An update on the genetics of polycystic ovary syndrome: progress and future directions

Michelle R. Jones, Ph.D.[a,b] and Mark O. Goodarzi, M.D., Ph.D.[a]

[a] Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine and [b] Bioinformatics and Computational Biology Research Center, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California

The field of the genetics of polycystic ovary syndrome (PCOS) has relatively recently moved into the era of genome-wide association studies. This has led to the discovery of 16 robust loci for PCOS. Some loci contain genes with clear roles in reproductive (*LHCGR*, *FSHR*, and *FSHB*) and metabolic (*INSR* and *HMGA2*) dysfunction in the syndrome. The next challenge facing the field is the identification of causal variants and genes and the role they play in PCOS pathophysiology. The potential for gene discovery to improve diagnosis and treatment of PCOS is promising, though there is much to be done in the field before the current findings can be translated to the clinic. (Fertil Steril® 2016;■:■–■. ©2016 by American Society for Reproductive Medicine.)

Key Words: PCOS, SNP, GWAS, linkage, association study

Discuss: You can discuss this article with its authors and with other ASRM members at http://fertstertforum.com/jonesm-genetics-pcos/

Use your smartphone to scan this QR code and connect to the discussion forum for this article now.*

* Download a free QR code scanner by searching for "QR scanner" in your smartphone's app store or app marketplace.

THE GENETICS OF PCOS

Polycystic ovary syndrome (PCOS) is a heterogeneous reproductive and metabolic endocrine disorder. The primary features of PCOS are hyperandrogenism and oligo- or anovulation. Elevated circulating testosterone levels manifest clinically as hirsutism and acne, and dysfunction in gonadotropin pathways contribute to oligo-ovulation. Polycystic ovaries featuring numerous immature antral follicles can be seen with the use of ultrasound (1). Two sets of diagnostic criteria developed by the National Institutes of Health (NIH)/National Institute of Child Health and Human Development (2) and European Society for Human Reproduction and Embryology/American Society for Reproductive Medicine (known as the Rotterdam criteria) (3) have been used and created some controversy in the field (4). The NIH criteria requires both oligo- or anovulation and either clinical or biochemical signs of hyperandrogenism, whereas the Rotterdam criteria also considers polycystic ovaries seen with the use of ultrasound and requires two of three signs (hyperandrogenism, oligo-ovulation, polycystic ovarian morphology). The multifactorial etiology of PCOS is underpinned by a complex genetic architecture that has only recently begun to be elucidated. The use of candidate gene analysis has provided several promising genes as PCOS risk loci or as genetic modifiers of component phenotypes of PCOS. Genome-wide association studies (GWAS) in very large case-control cohorts, made possible through the development of high-throughput genotyping methods, have proved to be successful in PCOS, as in other common complex diseases. The growing list of PCOS susceptibility genes contributes to our understanding of pathways and processes implicated in the syndrome's etiology and have revealed relatively homogenous genetic underpinnings of PCOS.

Received March 11, 2016; revised April 21, 2016; accepted April 26, 2016.
M.R.J. has nothing to disclose. M.O.G. has nothing to disclose.
Reprint requests: Michelle R. Jones, Ph.D., Division of Endocrinology, Diabetes, and Metabolism, Cedars-Sinai Medical Center, 8700 Beverly Blvd., Room B-131, Los Angeles, California 90048 (E-mail: michelle.jones@csmc.edu).

Fertility and Sterility® Vol. ■, No. ■, ■ 2016 0015-0282/$36.00

http://dx.doi.org/10.1016/j.fertnstert.2016.04.040

ESTABLISHING THE HERITABLE BASIS OF PCOS

Familial Aggregation Studies

The first evidence for the genetic basis of PCOS was reported in 1968 by Cooper et al. (5). Early studies reported increased prevalence of PCOS-related traits in the siblings of PCOS probands, with many indications for an autosomal dominant model of inheritance (6, 7). Several small family studies supported this hypothesis, reporting prevalence of PCOS ranging from 51% to 66% in first-degree relatives of probands (8, 9). The rate of polycystic ovarian morphology or male-pattern baldness in first-degree relatives of PCOS subjects was reported to be 51%, and it was subsequently suggested that a single gene might be responsible for oligomenorrhea

and hirsutism in PCOS women and male-pattern baldness in male family members of PCOS subjects (9). A large study of 250 consecutive PCOS probands and their families found that 75% of probands reported at least one family member with either hirsutism alone or PCOS including hirsutism (10). A strong familial aggregation for PCOS and its component traits was estimated in this study, with 50% of mothers or sisters, 25% of aunts, and 20% of grandmothers having hirsutism or hirsutism with oligomennorhea (10). This finding was supported by subsequent reports that established prevalences of PCOS in premenopausal mothers and sisters of probands of 35% and 40%, respectively (11). A study of hyperandrogenemia in sisters of PCOS probands found that although 22% of sisters had PCOS themselves, another 24% of sisters had hyperandrogenemia with regular menstrual cycles; the bimodal distribution of testosterone levels in the sisters led the authors to propose control by a single autosomal locus (12). The presence of a single-component phenotype in the absence of a diagnosis of PCOS suggested that phenotypic heterogeneity between patients may have a genetic basis, demonstrated by the fact that sisters of PCOS probands had inherited genetic risk factors for some but not all traits of PCOS, resulting in a partial phenotype.

Twin Studies

Examination of the rate of PCOS in a small cohort of both monozygotic and dizygotic twin pairs initially suggested that PCOS was unlikely to be an autosomal dominant disorder, owing to the high rate of discordance in phenotype, but rather a polygenic or X-linked disorder (13, 14). A subsequent very large twin study of more than 3,100 Dutch twins identified 92 subjects with PCOS with the use of a questionnaire (prevalence of 2.97%) (15), a lower prevalence than typically reported in population-based studies. This large twin study reported a monozygotic correlation for PCOS (r^2) of 0.72, and a dizygotic correlation for PCOS of 0.38. With the use of this twin approach, the authors were able to estimate that 72% of variance in risk of PCOS has a genetic basis (narrow-sense heritability calculated with the use of an additive genetic model), demonstrating that there is a significant genetic component to the disease (15). Androgen exposure in opposite-sex twin pairs has been proposed as a model for prenatal programming; however, increased risk to the female twin has not been consistently demonstrated, suggesting that the genetic risk for PCOS outweighs the environmental risk from prenatal androgen exposure in opposite-sex twin pairs (16).

IDENTIFYING PCOS RISK GENES

The Candidate Era

The genetics of PCOS was widely pursued by means of the candidate gene approach, which focuses on a gene of interest selected by its hypothesized role in the disease under study. This approach has been generally unsuccessful in complex disease, in part owing to incomplete understanding of disease pathophysiology and the inclusion of a single or very few markers in a gene. Another challenge of gene discovery in complex disease is inherent to the genetic architecture of multigenic diseases. Many variants each contribute a small amount of susceptibility risk to the disease, meaning that thousands or tens of thousands of samples need to be included to detect the effect of individual single-nucleotide polymorphisms (SNPs). Sample sizes of this magnitude have not been available in PCOS studies until very recently. More than 100 candidate genes have been examined in PCOS, but only a single candidate gene, the insulin receptor (*INSR*), went on to be validated as a risk locus in large well designed GWAS (17). Initial studies of the *INSR* locus described both linkage and association between the dinucleotide repeat microsatellite marker D19S884 in intron 55 of the fibrillin-3 gene (18), 1.2 cM from *INSR*, the intended candidate gene (18). Replication of association between D19S884 and PCOS in independent cohorts was subsequently reported by the original authors, with larger sample sizes gleaning increasingly significant *P* values (19–21). The function of this intronic microsatellite marker is unknown; however, promoter activity has been detected at the sequence encompassing D19S884 (20). The expression level of the *FBN3* gene itself has been reported to be extremely low in a number of ovarian cell types (22), leading some to conclude that the D19S884 association signal is likely a proxy signal for a causal variants in other genes elsewhere in the region (22). *FBN3* is expressed in the pituitary; however, its role there remains to be explored. After the initial reports of both linkage and association at D19S884, the *INSR* gene became a target for additional studies. Seven studies were able to find associations between SNPs in the *INSR* gene and PCOS risk, many of which are novel exonic SNPs (23–31). There have also been three reports failing to find association between this gene and PCOS risk; however, those studies were hampered by small sample sizes (31, 32) or included only a single marker (27). To date, the most extensive investigation of the *INSR* gene in PCOS interrogated the entire gene with a tagging approach (29). In the discovery cohort, five out of 30 SNPs across the *INSR* gene were associated with PCOS, four of which were carried forward to a replication study where association between rs2252673 and PCOS was successfully confirmed. This variant is located in intron 11 of the *INSR* gene, so it is unclear how this SNP might affect *INSR* expression or function to influence the PCOS phenotype. As with D19S884, rs2252673 may be a proxy marker for other functional causal variants in the region. The most important validation of the *INSR* locus as a PCOS risk locus is the discovery of an association signal at the *INSR* gene in a very large well designed case-control GWAS (17).

The role of body mass index (BMI)–associated variants has been widely studied in PCOS. Initial reports looking at the role of BMI loci in PCOS found little evidence to indicate shared risk between BMI and PCOS at these variants in early studies (33, 34). In contrast, the most recent and comprehensive analysis (in terms of number of BMI loci examined) found a positive relationship between effect size on BMI and odds ratio of PCOS (35). Association studies of the obesity locus known as *FTO* (fat mass–and obesity-associated) (36) with BMI in PCOS have yielded mixed results, and a meta-analysis found that *FTO* is not a PCOS susceptibility locus despite its increased effect on BMI in PCOS women compared with the general population (37). It has recently