CHAPTER

20

Genetics of Polycystic Ovary Syndrome

Tristan Hardy[*,†], *Robert J. Norman*[‡,§,¶,||]

[*]Genetics and Molecular Pathology, SA Pathology, Adelaide, SA, Australia [†]Repromed, Dulwich, SA, Australia [‡]University of Adelaide, Adelaide, SA, Australia [§]Fertility SA, Adelaide, SA, Australia [¶]Royal Adelaide Hospital, Adelaide, SA, Australia [||]NHMRC Centre for Research Excellence in Polycystic Ovary Syndrome, Adelaide, SA, Australia

OUTLINE

Abbreviations

AMH	antiMullerian hormone
BMI	body mass index
FSH	follicle stimulating hormone
GWAS	genome-wide association studies
hCG	human chorionic gonadotropin
LH	luteinizing hormone
NIH	National Institutes of Health
PCOS	polycystic ovary syndrome

https://doi.org/10.1016/B978-0-12-813570-9.00020-6

SNP single nucleotide polymorphism
T1DM type 1 diabetes mellitus
T2DM type 2 diabetes mellitus
WES whole exome sequencing
WGS whole genome sequencing

INTRODUCTION

Polycystic ovary syndrome (PCOS) is the most common endocrine disorder affecting women in their reproductive phase of life. The phenotype of PCOS is heterogeneous, manifesting in its more pronounced form primarily as chronic anovulation and/or hyperandrogenism. As a result of the clinical heterogeneity of the disorder, a variety of diagnostic criteria have been proposed that attempt to capture the salient elements of the phenotype.

It has been evident for some time that PCOS is a highly heritable condition, and numerous studies have been performed attempting to dissect the underlying genetic architecture of the condition. As a polygenic disorder involving a range of endocrine and metabolic pathways, it has been difficult to isolate individual genetic variants with evidence of high impact using the candidate gene approach. Genome-wide analytical techniques such as genome-wide association studies (GWAS) have identified numerous loci of interest, many of which have been confirmed across different populations, suggesting the impact of nearby genes on the underlying pathogenesis of the condition. However, these loci explain only a fraction of the estimated heritability of the condition, and it is likely that further studies using whole exome or whole genome sequencing (WES/WGS) will be required to identify additional genetic variants of high impact.

This chapter focuses on genomic approaches to the identification of important loci in PCOS pathogenesis, reviewing in detail the current GWAS studies and the emerging studies using WES/WGS.

DEFINING PCOS

Several well-known criteria for the diagnosis of PCOS have been enumerated [1], the most commonly used being the Rotterdam criteria that require two out of three of oligo/anovulation, clinical or biochemical hyperandrogenism, and polycystic ovarian morphology [2]. The variety of diagnostic criteria and the clinical heterogeneity of the condition has made it difficult to study the underlying genetic architecture of PCOS, as accurate phenotyping is a key component of any genetic study. Some studies have used polycystic ovaries alone to define the condition whereas others have demanded hyperandrogenism and menstrual disturbances to be included in any genetic examination. Regardless of the clinical criteria used, it has been evident to researchers for some time that the features of PCOS appear to have a familial relationship, suggesting a genetic component to the condition [3].

HERITABILITY OF PCOS

Initial studies of the heritability of PCOS concluded that the disorder was most likely a dominantly inherited trait with low penetrance and expressivity [4], although more recently these features have been thought to be consistent with a polygenic model of inheritance, along with the majority of complex phenotypic traits [5]. Cooper et al. [3] were the first to systematically analyze the potential genetic contributions to PCOS, ruling out a chromosomal origin of the disorder and suggesting that the pattern of symptoms (predominantly menstrual irregularities and hirsutism) in mothers and sisters of affected probands was most likely consistent with an autosomal dominant mode of inheritance. Expansion of inclusion criteria to patients with polycystic ovarian morphology on ultrasound similarly concluded that the pattern of features in family members was consistent with an autosomal dominant mode of transmission [6]. Subsequent studies confirmed that metabolic features of the disorder were also prevalent in family members, including male relatives. Hyperinsulinemia and hypertriglyceridemia affected 69% and 56% of family members overall, and polycystic ovaries were present in 74% of female relatives while male pattern baldness was found in 88% of male relatives [7]. Legro et al. [8] similarly demonstrated that sisters of probands with PCOS had an increased risk of both PCOS (22%) and hyperandrogenemia with regular menstrual cycles (24%).

Twin studies have been extremely effective in establishing the heritability of multiple conditions, and initial studies with small numbers concluded that PCOS was not likely to result from variants within a single gene but was either polygenic or multifactorial [9]. The most definitive twin family study analyzed 3205 females (comprising 1332 monozygotic twins, 680 dizygotic twins, 474 individuals from dizygotic opposite sex pairs, and 719 nontwin sisters), according to the Rotterdam criteria [10]. Monozygotic twin pairs had significantly higher correlation of the features of PCOS than dizygotic or nontwin sisters; the relative risk of oligomenorrhea was 0.67 versus 0.07, acne 0.78 versus 0.44, and hirsutism 0.86 versus 0.28 [10]. Overall, the risk of PCOS in monozygotic twin sisters was 0.71 (95% CI 0.43 to 0.88) versus 0.38 (95% CI 0.00 to 0.66) in dizygotic twin or nontwin sisters. After modeling the three variables of oligomenorrhea, acne, and hirsutism in an independent pathway model, the authors concluded that PCOS was affected predominantly by genetic variance (79%) and unique environmental influence (21%), with no role for shared environmental influence [10]. Although this and many other studies pointed to the strongly heritable nature of PCOS, the establishment of heritability gave no clues as to the putative genes involved.

CANDIDATE GENE STUDIES

The majority of genetic studies in PCOS have used the candidate gene approach [11], partly due to investigator-driven hypotheses regarding the pathogenesis of PCOS, but also as a reflection of the technology available for genetic analysis at the time of the studies. Overall, a recent database included information on 241 genes and 114 SNPs that have been associated with PCOS [12], demonstrating the variety of genetic variants that have been identified in the literature. Within these, a few promising candidate genes have been identified and replicated

in a number of studies, although their overall contribution to the pathogenesis of PCOS remains unclear. Generally there is a history of identifying gene relationships that do not stand the test of time.

The only candidate gene that has been subsequently validated by GWAS studies is the insulin receptor (*INSR*) [13]. Tucci et al. [14] identified the polymorphic variant (D19S884) upstream of *INSR* within the fibrillin 3 gene (*FBN3*) as a potential risk factor for PCOS, but could not validate whether the susceptibility locus was in the insulin gene receptor itself or within the gene region. The D19S884 marker was subsequently analyzed in a larger group of 367 families [15] and 453 families [16], with both authors concluding that there was a significant association with the transmission of PCOS. These findings were replicated in a study of Han Chinese patients with PCOS [17], although a smaller Spanish study concluded that there was no relationship between the D19S884 marker and PCOS [18].

Given the importance of insulin resistance to the overall phenotype of PCOS, there has been ongoing investigation into the role of polymorphisms in the insulin receptor gene itself with seven studies finding SNPs within *INSR* contributing to PCOS risk, one of which was eventually confirmed in a large study (rs2252673) [13]. Researchers have also postulated that the strong association between PCOS and obesity may imply a relationship between obesity-susceptibility variants and PCOS. Of these, the most studied is that of the fat mass and obesity associated (*FTO*) gene, in particular, the rs9939609 polymorphism [19]. Meta-analysis of five studies covering 5010 PCOS patients and 5300 controls demonstrated that rs9939609 was significantly different between groups, suggesting that the A allele was a risk factor for PCOS susceptibility in both Asian and Caucasian subgroups (OR 1.43 and OR 1.33, respectively) [19].

The genetic basis of hyperandrogenism in PCOS has led to a number of studies exploring variants in the enzymes involved in the steroidogenic pathway [6] and sensitivity to androgen signaling [20]. Early studies of the polymorphic trinucleotide repeat (CAG) in the androgen receptor gene suggested that infertile women with PCOS had a greater number of repeats compared with fertile control patients [21]. However, later meta-analysis of the role of CAG repeats within the androgen receptor gene concluded in favor of an association of the short repeat group with hyperandrogenism (56.25% vs. 29.14%, $P < 0.001$) [22]. Furthermore, in a discovery cohort of 354 PCOS and 161 control patients, SNPs within genes in the androgen receptor signaling pathway were analyzed (*HSPA1A*, *HSPA8*, *ST13*, *STIP1*, *PTGES3*, *FKBP4*, *BAG1*, and *STUB1*) and two SNPs in *FKBP4* were associated with a reduced odds ratio of PCOS. However, only one of these (rs4409904) was confirmed in a subsequent replication cohort (397 cases and 306 controls) [20]. A recent study also suggested that alternative splice variants of the androgen receptor may be associated with abnormal folliculogenesis and hyperandrogenemia in patients with PCOS [23]. Studies concurrently examining multiple genes in the steroidogenic pathway along with those involved with gonadotropin action, obesity, energy regulation, and insulin action have failed to demonstrate a compelling relationship between steroidogenic enzyme pathway variants and PCOS, but have suggested that variants in gonadotropin genes (Follistatin) may be associated with PCOS [24].

An inflammatory phenotype associated with PCOS and metabolic alterations has also been proposed, the genetic basis of which may be linked to polymorphisms in inflammatory cytokine genes such as *TNF-alpha*, *IL-6*, and *IL-1 beta* [25]. All reports examining polymorphisms

in these genes were subjected to meta-analysis, with 14 studies eligible for inclusion. For *TNF-alpha*, there were 802 cases and 802 controls in total, with no evidence of significance of the −308 G/A polymorphism with regard to PCOS or obesity. For *IL-6*, 351 cases and 464 controls were identified, examining the *−174 G/C* polymorphism and finding no significant association overall; however, using an allelic model led to some evidence of significance. Similarly, no obvious association between *IL-1 beta* polymorphism (−511C/T) was found, suggesting overall that previous studies using small numbers of cases and controls found associations that disappeared on wider meta-analysis [25].

Overall, despite many promising leads, the candidate gene approach is unsuited to the analysis of a polygenic condition such as PCOS, due to the extremely low likelihood of finding a high-impact variant in a single gene when comparing only small numbers of individuals with and without the disorder. In addition, previous studies have tested only for variants that are known in the population, and are therefore unlikely to discover new variants with high genetic impact. The transition to GWAS studies has proven useful for defining risk loci in PCOS due to the emergence of new genetic technologies.

GENOME-WIDE ASSOCIATION STUDIES

Improvements in the throughput of genetic analysis technologies have allowed the application of genome-wide techniques to analyze hundreds of thousands to millions of variants per individual, in significantly larger groups of cases and controls. There have been five major GWAS studies focusing on PCOS that have identified a number of promising genetic loci, many of which would not have been predicted by a hypothesis-driven candidate gene approach. The findings of these studies are worth reviewing in detail as they present the most up-to-date analysis of our current understanding of the heritability of PCOS.

Chen et al. [26] published the first GWAS study on PCOS in the Han Chinese population, analyzing 744 PCOS cases and 895 controls in a discovery set and further validating the susceptibility loci in 2840 PCOS cases and 5012 controls. The case definition in this study followed the 2003 Rotterdam Criteria requiring two of three criteria to be present: oligo- and/or anovulation, evidence of clinical and/or biochemical hyperandrogenism, and polycystic ovarian morphology, with the exclusion of other causes of oligomenorrhea or hyperandrogenism. Single nucleotide polymorphisms (SNPs) were analyzed using Affymetrix SNP 6.0 chips, with a total of 611,633 SNPs subjected to analysis over the population. In the discovery set, four distinct regions with 29 SNPs showed strong evidence of association: 2p16.3, 2p21, 5q14.3, and 9q33.3. Two independent replication sets from slightly different populations (northern Han Chinese, and southern and central Han Chinese) were used to validate the regions identified in the discovery set, with 28 of the 29 SNPs showing genome-wide significance, and 1 SNP representing the 5q14.3 region being removed from analysis. The three leading SNPs identified in the study were representative of each of the loci: rs13405728 at 2p16.3, rs13429458 at 2p21, and rs2479106 at 9q33.3. These regions remained significant when logistic regression analysis was performed to remove the effects of age and body mass index (BMI). Also, the regions did not overlap with those identified on previous GWAS studies analyzing BMI, suggesting an independent association with PCOS rather than metabolic dysfunction.

Analyzing nearby genes (±500 kb) and linkage disequilibrium (LD) blocks involving the most significant SNPs, Chen et al. [26] identified a number of candidate genes of interest in the pathogenesis of PCOS. At 2p16.3, the most significant nearby genes were the TFIIA-alpha and beta-like factor (*GTF2A1L*) as well as the luteinizing hormone and human chorionic gonadotropin gene receptor (*LHCGR*), both of which have a documented role in human reproduction. *GTF2A1L* had previously been reported to be a potential cause of human infertility, being expressed during germ cell development and playing a role in the testis [26]. The *LHCGR* gene has a central role in determining the sensitivity of granulosa cells to the luteinizing hormone, with previous studies suggesting inactivating mutations may result in LH resistance (manifesting as increased LH levels, enlarged ovaries, and oligomenorrhea) and activating mutations resulting in hyperandrogenism without infertility [26]. The association of SNP rs13405728 in this region did not significantly differ according to the BMI of the cases and controls, suggesting that the SNP independently predicted PCOS as defined by the Rotterdam criteria through a mechanism separate to the metabolic associations of PCOS. The authors also noted that the *FSHR* gene encoding the FSH receptor was nearby (211 kb downstream of rs13405728), although this location was beyond a strong recombination hotspot. Targeted analysis of 65 intragenic SNPs within *FSHR* demonstrated 13 with a PCA-adjusted *P*-value between 2×10^{-3} and 4×10^{-4}, lower than that required for genome-wide significance but potentially significant in combination, thus not ruling out a role for variants in the FSH receptor in susceptibility to PCOS.

The 2p21 region comprised the majority of significant SNPs (21 out of 28), spanning 304 kb and located within two different LD blocks [26]. The most significant independent associations in this region involved the *THADA* gene, which was originally identified in thyroid adenomas but may have a role in susceptibility to insulin resistance/type II diabetes mellitus (T2DM). Subjects with PCOS carrying these SNPs did not, however, have an increased risk of insulin resistance in this study population. The 9q33.3 region contained six significant SNPs spanning 42.3 kb within *DENND1A* (differentially expressed in normal and neoplastic cells domain containing 1A) [26]. This protein is a negative regulator of endoplasmic reticulum aminopeptidase 1 (ERAP1), which has previously been associated with PCOS and obesity [27].

A follow-up study from the same group [28] studied a larger cohort of Han Chinese (8226 cases and 7578 controls) in order to confirm the three previously reported loci, and independently identified eight new PCOS association signals. In addition, the authors identified an independent signal at the previously reported 2p16.3 region, located within *FSHR*. This study used the Affymetrix Axiom array and combined the data from genotyped and imputed SNPs from the Chen et al. [26] study to maximize statistical power. The study analyzed 2254 PCOS cases defined by the Rotterdam criteria and 3001 controls in the discovery set meta-analysis and replicated the findings in two independent sets of 1908/6318 cases and 1913/5665 controls. The previously identified regions 2p16.3, 2p21, and 9q33.3 again reached genome-wide significance and a further 19 new regions were identified in the discovery meta-analysis stage. The 19 new regions and variants in *FSHR* were analyzed in the replication set, of which eight remained significant (2p.16.3/*FSHR*, 9q22.32, 11q22.1, 12q13.2, 12q14.3, 16q12.1, 19p13.3, and 20q13.2), the majority containing candidate genes related to hormones, insulin resistance, and organ growth.

The most significant SNP at 9q22.32 was located in an intron of *C9orf3* that encodes for a protein within the zinc aminopeptidase family that had previously been associated with erectile dysfunction following radiotherapy for prostate cancer, alongside *FSHR* variants that have also been strongly associated with erectile dysfunction. A SNP within *YAP1* at 11q22.1 (rs1894116) was posited to have a role in altering the expression of genes associated with cell proliferation and organ size control, although the function within the ovary is not defined. The most significant SNP at 12q13.2 was located between *RAB5B* and *SUOX*, which had previously been associated with type 1 diabetes mellitus (T1DM). Another SNP in this region (rs2292239) located within the *ERBB3* gene showed evidence of association with PCOS, and identified a family of genes that have been identified in subsequent studies. The nearby region of 12q14.3 had an independent signal from a SNP within *HMGA2*, which encodes a protein involved in DNA transcription regulation and had previously been associated with T2DM.

The 19p13.3 region identified as associated with PCOS was significant for containing the insulin receptor gene (*INSR*), with a signal from an intronic SNP showing genome-wide significance (rs2059807). This ties in strongly with the common understanding of one of the primary pathogenetic mechanisms of PCOS being insulin resistance, as mutations in the *INSR* gene have previously been shown to be associated with severe hyperinsulinemia and insulin resistance. In addition, variants within the *FSHR* gene at 2p16.3 that had previously failed to meet genome-wide significance in the Chen et al. [26] study were identified as significant variants, independent from the previously identified signals at *LHCGR*. The addition of these regions to the GWAS loci known to be associated with PCOS added further weight to the most common explanations regarding the pathogenesis of PCOS. However, the strength of GWAS studies is in identifying regions that would not have been identified based on a hypothesis-driven approach. Shi et al. [28] added other candidate genes at 16q12.1 (*TOX3*) and 20q13.2 (located between *SUMO1P1* and *ZNF217*) that do not have roles in growth and metabolism at the endocrine level, but are related to cell growth and DNA modification.

A GWAS involving Korean women with PCOS by Rotterdam or NIH criteria subsequently identified a novel gene associated with PCOS in combination with obesity [29]. In this study, the PCOS cohort consisted of 774 patients with 967 controls and genotyping was performed using the Illumina HumanOmni1 Quad v1. In total, 619,339 SNPs were analyzed for association with PCOS. Three PCOS-associated SNPs were identified at 12p12.2 (rs10841843, rs6487237, rs7485509) that were related to *GYS2*, a glycogen synthase gene with a role in glycogen storage disease 0 (OMIM 138571). The variants in this gene were subjected to further analysis in relation to the BMI of the subjects, identifying an additional 14 significant variants. However, the initial three SNPs remained significant after adjustment for BMI, suggesting an independent risk for PCOS rather than being driven by an association with BMI. The authors then proceeded to analyze a cohort of childhood obesity (482 children) and gestational diabetes mellitus (1710 women), suggesting that these variants had a pleiotropic effect on obesity-related conditions across the lifespan. The authors analyzed seven of the SNPs reported by Chen et al. [26] at 2p16.3, 2p21, and 9q33.3 and concluded that six (with the exception of rs2479106 in *DENND1A*) had a *P*-value between 2×10^{-2} and 8×10^{-4} that failed to reach genome-wide significance. The authors also analyzed variants in *FSHR*, concluding that there was an association in this cohort of Korean women (*P*-value between 2.2×10^{-3} and 5.9×10^{-4}).

Lee et al. [30] published a further GWAS study in a population of Korean women with PCOS, according to the Rotterdam criteria. The discovery cohort consisted of 1000 PCOS cases and 1000 controls, with a replication study of 249 cases and 778 controls. Genotyping was performed using the HumanOmni1-Quad v1 array with 636,870 SNPs analyzed for association with PCOS. The discovery stage revealed 56 SNPs over 24 regions, of which 21 were analyzed in the replication stage using TaqMan technology. Twelve SNPs demonstrated the same direction of effect in the discovery and replication phases, of which one remained significant in decreasing the risk of PCOS for those carrying the variant rs10505648 at 8q24.2. This signal was located 487 kb upstream of *KHDRBS3*, which may be related telomerase activity, a pathogenetic mechanism explored in PCOS by Li et al. [31]. Several moderate associations with PCOS were shown in other loci. However, the small size of the replication cohort may have limited the ability of the study to detect an association. The authors were only able to analyze 10 of the 11 PCOS loci identified in the Chen et al. [26] and Shi et al. [28] studies due to a different genotyping method, and found that seven were associated with a consistent direction of effect at statistical significance. A follow-up pathway analysis utilizing the same dataset also identified variants in oocyte meiosis as the top-ranking pathway associated with PCOS [32].

Day et al. [33] have reported the largest GWAS study to date in PCOS with analysis of 5184 self-reported cases of PCOS with European ancestry and 82,759 controls and a replication set of clinically validated cases (1875 according to Rotterdam criteria and 861 according to NIH criteria) and 181,645 controls. Six independent signals were identified at genome-wide significance, of which four were novel and two were located within previously reported genes (*YAP1* and *THADA*). All six were independently associated with PCOS risk and were not associated with BMI, suggesting an impact on PCOS-specific pathways. The strongest novel signal was an intronic variant in *ERBB4/HER4*, with signals in *ERBB3/HER3* and *ERRB2/HER2* also nearing genome-wide significance. All those are members of the epidermal growth factor receptor family and may have a role in mediating LH-induced steroidogenesis in the ovary. Another variant rs11031006 near *FSHB* (encoding the FSH-specific beta subunit) reached genome-wide significance, adding further evidence to the importance of the FSH hormone or receptor gene variants in modulating the process of follicular maturation. The *FSHR* variant reported in the Shi et al. [28] study was detected but only weakly associated with PCOS in this population of Caucasian European women.

Other novel signals were near *RAD50* (a dsDNA break repair gene) and *KRR1* (a ribosome assembly factor) [33]. Similar to the previous Korean GWAS studies [29, 30], this European study found associations between 10 of the 11 reported variants in the Han Chinese population [28], of which six were significant but not at the genome-wide level. Two genes previously reported (*YAP1* and *THADA*) in the Han Chinese cohorts were identified as significant loci but with novel variants, and a further variant in *DENND1A* was not confirmed in the replication phase of the study. The authors also report an association between PCOS risk and delayed menopause in this cohort, suggesting a possible evolutionary mechanism for the persistence of PCOS risk alleles in the population.

Hayes et al. [34] undertook a discovery GWAS in 984 PCOS cases defined by NIH criteria and 2694 controls, followed by two replication studies in 1799/217 cases and 1231/1335 controls. Three loci associated with PCOS at genome-wide significance were identified: 8p32.1,

11p14.1, and 9q22.32. Of these, 8p32.1 was entirely novel and was near the *GATA4/NEIL2* genes, which have a role in steroidogenesis and the repair of DNA damage, respectively, as well as *FDFT1*, which is involved in the cholesterol-biosynthesis pathway. The 9q22.32 locus included the previously reported C9orf3 gene, with the most strongly associated SNP in this study (rs10993397) being independent of the association with rs3802457 reported in Shi et al. [28]. The 11p14.1 locus identified the rs11031006 variant in the *FSHB* gene reported in Day et al. [33] and in a quantitative trait analysis, demonstrated that this variant was strongly associated with LH levels. Again, seven of the 11 loci identified in the Shi et al. [28] study were confirmed, although the 8q24.2 locus identified by Lee et al. [30] was not found to be significant.

Overall, the contribution of the GWAS methodology to the understanding of PCOS has been moderate (Table 20.1). The recurrence of variants in the regions of gonadotropin or gonadotropin receptor genes across multiple studies confirms the importance of the hypothalamic-pituitary-ovarian axis in the development of PCOS. In addition, multiple regions with genes related to metabolism and cellular proliferation have been identified, in keeping with the hypotheses regarding insulin resistance, obesity, and the development of PCOS. Novel variants in genes such as *EGFR*, *DENND1A*, *THADA*, and C9orf3 have also appeared in diverse populations, suggesting a potential role for hitherto unexplored pathways and the possibility of new pharmacological interventions for PCOS. However, it must be remembered that the overall contribution of each of these variants and their effect on PCOS risk is very small, accounting for <10% of the overall ~80% heritability of the condition [35]. In addition, the variants identified are only effective in highlighting genomic regions of interest, and do not in themselves represent pathological variants. However, the confirmation of multiple loci across studies in different populations and using different diagnostic criteria has been a positive impact of the GWAS study era [13].

TABLE 20.1 Pathways and Genes Identified in PCOS GWAS Studies

Gene Family	Genes	Evidence From GWAS Studies
Gonadotropin function	*FSHR, FSHB, LHCGR*	+++
Insulin function, obesity	*INSR, THADA, DENND1A, RAB5B/SUOX*	+++
Steroidogenesis	*GATA4, FDFT1*	++
Epidermal growth factor receptors	*ERRB3/HER3, ERRB4/HER4, ERRB2/HER2*	++
Cell growth, DNA repair, telomerase activity	*C9orf3, YAP1, NEIL2, RAD50, KRR1, KHDRBS3, TOX3, SUMO1P1, ZNF217, HMGA2*	++
Reproductive tract development	*GTF2A1L*	+
Glycogen storage	*GYS2*	+

POST-GWAS STUDIES

The GWAS studies detailed above have provided a significant resource for researchers attempting to understand the pathogenesis of PCOS as well as those attempting to find or replicate population-specific variants. Numerous studies using the candidate gene approach have attempted to confirm the findings of the GWAS studies above in different populations. Goodarzi et al. [36] investigated the three loci from the original Chen et al. [26] GWAS in a population of European-derived PCOS patients defined by NIH criteria (939/535 cases and 957/845 controls across two cohorts). Variants in *DENND1A* and *THADA* were associated with PCOS; however, there was no evidence for association of the *LHCGR* variation. An Icelandic study replicated the 9q33.3 signals found in the Chen et al. [26] study, and was associated with the risk of hyperandrogenism in women without PCOS [37].

Studies involving the gonadotropin polypeptides and receptors have similarly confirmed involvement across multiple populations. Capalbo et al. [38] studied a Sardinian population with 198 PCOS patients and 187 controls, demonstrating a 2.0/2.7-fold risk with the heterozygous/homozygous S312N variant at the *LHCGR* locus. Almawi et al. [39] genotyped *FSHR* and *LHCGR* in 203 women with PCOS and 211 controls in a Bahraini population, demonstrating novel SNPs in both genes (rs7371084, rs4953616, and rs11692782) associated with PCOS risk. Ha et al. [40] demonstrated that variants in the *LHCGR* gene were associated with PCOS in Hui Chinese women, a separate ethnic group to the original Han Chinese study. However, numbers were small (151 cases and 99 controls) and the variants in *DENND1A* and *THADA* that had been identified in Chen et al. [26] were not found to be associated with PCOS in this population. Another study examined the novel loci identified in the European populations [33, 34] in a case-control cohort of Han Chinese, with 1500 PCOS cases and 1220 controls [41]. Marker SNPs in *ERBB4* again reached significance in the Chinese population, including after adjustment for BMI.

In addition, functional studies have added to our understanding of these risk loci and their differential expression in patients with and without PCOS. Jones et al. [42] studied methylation and mRNA expression in the regions surrounding the 11 risk loci identified in the Shi et al. [28] study, generating functional maps of adipose tissue gene expression in subjects with and without PCOS. Their finding of *LHCGR* overexpression in the adipose tissue of subjects with PCOS correlated with previous studies of granulosa and theca cell expression [43] and again suggests that variants in *LHCGR* identified in GWAS studies may have a key role in the pathogenesis of PCOS. Subjects with PCOS had lower *INSR* expression in adipose tissue, although cumulus cells of obese PCOS subjects had overexpressed *INSR*, suggesting that the sensitivity of ovarian tissue to insulin may differ from that in the periphery, allowing continuation of ovarian steroidogenesis.

Another study examining gene expression and DNA methylation in adipose tissue demonstrated significant expression differences in multiple genes, including those previously linked to PCOS (*RAB5B*) and those associated with insulin resistance, adipocyte size, and hyperandrogenism. However, this study did not find differences in *LHCGR* or *INSR* expression [44]. A study using differential in-gel electrophoresis analysis and mass spectroscopy of ovarian tissue in women with PCOS showed 18 differentially expressed proteins, including progesterone receptor component 1 and retinol-binding protein 1 (PGRMC1 and RBP1), which may serve as potential biomarkers to aid in identification of cases [45].

A functional study involving RNA sequencing of adipose tissue in subjects with PCOS before and after treatment with metformin demonstrated that one of the variants identified in the Han Chinese population (*THADA*, rs12478601) was related to a greater response to metformin [46]. The study also confirmed the relationship between variants near the *FSHB* gene and LH levels and the LH:FSH ratio. The findings suggest that susceptibility loci may not only have the ability to predict the likelihood of developing PCOS, but may also guide treatment based on pharmacogenomics.

Epigenetic studies have also attempted to explain features of PCOS and the association with metabolic syndrome. Zhao et al. [47] studied *PPARGC1A* promoter methylation and mitochondrial DNA (mtDNA) content in peripheral blood leukocytes of women with PCOS and concluded that there was a significant association with increased *PPARGC1A* promoter methylation and decreased mtDNA content with increasing metabolic risk. The contribution of dynamic changes in genomic expression to the phenotype of PCOS requires significant further exploration.

GENOMIC APPROACHES

The emergence of genomic technologies capable of whole exome and whole genome sequencing greatly increases the potential for the identification of pathogenic variants associated with PCOS. This allows the analysis of rare sequence variants rather than identification of associated variants (which by inclusion on a SNP array are by definition common population variants). A recent study by Gorsic et al. [48] used whole genome sequencing in 80 patients with PCOS according to NIH criteria and compared to 24 controls without PCOS. Three rare, putative functional coding variants were identified in the *AMH* gene in five of the women with PCOS, and were not identified in controls. Targeted resequencing of a larger PCOS cohort (643 cases and 153 controls) identified 21 additional rare coding variants in *AMH*, 18 of which were found only in patients with PCOS. Of these, 17 of the variants were shown to decrease AMH activity in a functional assay. The variants that were found in the control group did not have a functional impact. The findings also contrasted with previous metaanalyses of common variants in the AMH gene [49], suggesting that rare but strongly deleterious mutations may have a role in the pathogenesis of PCOS, demonstrating the potential for sequencing technologies in future gene discovery.

Another aspect of genome-wide analysis in future clinical practice will be the prospective identification of individuals at risk of PCOS based on known risk loci. A study by Lee et al. [50] analyzed a cohort of 862 PCOS and 860 control patients and assigned a genetic risk score based on the 11 susceptibility loci identified by Shi et al. [28]. The risk score was calculated by the addition of the number of susceptibility loci and was significantly higher on average in women with PCOS rather than controls. The odds ratio of having PCOS in the highest quartile of risk scores (>12) was 6.28 ($P < 0.001$), suggesting that the current loci had substantial predictive power. The genetic risk score for menopause may also be a surrogate marker for PCOS, with variants predicting late menopause also having an association with PCOS [51].

Cui et al. [51] analyzed the relative contribution of each of the SNPs identified in the Han Chinese population to the individual clinical features of PCOS, rather than PCOS as a

syndrome. Individuals with oligo-anovulation only (746), hyperandrogenism only (278), and polycystic ovarian morphology only (536) were compared to 1790 healthy controls. Individual SNPs were identified to have a unique relationship with each phenotype, with the exception of rs4385527 in *C9orf3*, which had a relationship with all three phenotypes. Variants in *LHCGR* and *INSR* were associated with oligo-anovulation and variants in *THADA and DENND1A* with polycystic ovarian morphology. The identification of individuals at risk of PCOS using genotypic data may allow early lifestyle intervention and is a key goal of the genomic era, alongside understanding the underlying pathogenesis of the condition and the development of new therapeutic agents.

CONCLUSION

In many ways, PCOS represents an evolutionary puzzle: why is it such a common disorder, if the ultimate effect is a reduction in reproductive fitness? Many potential reasons for the prevalence of PCOS have been identified, including the beneficial impact of hyperandrogenemia in improving muscle strength and fitness in women, the potential for insulin resistance in previous fasting periods, the lengthened period of fertility associated with delayed menopause, and the reduced exposure to pregnancy events that likely had a high mortality in the evolutionary past [52]. It seems likely that the variants affecting our susceptibility to PCOS arose in prehistoric times, prior to the migration out of Africa [53].

Ultimately, it is the complexity of mammalian female reproductive endocrinology that predisposes one to a risk of suboptimal functioning across multiple pathways, and it is likely that individual susceptibility to PCOS will have contributions from multiple genes [54]. Although the role of key pathways such as insulin resistance and gonadotropin secretion and receptor function have been confirmed by genetic studies, there are a number of pathways with clear signals but that need further functional studies to understand their role in pathogenesis. In addition, the overall heritability of PCOS remains unexplained by current studies, with analysis of the 11 loci predicted in the Han Chinese population estimated to explain 2.4% of the variance in risk of PCOS [55]. The application of whole exome/genome sequencing technologies has the potential to identify informative variants at the level of both the individual and the population, promising significant advances in our understanding of PCOS.

References

[1] Azziz R, Carmina E, Chen Z, Dunaif A, Laven JSE, Legro RS, et al. Polycystic ovary syndrome. Nat Rev: Disease Primers 2016;2:16057.
[2] Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. Fertil Steril 2004;81(1):19–25.
[3] Cooper HE, Spellacy WN, Prem KA, Cohen WD. Hereditary factors in the Stein-Leventhal syndrome. Am J Obstet Gynaecol 1968;100(3):371–87.
[4] Amato P, Simpson JL. The genetics of polycystic ovary syndrome. Best Pract Res Clin Obstet Gynaecol 2004;18(5):707–18.
[5] De Leo V, Musacchio MC, Cappelli V, Massaro MG, Morgante G, Petraglia F. Genetic, hormonal and metabolic aspects of PCOS: an update. Reprod Biol Endocrinol 2016;14:38.

[6] Franks S, Gharani N, Waterworth D, Batty S, White D, Williamson R, et al. The genetic basis of polycystic ovary syndrome. Hum Reprod 1997;12(12):2641–8.
[7] Norman RJ, Masters S, Hague W. Hyperinsulinemia is common in family members of women with polycystic ovary syndrome. Fertil Steril 1996;66(6):942–7.
[8] Legro RS, Driscoll D, Strauss JF, Fox J, Dunaif A. Evidence for a genetic basis for hyperandrogenemia in polycystic ovary syndrome. Proc Natl Acad Sci U S A 1998;95:14956–60.
[9] Jahanfar S, Seppala M, Eden JA, Nguyen TV, Warren P. A twin study of polycystic ovary syndrome. Fertil Steril 1995;63(3):478–86.
[10] 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.
[11] Dumesic DA, Oberfield SE, Stener-Victorin E, Marshall JC, Laven JS, Legro RS. Scientific statement on the diagnostic criteria, epidemiology, pathophysiology, and molecular genetics of polycystic ovary syndrome. Endocr Rev 2015;36(5):487–525.
[12] Joseph S, Barai RS, Bhujbalrao R, Idicula-Thomas S. PCOSKB: a KnowledgeBase on genes, diseases, ontology terms and biochemical pathways associated with polycystic ovary syndrome. Nucleic Acids Res 2016;44:D1032–5.
[13] Jones MR, Goodarzi MO. Genetic determinants of polycystic ovary syndrome: progress and future directions. Fertil Steril 2016;106(1):25–32.
[14] Tucci S, Futterweit W, Conception 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.
[15] Urbanek M, Woodroff A, Ewens KG, Diamanti-Kandarakis E, Legro RS, Strauss JF, et al. Candidate gene region for polycystic ovary syndrome on chromosome 19p13.2. J Clin Endocrinol Metab 2005;90(12):6623–9.
[16] Ewens KG, Stewart DR, Ankener W, Urbanek M, McAllister JM, Chen C, et al. Family-based analysis of candidate genes for polycystic ovary syndrome. J Clin Endocrinol Metab 2010;95:2306–15.
[17] Xie G, Xu P, Che Y, Xia Y, Cao Y, Wang W, et al. Microsatellite polymorphism in the fibrillin 3 gene and susceptibility to PCOS: a case-control study and meta-analysis. Reprod Biomed Online 2013;26:168–74.
[18] Villuendas G, Escobar-Morreale HF, Tosi F, Sancho J, Moghetti P, San Millan JL. Association between the D19S884 marker at the insulin receptor gene locus and polycystic ovary syndrome. Fertil Steril 2003;79(1):219–20.
[19] Liu AL, Xie HJ, Xie HY, Liu J, Yin J, Hu JS, et al. Association between fat mass and obesity associated (FTO) gene rs9939609 A/T polymorphism and polycystic ovary syndrome: a systematic review and meta-analysis. BMC Med Genet 2017;18:89.
[20] Ketefian A, Jones MR, Krauss RM, Chen YI, Legro RS, Azziz R, et al. Association study of androgen signaling pathway genes in polycystic ovary syndrome. Fertil Steril 2016;105(2):467–73.
[21] Hickey T, Chandy A, Norman RJ. The androgen receptor CAG repeat polymorphism and X-chromosome inactivation in Australian Caucasian women with infertility related to polycystic ovary syndrome. J Clin Endocrinol Metab 2002;87(1):161–5.
[22] Peng CY, Xie HJ, Guo ZF, Nie YL, Chen J, Zhou JM, et al. The association between androgen receptor gene CAG polymorphism and polycystic ovary syndrome: a case-control study and meta-analysis. J Assist Reprod Genet 2014;31:1211–9.
[23] Wang F, Pan J, Liu Y, Meng Q, Lv P, Qu F, et al. Alternative splicing of the androgen receptor in polycystic ovary syndrome. Proc Natl Acad Sci U S A 2015;112(15):4743–8.
[24] Urbanek M, Legro RS, Driscoll D, 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;96:8573–8.
[25] Guo R, Zheng Y, Yang J, Zheng N. Association of TNF-alpha, IL-6 and Il-1beta gene polymorphisms with polycystic ovary syndrome: a meta-analysis. BMC Genet 2015;16:5.
[26] Chen ZJ, Zhao H, He L, Shi Y, Qin Y, Shi Y, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. Nat Genet 2011;43(1):55–9.
[27] Olszanecka-Glinianowicz M, Banas M, Zahorska-Markiewicz B, Janowska J, Kocelak P, Madej P, et al. Is the polycystic ovary syndrome associated with chronic inflammation *per se*? Eur J Obstet Gynecol Reprod Biol 2007;133:197–202.
[28] 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.

[29] Hwang JY, Lee EJ, Go MJ, Sung YA, Lee HJ, Kwak SH, 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;57:660–4.
[30] 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.
[31] Li Q, Du J, Feng R, Xu Y, Wang H, Sang Q, et al. A possible new mechanism in the pathophysiology of polycystic ovary syndrome (PCOS): the discovery that leukocyte telomere length is strongly associated with PCOS. J Clin Endocrinol Metab 2014;99(2):E234–40.
[32] Shim U, Kim HN, Lee H, Oh JY, Sung YA, Kim HL. Pathway analysis based on a genome-wide association study of polycystic ovary syndrome. PLoS ONE 2015;10(8):e0136609.
[33] Day FR, Hinds DA, Tung JY, Stolk L, Styrkarsdottir U, Saxena R, et al. Causal mechanisms and balancing selection inferred from genetic association with polycystic ovary syndrome. Nat Commun 2015;6:8464.
[34] Hayes MG, Urbanek M, Ehrmann DA, Armstrong LL, Lee JY, Sisk R, et al. Genome-wide association of polycystic ovarian syndrome implicates alterations in gonadotropin secretion in European ancestry populations. Nat Commun 2015;6:7502.
[35] Dunaif A. Perspectives in polycystic ovary syndrome: From hair to eternity. J Clin Endocrinol Metab 2016;101(3):759–68.
[36] Goodarzi MO, Jones MR, Li X, Chua AK, Garcia O, Chen YI, 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.
[37] Welt CK, Styrkarsdottir U, Ehrmann DA, Thorleifsson G, Arason G, Gudmundsson JA, 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.
[38] Capalbo A, Sagnella F, Apa R, Fulghesu AM, Lanzones A, Morciano A, et al. The 312N variant of the luteinizing hormone/choriogonadotropin receptor gene (*LHCGR*) confers up to 2.7-fold increased risk of polycystic ovary syndrome in a Sardinian population. Clin Endocrinol (Oxf) 2012;77:113–9.
[39] Almawi WY, Hubail B, Arekat DZ, Al-Farsi SM, Al-Kindi SK, Arekat MR, et al. Luteinizing hormone/choriogonadotropin receptor and follicle stimulating hormone receptor gene variants in polycystic ovary syndrome. J Assist Reprod Genet 2015;32:607–14.
[40] Ha L, Shi Y, Zhao J, Li T, Chen ZJ. Association study between polycystic ovarian syndrome and the susceptibility genes polymorphisms in hui Chinese women. PLoS ONE 2015;10(5)e0126505.
[41] Peng Y, Zhang W, Yang P, Tian Y, Su S, Zhang C, et al. *ERBB4* confers risk for polycystic ovary syndrome in Han Chinese. Sci Rep 2017;7:42000.
[42] Jones MR, Brower MA, Xu N, Cui J, Mengesha E, Chen YI, et al. Systems genetics reveals the functional context of PCOS loci and identifies genetic and molecular mechanisms of disease heterogeneity. PLoS Genet 2015;11(8)e1005455.
[43] Wang P, Zhao H, Li T, Zhang W, Wu K, Li M, et al. Hypomethylation of the LH/choriogonadotropin receptor promoter region is a potential mechanism underlying susceptibility to polycystic ovary syndrome. Endocrinology 2014;155:1445–52.
[44] Kokosar M, Benrick A, Perfilyev A, Fornes R, Nilsson E, Maliqueo M, et al. Epigenetic and transcriptional alterations in human adipose tissue of polycystic ovary syndrome. Sci Rep 2016;6:22883.
[45] Li L, Zhang J, Deng Q, Li J, Li Z, Xiao Y, et al. Proteomic profiling for identification of novel biomarkers differentially expressed in human ovaries from polycystic ovary syndrome patients. PLoS ONE 2016;11(11):e0164538.
[46] Pau CT, Mosbruger T, Saxena R, Welt CK. Phenotype and tissue expression as a function of genetic risk in polycystic ovary syndrome. PLoS ONE 2017;12(1):e0168870.
[47] Zhao H, Zhao Y, Ren Y, Li M, Li T, Li R, et al. Epigenetic regulation of an adverse metabolic phenotype in polycystic ovary syndrome: the impact of the leukocyte methylation of *PPARGC1A* promoter. Fertil Steril 2017;107:467–74.
[48] Gorsic LK, Kosova G, Werstein B, Sisk R, Legro RS, Hayes MG, et al. Pathogenic anti-Mullerian hormone variants in polycystic ovary syndrome. J Clin Endocrinol Metab 2017;102(8):2862–72.
[49] Pabalan N, Montagna E, Singian E, Tabangay L, Jarjanazi H, Barbosa CP, et al. Associations of polymorphisms in anti-mullerian hormone (*AMH* Ile49Ser) and its type II receptor *AMHRII* -482 A>G on reproductive outcomes and polycystic ovary syndrome: a systematic review and meta-analysis. Cell Physiol Biochem 2016;39:2249–61.
[50] Lee H, Oh JY, Sung YA, Chung HW. A genetic risk score is associated with polycystic ovary syndrome-related traits. Hum Reprod 2016;31(1):209–15.

[51] Cui L, Li G, Zhong W, Bian Y, Su S, Sheng Y, et al. Polycystic ovary syndrome susceptibility single nucleotide polymorphisms in women with a single PCOS clinical feature. Hum Reprod 2015;30(3):732–6.
[52] Fessler DM, Natterson-Horowitz B, Azziz R. Evolutionary determinants of polycystic ovary syndrome: part 2. Fertil Steril 2016;106:42–7.
[53] Casarini L, Brigante G. The polycystic ovary syndrome evolutionary paradox: a genome-wide association studies-based, in silico evolutionary explanation. J Clin Endocrinol Metab 2014;99(11):E2412–20.
[54] Joshi N, Chan JL. Female genomics: infertility and overall health. Semin Reprod Med 2017;35:217–24.
[55] Xu Y, Li Z, Ai F, Chen J, Xing Q, Zhou P, et al. Systematic evaluation of genetic variants for polycystic ovary syndrome in a Chinese population. PLoS ONE 2015;10(10)e0140695.

Further Reading

Saxena R, Bjonnes AC, Georgopoulos NA, Koika V, Panidis D, Welt CK. Gene variants associated with age at menopause are also associated with polycystic ovary syndrome, gonadotrophins and ovarian volume. Hum Reprod 2015;30(7):1697–703.

Glossary

Oligo/anovulation	Failure to achieve ovulation on a regular basis/failure to achieve ovulation; manifests as menstrual disturbance and subfertility.
Hyperandrogenism	Elevated levels of androgens measured either by clinical means (hirsutism, acne) or biochemical tests (increased testosterone, elevated free androgen index).
Polycystic ovarian morphology	Appearance of polycystic ovaries on transvaginal ultrasound, various definitions including increased number of peripheral small follicles and increased ovarian volume overall.