovarian and adrenal *Cyp17a1* mRNA levels as well as transgenic ovarian thecal cell steroid production was observed in those mice expressing h*DENND1A.V2* transcripts demonstrating impacts on both ovarian and adrenal steroidogenesis [24].

Rare variants within *DENND1A*, identified through whole genome sequencing techniques, have also been found to be associated with certain quantitative traits within PCOS-affected women, specifically higher luteinising hormone (LH): follicle stimulating hormone (FSH) ratios [25]. Clustering analysis, demonstrating 'reproductive' and 'metabolic' PCOS subgroups, has demonstrated carriers of rare variants in *DENND1A* were more likely to have a reproductive subtype, characterised by lower BMI and insulin levels and higher LH and sex hormone binding globulin levels [26]. The finding of a strong association and large effect size for rs12000707 in the lean PCOS strata with *DENND1A* warrants further investigation. This builds on recent findings of specific *DENND1A* variants being more prevalent within a reproductive PCOS phenotype, with lower BMI [26]. Further studies are necessary to replicate these signals and explore the biology of this gene in PCOS subtypes.

The other genome-wide significant signal in the lean study, led by rs2228260, was located on chromosome 22, within *XBPI*. This signal is composed of a large block of genetic variants in strong LD spanning multiple genes, any one of which could be the effector gene. The SNP identified at this locus is a synonymous coding variant i.e., a codon change that does not alter the encoded amino acid [27]. Synonymous coding variants are not generally regarded as the most likely effectors for transcriptional regulation or altered protein function, but nevertheless, can have effects on protein expression and function. Previously considered 'silent' variants, it is now appreciated that these variants can affect mRNA stability and structure, protein folding, conformation and function [28, 29]. Alternatively, this variant may simply be tagging a functional variant that is yet to be identified.

The product of the X-box binding protein 1 (*XBPI*) gene is a transcription factor involved in the 'Unfolded Protein Response' (UPR), a series of finely tuned homeostatic mechanisms triggered by stress within the endoplasmic reticulum (ER) [30]. Dysfunctional ER response has been highlighted as a contributor to the pathogenesis of metabolic disease such as type 2 diabetes, obesity and atherosclerotic cardiovascular disease [30]. Components of the UPR are also known to be involved in the upregulation of metabolic processes, including gluconeogenesis and lipid synthesis, which can be perturbed in PCOS [30]. The protein product of *XBPI* may alter adipocyte, hepatocyte and pancreatic

cell signalling pathways to regulate glucose homeostasis and improve insulin sensitivity [30–32]. Deficiency of XBP1 in pancreatic alpha and beta cells has been implicated in impaired insulin secretion and signalling [32]. Increased UPR gene expression has also been seen in granulosa cells in PCOS-affected women and these processes, involving ER stress and associated adaptational mechanisms, have been highlighted as regulators of ovarian physiological and pathophysiological outcomes [33].

XBP1 levels have been reported as higher in women with PCOS [34]. A recent study examining XBP1 levels in three study groups of women: obese PCOS patients, non-obese PCOS patients and normal weight controls, found significantly higher levels in PCOS patients. Comparison between obese and non-obese PCOS affected women found higher levels in the former group and a significant positive correlation was seen between XBP1 levels and BMI, waist circumference, fasting plasma glucose and triglyceride levels [34]. In this context, links with obese PCOS and metabolic characteristics mean that the biological effects of XBP1 in the lean phenotype are not immediately obvious and merit further research. Overall, XBP1 appears to be involved in several processes that are perturbed in PCOS patients including oocyte maturation and aberrant glucose and lipid metabolism. Involvement in the lean phenotype appears to be a novel observation based on the literature to date and warrants further study.

Whilst our top lean PCOS SNP on chromosome 22, rs2228260, is located within *XBP1*, other genes in the region may also be of relevance to this association signal. *CHEK2*, or *Checkpoint Kinase 2*, is one of 7 other genes at this locus harbouring variants in strong LD with rs2228260, and has been associated with PCOS in Finnish and Estonian cohorts [15]. Indeed, rs2228260 has been reported as an eQTL for *CHEK2* in adrenal gland tissue [13]. Furthermore, existing research reporting an association between this gene and PCOS found that it remained significant in the Finnish population after including BMI as a covariate [15]. Given the notable proportion of subjects from these cohorts included in this meta-analysis, it is perhaps unsurprising that this signal is present. It should be noted however that the lead SNP identified in this study differs from that reported in the previous Finnish/Estonian research (rs182075939), the SNPs are not in particularly strong LD ( $r^2 < 0.2$ ) [16] and conditioning on rs182075939 did not completely remove the association for the lead SNP identified in this study (although it was no longer genome-wide significant). A recent GWAS found several variants within *CHEK2* to be associated with age at natural menopause (ANM) [35]. There is some evidence of LD between rs5762852 found in that study and our lead PCOS SNP rs2228260 ( $r^2 < 0.2$ ,

$D' = 1$ ), potentially suggesting shared biology between PCOS and ANM. *CHEK2* is involved in a number of reproductive physiological processes concerning oocyte numbers, follicle atresia, later age at menopause and anti-mullerian hormone (AMH) levels, providing plausible biological links to PCOS [35, 36]. It is possible that there is more than one gene in this chromosomal region driving the associations seen in this and previous studies.

Among the genome-wide suggestive lean PCOS loci, *YAPI*, *KRRI*, *IRFI* and *BLK* are of particular interest. The identification of variants meeting criteria for genome-wide suggestive association with lean PCOS within genes that have previous links to PCOS is encouraging and supports the validity of results. These signals have association with PCOS itself as well as traits involved including ovulatory dysfunction and insulin signalling [4, 37]. Polycystic ovary morphology is also associated with some of these signals, specifically with *YAPI* [4]. Interestingly, the two variants in *YAPI* previously identified as associated with PCOS, rs1894116 and rs11225154 [4, 7] both demonstrated genome-wide suggestive associations in the lean PCOS group but did not reach even nominal significance in the overweight or obese groups. This could suggest that the associations previously reported between this locus and PCOS may be driven primarily by lean PCOS patients. *YAPI* (Yes-associated protein 1) has been linked to PCOS pathogenesis through its role in maintaining normal ovarian function, response to gonadotrophins and susceptibility to the effects of androgens [38]. It is involved in a signalling cascade necessary for ovarian function and ovulation, and previous research has supported a *YAPI* mediated mechanism for cell survival and differentiation of granulosa cells during ovulation [39]. This may suggest that this gene contributes to oligoamenorrhoea, and resultant infertility seen in PCOS.

There were 18 SNPs meeting genome wide suggestive association with the combined overweight and obese PCOS strata. Among these, the *TEX41* (testis expressed 41) gene is a non-coding RNA gene, and has been identified as a locus associated with circulating AMH levels in women [40]. The SNP found to be associated with AMH levels, rs13009019, is in strong LD in Europeans ( $r^2 = 0.81$ ) with the SNP identified in this study, rs813684 [16], and is thus likely representative of the same signal. On chromosome 22, rs9613552 is found within the gene *TTC28-AS1* (TTC28-Antisense RNA 1). This long non-coding RNA gene has been shown to be downregulated in type 2 diabetes and decreased expression is potentially related to higher risk of developing type 2 diabetes [41]. *CDH18* (cadherin 18 type 2) is one of the closest genes to rs77388455 on chromosome 5. This gene has been reported as associated with phenotypic characteristics common to metabolic syndrome and therefore PCOS,

including insulin resistance, glucose intolerance, type 2 diabetes mellitus (T2DM) and obesity [42]. The top SNP for this gene had a much lower  $P$ -value and greater effect size in the non-lean cohort relative to the lean group ( $P=9.01 \times 10^{-7}$  vs.  $P=0.64$ , and beta 0.39 vs 0.03, respectively). Given the increased propensity for PCOS affected women to develop T2DM, and the associated metabolic syndrome type clinical features, it is possible that *CDH18* has some link to PCOS, particularly overweight/obese PCOS. Although signals meeting GW significance were also identified in the overweight/obese groups combined, including *DENNDIA*, the strength of association was lower than that seen in the lean cohort. This may imply that PCOS in overweight/obese women is influenced by environment as well as by genetics. Weight gain and high BMI are associated with PCOS-like features such as insulin resistance and oligomenorrhoea [43, 44]. It is possible that in a proportion of overweight/obese women the strength of the association with genetics is diluted by environmental factors. The relationship between obesity and genetics also needs to be considered, whereby obesity may be regarded as an environmental modifier of PCOS, affecting the emergence of an underlying genetic predisposition as body weight increases.

The gene-based analysis in this study found *LINC02905* to be significantly associated with lean PCOS, in addition to *DENNDIA*. *LINC02905* is a small uncharacterised gene located in between *GATA4* and *NEIL2* in a well-established PCOS susceptibility locus. *LINC02905* is also known as *GATA4* downstream membrane gene (*G4DM*) and is considered to be one of the target genes of *GATA4* [45]. *GATA4* (Gata Binding Protein 4) encodes a member of the GATA family of zinc-finger transcription factors, which is thought to play a role in embryogenesis, myocardial differentiation and function and normal testicular development [46]. Alterations in the expression of *GATA4* have been associated with different types of cancer, including ovarian cancer [45, 46]. Interestingly, this locus has shown heterogeneity of effect in previous research when analyses were compared according to PCOS subtypes, based on different diagnostic criteria. This signal showed stronger association with PCOS defined by NIH criteria (i.e., hyperandrogenism and oligomenorrhoea) [4].

The other gene highlighted in gene-based analyses was *ERBB4* (*v-erb-a* erythroblastic leukemia viral oncogene), which was found to be associated with PCOS in the combined overweight/obese group. This gene has previously been associated with PCOS in both women of European ancestry and Han Chinese women [4, 20]. *ERBB4* is a member of the tyrosine protein kinase family and epidermal growth factor receptor subfamily. This gene has been associated with both ovulatory dysfunction and polycystic

ovarian morphology, and is hypothesised to be involved in oligomenorrhoea and infertility aspects of this condition [4, 20]. Furthermore, a murine model, involving *ErbB4* deletion has demonstrated the emergence of various characteristics seen in PCOS patients, specifically disrupted ovulatory cycles with oligomenorrhoea, obesity and impaired oocyte development. In addition, hormonal disturbances included increases in LH and AMH levels, as well as hyperandrogenism [47]. These findings suggest that *ERBB4* may play a key role in PCOS pathophysiology and this is supported by a demonstrated functional role for this gene in ovarian homeostasis and folliculogenesis [47]. The association of this locus specifically with overweight/obese PCOS subjects is a novel finding.

All SNPs previously associated with PCOS were found to be nominally associated with PCOS in at least one of the three BMI stratifications included in this study. A higher proportion of SNPs met criteria for replication in the lean group compared to the individual or combined overweight and obese strata. This is unlikely to be purely due to sample size as the lean and non-lean (combined overweight/obese) groups contained comparable numbers. Some of the SNPs replicated within the lean group have been associated with ovulatory dysfunction and hyperandrogenism, supporting the concept that the lean phenotype is typified by hormonal disturbance and ovarian abnormality, as opposed to the overweight/obese phenotype, which may display a predominance of metabolic disturbance, such as insulin resistance. For example, rs2349415 within *FSHR* has been reported as associated with higher FSH levels as well as ovulatory dysfunction [7, 48]. This SNP demonstrated strongest association with PCOS in the lean group, followed by the overweight group and no association in the obese cohort ( $P=3.99 \times 10^{-3}$ , 0.02 and 0.96, respectively). Similarly, SNPs within *YAPI*, *TOX3* and *IRF1/RAD50* were only significant in the lean group, and have previous association with ovulatory dysfunction, hyperandrogenism and increased testosterone levels respectively [4, 7].

The population used for this study comprised a higher proportion of lean women than is described epidemiologically. PCOS-affected women are mostly overweight or obese, with 16–30% falling into a normal or lean BMI category, though around half of cases in this study were BMI < 25 kg/m<sup>2</sup>. Examination of the different cohorts included in the meta-analysis demonstrated varying proportions of women in the BMI strata. The Western Australian PCOS research group aimed to recruit lean women to the study wherever possible to maximise sample size for this group in genetic research, though the proportion of lean/normal BMI women included was just under one third of the WA cohort. The Estonian and Rotterdam cohorts comprised higher proportions of lean



women, with 64% and 55% lean cases included respectively. It is noted that these proportions are different to epidemiological reports, though may, in part, be reflective of obesity epidemiology differences between the various cohorts included. Estonia and The Netherlands are known to have lower rates of obesity than Australia and the United States, with data from the US indicating over one third of women in general are obese, compared to 20% in the Netherlands [49]. Overall, a higher number of lean women included in the study improves power for analysis. The previous GWAS available in the literature were conducted in women with BMI within the overweight to obese range, based on WHO definitions, hence this study is different to those previously reported and needs to be considered when drawing comparison. The two large studies in Han Chinese women reported mean BMI within the overweight range for Asian populations, ranging 23.28–24.76 between the two studies and discovery and replication cohorts [3, 7]. The studies conducted in women of European/Caucasian ethnicity included women classified as overweight or obese, with all cases reported to be BMI > 25 [4–6].

### Conclusion

The results from this study provide further evidence to support the theory of genetic differences between lean and overweight/obese PCOS-affected women. Whilst the exact mechanisms by which these signals are contributing to the pathophysiology of this condition are yet to be elucidated, the locations and proximity to a number of genes previously linked with features of PCOS, including ovulatory dysfunction and aberrant metabolism, intimates their potential involvement. Many of the variants identified in this study were intronic, suggesting that they are exerting an effect through modification or enhancement of transcriptional regulation of genes in close proximity (i.e., most often within 200 kb or less) [50], thus influencing the differences in expression of phenotype in these subjects. The findings reported in this study are unique and add to the growing body of evidence supporting both a genetic basis for PCOS as well as differences in genetic patterns based on PCOS phenotype.

### Methods

In this study, a meta-analysis of case–control GWAS data stratified according to BMI was performed. Study subjects were allocated into three groups according to BMI, based on WHO definitions. Lean PCOS was defined by BMI  $\leq 25$  kg/m<sup>2</sup>, overweight PCOS was defined by BMI  $> 25$ – $< 30$  kg/m<sup>2</sup> and obese PCOS was defined by BMI  $\geq 30$  kg/m<sup>2</sup>; control subjects were similarly stratified.

The study subjects used for this analysis were sourced from six separate international cohorts, from the United

States of America, Australia, Estonia, Finland and the Netherlands. Each centre recruited PCOS affected women and control subjects of European ancestry or identified them from an existing biobank. For the purposes of study inclusion, PCOS cases were defined according to NIH or Rotterdam criteria, or based on ICD codes and questionnaires (“PCOS coded/self-reported”) depending on the criteria stipulated by each individual centre. Controls were defined as women who did not have a PCOS diagnosis, recruited from population-based samples. Research contributions from the United States of America included the Cedars Sinai ( $n = 359$  cases and  $n = 276$  controls) [51] and BioVU ( $n = 365$  cases and  $n = 6535$  controls) [20] cohorts. The Australian cohort was from WA-PCOS ( $n = 271$  cases and  $n = 2492$  controls) [18, 52]. The contribution from Estonia was from Estonian Biobank ( $n = 3665$  cases and  $n = 113,878$  controls) [15]. The cohort from Finland was FinnGen ( $n = 643$  cases and  $n = 117,794$  controls) [15] and from the Netherlands was the Rotterdam PCOS Cohort, with PCOS cases diagnosed in Erasmus Medical Centre, Rotterdam by thorough standardized screening [4] and controls provided by the Lifelines Cohort Study ( $n = 634$  cases and  $n = 7685$  controls) [53]. All these participating cohorts have been described in detail previously.

### Genotyping, quality control and imputation

Cohort-specific information is summarised in Supplementary Table 5. Only individuals from European ancestry were included in the meta-analysis, with each cohort performing adjustment for principal components to correct for any population stratification. GWAS was performed for each cohort using either the SAIGE [54] or SNPTEST [55] software packages. SAIGE software accounts for imbalance in case control ratios, and uses a random effect model. Summary results were supplied from each cohort for meta-analysis, with quality control of the supplied results files performed using EasyQC [56].

### Meta-analysis

Meta-analysis was performed using the METAL software [57]. METAL effectively handles analyses where studies contain disproportionate numbers of cases and controls, thus allowing flexibility, and performs tests for heterogeneity to ensure participating studies demonstrate consistent effects [57]. Meta-analysis was performed using the METAL software using a fixed effects model weighted by standard error [57]. METAL effectively handles analyses where studies contain disproportionate numbers of cases and controls, thus allowing flexibility, and performs tests for heterogeneity to ensure participating studies demonstrate consistent effects [57].

### Annotation and bioinformatics analysis of meta-analysis results

Meta-analysis results were annotated using FUMA software [14]. This platform performs functional mapping and annotation of GWAS results to facilitate interpretation and provide biological context, thus helping to identify causal variants. Both single-variant based annotation and gene-based testing approaches were employed. The FUMA module, SNP2GENE approach, uses submitted GWAS summary statistics to identify lead SNPs and perform functional annotation of all variants in the surrounding genomic regions. Three mapping processes, specifically positional, eQTL and chromatin mapping, work in concert to create a mapped genes table, which in turn is used for the next major function of the FUMA software suite, GENE2FUNC. This process annotates the biological context of these genes thus providing insight into the potential mechanisms of the involved loci [14]. Conditional analysis of the lean PCOS meta-analysis results was performed using the COJO function of the GCTA package [17], which uses GWAS summary statistics and estimated LD from a sample population to identify independent association signals within a genetic locus [58]. The sample genotypes used for LD estimation in the conditional analysis were from the Western Australia cohort. Replication analysis of previously identified PCOS loci was performed for each BMI stratum, with  $P < 0.05$  considered nominally significant evidence of replication. Beta values/odds ratios were then examined to confirm a consistent allelic effect to that previously reported. Analysis of the linkage disequilibrium (LD) in regions of interest was performed using LDlink (1000 Genomes Project Phase 3 EUR population) [16]. Expression quantitative trait locus (eQTL) associations were assessed using the GTEx dataset [13]. Co-localisation analysis of GWAS results was performed using the coloc package in R [12], which uses a Bayesian framework to calculate posterior probabilities for 5 different scenarios regarding the presence and co-localisation of association signals in two datasets.

#### Abbreviations

BMI	Body mass index
GWAS	Genome-wide association study
PCOS	Polycystic ovary syndrome
LH	Luteinising hormone
FSH	Follicle stimulating hormone
SHBG	Sex hormone binding globulin
ER	Endoplasmic reticulum
UPR	Unfolded Protein Response
AMH	Anti-Müllerian hormone
FOA	Foetal oocyte attrition
OD	Ovulatory dysfunction
GDM	Gestational diabetes mellitus
eQTL	Expression quantitative trait locus

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-09990-w>.

**Additional file 1: Supplementary Figure 1A.** QQ plots for SNP-based analyses for each BMI strata in the meta-analysis of genome-wide association studies for PCOS. A) Lean BMI  $\leq 25$  kg/m<sup>2</sup> ( $\lambda=1.01$ ) B) overweight 25 < BMI < 30 kg/m<sup>2</sup> ( $\lambda=1.01$ ) C) obese BMI  $\geq 30$  kg/m<sup>2</sup> ( $\lambda=1.02$ ) D) combined overweight/obese (non-lean) groups ( $\lambda=1.01$ ) E) all groups combined ( $\lambda=1.04$ ). The lean group demonstrates a greater number of highly significant  $p$ -values than the overweight and obese groups, likely due to the comparatively larger sample size. **Supplementary Figure 1B.** QQ plots for gene-based analyses for each BMI strata in the meta-analysis of genome-wide association studies for PCOS A) Lean BMI  $\leq 25$  kg/m<sup>2</sup> gene-based analysis, B) overweight 25 < BMI < 30 kg/m<sup>2</sup> gene-based analysis, C) obese BMI  $\geq 30$  kg/m<sup>2</sup> gene-based analysis D) combined overweight/obese (non-lean) groups (gene-based analysis), E) all groups combined gene-based analysis.

**Additional file 2: Supplementary Figure 2.** Manhattan plot displaying the results of the lean PCOS single-variant based meta-analysis. Genome-wide significant loci labelled and the threshold for genome wide significance ( $P < 5 \times 10^{-8}$ ) is shown in red.

**Additional file 3: Supplementary Figure 3.** Miami plot depicting the meta-analysis results for the lean (upper panel) and combined overweight/obese PCOS (lower panel) strata. Genome-wide significant loci are labelled and the thresholds for genome-wide significance ( $P < 5 \times 10^{-8}$ ) and genome-wide suggestive significance ( $P < 5 \times 10^{-6}$ ) are shown in red and orange respectively.

**Additional file 4: Supplementary Figure 4.** Manhattan plot displaying the results of the combined overweight/obese PCOS single-variant based meta-analysis. The genome-wide significant locus is labelled and the threshold for genome wide significance ( $P < 5 \times 10^{-8}$ ) is shown in red.

**Additional file 5: Supplementary Figure 5.** Annotation of genome-wide significant signals from meta-analysis of the lean PCOS strata using FUMA software [10]. Plots show the characteristics of each genome-wide significant locus in terms of the physical size, number of potentially relevant SNPs and genes at each locus.

**Additional file 6: Supplementary Figure 6.** Manhattan plot displaying the results from the lean PCOS gene-based meta-analysis. The genome-wide significant genes are labelled and the threshold for genome wide significance ( $P < 1.96 \times 10^{-6}$ ) is shown in red.

**Additional file 7: Supplementary Figure 7.** Manhattan plot displaying the results from the combined overweight/obese PCOS gene-based meta-analysis with a single genome-wide significant gene labelled. The threshold for genome wide significance ( $P < 1.96 \times 10^{-6}$ ) is shown in red.

**Additional file 8: Supplementary Table 1.** Results in each BMI subgroup for loci demonstrating genome-wide suggestive association ( $P < 5 \times 10^{-6}$ ) with lean PCOS in the individual-variant meta-analysis.

**Additional file 9: Supplementary Table 2.** Cohort-specific results for genome-wide significant association signals identified in the lean PCOS meta-analysis.

**Additional file 10: Supplementary Table 3.** Loci meeting genome-wide suggestive significance ( $P < 5 \times 10^{-6}$ ) in the combined overweight/obese strata individual-variant based meta-analysis, with comparative data from the lean stratum.

**Additional file 11: Supplementary Table 4.** Summary of loci associated with PCOS from previous GWAS reported in the literature showing the level of significance in this BMI-stratified PCOS meta-analysis.

**Additional file 12: Supplementary Table 5.** Descriptive information for the six cohorts included in the meta-analyses.

**Additional file 13: Supplementary Table 6.** FinnGen Banner Authors 2023.

## Acknowledgements

BioVU—N/A

This study presents independent research supported by the Sir Charles Gairdner Osborne Park Health Care Group, Department of Health (Western Australia), the National Health and Medical Research Council of Australia, Health National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London and the IoPPN Genomics & Biomarker Core Facility, King's College London. The views expressed are those of the author(s) and not necessarily those of the organisations listed.

Estonian Biobank—Data analysis was carried out in part in the High-Performance Computing Center of University of Tartu

– FinnGen—We want to acknowledge the participants and investigators of FinnGen study. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank ([www.auria.fi/biopankki](http://www.auria.fi/biopankki)), THL Biobank ([www.thl.fi/biobank](http://www.thl.fi/biobank)), Helsinki Biobank ([www.helsinginbiopankki.fi](http://www.helsinginbiopankki.fi)), Biobank Borealis of Northern Finland (<https://www.ppsph.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx>), Finnish Clinical Biobank Tampere ([www.tays.fi/en-US/Research\\_and\\_development/Finnish\\_Clinical\\_Biobank\\_Tampere](http://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere)), Biobank of Eastern Finland ([www.ita-suomenbiopankki.fi/en](http://www.ita-suomenbiopankki.fi/en)), Central Finland Biobank ([www.ksshp.fi/fi-FI/Potilaalle/Biopankki](http://www.ksshp.fi/fi-FI/Potilaalle/Biopankki)), Finnish Red Cross Blood Service Biobank ([www.veripalvelu.fi/verenluovutus/biopankkitoiminta](http://www.veripalvelu.fi/verenluovutus/biopankkitoiminta)), Terveystalo Biobank ([www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/](http://www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/)) and Arctic Biobank (<https://www.oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank>). All Finnish Biobanks are members of BBMRI.fi infrastructure ([www.bbmrif.org](http://www.bbmrif.org)). Finnish Biobank Cooperative -FINBB (<https://finbb.fi/>) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious® services (<https://site.fingenious.fi/en/>) managed by FINBB.

Lifelines Cohort Study—Lifelines is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviours of 167,729 persons living in the North of the Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioural, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multi-morbidity and complex genetics.

## Consortium name

### Estonian Biobank Research Team:

Andres Metspalu<sup>1</sup>, Lili Milani<sup>1</sup>, Tõnu Esko<sup>1</sup>, Mari Nelis<sup>1</sup>, Georgi Hudjashov<sup>1</sup>  
Estonian Biobank Research Team Affiliations.

<sup>1</sup> Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia,

### FinnGen:

See attached Supplementary Table 6 for 2023 FinnGen Banner Authors.

### International PCOS Consortium:

Felix R. Day<sup>1</sup>  
Tugce Karaderi<sup>2,3</sup>  
Michelle R. Jones<sup>4</sup>  
Cindy Meun<sup>5</sup>  
Chunyan He<sup>6,7</sup>  
Alex Drong<sup>2</sup>  
Peter Kraft<sup>8</sup>  
Nan Lin<sup>6,7</sup>  
Hongyan Huang<sup>8</sup>  
Linda Broer<sup>9</sup>  
Reedik Magi<sup>10</sup>  
Richa Saxena<sup>11</sup>  
Jaakko S. Tyrmi<sup>12,13</sup>  
Triin Laisk<sup>10,14</sup>  
Andres Metspalu<sup>10</sup>  
Lili Milani<sup>10</sup>  
Tõnu Esko<sup>10</sup>  
Mari Nelis<sup>10</sup>  
Georgi Hudjashov<sup>10</sup>  
Margrit Urbanek<sup>15,16</sup>  
M. Geoffrey Hayes<sup>15,16,17</sup>

Gudmar Thorleifsson<sup>18</sup>

Juan Fernandez-Tajes<sup>2</sup>

Anubha Mahajan<sup>2,19</sup>

Kharis A. Burns<sup>20,21</sup>

Benjamin H. Mullin<sup>22</sup>

Bronwyn G. A. Stuckey<sup>21,22,23</sup>

Timothy D. Spector<sup>24</sup>

Scott G. Wilson<sup>22,24</sup>

Frank Dudbridge<sup>25</sup>

Jinrui Cui<sup>26</sup>

Mark O. Goodarzi<sup>26</sup>

Ky'Era Actkins<sup>27,28,29</sup>

Lea K. Davis<sup>28,29</sup>

Barbara Obermayer-Pietsch<sup>30</sup>

André G. Uitterlinden<sup>9</sup>

Verneri Anttila<sup>28,31,32</sup>

Benjamin M. Neale<sup>31,32</sup>

Marjo-Riitta Jarvelin<sup>33,34,35,36</sup>

Hannele Laivuori<sup>12, 37,38,39</sup>

Mark Daly<sup>11,32,39</sup>

Bart Fauser<sup>36</sup>

Irina Kowalska<sup>40</sup>

Loes M.E. Moolhuijsen<sup>9</sup>

Yvonne Louwers<sup>5</sup>

Jenny A. Visser<sup>9</sup>

Marianne Andersen<sup>41</sup>

Ken Ong<sup>1</sup>

Elisabet Stener-Victorin<sup>42</sup>

David Ehrmann<sup>43</sup>

Richard S. Legro<sup>44</sup>

Andres Salumets<sup>12,45,46,47</sup>

Mark I. McCarthy<sup>2,19,48</sup>

Laure Morin-Papunen<sup>49</sup>

Unnur Thorsteinsdottir<sup>18,50</sup>

Kari Stefansson<sup>18,50</sup>

The 23andMe Research Team<sup>®</sup>

Unnur Styrkarsdottir<sup>18</sup>

John R. B. Perry<sup>1</sup>

Andrea Dunaif<sup>1,5,51</sup>

Joop Laven<sup>5</sup>

Steve Franks<sup>52</sup>

Cecilia M. Lindgren<sup>2,11,53</sup>

Corrine K. Welt<sup>54,55</sup>

International PCOS Consortium Affiliations.

1 MRC Epidemiology Unit, Cambridge Biomedical Campus, University of Cambridge School of Clinical Medicine, Cambridge, United Kingdom

2 The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

3 Department of Biological Sciences, Faculty of Arts and Sciences, Eastern Mediterranean University, Famagusta, Cyprus

4 Center for Bioinformatics & Functional Genomics, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California, United States of America

5 Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynaecology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

6 Department of Internal Medicine, University of Kentucky College of Medicine, Lexington, Kentucky, United States of America

7 University of Kentucky Markey Cancer Center, Lexington, Kentucky, United States of America

8 Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America

9 Department of Internal Medicine, Section Endocrinology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

10 Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia

11 Broad Institute of Harvard and MIT and Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, United States of America.

12 Center for Child, Adolescent and Maternal Health Research, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.

13 Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland.

14 Department of Obstetrics and Gynaecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia

15 Division of Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, United States of America

16 Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, United States of America

17 Department of Anthropology, Northwestern University, Evanston, Illinois, United States of America

18 deCODE genetics/Amgen, Reykjavik, Iceland

19 Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom

20 Department of Endocrinology and Diabetes, Royal Perth Hospital, Perth, WA 6009, Australia.

21 Medical School, University of Western Australia, Nedlands, WA, Australia.

22 Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia

23 Keogh Institute for Medical Research, Nedlands, Western Australia, Australia

24 Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom

25 Department of Health Sciences, University of Leicester, Leicester, UK.

26 Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, California, United States of America

27 Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC.

28 Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America

29 Vanderbilt Genomics Institute, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America

30 Division of Endocrinology and Diabetology, Department of Internal Medicine Medical University of Graz, Graz, Austria

31 Stanley Center for Psychiatric Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America

32 Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, United States of America

33 Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, United Kingdom

34 Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland

35 Biocenter Oulu, University of Oulu, Oulu, Finland, 31 Unit of Primary Care, Oulu University Hospital, Oulu, Finland

36 Department of Reproductive Medicine and Gynaecology, University Medical Center, Utrecht, The Netherlands

37 Department of Obstetrics and Gynecology, Tampere University Hospital, Tampere, Finland.

38 Medical and Clinical Genetics, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.

39 Institute for Molecular Medicine Finland, FIMM, hiLIFE, University of Helsinki, Helsinki, Finland.

40 Department of Internal Medicine and Metabolic Diseases, Medical University of Białystok, Białystok, Poland

41 Odense University Hospital, University of Southern Denmark, Odense, Denmark

42 Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden

43 Department of Medicine, Section of Adult and Paediatric Endocrinology, Diabetes, and Metabolism, The University of Chicago, Chicago, Illinois, United States of America

44 Department of Obstetrics and Gynecology and Public Health Sciences, Penn State University College of Medicine, Hershey, Pennsylvania, United States of America

45 Competence Centre on Health Technologies, Tartu, Estonia

46 Institute of Bio- and Translational Medicine, University of Tartu, Tartu, Estonia

47 Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

48 Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom

49 Department of Obstetrics and Gynecology, University of Oulu and Oulu University Hospital, Medical Research Center, PEDEGO Research Unit, Oulu, Finland

50 Faculty of Medicine, University of Iceland, Reykjavik, Iceland

51 Division of Endocrinology, Diabetes and Bone Disease, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America

52 Institute of Reproductive & Developmental Biology, Department of Surgery & Cancer, Imperial College London, London, United Kingdom

53 Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

54 Division of Endocrinology, Metabolism and Diabetes, University of Utah, Salt Lake City, Utah, United States of America

55 Reproductive Endocrine Unit, Massachusetts General Hospital, Boston, Massachusetts, United States of America.

**23andMe Research Team** (23andMe, Inc., Mountain View, California, United States of America): Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Karen E. Huber, Aaron Kleinman, Nadia Kenref. Litterman, Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, Catherine H. Wilson.

#### Authors' contributions

Conceived and designed the study: KAB, BHM, FD, SGW, BGAS. Performed the genotyping and genotype imputation: BHM, LMEM, JST, Finngen, TL, KA, JC, Estonian Biobank Research Team. Performed the phenotyping: KAB, SGW, BGAS, Estonian Biobank Research Team, HL, JST, Finngen. Analysed the data/performed the statistical analysis: BHM, FD, LMEM, TL, RM. Drafted the manuscript: KAB, BHM, FD, SGW. Reviewed and edited the manuscript: HL, JAV, JST, YL, LMEM, LKD, TL, KA, JL, RM, MG, RA, FRD, BGAS. All authors read and approved the final manuscript.

#### Funding

Cedars Sinai – Mark Goodarzi was supported by the Eris M.Field Chair in Diabetes Research and NIDDK P30-DK063481. Ricardo Azziz was supported by NICHD grants R01-HD29364 and K24-HD01346.

BioVU—N/A

WA-PCOS—This study was supported by a grant from the Sir Charles Gairdner Osborne Park Health Care Group Research Advisory Committee, (Grant number: RAC2019-20/029) and, in part, by funding from the National Health and Medical Research Council of Australia (APP2003629 to B.H.M) and a Department of Health (Western Australia) Merit Award (No. 1186046 to B.H.M).

The Estonian Biobank work was supported by the Estonian Research council grant PRG1911 and by European Union through the European Regional Development Fund Project No. 2014–2020.4.01.15–0012 GENTRANSMED.

The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sarl), Genentech Inc., Merck Sharp & Dohme Corp., Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis AG, and Boehringer Ingelheim International GmbH.

The Lifelines Cohort Study "The Lifelines initiative has been made possible by subsidy from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the University Medical Center Groningen (UMCG), Groningen University and the Provinces in the North of the Netherlands (Drenthe, Friesland, Groningen)."

#### Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due to embargo instituted by the International PCOS Consortium but are available from the corresponding author on reasonable request.

## Declarations

### Ethics approval and consent to participate

All subjects in the study provided written informed consent and the experimental protocols were approved for individual cohorts as follows: 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).

WA-PCOS by the SCGOPHCG Human Research Ethics Committee (RGS0000001467) and controls by HRA North West – Liverpool East Research Ethics Committee (19/NW/0187; TwinsUK).

-Estonian Biobank—All biobank participants have signed a broad informed consent form and analyses were carried out under ethical approval 1.1–12/624 from the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs) and data release N05 from the EstBB.

-FinnGen—Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act came into effect (in September 2013) and start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea (Finnish Medicines Agency), the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr HUS/990/2017.

The FinnGen study is 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 and THL/1524/5.05.00/2020), 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 134/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020), Findata permit numbers THL/2364/14.02/2020, THL/4055/14.06.00/2020 „THL/3433/14.06.00/2020, THL/4432/14.06/2020, THL/5189/14.06/2020, THL/5894/14.06.00/2020, THL/6619/14.06.00/2020, THL/209/14.06.00/2021, THL/688/14.06.00/2021, THL/1284/14.06.00/2021, THL/1965/14.06.00/2021, THL/5546/14.02.00/2020 and Statistics Finland (permit numbers: TK-53–1041-17 and TK/143/07.03.00/2020 (earlier TK-53–90-20)).

The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 7 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 and amendment #1 (August 17 2020), Biobank Borealis of Northern Finland\_2017\_1013, Biobank of Eastern Finland 1186/2018 and amendment 22 § /2020, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), Central Finland Biobank 1–2017, and Terveystalo Biobank STB 2018001.

- Rotterdam—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.

All methods were carried out in accordance with relevant guidelines and regulations.

### Consent for publication

Not applicable.

### Competing interests

The FinnGen project is funded by the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sarl), Genentech Inc., Merck Sharp & Dohme Corp., Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis AG, and Boehringer Ingelheim International GmbH. All other authors declare that they have no competing interests.

## Author details

<sup>1</sup>Department of Endocrinology and Diabetes, Royal Perth Hospital, Perth, WA 6009, Australia. <sup>2</sup>Medical School, University of Western Australia, Nedlands, WA, Australia. <sup>3</sup>Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands, WA, Australia. <sup>4</sup>School of Biomedical Sciences, University of Western Australia, Perth, WA, Australia. <sup>5</sup>Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. <sup>6</sup>Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia. <sup>7</sup>Center for Child, Adolescent and Maternal Health Research, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. <sup>8</sup>Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland. <sup>9</sup>Division of Endocrinology, Diabetes, and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, USA. <sup>10</sup>Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA. <sup>11</sup>Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>12</sup>Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>13</sup>Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynaecology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands. <sup>14</sup>Population Health Sciences, University of Leicester, Leicester, UK. <sup>15</sup>Obstetrics & Gynecology, Medicine, and Healthcare Organization & Policy, Schools of Medicine and Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. <sup>16</sup>Department of Obstetrics and Gynecology, Tampere University Hospital, Tampere, Finland. <sup>17</sup>Institute for Molecular Medicine Finland, FIMM, hiLIFE, University of Helsinki, Helsinki, Finland. <sup>18</sup>Medical and Clinical Genetics, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. <sup>19</sup>Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. <sup>20</sup>MRC Epidemiology Unit, Cambridge Biomedical Campus, University of Cambridge School of Clinical Medicine, Cambridge, UK. <sup>21</sup>Keogh Institute for Medical Research, Nedlands, WA, Australia.

Received: 2 May 2023 Accepted: 8 January 2024

Published online: 26 February 2024

## References

- Azziz R. PCOS in 2015: New insights into the genetics of polycystic ovary syndrome. *Nat Rev Endocrinol*. 2016;12(2):74–5. <https://doi.org/10.1038/nrendo.2015.1230>. (Epub 2016 Jan 1034).
- Zawadzki JK, Dunaif A: Diagnostic criteria for polycystic ovary syndrome; towards a rational approach. In *Polycystic Ovary Syndrome*, Dunaif A, Givens JR, Haseltine F, editors. : Boston: Blackwell Scientific Publications, 1992; 377–84.
- Chen ZJ, Zhao H, He L, Shi Y, Qin Y, Shi Y, Li Z, You L, Zhao J, Liu J, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat Genet*. 2011;43(1):55–9.
- Day F, Karaderi T, Jones MR, Meun C, He C, Drong A, Kraft P, Lin N, Huang H, Broer L, et al. Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *Plos Genet*. 2018;14(12):e1007813.
- Day F, Hinds DA, Tung JY, Stolk L, Stykarskiddottir U, Saxena R, Bjornnes A, Broer L, Dunger DB, Halldorsson BV, et al. Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nat Commun*. 2015;6:8464.
- Hayes G, Urbaneck M, Ehrmann DA, Armstrong LL, Young Lee J, Sisk R, Karaderi T, Barber TM, McCarthy MI, Franks S, et al. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. *Nat Commun*. 2015;6:7502.
- Shi Y, Zhao H, Shi Y, Cao Y, Yang D, Li Z, Zhang B, Liang X, Li T, Chen J, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet*. 2012;44(9):1020–5.
- Cussons AJ, Watts GF, Burke V, Shaw JE, Zimmet PZ, Stuckey BG. Cardio-metabolic risk in polycystic ovary syndrome: a comparison of different approaches to defining the metabolic syndrome. *Hum Reprod (Oxford, England)*. 2008;23(10):2352–8.
- Haider S, Mannan N, Khan A, Qureshi MA. Influence of anthropometric measurements on abnormal gonadotropin secretion in women with polycystic ovary syndrome. *J Coll Phys Surg-Pakistan*. 2014;24(7):463–6.

10. Morciano A, Romani F, Sagnella F, Scarinci E, Palla C, Moro F, Tropea A, Policola C, Della Casa S, Guido M, et al. Assessment of insulin resistance in lean women with polycystic ovary syndrome. *Fertil Steril*. 2014;102(1):250–256 e253.
11. Stuckey BG, Opie N, Cussons AJ, Watts GF, Burke V. Clustering of metabolic and cardiovascular risk factors in the polycystic ovary syndrome: a principal component analysis. *Metabolism*. 2014;63(8):1071–7.
12. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *Plos Genet*. 2014;10(5):e1004383.
13. Consortium G: The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580–5.
14. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826.
15. Tyrmi JS, Arffman RK, Pujol-Gualdo N, Kurra V, Morin-Papunen L, Sliz E, Piltonen TT, Laisk T, Kettunen J, Laivuori H. Leveraging Northern European population history: novel low-frequency variants for polycystic ovary syndrome. *Hum Reprod (Oxford, England)*. 2022;37(2):352–65.
16. Machiela MJ, Chanock SJ. LDlink a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31:3555.
17. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76–82.
18. Ruth KS, Campbell PJ, Chew S, Lim EM, Hadlow N, Stuckey BG, Brown SJ, Feenstra B, Joseph J, Surdulescu GL, et al. Genome-wide association study with 1000 genomes imputation identifies signals for nine sex hormone-related phenotypes. *Eur J Hum Genet*. 2016;24(2):284–90.
19. Peng Y, Zhang W, Yang P, Tian Y, Su S, Zhang C, Chen ZJ, Zhao H. ERBB4 confers risk for polycystic ovary syndrome in Han Chinese. *Sci Rep*. 2017;7:42000.
20. Zhang Y, Ho K, Keaton JM, Hartzel DN, Day F, Justice AE, Josyula NS, Pendergrass SA, Actkins K, Davis LK, et al. A genome-wide association study of polycystic ovary syndrome identified from electronic health records. *Am J Obstet Gynecol*. 2020;223(4):559.e551–559.e521.
21. Welt CK, Stykarsdottir U, Ehrmann DA, Thorleifsson G, Arason G, Gudmundsson JA, Ober C, Rosenfield RL, Saxena R, Thorsteinsdottir U, et al. Variants in DENND1A are associated with polycystic ovary syndrome in women of European ancestry. *J Clin Endocrinol Metab*. 2012;97(7):E1342–7.
22. Wickenheisser JK, Quinn PG, Nelson VL, Legro RS, Strauss JF 3rd, McAllister JM. Differential activity of the cytochrome P450 17 $\alpha$ -hydroxylase and steroidogenic acute regulatory protein gene promoters in normal and polycystic ovary syndrome theca cells. *J Clin Endocrinol Metab*. 2000;85(6):2304–11.
23. McAllister JM, Modi B, Miller BA, Biegler J, Bruggeman R, Legro RS, Strauss JF 3rd. Overexpression of a DENND1A isoform produces a polycystic ovary syndrome theca phenotype. *Proc Natl Acad Sci USA*. 2014;111(15):E1519–1527.
24. Teves ME, Modi BP, Kulkarni R, Han AX, Marks JS, Subler MA, Windle J, Newall JM, McAllister JM, Strauss JF 3rd. Human DENND1A2 drives Cyp17a1 expression and androgen production in mouse ovaries and adrenals. *Int J Mol Sci*. 2020;21(7):2545.
25. Dapas M, Sisk R, Legro RS, Urbanek M, Dunaif A, Hayes MG. Family-based quantitative trait meta-analysis implicates rare noncoding variants in DENND1A in polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2019;104(9):3835–50.
26. Dapas M, Lin FTJ, Nadkarni GN, Sisk R, Legro RS, Urbanek M, Hayes MG, Dunaif A. Distinct subtypes of polycystic ovary syndrome with novel genetic associations: an unsupervised, phenotypic clustering analysis. *Plos Med*. 2020;17(6): e1003132.
27. Edwards NC, Hing ZA, Perry A, Blaisdell A, Kopelman DB, Fathke R, Plum W, Newell J, Allen CE, Geetha S, et al. Characterization of coding synonymous and non-synonymous variants in ADAMTS13 using ex vivo and in silico approaches. *Plos one*. 2012;7(6):e38864.
28. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM. A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science (New York, NY)*. 2007;315(5811):525–8.
29. Chamary JV, Hurst LD. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol*. 2005;6(9):R75.
30. Piperi C, Adamopoulos C, Papavassiliou AG. XBP1: a pivotal transcriptional regulator of glucose and lipid metabolism. *Trends Endocrinol Metab*. 2016;27(3):119–22.
31. Zhou Y, Lee J, Reno CM, Sun C, Park SW, Chung J, Lee J, Fisher SJ, White MF, Biddinger SB, et al. Regulation of glucose homeostasis through a XBP-1-FoxO1 interaction. *Nat Med*. 2011;17(3):356–65.
32. Akiyama M, Liew CW, Lu S, Hu J, Martinez R, Hambro B, Kennedy RT, Kulkarni RN. X-box binding protein 1 is essential for insulin regulation of pancreatic  $\alpha$ -cell function. *Diabetes*. 2013;62(7):2439–49.
33. Sun HL, Tian MM, Jiang JX, Liu CJ, Zhai QL, Wang CY, Li QC, Wang YL. Does endoplasmic reticulum stress stimulate the apoptosis of granulosa cells in polycystic ovary syndrome?. *J Physiol Pharmacol*. 2021;72(5):785–92.
34. Bahçeci E, Kaya C, Karakaş S, Yıldız S, Hoşgören M, Ekin M. Serum X-box-binding protein 1 levels in PCOS patients. *Gynecol Endocrinol*. 2021;37(10):920–4.
35. Ruth KS, Day FR, Hussain J, Martínez-Marchal A, Aiken CE, Azad A, Thompson DJ, Knoblochova L, Abe H, Tarry-Adkins JL, et al. Genetic insights into biological mechanisms governing human ovarian ageing. *Nature*. 2021;596(7872):393–7.
36. Ward LD, Parker MM, Deaton AM, Tu H-C, Flynn-Carroll AO, Hinkle G, Nioi P. Rare coding variants in DNA damage repair genes associated with timing of natural menopause. *Hum Genet Genomics Adv*. 2022;3(2):100079.
37. Borowiec M, Liew CW, Thompson R, Boonyasrisawat W, Hu J, Mlynarski WM, El Khattabi I, Kim SH, Marselli L, Rich SS, et al. Mutations at the BLK locus linked to maturity onset diabetes of the young and beta-cell dysfunction. *Proc Natl Acad Sci USA*. 2009;106(34):14460–5.
38. Ji SY, Liu XM, Li BT, Zhang YL, Liu HB, Zhang YC, Chen ZJ, Liu J, Fan HY. The polycystic ovary syndrome-associated gene Yap1 is regulated by gonadotropins and sex steroid hormones in hyperandrogenism-induced oligo-ovulation in mouse. *Mol Hum Reprod*. 2017;23(10):698–707.
39. Sun T, Diaz FJ. Ovulatory signals alter granulosa cell behavior through YAP1 signaling. *Reprod Biol Endocrinol*. 2019;17(1):113.
40. Verdiesen RM, van der Schouw YT, van Gils CH, Verschuren WM, Broekmans FJ, Borges MC, Soares AL, Lawlor DA, Eliassen AH, Kraft P, et al. Genome-wide association study meta-analysis identifies three novel loci for circulating anti-Müllerian hormone levels in women. *MedRxiv : Preprint Server for Health Sci*. 2020;37:1069.
41. Mohamadi M, Ghaedi H, Kazerouni F, Erfanian Omidvar M, Kalbasi S, Shanaki M, Miralamy G, Rahimpour A. Deregulation of long noncoding RNA SNHG17 and TTC28-AS1 is associated with type 2 diabetes mellitus. *Scand J Clin Lab Invest*. 2019;79(7):519–23.
42. Zhang Y, Kent JW Jr, Olivier M, Ali O, Cerjak D, Broeckel U, Abdou RM, Dyer TD, Comuzzie A, Curran JE, et al. A comprehensive analysis of adiponectin QTLs using SNP association, SNP cis-effects on peripheral blood gene expression and gene expression correlation identified novel metabolic syndrome (MetS) genes with potential role in carcinogenesis and systemic inflammation. *BMC Med Genomics*. 2013;6:14.
43. Wei S, Schmidt MD, Dwyer T, Norman RJ, Venn AJ. Obesity and menstrual irregularity: associations with SHBG, testosterone, and insulin. *Obesity (Silver Spring)*. 2009;17(5):1070–6.
44. Castillo-Martínez L, López-Alvarenga JC, Villa AR, González-Barranco J. Menstrual cycle length disorders in 18- to 40-year-old obese women. *Nutrition*. 2003;19(4):317–20.
45. Xia L, Wang Y, Meng Q, Su X, Shen J, Wang J, He H, Wen B, Zhang C, Xu M. Integrated bioinformatic analysis of a competing Endogenous RNA network reveals a prognostic signature in endometrial cancer. *Front Oncol*. 2019;9:448.
46. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996–1006.
47. Veikkolainen V, Ali N, Doroszkó M, Kiviniemi A, Miinalainen I, Ohlsson C, Poutanen M, Rahman N, Elenius K, Vainio SJ, et al. Erbb4 regulates the oocyte microenvironment during folliculogenesis. *Hum Mol Genet*. 2020;29(17):2813–30.
48. Welt CK. Genetics of polycystic ovary syndrome: what is New? *Endocrinol Metab Clin North Am*. 2021;50(1):71–82.

49. World Health Organisation. (2017, September 22). Prevalence of obesity among adults, BMI  $\geq$  30 (age-standardized estimate) (%). <https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-obesity-among-adults-bmi-30-age-standardized-estimate>-(-).
50. Keildson S, Fadista J, Ladenvall C, Hedman ÅK, Elgzyri T, Small KS, Grundberg E, Nica AC, Glass D, Richards JB, et al. Expression of phosphofructokinase in skeletal muscle is influenced by genetic variation and associated with insulin sensitivity. *Diabetes*. 2014;63(3):1154–65.
51. Goodarzi MO, Jones MR, Li X, Chua AK, Garcia OA, Chen YD, Krauss RM, Rotter JJ, Ankener W, Legro RS, et al. Replication of association of DENND1A and THADA variants with polycystic ovary syndrome in European cohorts. *J Med Genet*. 2012;49(2):90–5.
52. Jones MR, Italiano L, Wilson SG, Mullin BH, Mead R, Dudbridge F, Watts GF, Stuckey BG. Polymorphism in HSD17B6 is associated with key features of polycystic ovary syndrome. *Fertil Steril*. 2006;86(5):1438–46.
53. Sijtsma A, Rienks J, van der Harst P, Navis G, Rosmalen JGM, Dotinga A. Cohort profile update: lifelines, a three-generation cohort study and biobank. *Int J Epidemiol*. 2022;51(5):e295–302.
54. Zhou W, Zhao Z, Nielsen JB, Fritsche LG, LeFaive J, Gagliano Taliun SA, Bi W, Gabrielsen ME, Daly MJ, Neale BM, et al. Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. *Nat Genet*. 2020;52(6):634–9.
55. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;39(7):906–13.
56. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R, Ferreira T, Fall T, Graff M, Justice AE, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc*. 2014;9(5):1192–212.
57. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)*. 2010;26(17):2190–1.
58. Yang J, Ferreira T, Morris AP, Medland SE, Madden PA, Heath AC, Martin NG, Montgomery GW, Weedon MN, Loos RJ, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. 2012;44(4):369–75.

## Publisher's Note

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