



Recent Advances in the Genetics of Polycystic Ovary Syndrome

4

Michelle R. Jones and Mark O. Goodarzi

Key Points

- Genome-wide association studies in large sample sizes have identified, with high confidence, about 20 susceptibility loci for PCOS.
- Robust susceptibility variants for PCOS have been used in Mendelian randomization studies to identify causes and consequences of PCOS.
- Identification of PCOS susceptibility genes will expand our understanding of pathways and processes implicated in the syndrome's etiology, allowing development of new diagnostic and treatment modalities.

The Heritable Basis of PCOS

In recent years the complex genetic architecture of polycystic ovary syndrome (PCOS) has begun to come into focus. Early family aggregation studies focused on the prevalence of PCOS-related traits in the siblings of PCOS cases and provided the first evidence for a genetic basis to the disorder [1–3]. These studies suggested an autosomal-dominant mode of inheritance based on the incidence of PCOS-related traits in the first-degree relatives of probands of 51–66% [4, 5]. Larger studies provided further evidence for an autosomal-dominant model of inheritance, with as many as 50% of mothers or sisters, 25% of aunts, and 20% of grandmothers of 250 PCOS probands having either hirsutism alone or hirsutism with

M. R. Jones

Center for Bioinformatics & Functional Genomics, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA
e-mail: michelle.jones@csmc.edu

M. O. Goodarzi (✉)

Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA
e-mail: mark.goodarzi@cshs.org

oligomenorrhea [6]. Following the initial reports, however, systematic genetic investigations failed to support an autosomal-dominant mode of inheritance; rather, PCOS appears to be inherited as a common complex disorder, with multiple susceptibility loci. Twin studies that used a large cohort of more than 3000 Danish twins identified a small number of self-reported PCOS cases ($n = 92$), with an estimate of the monozygotic twin correlation for PCOS of 0.72 and a dizygotic correlation of 0.39 [7]. The identification of such a large proportion of variance in risk for PCOS in monozygotic twins provided strong evidence that there is a significant genetic component to the disease.

Candidate Gene Approaches Revealed an Incomplete Understanding of PCOS Biology

More than 100 candidate genes were studied as potential causal risk genes for PCOS; however, only the region surrounding the gene encoding the insulin receptor, *INSR*, was replicated in subsequent large, well-powered genome-wide association studies (GWAS) [8]. The initial studies of the region combined linkage and association analyses to identify the microsatellite marker D19S884, located in intron 55 of the fibrillin-3 gene (*FBN3*), which is 1.3 cM distal to *INSR*, the candidate gene targeted with this variant [9]. It remains unclear whether the causal gene at this locus is *FBN3*, or in fact *INSR*. *FBN3* was known to be expressed in the pituitary, but its role there is unknown. Contemporary epigenomic datasets from the ENCODE project [10] provide strong evidence to suggest this microsatellite is within an active gene regulatory element, but its target remains unknown. Histone modification data indicates likely promoter and/or enhancer activity across the region spanning the microsatellite, with clear cell type-specific modification of histones H3k4me3, H3K27ac, and H3K4me1 in conjunction with open chromatin identified using DNase hypersensitivity site analysis. There is currently no transcriptional isoform of the *FBN3* gene with a promoter position overlapping this microsatellite and active regulatory region, but it is plausible that an isoform with corresponding promoter and transcriptional start site may exist in a cell type not yet comprehensively assayed as part of the ENCODE project. The close proximity of this marker to the *INSR* gene made it a popular target in candidate gene studies. Seven individual studies identified an association between single nucleotide polymorphisms (SNPs) across the *INSR* locus and PCOS risk [11–19]. Many of these studies included a small number of polymorphisms, and modest sample sizes, as did three additional studies that were not able to replicate a significant association between PCOS risk and variants at the *INSR* locus [11, 14, 20].

Additional candidate gene studies focused on genes with known roles in obesity [21–24], type 2 diabetes [25–30], hormone metabolism, and synthesis and ovarian biology [31–35] did not yield any robust loci for PCOS. These studies were largely hampered by small sample sizes and small numbers of variants that provided incomplete tagging across the locus, focusing on coding regions which we now know are unlikely to harbor causal variants for complex traits [36].

GWAS Studies in PCOS

High-throughput genotyping platforms have enabled GWAS and facilitated rapid advancement in the understanding of the complex genetic architecture of many common traits. The first GWAS in PCOS reported in 2011 identified three risk loci: at 2p16.3 (*LHCGR*), 2p21 (*THADA*), and 9q33.3 (*DENND1A*) in Chinese PCOS cases and healthy controls [37]. This three-stage study used a modestly sized discovery cohort of 744 PCOS cases and 895 controls in the GWAS, with replication of suggestive risk loci in a two-stage approach in two cohorts: cohort I, 2840 PCOS cases and 5012 controls; cohort II, 498 PCOS cases and 780 controls [37]. A second study, also performed in Chinese PCOS cases and controls, identified an additional eight risk loci: 2p16.3, 9q22.32, 11q22.1, 12q13.2, 12q14.3, 16q12.1, 19p13.3, and 20q13.2 [8]. This study identified a second, independent risk signal at the 2p16.3 locus, implicating both *LHCGR* and *FSHR* as potential causal genes in the region. *LHCGR* and *FSHR* encode the luteinizing hormone/choriogonadotropin receptor and the follicle-stimulating hormone receptor, which play important roles in hormone signaling in the gonads, making them very plausible susceptibility genes for PCOS. The 2p16.3 region had been the focus of candidate gene studies that profiled only coding variants, without success [20, 38, 39], highlighting the importance of haplotype tagging approaches that include extensive coverage of non-coding variants at gene regions to enable risk locus discovery. The *INSR* locus at 19p13.3 was discussed above. Additional signals identified in the two Chinese GWAS (*THADA* and *HMGA2* associated with type 2 diabetes [40], *RAB5B/SUOX* associated with type 1 diabetes [41]) are near genes from insulin and glucose metabolism pathways, supporting the importance of insulin resistance and metabolic disturbance in PCOS [42]. Two subsequent GWAS performed in Korean cases and controls did not identify any genome-wide significant loci, likely due to small sample size [43, 44].

The first two GWAS for PCOS performed in European-origin populations were published in 2015 [45, 46]. These analyses provided replication of loci reported by Chen and Shi [8, 37] and identified novel loci not previously identified as risk loci for PCOS (Table 4.1). In an initial study that used discovery and replication cohorts of European descent from North America that included a total of 3000 PCOS cases and more than 5000 controls, two novel risk loci were identified: 8p23.1 (*GATA4/NEIL2*) and 11p14.1 (*FSHB*) [45]. The potential causal gene at 8p23.1 is not immediately apparent. Due to linkage disequilibrium (LD) across the region, the association interval spans almost 30 kb. The lead SNP resides between *GATA4* and *NEIL2*, and SNPs in LD with this variant intersect known regulatory regions that connect to the promoters of *C8orf49*, *NEIL2*, and *FDFT1*. *NEIL2* is a transcription factor that is ubiquitously expressed [47, 48] and targets the promoter of more than 240 genes [47, 48], many of which are themselves transcription factors and are important in pathways that include the regulation of development that are dysfunctional in cancer (e.g., *HOX* family of genes) [49] and in hormone signaling (e.g., *FST*, which inhibits FSH release). Both *C8orf49* and *GATA4* are highly expressed in the ovary [47] and present possible causal genes at this locus. The association signal identified by Hayes et al. [45] at 11p14.1 intersects with the coding region for

Table 4.1 Loci associated with PCOS in genome-wide association studies

Locus	Nearest gene	First GWAS report	GWAS replication
2p21	<i>THADA</i>	Chen 2011	Shi 2012, Day 2015, Meta 2018
2p16.3 (A)	<i>LHCGR</i>	Chen 2011	Shi 2012
2p16.3 (B)	<i>FSHR</i>	Shi 2012	
2q34	<i>ERBB4</i>	Day 2015	Meta 2018, Zhang 2020
5q31.1	<i>IRF1/RAD50</i>	Day 2015	Meta 2018
6q25.3	<i>FNDC1/SOD2</i>	Zhang 2020	
8p23.1	<i>GATA4/NEIL2</i>	Hayes 2015	Meta 2018
9p24.1	<i>PLGRKT</i>	Meta 2018	
9q33.3	<i>DENND1A</i>	Chen 2011	Shi 2012, Meta 2018
9p24.1	<i>C9orf3</i>	Hayes 2015	Meta 2018
11p14.1	<i>FSHB</i>	Hayes 2015	Day 2015, Meta 2018
11q22.1	<i>YAP1</i>	Shi 2012	Day 2015, Meta 2018
11q23.2	<i>ZBTB16</i>	Meta 2018	
12q13.2	<i>RAB5B/SUOX</i>	Shi 2012	Meta 2018
12q14.3	<i>HMGA2</i>	Shi 2012	
16q21.1	<i>TOX3</i>	Shi 2012	Meta 2018
19p13.3	<i>INSR</i>	Shi 2012	
20q13.2	<i>SUMO1P1</i>	Shi 2012	
12q21.2	<i>KRR1</i>	Day 2015	Meta 2018
20q11.21	<i>MAPRE1</i>	Meta 2018	

Chen 2011 refers to [37], Shi 2012 refers to [8], Hayes 2015 refers to [45], Day 2015 refers to [46], Meta 2018 refers to [52], Zhang 2020 refers to [55]

FSHB, the gene encoding follicle-stimulating hormone beta subunit, which is a strong candidate as the causal gene at this locus. Genome-wide significant association signals were reported across a 300 kb interval at this locus, and the lead SNP is located >20 kb upstream of the *FSHB* gene within a highly conserved 450 bp region upstream of the coding region for *FSHB*. In vitro studies have since shown this region binds the transcription factor steroidogenic factor 1 (SF1) and enhances the transcription of *FSHB* in an allele-specific manner, supporting the hypothesis that the risk allele at rs11031006 upregulates *FSHB* expression [50]. In this GWAS of European cohorts, more than half of the loci discovered in GWAS of Chinese cohorts exhibited nominal ($P < 0.05$) association with PCOS.

A second GWAS performed in PCOS cases and controls of European descent was published in 2015, by Day et al. [46]. In this study the discovery analysis was performed in a cohort of more than 5000 self-reported PCOS cases and 82,000 healthy controls from the 23andme research resource, with replication performed in 2000 clinically identified cases and nearly 100,000 controls. This analysis successfully replicated genome-wide significant signals at 2p21 (*THADA*) and 11q22.1 (*YAP*), initially reported as PCOS risk loci in Chinese populations [8, 37] and 11p14.1 (*FSHB*), previously reported as a risk locus in European PCOS cases [45]. In this analysis there was directional consistency in effect on PCOS risk at 10 of the initially reported 11 signals identified in Chinese PCOS cohorts; however, only 6 were nominally ($P < 0.05$) associated, and due to consistently smaller effect sizes, none were genome-wide significant in the discovery GWAS. The effects of different LD structures between Han Chinese and European populations resulted in three of

these loci (2p21 (*THADA*), 9q33.3 (*DENND1A*), and 11q22.1 (*YAP1*)) having different lead SNPs, only one of which (rs11225154; *YAP1*) is in LD with the lead SNP reported in Chinese PCOS cases [46]. Three novel loci were identified in this GWAS at 2q34 (*ERBB4*), 5q31 (*IRF1/RAD50*), and 12q21.2 (*KRR1*) as PCOS risk regions at genome-wide significance. Three members of the EGFR gene family (*ERBB4*, *ERBB3*, and *ERBB2*) were identified as risk loci at, or close to, genome-wide significance in this analysis. Recent studies identified a role for *ErbB4* in the ovary, where it regulates anti-Müllerian hormone (AMH) level and folliculogenesis [51]. The risk association signal detected at 5q31 is within a complex, gene dense region. The index SNP lies within intron 3 of *C5orf55* and intron 4 of *IRF1* as well as within the reading frame for an uncharacterized protein-coding transcript AC116366.3. Nearby genes also include the transporter *SLC22A5*, an anti-sense RNA to the nearby gene *IRF1*, *IRF1-AS1*, the B cell growth factor *IL5*, and the double strand break repair gene *RAD50*. It is difficult to identify a candidate causal transcript at this locus given its complexity and what is known about the function of the genes in this region. To further identify potential biological mechanisms by which identified risk variants may impact PCOS biology, a quantitative analysis of the six genome-wide significant loci identified by Day et al. 2015 revealed an association between these six PCOS risk alleles and AMH levels in girls [46], suggesting that PCOS risk alleles from across the genome act through endocrine and reproductive pathways.

An international collaborative consortium assembled the largest GWAS of PCOS to date in order to identify risk loci in PCOS cases of European descent [52]. This analysis included more than 10,000 cases and 100,000 controls from seven cohorts (effective sample size 18,000), including a large proportion of previously analyzed cases [45, 46]. Imputation was conducted using the 1000 Genomes database, yielding over ten million SNPs for the GWAS. Fourteen risk loci were identified in this consortium effort. Three loci initially reported in GWAS studies of Chinese PCOS cases were replicated at genome-wide significance: 2p21 (*THADA*), 9q33.3 (*DENND1A*), and 16q21.1 (*TOX3*). The two risk loci, located at 8p23.1 (*GATA4/NEIL2*) and 11p14.1 (*FSHB*), reported by Hayes et al. [45] were confirmed in this large meta-analysis, as were the three risk loci at 2q34 (*ERBB4*), 5q31.1 (*IRF1/RAD50*), and 12q21.2 (*KRR1*) reported by Day et al. [46]. Three novel loci were identified in this collaborative meta-analysis at 9p24.1 (*PLGRKT*), 11q23.2 (*ZBTB16*), and 20q11.21 (*MAPRE1*). An additional novel genome-wide significant locus was identified on the X chromosome at the *ARSD* locus but was excluded from the formal results of the analysis due to low imputation quality, low minor allele frequency, and heterogeneity of effect across the three cohorts that had SNP data available for the X chromosome [52]. Additional analyses of this region in a larger sample size are needed to resolve the potential role of this locus in PCOS risk. Given that this GWAS included PCOS cases identified by self-report and two different clinical diagnostic criteria, heterogeneity analysis was performed to identify loci that demonstrated a difference in effect by these strata. The analysis identified heterogeneity at a single locus, 8p23.1 (*GATA4/NEIL2*), where the effect size associated with the risk allele was significantly less in self-reported PCOS cases and significantly greater in PCOS cases diagnosed using the NIH criteria [52]. For the

remaining 13 loci, the magnitude of association with PCOS was similar regardless of mode of diagnosis. This lack of heterogeneity across PCOS cases identified using these different criteria, along with the consistent replication of PCOS risk loci across individual studies, underscores a conserved shared genetic architecture for this phenotype.

Day et al. 2018 combined the PCOS GWAS data with results from GWAS for other traits to carry out genetic correlation analyses [52]. Such analyses suggest shared etiology but do not indicate directionality or causality. This investigation found genetic correlation between PCOS and body mass index (the most correlated trait), childhood obesity, fasting insulin, type 2 diabetes, high-density lipoprotein cholesterol, triglyceride levels, age of menarche, coronary artery disease, and depression. No genetic correlation was observed between PCOS and age of menopause or male pattern balding.

As the use of research biobanks has grown over recent years, the ability for case identification via electronic medical records has facilitated the analysis of population-based cohorts recruited through large medical care systems. Two such systems are the Geisinger MyCode Community Health Initiative that has recruited more than 250,000 research participants throughout the care system in Pennsylvania [53] and the collaborative eMERGE (electronic MEDical Records and GENomics) network that combines biobanks or studies with clinical data derived from medical records from across many sites [54]. Two such programs performed a GWAS in close to 3000 PCOS cases that met two of the following: (a) diagnosis of PCOS or polycystic ovaries; (b) hyperandrogenism or its related signs, or hyperandrogenemia; and (c) oligomenorrhea, amenorrhea, or infertility (i.e., Rotterdam diagnosis criteria) and 53,000 controls that did not meet any of the three criteria [55]. A small validation cohort of 253 cases and 2161 controls was available from the Vanderbilt BioVu study. This analysis identified three genome-wide significant signals (at 6q25.3, 2q34, and 3q25.1). The locus at 6q25.3 had not been detected in prior studies. The index SNP at this locus is more than 200 kb from the nearest genes (*FNDCl* and *SOD2*) and does not overlap known regulatory elements from ENCODE or 3D chromatin interactions reported by GeneHancer. It is not immediately apparent what the causal gene is at this locus. The previously reported risk signal at 2q34 (*ERBB4*) was identified in this study at a suggestive level of significance, and additionally a novel independent risk variant was identified at this locus at genome-wide significance. A third locus at 3q25.1 (*WWTR1*) was reported as nearing genome-wide significance; this locus has not been previously reported as a risk locus for PCOS [55]. It should be noted that 17% of the total cohort in this study was listed as African American, although the numbers of cases and controls were not provided. A lookup of the three reported risk loci identified in this study was performed in an analysis of only African American participants, and only the novel risk SNP identified at 2q34 (*ERBB4*) passed quality control metrics. Despite having a higher minor allele frequency in African American populations, this SNP was only nominally associated with PCOS risk ($P > 0.01$) [55]. Genome-wide association studies in populations of other ethnicities have not been performed. Our lack of understanding of the shared or differing genetic architecture of PCOS in populations that are

not of Chinese or European ancestry represents a significant deficit in our understanding. A major focus of ongoing research should prioritize the recruitment and profiling of PCOS cases and controls of other ancestries (e.g., Hispanic, African) to address this lack of knowledge.

To better identify the biological pathways through which susceptibility loci act to increase risk of PCOS, association of these loci with phenotypic traits related to PCOS has been performed in several studies, including the recent meta-analysis. Significant associations between known risk loci and polycystic ovarian morphology, ovulatory dysfunction, and hyperandrogenism were all identified [52]. GWAS analyses within PCOS cases also found that the allele associated with increased risk of PCOS at the *FSHB* locus was also associated with increased circulating LH level, decreased FSH level, and increased ratio of LH to FSH [45, 46]. Taken together these analyses further support the role for much of the genetic basis for PCOS to act through disrupting hormone pathways.

Polygenic Risk Scores for Disease Risk Prediction in PCOS

Polygenic risk scores (PRS) have been under active development in recent years, leveraging the increasing pace of discovery of the polygenic genetic architecture of many complex traits and the increasing sample sizes that are becoming available for testing and validation of such scores. The development of methods used to generate such scores is an active area, with empirical and Bayesian methods currently being applied. The long-term goal of PRS application in the population is to allow the early detection of risk for disease prevention strategies to be deployed [56]. This strategy is underway in cardiovascular traits, where the polygenic genetic risk estimated by GWAS equals the known monogenic risk and clinical risk factors [57]. A polygenic risk score for PCOS was developed based on the meta-analysis performed on clinically diagnosed cases included in the collaborative meta-analysis [52] and applied to a cohort of more than 120,000 individuals for whom electronic health records were available through the eMERGE network [58]. The best performing PRS in this analysis demonstrated a prediction accuracy of PCOS cases of 0.55 with an area under the curve (AUC) of 0.715 in eMERGE participants of European ancestry. When combined with information available based on PCOS component phenotypes, the PRS plus phenotype model performed with an accuracy of 0.873 and an AUC of 0.87, indicating that the PRS model built from this analysis is able to predict PCOS phenotype in individuals of European ancestry [58]. This genetic PRS model was also used to perform a phenome-wide association study (PheWAS), where the genetic risk score of an individual is used to identify anthropometric and clinical traits that are enriched in individuals of high genetic risk. This analysis can identify cross phenotype associations that may be the result of pleiotropy – whereby risk alleles impact multiple traits or phenotypes [59]. A significant PheWAS relationship was identified between the PCOS PRS and traits related to endocrine and metabolic traits (obesity, lipid dysfunction, type 2 diabetes), neurological traits (sleep apnea), circulatory system (hypertension), and digestive traits (esophageal

disease) [58]. Many of these associations remained significant after the analysis was repeated without any PCOS cases included in the cohort, suggesting that there are likely undiagnosed PCOS cases within the eMERGE network.

Mendelian Randomization Using GWAS Signals

Even before causal genes are identified at risk loci, GWAS information can be used to dissect the biology of disease. A major example is that robust loci identified by GWAS can be used to interrogate causality between an exposure and an outcome using Mendelian randomization (MR). In this approach, SNPs associated with the exposure are used as instrument variables to estimate the genetically driven effect of the exposure on the outcome, yielding causal effect estimates. Reports of PCOS GWAS included MR analyses that suggested increased body mass index (BMI), age at menopause, decreased sex hormone-binding globulin (SHBG), fasting insulin, male pattern balding, and depression were causal factors for PCOS [46, 52]. The relationship between BMI and PCOS has been extensively investigated using MR, with results finding that while obesity appears to be causal for PCOS, PCOS does not cause obesity [60, 61]. MR studies found that testosterone levels, but not AMH levels, are causal for PCOS [62, 63].

A series of MR studies examined PCOS as the exposure against various outcomes, using PCOS SNPs from the largest GWAS for PCOS [52] as instrument variables. PCOS was found not to have a genetic causal effect on type 2 diabetes, coronary heart disease, or stroke [64]. Given that prior MR studies had demonstrated causal effects of BMI, higher testosterone, and lower SHBG on diabetes and/or cardiovascular disease, the authors concluded that these features commonly present in PCOS, rather than PCOS in and of itself, explain the association between PCOS and cardiometabolic disease. Genetically predicted PCOS was associated with increased risk of breast cancer overall and estrogen receptor-positive breast cancer; no effect on estrogen receptor-negative breast cancer was observed [65]. Consistent results were observed in a study that examined several subtypes of breast cancer [66]. MR studies found a protective effect of PCOS against invasive ovarian cancer and endometrioid ovarian cancer [67, 68]. These MR studies yielded key insights on causes and consequences of PCOS, avoiding confounding variables that affect epidemiological association studies.

Identifying Causal Genes at PCOS Risk Loci

Colocalization analysis of disease and intermediate cellular phenotypes (e.g., gene expression and protein level across different relevant tissues) is performed by measuring the probability that the two traits share a causal variant [69]. A recent analysis applied this approach and successfully identified seven proteins with strong evidence of colocalization [70]. The FSH protein was clearly implicated at the 11p14.1 locus where the significant correlation between genotype at risk-associated

SNPs and circulating FSH level presents a clear colocalization of the same causal SNPs acting on both PCOS and FSH level. This approach was unable to resolve a single likely causal transcript at the 12q13.2 locus but implicated *SUOX*, *ERBB3*, *IKZF4*, *RPS26*, and *GDF11* as potential causal genes. A single likely causal gene, *ZFP36L2*, was identified at 2q21 (*THADA* locus), and *C9orf3* was implicated at 9p24.1. Colocalization analysis at 8p23.1 identified both *C8orf49* and *NEIL2* as potential causal transcripts [70].

Conclusion and Future Directions

Advances in genomic technology have led to rapid progress in our understanding of the genetic architecture of PCOS. Though PCOS is clinically heterogeneous, GWAS have found little genetic heterogeneity across PCOS diagnostic criteria. Twenty loci across the genome have been identified at genome-wide significance in Chinese and/or European cohorts (Table 4.1). The causal gene at many of these loci is unknown; however, genomic analysis and in vitro studies have provided some suggestion of the likely causal gene at specific loci. These results indicate that disruption of hormone signaling pathways, particularly related to the synthesis and signaling of FSH and the signaling of the LH receptor, are key to the pathogenesis of PCOS. As with many complex traits, much of the heritability for PCOS has yet to be identified. Identifying additional risk alleles will contribute to improved PRS accuracy and sensitivity and may identify further biological pathways to be targeted for the treatment of PCOS symptoms. Increasing sample sizes will be required for the discovery of additional risk alleles, and the continued efforts of the International PCOS Consortium (iPCOS) are focused on including increasing numbers of PCOS cases and controls for ongoing meta-analysis for risk allele discovery. A second focus of the iPCOS consortium is to foster the inclusion of PCOS cases and controls of Hispanic and African ancestry, so that we may begin to understand the shared and differing genetic architecture of PCOS between these populations and those already studied. The current move of genomic technologies beyond array-based genotyping into population-level whole genome sequencing will provide opportunities to discover additional types of risk variants (e.g., structural variants) and variants with rare and very rare risk allele frequencies, allowing a deeper understanding of the complex genetic underpinnings of PCOS.

References

1. Hague WM, Adams J, Reeders ST, Peto TE, Jacobs HS. Familial polycystic ovaries: a genetic disease? *Clin Endocrinol.* 1988;29(6):593–605. <https://doi.org/10.1111/j.1365-2265.1988.tb03707.x>.
2. Ferriman D, Purdie AW. The inheritance of polycystic ovarian disease and a possible relationship to premature balding. *Clin Endocrinol.* 1979;11(3):291–300. <https://doi.org/10.1111/j.1365-2265.1979.tb03077.x>.

3. Cooper HE, Spellacy WN, Prem KA, Cohen WD. Hereditary factors in the Stein-Leventhal syndrome. *Am J Obstet Gynecol.* 1968;100(3):371–87. [https://doi.org/10.1016/s0002-9378\(15\)33704-2](https://doi.org/10.1016/s0002-9378(15)33704-2).
4. Govind A, Obhrai MS, Clayton RN. Polycystic ovaries are inherited as an autosomal dominant trait: analysis of 29 polycystic ovary syndrome and 10 control families. *J Clin Endocrinol Metab.* 1999;84(1):38–43. <https://doi.org/10.1210/jcem.84.1.5382>.
5. Carey AH, Chan KL, Short F, White D, Williamson R, Franks S. Evidence for a single gene effect causing polycystic ovaries and male pattern baldness. *Clin Endocrinol.* 1993;38(6):653–8. <https://doi.org/10.1111/j.1365-2265.1993.tb02150.x>.
6. Kashar-Miller M, Azziz R. Heritability and the risk of developing androgen excess. *J Steroid Biochem Mol Biol.* 1999;69(1–6):261–8. [https://doi.org/10.1016/s0960-0760\(99\)00043-6](https://doi.org/10.1016/s0960-0760(99)00043-6).
7. Vink JM, Sadrzadeh S, Lambalk CB, Boomsma DI. Heritability of polycystic ovary syndrome in a Dutch twin-family study. *J Clin Endocrinol Metab.* 2006;91(6):2100–4. <https://doi.org/10.1210/jc.2005-1494>.
8. Shi Y, Zhao H, Cao Y, Yang D, Li Z, Zhang B, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet.* 2012;44(9):1020–5. <https://doi.org/10.1038/ng.2384>.
9. Urbanek M, Legro RS, Driscoll DA, Azziz R, Ehrmann DA, Norman RJ, et al. Thirty-seven candidate genes for polycystic ovary syndrome: strongest evidence for linkage is with follistatin. *Proc Natl Acad Sci U S A.* 1999;96(15):8573–8. <https://doi.org/10.1073/pnas.96.15.8573>.
10. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57–74. <https://doi.org/10.1038/nature11247>.
11. Xu X, Zhao H, Shi Y, You L, Bian Y, Zhao Y, et al. Family association study between INSR gene polymorphisms and PCOS in Han Chinese. *Reprod Biol Endocrinol.* 2011;9:76. <https://doi.org/10.1186/1477-7827-9-76>.
12. Goodarzi MO, Louwers YV, Taylor KD, Jones MR, Cui J, Kwon S, et al. Replication of association of a novel insulin receptor gene polymorphism with polycystic ovary syndrome. *Fertil Steril.* 2011;95(5):1736–41.e1–11. <https://doi.org/10.1016/j.fertnstert.2011.01.015>.
13. Mukherjee S, Shaikh N, Khavale S, Shinde G, Meherji P, Shah N, et al. Genetic variation in exon 17 of INSR is associated with insulin resistance and hyperandrogenemia among lean Indian women with polycystic ovary syndrome. *Eur J Endocrinol.* 2009;160(5):855–62. <https://doi.org/10.1530/EJE-08-0932>.
14. Lee EJ, Oh B, Lee JY, Kimm K, Lee SH, Baek KH. A novel single nucleotide polymorphism of INSR gene for polycystic ovary syndrome. *Fertil Steril.* 2008;89(5):1213–20. <https://doi.org/10.1016/j.fertnstert.2007.05.026>.
15. Lee EJ, Yoo KJ, Kim SJ, Lee SH, Cha KY, Baek KH. Single nucleotide polymorphism in exon 17 of the insulin receptor gene is not associated with polycystic ovary syndrome in a Korean population. *Fertil Steril.* 2006;86(2):380–4. <https://doi.org/10.1016/j.fertnstert.2005.12.073>.
16. Jin L, Zhu XM, Luo Q, Qian Y, Jin F, Huang HF. A novel SNP at exon 17 of INSR is associated with decreased insulin sensitivity in Chinese women with PCOS. *Mol Hum Reprod.* 2006;12(3):151–5. <https://doi.org/10.1093/molehr/gal022>.
17. Chen ZJ, Shi YH, Zhao YR, Li Y, Tang R, Zhao LX, et al. Correlation between single nucleotide polymorphism of insulin receptor gene with polycystic ovary syndrome. *Zhonghua Fu Chan Ke Za Zhi.* 2004;39(9):582–5.
18. Siegel S, Futterweit W, Davies TF, Concepcion ES, Greenberg DA, Villanueva R, et al. A C/T single nucleotide polymorphism at the tyrosine kinase domain of the insulin receptor gene is associated with polycystic ovary syndrome. *Fertil Steril.* 2002;78(6):1240–3. [https://doi.org/10.1016/s0015-0282\(02\)04241-3](https://doi.org/10.1016/s0015-0282(02)04241-3).
19. Tucci S, Futterweit W, Concepcion ES, Greenberg DA, Villanueva R, Davies TF, et al. Evidence for association of polycystic ovary syndrome in Caucasian women with a marker at the insulin receptor gene locus. *J Clin Endocrinol Metab.* 2001;86(1):446–9. <https://doi.org/10.1210/jcem.86.1.7274>.
20. Unsal T, Konac E, Yesilkaya E, Yilmaz A, Bideci A, Ilke Onen H, et al. Genetic polymorphisms of FSHR, CYP17, CYP11A1, CAPN10, INSR, SERPINE1 genes in adolescent girls

- with polycystic ovary syndrome. *J Assist Reprod Genet.* 2009;26(4):205–16. <https://doi.org/10.1007/s10815-009-9308-8>.
21. Li T, Wu K, You L, Xing X, Wang P, Cui L, et al. Common variant rs9939609 in gene FTO confers risk to polycystic ovary syndrome. *PLoS One.* 2013;8(7):e66250. <https://doi.org/10.1371/journal.pone.0066250>.
 22. Wojciechowski P, Lipowska A, Rys P, Ewens KG, Franks S, Tan S, et al. Impact of FTO genotypes on BMI and weight in polycystic ovary syndrome: a systematic review and meta-analysis. *Diabetologia.* 2012;55(10):2636–45. <https://doi.org/10.1007/s00125-012-2638-6>.
 23. Ewens KG, Jones MR, Ankener W, Stewart DR, Urbanek M, Dunaif A, et al. FTO and MC4R gene variants are associated with obesity in polycystic ovary syndrome. *PLoS One.* 2011;6(1):e16390. <https://doi.org/10.1371/journal.pone.0016390>.
 24. Barber TM, Bennett AJ, Groves CJ, Sovio U, Ruokonen A, Martikainen H, et al. Association of variants in the fat mass and obesity associated (FTO) gene with polycystic ovary syndrome. *Diabetologia.* 2008;51(7):1153–8. <https://doi.org/10.1007/s00125-008-1028-6>.
 25. Ben-Salem A, Ajina M, Suissi M, Daher HS, Almawi WY, Mahjoub T. Polymorphisms of transcription factor-7-like 2 (TCF7L2) gene in Tunisian women with polycystic ovary syndrome (PCOS). *Gene.* 2014;533(2):554–7. <https://doi.org/10.1016/j.gene.2013.09.104>.
 26. Dasgupta S, Sirisha PV, Neelaveni K, Anuradha K, Reddy BM. Association of CAPN10 SNPs and haplotypes with polycystic ovary syndrome among South Indian Women. *PLoS One.* 2012;7(2):e32192. <https://doi.org/10.1371/journal.pone.0032192>.
 27. Ewens KG, Jones MR, Ankener W, Stewart DR, Urbanek M, Dunaif A, et al. Type 2 diabetes susceptibility single-nucleotide polymorphisms are not associated with polycystic ovary syndrome. *Fertil Steril.* 2011;95(8):2538–41.e6. <https://doi.org/10.1016/j.fertnstert.2011.02.050>.
 28. Liu X, Li L, Chen ZJ, Lu Z, Shi Y, Zhao Y. Genetic variants of cyclin-dependent kinase 5 regulatory subunit associated protein 1-like 1 and transcription factor 7-like 2 are not associated with polycystic ovary syndrome in Chinese women. *Gynecol Endocrinol.* 2010;26(2):129–34. <https://doi.org/10.3109/09513590903215490>.
 29. Haddad L, Evans JC, Gharani N, Robertson C, Rush K, Wiltshire S, et al. Variation within the type 2 diabetes susceptibility gene calpain-10 and polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2002;87(6):2606–10. <https://doi.org/10.1210/jcem.87.6.8608>.
 30. Ehrmann DA, Schwarz PE, Hara M, Tang X, Horikawa Y, Imperial J, et al. Relationship of calpain-10 genotype to phenotypic features of polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2002;87(4):1669–73. <https://doi.org/10.1210/jcem.87.4.8385>.
 31. Wickham EP 3rd, Ewens KG, Legro RS, Dunaif A, Nestler JE, Strauss JF 3rd. Polymorphisms in the SHBG gene influence serum SHBG levels in women with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2011;96(4):E719–27. <https://doi.org/10.1210/jc.2010-1842>.
 32. Goodarzi MO, Antoine HJ, Azziz R. Genes for enzymes regulating dehydroepiandrosterone sulfonation are associated with levels of dehydroepiandrosterone sulfate in polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2007;92(7):2659–64. <https://doi.org/10.1210/jc.2006-2600>.
 33. Jones MR, Wilson SG, Mullin BH, Mead R, Watts GF, Stuckey BG. Polymorphism of the foliastatin gene in polycystic ovary syndrome. *Mol Hum Reprod.* 2007;13(4):237–41. <https://doi.org/10.1093/molehr/gal120>.
 34. Qin K, Ehrmann DA, Cox N, Refetoff S, Rosenfield RL. Identification of a functional polymorphism of the human type 5 17beta-hydroxysteroid dehydrogenase gene associated with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2006;91(1):270–6. <https://doi.org/10.1210/jc.2005-2012>.
 35. Petry CJ, Ong KK, Michelmore KF, Artigas S, Wingate DL, Balen AH, et al. Association of aromatase (CYP 19) gene variation with features of hyperandrogenism in two populations of young women. *Hum Reprod.* 2005;20(7):1837–43. <https://doi.org/10.1093/humrep/deh900>.
 36. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012;337(6099):1190–5. <https://doi.org/10.1126/science.1222794>.

37. Chen ZJ, Zhao H, He L, Shi Y, Qin Y, Li Z, 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. <https://doi.org/10.1038/ng.732>.
38. Du J, Wang J, Sun X, Xu X, Zhang F, Wang B, et al. Family-based analysis of INSR polymorphisms in Chinese PCOS. *Reprod Biomed Online.* 2014;29(2):239–44. <https://doi.org/10.1016/j.rbmo.2014.03.028>.
39. Fu L, Zhang Z, Zhang A, Xu J, Huang X, Zheng Q, et al. Association study between FSHR Ala307Thr and Ser680Asn variants and polycystic ovary syndrome (PCOS) in Northern Chinese Han women. *J Assist Reprod Genet.* 2013;30(5):717–21. <https://doi.org/10.1007/s10815-013-9979-z>.
40. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet.* 2018;50(11):1505–13. <https://doi.org/10.1038/s41588-018-0241-6>.
41. Zhu M, Xu K, Chen Y, Gu Y, Zhang M, Luo F, et al. Identification of novel T1D risk loci and their association with age and islet function at diagnosis in autoantibody-positive T1D individuals: based on a two-stage genome-wide association study. *Diabetes Care.* 2019;42(8):1414–21. <https://doi.org/10.2337/dc18-2023>.
42. Goodarzi MO, Korenman SG. The importance of insulin resistance in polycystic ovary syndrome. *Fertil Steril.* 2003;80(2):255–8. [https://doi.org/10.1016/s0015-0282\(03\)00734-9](https://doi.org/10.1016/s0015-0282(03)00734-9).
43. Lee H, Oh JY, Sung YA, Chung H, Kim HL, Kim GS, et al. Genome-wide association study identified new susceptibility loci for polycystic ovary syndrome. *Hum Reprod.* 2015;30(3):723–31. <https://doi.org/10.1093/humrep/deu352>.
44. Hwang JY, Lee EJ, Jin Go M, Sung YA, Lee HJ, Heon Kwak S, et al. Genome-wide association study identifies GYS2 as a novel genetic factor for polycystic ovary syndrome through obesity-related condition. *J Hum Genet.* 2012;57(10):660–4. <https://doi.org/10.1038/jhg.2012.92>.
45. Hayes MG, Urbanek M, Ehrmann DA, Armstrong LL, Lee JY, Sisk R, et al. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. *Nat Commun.* 2015;6:7502. <https://doi.org/10.1038/ncomms8502>.
46. Day FR, Hinds DA, Tung JY, Stolk L, Styrkarsdottir U, Saxena R, et al. Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nat Commun.* 2015;6:8464. <https://doi.org/10.1038/ncomms9464>.
47. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature.* 2017;550(7675):204–13. <https://doi.org/10.1038/nature24277>.
48. Wingender E, Dietze P, Karas H, Knuppel R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res.* 1996;24(1):238–41. <https://doi.org/10.1093/nar/24.1.238>.
49. Li B, Huang Q, Wei GH. The role of HOX transcription factors in cancer predisposition and progression. *Cancers (Basel).* 2019;11(4) <https://doi.org/10.3390/cancers11040528>.
50. Bohaczuk SC, Thackray VG, Shen J, Skowronska-Krawczyk D, Mellon PL. FSHB transcription is regulated by a novel 5' distal enhancer with a fertility-associated single nucleotide polymorphism. *Endocrinology.* 2021;162(1) <https://doi.org/10.1210/endoocr/bqaa181>.
51. Veikkolainen V, Ali N, Doroszko M, Kiviniemi A, Miinalainen I, Ohlsson C, et al. Erbb4 regulates the oocyte microenvironment during folliculogenesis. *Hum Mol Genet.* 2020;29(17):2813–30. <https://doi.org/10.1093/hmg/ddaa161>.
52. Day F, Karaderi T, Jones MR, Meun C, He C, Drong A, et al. Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PLoS Genet.* 2018;14(12):e1007813. <https://doi.org/10.1371/journal.pgen.1007813>.
53. Carey DJ, Fetterolf SN, Davis FD, Faucett WA, Kirchner HL, Mirshahi U, et al. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet Med.* 2016;18(9):906–13. <https://doi.org/10.1038/gim.2015.187>.

54. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, et al. The eMERGE network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genet.* 2011;4:13. <https://doi.org/10.1186/1755-8794-4-13>.
55. Zhang Y, Ho K, Keaton JM, Hartzel DN, Day F, Justice AE, et al. A genome-wide association study of polycystic ovary syndrome identified from electronic health records. *Am J Obstet Gynecol.* 2020;223(4):559.e1–e21. <https://doi.org/10.1016/j.ajog.2020.04.004>.
56. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet.* 2018;19(9):581–90. <https://doi.org/10.1038/s41576-018-0018-x>.
57. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet.* 2018;50(9):1219–24. <https://doi.org/10.1038/s41588-018-0183-z>.
58. Joo YY, Actkins K, Pacheco JA, Basile AO, Carroll R, Crosslin DR, et al. A polygenic and phenotypic risk prediction for polycystic ovary syndrome evaluated by phenome-wide association studies. *J Clin Endocrinol Metab.* 2020;105(6) <https://doi.org/10.1210/clinem/dgz326>.
59. Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nat Rev Genet.* 2016;17(3):129–45. <https://doi.org/10.1038/nrg.2015.36>.
60. Brower MA, Hai Y, Jones MR, Guo X, Chen YDI, Rotter JJ, et al. Bidirectional Mendelian randomization to explore the causal relationships between body mass index and polycystic ovary syndrome. *Hum Reprod.* 2019;34(1):127–36. <https://doi.org/10.1093/humrep/dey343>.
61. Zhao YL, Xu YP, Wang XM, Xu L, Chen JH, Gao CW, et al. Body mass index and polycystic ovary syndrome: a 2-sample bidirectional Mendelian randomization study. *J Clin Endocrinol Metab.* 2020;105(6):1778–84. <https://doi.org/10.1210/clinem/dgaa125>.
62. Ruth KS, Day FR, Tyrrell J, Thompson DJ, Wood AR, Mahajan A, et al. Using human genetics to understand the disease impacts of testosterone in men and women. *Nat Med.* 2020;26(2):252. <https://doi.org/10.1038/s41591-020-0751-5>.
63. Verdiesen RM, van der Schouw YT, van Gils CH, Verschuren WM, Broekmans FJ, Borges MC, et al. Genome-wide association study meta-analysis identifies three novel loci for circulating anti-Müllerian hormone levels in women. *medRxiv.* 2020; <https://doi.org/10.1101/2020.10.29.20221390>.
64. Zhu T, Cui J, Goodarzi MO. Polycystic ovary syndrome and risk of type 2 diabetes, coronary heart disease, and stroke. *Diabetes.* 2021;70(2):627–37. <https://doi.org/10.2337/db20-0800>.
65. Wu PF, Li RZ, Zhang W, Hu HY, Wang W, Lin Y. Polycystic ovary syndrome is causally associated with estrogen receptor-positive instead of estrogen receptor-negative breast cancer: a Mendelian randomization study. *Am J Obstet Gynecol.* 2020;223(4):583–5. <https://doi.org/10.1016/j.ajog.2020.05.016>.
66. Zhu T, Cui J, Goodarzi MO. Polycystic ovary syndrome and breast cancer subtypes: a Mendelian randomization study. *Am J Obstet Gynecol.* 2021; <https://doi.org/10.1016/j.ajog.2021.03.020>.
67. Harris HR, Cushing-Haugen KL, Webb PM, Nagle CM, Jordan SJ, Risch HA, et al. Association between genetically predicted polycystic ovary syndrome and ovarian cancer: a Mendelian randomization study. *Int J Epidemiol.* 2019;48(3):822–30. <https://doi.org/10.1093/ije/dyz113>.
68. Yarmolinsky J, Relton CL, Lophatananon A, Muir K, Menon U, Gentry-Maharaj A, et al. Appraising the role of previously reported risk factors in epithelial ovarian cancer risk: a Mendelian randomization analysis. *PLoS Med.* 2019;16(8):e1002893. <https://doi.org/10.1371/journal.pmed.1002893>.
69. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10(5):e1004383. <https://doi.org/10.1371/journal.pgen.1004383>.
70. Censin JC, Bovijn J, Holmes MV, Lindgren CM. Colocalization analysis of polycystic ovary syndrome to identify potential disease-mediating genes and proteins. *Eur J Hum Genet.* 2021; <https://doi.org/10.1038/s41431-021-00835-8>.