

Deconstructing a Syndrome: Genomic Insights Into PCOS Causal Mechanisms and Classification

Matthew Dapas¹  and Andrea Dunaif² 

¹Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA

²Division of Endocrinology, Diabetes and Bone Disease, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

Correspondence: Andrea Dunaif, MD, Division of Endocrinology, Diabetes and Bone Disease, Department of Medicine, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1055, New York, NY 10029, USA. Email: andrea.dunaif@mssm.edu.

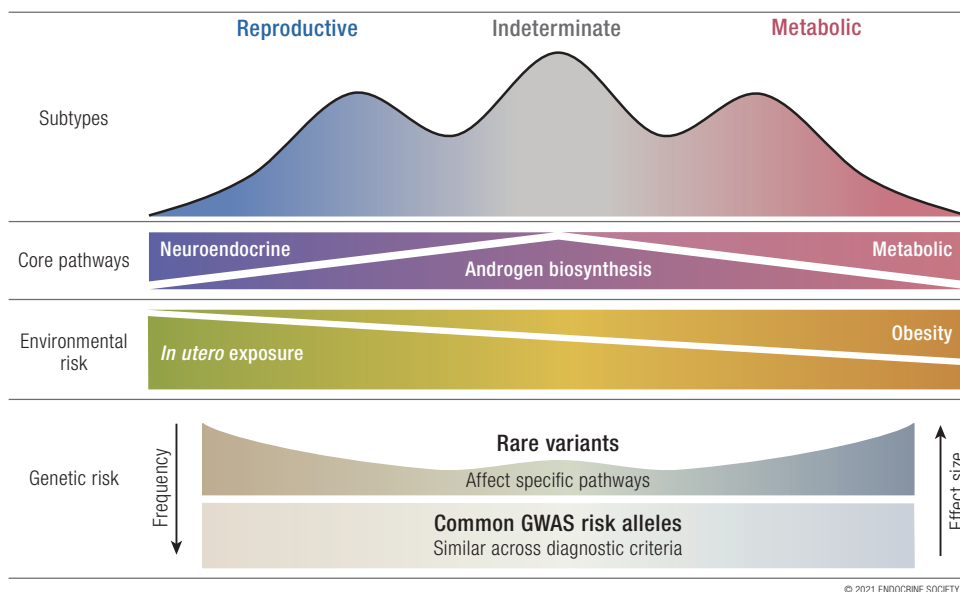
Abstract

Polycystic ovary syndrome (PCOS) is among the most common disorders in women of reproductive age, affecting up to 15% worldwide, depending on the diagnostic criteria. PCOS is characterized by a constellation of interrelated reproductive abnormalities, including disordered gonadotropin secretion, increased androgen production, chronic anovulation, and polycystic ovarian morphology. It is frequently associated with insulin resistance and obesity. These reproductive and metabolic derangements cause major morbidities across the lifespan, including anovulatory infertility and type 2 diabetes (T2D).

Despite decades of investigative effort, the etiology of PCOS remains unknown. Familial clustering of PCOS cases has indicated a genetic contribution to PCOS. There are rare Mendelian forms of PCOS associated with extreme phenotypes, but PCOS typically follows a non-Mendelian pattern of inheritance consistent with a complex genetic architecture, analogous to T2D and obesity, that reflects the interaction of susceptibility genes and environmental factors. Genomic studies of PCOS have provided important insights into disease pathways and have indicated that current diagnostic criteria do not capture underlying differences in biology associated with different forms of PCOS.

We provide a state-of-the-science review of genetic analyses of PCOS, including an overview of genomic methodologies aimed at a general audience of non-geneticists and clinicians. Applications in PCOS will be discussed, including strengths and limitations of each study. The contributions of environmental factors, including developmental origins, will be reviewed. Insights into the pathogenesis and genetic architecture of PCOS will be summarized. Future directions for PCOS genetic studies will be outlined.

Graphical Abstract



Key Words: polycystic ovary syndrome, PCOS, androgen excess, genetics, genomics, epigenetics

Abbreviations: AES, Androgen Excess Society; AMH, anti-Müllerian hormone; BMI, body mass index; ChIP-seq, chromatin immunoprecipitation followed

Received: 15 March 2021. Editorial Decision: 4 January 2022. Corrected and Typeset: 3 February 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of the Endocrine Society. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

by NGS; DHEAS, dehydroepiandrosterone sulfate; DZ, dizygotic; EHR, electronic health record; FSH, follicle-stimulating hormone; FSHR, follicle-stimulating hormone receptor; GnRH, gonadotropin-releasing hormone; GWAS, genome-wide association studies; HA, hyperandrogenism; LD, linkage disequilibrium; LH, luteinizing hormone; LHCGR, luteinizing hormone/human chorionic gonadotrophin receptor; MAF, minor allele frequency; MZ, monozygotic; NGS, next-generation sequencing; NICHD, National Institute of Child Health and Human Development; NIH, National Institutes of Health; OD, ovulatory dysfunction; PCO, polycystic ovaries; PCOM, polycystic ovary morphology; PCOS, polycystic ovary syndrome; PheWAS, phenome-wide association study; RCT, randomized controlled trial; SHBG, sex hormone-binding globulin; SNP, single-nucleotide polymorphism; T2D, type 2 diabetes; TDT, transmission disequilibrium test; WES, whole exome sequencing; WGS, whole genome sequencing.

Essential Points

- PCOS is a highly heritable complex trait with ~20 identified common susceptibility variants from genome-wide association studies and rare variants in *DENND1A*, *AMH*, and *AMHR2* discovered by next-generation sequencing
- The genetic architecture of NIH PCOS, non-NIH Rotterdam PCOS, and self-reported PCOS is generally similar, suggesting that these diagnostic criteria do not identify biologically distinct PCOS phenotypes
- Unsupervised clustering analysis has identified discrete and reproducible reproductive and metabolic PCOS subtypes with preliminary evidence for distinct genetic architectures
- Environmental factors, such as intrauterine androgen excess, may act through epigenetic mechanisms in concert with susceptibility variants to produce PCOS phenotypes
- Ongoing genetic analyses promise to elucidate the distinct etiologies of PCOS, enabling the transition toward precision medicine for PCOS

Polycystic ovary syndrome (PCOS) was originally described as a reproductive disorder characterized by enlarged, smooth polycystic ovaries, menstrual irregularity, infertility, and hirsutism (1, 2). Stein and Leventhal (3) are credited with the first report that the clinical features of menstrual regularity and infertility could be improved by removal of portions of the enlarged ovaries in a procedure known as *bilateral ovarian wedge resection*.

In 1935, Stein and Leventhal called attention to a syndrome characterized by amenorrhea, occasionally menometrorrhagia, sterility, and hirsutism. The pathological findings were large, pale, polycystic ovaries with thickened capsules. Therapeutically, it was found that with adequate wedge resection of both ovaries, fertility and regular menstrual periods could be achieved in many instances (4).

As a result, this constellation of findings became known as the Stein-Leventhal syndrome (2, 5). Over the ensuing decades, PCOS has replaced Stein-Leventhal syndrome as the preferred terminology (2, 6). With the advent of hormone measurements in the 1950s, the biochemical features of PCOS began to be characterized.

Despite derangements in gonadotropin secretion and androgen production, PCOS remained an enigmatic reproductive disorder apparently amenable to surgical cure. The discovery in the 1970s of the monogenic syndromes of extreme insulin resistance, acanthosis nigricans, and hyperandrogenism, due to mutations in the insulin

receptor gene, suggested a possible causal link between hyperinsulinemia and hyperandrogenism (7). In 1980, Burghen and colleagues (8) reported that women with PCOS had increased insulin responses during oral glucose tolerance testing independent of obesity, which suggested for the first time that PCOS might be related to these insulin resistance-hyperandrogenism syndromes. Subsequent studies reported that acanthosis nigricans was a common finding in women with typical PCOS (9, 10). Studies of glucose homeostasis indicated that affected women were substantially insulin resistant, independent of obesity, but phenotypically distinct from women with the monogenic disorders of extreme insulin resistance (reviewed in (7, 11)). Nevertheless, these similarities suggested that variation in the insulin receptor gene might also contribute to typical PCOS.

Pathophysiology of PCOS

Reproductive phenotype

PCOS is characterized by a constellation of interrelated reproductive hormone alterations (12) (Fig. 1) designated as a “vicious cycle” by Rebar and colleagues in 1976 (13). There is enhanced luteinizing hormone (LH) relative to follicle-stimulating hormone (FSH) release. However, circulating FSH levels are in the low-normal range compared to normally cycling women in the early follicular phase of the menstrual cycle and do not exhibit cyclic increases in

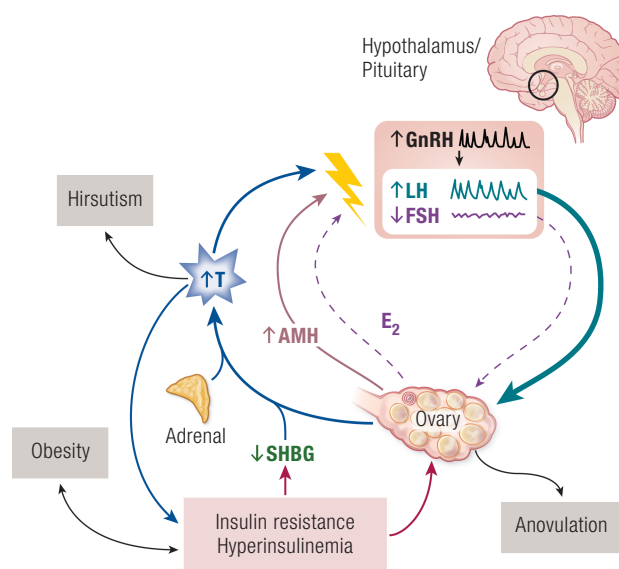


Figure 1. Pathophysiology of PCOS. The “vicious cycle” of PCOS (13), a self-sustaining constellation of reproductive abnormalities. The association with insulin resistance and hyperinsulinemia was discovered in the 1980s (reviewed in (7, 21)). Abbreviations: AMH, anti-Müllerian hormone; E₂, estradiol; FSH, follicle-stimulating hormone; GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone; PCOS, polycystic ovary syndrome; SHBG, sex hormone-binding globulin; T, testosterone.

anovulatory PCOS (13, 14). Further, the characteristic increased LH:FSH ratio may escape detection in a single random blood sample due to pulsatile LH release (15). The most frequent androgen abnormality in PCOS is increased free testosterone levels because sex hormone-binding globulin (SHBG) levels are reduced (7, 16-19). In anovulatory PCOS, estradiol levels are tonically in the mid-follicular range and progesterone levels are low (20).

Anti-Müllerian hormone (AMH) is secreted by the granulosa cells of preantral and small antral follicles and is a regulator of ovarian folliculogenesis (22). Circulating AMH levels are proportional to the number of growing ovarian follicles (23). Accordingly, AMH levels are increased in PCOS and have been considered to be a marker for the syndrome's distinctive disorder of folliculogenesis (22, 23). However, the receptor for AMH, AMHR2, is also expressed on a subset of hypothalamic gonadotropin-releasing hormone (GnRH) neurons in mice and humans (24). In mice, AMH activates GnRH neuron firing (24, 25). Thus, it has been proposed that AMH may contribute to PCOS through extragonadal actions (24, 25).

The frequency and amplitude of pulsatile GnRH secretion is increased (26, 27). This derangement selectively increases LH while simultaneously suppressing FSH release (28). Testosterone decreases the sensitivity of the hypothalamic GnRH pulse generator to the normal feedback actions of estradiol and progesterone to slow pulse frequency (29). In addition, enhanced gonadotrope sensitivity to GnRH resulting from tonic estrogen feedback contributes to increased LH pulses (30). LH stimulates ovarian theca cell testosterone production; there is decreased granulosa aromatization of testosterone to estradiol due to relative FSH deficiency (31, 32). Intraovarian factors, such as increased AMH (33) and, perhaps, alternative splicing of the androgen receptor (34), may contribute to the inhibition of granulosa cell aromatase activity. There are also constitutive increases across multiple theca cell steroidogenic enzyme pathways regulating androgen biosynthesis (35). These steroidogenic enzymes are shared by the adrenal gland so that this alteration likely contributes to increased adrenal androgen production that is a common feature of PCOS (18, 19, 36). There is also evidence for enhanced adrenal sensitivity to ACTH in PCOS (37).

Polycystic ovaries (PCO) are characterized by an increase in antral follicles and ovarian stroma as well as by theca cell hyperplasia and ovarian cortical thickening (21, 38). Theca cells from PCO secrete more androgens, basally and in response to LH (39), consistent with the constitutive activation of steroidogenic enzymes (40). Increases in androgen production are found in theca cells isolated from ovulatory as well as anovulatory women with PCOS (39). PCO have an excess of growing follicles, with a reciprocal decrease in the proportion of primordial follicles compared with normal ovaries (38, 41). The excess of follicles could result from accelerated follicle growth and/or prolonged survival of small follicles in comparison to follicles from normal ovaries (41-43). Thus, the gonadotropin-independent development of preantral follicles appears to be disordered in PCO, but this process remains poorly understood (43). Aberrant responsiveness of granulosa cells to LH (44, 45), altered expression of luteinizing hormone/human chorionic gonadotrophin receptor (LHCGR) and follicle-stimulating hormone receptor (FSHR) (46), androgens (47) and intraovarian actions of AMH and other members of the transforming growth factor beta

(TGF- β) signaling family (22, 46) may contribute to arrested follicular development characteristic of PCO. However, FSH administration can produce normal follicular maturation and ovulation (43, 48, 49).

Polycystic ovarian morphology (PCOM) detected by ovarian ultrasound examination is a very common finding in adult women with regular menstrual cycles, with prevalence rates as high as 32% in a large population-based cohort (50). The prevalence of PCOM is age-related and it decreases with increasing age (50). Indeed, PCOM prevalence rates are so high in adolescent girls that this finding is not recommended as a diagnostic criterion until 8 years after menarche (51-53). Women with PCOM and regular menstrual cycles may have reproductive hormone alterations (54, 55), most commonly higher testosterone levels, compared to reproductively normal women without PCOM (56, 57). Some women with PCOM have normal baseline reproductive hormone levels (56, 58) but increased androgen (59) and 17-hydroxyprogesterone responses (59, 60) to GnRH analog stimulation. Further, otherwise reproductively normal women with PCOM are at increased risk for ovarian hyperstimulation during ovulation induction with exogenous gonadotropins, analogous to women with PCOS (61-63). Family studies suggest that PCOM is a heritable trait (64, 65). Taken together, these observations suggest that PCOM reflects intrinsic ovarian abnormalities. Nevertheless, one study suggests that women with PCOM do not develop PCOS during follow-up and that ~50% of women with PCOM at baseline no longer met criteria for PCOM at follow-up (66).

Metabolic phenotype

In the 1980s, comprehensive studies of whole-body insulin action using the gold standard glucose clamp technique demonstrated that insulin-mediated glucose disposal, which reflects primarily skeletal muscle, was significantly and substantially decreased (~35%-40%) in women with PCOS, independent of obesity and body fat topography (67, 68). The decrease was of a similar magnitude to that reported in type 2 diabetes (T2D) (69). This finding has been replicated in multiple studies, although there are some studies in which nonobese women with PCOS have had normal insulin sensitivity; an observation that may reflect ethnic/racial differences (70). Hepatic insulin resistance is found only in obese women with PCOS, reflecting a synergistic deleterious effect of adiposity on hepatic insulin action.

The predominant cellular defect in classic insulin target tissues, adipocytes, and skeletal muscle, is a post-binding defect in insulin-mediated signaling (7). Most studies have also found a less striking, but significant, decrease in maximal rates of insulin-stimulated glucose transport (68, 71) suggesting a decrease in post-receptor events (68, 72). Significant decreases in the abundance of GLUT4 glucose transporters in subcutaneous adipocytes accounted for this defect in some (73, 74) but not all studies (75).

The insulin receptor-signaling defect is due to decreased tyrosine autophosphorylation of the receptor and downstream signaling molecules. There is increased constitutive serine phosphorylation of the insulin receptor and insulin receptor substrates due to serine kinases; serine inhibits tyrosine phosphorylation (7). One serine kinase responsible for this abnormal phosphorylation is in the MEK-ERK1/2 pathway, which is constitutively activated in PCOS skeletal muscle (76). In at least some tissues, such as skin fibroblasts

(77), skeletal muscle (76), and ovarian granulosa lutein cells (see below and (78)), insulin resistance in PCOS is selective, affecting metabolic but no other actions of insulin. However, another study suggested that both metabolic and mitogenic pathways were compromised in PCOS skeletal muscle (79).

Under normal circumstances, insulin secretion increases to compensate for peripheral insulin resistance (80). Glucose tolerance decompensates when the β -cell is no longer able to secrete sufficient amounts of insulin to meet the increased requirements (81, 82). Despite fasting hyperinsulinemia, which reflects both insulin secretion and clearance, women with PCOS have evidence for β -cell dysfunction when it is appropriately assessed (83-85). These defects are independent of obesity and dysglycemia but are more pronounced in women with PCOS who have a first-degree relative with T2D (85, 86). Evidence for β -cell dysfunction is present in daughters of women with PCOS prior to menarche (87). However, increased basal secretion rates of insulin (84) contribute to the fasting hyperinsulinemia that is characteristic of obese women with PCOS (8, 88). Studies that have reported increased insulin secretion in PCOS have been constrained by failing to examine insulin secretion in the context of insulin sensitivity (89-93).

Diagnosis of PCOS

Given the lack of information on the cause(s) of PCOS, its diagnosis has been based on the phenotypic features of the syndrome (94), as with the diagnosis of many medical disorders.

National Institutes of Health criteria

The first widely accepted diagnostic criteria for PCOS came from a survey of a small group of experts at the 1990 National Institutes of Child Health and Human Development (NICHD) Conference on PCOS (6); a scientific meeting that reviewed various features of the syndrome. These participants were asked to vote on potential diagnostic features and those receiving the most votes, hyperandrogenism (clinical and/or biochemical) and chronic anovulation, with the exclusion of secondary causes, became what are known as the NICHD or National Institutes of Health (NIH) criteria (95). The NIH criteria did not include ovarian morphology because of the lack of specificity of this finding (95). It was clear at that time that 20% to 30% of women with regular menses and no androgenic symptoms had PCOM on ultrasound examination (54). An increased LH:FSH ratio had been a common diagnostic criterion for PCOS, but it was not included in the NIH criteria because it was thought that it could escape reliable detection in a single blood sample due to the pulsatility of LH release (15, 95). The widespread adoption of the NIH criteria standardized the diagnosis of PCOS. Importantly, these criteria identified a single phenotype of women with chronic anovulation, also known as *ovulatory dysfunction* (OD), and *hyperandrogenism* (HA). In doing so, these criteria enabled research in the field by providing a common definition of PCOS.

Rotterdam criteria

In 2003, a conference specifically addressing the PCOS diagnostic criteria was convened in Rotterdam, the Netherlands (96, 97). Although designated as a consensus conference, the conference did not follow any formal consensus development

process nor were the recommended criteria evidence-based. The result of the conference was that PCOM on ultrasound examination was added to the NIH diagnostic criteria (96, 97). Accordingly, the Rotterdam criteria (96, 97) for the diagnosis of PCOS required the presence of 2 of the 3 following findings, after the exclusion of disorders of the pituitary, ovary, or adrenals that could present in a manner similar to PCOS: 1) HA; 2) chronic anovulation or OD; and 3) PCOM. These criteria include the NIH criteria but extended the diagnosis to include 2 new groups of affected women: 1) PCOM+HA without OD; and 2) PCOM+OD without HA (Table 1).

Androgen Excess Society

In 2006, the Androgen Excess Society (AES) charged a task force of experts in the field with the development of an evidence-based definition of PCOS (18, 19). The task force recommended that hyperandrogenism be considered as an essential component of PCOS. However, robust human data supporting an essential role for HA in the pathogenesis of PCOS was not provided, although such data were beginning to emerge (98, 99). According to the AES criteria, the diagnosis of PCOS required biochemical or clinical HA with 1) OD \pm PCOM; or with 2) PCOM without OD (18, 19). In addition, the task force proposed further stratifying these AES phenotypes by the presence or absence of hirsutism. The AES criteria eliminated the Rotterdam phenotype of OD+PCOM (18, 19) (Table 1). The AES criteria have not been widely adopted (100).

NIH-sponsored evidence-based methodology workshop on PCOS

In 2012, the NIH sponsored an evidence-based methodology workshop on PCOS, in which a panel of experts reviewed the current state-of-the-science (101). Although it was not an official NIH Consensus Development Conference (<https://consensus.nih.gov/>), the meeting followed the same “court” model where the evidence was presented to a panel that functioned as a jury. The panel members were experts in relevant topic areas, such as Gynecology, Diabetes and Metabolism, Cardiology, and Primary Care, but were not engaged in PCOS research. As noted in the panel’s report (102), “invited experts discussed the body of evidence and attendees had opportunities to provide comments during open discussion periods.” The panel’s final report noted that “the name ‘PCOS’ was a distraction and an impediment to progress,” and that the emphasis on PCOM created confusion because it was neither necessary nor sufficient for the diagnosis of PCOS. The panel recommended using the Rotterdam criteria with precise specification of the phenotype in research studies. Further, they proposed a comprehensive research agenda that included assessment of the epidemiology and long-term health outcomes of the PCOS phenotypes.

Evidence-based guidelines

There are 2 major evidence-based guidelines for the diagnosis and management of PCOS: the Endocrine Society’s 2013 Clinical Practice Guideline (103) and the 2018 International Evidence-Based Guideline (51-53). However, the quality of the evidence upon which these guidelines are based is predominantly low, due to a paucity of randomized clinical trials (RCTs) in the field. Of the 34 recommendations in the Endocrine Society Clinical Practice Guideline, the evidence supporting 24 of these was rated as *low* or *very low*.

There were almost 175 recommendations in the International Guideline, of which only 31 were ranked as evidence-based, the remainder were clinical consensus recommendations or clinical practice points. Both guidelines recommended the use of the Rotterdam criteria for the diagnosis of PCOS, which, as outlined above, are based on expert opinion rather than on RCTs or high-quality observational studies.

Assessment of PCOS phenotypes

There have been numerous cross-sectional studies comparing the PCOS phenotypes created by application of the PCOS diagnostic criteria (Table 1) (reviewed in (7, 104)). There is agreement that the NIH phenotype is associated with substantially more insulin resistance and related cardiometabolic abnormalities than the non-NIH Rotterdam phenotypes of HA+PCOM or OD+PCOM (7, 50, 59, 101, 104-111). Women with HA+PCOM and with OD+PCOM tend to be leaner than those with HA+OD (105, 107, 110, 112-114). Women with HA+PCOM and with OD+PCOM are similar with regard to body mass index (BMI) and metabolic abnormalities (105, 107), although some studies have failed to detect metabolic abnormalities in women with OD+PCOM compared to reproductively normal control women (109). There are no reproductive or metabolic differences in NIH phenotype PCOS (HA+OD) with PCOM as compared to those without (115). In fact, studies report that PCOM is present in > 90% of women with HA+OD, that is, the NIH phenotype (105, 116). The strong correlation between HA+OD and PCOM is supported by genetic analyses suggesting that these traits are associated with many of the same genome-wide association studies (GWAS) loci (117) (see “*Genetic Architecture of PCOS NIH and Rotterdam Phenotypes*”). Stratification of PCOS according to Rotterdam phenotypes is also affected by the accuracy of the androgen assays utilized, whether ovulation is formally assessed, the technical capacity of the ultrasound equipment as well as the approach (transvaginal vs transabdominal), and operator expertise in performing pelvic ultrasound (18, 19, 118-122).

Epidemiology

The prevalence of the NIH PCOS phenotype is 5% to 8% of premenopausal women (123-127). These prevalence estimates are remarkably consistent across racial and ethnic groups (123, 124, 126, 128, 129). Nevertheless, racial and ethnic variation in the phenotypic features of PCOS has been reported in Latinas (130, 131), African Americans (132, 133), Icelanders

(112), Sri Lankans (113), Koreans (134), and Chinese (110). PCOS is the most common cause of normogonadotropic anovulation, accounting for 55% to 91% of the entire World Health Organization-II (WHO-II) cohort (135). The prevalence of PCOS is higher, 8% to 13%, using the 2003 Rotterdam criteria since these criteria include additional phenotypes (127, 135) (Table 1). PCOS is commonly associated with obesity, and Mendelian randomization studies suggest that BMI is causally related to PCOS (117, 136, 137) (see “*Mendelian Randomization with PCOS*”). However, one study failed to detect a significant increase in the prevalence of PCOS with the increasing prevalence of obesity (138). Additional adequately powered analyses are needed to robustly investigate this question. It is noteworthy that the prevalence of PCOS is similar in diverse populations with markedly different prevalence rates of obesity (123, 124, 126, 128, 129).

Evidence for a Genetic Contribution to PCOS

Phenotypic variance, or the observable characteristic differences between individuals, results from the combined variation of genetic, environmental, and random factors in any given population. Quantifying the relative extent that these factors can account for differences in disease prevalence is an important first step in understanding the nature of a disease. Such analyses can in turn inform the prediction of clinical outcomes. For example, in conditions that are more heavily influenced by inherited genetic factors, family history and/or genetic testing may be more predictive of disease risk and treatment outcomes than environmental or behavioral measures. Examining trait correlations within families can indicate to what extent different PCOS phenotypes are caused by inherited genetic factors.

Familial Clustering

A possible genetic susceptibility to PCOS was first suggested in the 1960s, when families with multiple affected women were reported (139-143). The phenotypic similarity between PCOS and the rare syndromes of extreme insulin resistance and hyperandrogenism (144, 145) suggested that insulin receptor mutations might also be present in PCOS (130, 146). Further, defects in insulin action persisted in cultured cells from women with PCOS, suggesting that they were genetically programmed (40, 147).

Systematic studies of relatives of women with PCOS indicated that there was familial clustering of reproductive features of the syndrome (64, 65, 98, 99, 148, 149) consistent

Table 1. Diagnostic criteria for PCOS and corresponding phenotypes

Diagnostic criteria	NIH (1990)	Rotterdam (2003)	AE-PCOS (2006)
1.HA			
2.OD	+	+	+
3.± PCOM			
1.HA			
2.PCOM		+	+
1.OD			
3.PCOM		+	

Sets of diagnostic criteria for PCOS are shown relative to their corresponding applicable phenotypes. Abbreviations: AE-PCOS, Society of Androgen Excess and Polycystic Ovary Syndrome; HA, hyperandrogenism (clinical and/or biochemical); NIH, National Institutes of Health; OD, ovulatory dysfunction; PCOM, polycystic ovarian morphology; ±, with or without.

with a genetic susceptibility to these traits. When premature balding was used to assign male affected status (64, 149), a dominant mode of inheritance was suggested but the studies were constrained by small sample size and a failure to examine all relatives (150). Nevertheless, elevated androgen levels were highly prevalent (64, 65, 98), affecting up to ~40% of reproductive-age sisters. Non-SHBG-bound testosterone levels had a bimodal distribution in sisters of women with PCOS compared to a unimodal distribution observed in control women (98, 151), suggesting that elevated testosterone levels in PCOS were a monogenic trait within families (152) and/or self-reinforcing, as in a positive feedback loop, beyond a particular threshold (153). Similarly, bimodality of insulin levels was observed in sisters of women with PCOS (151). Brothers (154), mothers (155), and premenarchal daughters (87) had hyperandrogenemia, suggesting it was a consistent reproductive endophenotype. Elevated AMH levels were also present in male (156) as well as female relatives, including children (157-159). Metabolic features of the syndrome, including hyperinsulinemia (98, 151, 155), T2D (160, 161), and obesity (162), were present in both male and female relatives.

Heritability

The proportion of trait variation in a population that can be explained by genetic differences between individuals is referred to as a trait's heritability (163). The higher the heritability is for a given trait or disease, the more informative genetic studies can be for understanding its underlying causes. Twins are especially informative for estimating trait heritability, because they share the same environmental influences (164). Heritability can be derived by simply comparing trait correlations between pairs of monozygotic (MZ) twins, who share identical DNA, to those between pairs of dizygotic (DZ) twins, who share half of their DNA. In a large study of Dutch twins, Vink and colleagues estimated the heritability of PCOS (165). They defined PCOS as having fewer than 9 menstrual cycles annually, combined with hirsutism or acne. They found that PCOS defined by these criteria co-occurred in MZ twins with a rate of 0.71, compared to 0.38 in DZ twins (165). These correlations suggest an additive heritability of 0.66 for PCOS, but by using a common pathway model that accounted for individual genetic effects from oligomenorrhea, acne, and hirsutism, they estimated the overall heritability of PCOS to be 0.79 (165).

Other studies have shown that individual hormonal components of PCOS are also highly heritable. Family studies of testosterone heritability in women have yielded estimates of 0.26 to 0.50 (65, 166-168). Correlations of dehydroepiandrosterone sulfate (DHEAS) levels between PCOS probands and their sisters indicated a heritability of 0.44 for BMI-adjusted DHEAS (169). Heritability estimates for SHBG in women have ranged from 0.56 to 0.63 (65, 168). Metabolic factors such as BMI and insulin resistance have also been found to be highly heritable in sisters of women with PCOS (65). New computational methods have recently been developed that can use large, unrelated genomic datasets to estimate the heritability of traits due to common genetic differences between individuals (170). Such methods have not yet been applied with adequate power to PCOS, itself, but have produced common-variant heritability estimates of 0.13 to 0.20 for testosterone and SHBG levels in women (171).

Genetics for Non-Geneticists

Detecting Genetic Variation

The familial clustering and estimated heritability of PCOS and its component traits provide compelling evidence for genetic susceptibility to PCOS and provide a rationale for studies investigating associated genetic variation. Genetic variation arises from naturally occurring random mutations in germ cells that are passed to future generations, ultimately becoming distributed among populations via random drift, migration, and selection (172-174). By producing phenotypic changes, genetic variants are selected for or against depending on the survival advantages or disadvantages they may confer under different environmental pressures (175). Importantly, the phenotypic effects from individual genetic variants vary widely and depend on many interacting factors, such as age (176), environmental influences (177), and the co-occurrence of other genetic variants (178). Natural selection generally acts to prevent the proliferation of deleterious alleles within populations (175, 179, 180). Deleterious variants with large phenotypic effects—the variants that can most directly point to specific disease mechanisms—therefore, tend to be rare (175, 181, 182). Consequently, there is generally an inverse relationship between variant frequency and effect size. Rare diseases are typically caused by *de novo* mutations or from co-occurrence of rare, recessive genetic variants within a single gene that significantly impair important biological functions. Such monogenic diseases follow classic Mendelian patterns of inheritance and are referred to as Mendelian disorders. Most of the genetic risk for common diseases, however, results from the cumulative influence of many variants with individually tiny effect sizes that interact in complex ways (183, 184). Such conditions, including PCOS, are therefore much more difficult to understand from a genetics perspective (185, 186) and are collectively referred to as complex traits. There are a number of different methods for detecting genetic variation and many different analytical approaches for testing genetic associations with a given trait or disease. We will briefly review some of the primary methods prior to discussing their specific applications in studying PCOS.

Linkage Analysis

Humans are diploid organisms with 2 pairs of every chromosome, having inherited 1 set of chromosomes from each parent. During meiosis, a single set of chromosomes carried in each gamete (sperm or egg cell) is produced in such a way that crossovers may occur between chromosome pairs, resulting in single chromosomes of alternating segments from each of the parental chromosomes. DNA is, therefore, inherited in blocks, and genes that lie nearer to each other on a chromosome tend to be inherited together on the same haplotypes (ie, are “linked”) more often than distant genes are (Fig. 2). Linkage is therefore a function of the expected rate of recombination between loci. Prior to the sequencing of the human genome, genetic distances were measured in terms of recombination frequencies.

Linkage analysis can be applied to family data to identify disease loci. By measuring polymorphic sequences in large families (using polymerase chain reaction [PCR], for example), co-segregation patterns can be studied, and specific alleles can be mapped to different phenotypes (188-190). Whether by specifying an inheritance model or by using a nonparametric approach, linkage analysis essentially

measures the relative frequencies with which alleles at different loci are shared between relatives. In theory, haplotypes carrying disease variants are shared between concordantly affected relatives more often than between discordantly affected ones. If certain disease variants consistently occur in particular regions of the genome, then linkage analyses can be used to identify those regions (191). Linkage analyses are ideal for studying Mendelian disorders that run strongly in families, but these analyses are less well suited for studying complex traits, which feature greater polygenicity, smaller allelic effect sizes, and more genetic heterogeneity (192).

Association Testing

An alternative method for identifying disease genes is the genetic association study. Rather than tracing haplotypes through families and calculating allele-sharing probabilities, genetic association studies simply compare allele counts between affected and unaffected individuals (ie, cases and controls). Association studies can therefore be performed using case-control cohorts of unrelated subjects. Association studies likewise rely on the concept of linkage to identify candidate genes, as measured genotype markers tend to be co-inherited with nearby alleles. The extent to which alleles at different loci are correlated in a population is referred to as *linkage disequilibrium* (LD). Because the average number of meioses between unrelated individuals is much greater than between relatives, the sizes of the nonrecombinant regions, or *LD blocks*, measured by markers in association studies, are much smaller than those measured in linkage analyses (Fig. 3). Accordingly, the candidate regions tagged in association studies are more precise, but many more markers are required to cover the full extent of the genome.

Prior to the advent of technologies that could efficiently genotype hundreds of thousands of markers simultaneously, association studies were typically limited to candidate genes with putative involvement in a particular disease pathway. Given large enough sample sizes, association tests can detect any number of genetic associations with more modest effect sizes. However, association tests are extremely sensitive to genetic ancestry differences between cases and controls, an effect known as *population stratification*, as alleles frequencies and LD structures vary greatly between ancestral populations (Fig. 4). These differences have arisen over time due to random genetic drift and local selection and are reflective of the migration history of modern humans. Failure to adequately control for population structure, which typically requires accurate genetic ancestry information, can lead to false positive associations (193, 194). Confounding can also occur from unaccounted-for relatedness between study participants, so-called *cryptic relatedness*, which violates assumptions of independence underlying basic association test statistics. Consequently, results from many early genetic association tests were never successfully replicated (195).

Family-Based Association Testing

Under the null hypothesis in population-based association tests, variant allele frequencies are the same between the cases and controls (196). With family-based data, however, association test statistics can be computed conditionally on observed parental genotypes, thereby eliminating bias due to population stratification and/or admixture (197-200). In other words, the null hypotheses in family-based association tests are a function of which alleles each set of parents have.

The most basic such test, known as the *transmission disequilibrium test* (TDT), simply compares the relative transmission frequencies of different marker alleles between parents and affected offspring using a χ^2 statistic, with excess transmission of an allele suggesting a disease association (201). The family-based association test (FBAT) is a generalization of the TDT method, allowing for much greater flexibility in modeling family genotype and phenotype distributions (202, 203). The FBAT can be configured for different genetic models (eg, dominant or recessive or additive), more complex pedigree structures with affected and unaffected offspring, pedigrees with missing parental genotypes, and multi-allelic markers for both binary and quantitative traits. Large pedigrees are ideal for tracing causal variants in rare diseases (204), but family-based designs can also increase the power to detect genetic associations in common diseases by inherently enriching the frequencies of causal rare variants in the study population (205, 206).

Genome-Wide Association Studies

Following the sequencing of the human genome (207, 208), the development of commercial genotyping arrays that could simultaneously assay hundreds of thousands of common DNA variants established a new era of genetic discovery (209, 210). Association testing could now be performed economically across the entire genome, thereby negating the need for *a priori* hypotheses about candidate disease genes. These analyses came to be known as *genome-wide association studies* (GWAS). The new technology held particular promise for studying the genetics of common diseases, which were hypothesized to be a function of common alleles with relatively small individual effect sizes (211, 212). By cataloging genetic variation from numerous individuals from multiple populations (213), panels of single nucleotide polymorphisms (SNPs) were developed that could most efficiently tag different haplotypes in each LD block across the entire genome. In this way, many variants can be inferred, or “imputed,” from a more limited set of genotyped SNPs. As a consequence of this design, accurate imputation of tagged variants in GWAS is predominantly limited to common variants, as LD mapping for lower frequency variants is limited by haplotype reference panel sizes (214). Furthermore, trait-associated SNPs in GWAS are rarely causal variants themselves; rather, they are linked to causal variants within the same haplotype (Figs. 2, 3).

The identification of causal variants within an associated haplotype requires more intensive follow-up study of the implicated genomic region, a process known as *fine mapping* (215). Nevertheless, the efficient identification of candidate gene regions by GWAS has greatly facilitated disease gene discovery. To date, thousands of GWAS have been published and hundreds of thousands of variant-trait associations have been identified (216). Furthermore, GWAS have enabled individual genetic risk prediction for complex diseases. By considering the combined effects from trait-associated variants across the genome, a polygenic risk score can be derived for a given individual. Genetic risk scores facilitate disease risk stratification and can inform disease screening and lifestyle modification (217). However, the predictive power of polygenic risk scores is inherently limited by a condition's heritability and the quantity of underlying genetic risk that has been identified via GWAS (218), which is a limitation for PCOS given the limited sample sizes studied to date. Furthermore, risk score models

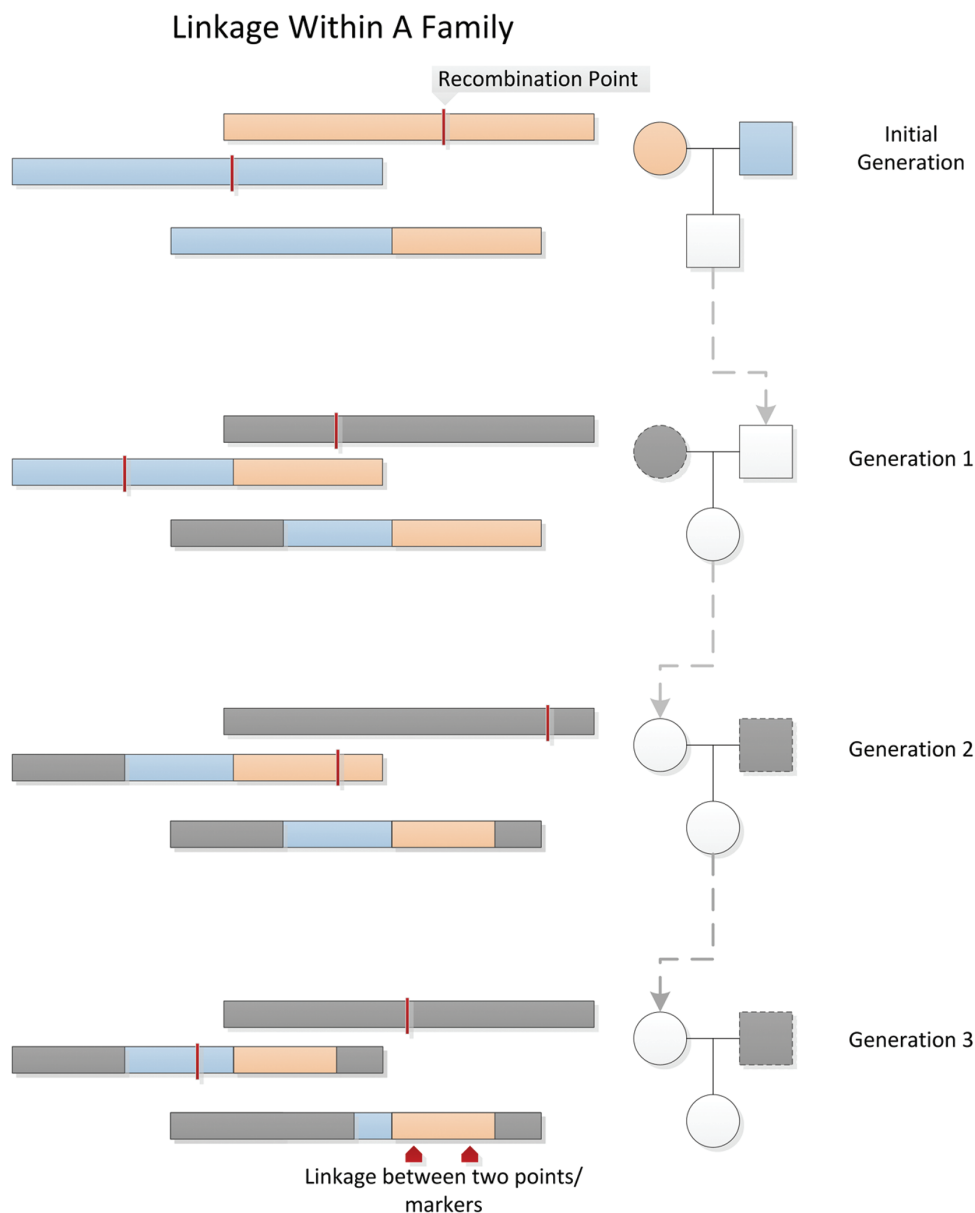


Figure 2. Linkage. Alleles that are inherited together are “linked.” The probability that any 2 markers on the same chromosome are linked is inversely proportional to the distance between the 2 markers. Figure reproduced from Bush & Moore, 2012 (187).

do not transfer well across populations due to differences in LD and allele frequencies (219).

Collectively, GWAS have also revealed a number of important characteristics regarding the genetic architecture of complex traits and common diseases in general, namely that complex traits are highly polygenic, single gene variants often confer multiple phenotypic effects (a phenomenon called *pleiotropy*), and that most of the heritability of common diseases can be explained by the cumulative effects of common genetic variation (184, 220).

Phenome-Wide Association Studies

More recently, the integration of genetic data with electronic health records (EHRs) has enabled a different kind of genetic association study, the phenome-wide association study (PheWAS). Instead of systematically testing SNP associations with a given phenotype, as in a GWAS, the paradigm is reversed in PheWAS such that a catalog of phenotypes

is systematically tested for association with a given SNP. Phenotypes included in a PheWAS could be International Classification of Diseases (ICD) codes, for example, or more carefully defined algorithmic-based conditions. PheWAS need not be limited to individual SNP associations, either; genetic risk scores can also be tested against multiple phenotypes. PheWAS are ideal for identifying pleiotropic gene effects. Identifying genetic variants with multiple disease associations using PheWAS can be informative for understanding common disease pathways, including potentially revealing new drug indications or off-target drug effects (221).

Mendelian Randomization and LD Score Regression

Genetics can also be used to assess causality between correlated phenotypes. Observational epidemiologic studies, which test for associations between diseases and various clinical, behavioral, and/or environmental factors, are susceptible to

Linkage Disequilibrium Within A Population

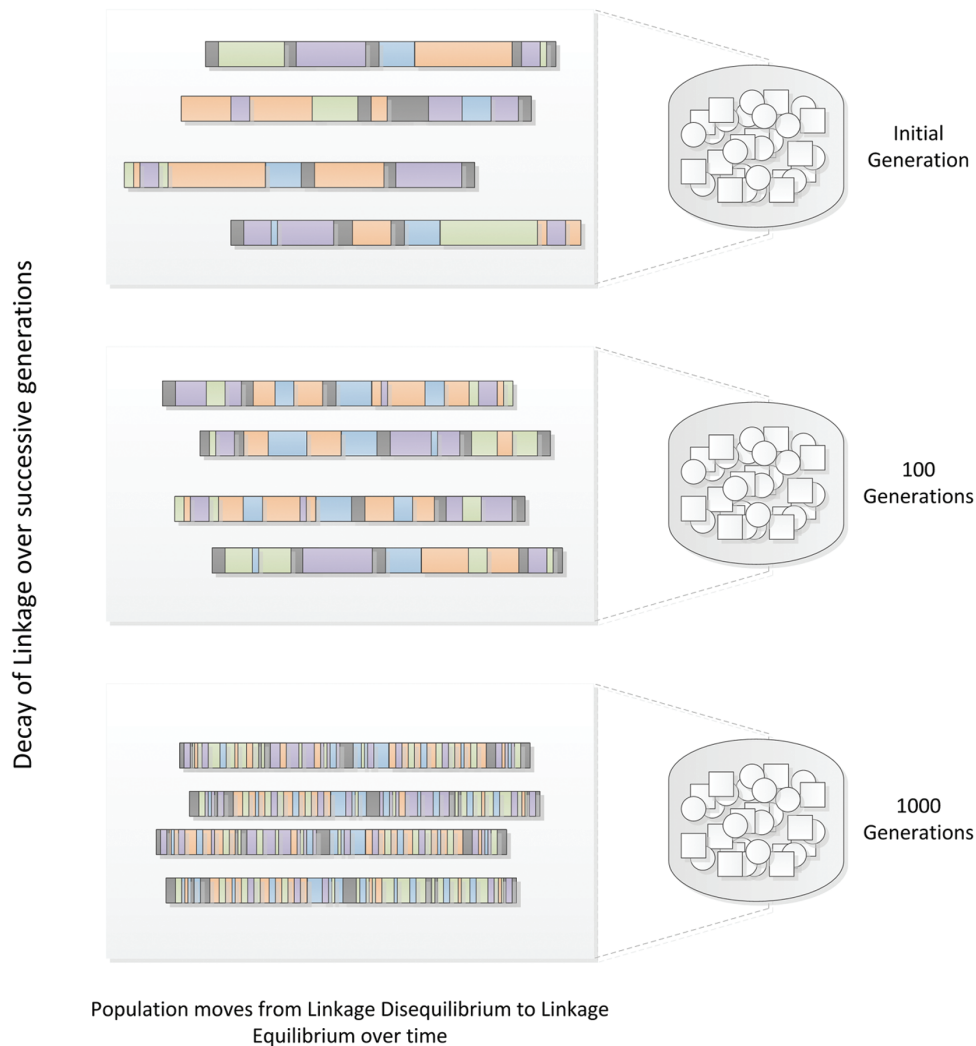


Figure 3. Linkage disequilibrium. Over many generations, linked chromosomal segments undergo successive recombination until all nonadjacent allele pairs reach equilibrium. The extent to which any 2 alleles coexist nonrandomly in a population is referred to as “linkage disequilibrium.” Figure reproduced from Bush & Moore, 2012 (187).

confounders and reverse causation, in which a disease influences the apparent exposure. Genetic alleles, however, are randomly allocated during gamete formation and are not influenced by subsequent environmental exposures or reverse causation, akin to study arm assignment in a randomized control trial (222). Genetic variation can therefore serve as proxy variants for inferring the extent to which a disease is caused by an environmental exposure in an approach called Mendelian randomization (223, 224). Mendelian randomization can likewise be used to determine whether the effects of disease-associated variants independently influence disease risk or whether they are mediated through certain intermediary risk factors (eg, BMI) (Fig. 5).

Mendelian randomization requires that the instrumental genetic variants that predict an intermediate exposure variable do not affect the outcome directly or through some other confounding relationship. For example, to determine whether BMI is causal for PCOS, the genetic variants that predict BMI should not be independently associated with PCOS or any

confounding factors like insulin resistance (Fig. 5). Violation of this assumption, known as *horizontal pleiotropy*, is common and can produce severely biased results in Mendelian randomization studies, if not properly controlled (225). Additionally, the ability of Mendelian randomization to determine causality is limited by the extent to which the exposure variable is genetically determined by the instrument variants.

GWAS remain susceptible to confounding from population stratification (226) and cryptic relatedness (194) among samples. The net confounding bias from these effects can be estimated, however, using LD score regression (227). The basic principle behind LD score regression is that the distribution of test statistics in a GWAS is a function of LD: SNPs tagging a causal variant have elevated association statistics proportional to their relative LD with the causal variant (226). Therefore, by regressing the association test statistics against the amount of genetic variation tagged by each variant (its LD score), inflation of test statistics due to LD is accounted for, and a more accurate estimate of confounding due to

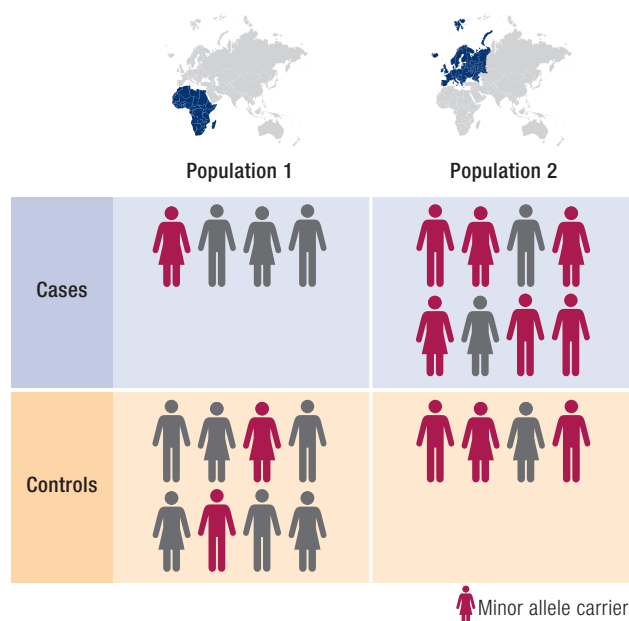


Figure 4. Population stratification. Allele frequencies often vary by ancestry. Failing to control for unequal distributions of ancestry between cases and controls can result in false positive associations.

population stratification and/or cryptic relatedness can be obtained (227). Further, LD score regression can be extended to study the genetic overlap between traits. By regressing the product of normalized effect sizes (z-scores) derived from each of 2 GWAS against variant LD scores, the resultant slope is an estimate of genetic covariance, from which the genetic correlation between the corresponding traits can be derived (228). The genetic correlation between 2 traits is essentially the extent to which the set of genetic variants that contribute to one trait also contribute to the other, the net result of pleiotropy and LD between the 2 traits. Unlike Mendelian randomization, which can probe overlapping contributions between traits for a limited number of SNPs, cross-trait LD score regression provides genome-wide estimates of genetic correlation. Quantifying the shared additive genetic effects between traits can yield broader insights into shared biological pathways and causal relationships.

Next-Generation Sequencing

The advent of next-generation sequencing (NGS) technologies, in which millions of sequence reads can be processed simultaneously, led a dramatic reduction in the time and cost associated with performing genomic sequencing (229). Unlike other genotyping methods that measure a limited set of SNPs and rely on LD patterns to fill in the blanks, as it were, NGS can directly detect nearly all positions in the genome, thereby enabling researchers to genotype rare variants and/or private mutations. Accordingly, NGS is especially useful in identifying pathogenic variants in rare diseases (230), in fine mapping of candidate gene regions (215), or in measuring the expression of novel gene variants (231). NGS has also enabled researchers to study to what extent rare alleles contribute to common disease (212). Many different experimental approaches have been developed that utilize NGS, including measuring genetic variation across the genome with whole genome sequencing (WGS) or within protein-coding regions using whole exome sequencing (WES) (232) or within specific regions with targeted sequencing; quantifying

gene expression by measuring transcribed DNA using RNA sequencing (RNA-seq) (233); identifying histone modifications using chromatin immunoprecipitation followed by NGS (ChIP-seq) (234); measuring DNA methylation using whole-genome bisulfite sequencing (WGBS) (235); and studying chromatin structure with transposase accessibility profiling, (ATAC-seq) (236), among others (237) (Table 2). Recent efforts to amass extremely large databases of whole genome sequencing data are facilitating great advances in our collective understanding around the genetics of human populations and of complex traits (238, 239).

The additional genomic resolution afforded by WGS does come at a cost, however, beyond that associated with the sequencing itself. Due to the extensive repetitive sequences found throughout the human genome and the sheer volume associated with WGS output in general, analyzing WGS data is extremely computationally arduous. Efficient processing of NGS data of any kind typically requires the use of high-performance computing clusters. A typical WGS study will require many terabytes (1 terabyte = 1024 gigabytes) of storage space (251). There are numerous open-source software tools for processing sequencing data and performing various statistical analyses, but advanced bioinformatics training is required to ensure such analyses are performed correctly with the application of appropriate quality control measures.

Genetic Analyses of PCOS

PCOS Genetics Before GWAS

Historically, most genetic studies of PCOS used a candidate gene approach, in which polymorphic sites near specific genes of interest—chosen based on the gene's hypothesized role in disease-related pathways—were tested for associations with PCOS. Such studies are inherently limited by *a priori* assumptions about gene functions. Based on the primary characteristics of the disease, early PCOS candidate genes included those involved with steroidogenesis (eg, *CYP11A1*, *CYP17A1*, *STAR*), androgen and gonadotropin action (eg, *FSHR*, *LHCGR*, *SHBG*, *AR*), and insulin resistance (eg, *INSR*, *INS-VNTR*, *IGF1*, *IGF1R*, *IRS1*, *PPARG*), among others (252). Collectively, these candidate gene studies identified hundreds of associations, but the findings were largely inconsistent, with

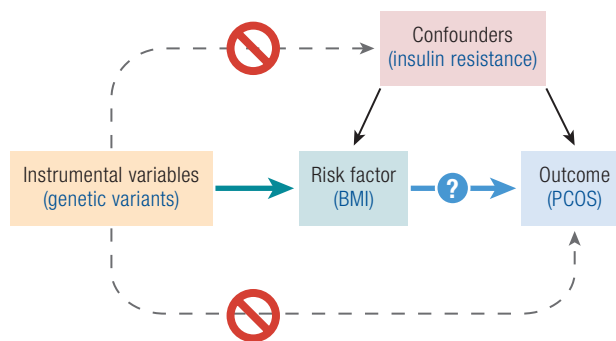


Figure 5. Mendelian randomization. In Mendelian randomization analyses, genetic variants that are predictive of an intermediate risk factor but are otherwise independent of the outcome of interest can be used to determine to what extent the intermediate factor is causal for the outcome in question. Example variables are shown in parentheses: to determine whether BMI is causal for PCOS, the genetic variants that predict BMI should not be independently associated with PCOS or any confounding factors like insulin resistance.

Table 2. Common next-generation sequencing applications

System	Method	Information	References
<i>Genome</i>	WGS (whole genome sequencing)	Genetic variation across the genome	(240)
	WES (whole exome sequencing)	Genetic variation in gene-coding regions	(232)
<i>Epigenome</i>	Targeted sequencing	Genetic variation in specific regions of interest	(241, 242)
	ChIP-seq (chromatin immunoprecipitation sequencing)	Protein-DNA interactions	(234)
	WGBS (whole-genome bisulfite sequencing)	DNA methylation sites	(235)
	DNase-seq (DNase I hypersensitivity sequencing)	Open chromatin sites	(243, 244)
	FAIRE-seq (formaldehyde-assisted isolation of regulatory elements sequencing)	Open chromatin sites	(245, 246)
	ATAC-seq (assay for transposase-accessible chromatin using sequencing)	Open chromatin sites	(236)
	Hi-C	Chromatin interactions	(247)
<i>Transcriptome</i>	ChIA-PET (chromatin immunoprecipitation and paired-end tag sequencing)	Protein-mediated chromatin interactions	(248)
	RNA-seq (RNA sequencing)	Gene expression	(233)
	smRNA-seq (small RNA sequencing)	Expression of microRNAs and other small RNAs	(249)
	CLIP-seq (cross-linking immunoprecipitation sequencing)	Protein-RNA interactions	(250)

few associations being replicated in subsequent studies for a variety of reasons (16, 253). Many failed to adequately control for population stratification between cases and controls (210), which could have produced false associations due to ethnic/racial differences in allele frequencies (254). Likewise, failure to control for comorbid phenotypes, such as obesity, could produce misleading association signals (16). These confounding effects were exacerbated by the small sample sizes in most of these studies, which also resulted in limited statistical power to detect associations with modest effect sizes (255). Most studies also failed to adequately correct for multiple hypothesis testing (256), which is the simple reality that if you perform enough statistical tests, you will inevitably observe improbable associations under the null. Finally, because PCOS phenotypes will vary according to the diagnostic criteria applied (254, 257, 258), candidate gene studies were doubtlessly affected by the populations chosen for each study.

In an effort to assess the relative reliability of different PCOS candidate gene study findings, Hiam and colleagues recently performed a systematic review of candidate gene meta-analyses in PCOS (259). Of 21 qualified candidate gene meta-analyses, only 5 were deemed “high quality.” Most of the meta-analyses did not describe inclusion criteria for control groups or exclude studies with statistically implausible distributions of alleles (260). Moreover, the quality assessment used to systematically evaluate the meta-analyses (261) did not include specific considerations for genetic association studies, such as those detailed above, so even candidate gene meta-analyses that were deemed to be of relatively high quality could still suffer from underlying issues that affect candidate gene studies in general.

The first replicated candidate risk locus for PCOS was in a region on chromosome 19 linked to the insulin receptor gene, *INSR*. In linkage and TDT analyses of 150 families with at least one affected case, we (252) identified an allele located near *INSR* that was significantly associated with NIH PCOS or hyperandrogenemia. The NIH PCOS or hyperandrogenemia phenotype was used based on our study suggesting that hyperandrogenemia was a major endophenotype in the sisters of women with PCOS (98). This finding was then replicated by us in several independent family-based cohorts (262-265), as well as in subsequent case-control analyses (266, 267). Although *INSR* was the intended target in the initial study, the associated allele was actually located in an intron of the fibrillin 3 gene, *FBN3*, about 800kb from *INSR*. The risk marker could conceivably have been tagging an allele in *FBN3* or any one of several other genes in the genomic vicinity, despite the physiologic evidence in support of *INSR*. However, by measuring a set of SNPs across the entire *INSR* gene in a relatively large case-control association study (799 cases, 3758 controls), Goodarzi and colleagues did identify an intronic variant within *INSR* that was significantly associated with PCOS, independent of BMI, which lent additional evidence in support of *INSR* being the functional gene in the region (268), a finding subsequently supported by GWAS (269). Nevertheless, until one or more functional *INSR* variants are experimentally validated, it remains unknown which gene(s) in this gene-rich region of chromosome 19 are actually contributing to PCOS.

GWAS of PCOS

The development of cost-efficient, high-throughput genotyping platforms marked a turning point in the study

of complex trait genetics, including PCOS. Genetic variation could now be measured across the genome using a hypothesis-free approach in thousands of individuals in order to identify alleles associated with the disease or trait of interest. The first 2 PCOS GWAS were completed in Han Chinese cohorts with PCOS cases defined according to the Rotterdam criteria (269, 270). These GWAS collectively identified 11 significant association signals with PCOS (Table 3 (117, 269-273)). Two subsequent GWAS were performed in women of European ancestry, with the first using the NIH definition for PCOS (271) and the second using self-reported PCOS cases validated by replication in additional cohorts of Rotterdam and NIH phenotype cases (272). These European GWAS identified 5 novel associations—including a locus encompassing the *FSHB* gene, which encodes the beta subunit of FSH—in addition to replicating 2 Han Chinese association signals. Two GWAS have also been performed in Korean women using the Rotterdam criteria (274, 275), but neither identified any loci significantly associated with PCOS, likely due to limited sample sizes and the prevalence of hyperandrogenemia in the controls (16). A more recent, large-scale meta-analysis of PCOS in European ancestry identified 3 novel loci and replicated 11 of the previously reported loci (117). Finally, a meta-analysis in mixed ancestries from multiple biobanks identified a novel locus associated with algorithmically defined Rotterdam PCOS cases from EHRs (273) (Table 3), but the accuracy of the EHR-based algorithm was not validated by manual chart review or patient recall (273). Further, none of the previously identified PCOS GWAS loci were replicated with genome-wide significance, which is surprising considering multiple loci have been otherwise replicated in PCOS cohorts of different ancestries (Table 3). Recent studies have validated EHR-based algorithms for effective identification of PCOS with positive-predictive values over 90% in populations enriched for PCOS (276, 277), which should facilitate more biobank-based genetic analyses. However, the best performing algorithms have low sensitivity (~50%) (276). Therefore, EHR-predicted case cohorts may represent a distinct subset of women with PCOS compared to traditionally defined PCOS case cohorts.

In total, 19 loci significantly associated with PCOS risk have been identified via GWAS in Chinese and European cohorts, with one additional locus identified in a mixed cohort. Twelve loci have been replicated with genome-wide significance in more than one GWAS, with 6 across different ancestries (Table 3). Further, most loci have now been replicated across ancestries in smaller replication studies (278). The fact that many of the top PCOS risk loci are shared between European and Chinese populations suggests that PCOS is an ancient trait and that aspects of PCOS were selected for prior to human migration out of Africa (279).

A number of risk loci identified by these PCOS GWAS contain genes that had already been recognized as candidate genes for PCOS, namely the receptors for LH (*LHCGR*), FSH (*FSHR*), and insulin (*INSR*). Granulosa cell expression of *FSHR* was previously found to be higher in women with PCOS (280), and certain *FSHR* haplotypes had been identified that were associated with PCOS or its cardinal features in smaller, candidate gene studies (281-283). The gene encoding the beta subunit of FSH, *FSHB*, is also at a well-replicated PCOS GWAS risk locus (117, 271, 272, 284). The primary risk variant in the *FSHB* locus resides in a highly conserved regulatory region

upstream of *FSHB* (285, 286) and upregulated *FSHB* transcription via enhancer activity in mouse gonadotropes (285, 287). *FSHB* variants were significantly associated with LH and FSH levels in women of European ancestry (271, 272), and with LH levels in women of Han Chinese ancestry (284). *LHCGR* was identified as a candidate gene due to its role in gonadotropin action (288). Its expression was higher in theca and granulosa cells from polycystic ovaries (289), and the gene was significantly demethylated, thereby facilitating expression (290, 291), in an animal model of PCOS (292). Importantly, the region of LD for the GWAS locus over *INSR* did not extend beyond the bounds of the *INSR* gene, itself (269). Therefore, it remains possible that the previously identified risk allele in *FBN3* (264) marks an independent risk allele. Nonetheless, the *INSR* GWAS signal provided the strongest evidence yet for *INSR*'s role in PCOS pathogenesis. Indeed, although each of these genes had already been implicated in PCOS, for reasons noted above, such findings were not always consistent (256, 282, 293-295). The identification of these loci in PCOS GWAS served to further corroborate their putative involvement in PCOS pathogenesis.

Other GWAS loci include genes that had not yet been considered in PCOS but have since been connected to specific disease pathways. Risk alleles for T2D had already been identified at the 2 loci containing the *THADA* and *HMGA2* genes (296, 297). The T2D risk alleles in *THADA* are independent of PCOS risk alleles (298), but *THADA* is now known to regulate thermogenesis (299, 300). *HMGA2* has been found to promote adipogenesis (301, 302) and granulosa cell proliferation (303). The *DENND1A* gene was discovered to be an important regulator of theca cell androgen biosynthesis, with upregulated expression in PCOS (304). The *YAP1* gene was also found to play a key role in ovarian follicle development (305, 306). The GWAS locus spanning the *RAB5B*, *SUOX*, and *ERBB3* genes was previously identified as containing a risk allele for type 1 diabetes (T1D) (307, 308). *ERBB3* has since been implicated in β -cell apoptosis (309). *ERBB4*, which is also at a risk locus for BMI (310, 311), encodes a receptor in the epidermal growth factor receptor (EGFR) superfamily, like *ERBB3*. *ERBB4* has been found to regulate the oocyte environment during folliculogenesis (312). Both *ERBB3* and *ERBB4* bind neuregulins 1 (NRG1) and 2 (NRG2) (313). NRG1 is expressed and secreted by granulosa cells in response to the ovulatory LH surge regulates luteinization and oocyte maturation (314, 315). NRG1 has also been found to affect glucose metabolism in rodents (316, 317). These results suggest that the *ERBB3* and *ERBB4* GWAS loci may contribute to PCOS risk via reproductive and metabolic pathways. The lead variant in the chr5q13.1 risk allele in Europeans is associated with alternative splicing of the nearby *RAD50* gene in the ovary, according to estimates from the Genotype-Tissue Expression (GTEx) project database (318, 319). *RAD50* is involved in DNA damage response signaling (320). DNA repair pathways have been implicated in age of menopause GWAS (321, 322). Therefore, alleles at the *RAD50* locus may include a common mechanism between PCOS and ovarian aging, although this connection is speculative.

The potential mechanisms through which other PCOS GWAS loci affect PCOS risk remain to be determined, but colocalization studies are beginning to connect these GWAS loci to different pathways (323, 324). Collectively, PCOS GWAS have substantially advanced our understanding of

Table 3. PCOS GWAS susceptibility loci

Locus	Implicated genes	Han Chinese Rotterdam; Chen, 2011 (270)	Han Chinese Rotterdam; Shi, 2012 (269)	European NIH; Hayes & Urbanek, 2015 (271)	European self-report, Rotterdam, NIH; Day, 2015 (272)	European self-report, Rotterdam, NIH; Day, 2018 (117)	Mixed ancestry EHR-based Rotterdam; Zhang & Ho, 2020 (273)
2p16.3	<i>LHCGR</i>	x					
2p16.3	<i>FSHR</i>	x					
2p21	<i>THADA</i>	x			x	x	
2q34	<i>ERBB4</i>				x	x	
5q13.1	<i>IRF1/RAD50</i>				x	x	
6q25.3	<i>SOD2</i>						x
8p32.1	<i>GATA4/NEIL2</i>			x		x	
9p24.1	<i>PLGRKT</i>					x	
9q22.32	<i>C9orf3</i>		x	x		x	
9q33.3	<i>DENND1A</i>	x				x	
11p14.1	<i>FSHB</i>			x		x	
11q22.1	<i>YAP1</i>	x			x	x	
11q23.1	<i>ZBTB16</i>					x	
12q13.2	<i>RAB5B/SUOX/ERBB3</i>	x				x	
12q14.3	<i>HMG2</i>		x				
12q21.2	<i>KRR1</i>				x	x	
16q12.1	<i>TOX3</i>		x			x	
19q13.3	<i>INSR</i>	x					
20q11.21	<i>SUMO1P1</i>	x					
20q13.2	<i>MAPRE1</i>					x	

Loci with reported associations with PCOS at genome-wide significance in GWAS. Loci identified in multiple studies are shown in bold. Loci identified in multiple ancestries are highlighted.

PCOS pathophysiology by identifying new candidate genes and by implicating various causal pathways, including gonadotropin secretion (*FSHB*) and action (*LHCGR*, *FSHR*), androgen biosynthesis (*DENND1A*), metabolic regulation (*THADA*, *INSR*, *HMG2*), follicle development (*HMG2*, *YAP1*), and age of menopause (*FSHB*, *RAD50*). Additional fine mapping and functional studies are needed to confirm the roles of these putative disease genes and uncover the specific molecular mechanisms through which variation in these GWAS loci confer risk for PCOS.

Case-control GWAS of PCOS have been informative in identifying disease loci, but there have also been GWAS performed for PCOS-related quantitative traits. Our European ancestry GWAS (271) included genome-wide association tests with testosterone, LH, FSH, DHEAS, and SHBG. There was a significant association between the PCOS risk allele at the *FSHB* locus and LH levels. This finding was replicated in Han Chinese ancestry PCOS with a variant in *FSHB* that was the lead SNP in this ancestry (284). A GWAS using data from the Twins UK cohort, featuring ~2600 women and ~300 men, replicated this finding in a non-PCOS population by finding that a highly correlated variant at the same locus was associated with higher LH and lower FSH levels (325). In a GWAS limited to mothers of dizygotic (DZ) twins, the *FSHB* PCOS variant was also associated with higher FSH levels and DZ twinning, as well as with earlier ages at menarche, menopause, and first child (326). These findings suggest that this *FSHB* locus is a general regulator of fertility in women rather than a PCOS-specific variant. A quantitative trait analysis in Han Chinese PCOS cases and control women identified variants in the *FSHR* gene that were associated with FSH levels, independent of PCOS status (327).

Using data from the UK Biobank, a large-scale biomedical database containing genetic and health information from half a million participants (328), Ruth, Day, Tyrrell, Thompson, and colleagues performed GWAS of testosterone and SHBG levels in more than 188 000 women (171). They found 254 significant genetic associations with total testosterone and 359 with SHBG, which demonstrates that these traits are highly polygenic. In a GWAS of steroid hormones in Germans, a low-frequency variant in *THADA* was significantly associated with DHEAS levels in women (329). The DHEAS-increasing allele is found exclusively on the same haplotype as the European PCOS-risk allele in *THADA* (117, 272, 330, 331). Large-scale quantitative trait studies such as these may ultimately help in mapping the biological pathways surrounding PCOS disease genes. These results also demonstrate that PCOS risk alleles may affect general regulators of circulating hormone levels and that such variants are not necessarily specific to PCOS.

Mendelian Randomization With PCOS

The reproductive and metabolic abnormalities of PCOS are interrelated and the primary defect(s) remains unknown. Mendelian randomization studies have enabled the investigation of putative causal pathways and complement the insights provided by GWAS (Fig. 6). Genetically determined BMI is a significant risk factor for PCOS, with reported odds ratios of 1.90 to 4.89 per standard deviation increase in BMI (136, 137, 272). Conversely, genetically determined PCOS is not associated with increased BMI (137). A Mendelian randomization analysis conducted in Korean women did not find a

causal relationship between BMI and PCOS, but the genetic instrument for BMI consisted of only 3 SNPs, and the PCOS GWAS sample size was small, with only 1000 cases and 1000 controls (332). Further, Mendelian randomization suggests that higher insulin and lower SHBG levels have causal effects on PCOS (117, 272, 333). These studies provide robust genetic evidence that cardiometabolic abnormalities contribute to the development of PCOS.

Similarly, Mendelian randomization studies have elucidated reproductive causal pathways. Both genetically determined total and non-SHBG-bound testosterone contribute to the development of PCOS (171), providing support for the hypothesis that testosterone plays a primary role in PCOS pathogenesis (18, 19, 98, 335). In addition, a Mendelian randomization analysis of genetically determined serum metabolites found that higher levels of epiandrosterone sulfate (EPIA-S) were significantly associated with risk of developing PCOS (334). Genetically determined later age of menopause was also causally associated with PCOS (117, 272), which may be attributable to the PCOS risk variants at the *FSHB* locus (271, 284, 321, 325, 326, 336, 337) and/or perhaps the DNA repair gene *RAD50* (272). Male pattern balding has been proposed to be a PCOS male phenotype (142, 149); this hypothesis is supported by Mendelian randomization showing a causal association between male pattern balding and PCOS (117). Of course, menopause and male pattern balding do not precede PCOS; therefore, their causal associations must represent overlapping biological pathways, as opposed to other traits that may directly influence PCOS risk, such as testosterone or BMI.

There have been no adequately powered prospective studies to determine the long-term morbidities of PCOS, such as T2D, cardiovascular events, or cancer (338, 339). Many cross-sectional and patient registry studies have shown that PCOS is associated with an increased risk of T2D (338-344), independent of obesity. However, obesity substantially increases this risk (338-342). Although PCOS is associated with multiple risk factors for atherosclerosis, including metabolic syndrome (345), endothelial dysfunction (346), and increased carotid intima thickness (347), no study has demonstrated an increase in cardiovascular events (348). A recent Mendelian randomization study found that genetically determined PCOS did not increase risk for coronary heart disease or stroke in Europeans, nor for T2D in both European and East Asian ancestry cohorts (349). However, since increased BMI, low SHBG, and higher testosterone levels in women increase risk for both T2D and PCOS (117, 137, 171, 272, 350), these common features may account for the association of PCOS with T2D (349). Because Mendelian randomization is confounded by shared genetic etiology, also known as horizontal pleiotropy, it may be difficult to eliminate all potentially confounding associations when assessing causality between 2 highly correlated complex traits like T2D and PCOS (351). Indeed, LD score regression suggests substantial shared genetic architecture between PCOS and T2D (117). Recently, several methods have been developed to account for horizontal pleiotropy in Mendelian randomization analyses (352-354). These methods may help improve our understanding of causal relationships between PCOS and other complex traits, but to our knowledge none have yet been applied in studying PCOS.

Mendelian randomization studies have investigated whether PCOS is a risk factor for certain cancers. There

have been conflicting data regarding the association between PCOS and ovarian cancer (355), but a recent Mendelian randomization study found that PCOS had a modest protective effect against ovarian cancer (356). Further, genetically determined PCOS was causally associated for estrogen receptor-positive, but not estrogen receptor-negative breast cancer (357).

LD Score Regression With PCOS

In their large-scale meta-analysis of PCOS, Day and colleagues (117) performed LD score regression to investigate the genetic correlations between PCOS and other traits (Fig. 7). The analysis revealed strong genetic correlations between PCOS and BMI, childhood obesity, fasting insulin, T2D, triglyceride levels, high-density lipoprotein levels, and cardiovascular disease. These correlations indicate there is substantial shared genetic architecture between PCOS and cardiometabolic disorders. Additionally, there was a significant genetic correlation between PCOS and menarche timing. A previous study found that PCOS was genetically related with earlier onset of puberty, as well, demonstrating significant genetic correlations with age at menarche in women and age at voice breaking in men (358), consistent with epidemiological findings (359-361). There was also shared genetic architecture between PCOS and depression according to Mendelian randomization and LD score regression estimates (117), which supports observational studies suggesting an increased risk of depression and anxiety in women with PCOS (362). The connection between PCOS and depression is likely mediated to some extent through obesity, as BMI is causally related to both conditions (363). Indeed, a recent study investigating shared genetic effects between PCOS and psychiatric disorders found that the causal association between

PCOS and depression was nullified after controlling for BMI (364). Male pattern baldness was associated with PCOS in a Mendelian randomization analysis, but its overall genetic correlation with PCOS was not significant according to LD score regression (117). These contrasting results suggest that the shared etiology between the 2 conditions may be limited to a few key pathways. In their UK Biobank study, Ruth, Day, Tyrrell, Thompson, and colleagues calculated genetic correlations between different sex hormone levels in women and reported correlations between total testosterone and SHBG ($r = -0.06$), between testosterone and estradiol ($r = -0.25$), and between estradiol and SHBG ($r = 0.45$), indicating that these traits are functions of overlapping biological pathways (171).

PheWAS of PCOS

PheWAS offer yet another way of investigating genetic overlap between conditions. In a large-scale PheWAS ($n_{\text{discovery}} = 49\,343$, $n_{\text{replication}} = 18\,096$) that utilized genotype data linked with electronic health record (EHR) data from the eMERGE network (365), Joo and colleagues (366) tested for associations of 1711 EHR phenotypes, classified according to ICD codes, against a PCOS polygenic risk score generated from the recent PCOS meta-GWAS summary statistics (117). They identified 13 replicated phenome-wide signals with consistent directions of effect. Some of these signals represent the same disorders, such as “obesity” and “overweight” or “sleep apnea” and “obstructive sleep apnea,” but all of the significant associations appear to be obesity-related (Table 4). These findings reinforce the comorbid relationship of BMI with PCOS. PheWAS of individual GWAS SNPs have likewise identified correlated traits, including menstrual cycle length, obesity, and age at menopause (324).

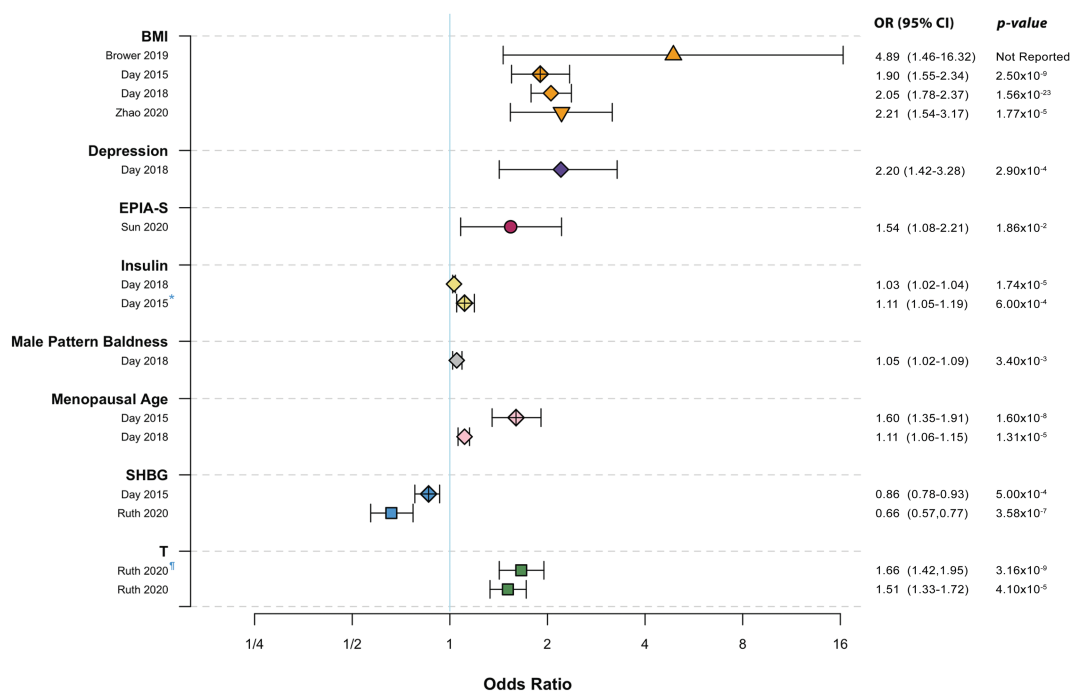


Figure 6. Significant causal factors for PCOS. Odds ratios for PCOS risk per standard deviation increase are shown for significant causal factors according to Mendelian randomization analyses (117, 136, 137, 171, 272, 334). Abbreviations: BMI, body mass index; EPIA-S, epiandrosterone sulfate; PCOS, polycystic ovary syndrome; SHBG, sex hormone-binding globulin; T, testosterone. *Insulin resistance was the tested causal factor. [†]Bioavailable testosterone was the tested causal factor.

Next-Generation Sequencing in PCOS

In recent years, the use of NGS technologies has become more common in PCOS research and has been applied numerous different ways, yielding new insights into the disease's genetic origins. We used genomic sequencing to study the role that rare variants play in PCOS, using both candidate gene and genome-wide approaches. The common PCOS-associated alleles identified in GWAS could only explain a relatively small fraction of overall PCOS heritability (271, 367); therefore, we hypothesized that rare variants, which are not reliably tagged by GWAS arrays, could account for a significant proportion of the unexplained heritability (186).

Although next-generation sequencing can be used to identify rare variants, the inherently low population frequencies of rare variants necessitate extremely large cohorts to detect variant effects via standard case vs control genetic association tests (220). For example, in their large WES study of T2D, Flannick and colleagues estimated that at least 150 000 to 370 000 sequenced exomes would be required to detect individual T2D-associated rare variants at exome-wide significance with 80% power (368). Populations can be enriched for causal rare variants by studying extreme phenotypes (369) or by studying families with multiple affected individuals (205, 370). Unlike with Mendelian disorders, however, pathogenic alleles in complex traits tend to have small phenotypic effects (371, 372). Even in enriched cohorts, most rare alleles remain too scarce for effective variant association testing (206). Therefore, rare variant association studies are often either limited to specific candidate genes or use methods that aggregate sets of rare variants together for association testing, such as with gene-based burden tests (373) or sequence kernel association tests (374). To study rare variants in PCOS, we used a candidate gene approach to study individual rare variants and a family-based analysis to study rare variants across the genome at the gene level.

In 2 targeted sequencing studies of the PCOS candidate genes *AMH* and its type 2 receptor, *AMHR2*, we measured the functional impact of rare variants (minor allele frequency [MAF] $\leq 1\%$) that were present in PCOS cases using AMH-mediated luciferase assays (375, 376). AMH plays a central role in folliculogenesis and is typically overexpressed in women with PCOS (22, 377, 378). First, we sequenced the *AMH* gene in a cohort of 643 PCOS cases and 153 controls and identified 18 rare coding variants that were present in PCOS cases but not in controls, 4 that were present in both cases and controls, and 2 found in controls only (375). Seventeen of the 18 PCOS-specific variants decreased AMH-mediated signaling capacity in transfected COS7 cells. None of the rare variants present in controls impacted AMH signaling. Several computational tools were applied to assess variant deleteriousness (379–382), but their scores were poorly correlated with the assay results, emphasizing the importance of using functional assays to confirm the biologic relevance of genetic variants. We subsequently investigated rare variation in *AMH* regulatory regions and *AMHR2* (376). We identified 20 additional PCOS-specific variants in or near *AMH* and *AMHR2* that significantly reduced AMH signaling activity. These variants included 3 noncoding variants upstream of *AMH*, 16 noncoding or splicing variants in or upstream of *AMHR2*, and 1 missense variant in *AMHR2*. Interestingly, women with PCOS and 1 or more functional *AMH/AMHR2* variants had significantly lower AMH levels than other women with PCOS.

In total, we identified 37 PCOS-specific rare variants that significantly impaired AMH signaling activity. The variants were also collectively associated with PCOS at the population level (375, 376). All PCOS cases with functional *AMH/AMHR2* variants were heterozygous carriers. Five of the functional *AMH* variants have previously been identified in men with persistent Müllerian duct syndrome (PMDS), which

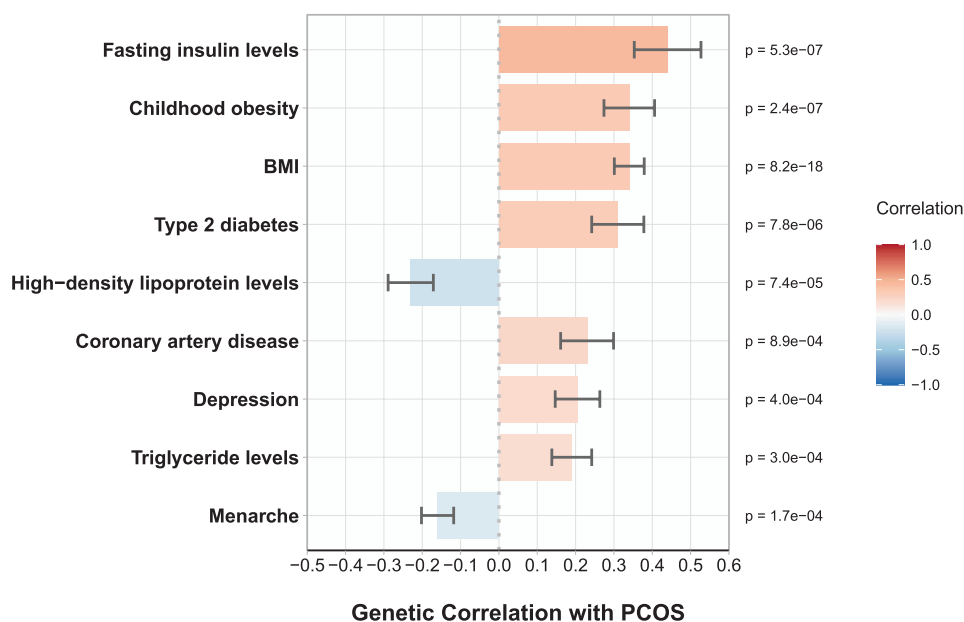


Figure 7. Genetic correlations with PCOS. Genetic correlations between PCOS risk and various hormonal and metabolic traits, according to Day and colleagues (117). Positive and negative correlations correspond to the direction of effect. Correlation estimates are shown with standard error bars. Menarche and Menopause correspond to age of onset. Other binary traits correspond to risk of occurrence. Abbreviations: BMI, body mass index; PCOS, polycystic ovary syndrome.

is a rare disorder where men retain internal Müllerian duct structures (383). Men with PMDS caused by *AMH* mutations typically have low or undetectable AMH levels (384). Loss of CYP17 inhibition by AMH, however, could contribute to PCOS, as CYP17 is a key enzyme in androgen biosynthesis (385). Indeed, we found that *AMH* variants with reduced AMH signaling showed a significant reduction in CYP17 α 1 expression inhibition compared with wild-type *AMH* (376). Testosterone levels did not differ significantly between carriers of *AMH* variants and other women with PCOS. Collectively, about 6.7% of PCOS cases from these cohorts had one or more of the *AMH/AMHR2* rare variants (375, 376), which represents a substantial minority of cases, but the PCOS risk effect sizes and heritability attributable to these variants has not been estimated. Although AMH overexpression is a more consistent feature of PCOS (22, 377, 378), these rare variant studies indicate that the role of AMH in PCOS is more complex than previously thought and may vary between affected women. The specific mechanism through which impaired AMH signaling may lead to PCOS requires further study.

In a WGS study of 261 individuals from 62 families with one or more daughters with PCOS, we tested for gene-level associations of deleterious rare variants (MAF \leq 2%) with PCOS (386). In order to increase the study's power to detect PCOS-associated variants, we first tested each set of rare variants against PCOS-related quantitative traits (testosterone, DHEAS, insulin, glucose, LH, FSH, SHBG) and then combined the trait association results into a meta-statistic. The meta-statistic served to combine information from multiple association tests and thereby reduce the penalty for multiple hypothesis testing. We identified a collection of rare variants in the *DENND1A* gene that were significantly associated with altered levels of these reproductive and metabolic traits within the families. *DENND1A* was first implicated as a PCOS candidate gene in GWAS (269). Subsequently, it was shown to influence androgen production in ovarian theca cells according to the relative abundance of a particular

splicing isoform, *DENND1A.V2* (304). Most of the rare variants identified in our family study were not in LD with the *DENND1A* GWAS variants (117, 269, 270), which have no impact on splicing or apparent functional significance (304, 387), but the rare variants were predicted to significantly disrupt transcription factor binding and/or RNA-binding protein motifs. Thus far, targeted sequencing (304, 387, 388) and whole exome sequencing (389) have failed to identify any variants in *DENND1A* associated with PCOS or with *DENND1A.V2* expression, but these studies only examined limited regions of the gene and were in small case-control cohorts. Replication and functional studies are needed to confirm whether any of the rare *DENND1A* variants identified in the family study drive *DENND1A.V2* expression and/or PCOS risk. Notably, subjects from 50% of families in the family study had at least one of the rare *DENND1A* variants (386). These results support a model in which causal variants may be individually rare, but they are collectively common in certain disease genes.

No other gene associations reached genome-wide significance in the family-based meta-analysis, but 2 additional genes among the top 5 associations are established PCOS candidate genes: *C9orf3*, which is a PCOS GWAS gene (117, 269-271), and *BMP6*, which regulates folliculogenesis in granulosa cells (390) and is overexpressed in PCOS women (391). Taken together, these genomic sequencing studies reinforce the putative role these genes play in PCOS pathogenesis and suggest that family-specific variants affecting key genes may be largely responsible for PCOS risk.

Numerous studies of PCOS-related pathways have measured gene expression in various tissues under different conditions using RNA-seq. These studies have identified different gene networks and pathways that are disrupted in PCOS, including MAPK signaling (392, 393), androgen receptor signaling (34), metabolic processes (394), and inflammatory and immune responses (395, 396). In gene expression studies, however, it is difficult to discern what changes are adaptive effects of the disease as opposed to causal aberrations (397).

Table 4. Replicated phenome-wide associations with polygenic PCOS risk

Phenotype code	Discovery (n = 49 343)			Replication (n = 18 096)		
	Prevalence	OR (95% CI)	P	Prevalence	OR (95% CI)	P
Morbid obesity	0.18	1.010 (1.008–1.013)	9.74×10^{-18}	0.11	1.116 (1.054–1.182)	1.64×10^{-4}
Obesity	0.32	1.008 (1.006–1.009)	4.14×10^{-17}	0.20	1.087 (1.042–1.134)	1.29×10^{-4}
Overweight	0.37	1.007 (1.005–1.009)	2.20×10^{-16}	0.25	1.077 (1.037–1.120)	1.44×10^{-4}
T2D	0.25	1.007 (1.005–1.009)	8.18×10^{-13}	0.22	1.081 (1.036–1.128)	3.70×10^{-4}
Sleep apnea	0.16	1.008 (1.006–1.010)	4.71×10^{-12}	0.12	1.096 (1.036–1.158)	1.33×10^{-3}
Diabetes mellitus	0.26	1.007 (1.005–1.009)	5.39×10^{-12}	0.23	1.079 (1.035–1.125)	3.56×10^{-4}
Chronic liver disease and cirrhosis	0.11	1.008 (1.005–1.011)	4.17×10^{-9}	0.10	1.093 (1.028–1.163)	4.64×10^{-3}
Bariatric surgery	0.04	1.012 (1.008–1.016)	7.59×10^{-9}	0.02	1.202 (1.079–1.339)	8.00×10^{-4}
Obstructive sleep apnea	0.13	1.007 (1.005–1.010)	1.16×10^{-8}	0.09	1.098 (1.030–1.170)	3.98×10^{-3}
Other chronic nonalcoholic liver disease	0.11	1.008 (1.005–1.011)	2.13×10^{-8}	0.09	1.112 (1.042–1.187)	1.38×10^{-3}
Polycystic ovaries	0.02	1.015 (1.009–1.020)	3.16×10^{-7}	0.02	1.174 (1.026–1.343)	1.93×10^{-2}
Insulin pump user	0.09	1.008 (1.005–1.011)	2.25×10^{-6}	0.08	1.136 (1.060–1.219)	3.42×10^{-4}
T2D with ophthalmic manifestations	0.05	1.010 (1.005–1.014)	9.20×10^{-6}	0.03	1.221 (1.082–1.377)	1.20×10^{-3}

Replicated phenotype associations demonstrating a consistent direction of effect with PCOS polygenic risk score, according to Joo and colleagues (366). Patients without the phenotype code of interest but with clinically related phenotypes were excluded from consideration. Abbreviations: OR, odds ratio; PCOS, polycystic ovary syndrome; T2D, type 2 diabetes.

SNP/Analysis

OR [95% CI]

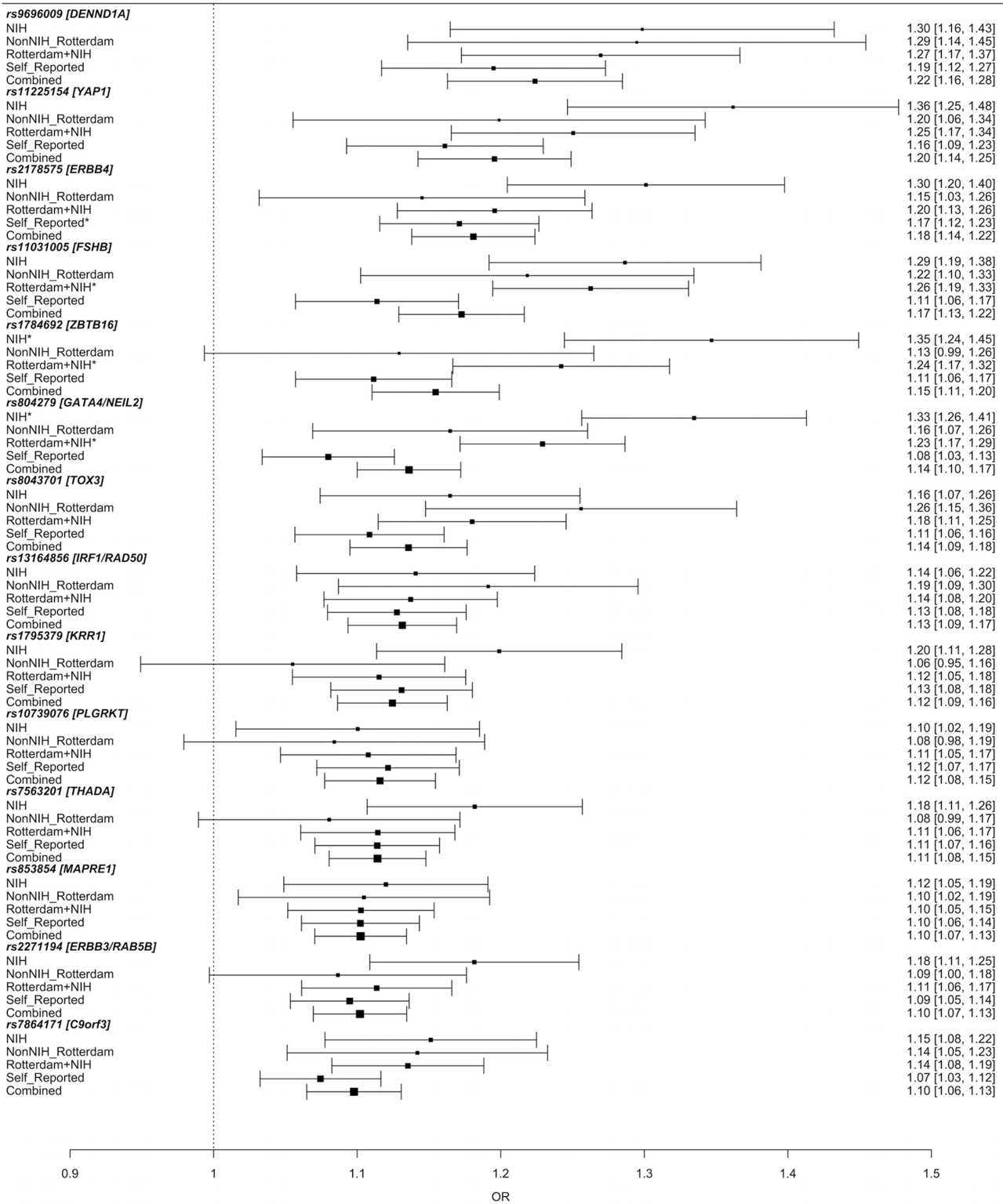


Figure 8. Odds ratio of PCOS as a function of diagnostic criteria applied. The odds ratios (OR) and 95% CI are shown for each significant GWAS locus from the PCOS meta-analysis, stratified by case definitions according to different diagnostic criteria. NIH: groups recruiting only NIH diagnostic criteria; NonNIH_Rotterdam: Rotterdam diagnostic criteria excluding the subset fulfilling NIH diagnostic criteria; Rotterdam+NIH: all groups except self-reported; self-reported: 23andMe. rs804279 at the *GATA4/NEIL2* locus demonstrated significant heterogeneity (Het $P = 2.6 \times 10^{-5}$). The * indicates statistically significant associations for PCOS. Abbreviations: NIH, National Institutes of Health; PCOS, polycystic ovary syndrome. Figure reproduced from Day et al, 2018 (117).

Investigators have also begun to map DNA methylation changes to gene expression changes in PCOS, in an effort to connect environmental exposures to transcriptional changes (398, 399).

Genetic Architecture of PCOS

Frequency and Function of Disease Variants

There is emerging consensus that complex traits are primarily driven by common noncoding variation (184, 400, 401). Around 90% of disease-associated GWAS variants are in noncoding regions of the genome (402, 403), and although lead GWAS SNPs are seldom causal (404), fine-mapping studies have confirmed that putatively causal SNPs in GWAS loci are significantly enriched in regulatory regions (405). Furthermore, most trait-associated variants either directly influence nearby gene expression or co-localize with other variants that influence expression of one or more genes (318).

As mentioned previously, common SNPs in GWAS loci typically account for only a small fraction of the estimated heritability for a given complex disease, an observation that came to be known as the problem of “missing heritability” (186). This mystery has since been largely resolved, as common SNPs collectively (not just those with significant association signals in GWAS) account for most of the remaining heritability in complex diseases (406–408). Accordingly, as the sample sizes of GWAS have increased, the number of significant associations discovered per study has increased proportionally (409). This widespread distribution of heritability—apart from the inherent complexity of biological networks—likely results from the fact that pathogenic variants in disease pathways are more strongly selected against via natural selection (183, 401). All of the most significant PCOS GWAS index variants have been located in noncoding regions, primarily intronic, and are common in their respective populations (117, 269–272). Based on these observations and findings across other complex traits, it is likely that the causal variants tagged by PCOS GWAS SNPs influence the expression of one or more nearby gene transcripts in one or more specific tissues, although some may affect more distant genes through “*trans*” effects (401).

Although rare variants may contribute less to the phenotypic variance of complex traits overall, the identification of disease-associated rare variants can point more directly to causal mechanisms and key disease pathways (410). In fact, gene expression variation attributable to nearby genetic variation is driven most strongly by rare variants (411). In our family-based WGS study of rare variants in PCOS, 30 of the 32 identified rare variants in *DENND1A* were noncoding, most of which were predicted to impact transcription factor and/or RNA-binding protein binding sites (386). Among the other top gene associations in that study were several other GWAS candidate genes. These results suggest that rare noncoding variants within GWAS loci likely contribute to PCOS risk, in addition to the common alleles tagged from GWAS. GWAS loci have been found to harbor independent trait-associated rare variants in other complex traits as well (412–414). Therefore, future studies in PCOS would likely benefit from considering both common and rare variants in their statistical models.

Nonsynonymous protein-coding variants, by altering the encoding of amino acids, can directly impact protein function

significantly. Coding variants, therefore, are more often deleterious and, consequently, are relatively rare in the genome (415). Disruptive mutations in protein-coding regions are typical of rare monogenic diseases, so-called Mendelian disorders. Such variants, which individually cause significant individual phenotypic consequences, are said to be highly penetrant. While common diseases are characterized by highly polygenic architectures driven by noncoding variants with small individual variant effects, protein-coding variants are nonetheless significantly enriched for contributing to complex traits (405). Indeed, coding variant associations have been validated in a number of complex phenotypes, including LDL-cholesterol levels (416–418), Alzheimer’s disease (419, 420), and even height (413, 421). Recent large-scale whole exome association studies have identified rare alleles with large effect sizes that are associated with BMI (422) and T2D (368).

Our sequencing studies of rare variants in the *AMH* and *AMHR2* genes identified 18 coding variants that impaired AMH-mediated signaling capacity and were collectively associated with PCOS in an independent cohort (375, 376). Missense mutations in other candidate genes have been reported in PCOS (388, 423–425), but functional and/or replication studies are needed to confirm their disease associations. As more candidate genes are sequenced, it is likely that additional rare coding variants contributing to PCOS will be identified. Although such variants contribute modestly to complex trait heritability (368, 426), they still may play a significant role in efforts to better understand PCOS disease etiology, because their biological consequences are more substantial and easier to interpret.

Similarly, identifying causal variants in extreme phenotypes can be an effective way of identifying candidate genes for corresponding complex traits. Genes linked to monogenic forms of complex traits are significantly enriched in corresponding complex trait GWAS gene sets (371). Genetic risk in Mendelian compared to complex diseases may simply be a function of penetrance determined by the nature of individual risk variants and the relative essentiality of their affected genes (427). For example, common variants in maturity-onset diabetes of the young (MODY) genes are associated with T2D (428–430). Therefore, studying familial genetics of extreme PCOS phenotypes could help identify genes and pathways commonly disrupted in PCOS more generally.

The genetic studies of PCOS indicate that common and rare variants, both coding and noncoding, contribute to PCOS pathogenesis. Although their relative contributions to PCOS heritability have yet to be precisely quantified, based on findings in other complex diseases, the majority of heritability likely comes from a highly polygenic network of common variants with small effect sizes, followed by the contribution of less frequent variants found in more central disease genes (184, 401). Collectively, 50% of families in our family-based WGS study had one or more of the rare variants identified in *DENND1A* (386). The high prevalence of hormone-associated rare variants in *DENND1A*, together with our earlier family studies in PCOS that observed a bimodal distribution of testosterone levels in sisters of women with PCOS compared to a unimodal distribution observed in control women (98), suggests that ovarian androgen biosynthesis is a central pathway in PCOS pathogenesis, with *DENND1A* as a key mediator. The discovery of functional rare variants that decrease AMH signaling implicates this

pathway in PCOS pathogenesis, as well (375, 376). The findings suggest that AMH is more than a biomarker for antral follicle number (431). Conceivably then, there is a spectrum of underlying genetic risk for PCOS among any given population driven by common variation, but heritability within individual families or certain subpopulations of women with PCOS may be driven more by rare alleles with larger effects.

Genetic Architecture of PCOS NIH and Rotterdam Phenotypes

The most appropriate diagnostic criteria for PCOS have always been controversial since they are all based on subjective assessment of which clinical features of the syndrome reflect core biologic derangements (102) (Table 1). The PCOS GWAS meta-analysis (117) had adequate statistical power to objectively assess for the first time whether there were genetic differences among the PCOS phenotypes defined by the NIH compared to the Rotterdam criteria. Specifically, cases defined by the NIH criteria, HA+OD ± PCOM, were compared to the non-NIH Rotterdam cases, HA+PCOM and OD+PCOM, as well as to self-reported cases. The 14 GWAS loci that were significantly associated with PCOS according to any case definition were tested for heterogeneity of effect sizes across the cases stratified by phenotype. Only one locus, near *GATA4/NEIL2*, showed significant evidence for heterogeneity between the different diagnostic criteria groups, being most strongly associated with the NIH phenotype (Fig. 8 [(117)]). Since the NIH phenotype is substantially more insulin resistant than the non-NIH Rotterdam phenotypes (7), this locus may be involved in pathways regulating insulin sensitivity. However, the absence of heterogeneity among the phenotypes for the other 13 loci suggests that the genetic architecture of these phenotypes is generally similar. These findings imply that the current diagnostic criteria do not identify biologically distinct phenotypes.

The PCOS GWAS meta-analysis (117) also examined associations for each of the lead PCOS susceptibility variants with the individual diagnostic features, PCOM, OD, HA, as well as with PCOS-related reproductive traits including ovarian volume and testosterone, LH, and FSH levels. For the 3 PCOS diagnostic features, 4 SNPs were associated with HA, 9 with OD, and 8 with PCOM. The *FSHB* locus contained the only SNP associated with LH and FSH levels, consistent with findings in previous European ancestry PCOS GWAS (271, 272). There was only 1 SNP, at the *IRF1/RAD50* locus, that was associated with testosterone levels. SNPs from 9 loci were associated with OD, 7 of these were also associated with PCOM. There was only 1 locus, near *THADA*, that was associated with PCOM and no other PCOS-related traits. There were no SNPs associated with ovarian volume. These findings suggest that the genetic architecture of PCOM and OD are similar. Taken together with the shared genetic architecture of NIH, non-NIH Rotterdam, and self-reported PCOS, these findings support the use of OD as a proxy for ultrasound assessment of ovarian morphology for the diagnosis of PCOS. Limitations of the PCOS-related trait analysis include the relatively small sample size with available data ($n = \sim 2000$ for PCOM and $n = \sim 3000$ for HA, OD, and quantitative hormonal traits) and the overlap between groups (117). Thus, it remains possible that additional unique association signals will be discovered as sample sizes increase.

Novel PCOS Subtypes

Genetic heterogeneity in a disease is the notion that different sets of biological aberrations in different individuals can lead to convergent phenotypes under the same clinical diagnosis. The inclusion of genetically discrete disorders within one disease cohort would significantly compromise the power of disease association studies and could account for missing heritability by masking true allele effects (432-434). More precise disease phenotyping resulting in more homogeneous study populations would not only empower the discovery of causal mechanisms (434) but would also lead to more effective care for patients along the path toward precision medicine (435).

PCOS is frequently referred to as a heterogeneous disorder (436, 437). It has been suggested that there are subtypes of PCOS, a leaner type with increased LH:FSH ratios and a heavier type with insulin resistance (438). However, there have been limited objective assessments of putative PCOS subtypes. Cluster analysis has been previously performed on PCOS quantitative traits (120, 439) but there has been no validation that the clusters thus identified were biologically relevant. In an effort to identify subpopulations of women with PCOS with distinct underlying genetic risk factors, we applied an unsupervised clustering approach (440). Using anthropometric, reproductive, and metabolic data from multiple, independent PCOS cohorts, we identified 2 reproducible subtypes that had distinct phenotypic characteristics: a “reproductive” group (23%) characterized by higher LH and SHBG levels with relatively low BMI and insulin levels, and a “metabolic” group (37%) characterized by high BMI, glucose, and insulin levels with lower LH and SHBG levels. The remaining cases were designated as “indeterminate” (40%) (Fig. 9). These clusters were additionally validated using bootstrap resampling (441).

Subsequent GWAS, using a data subset from our European PCOS GWAS (243), revealed novel loci significantly associated with each subtype; these loci included genes with putative functions in pathways relevant to PCOS (Fig. 10). One locus significantly associated with the reproductive subtype was located in the type-I AMH receptor *BMPRI1B* (bone morphogenetic protein receptor type 1B). The *BMPRI1B* receptor, which is highly expressed in granulosa cells and GnRH neurons (24), forms a heterodimer with transforming growth factor beta (TGF- β) type-II receptors, including *AMHR2*, and binds AMH and other BMP ligands to initialize TGF- β signaling (442). *BMPRI1B* regulates follicular development and mediate granulosa AMH response in sheep (443), and *BMPRI1B* deficiency led to infertility and functional ovarian defects in mice, including lower aromatase production in granulosa cells (444). *BMPRI1B* was also identified as a key driver in an androstenedione-related lncRNA-mRNA network in PCOS (445). Interestingly, in our family-based rare variant study (386), one of the *BMPRI1B* ligand genes, *BMP6*, had the third-strongest gene-level association with altered hormone levels, out of 339 genes. *BMP6* expression was significantly higher in granulosa cells from women with PCOS compared to those from healthy controls and significantly inhibited FSH-induced estradiol production (391). Similarly, the strongest association with the reproductive subtype was just upstream of and within the same topologically associated domain (TAD) as *PRDM2*, the gene with the fifth-strongest association signal in the family-based rare variant study. *PRDM2* is an estrogen receptor coactivator (446) highly expressed in the pituitary gland (447) and the ovary (448), and

its protein also binds with the retinoblastoma protein (449), which is known to play a role in ovarian granulosa cell development (450, 451). The one locus that was significantly associated with the metabolic subtype included a number of potential disease genes in the surrounding region, such as *GRB14*, *FIGN*, and *KCNH7* (440). The case sample sizes of the subtype GWAS were relatively small, however ($n = 207$ reproductive, 329 metabolic), so until these association results have been replicated, they should be considered preliminary.

Association testing on the indeterminate subtype replicated the *FSHB* locus association from the original PCOS GWAS (271). This result suggests that the indeterminate group captures relatively generic PCOS cases. However, the association signal with indeterminate PCOS cases was actually stronger than in the larger, original GWAS (271), suggesting that the indeterminate group was more genetically homogenous after the reproductive and metabolic subtypes were removed. It is possible that the indeterminate subtype or some other undefined subset is associated with characteristics that were not included in the clustering analysis, such as AMH levels and/or follicle numbers (55). Regardless, the association between the *FSHB* locus and the indeterminate cluster indicates that the locus is less likely to drive the forms of PCOS associated with the reproductive or metabolic subtypes.

The reproductive and metabolic clusters identified in the subtyping study appear to capture the opposite ends of the PCOS phenotypic spectrum. The largest dimension of phenotypic variation between women with PCOS was an axis with high SHBG levels on one end and high BMI and fasting insulin levels on the other, just as in a previous analysis of PCOS phenotypic variation by Dewailly and colleagues (55). The reproductive and metabolic subtypes clustered on either end of this axis, respectively (Fig. 9). Overlaying these results with the pathophysiology of PCOS, as illustrated in Fig. 1, the reproductive subtype aligns with LH-dependent androgen production resulting from aberrations in the hypothalamic-pituitary-gonadal axis, whereas the metabolic subtype is consistent with hyperandrogenemia driven primarily by insulin resistance. This concept of distinct causal pathways underlying the subtypes is supported by the fact that PCOS GWAS loci appear to confer risk for PCOS through either reproductive or metabolic mechanisms (452). Indeed, Dewailly and colleagues found that metabolic-, androgen-, and follicle-related variables were all independently correlated with high total testosterone levels in women with PCOS (55). Moreover, affected sisters tended to be concordant for either reproductive or metabolic subtypes. Furthermore, the carriers of the rare variants reported in *DENND1A* (386), a gene that regulates androgen biosynthesis, were significantly more likely to be classified with the reproductive subtype of PCOS (440). In contrast to the PCOS diagnostic criteria that do not result in genetically discrete phenotypes (117), these reproductive and metabolic subtypes appear to identify distinct genetic architectures and therefore promise to be more biologically relevant to PCOS.

Environmental Contributors to PCOS

Obesity

Obesity is an obvious environmentally determined, putative contributor to PCOS. It is a common feature of PCOS with prevalence rates as high as 80% to 90% in affected women in the United States (453, 454). However, there is clearly a

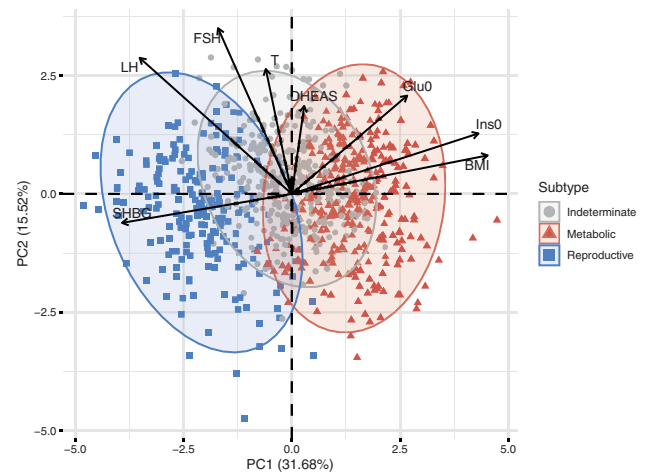


Figure 9. PCA plot of novel PCOS clusters. Clustered PCOS cases are plotted on the first 2 PCs of adjusted quantitative trait data, colored according to their identified subtype with 95% concentration ellipses. The relative magnitude and direction of trait correlations with the PCs are shown with black arrows. Abbreviations: BMI, body mass index; DHEAS, dehydroepiandrosterone sulfate; FSH, follicle-stimulating hormone; Glu0, fasting glucose; Ins0, fasting insulin; LH, luteinizing hormone; PC, principal component; PCA, principal component analysis; PCOS, polycystic ovary syndrome; SHBG, sex hormone-binding globulin; T, testosterone. Figure reproduced from Dapas et al, 2020 (440).

large population of women with PCOS who are not obese. Furthermore, the prevalence rates of PCOS are similar in diverse populations with varying prevalence rates of obesity (124, 127, 138, 455). Additionally, despite increasing prevalence rates of obesity worldwide, the prevalence of PCOS has remained relatively stable (138). Obesity exacerbates androgenic symptoms (456, 457), chronic anovulation (458), insulin resistance (67, 68) and dysglycemia (340, 341). Accordingly, women who present for medical evaluation of PCOS are more likely to be obese (459).

The strongest obesity susceptibility gene, *FTO* (460), is associated with PCOS (461, 462), although it has not appeared as a genome-wide significant PCOS susceptibility locus in GWAS conducted to date (117, 269). Its association with PCOS appears to be mediated through BMI. *FTO* and other robustly replicated BMI-increasing alleles are not associated with PCOS independent of BMI (463), but recent Mendelian randomization analyses indicate that genetically determined BMI is significantly associated with PCOS (117, 136, 137, 358). Conversely, genetically determined PCOS is not associated with BMI (136, 137). Taken together, the genetic data support the hypothesis that obesity does contribute to the development of PCOS.

Developmental Origins of PCOS

It is now widely recognized that environmental factors acting in early life can contribute to the development of a number of adult chronic diseases (464-466), a field known as the *developmental origins of health and disease* (DOHaD). There is compelling evidence that suggests early-life exposures contribute to PCOS risk, as well. The co-occurrence of PCOS in DZ twins is greater than half the rate observed in MZ twins (165), which is indicative of shared environmental influences. Furthermore, maternal parent-of-origin effects on PCOS phenotype were found to be much stronger than paternal effects (467), suggesting that the intrauterine environment may play a role in PCOS.

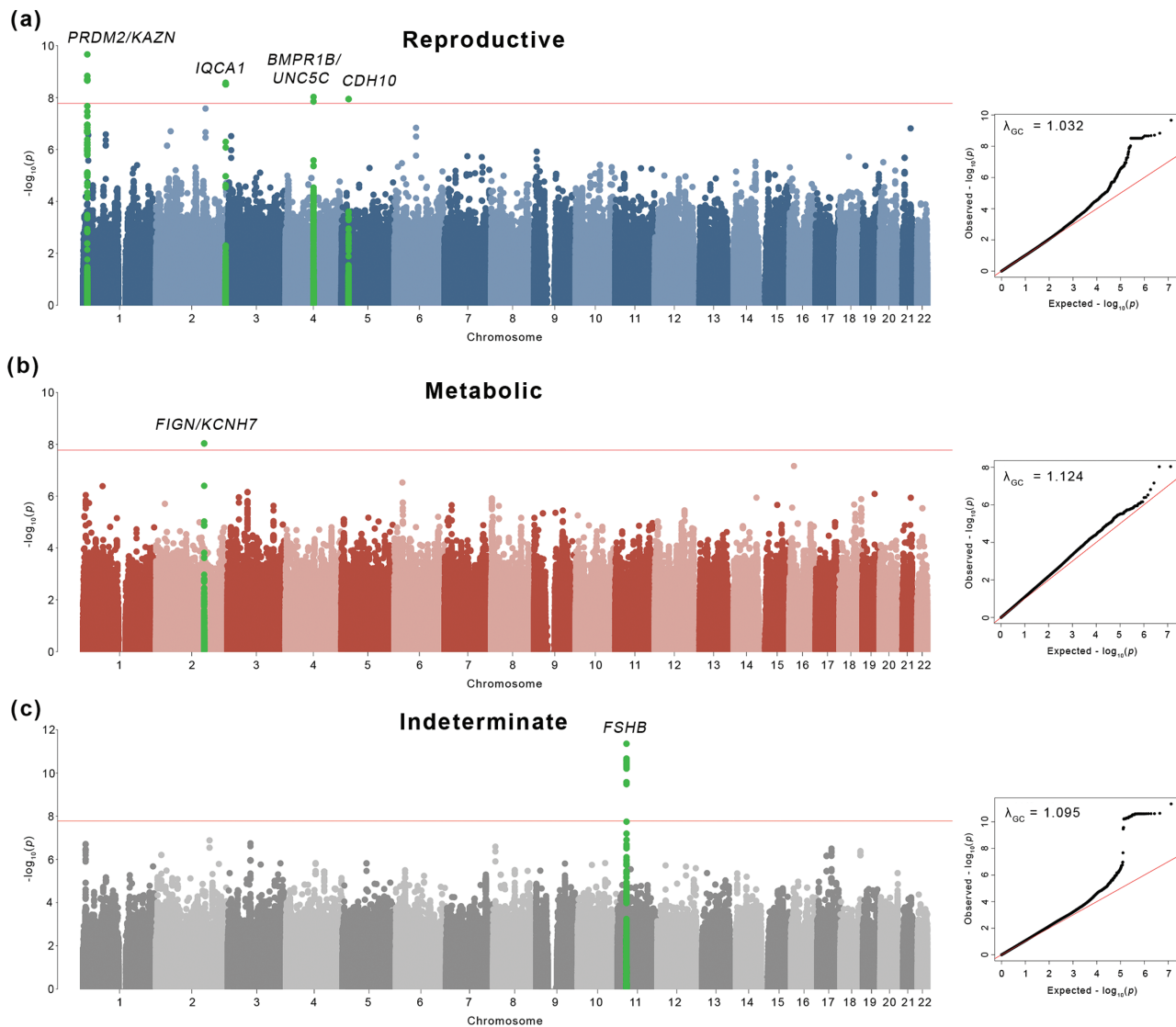


Figure 10. Novel PCOS cluster GWAS results. Manhattan plots for (a) reproductive, (b) metabolic, and (c) indeterminate PCOS subtypes. The red horizontal line indicates genome-wide significance ($P \leq 1.67 \times 10^{-8}$). Variants proximal to genome-wide significant loci (± 200 kb) are colored in green and labeled according to nearby gene(s). Quantile–quantile plots with genomic inflation factor, λ_{GC} , are shown adjacent to corresponding Manhattan plots. Abbreviations: GWAS, genome-wide association studies; PCOS, polycystic ovary syndrome. Figure reproduced from Dapas et al, 2020 (440).

Based on epidemiologic studies, Barker first proposed the thrifty phenotype hypothesis that fetal undernutrition results in adult cardiometabolic disorders (466, 468). Low birth weight is a consequence of poor fetal nutrition and is highly correlated with adult cardiovascular disease (469). Ibáñez and colleagues proposed that low birth weight was associated with the later development of premature pubarche and PCOS (470, 471). However, studies examining the association between birth weight and PCOS have yielded conflicting results. There was no association between low birth weight and the development of PCOS in a number of studies (472–477), whereas others found such an association (478, 479). The reasons for these discrepant findings are unclear but ethnic/racial differences could account for some of the heterogeneity in study outcomes (480–482).

There is considerably more evidence to support the role of another extrinsic factor, intrauterine androgens, in the pathogenesis of PCOS. Abbott, Dumesic, and colleagues reported that female rhesus monkeys who had been exposed

to testosterone *in utero* developed both reproductive and metabolic features of PCOS after puberty (483–488). Similar phenocopies of PCOS could be created in other animal species, including sheep (489–491) and rodents (492–494), by prenatal androgen exposure. These observations have led to the hypothesis that intrauterine exposure to androgens is a final common path for the development of PCOS (495). Maternal circulating androgen levels are increased in pregnant women with PCOS (496). Maternal PCOS was associated with greater risk for developing androgen-associated neuropsychiatric disorders, after accounting for genetic relatedness (497). However, given the tremendous capacity of the placenta to aromatize androgens (498–500), the source of intrauterine androgens in PCOS could be fetal rather than maternal. The fetal adrenal is steroidogenically active and an important source of fetal androgens (501). The fetal ovary is minimally steroidogenically active under normal circumstances (501). Nevertheless, it has the steroidogenic enzyme capacity for androgen biosynthesis (502, 503). Therefore, it is

biologically plausible that there is ovarian as well as adrenal androgen secretion by the female fetus.

Based on our findings that hyperandrogenemia was a consistent PCOS endophenotype (87, 98, 154, 155), we proposed that variation in a gene regulating androgen biosynthesis could contribute to PCOS (98), perhaps by causing intra-uterine androgen excess (504). The discovery that the PCOS GWAS candidate gene, *DENND1A*, is a key regulator of androgen biosynthesis (304) and that rare variants in it are present in ~50% of families with PCOS (386) supports this hypothesis. Moreover, a recent Mendelian randomization found that genetically determined total and bioavailable testosterone levels were causally associated with PCOS (171).

Tata and colleagues (25) have proposed an alternative mechanism for prenatal androgen excess. They demonstrated that AMH administration to pregnant mice resulted in increased maternal LH-dependent testosterone secretion and a PCOS phenotype in adult female offspring. These offspring have hyperactivated GnRH neurons, an established extragonadal action of AMH (24). The GnRH hyperactivation was normalized with intermittent GnRH antagonist treatment. They found that AMH levels were increased during pregnancy in women with PCOS, supporting the potential physiologic relevance of AMH-mediated increased maternal LH secretion. However, our findings that PCOS-specific rare genetic variants in *AMH* (375) or in *AMHR2* (376) reduce signaling through this pathway in a substantial minority of cases suggest that there may be different AMH-mediated pathways that contribute to PCOS (505).

Nevertheless, human studies attempting to find evidence for prenatal androgen exposure have provided conflicting results. Cord blood androgen levels from female infants of women with PCOS have been reported to be increased (506, 507), decreased (476, 501, 508), or unchanged (509). Differences in steroid assay methodology may have contributed to these discrepant results, with studies employing the gold standard method, liquid chromatography with mass spectrometry, failing to find increased androgen levels (476, 509). Further, cord blood androgens levels did not predict the development of PCOS in a prospective cohort study (510). However, cord blood steroid levels at birth may not reflect those levels in the second trimester during the putative period for androgen programming of PCOS (495). A lower ratio of the index (digit 2) to ring (digit 4) finger (511, 512) or a longer anogenital distance (513) provide anatomical evidence for prenatal androgen exposure. Digit ratio studies in PCOS have been conflicting, with reports of lower ratios (514) and unchanged ratios (515). Similarly, there have been conflicting reports regarding anogenital distance in PCOS, with several studies finding increased anogenital distance in adult women with PCOS (516-518) as well as in their daughters (518). However, another study in infant daughters of women with PCOS failed to find differences in anogenital distance compared to control girls (519).

Epigenetic Inheritance

Gene expression is controlled by mechanisms that alter the local structure of DNA folding within chromosomes and/or recruit certain proteins that promote or repress transcription. Such modifications that persist through mitosis are referred to as *epigenetic* and contribute to trait variance beyond genetic differences. Two of the most common epigenetic mechanisms are DNA methylation, in which methyl groups are added

directly to certain DNA nucleotides leading to gene silencing (520), and histone modifications, where chemical modifications to histone tails can distort local chromatin structure or recruit remodeling enzymes (521). Epigenetic modifications are the mechanistic link between nature and nurture, because epigenetic states are shaped by both intrinsic genetic variants and extrinsic environmental exposures (522). Epigenetic modifications are how early-life experiences can lead to specific phenotypic changes in adulthood (522). Furthermore, epigenetic programming is often disrupted in disease (520, 523). Epigenetic markers are globally erased in primordial germ cells and reset following fertilization to allow for organism development (524). However, an accumulation of studies, particularly in rodents, have demonstrated that epigenetic modifications can persist across multiple generations (525), whether through incomplete developmental reprogramming or through transgenerational fetal programming.

Evidence for transgenerational epigenetic, non-DNA-sequence-based inheritance in PCOS, such as persistent DNA methylation patterns, could connect numerous findings concerning environmental influences, developmental origins, and unexplained heritability in PCOS pathogenesis (526, 527). It is possible the epigenetic transgenerational changes contribute to some of the unexplained heritability of PCOS (526, 528). Correlation of DNA methylation between MZ twins is greater than between DZ twins, but most of the difference can be attributed to additive genetic influences (529-531). Epigenetic modifications caused by genetic variation would inherently be accounted for in heritability estimates from both twin studies and GWAS, and therefore would not contribute to unexplained heritability. However, epigenetic effects divorced from additive genetic variation at the population level would not be quantified in GWAS-based heritability estimates, and therefore could contribute to unexplained heritability (532).

Several studies have provided evidence for physiologically relevant epigenetic changes in women with PCOS by demonstrating alterations in gene expression associated with methylation changes in granulosa cells (395, 533), adipose tissue (534), and skeletal muscle (535). There were also differences in DNA methylation patterns in a pilot study of cord blood from infants of women with PCOS compared to that from control women (536). Particularly compelling evidence for an epigenetic contribution to PCOS heritability mediated by *in utero* exposures has been provided recently by complementary studies examining transgenerational effects of prenatally exposed murine PCOS models (528, 537). These studies identified persistent PCOS-like transgenerational effects in the third generation of offspring from mice either prenatally androgenized by dihydrotestosterone (528) or exposed to AMH (537). These changes included persistent alterations in gene expression across generations. Parallel human studies supported these findings by demonstrating that some of the same genes were differentially methylated (538) or expressed (528) in cross-sectional PCOS case-control cohorts and in daughters of women with PCOS. It is noteworthy that the genes with altered expression, as well as their associated Gene Ontology terms (539), differed between the 2 studies, although elevated prenatal androgen levels were a common feature presumably mediating expression changes in both models (25, 538).

These findings should be considered in the context of the robustly replicated evidence for substantial genetic contributions

to PCOS. The increased risk of diagnosis in daughters of mothers with PCOS observed by Risal and colleagues (528) is consistent with genetic inheritance. Sisters of women with PCOS are more likely to be affected than their premenopausal mothers (540), MZ twins are much more likely to both have PCOS than DZ twins (165), and the rate of disease concordance between DZ twins and non-twin sisters is similar (165, 540), all of which indicates that intrauterine exposure does not drive PCOS inheritance in humans. Family studies have also shown that there are paternal effects on PCOS risk and phenotype (541, 542), which is inconsistent with an intrauterine exposure model. It is possible that paternal effects can be transmitted via epigenetic inheritance (543), but it is likely that perceived inheritance of DNA methylation in mammals is primarily driven by genetics and intrauterine exposure rather than incomplete erasure of epigenetic signatures during developmental reprogramming (544). Therefore, rather than being exclusive mechanisms, it is more plausible that genetically determined core pathways, eg, testosterone biosynthesis (98, 171, 386, 401), act in concert with epigenetic actions of these androgens to produce PCOS phenotypes. This hypothesis aligns with the current understanding of complex traits, in which heritability is predominantly driven by numerous common variants with small effects (401). Collectively, these findings emphasize the need to better understand how both genetic and environmental factors lead to specific epigenetic changes that can increase risk for developing PCOS.

Summary and Future Directions

In less than 10 years, modern genetic analyses have confirmed an important contribution of genetic variation to the pathogenesis of PCOS. GWAS have identified numerous common risk alleles, implicating several plausible etiologic pathways related to neuroendocrine, reproductive, and metabolic function. Mendelian randomization studies have used genetic data to support epidemiologic findings linking BMI, insulin, age at menopause, depression, SHBG, and male pattern balding to PCOS risk. LD score regression analyses have indicated that PCOS shares genetic architecture with T2D, coronary artery disease, BMI, insulin levels, HDL levels, triglyceride levels, depression, and age at menarche. Candidate gene sequencing and functional studies have identified rare coding and noncoding variants in *AMH* and *AMHR2* in a substantial minority of European ancestry PCOS cases. Whole genome sequencing studies have found rare, mainly noncoding, hormone-associated variants in *DENND1A* in ~50% of families with PCOS. Genetic subtyping has indicated there are reproductive and metabolic subtypes of PCOS that appear to have distinct genetic architectures. Meanwhile, animal studies have demonstrated that epigenetic changes incurred by environmental exposures *in utero* can perpetuate PCOS phenotypes across multiple generations, offering a possible mechanism for genetic and/or non-genetic familial PCOS risk.

Taken together with prevailing models in complex trait genetics (183, 184, 545), results from recent PCOS genetics studies have yielded important insights that should inform future directions of research. PCOS GWAS have identified many risk loci (Table 3), but it is important to emphasize that GWAS SNPs are rarely the disease-causing variants themselves but rather are tagging regions of the genome containing the actual causal variants (220). Extensive fine-mapping studies with subsequent functional analyses will be needed

to determine how genetic variation within these loci contributes to the development of PCOS. Indeed, determining the biological significance of genetic variation remains a central challenge in elucidating complex traits such as PCOS (220).

There is a pressing need to expand genetic analyses to PCOS cohorts of diverse racial and ethnic backgrounds. Because genetic variation is population-specific, risk loci identified in genetic association studies do not necessarily apply to individuals with a different ancestry (Table 3). Moreover, for shared risk loci, individual risk variants are likely to vary across populations. For example, the top risk SNPs differ between European and Han Chinese PCOS susceptibility loci (117, 546, 547). In general, the proportion of shared variants between 2 populations is a function of time since their divergence, and common variants tend to be older (548). Accordingly, there is common genetic variation shared between the 2 major PCOS ancestry groups studied to date, European and Han Chinese (117, 271, 549, 550), but rare genetic variation appears to be population-specific (551). Multiethnic studies can utilize population information for fine mapping causal variants (552-554). Ancestral differences in LD structure in shared risk loci can also help to pinpoint the location of causal variants (555). For example, LD blocks among individuals with African ancestry tend to be much smaller, as human migration across the rest of the world was associated with a corresponding population bottleneck and reduction in genetic diversity (556). Importantly, performing genetic studies in minority populations would not only help researchers to better understand PCOS pathogenesis, but it would also help ensure that these historically understudied populations will be more likely to share in the potential benefits of PCOS genetics research in the future.

Results from whole genome sequencing studies indicate that set-based rare variant association tests in larger PCOS cohorts could be used to discover more functional variants. Family-based cohorts can be more effective for identifying rare variants relative to population-based cohorts (205), because families are enriched for the same variants and Mendelian inheritance filters can further refine the selection of candidate variants (206, 220, 386). However, targeted sequencing of candidate genes may be a more cost-effective and reliable approach for identifying pathogenic rare variants (557), as disease risk genes harbor both common and rare risk variants (183, 545).

Methods for both fine mapping and rare variant association testing typically utilize information from large databases that combine association study results and functional data to predict the relative effects of different variants (318, 558). For example, by aggregating data from numerous ChIP-seq experiments in various tissues (559), variant effects on transcription factor binding can subsequently be predicted “*in silico*” (560). Regulatory variant effects are tissue- and cell-type-specific (561), however, so the relevance of available genome annotation data for PCOS studies may depend on the particular gene(s) under investigation. Moreover, specific cell populations within tissues are often of pathophysiologic importance to PCOS, eg, theca cells in the ovary, gonadotropes in the pituitary or GnRH neurons in the arcuate nucleus of the hypothalamus; whole-tissue gene expression may not be sufficiently sensitive to detect changes at the cellular level. Single-cell RNA sequencing could be used to distinguish cell-type-specific effects (562) but is likewise dependent on the ability to procure relevant human tissue samples.

Genetic analyses have also enabled the first objective assessment of the PCOS diagnostic criteria. The criterion of PCOM was added to PCOS diagnosis with the Rotterdam criteria (96, 97). However, in the PCOS meta-analysis of more than 10 000 cases (117), the genetic architecture was similar among nonoverlapping PCOS cases defined by NIH criteria, non-NIH Rotterdam criteria, or by self-report for 13 of 14 susceptibility loci. In the PCOS-related trait analysis, PCOM was genetically similar to OD for 7 of the 8 loci with which it was associated; there was only 1 locus associated solely with PCOM (117). Taken together, these findings suggest that assessment of PCOM is not needed for the diagnosis of PCOS. In contrast, unsupervised clustering identified reproductive and metabolic subtypes that appeared to have distinct genetic architectures (440). Although these findings require replication, they provide an example of modern disease classification based on objective biologic differences (94).

Current genetic risk scores for PCOS derived from genome-wide significant PCOS susceptibility loci (117) have limited predictive power given the small effect sizes of common genetic variants and the fact that variants below the genome-wide significance threshold collectively contribute greatly to disease risk (184, 563). It remains to be determined whether expanded polygenic risk scores, or subtype-specific risk scores, or rare variants contributing to PCOS pathogenesis (375, 376, 386) have greater predictive power. Integrated risk models that consider genetic and clinical factors may prove most effective for risk stratification (564-566).

The extensive body of animal data provides compelling support for the role of intrauterine androgen programming in PCOS pathogenesis with recent murine models demonstrating evidence for transgenerational androgen effects (528, 537). These mechanisms likely act in concert with genetic variation, leading to intrauterine androgen excess that contributes to the development of PCOS through epigenetic effects *in utero*. Numerous studies have reported differentially methylated and/or expressed genes in PCOS, but few have tied these changes to specific genetic variants. Jones and colleagues (398) found variants at PCOS loci that were associated with adjacent, differentially methylated regions in adipose tissue. Another study by Makrinou and colleagues (567) identified sites at PCOS risk loci that were differentially methylated in granulosa lutein cells of women with PCOS. Neither of these studies, however, could attribute any epigenetic differences to PCOS risk variants. Pau and colleagues (452) identified differentially expressed genes in adipose tissue associated with specific PCOS risk genotypes, including *BLK* and *NEIL2* at the *NEIL2/GATA4* locus, and *GLIPR1* and *PHLDA1* at the *KRR1* locus (Table 3), but the intermediary mechanisms remain unknown. Therefore, in an effort to connect the parallel avenues of research exploring genetic and epigenetic risk factors, future studies should strive to identify risk variants that associate with the epigenetic changes observed in PCOS.

While many opportunities remain for future research in PCOS genetics, we can now propose an updated model of PCOS pathogenesis that incorporates findings from the genetic studies reviewed herein. Because the characteristic hormonal disruptions in PCOS self-propagate in a feedback loop along the hypothalamic-pituitary-gonadal axis (Fig. 1), the underlying causes of this syndrome can vary in their tissues and pathways of origin and nonetheless result in the same PCOS phenotype. This notion has been supported empirically

by PCOS GWAS, which have implicated genes in neuroendocrine, metabolic, and reproductive pathways (117, 269-272, 452). The different phenotypes described by PCOS diagnostic criteria do not appear to capture this genetic heterogeneity (117), but by characterizing women by their relative hormonal profiles using machine learning methods (440), we can identify subsets of women with PCOS rooted in one or another of these core pathways. These subsets include reproductive and metabolic subtypes with unique genetic associations. Rare genetic variation in *DENND1A*, *AMH*, and *AMHR2* implicate androgen biosynthesis and AMH signaling as 2 of the core reproductive pathways in PCOS pathogenesis. Similarly, certain environmental risk factors are likely to be more or less relevant to these different forms of PCOS, as evidenced by the specific phenotypes that result from different *in utero* exposures (25, 495, 538). Ultimately, these relative genetic and environmental contributions occur along a spectrum. It is likely that many women with PCOS have genetic risk alleles from multiple core pathways. There are likely genetic risk factors that transcend subtypes, as well.

PCOS can no longer be dismissed as a perplexing vicious cycle of hormonal disturbances (16). Its phenotypic and genetic heterogeneity clearly indicate the need to shift away from PCOS diagnosis based on expert opinion to criteria based on biologic mechanisms. By deconstructing the syndrome into its core components, genomics is leading the transition toward precision medicine for PCOS. Ongoing genetic analyses promise to elucidate the distinct etiologies of PCOS, enabling the development of targeted therapies to reverse and ultimately prevent the development of the syndrome. Affected women and their families, who have been woefully underserved by the medical community (568, 569), may at last look forward to reaping the benefits of precision medicine in predicting, treating, and preventing PCOS.

Financial Support

National Institutes of Health funding sources: P50 HD044405, R01 HD085227, and R01 HD100812 (to A.D.); 5TL1TR002388 (to M.D.)

Disclosures

The authors have nothing to disclose.

References

- Chereau A. *Mémoires pour servir à l'étude des malades ovaires*. Paris: Fortin, Masson & Cie; 1844.
- Goldzieher J. Historical perspectives. In: Chang RJ, Heindel J, Dunaif A, eds. *Polycystic Ovary Syndrome*. New York: Marcel Dekker, Inc; 2002:1-14.
- Stein I, Leventhal M. Amenorrhea associated with bilateral polycystic ovaries. *Am J Obstet Gynecol*. 1935;29:181-191.
- Keettel WC, Bradbury JT, Stoddard FJ. Observations on the polycystic ovary syndrome. *Am J Obstet Gynecol*. 1957;73(5):954-962.
- Goldzieher JW, Green JA. The polycystic ovary. I. Clinical and histologic features. *J Clin Endocrinol Metab*. 1962;22:325-338.
- Dunaif A, Givens JR, Haseltine FP, Merriam GR. *Polycystic Ovary Syndrome*. Cambridge, MA: Blackwell Scientific Publications, Inc; 1992.
- Diamanti-Kandaraki E, Dunaif A. Insulin resistance and the polycystic ovary syndrome revisited: an update on mechanisms and implications. *Endocr Rev*. 2012;33(6):981-1030.

8. Burghen GA, Givens JR, Kitabchi AE. Correlation of hyperandrogenism with hyperinsulinism in polycystic ovarian disease. *J Clin Endocrinol Metab.* 1980;50(1):113-116.
9. Dunaif A, Hoffman AR, Scully RE, et al. Clinical, biochemical, and ovarian morphologic features in women with acanthosis nigricans and masculinization. *Obstet Gynecol.* 1985;66(4):545-552.
10. Flier JS, Eastman RC, Minaker KL, Matteson D, Rowe JW. Acanthosis nigricans in obese women with hyperandrogenism. Characterization of an insulin-resistant state distinct from the type A and B syndromes. *Diabetes.* 1985;34(2):101-107.
11. Semple RK, Savage DB, Cochran EK, Gorden P, O'Rahilly S. Genetic syndromes of severe insulin resistance. *Endocr Rev.* 2011;32(4):498-514.
12. Legro RS, Spielman R, Urbanek M, Driscoll D, Strauss JF 3rd, Dunaif A. Phenotype and genotype in polycystic ovary syndrome. *Recent Prog Horm Res.* 1998;53:217-256.
13. Rebar R, Judd HL, Yen SS, Rakoff J, Vandenberg G, Naftolin F. Characterization of the inappropriate gonadotropin secretion in polycystic ovary syndrome. *J Clin Invest.* 1976;57(5):1320-1329.
14. Yen SS, Vela P, Rankin J. Inappropriate secretion of follicle-stimulating hormone and luteinizing hormone in polycystic ovarian disease. *J Clin Endocrinol Metab.* 1970;30(4):435-442.
15. Taylor AE, McCourt B, Martin KA, et al. Determinants of abnormal gonadotropin secretion in clinically defined women with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 1997;82(7):2248-2256.
16. Dunaif A. Perspectives in polycystic ovary syndrome: from hair to eternity. *J Clin Endocrinol Metab.* 2016;101(3):759-768.
17. Wild RA, Umstot ES, Andersen RN, Ranney GB, Givens JR. Androgen parameters and their correlation with body weight in one hundred thirty-eight women thought to have hyperandrogenism. *Am J Obstet Gynecol.* 1983;146(6):602-606.
18. Azziz R, Carmina E, Dewailly D, et al.; Androgen Excess Society. Positions statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an Androgen Excess Society guideline. *J Clin Endocrinol Metab.* 2006;91(11):4237-4245.
19. Azziz R, Carmina E, Dewailly D, et al.; Task Force on the Phenotype of the Polycystic Ovary Syndrome of The Androgen Excess and PCOS Society. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. *Fertil Steril.* 2009;91(2):456-488.
20. Baird DT, Corker CS, Davidson DW, Hunter WM, Michie EA, Van Look PF. Pituitary-ovarian relationships in polycystic ovary syndrome. *J Clin Endocrinol Metab.* 1977;45(4):798-801.
21. Dunaif A. Insulin resistance and the polycystic ovary syndrome: mechanism and implications for pathogenesis. *Endocr Rev.* 1997;18(6):774-800.
22. Dewailly D, Barbotin AL, Dumont A, Catteau-Jonard S, Robin G. Role of anti-Müllerian hormone in the pathogenesis of polycystic ovary syndrome. *Front Endocrinol (Lausanne).* 2020;11:641.
23. Dumont A, Robin G, Catteau-Jonard S, Dewailly D. Role of anti-Müllerian hormone in pathophysiology, diagnosis and treatment of polycystic ovary syndrome: a review. *Reprod Biol Endocrinol.* 2015;13:137.
24. Cimino I, Casoni F, Liu X, et al. Novel role for anti-Müllerian hormone in the regulation of GnRH neuron excitability and hormone secretion. *Nat Commun.* 2016;7:10055.
25. Tata B, Mimouni NEH, Barbotin AL, et al. Elevated prenatal anti-Müllerian hormone reprograms the fetus and induces polycystic ovary syndrome in adulthood. *Nat Med.* 2018;24(6):834-846.
26. Waldstreicher J, Santoro NF, Hall JE, Filicori M, Crowley WF Jr. Hyperfunction of the hypothalamic-pituitary axis in women with polycystic ovarian disease: indirect evidence for partial gonadotroph desensitization. *J Clin Endocrinol Metab.* 1988;66(1):165-172.
27. Silva MSB, Giacobini P. New insights into anti-Müllerian hormone role in the hypothalamic-pituitary-gonadal axis and neuroendocrine development. *Cell Mol Life Sci.* 2021;78(1):1-16.
28. Marshall JC, Eagleson CA. Neuroendocrine aspects of polycystic ovary syndrome. *Endocrinol Metab Clin North Am.* 1999;28(2):295-324.
29. Eagleson CA, Gingrich MB, Pastor CL, et al. Polycystic ovarian syndrome: evidence that flutamide restores sensitivity of the gonadotropin-releasing hormone pulse generator to inhibition by estradiol and progesterone. *J Clin Endocrinol Metab.* 2000;85(11):4047-4052.
30. Coyle C, Campbell RE. Pathological pulses in PCOS. *Mol Cell Endocrinol.* 2019;498:110561.
31. Ehrmann DA, Barnes RB, Rosenfield RL. Polycystic ovary syndrome as a form of functional ovarian hyperandrogenism due to dysregulation of androgen secretion. *Endocr Rev.* 1995;16(3):322-353.
32. Franks S, Mason H, Willis D. Follicular dynamics in the polycystic ovary syndrome. *Mol Cell Endocrinol.* 2000;163(1-2):49-52.
33. Franks S, Hardy K. What causes anovulation in PCOS? *Curr Opin Endocr Metab Res.* 2020;12:59-65.
34. Wang F, Pan J, Liu Y, et al. Alternative splicing of the androgen receptor in polycystic ovary syndrome. *Proc Natl Acad Sci U S A.* 2015;112(15):4743-4748.
35. Nelson VL, Legro RS, Strauss JF 3rd, McAllister JM. Augmented androgen production is a stable steroidogenic phenotype of propagated theca cells from polycystic ovaries. *Mol Endocrinol.* 1999;13(6):946-957.
36. Rosenfield RL, Ehrmann DA. The pathogenesis of polycystic ovary syndrome (PCOS): the hypothesis of PCOS as functional ovarian hyperandrogenism revisited. *Endocr Rev.* 2016;37(5):467-520.
37. Lachelin GC, Barnett M, Hopper BR, Brink G, Yen SS. Adrenal function in normal women and women with the polycystic ovary syndrome. *J Clin Endocrinol Metab.* 1979;49(6):892-898.
38. Hughesdon PE. Morphology and morphogenesis of the Stein-Leventhal ovary and of so-called "hyperthecosis". *Obstet Gynecol Surv.* 1982;37(2):59-77.
39. Gilling-Smith C, Willis DS, Beard RW, Franks S. Hypersecretion of androstenedione by isolated thecal cells from polycystic ovaries. *J Clin Endocrinol Metab.* 1994;79(4):1158-1165.
40. Nelson VL, Qin KN, Rosenfield RL, et al. The biochemical basis for increased testosterone production in theca cells propagated from patients with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2001;86(12):5925-5933.
41. Webber LJ, Stubbs S, Stark J, et al. Formation and early development of follicles in the polycystic ovary. *Lancet.* 2003;362(9389):1017-1021.
42. Webber LJ, Stubbs SA, Stark J, et al. Prolonged survival in culture of preantral follicles from polycystic ovaries. *J Clin Endocrinol Metab.* 2007;92(5):1975-1978.
43. Franks S, Stark J, Hardy K. Follicle dynamics and anovulation in polycystic ovary syndrome. *Hum Reprod Update.* 2008;14(4):367-378.
44. Mason HD, Willis DS, Beard RW, Winston RM, Margara R, Franks S. Estradiol production by granulosa cells of normal and polycystic ovaries: relationship to menstrual cycle history and concentrations of gonadotropins and sex steroids in follicular fluid. *J Clin Endocrinol Metab.* 1994;79(5):1355-1360.
45. Willis DS, Watson H, Mason HD, Galea R, Brincat M, Franks S. Premature response to luteinizing hormone of granulosa cells from anovulatory women with polycystic ovary syndrome: relevance to mechanism of anovulation. *J Clin Endocrinol Metab.* 1998;83(11):3984-3991.
46. Owens LA, Kristensen SG, Lerner A, et al. Gene expression in granulosa cells from small antral follicles from women with or without polycystic ovaries. *J Clin Endocrinol Metab.* 2019;104(12):6182-6192.
47. Franks S. Androgen production and action in the ovary. *Curr Opin Endocr Metab Res.* 2021;18:48-53.
48. Erickson GF, Magoffin DA, Garzo VG, Cheung AP, Chang RJ. Granulosa cells of polycystic ovaries: are they normal or abnormal? *Hum Reprod.* 1992;7(3):293-299.
49. White DM, Polson DW, Kiddy D, et al. Induction of ovulation with low-dose gonadotropins in polycystic ovary syndrome: an analysis of 109 pregnancies in 225 women. *J Clin Endocrinol Metab.* 1996;81(11):3821-3824.

50. Johnstone EB, Rosen MP, Neril R, *et al.* The polycystic ovary post-Rotterdam: a common, age-dependent finding in ovulatory women without metabolic significance. *J Clin Endocrinol Metab.* 2010;95(11):4965-4972.
51. Teede HJ, Misso ML, Costello MF, *et al.*; International PCOS Network. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Clin Endocrinol (Oxf).* 2018;89(3):251-268.
52. Teede HJ, Misso ML, Costello MF, *et al.*; International PCOS Network. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Fertil Steril.* 2018;110(3):364-379.
53. Teede HJ, Misso ML, Costello MF, *et al.*; International PCOS Network. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum Reprod.* 2018;33(9):1602-1618.
54. Polson DW, Adams J, Wadsworth J, Franks S. Polycystic ovaries—a common finding in normal women. *Lancet.* 1988;1(8590):870-872.
55. Dewailly D, Pigny P, Soudan B, *et al.* Reconciling the definitions of polycystic ovary syndrome: the ovarian follicle number and serum anti-Müllerian hormone concentrations aggregate with the markers of hyperandrogenism. *J Clin Endocrinol Metab.* 2010;95(9):4399-4405.
56. Carmina E, Wong L, Chang L, *et al.* Endocrine abnormalities in ovulatory women with polycystic ovaries on ultrasound. *Hum Reprod.* 1997;12(5):905-909.
57. Adams JM, Taylor AE, Crowley WF Jr, Hall JE. Polycystic ovarian morphology with regular ovulatory cycles: insights into the pathophysiology of polycystic ovarian syndrome. *J Clin Endocrinol Metab.* 2004;89(9):4343-4350.
58. Mortensen M, Rosenfield RL, Littlejohn E. Functional significance of polycystic-size ovaries in healthy adolescents. *J Clin Endocrinol Metab.* 2006;91(10):3786-3790.
59. Chang PL, Lindheim SR, Lowre C, *et al.* Normal ovulatory women with polycystic ovaries have hyperandrogenic pituitary-ovarian responses to gonadotropin-releasing hormone-agonist testing. *J Clin Endocrinol Metab.* 2000;85(3):995-1000.
60. Mortensen M, Ehrmann DA, Littlejohn E, Rosenfield RL. Asymptomatic volunteers with a polycystic ovary are a functionally distinct but heterogeneous population. *J Clin Endocrinol Metab.* 2009;94(5):1579-1586.
61. MacDougall MJ, Tan SL, Jacobs HS. In-vitro fertilization and the ovarian hyperstimulation syndrome. *Hum Reprod.* 1992;7(5):597-600.
62. Swanton A, Storey L, McVeigh E, Child T. IVF outcome in women with PCOS, PCO and normal ovarian morphology. *Eur J Obstet Gynecol Reprod Biol.* 2010;149(1):68-71.
63. Fauser BC, Diedrich K, Devroey P; Evian Annual Reproduction Workshop Group 2007. Predictors of ovarian response: progress towards individualized treatment in ovulation induction and ovarian stimulation. *Hum Reprod Update.* 2008;14(1):1-14.
64. 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.
65. Franks S, Webber LJ, Goh M, *et al.* Ovarian morphology is a marker of heritable biochemical traits in sisters with polycystic ovaries. *J Clin Endocrinol Metab.* 2008;93(9):3396-3402.
66. Murphy MK, Hall JE, Adams JM, Lee H, Welt CK. Polycystic ovarian morphology in normal women does not predict the development of polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2006;91(10):3878-3884.
67. Dunaif A, Segal KR, Futterweit W, Dobrjansky A. Profound peripheral insulin resistance, independent of obesity, in polycystic ovary syndrome. *Diabetes.* 1989;38(9):1165-1174.
68. Dunaif A, Segal KR, Shelley DR, Green G, Dobrjansky A, Licholai T. Evidence for distinctive and intrinsic defects in insulin action in polycystic ovary syndrome. *Diabetes.* 1992;41(10):1257-1266.
69. DeFronzo RA. Lilly lecture 1987. The triumvirate: beta-cell, muscle, liver. A collusion responsible for NIDDM. *Diabetes.* 1988;37(6):667-687.
70. Ovesen P, Moller J, Ingerslev HJ, *et al.* Normal basal and insulin-stimulated fuel metabolism in lean women with the polycystic ovary syndrome. *J Clin Endocrinol Metab.* 1993;77(6):1636-1640.
71. Chang W, Goodarzi MO, Williams H, Magoffin DA, Pall M, Azziz R. Adipocytes from women with polycystic ovary syndrome demonstrate altered phosphorylation and activity of glycogen synthase kinase 3. *Fertil Steril.* 2008;90(6):2291-2297.
72. Ciaraldi TP, el-Roeiy A, Madar Z, Reichart D, Olefsky JM, Yen SS. Cellular mechanisms of insulin resistance in polycystic ovarian syndrome. *J Clin Endocrinol Metab.* 1992;75(2):577-583.
73. Rosenbaum D, Haber RS, Dunaif A. Insulin resistance in polycystic ovary syndrome: decreased expression of GLUT-4 glucose transporters in adipocytes. *Am J Physiol.* 1993;264(2 Pt 1):E197-E202.
74. Seow KM, Juan CC, Hsu YP, Hwang JL, Huang LW, Ho LT. Amelioration of insulin resistance in women with PCOS via reduced insulin receptor substrate-1 Ser312 phosphorylation following laparoscopic ovarian electrocautery. *Hum Reprod.* 2007;22(4):1003-1010.
75. Ciaraldi TP, Aroda V, Mudaliar S, Chang RJ, Henry RR. Polycystic ovary syndrome is associated with tissue-specific differences in insulin resistance. *J Clin Endocrinol Metab.* 2009;94(1):157-163.
76. Corbould A, Zhao H, Mirzoeva S, Aird F, Dunaif A. Enhanced mitogenic signaling in skeletal muscle of women with polycystic ovary syndrome. *Diabetes.* 2006;55(3):751-759.
77. Book CB, Dunaif A. Selective insulin resistance in the polycystic ovary syndrome. *J Clin Endocrinol Metab.* 1999;84(9):3110-3116.
78. Rice S, Christoforidis N, Gadd C, *et al.* Impaired insulin-dependent glucose metabolism in granulosa-lutein cells from anovulatory women with polycystic ovaries. *Hum Reprod.* 2005;20(2):373-381.
79. Rajkhowa M, Brett S, Cuthbertson DJ, *et al.* Insulin resistance in polycystic ovary syndrome is associated with defective regulation of ERK1/2 by insulin in skeletal muscle in vivo. *Biochem J.* 2009;418(3):665-671.
80. Bergman RN, Finegood DT, Kahn SE. The evolution of beta-cell dysfunction and insulin resistance in type 2 diabetes. *Eur J Clin Invest.* 2002;32 Suppl 3:35-45.
81. Bergman RN, Phillips LS, Cobelli C. Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose. *J Clin Invest.* 1981;68(6):1456-1467.
82. Bergman RN. Orchestration of glucose homeostasis: from a small acorn to the California oak. *Diabetes.* 2007;56(6):1489-1501.
83. Ehrmann DA, Sturis J, Byrne MM, Karrison T, Rosenfield RL, Polonsky KS. Insulin secretory defects in polycystic ovary syndrome. Relationship to insulin sensitivity and family history of non-insulin-dependent diabetes mellitus. *J Clin Invest.* 1995;96(1):520-527.
84. O'Meara NM, Blackman JD, Ehrmann DA, *et al.* Defects in beta-cell function in functional ovarian hyperandrogenism. *J Clin Endocrinol Metab.* 1993;76(5):1241-1247.
85. Dunaif A, Finegood DT. Beta-cell dysfunction independent of obesity and glucose intolerance in the polycystic ovary syndrome. *J Clin Endocrinol Metab.* 1996;81(3):942-947.
86. Fehmann HC, Göke R, Göke B. Cell and molecular biology of the incretin hormones glucagon-like peptide-I and glucose-dependent insulin releasing polypeptide. *Endocr Rev.* 1995;16(3):390-410.
87. Torchen LC, Fogel NR, Brickman WJ, Papanodis R, Dunaif A. Persistent apparent pancreatic β -cell defects in premenarchal PCOS relatives. *J Clin Endocrinol Metab.* 2014;99(10):3855-3862.
88. Dunaif A, Graf M, Mandeli J, Laumas V, Dobrjansky A. Characterization of groups of hyperandrogenic women with acanthosis nigricans, impaired glucose tolerance, and/or hyperinsulinemia. *J Clin Endocrinol Metab.* 1987;65(3):499-507.
89. Holte J, Bergh T, Berne C, Berglund L, Lithell H. Enhanced early insulin response to glucose in relation to insulin resistance in women with polycystic ovary syndrome and normal glucose tolerance. *J Clin Endocrinol Metab.* 1994;78(5):1052-1058.
90. Weber RF, Pache TD, Jacobs ML, *et al.* The relation between clinical manifestations of polycystic ovary syndrome and beta-cell function. *Clin Endocrinol (Oxf).* 1993;38(3):295-300.

91. Ciampelli M, Fulghesu AM, Cucinelli F, *et al.* Heterogeneity in beta cell activity, hepatic insulin clearance and peripheral insulin sensitivity in women with polycystic ovary syndrome. *Hum Reprod.* 1997;12(9):1897-1901.
92. Ciampelli M, Fulghesu AM, Murgia F, *et al.* Acute insulin response to intravenous glucagon in polycystic ovary syndrome. *Hum Reprod.* 1998;13(4):847-851.
93. Goodarzi MO, Erickson S, Port SC, Jennrich RI, Korenman SG. beta-Cell function: a key pathological determinant in polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2005;90(1):310-315.
94. Haendel MA, Chute CG, Robinson PN. Classification, ontology, and precision medicine. *N Engl J Med.* 2018;379(15):1452-1462.
95. Zawadzki JK, Dunaif A. Diagnostic criteria for polycystic ovary syndrome; towards a rational approach. In: Dunaif A, Givens JR, Haseltine F, Merriam G, eds. *Polycystic Ovary Syndrome*. Boston: Blackwell Scientific; 1992:377-384.
96. 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.
97. Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod.* 2004;19(1):41-47.
98. Legro RS, Driscoll D, Strauss JF 3rd, Fox J, Dunaif A. Evidence for a genetic basis for hyperandrogenemia in polycystic ovary syndrome. *Proc Natl Acad Sci U S A.* 1998;95(25):14956-14960.
99. Kashar-Miller M, Azziz R. Heritability and the risk of developing androgen excess. *J Steroid Biochem Mol Biol.* 1999;69(1-6):261-268.
100. Dokras A, Saini S, Gibson-Helm M, Schulkin J, Cooney L, Teede H. Gaps in knowledge among physicians regarding diagnostic criteria and management of polycystic ovary syndrome. *Fertil Steril.* 2017;107(6):1380-1386.e1.
101. Amsterdam ESHRE/ASRM-Sponsored 3rd PCOS Consensus Workshop Group. Consensus on women's health aspects of polycystic ovary syndrome (PCOS). *Hum Reprod.* 2012;27(1):14-24.
102. National Institutes of Health. *Evidence-based Methodology Workshop on Polycystic Ovary Syndrome - Final Report*. December 3-5, 2012. <https://prevention.nih.gov/research-priorities/research-needs-and-gaps/pathways-prevention/evidence-based-methodology-workshop-polycystic-ovary-syndrome-pcos>
103. Legro RS, Arslanian SA, Ehrmann DA, *et al.*; Endocrine Society. Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab.* 2013;98(12):4565-4592.
104. Moran L, Teede H. Metabolic features of the reproductive phenotypes of polycystic ovary syndrome. *Hum Reprod Update.* 2009;15(4):477-488.
105. Welt CK, Gudmundsson JA, Arason G, *et al.* Characterizing discrete subsets of polycystic ovary syndrome as defined by the Rotterdam criteria: the impact of weight on phenotype and metabolic features. *J Clin Endocrinol Metab.* 2006;91(12):4842-4848.
106. Carmina E, Chu MC, Longo RA, Rini GB, Lobo RA. Phenotypic variation in hyperandrogenic women influences the findings of abnormal metabolic and cardiovascular risk parameters. *J Clin Endocrinol Metab.* 2005;90(5):2545-2549.
107. Dewailly D, Catteau-Jonard S, Reyss AC, Leroy M, Pigny P. Oligoanovulation with polycystic ovaries but not overt hyperandrogenism. *J Clin Endocrinol Metab.* 2006;91(10):3922-3927.
108. Diamanti-Kandarakis E, Panidis D. Unravelling the phenotypic map of polycystic ovary syndrome (PCOS): a prospective study of 634 women with PCOS. *Clin Endocrinol (Oxf).* 2007;67(5):735-742.
109. Barber TM, Wass JA, McCarthy MI, Franks S. Metabolic characteristics of women with polycystic ovaries and oligo-amenorrhoea but normal androgen levels: implications for the management of polycystic ovary syndrome. *Clin Endocrinol (Oxf).* 2007;66(4):513-517.
110. Zhang HY, Zhu FF, Xiong J, Shi XB, Fu SX. Characteristics of different phenotypes of polycystic ovary syndrome based on the Rotterdam criteria in a large-scale Chinese population. *BJOG.* 2009;116(12):1633-1639.
111. Goverde AJ, van Koert AJ, Eijkemans MJ, *et al.* Indicators for metabolic disturbances in anovulatory women with polycystic ovary syndrome diagnosed according to the Rotterdam consensus criteria. *Hum Reprod.* 2009;24(3):710-717.
112. Welt CK, Arason G, Gudmundsson JA, *et al.* Defining constant versus variable phenotypic features of women with polycystic ovary syndrome using different ethnic groups and populations. *J Clin Endocrinol Metab.* 2006;91(11):4361-4368.
113. Wijeyaratne CN, Seneviratne Rde A, Dahanayake S, *et al.* Phenotype and metabolic profile of South Asian women with polycystic ovary syndrome (PCOS): results of a large database from a specialist Endocrine Clinic. *Hum Reprod.* 2011;26(1):202-213.
114. Wild RA, Carmina E, Diamanti-Kandarakis E, *et al.* Assessment of cardiovascular risk and prevention of cardiovascular disease in women with the polycystic ovary syndrome: a consensus statement by the Androgen Excess and Polycystic Ovary Syndrome (AE-PCOS) Society. *J Clin Endocrinol Metab.* 2010;95(5):2038-2049.
115. Legro RS, Chiu P, Kunselman AR, Bentley CM, Dodson WC, Dunaif A. Polycystic ovaries are common in women with hyperandrogenic chronic anovulation but do not predict metabolic or reproductive phenotype. *J Clin Endocrinol Metab.* 2005;90(5):2571-2579.
116. Rosenfield RL. The polycystic ovary morphology-polycystic ovary syndrome spectrum. *J Pediatr Adolesc Gynecol.* 2015;28(6):412-419.
117. Day F, Karaderi T, Jones MR, *et al.*; 23andMe Research Team. Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *Plos Genet.* 2018;14(12):e1007813.
118. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab.* 2007;92(2):405-413.
119. Salameh WA, Redor-Goldman MM, Clarke NJ, Mathur R, Azziz R, Reitz RE. Specificity and predictive value of circulating testosterone assessed by tandem mass spectrometry for the diagnosis of polycystic ovary syndrome by the National Institutes of Health 1990 criteria. *Fertil Steril.* 2014;101(4):1135-1141 e1132.
120. Dewailly D, Alebić MŠ, Duhamel A, Stojanović N. Using cluster analysis to identify a homogeneous subpopulation of women with polycystic ovarian morphology in a population of non-hyperandrogenic women with regular menstrual cycles. *Hum Reprod.* 2014;29(11):2536-2543.
121. Chang S, Dunaif A. Diagnosis of polycystic ovary syndrome: which criteria to use and when? *Endocrinol Metab Clin North Am.* 2021;50(1):11-23.
122. Dewailly D, Lujan ME, Carmina E, *et al.* Definition and significance of polycystic ovarian morphology: a task force report from the Androgen Excess and Polycystic Ovary Syndrome Society. *Hum Reprod Update.* 2014;20(3):334-352.
123. Knochenhauer ES, Key TJ, Kahsar-Miller M, Waggoner W, Boots LR, Azziz R. Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States: a prospective study. *J Clin Endocrinol Metab.* 1998;83(9):3078-3082.
124. Diamanti-Kandarakis E, Kouli CR, Bergiele AT, *et al.* A survey of the polycystic ovary syndrome in the Greek island of Lesbos: hormonal and metabolic profile. *J Clin Endocrinol Metab.* 1999;84(11):4006-4011.
125. Carmina E, Lobo RA. Polycystic ovary syndrome (PCOS): arguably the most common endocrinopathy is associated with significant morbidity in women. *J Clin Endocrinol Metab.* 1999;84(6):1897-1899.
126. Azziz R, Woods KS, Reyna R, Key TJ, Knochenhauer ES, Yildiz BO. The prevalence and features of the polycystic ovary

- syndrome in an unselected population. *J Clin Endocrinol Metab.* 2004;89(6):2745-2749.
127. Bozdag G, Mumusoglu S, Zengin D, Karabulut E, Yildiz BO. The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod.* 2016;31(12):2841-2855.
 128. Asunción M, Calvo RM, San Millán JL, Sancho J, Avila S, Escobar-Morreale HF. A prospective study of the prevalence of the polycystic ovary syndrome in unselected Caucasian women from Spain. *J Clin Endocrinol Metab.* 2000;85(7):2434-2438.
 129. Kumarapeli V, Seneviratne Rde A, Wijeyaratne CN, Yapa RM, Dodampahala SH. A simple screening approach for assessing community prevalence and phenotype of polycystic ovary syndrome in a semi-urban population in Sri Lanka. *Am J Epidemiol.* 2008;168(3):321-328.
 130. Sorbara LR, Tang Z, Cama A, et al. Absence of insulin receptor gene mutations in three insulin-resistant women with the polycystic ovary syndrome. *Metabolism.* 1994;43(12):1568-1574.
 131. Goodarzi MO, Quiñones MJ, Azziz R, Rotter JL, Hsueh WA, Yang H. Polycystic ovary syndrome in Mexican-Americans: prevalence and association with the severity of insulin resistance. *Fertil Steril.* 2005;84(3):766-769.
 132. Ladson G, Dodson WC, Sweet SD, et al. Racial influence on the polycystic ovary syndrome phenotype: a black and white case-control study. *Fertil Steril.* 2011;96(1):224-229.e2.
 133. Engmann L, Jin S, Sun F, et al.; Reproductive Medicine Network. Racial and ethnic differences in the polycystic ovary syndrome metabolic phenotype. *Am J Obstet Gynecol.* 2017;216(5):493.e1-493.e13.
 134. Park HR, Choi Y, Lee HJ, Oh JY, Hong YS, Sung YA. Phenotypic characteristics according to insulin sensitivity in non-obese Korean women with polycystic ovary syndrome. *Diabetes Res Clin Pract.* 2007;77 Suppl 1:S233-S237.
 135. Broekmans FJ, Knauff EA, Valkenburg O, Laven JS, Eijkemans MJ, Fauser BC. PCOS according to the Rotterdam consensus criteria: change in prevalence among WHO-II anovulation and association with metabolic factors. *BJOG.* 2006;113(10):1210-1217.
 136. Brower MA, Hai Y, Jones MR, et al. Bidirectional Mendelian randomization to explore the causal relationships between body mass index and polycystic ovary syndrome. *Hum Reprod.* 2019;34(1):127-136.
 137. Zhao Y, Xu Y, Wang X, et al. Body mass index and polycystic ovary syndrome: a 2-sample bidirectional mendelian randomization study. *J Clin Endocrinol Metab.* 2020;105(6):1778-1784.
 138. Yildiz BO, Knochenhauer ES, Azziz R. Impact of obesity on the risk for polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2008;93(1):162-168.
 139. Cooper HE, Spellacy WN, Prem KA, Cohen WD. Hereditary factors in the Stein-Leventhal syndrome. *Am J Obstet Gynecol.* 1968;100(3):371-387.
 140. Judd HL, Scully RE, Herbst AL, Yen SS, Ingersol FM, Kliman B. Familial hyperthecosis: comparison of endocrinologic and histologic findings with polycystic ovarian disease. *Am J Obstet Gynecol.* 1973;117(7):976-982.
 141. Cohen PN, Givens JR, Wisner WL, et al. Polycystic ovarian disease, maturation arrest of spermiogenesis, and Klinefelter's syndrome in siblings of a family with familial hirsutism. *Fertil Steril.* 1975;26(12):1228-1238.
 142. Ferriman D, Purdie AW. The inheritance of polycystic ovarian disease and a possible relationship to premature balding. *Clin Endocrinol (Oxf).* 1979;11(3):291-300.
 143. Givens JR. Familial polycystic ovarian disease. *Endocrinol Metab Clin North Am.* 1988;17(4):771-783.
 144. Kahn CR, Flier JS, Bar RS, et al. The syndromes of insulin resistance and acanthosis nigricans. Insulin-receptor disorders in man. *N Engl J Med.* 1976;294(14):739-745.
 145. Taylor SI, Cama A, Accili D, et al. Mutations in the insulin receptor gene. *Endocr Rev.* 1992;13(3):566-595.
 146. Stuart CA, Peters EJ, Prince MJ, Richards G, Cavallo A, Meyer WJ 3rd. Insulin resistance with acanthosis nigricans: the roles of obesity and androgen excess. *Metabolism.* 1986;35(3):197-205.
 147. Dunaif A, Xia J, Book CB, Schenker E, Tang Z. Excessive insulin receptor serine phosphorylation in cultured fibroblasts and in skeletal muscle. A potential mechanism for insulin resistance in the polycystic ovary syndrome. *J Clin Invest.* 1995;96(2):801-810.
 148. Lunde O, Magnus P, Sandvik L, Høglø S. Familial clustering in the polycystic ovarian syndrome. *Gynecol Obstet Invest.* 1989;28(1):23-30.
 149. 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 (Oxf).* 1993;38(6):653-658.
 150. Azziz R, Kasher-Miller MD. Family history as a risk factor for the polycystic ovary syndrome. *J Pediatr Endocrinol Metab.* 2000;13 Suppl 5:1303-1306.
 151. Legro RS, Bentley-Lewis R, Driscoll D, Wang SC, Dunaif A. Insulin resistance in the sisters of women with polycystic ovary syndrome: association with hyperandrogenemia rather than menstrual irregularity. *J Clin Endocrinol Metab.* 2002;87(5):2128-2133.
 152. Weinshilboum RM, Raymond FA. Inheritance of low erythrocyte catechol-o-methyltransferase activity in man. *Am J Hum Genet.* 1977;29(2):125-135.
 153. Camussi A, Bianchi G. Genetics of essential hypertension. From the unimodal-bimodal controversy to molecular technology. *Hypertension.* 1988;12(6):620-628.
 154. Legro RS, Kunselman AR, Demers L, Wang SC, Bentley-Lewis R, Dunaif A. Elevated dehydroepiandrosterone sulfate levels as the reproductive phenotype in the brothers of women with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2002;87(5):2134-2138.
 155. Sam S, Legro RS, Essah PA, Apridonidze T, Dunaif A. Evidence for metabolic and reproductive phenotypes in mothers of women with polycystic ovary syndrome. *Proc Natl Acad Sci U S A.* 2006;103(18):7030-7035.
 156. Torchen LC, Kumar A, Kalra B, et al. Increased antiMüllerian hormone levels and other reproductive endocrine changes in adult male relatives of women with polycystic ovary syndrome. *Fertil Steril.* 2016;106(1):50-55.
 157. Sir-Petermann T, Codner E, Maliqueo M, et al. Increased anti-Müllerian hormone serum concentrations in prepubertal daughters of women with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2006;91(8):3105-3109.
 158. Recabarren SE, Sir-Petermann T, Rios R, et al. Pituitary and testicular function in sons of women with polycystic ovary syndrome from infancy to adulthood. *J Clin Endocrinol Metab.* 2008;93(9):3318-3324.
 159. Torchen LC, Legro RS, Dunaif A. Distinctive reproductive phenotypes in peripubertal girls at risk for polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2019;104(8):3355-3361.
 160. Sir-Petermann T, Angel B, Maliqueo M, Carvajal F, Santos JL, Pérez-Bravo F. Prevalence of type II diabetes mellitus and insulin resistance in parents of women with polycystic ovary syndrome. *Diabetologia.* 2002;45(7):959-964.
 161. Yildiz BO, Yarali H, Oguz H, Bayraktar M. Glucose intolerance, insulin resistance, and hyperandrogenemia in first degree relatives of women with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2003;88(5):2031-2036.
 162. Coviello AD, Sam S, Legro RS, Dunaif A. High prevalence of metabolic syndrome in first-degree male relatives of women with polycystic ovary syndrome is related to high rates of obesity. *J Clin Endocrinol Metab.* 2009;94(11):4361-4366.
 163. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era—concepts and misconceptions. *Nat Rev Genet.* 2008;9(4):255-266.
 164. Conley D, Rauscher E, Dawes C, Magnusson PK, Siegal ML. Heritability and the equal environments assumption: evidence from multiple samples of misclassified twins. *Behav Genet.* 2013;43(5):415-426.

165. 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):2100-2104.
166. Harris JA, Vernon PA, Boomsma DI. The heritability of testosterone: a study of Dutch adolescent twins and their parents. *Behav Genet.* 1998;28(3):165-171.
167. Hoekstra RA, Bartels M, Boomsma DI. Heritability of testosterone levels in 12-year-old twins and its relation to pubertal development. *Twin Res Hum Genet.* 2006;9(4):558-565.
168. Coviello AD, Zhuang WV, Lunetta KL, et al. Circulating testosterone and SHBG concentrations are heritable in women: the Framingham Heart Study. *J Clin Endocrinol Metab.* 2011;96(9):E1491-E1495.
169. Yildiz BO, Goodarzi MO, Guo X, Rotter JJ, Azziz R. Heritability of dehydroepiandrosterone sulfate in women with polycystic ovary syndrome and their sisters. *Fertil Steril.* 2006;86(6):1688-1693.
170. Speed D, Holmes J, Balding DJ. Evaluating and improving heritability models using summary statistics. *Nat Genet.* 2020;52(4):458-462.
171. Ruth KS, Day FR, Tyrrell J, et al.; Endometrial Cancer Association Consortium. Using human genetics to understand the disease impacts of testosterone in men and women. *Nat Med.* 2020;26(2):252-258.
172. Wright S. Evolution in Mendelian populations. *Genetics.* 1931;16(2):97-159.
173. Felsenstein J. The theoretical population genetics of variable selection and migration. *Annu Rev Genet.* 1976;10:253-280.
174. Lynch M, Ackerman MS, Gout JF, et al. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet.* 2016;17(11):704-714.
175. Fisher RA. *The Genetical Theory of Natural Selection.* Oxford: The Clarendon press; 1931.
176. Bickeböller H, Campion D, Brice A, et al. Apolipoprotein E and Alzheimer disease: genotype-specific risks by age and sex. *Am J Hum Genet.* 1997;60(2):439-446.
177. Qi Q, Chu AY, Kang JH, et al. Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies. *BMJ.* 2014;348:g1610.
178. Raghavan N, Tosto G. Genetics of Alzheimer's disease: the importance of polygenic and epistatic components. *Curr Neurol Neurosci Rep.* 2017;17(10):78.
179. Maruyama T. The age of a rare mutant gene in a large population. *Am J Hum Genet.* 1974;26(6):669-673.
180. Gazal S, Finucane HK, Furlotte NA, et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat Genet.* 2017;49(10):1421-1427.
181. Haldane JBS. The effect of variation on fitness. *Am Nat.* 1937;71:337-349.
182. Lohmueller KE. The distribution of deleterious genetic variation in human populations. *Curr Opin Genet Dev.* 2014;29:139-146.
183. O'Connor LJ, Schoech AP, Hormozdiari F, Gazal S, Patterson N, Price AL. Extreme polygenicity of complex traits is explained by negative selection. *Am J Hum Genet.* 2019;105(3):456-476.
184. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell.* 2017;169(7):1177-1186.
185. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nat Rev Genet.* 2009;10(4):241-251.
186. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461(7265):747-753.
187. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *Plos Comput Biol.* 2012;8(12):e1002822.
188. Elston RC, Stewart J. A general model for the genetic analysis of pedigree data. *Hum Hered.* 1971;21(6):523-542.
189. Hearne CM, Ghosh S, Todd JA. Microsatellites for linkage analysis of genetic traits. *Trends Genet.* 1992;8(8):288-294.
190. Ott J. *Analysis of Human Genetic Linkage.* 3rd ed. Baltimore: Johns Hopkins University Press; 1999.
191. Haseman JK, Elston RC. The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet.* 1972;2(1):3-19.
192. Bush WS, Haines J. Overview of linkage analysis in complex traits. *Curr Protoc Hum Genet.* 2010;Chapter 1:Unit 1.9.1-Unit 1.918.
193. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904-909.
194. Voight BF, Pritchard JK. Confounding from cryptic relatedness in case-control association studies. *Plos Genet.* 2005;1(3):e32.
195. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet.* 2003;361(9357):598-604.
196. Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. *Nat Protoc.* 2011;6(2):121-133.
197. Spielman RS, McGinnis RE, Ewens WJ. The transmission/disequilibrium test detects cosegregation and linkage. *Am J Hum Genet.* 1994;54(3):559-60; author reply 560.
198. Ewens WJ, Spielman RS. The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet.* 1995;57(2):455-464.
199. Spielman RS, Ewens WJ. The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet.* 1996;59(5):983-989.
200. Ewens WJ, Li M, Spielman RS. A review of family-based tests for linkage disequilibrium between a quantitative trait and a genetic marker. *Plos Genet.* 2008;4(9):e1000180.
201. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet.* 1993;52(3):506-516.
202. Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. *Genet Epidemiol.* 2000;19 Suppl 1:S36-S42.
203. Rabinowitz D, Laird N. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum Hered.* 2000;50(4):211-223.
204. Wijsman EM. The role of large pedigrees in an era of high-throughput sequencing. *Hum Genet.* 2012;131(10):1555-1563.
205. Ionita-Laza I, Ottman R. Study designs for identification of rare disease variants in complex diseases: the utility of family-based designs. *Genetics.* 2011;189(3):1061-1068.
206. Jun G, Manning A, Almeida M, et al. Evaluating the contribution of rare variants to type 2 diabetes and related traits using pedigrees. *Proc Natl Acad Sci U S A.* 2018;115(2):379-384.
207. Lander ES, Linton LM, Birren B, et al.; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature.* 2001;409(6822):860-921.
208. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science.* 2001;291(5507):1304-1351.
209. Syvänen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet.* 2001;2(12):930-942.
210. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet.* 2005;6(2):95-108.
211. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet.* 2001;17(9):502-510.
212. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev.* 2009;19(3):212-219.
213. International HapMap Consortium. The International HapMap Project. *Nature.* 2003;426(6968):789-796.
214. McCarthy S, Das S, Kretschmar W, et al.; Haplotype Reference Consortium. A reference panel of 64 976 haplotypes for genotype imputation. *Nat Genet.* 2016;48(10):1279-1283.

215. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet.* 2018;19(8):491-504.
216. Buniello A, MacArthur JAL, Cerezo M, *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47(D1):D1005-D1012.
217. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet.* 2018;19(9):581-590.
218. Khera AV, Chaffin M, Aragam KG, *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet.* 2018;50(9):1219-1224.
219. Duncan L, Shen H, Gelaye B, *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun.* 2019;10(1):3328.
220. Visscher PM, Wray NR, Zhang Q, *et al.* 10 Years of GWAS discovery: biology, function, and translation. *Am J Hum Genet.* 2017;101(1):5-22.
221. Rastegar-Mojarad M, Ye Z, Kolesar JM, Hebring SJ, Lin SM. Opportunities for drug repositioning from phenome-wide association studies. *Nat Biotechnol.* 2015;33(4):342-345.
222. Swanson SA, Tiemeier H, Ikram MA, Hernán MA. Nature as a Trialist?: deconstructing the analogy between Mendelian randomization and randomized trials. *Epidemiology.* 2017;28(5):653-659.
223. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003;32(1):1-22.
224. Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet.* 2008;123(1):15-33.
225. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet.* 2018;50(5):693-698.
226. Yang J, Weedon MN, Purcell S, *et al.*; GIANT Consortium. Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet.* 2011;19(7):807-812.
227. Bulik-Sullivan BK, Loh PR, Finucane HK, *et al.*; Schizophrenia Working Group of the Psychiatric Genomics Consortium. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47(3):291-295.
228. Bulik-Sullivan B, Finucane HK, Anttila V, *et al.*; ReproGen Consortium; Psychiatric Genomics Consortium; Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015;47(11):1236-1241.
229. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet.* 2010;11(1):31-46.
230. Bell CJ, Dinwiddie DL, Miller NA, *et al.* Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med.* 2011;3(65):65ra4.
231. Majewski J, Pastinen T. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet.* 2011;27(2):72-79.
232. Ng SB, Turner EH, Robertson PD, *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature.* 2009;461(7261):272-276.
233. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008;5(7):621-628.
234. Barski A, Cuddapah S, Cui K, *et al.* High-resolution profiling of histone methylations in the human genome. *Cell.* 2007;129(4):823-837.
235. Lister R, Pelizzola M, Dowen RH, *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 2009;462(7271):315-322.
236. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013;10(12):1213-1218.
237. Yang HJ, Ratnapriya R, Cogliati T, Kim JW, Swaroop A. Vision from next generation sequencing: multi-dimensional genome-wide analysis for producing gene regulatory networks underlying retinal development, aging and disease. *Prog Retin Eye Res.* 2015;46:1-30.
238. Kowalski MH, Qian H, Hou Z, *et al.*; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium; TOPMed Hematology & Hemostasis Working Group. Use of >100 000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *Plos Genet.* 2019;15(12):e1008500.
239. Kaiser J. 200,000 whole genomes made available for biomedical studies. *Science.* 2021;374(6571):1036.
240. Margulies M, Egholm M, Altman WE, *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* 2005;437(7057):376-380.
241. Binladen J, Gilbert MT, Bollback JP, *et al.* The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *Plos One.* 2007;2(2):e197.
242. Mamanova L, Coffey AJ, Scott CE, *et al.* Target-enrichment strategies for next-generation sequencing. *Nat Methods.* 2010;7(2):111-118.
243. Crawford GE, Holt IE, Whittle J, *et al.* Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res.* 2006;16(1):123-131.
244. Boyle AP, Davis S, Shulha HP, *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell.* 2008;132(2):311-322.
245. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 2007;17(6):877-885.
246. Gaulton KJ, Nammo T, Pasquali L, *et al.* A map of open chromatin in human pancreatic islets. *Nat Genet.* 2010;42(3):255-259.
247. Lieberman-Aiden E, van Berkum NL, Williams L, *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009;326(5950):289-293.
248. Fullwood MJ, Liu MH, Pan YF, *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature.* 2009;462(7269):58-64.
249. Morin RD, O'Connor MD, Griffith M, *et al.* Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* 2008;18(4):610-621.
250. Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol.* 2009;16(2):130-137.
251. Zhao S, Watrous K, Zhang C, Zhang B. Cloud computing for next-generation sequencing data analysis. In: Sen J, ed. *Cloud Computing-Architecture and Applications.* Rijeka: Intech; 2017:29-51.
252. Urbanek M, Legro RS, Driscoll DA, *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):8573-8578.
253. Goodarzi MO. Looking for polycystic ovary syndrome genes: rational and best strategy. *Semin Reprod Med.* 2008;26(1):5-13.
254. Bjornes AC, Saxena R, Welt CK. Relationship between polycystic ovary syndrome and ancestry in European Americans. *Fertil Steril.* 2016;106(7):1772-1777.
255. Kosova G, Urbanek M. Genetics of the polycystic ovary syndrome. *Mol Cell Endocrinol.* 2013;373(1-2):29-38.

256. Pau C, Saxena R, Welt CK. Evaluating reported candidate gene associations with polycystic ovary syndrome. *Fertil Steril*. 2013;99(6):1774-1778.
257. Carmina E, Napoli N, Longo RA, Rini GB, Lobo RA. Metabolic syndrome in polycystic ovary syndrome (PCOS): lower prevalence in southern Italy than in the USA and the influence of criteria for the diagnosis of PCOS. *Eur J Endocrinol*. 2006;154(1):141-145.
258. Wolf WM, Wattick RA, Kinkade ON, Olfert MD. Geographical prevalence of polycystic ovary syndrome as determined by region and race/ethnicity. *Int J Environ Res Public Health*. 2018;15(11):2589.
259. Hiam D, Moreno-Asso A, Teede HJ, et al. The genetics of polycystic ovary syndrome: an overview of candidate gene systematic reviews and genome-wide association studies. *J Clin Med*. 2019;8(10):1606.
260. Schaid DJ, Jacobsen SJ. Biased tests of association: comparisons of allele frequencies when departing from Hardy-Weinberg proportions. *Am J Epidemiol*. 1999;149(8):706-711.
261. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:j4008.
262. Urbanek M, Woodroffe A, Ewens KG, et al. Candidate gene region for polycystic ovary syndrome on chromosome 19p13.2. *J Clin Endocrinol Metab*. 2005;90(12):6623-6629.
263. Stewart DR, Dombroski BA, Urbanek M, et al. Fine mapping of genetic susceptibility to polycystic ovary syndrome on chromosome 19p13.2 and tests for regulatory activity. *J Clin Endocrinol Metab*. 2006;91(10):4112-4117.
264. Urbanek M, Sam S, Legro RS, Dunaif A. Identification of a polycystic ovary syndrome susceptibility variant in fibrillin-3 and association with a metabolic phenotype. *J Clin Endocrinol Metab*. 2007;92(11):4191-4198.
265. Ewens KG, Stewart DR, Ankener W, et al. Family-based analysis of candidate genes for polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2010;95(5):2306-2315.
266. Tucci S, Futterweit W, Concepcion ES, 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):446-449.
267. Xie GB, Xu P, Che YN, 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(2):168-174.
268. Goodarzi MO, Louwers YV, Taylor KD, et al. Replication of association of a novel insulin receptor gene polymorphism with polycystic ovary syndrome. *Fertil Steril*. 2011;95(5):1736-41.e1.
269. Shi Y, Zhao H, Shi Y, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet*. 2012;44(9):1020-1025.
270. Chen ZJ, Zhao H, He L, 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):55-59.
271. Hayes MG, Urbanek M, Ehrmann DA, et al.; Reproductive Medicine Network. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. *Nat Commun*. 2015;6:7502.
272. Day FR, Hinds DA, Tung JY, et al. Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nat Commun*. 2015;6:8464.
273. Zhang Y, Ho K, Keaton JM, et al. A genome-wide association study of polycystic ovary syndrome identified from electronic health records. *Am J Obstet Gynecol*. 2020;223(4):559.e1-559.e21.
274. Hwang JY, Lee EJ, Jin Go M, 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):660-664.
275. Lee H, Oh JY, Sung YA, et al. Genome-wide association study identified new susceptibility loci for polycystic ovary syndrome. *Hum Reprod*. 2015;30(3):723-731.
276. Actkins KV, Singh K, Hucks D, et al. Characterizing the clinical and genetic spectrum of polycystic ovary syndrome in electronic health records. *J Clin Endocrinol Metab*. 2021;106(1):153-167.
277. Castro V, Shen Y, Yu S, et al. Identification of subjects with polycystic ovary syndrome using electronic health records. *Reprod Biol Endocrinol*. 2015;13:116.
278. Jones MR, Goodarzi MO. Genetic determinants of polycystic ovary syndrome: progress and future directions. *Fertil Steril*. 2016;106(1):25-32.
279. Azziz R, Dumesic DA, Goodarzi MO. Polycystic ovary syndrome: an ancient disorder? *Fertil Steril*. 2011;95(5):1544-1548.
280. Carreau-Jonard S, Jamin SP, Leclerc A, Gonzalès J, Dewailly D, di Clemente N. Anti-Mullerian hormone, its receptor, FSH receptor, and androgen receptor genes are overexpressed by granulosa cells from stimulated follicles in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2008;93(11):4456-4461.
281. Du J, Zhang W, Guo L, et al. Two FSHR variants, haplotypes and meta-analysis in Chinese women with premature ovarian failure and polycystic ovary syndrome. *Mol Genet Metab*. 2010;100(3):292-295.
282. Valkenburg O, Uitterlinden AG, Piersma D, et al. Genetic polymorphisms of GnRH and gonadotrophic hormone receptors affect the phenotype of polycystic ovary syndrome. *Hum Reprod*. 2009;24(8):2014-2022.
283. Gu BH, Park JM, Baek KH. Genetic variations of follicle stimulating hormone receptor are associated with polycystic ovary syndrome. *Int J Mol Med*. 2010;26(1):107-112.
284. Tian Y, Zhao H, Chen H, et al. Variants in FSHB are associated with polycystic ovary syndrome and luteinizing hormone level in han Chinese women. *J Clin Endocrinol Metab*. 2016;101(5):2178-2184.
285. 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):bqaa181. doi:10.1210/endoctr/bqaa181
286. Ruf-Zamojski F, Zhang Z, Zamojski M, et al. Single nucleus multi-omics regulatory landscape of the murine pituitary. *Nat Commun*. 2021;12(1):2677.
287. Bohaczuk SC, Cassin J, Slaiwa TI, Thackray VG, Mellon PL. Distal enhancer potentiates activin- and GnRH-induced transcription of FSHB. *Endocrinology*. 2021;162(7):bqab069. doi:10.1210/endoctr/bqab069
288. Urbanek M. The genetics of the polycystic ovary syndrome. *Nat Clin Pract Endocrinol Metab*. 2007;3(2):103-111.
289. Jakimiuk AJ, Weitsman SR, Navab A, Magoffin DA. Luteinizing hormone receptor, steroidogenesis acute regulatory protein, and steroidogenic enzyme messenger ribonucleic acids are overexpressed in thecal and granulosa cells from polycystic ovaries. *J Clin Endocrinol Metab*. 2001;86(3):1318-1323.
290. Zhang Y, Fatima N, Dufau ML. Coordinated changes in DNA methylation and histone modifications regulate silencing/derepression of luteinizing hormone receptor gene transcription. *Mol Cell Biol*. 2005;25(18):7929-7939.
291. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol*. 2019;20(10):590-607.
292. Zhu JQ, Zhu L, Liang XW, Xing FQ, Schatten H, Sun QY. Demethylation of LHR in dehydroepiandrosterone-induced mouse model of polycystic ovary syndrome. *Mol Hum Reprod*. 2010;16(4):260-266.
293. 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):380-384.

294. Unsal T, Konac E, Yesilkaya E, *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):205-216.
295. Xu X, Zhao H, Shi Y, *et al.* Family association study between *INSR* gene polymorphisms and PCOS in Han Chinese. *Reprod Biol Endocrinol.* 2011;9:76.
296. Zeggini E, Scott LJ, Saxena R, *et al.*; Wellcome Trust Case Control Consortium. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* 2008;40(5):638-645.
297. Voight BF, Scott LJ, Steinthorsdottir V, *et al.*; MAGIC investigators; GIANT Consortium. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet.* 2010;42(7):579-589.
298. Saxena R, Welt CK. Polycystic ovary syndrome is not associated with genetic variants that mark risk of type 2 diabetes. *Acta Diabetol.* 2013;50(3):451-457.
299. Moraru A, Cakan-Akdogan G, Strassburger K, *et al.* THADA regulates the organismal balance between energy storage and heat production. *Dev Cell.* 2017;41(4):450.
300. Chatterjee N, Perrimon N. Thermogenesis by THADA. *Dev Cell.* 2017;41(1):1-2.
301. Markowski DN, Thies HW, Gottlieb A, Wenk H, Wischnewsky M, Bullerdiek J. HMGA2 expression in white adipose tissue linking cellular senescence with diabetes. *Genes Nutr.* 2013;8(5):449-456.
302. Xi Y, Shen W, Ma L, *et al.* HMGA2 promotes adipogenesis by activating C/EBP β -mediated expression of PPAR γ . *Biochem Biophys Res Commun.* 2016;472(4):617-623.
303. Li M, Zhao H, Zhao SG, *et al.* The HMGA2-IMP2 pathway promotes granulosa cell proliferation in polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2019;104(4):1049-1059.
304. McAllister JM, Modi B, Miller BA, *et al.* Overexpression of a DENND1A isoform produces a polycystic ovary syndrome theca phenotype. *Proc Natl Acad Sci U S A.* 2014;111(15):E1519-E1527.
305. Lv X, He C, Huang C, *et al.* Timely expression and activation of YAP1 in granulosa cells is essential for ovarian follicle development. *Faseb J.* 2019;33(9):10049-10064.
306. Gershon E, Dekel N. Newly identified regulators of ovarian folliculogenesis and ovulation. *Int J Mol Sci.* 2020;21(12):4565.
307. Hakonarson H, Qu HQ, Bradfield JP, *et al.* A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. *Diabetes.* 2008;57(4):1143-1146.
308. Barrett JC, Clayton DG, Concannon P, *et al.*; Type 1 Diabetes Genetics Consortium. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet.* 2009;41(6):703-707.
309. Kaur S, Mirza AH, Brorsson CA, *et al.*; Hvidoere International Study Group. The genetic and regulatory architecture of ERBB3-type 1 diabetes susceptibility locus. *Mol Cell Endocrinol.* 2016;419:83-91.
310. Locke AE, Kahali B, Berndt SI, *et al.*; LifeLines Cohort Study; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGEN Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; International Endogene Consortium. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518(7538):197-206.
311. Chiang KM, Chang HC, Yang HC, *et al.* Genome-wide association study of morbid obesity in Han Chinese. *BMC Genet.* 2019;20(1):97.
312. Veikkolainen V, Ali N, Doroszko M, *et al.* Erbb4 regulates the oocyte microenvironment during folliculogenesis. *Hum Mol Genet.* 2020;29(17):2813-2830.
313. Richani D, Gilchrist RB. The epidermal growth factor network: role in oocyte growth, maturation and developmental competence. *Hum Reprod Update.* 2018;24(1):1-14.
314. Noma N, Kawashima I, Fan HY, *et al.* LH-induced neuregulin 1 (NRG1) type III transcripts control granulosa cell differentiation and oocyte maturation. *Mol Endocrinol.* 2011;25(1):104-116.
315. Kawashima I, Umehara T, Noma N, *et al.* Targeted disruption of Nrg1 in granulosa cells alters the temporal progression of oocyte maturation. *Mol Endocrinol.* 2014;28(5):706-721.
316. Ennequin G, Caillaud K, Chavanelle V, *et al.* Neuregulin 1 treatment improves glucose tolerance in diabetic db/db mice, but not in healthy mice. *Arch Physiol Biochem.* 2020;126(4):320-325.
317. Zhang P, Kuang H, He Y, *et al.* NRG1-Fc improves metabolic health via dual hepatic and central action. *JCI Insight.* 2018;3(5):e98522.
318. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature.* 2017;550(7675):204-213.
319. Li YI, Knowles DA, Humphrey J, *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet.* 2018;50(1):151-158.
320. Lamarche BJ, Orazio NI, Weitzman MD. The MRN complex in double-strand break repair and telomere maintenance. *FEBS Lett.* 2010;584(17):3682-3695.
321. Stolk L, Perry JR, Chasman DI, *et al.*; LifeLines Cohort Study. Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat Genet.* 2012;44(3):260-268.
322. Perry JR, Hsu YH, Chasman DI, *et al.*; kConFab investigators; ReproGen Consortium. DNA mismatch repair gene MSH6 implicated in determining age at natural menopause. *Hum Mol Genet.* 2014;23(9):2490-2497.
323. Kulkarni R, Teves ME, Han AX, McAllister JM, Strauss JF 3rd. Colocalization of polycystic ovary syndrome candidate gene products in theca cells suggests novel signaling pathways. *J Endocr Soc.* 2019;3(12):2204-2223.
324. 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;29(9):1446-1454.
325. Ruth KS, Campbell PJ, Chew S, *et al.* Genome-wide association study with 1000 genomes imputation identifies signals for nine sex hormone-related phenotypes. *Eur J Hum Genet.* 2016;24(2):284-290.
326. Mbarek H, Steinberg S, Nyholt DR, *et al.* Identification of common genetic variants influencing spontaneous dizygotic twinning and female fertility. *Am J Hum Genet.* 2016;98(5):898-908.
327. Yan J, Tian Y, Gao X, *et al.* A genome-wide association study identifies *FSHR* rs2300441 associated with follicle-stimulating hormone levels. *Clin Genet.* 2020;97(6):869-877.
328. Bycroft C, Freeman C, Petkova D, *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203-209.
329. Pott J, Bae YJ, Horn K, *et al.* Genetic association study of eight steroid hormones and implications for sexual dimorphism of coronary artery disease. *J Clin Endocrinol Metab.* 2019;104(11):5008-5023.
330. Auton A, Brooks LD, Durbin RM, *et al.*; 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74.
331. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics.* 2015;31(21):3555-3557.
332. Ahn Y, Lee H, Cho YS. Identification of genetic variants for female obesity and evaluation of the causal role of genetically defined obesity in polycystic ovarian syndrome. *Diabetes Metab Syndr Obes.* 2020;13:4311-4322.
333. Ioannidis A, Ikonomi E, Dimou NL, Douma L, Bagos PG. Polymorphisms of the insulin receptor and the insulin receptor substrates genes in polycystic ovary syndrome: a Mendelian randomization meta-analysis. *Mol Genet Metab.* 2010;99(2):174-183.
334. Sun S, Jiao M, Han C, *et al.* Causal effects of genetically determined metabolites on risk of polycystic ovary syndrome: a

- Mendelian randomization study. *Front Endocrinol (Lausanne)*. 2020;11:621.
335. Dumesic DA, Schramm RD, Abbott DH. Early origins of polycystic ovary syndrome. *Reprod Fertil Dev*. 2005;17(3):349-360.
 336. Day FR, Ruth KS, Thompson DJ, *et al.*; PRACTICAL consortium; kConFab Investigators; AOCs Investigators; Generation Scotland; EPIC-InterAct Consortium; LifeLines Cohort Study. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet*. 2015;47(11):1294-1303.
 337. Ruth KS, Beaumont RN, Tyrrell J, *et al.* Genetic evidence that lower circulating FSH levels lengthen menstrual cycle, increase age at menopause and impact female reproductive health. *Hum Reprod*. 2016;31(2):473-481.
 338. Gilbert EW, Tay CT, Hiam DS, Teede HJ, Moran LJ. Comorbidities and complications of polycystic ovary syndrome: an overview of systematic reviews. *Clin Endocrinol (Oxf)*. 2018;89(6):683-699.
 339. Kakoly NS, Khomami MB, Joham AE, *et al.* Ethnicity, obesity and the prevalence of impaired glucose tolerance and type 2 diabetes in PCOS: a systematic review and meta-regression. *Hum Reprod Update*. 2018;24(4):455-467.
 340. Legro RS, Kunselman AR, Dodson WC, Dunaif A. Prevalence and predictors of risk for type 2 diabetes mellitus and impaired glucose tolerance in polycystic ovary syndrome: a prospective, controlled study in 254 affected women. *J Clin Endocrinol Metab*. 1999;84(1):165-169.
 341. Ehrmann DA, Barnes RB, Rosenfield RL, Cavaghan MK, Imperial J. Prevalence of impaired glucose tolerance and diabetes in women with polycystic ovary syndrome. *Diabetes Care*. 1999;22(1):141-146.
 342. Shi Y, Guo M, Yan J, *et al.* Analysis of clinical characteristics in large-scale Chinese women with polycystic ovary syndrome. *Neuro Endocrinol Lett*. 2007;28(6):807-810.
 343. Morgan CL, Jenkins-Jones S, Currie CJ, Rees DA. Evaluation of adverse outcome in young women with polycystic ovary syndrome versus matched, reference controls: a retrospective, observational study. *J Clin Endocrinol Metab*. 2012;97(9):3251-3260.
 344. Rubin KH, Glinborg D, Nybo M, Abrahamsen B, Andersen M. Development and risk factors of type 2 diabetes in a nationwide population of women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2017;102(10):3848-3857.
 345. Apridonidze T, Essah PA, Iuorno MJ, Nestler JE. Prevalence and characteristics of the metabolic syndrome in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2005;90(4):1929-1935.
 346. Paradisi G, Steinberg HO, Hempfling A, *et al.* Polycystic ovary syndrome is associated with endothelial dysfunction. *Circulation*. 2001;103(10):1410-1415.
 347. Orio F Jr, Palomba S, Cascella T, *et al.* Early impairment of endothelial structure and function in young normal-weight women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2004;89(9):4588-4593.
 348. Wekker V, van Dammen L, Koning A, *et al.* Long-term cardiometabolic disease risk in women with PCOS: a systematic review and meta-analysis. *Hum Reprod Update*. 2020;26(6):942-960.
 349. Zhu T, Cui J, Goodarzi MO. Polycystic ovary syndrome and risk of type 2 diabetes, coronary heart disease, and stroke. *Diabetes*. 2021;70(2):627-637.
 350. Ding EL, Song Y, Manson JE, *et al.* Sex hormone-binding globulin and risk of type 2 diabetes in women and men. *N Engl J Med*. 2009;361(12):1152-1163.
 351. O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet*. 2018;50(12):1728-1734.
 352. Cheng Q, Yang Y, Shi X, *et al.* MR-LDP: a two-sample Mendelian randomization for GWAS summary statistics accounting for linkage disequilibrium and horizontal pleiotropy. *NAR Genom Bioinform*. 2020;2(2):lqaa028.
 353. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat Genet*. 2020;52(7):740-747.
 354. Zhu X, Li X, Xu R, Wang T. An iterative approach to detect pleiotropy and perform Mendelian randomization analysis using GWAS summary statistics. *Bioinformatics*. 2021;37(10):1390-1400.
 355. Harris HR, Terry KL. Polycystic ovary syndrome and risk of endometrial, ovarian, and breast cancer: a systematic review. *Fertil Res Pract*. 2016;2:14.
 356. Harris HR, Cushing-Haugen KL, Webb PM, *et al.*; Australian Ovarian Cancer Study Group. Association between genetically predicted polycystic ovary syndrome and ovarian cancer: a Mendelian randomization study. *Int J Epidemiol*. 2019;48(3):822-830.
 357. 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):583-585.
 358. Day FR, Bulik-Sullivan B, Hinds DA, *et al.* Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. *Nat Commun*. 2015;6:8842.
 359. Ibañez L, Potau N, Virdis R, *et al.* Postpubertal outcome in girls diagnosed of premature pubarche during childhood: increased frequency of functional ovarian hyperandrogenism. *J Clin Endocrinol Metab*. 1993;76(6):1599-1603.
 360. Ibañez L, Potau N, Zampolli M, Street ME, Carrascosa A. Girls diagnosed with premature pubarche show an exaggerated ovarian androgen synthesis from the early stages of puberty: evidence from gonadotropin-releasing hormone agonist testing. *Fertil Steril*. 1997;67(5):849-855.
 361. Bronstein J, Tawdekar S, Liu Y, Pawelczak M, David R, Shah B. Age of onset of polycystic ovarian syndrome in girls may be earlier than previously thought. *J Pediatr Adolesc Gynecol*. 2011;24(1):15-20.
 362. Kerchner A, Lester W, Stuart SP, Dokras A. Risk of depression and other mental health disorders in women with polycystic ovary syndrome: a longitudinal study. *Fertil Steril*. 2009;91(1):207-212.
 363. Wray NR, Ripke S, Mattheisen M, *et al.*; eQTLGen; 23andMe; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668-681.
 364. Jiang X, Deng Q, Stener-Victorin E. Is there a shared genetic basis and causal relationship between polycystic ovary syndrome and psychiatric disorders: evidence from a comprehensive genetic analysis. *Hum Reprod*. 2021;36(8):2382-2391.
 365. Stanaway IB, Hall TO, Rosenthal EA, *et al.*; eMERGE Network. The eMERGE genotype set of 83 717 subjects imputed to ~40 million variants genome wide and association with the herpes zoster medical record phenotype. *Genet Epidemiol*. 2019;43(1):63-81.
 366. Joo YY, Actkins K, Pacheco JA, *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):1918-1936.
 367. Azziz R. PCOS in 2015: New insights into the genetics of polycystic ovary syndrome. *Nat Rev Endocrinol*. 2016;12(3):183.
 368. Flannick J, Mercader JM, Fuchsberger C, *et al.*; Broad Genomics Platform; DiscovEHR Collaboration; CHARGE; LuCamp; ProDiGY; GoT2D; ESP; SIGMA-T2D; T2D-GENES; AMP-T2D-GENES. Exome sequencing of 20 791 cases of type 2 diabetes and 24 440 controls. *Nature*. 2019;570(7759):71-76.
 369. Peloso GM, Rader DJ, Gabriel S, Kathiresan S, Daly MJ, Neale BM. Phenotypic extremes in rare variant study designs. *Eur J Hum Genet*. 2016;24(6):924-930.
 370. Shi G, Simino J, Rao DC. Enriching rare variants using family-specific linkage information. *BMC Proc*. 2011;5 Suppl 9:S82.
 371. Freund MK, Burch KS, Shi H, *et al.* Phenotype-specific enrichment of Mendelian disorder genes near GWAS regions across 62 complex traits. *Am J Hum Genet*. 2018;103(4):535-552.

372. Wright CF, West B, Tuke M, *et al.* Assessing the pathogenicity, penetrance, and expressivity of putative disease-causing variants in a population setting. *Am J Hum Genet.* 2019;104(2):275-286.
373. Morgenthaler S, Thilly WG. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res.* 2007;615(1-2):28-56.
374. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet.* 2011;89(1):82-93.
375. Gorsic LK, Kosova G, Werstein B, *et al.* Pathogenic anti-Müllerian hormone variants in polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2017;102(8):2862-2872.
376. Gorsic LK, Dapas M, Legro RS, Hayes MG, Urbanek M. Functional genetic variation in the anti-Müllerian hormone pathway in women with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2019;104(7):2855-2874.
377. Moolhuijsen LME, Visser JA. AMH in PCOS: Controlling the ovary, placenta, or brain? *Curr Opin Endocr Metab Res.* 2020;12:91-97.
378. di Clemente N, Racine C, Pierre A, Taieb J. Anti-Müllerian hormone in female reproduction. *Endocr Rev.* 2021;42(6):753-782.
379. Adzhubei IA, Schmidt S, Peshkin L, *et al.* A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248-249.
380. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003;31(13):3812-3814.
381. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol.* 2010;6(12):e1001025.
382. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46(3):310-315.
383. Josso N, Rey RA, Picard JY. Anti-Müllerian hormone: a valuable addition to the toolbox of the pediatric endocrinologist. *Int J Endocrinol.* 2013;2013:674105.
384. Josso N, Belville C, di Clemente N, Picard JY. AMH and AMH receptor defects in persistent Müllerian duct syndrome. *Hum Reprod Update.* 2005;11(4):351-356.
385. Vasaitis TS, Bruno RD, Njar VC. CYP17 inhibitors for prostate cancer therapy. *J Steroid Biochem Mol Biol.* 2011;125(1-2):23-31.
386. Dapas M, Sisk R, Legro RS, Urbanek M, Dunaif A, Hayes MG. Family-based quantitative trait meta-analysis implicates rare non-coding variants in DENND1A in polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2019;104(9):3835-3850.
387. Tee MK, Speck M, Legeza B, *et al.* Alternative splicing of DENND1A, a PCOS candidate gene, generates variant 2. *Mol Cell Endocrinol.* 2016;434:25-35.
388. Eriksen MB, Nielsen ME, Brusgaard K, *et al.* Genetic alterations within the DENND1A gene in patients with polycystic ovary syndrome (PCOS). *PLoS One.* 2013;8(9):e77186.
389. Khan MJ, Nazli R, Ahmed J, Basit S. Whole genome sequencing instead of whole exome sequencing is required to identify the genetic causes of polycystic ovary syndrome in Pakistani families. *Pak J Med Sci.* 2018;34(3):540-545.
390. Shi J, Yoshino O, Osuga Y, *et al.* Bone morphogenetic protein-6 stimulates gene expression of follicle-stimulating hormone receptor, inhibin/activin beta subunits, and anti-Müllerian hormone in human granulosa cells. *Fertil Steril.* 2009;92(5):1794-1798.
391. Khalaf M, Morera J, Bourret A, *et al.* BMP system expression in GCs from polycystic ovary syndrome women and the in vitro effects of BMP4, BMP6, and BMP7 on GC steroidogenesis. *Eur J Endocrinol.* 2013;168(3):437-444.
392. Liu S, Zhang X, Shi C, *et al.* Altered microRNAs expression profiling in cumulus cells from patients with polycystic ovary syndrome. *J Transl Med.* 2015;13:238.
393. Hu MH, Zheng SX, Yin H, *et al.* Identification of microRNAs that regulate the MAPK pathway in human cumulus cells from PCOS women with insulin resistance. *Reprod Sci.* 2020;27(3):833-844.
394. Zhao J, Huang J, Geng X, *et al.* Polycystic ovary syndrome: novel and hub lncRNAs in the insulin resistance-associated lncRNA-mRNA network. *Front Genet.* 2019;10:772.
395. Pan JX, Tan YJ, Wang FF, *et al.* Aberrant expression and DNA methylation of lipid metabolism genes in PCOS: a new insight into its pathogenesis. *Clin Epigenetics.* 2018;10:6.
396. Wang LP, Peng XY, Lv XQ, *et al.* High throughput circRNAs sequencing profile of follicle fluid exosomes of polycystic ovary syndrome patients. *J Cell Physiol.* 2019;234(9):15537-15547.
397. Walters KA, Handelsman DJ. Androgen receptor splice variants and polycystic ovary syndrome: cause or effect? *Asian J Androl.* 2016;18(3):442-443.
398. Jones MR, Brower MA, Xu N, *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.
399. Wang XX, Wei JZ, Jiao J, Jiang SY, Yu DH, Li D. Genome-wide DNA methylation and gene expression patterns provide insight into polycystic ovary syndrome development. *Oncotarget.* 2014;5(16):6603-6610.
400. Ma M, Ru Y, Chuang LS, *et al.* Disease-associated variants in different categories of disease located in distinct regulatory elements. *BMC Genomics.* 2015;16 Suppl 8:S3.
401. Liu X, Xu X, Hilger D, *et al.* Structural Insights into the Process of GPCR-G Protein Complex Formation. *Cell.* 2019;177(5):1243-1251.e12.
402. Gallagher MD, Chen-Plotkin AS. The Post-GWAS era: from association to function. *Am J Hum Genet.* 2018;102(5):717-730.
403. Maurano MT, Humbert R, Rynes E, *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012;337(6099):1190-1195.
404. van de Bunt M, Cortes A, Brown MA, Morris AP, McCarthy MI; IGAS Consortium. Evaluating the performance of fine-mapping strategies at common variant GWAS loci. *PLoS Genet.* 2015;11(9):e1005535.
405. Watanabe K, Stringer S, Frei O, *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet.* 2019;51(9):1339-1348.
406. Yang J, Benyamin B, McEvoy BP, *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010;42(7):565-569.
407. Shi H, Kichaev G, Pasaniuc B. Contrasting the genetic architecture of 30 complex traits from summary association data. *Am J Hum Genet.* 2016;99(1):139-153.
408. Speed D, Cai N, Johnson MR, Nejentsev S, Balding DJ; UCLEB Consortium. Reevaluation of SNP heritability in complex human traits. *Nat Genet.* 2017;49(7):986-992.
409. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet.* 2012;90(1):7-24.
410. Momozawa Y, Mizukami K. Unique roles of rare variants in the genetics of complex diseases in humans. *J Hum Genet.* 2020.
411. Hernandez RD, Uricchio LH, Hartman K, Ye C, Dahl A, Zaitlen N. Ultrarare variants drive substantial cis heritability of human gene expression. *Nat Genet.* 2019;51(9):1349-1355.
412. Trynka G, Hunt KA, Bockett NA, *et al.*; Spanish Consortium on the Genetics of Coeliac Disease (CEGEC); PreventCD Study Group; Wellcome Trust Case Control Consortium (WTCCC). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet.* 2011;43(12):1193-1201.
413. Marouli E, Graff M, Medina-Gomez C, *et al.*; EPIC-InterAct Consortium; CHD Exome+ Consortium; ExomeBP Consortium; T2D-Genes Consortium; GoT2D Genes Consortium; Global Lipids Genetics Consortium; ReproGen Consortium; MAGIC Investigators. Rare and low-frequency coding variants alter human adult height. *Nature.* 2017;542(7640):186-190.
414. Xue A, Wu Y, Zhu Z, *et al.*; eQTLGen Consortium. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun.* 2018;9(1):2941.

415. Marth GT, Yu F, Indap AR, *et al.*; 1000 Genomes Project. The functional spectrum of low-frequency coding variation. *Genome Biol.* 2011;12(9):R84.
416. Tajima T, Morita H, Ito K, *et al.* Blood lipid-related low-frequency variants in LDLR and PCSK9 are associated with onset age and risk of myocardial infarction in Japanese. *Sci Rep.* 2018;8(1):8107.
417. Natarajan P, Peloso GM, Zekavat SM, *et al.*; NHLBI TOPMed Lipids Working Group. Deep-coverage whole genome sequences and blood lipids among 16 324 individuals. *Nat Commun.* 2018;9(1):3391.
418. Lange LA, Hu Y, Zhang H, *et al.*; NHLBI Grand Opportunity Exome Sequencing Project. Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am J Hum Genet.* 2014;94(2):233-245.
419. Sims R, van der Lee SJ, Naj AC, *et al.*; ARUK Consortium; GERAD/PERADES, CHARGE, ADGC, EADI. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat Genet.* 2017;49(9):1373-1384.
420. Conway OJ, Carrasquillo MM, Wang X, *et al.* ABI3 and PLCG2 missense variants as risk factors for neurodegenerative diseases in Caucasians and African Americans. *Mol Neurodegener.* 2018;13(1):53.
421. Rotwein P. Revisiting the population genetics of human height. *J Endocr Soc.* 2020;4(4):bvaa025.
422. Turcot V, Lu Y, Highland HM, *et al.*; CHD Exome+ Consortium; EPIC-CVD Consortium; ExomeBP Consortium; Global Lipids Genetic Consortium; GoT2D Genes Consortium; EPIC InterAct Consortium; INTERVAL Study; ReproGen Consortium; T2D-Genes Consortium; MAGIC Investigators; Understanding Society Scientific Group. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet.* 2018;50(1):26-41.
423. 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):151-155.
424. Wang B, Zhou S, Wang J, *et al.* Identification of novel missense mutations of GDF9 in Chinese women with polycystic ovary syndrome. *Reprod Biomed Online.* 2010;21(3):344-348.
425. Li S, Zhao L, Wan XH. A missense variant rs4645843 in TNF- α gene is a risk factor of polycystic ovary syndrome in the Uygur population. *Toboku J Exp Med.* 2017;243(2):95-100.
426. Gusev A, Lee SH, Trynka G, *et al.*; Schizophrenia Working Group of the Psychiatric Genomics Consortium; SWE-SCZ Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; SWE-SCZ Consortium. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am J Hum Genet.* 2014;95(5):535-552.
427. Spataro N, Rodríguez JA, Navarro A, Bosch E. Properties of human disease genes and the role of genes linked to Mendelian disorders in complex disease aetiology. *Hum Mol Genet.* 2017;26(3):489-500.
428. Tallapragada DS, Bhaskar S, Chandak GR. New insights from monogenic diabetes for "common" type 2 diabetes. *Front Genet.* 2015;6:251.
429. Yang Y, Chan L. Monogenic diabetes: what it teaches us on the common forms of type 1 and type 2 diabetes. *Endocr Rev.* 2016;37(3):190-222.
430. Winckler W, Weedon MN, Graham RR, *et al.* Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes.* 2007;56(3):685-693.
431. Pigny P, Merlen E, Robert Y, *et al.* Elevated serum level of anti-Müllerian hormone in patients with polycystic ovary syndrome: relationship to the ovarian follicle excess and to the follicular arrest. *J Clin Endocrinol Metab.* 2003;88(12):5957-5962.
432. Wray NR, Maier R. Genetic basis of complex genetic disease: the contribution of disease heterogeneity to missing heritability. *Curr Epidemiol Rep.* 2014;1(4):220-227.
433. Manchia M, Cullis J, Turecki G, Rouleau GA, Uher R, Alda M. The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. *PLoS One.* 2013;8(10):e76295.
434. Traylor M, Markus H, Lewis CM. Homogeneous case subgroups increase power in genetic association studies. *Eur J Hum Genet.* 2015;23(6):863-869.
435. Saria S, Goldenberg A. Subtyping: what it is and its role in precision medicine. *IEEE Intell Syst.* 2015;30(4):70-75.
436. Goldzieher JW, Axelrod LR. Clinical and biochemical features of polycystic ovarian disease. *Fertil Steril.* 1963;14:631-653.
437. 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.
438. Goodarzi MO. Genetics of common endocrine disease: the present and the future. *J Clin Endocrinol Metab.* 2016;101(3):787-794.
439. Tzeng CR, Chang YC, Chang YC, Wang CW, Chen CH, Hsu MI. Cluster analysis of cardiovascular and metabolic risk factors in women of reproductive age. *Fertil Steril.* 2014;101(5):1404-1410.
440. Dapas M, Lin FTJ, Nadkarni GN, *et al.* Distinct subtypes of polycystic ovary syndrome with novel genetic associations: An unsupervised, phenotypic clustering analysis. *PLoS Med.* 2020;17(6):e1003132.
441. Hennig C. Cluster-wise assessment of cluster stability. *Comput Stat Data An.* 2007;52(1):258-271.
442. Shimasaki S, Moore RK, Otsuka F, Erickson GF. The bone morphogenetic protein system in mammalian reproduction. *Endocr Rev.* 2004;25(1):72-101.
443. Estienne A, Pierre A, di Clemente N, *et al.* Anti-Müllerian hormone regulation by the bone morphogenetic proteins in the sheep ovary: deciphering a direct regulatory pathway. *Endocrinology.* 2015;156(1):301-313.
444. Yi SE, LaPolt PS, Yoon BS, Chen JY, Lu JK, Lyons KM. The type I BMP receptor Bmpr1B is essential for female reproductive function. *Proc Natl Acad Sci U S A.* 2001;98(14):7994-7999.
445. Jiao J, Shi B, Wang T, *et al.* Characterization of long non-coding RNA and messenger RNA profiles in follicular fluid from mature and immature ovarian follicles of healthy women and women with polycystic ovary syndrome. *Hum Reprod.* 2018;33(9):1735-1748.
446. Di Zazzo E, De Rosa C, Abbondanza C, Monchamont B. PRDM proteins: molecular mechanisms in signal transduction and transcriptional regulation. *Biology (Basel).* 2013;2(1):107-141.
447. Forrest AR, Kawaji H, Rehli M, *et al.*; FANTOM Consortium and the RIKEN PMI and CLST (DGT). A promoter-level mammalian expression atlas. *Nature.* 2014;507(7493):462-470.
448. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-585.
449. Liu L, Shao G, Steele-Perkins G, Huang S. The retinoblastoma interacting zinc finger gene RIZ produces a PR domain-lacking product through an internal promoter. *J Biol Chem.* 1997;272(5):2984-2991.
450. Andreu-Vieyra C, Chen R, Matzuk MM. Conditional deletion of the retinoblastoma (Rb) gene in ovarian granulosa cells leads to premature ovarian failure. *Mol Endocrinol.* 2008;22(9):2141-2161.
451. Yang QE, Nagaoka SI, Gwost I, Hunt PA, Oatley JM. Inactivation of retinoblastoma protein (Rb1) in the oocyte: evidence that dysregulated follicle growth drives ovarian teratoma formation in mice. *PLoS Genet.* 2015;11(7):e1005355.
452. Pau CT, Mosbrugger 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.
453. Carmina E, Legro RS, Stamets K, Lowell J, Lobo RA. Difference in body weight between American and Italian women with

- polycystic ovary syndrome: influence of the diet. *Hum Reprod.* 2003;18(11):2289-2293.
454. Pasquali R, Gambineri A, Pagotto U. The impact of obesity on reproduction in women with polycystic ovary syndrome. *BJOG.* 2006;113(10):1148-1159.
 455. Jiao J, Fang Y, Wang T, Wang Z, Zhou M, Wang X. Epidemiologic investigation of polycystic ovarian syndrome (PCOS) in Han ethnic women of reproductive age in Liaoning Province, China. *Clin Exp Obstet Gynecol.* 2014;41(3):304-309.
 456. Balen AH, Conway GS, Kaltsas G, et al. Polycystic ovary syndrome: the spectrum of the disorder in 1741 patients. *Hum Reprod.* 1995;10(8):2107-2111.
 457. Holte J, Bergh T, Gennarelli G, Wide L. The independent effects of polycystic ovary syndrome and obesity on serum concentrations of gonadotrophins and sex steroids in premenopausal women. *Clin Endocrinol (Oxf).* 1994;41(4):473-481.
 458. Barber TM, McCarthy MI, Wass JA, Franks S. Obesity and polycystic ovary syndrome. *Clin Endocrinol (Oxf).* 2006;65(2):137-145.
 459. Ezeh U, Yildiz BO, Azziz R. Referral bias in defining the phenotype and prevalence of obesity in polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2013;98(6):E1088-E1096.
 460. Speliotes EK, Willer CJ, Berndt SI, et al.; MAGIC; Procardis Consortium. Association analyses of 249 796 individuals reveal 18 new loci associated with body mass index. *Nat Genet.* 2010;42(11):937-948.
 461. Barber TM, Bennett AJ, Groves CJ, et al. Association of variants in the fat mass and obesity associated (FTO) gene with polycystic ovary syndrome. *Diabetologia.* 2008;51(7):1153-1158.
 462. Wojciechowski P, Lipowska A, Rys P, et al.; GIANT Consortium. Impact of FTO genotypes on BMI and weight in polycystic ovary syndrome: a systematic review and meta-analysis. *Diabetologia.* 2012;55(10):2636-2645.
 463. Louwers YV, Rayner NW, Herrera BM, et al. BMI-associated alleles do not constitute risk alleles for polycystic ovary syndrome independently of BMI: a case-control study. *PLoS One.* 2014;9(1):e87335.
 464. Simmons R. Developmental origins of adult metabolic disease: concepts and controversies. *Trends Endocrinol Metab.* 2005;16(8):390-394.
 465. Hanson MA, Gluckman PD. Early developmental conditioning of later health and disease: physiology or pathophysiology? *Physiol Rev.* 2014;94(4):1027-1076.
 466. King SE, Skinner MK. Epigenetic Transgenerational Inheritance of Obesity Susceptibility. *Trends Endocrinol Metab.* 2020;31(7):478-494.
 467. Kobaly K, Vellanki P, Sisk RK, et al. Parent-of-origin effects on glucose homeostasis in polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2014;99(8):2961-2966.
 468. Barker DJ, Osmond C. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *Lancet.* 1986;1(8489):1077-1081.
 469. Rich-Edwards JW, Stampfer MJ, Manson JE, et al. Birth weight and risk of cardiovascular disease in a cohort of women followed up since 1976. *BMJ.* 1997;315(7105):396-400.
 470. Ibáñez L, de Zegher F, Potau N. Premature pubarche, ovarian hyperandrogenism, hyperinsulinism and the polycystic ovary syndrome: from a complex constellation to a simple sequence of prenatal onset. *J Endocrinol Invest.* 1998;21(9):558-566.
 471. Ibáñez L, Potau N, Francois I, de Zegher F. Precocious pubarche, hyperinsulinism, and ovarian hyperandrogenism in girls: relation to reduced fetal growth. *J Clin Endocrinol Metab.* 1998;83(10):3558-3562.
 472. Sadrzadeh S, Klip WA, Broekmans FJ, et al.; OMEGA Project group. Birth weight and age at menarche in patients with polycystic ovary syndrome or diminished ovarian reserve, in a retrospective cohort. *Hum Reprod.* 2003;18(10):2225-2230.
 473. Boomsma CM, Eijkemans MJ, Hughes EG, Visser GH, Fauser BC, Macklon NS. A meta-analysis of pregnancy outcomes in women with polycystic ovary syndrome. *Hum Reprod Update.* 2006;12(6):673-683.
 474. Legro RS, Roller RL, Dodson WC, Stetter CM, Kunselman AR, Dunaif A. Associations of birthweight and gestational age with reproductive and metabolic phenotypes in women with polycystic ovarian syndrome and their first-degree relatives. *J Clin Endocrinol Metab.* 2010;95(2):789-799.
 475. Anderson H, Fogel N, Grebe SK, Singh RJ, Taylor RL, Dunaif A. Infants of women with polycystic ovary syndrome have lower cord blood androstenedione and estradiol levels. *J Clin Endocrinol Metab.* 2010;95(5):2180-2186.
 476. Paschou SA, Ioannidis D, Vassilatou E, et al. Birth weight and polycystic ovary syndrome in adult life: is there a causal link? *PLoS One.* 2015;10(3):e0122050.
 477. Sadrzadeh S, Painter RC, Lambalk CB. Developmental origins of polycystic ovary syndrome (PCOS), a case-control study comparing birth weight in women with PCOS and control group. *Gynecol Endocrinol.* 2016;32(10):856-859.
 478. de Zegher F, Reinehr T, Malpique R, Darendeliler F, López-Bermejo A, Ibáñez L. Reduced prenatal weight gain and/or augmented postnatal weight gain precedes polycystic ovary syndrome in adolescent girls. *Obesity (Silver Spring).* 2017;25(9):1486-1489.
 479. Melo AS, Vieira CS, Barbieri MA, et al. High prevalence of polycystic ovary syndrome in women born small for gestational age. *Hum Reprod.* 2010;25(8):2124-2131.
 480. Reagan PB, Salsberry PJ, Fang MZ, Gardner WP, Pajer K. African-American/white differences in the age of menarche: accounting for the difference. *Soc Sci Med.* 2012;75(7):1263-1270.
 481. Terry MB, Ferris JS, Tehranifar P, Wei Y, Flom JD. Birth weight, postnatal growth, and age at menarche. *Am J Epidemiol.* 2009;170(1):72-79.
 482. Zhao Y, Qiao J. Ethnic differences in the phenotypic expression of polycystic ovary syndrome. *Steroids.* 2013;78(8):755-760.
 483. Dumesic DA, Abbott DH, Eisner JR, Goy RW. Prenatal exposure of female rhesus monkeys to testosterone propionate increases serum luteinizing hormone levels in adulthood. *Fertil Steril.* 1997;67(1):155-163.
 484. Abbott DH, Dumesic DA, Eisner JR, Colman RJ, Kennitz JW. Insights into the development of polycystic ovary syndrome (PCOS) from studies of prenatally androgenized female rhesus monkeys. *Trends Endocrinol Metab.* 1998;9(2):62-67.
 485. Abbott DH, Dumesic DA, Franks S. Developmental origin of polycystic ovary syndrome - a hypothesis. *J Endocrinol.* 2002;174(1):1-5.
 486. Eisner JR, Dumesic DA, Kennitz JW, Colman RJ, Abbott DH. Increased adiposity in female rhesus monkeys exposed to androgen excess during early gestation. *Obes Res.* 2003;11(2):279-286.
 487. Abbott DH, Barnett DK, Bruns CM, Dumesic DA. Androgen excess fetal programming of female reproduction: a developmental aetiology for polycystic ovary syndrome? *Hum Reprod Update.* 2005;11(4):357-374.
 488. Zhou R, Bird IM, Dumesic DA, Abbott DH. Adrenal hyperandrogenism is induced by fetal androgen excess in a rhesus monkey model of polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2005;90(12):6630-6637.
 489. Sharma TP, Herkimer C, West C, et al. Fetal programming: prenatal androgen disrupts positive feedback actions of estradiol but does not affect timing of puberty in female sheep. *Biol Reprod.* 2002;66(4):924-933.
 490. Dumesic DA, Abbott DH, Padmanabhan V. Polycystic ovary syndrome and its developmental origins. *Rev Endocr Metab Disord.* 2007;8(2):127-141.
 491. Padmanabhan V, Veiga-Lopez A. Developmental origin of reproductive and metabolic dysfunctions: androgenic versus estrogenic reprogramming. *Semin Reprod Med.* 2011;29(3):173-186.
 492. Demissie M, Lazic M, Foecking EM, Aird F, Dunaif A, Levine JE. Transient prenatal androgen exposure produces metabolic syndrome in adult female rats. *Am J Physiol Endocrinol Metab.* 2008;295(2):E262-E268.

493. Shi D, Vine DF. Animal models of polycystic ovary syndrome: a focused review of rodent models in relationship to clinical phenotypes and cardiometabolic risk. *Fertil Steril*. 2012;98(1):185-193.
494. Motta AB. Dehydroepiandrosterone to induce murine models for the study of polycystic ovary syndrome. *J Steroid Biochem Mol Biol*. 2010;119(3-5):105-111.
495. Stener-Victorin E, Padmanabhan V, Walters KA, et al. Animal models to understand the etiology and pathophysiology of polycystic ovary syndrome. *Endocr Rev*. 2020;41(4):bnaa010.
496. Sir-Petermann T, Maliqueo M, Angel B, Lara HE, Pérez-Bravo F, Recabarren SE. Maternal serum androgens in pregnant women with polycystic ovarian syndrome: possible implications in prenatal androgenization. *Hum Reprod*. 2002;17(10):2573-2579.
497. Cesta CE, Öberg AS, Ibrahimson A, et al. Maternal polycystic ovary syndrome and risk of neuropsychiatric disorders in offspring: prenatal androgen exposure or genetic confounding? *Psychol Med*. 2020;50(4):616-624.
498. Smith SW, Axelrod LR. Studies on the metabolism of steroid hormones and their precursors by the human placenta at various stages of gestation. II. In vitro metabolism of 3 beta-hydroxyandrost-5-en-17-one. *J Clin Endocrinol Metab*. 1969;29(9):1182-1190.
499. Payne AH, Hales DB. Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. *Endocr Rev*. 2004;25(6):947-970.
500. Maliqueo M, Lara HE, Sánchez F, Echiburú B, Crisosto N, Sir-Petermann T. Placental steroidogenesis in pregnant women with polycystic ovary syndrome. *Eur J Obstet Gynecol Reprod Biol*. 2013;166(2):151-155.
501. Longcope C. Adrenal and gonadal androgen secretion in normal females. *Clin Endocrinol Metab*. 1986;15(2):213-228.
502. Payne AH, Jaffe RB. Androgen formation from pregnenolone sulfate by the human fetal ovary. *J Clin Endocrinol Metab*. 1974;39(2):300-304.
503. Cole B, Hensinger K, Maciel GA, Chang RJ, Erickson GF. Human fetal ovary development involves the spatiotemporal expression of p450c17 protein. *J Clin Endocrinol Metab*. 2006;91(9):3654-3661.
504. Torchen LC, Idkowiak J, Fogel NR, et al. Evidence for increased 5 α -reductase activity during early childhood in daughters of women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2016;101(5):2069-2075.
505. Abbott DH, Dumesic DA. Passing on PCOS: new insights into its epigenetic transmission. *Cell Metab*. 2021;33(3):463-466.
506. Barry JA, Kay AR, Navaratnarajah R, et al. Umbilical vein testosterone in female infants born to mothers with polycystic ovary syndrome is elevated to male levels. *J Obstet Gynaecol*. 2010;30(5):444-446.
507. Daan NM, Koster MP, Steegers-Theunissen RP, Eijkemans MJ, Fauser BC. Endocrine and cardiometabolic cord blood characteristics of offspring born to mothers with and without polycystic ovary syndrome. *Fertil Steril*. 2017;107(1):261-268.e3.
508. Duan C, Pei T, Li Y, Cao Q, Chen H, Fu J. Androgen levels in the fetal cord blood of children born to women with polycystic ovary syndrome: a meta-analysis. *Reprod Biol Endocrinol*. 2020;18(1):81.
509. Caanen MR, Kuijper EA, Hompes PG, et al. Mass spectrometry methods measured androgen and estrogen concentrations during pregnancy and in newborns of mothers with polycystic ovary syndrome. *Eur J Endocrinol*. 2016;174(1):25-32.
510. Hickey M, Sloboda DM, Atkinson HC, et al. The relationship between maternal and umbilical cord androgen levels and polycystic ovary syndrome in adolescence: a prospective cohort study. *J Clin Endocrinol Metab*. 2009;94(10):3714-3720.
511. Manning JT, Scutt D, Wilson J, Lewis-Jones DI. The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Hum Reprod*. 1998;13(11):3000-3004.
512. McIntyre MH. The use of digit ratios as markers for perinatal androgen action. *Reprod Biol Endocrinol*. 2006;4:10.
513. Dean A, Sharpe RM. Clinical review: Anogenital distance or digit length ratio as measures of fetal androgen exposure: relationship to male reproductive development and its disorders. *J Clin Endocrinol Metab*. 2013;98(6):2230-2238.
514. Cattrall FR, Vollenhoven BJ, Weston GC. Anatomical evidence for in utero androgen exposure in women with polycystic ovary syndrome. *Fertil Steril*. 2005;84(6):1689-1692.
515. Lujan ME, Bloski TG, Chizen DR, Lehotay DC, Pierson RA. Digit ratios do not serve as anatomical evidence of prenatal androgen exposure in clinical phenotypes of polycystic ovary syndrome. *Hum Reprod*. 2010;25(1):204-211.
516. Wu Y, Zhong G, Chen S, Zheng C, Liao D, Xie M. Polycystic ovary syndrome is associated with anogenital distance, a marker of prenatal androgen exposure. *Hum Reprod*. 2017;32(4):937-943.
517. Sánchez-Ferrer ML, Mendiola J, Hernández-Peñalver AI, et al. Presence of polycystic ovary syndrome is associated with longer anogenital distance in adult Mediterranean women. *Hum Reprod*. 2017;32(11):2315-2323.
518. Barrett ES, Hoeger KM, Sathyanarayana S, et al. Anogenital distance in newborn daughters of women with polycystic ovary syndrome indicates fetal testosterone exposure. *J Dev Orig Health Dis*. 2018;9(3):307-314.
519. Glinborg D, Jensen RC, Schmedes AV, et al. Anogenital distance in children born of mothers with polycystic ovary syndrome: the Odense Child Cohort. *Hum Reprod*. 2019;34(10):2061-2070.
520. Robertson KD. DNA methylation and human disease. *Nat Rev Genet*. 2005;6(8):597-610.
521. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res*. 2011;21(3):381-395.
522. Tammen SA, Friso S, Choi SW. Epigenetics: the link between nature and nurture. *Mol Aspects Med*. 2013;34(4):753-764.
523. Cavalli G, Heard E. Advances in epigenetics link genetics to the environment and disease. *Nature*. 2019;571(7766):489-499.
524. Zeng Y, Chen T. DNA methylation reprogramming during mammalian development. *Genes (Basel)*. 2019;10(4):257.
525. Horsthemke B. A critical view on transgenerational epigenetic inheritance in humans. *Nat Commun*. 2018;9(1):2973.
526. Stener-Victorin E. Epigenetic and transgenerational transmission of polycystic ovary syndrome. *Curr Opin Endocr Metab Res*. 2020;12:72-77.
527. Nilsson EE, Sadler-Riggleman I, Skinner MK. Environmentally induced epigenetic transgenerational inheritance of disease. *Environ Epigenet*. 2018;4(2):dvy016.
528. Risal S, Pei Y, Lu H, et al. Prenatal androgen exposure and transgenerational susceptibility to polycystic ovary syndrome. *Nat Med*. 2019;25(12):1894-1904.
529. McRae AF, Powell JE, Henders AK, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol*. 2014;15(5):R73.
530. van Dongen J, Nivard MG, Willemsen G, et al.; BIOS Consortium. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun*. 2016;7:11115.
531. Hannon E, Knox O, Sugden K, et al. Characterizing genetic and environmental influences on variable DNA methylation using monozygotic and dizygotic twins. *PLoS Genet*. 2018;14(8):e1007544.
532. Bourrat P, Lu Q. Dissolving the missing heritability problem. *Philos Sci*. 2017;84(5):1055-1067.
533. Sagvekar P, Kumar P, Mangoli V, Desai S, Mukherjee S. DNA methylome profiling of granulosa cells reveals altered methylation in genes regulating vital ovarian functions in polycystic ovary syndrome. *Clin Epigenetics*. 2019;11(1):61.
534. Kokosar M, Benrick A, Perfilyev A, et al. Epigenetic and transcriptional alterations in human adipose tissue of polycystic ovary syndrome. *Sci Rep*. 2016;6:22883.
535. Nilsson E, Benrick A, Kokosar M, et al. Transcriptional and epigenetic changes influencing skeletal muscle metabolism in women

- with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2018;103(12):4465-4477.
536. Lambertini L, Saul SR, Copperman AB, *et al.* Intrauterine reprogramming of the polycystic ovary syndrome: evidence from a pilot study of cord blood global methylation analysis. *Front Endocrinol (Lausanne).* 2017;8:352.
537. Mimouni NEH, Paiva I, Barbotin AL, *et al.* Polycystic ovary syndrome is transmitted via a transgenerational epigenetic process. *Cell Metab.* 2021;33(3):513-530.e8.
538. Roland AV, Nunemaker CS, Keller SR, Moenter SM. Prenatal androgen exposure programs metabolic dysfunction in female mice. *J Endocrinol.* 2010;207(2):213-223.
539. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* 2017;45(D1):D331-D338.
540. Kahsar-Miller MD, Nixon C, Boots LR, Go RC, Azziz R. Prevalence of polycystic ovary syndrome (PCOS) in first-degree relatives of patients with PCOS. *Fertil Steril.* 2001;75(1):53-58.
541. Kahsar-Miller M, Azziz R. The development of the polycystic ovary syndrome: family history as a risk factor. *Trends Endocrinol Metab.* 1998;9(2):55-58.
542. Cheng C, Zhang H, Zhao Y, Li R, Qiao J. Paternal history of diabetes mellitus and hypertension affects the prevalence and phenotype of PCOS. *J Assist Reprod Genet.* 2015;32(12):1731-1739.
543. Champroux A, Cocquet J, Henry-Berger J, Drevet JR, Kocer A. A decade of exploring the mammalian sperm epigenome: paternal epigenetic and transgenerational inheritance. *Front Cell Dev Biol.* 2018;6:50.
544. van Otterdijk SD, Michels KB. Transgenerational epigenetic inheritance in mammals: how good is the evidence? *Faseb J.* 2016;30(7):2457-2465.
545. Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM. Common disease is more complex than implied by the core gene omnigenic model. *Cell.* 2018;173(7):1573-1580.
546. Mutharasan P, Galdones E, Peñalver Bernabé B, *et al.* Evidence for chromosome 2p16.3 polycystic ovary syndrome susceptibility locus in affected women of European ancestry. *J Clin Endocrinol Metab.* 2013;98(1):E185-E190.
547. Laven JSE. Follicle stimulating hormone receptor (FSHR) polymorphisms and polycystic ovary syndrome (PCOS). *Front Endocrinol (Lausanne).* 2019;10:23.
548. Gravel S, Henn BM, Gutenkunst RN, *et al.*; 1000 Genomes Project. Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci U S A.* 2011;108(29):11983-11988.
549. Goodarzi MO, Jones MR, Li X, *et al.* Replication of association of DENND1A and THADA variants with polycystic ovary syndrome in European cohorts. *J Med Genet.* 2012;49(2):90-95.
550. Louwers YV, Stolk L, Uitterlinden AG, Laven JS. Cross-ethnic meta-analysis of genetic variants for polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2013;98(12):E2006-E2012.
551. Qin L, Zhao S, Yang P, *et al.* Variation analysis of anti-Müllerian hormone gene in Chinese women with polycystic ovary syndrome. *Endocrine.* 2021;72(1):287-293.
552. Bien SA, Pankow JS, Haessler J, *et al.* Transethnic insight into the genetics of glycaemic traits: fine-mapping results from the Population Architecture using Genomics and Epidemiology (PAGE) consortium. *Diabetologia.* 2017;60(12):2384-2398.
553. Fernández-Rhodes L, Gong J, Haessler J, *et al.* Trans-ethnic fine-mapping of genetic loci for body mass index in the diverse ancestral populations of the Population Architecture using Genomics and Epidemiology (PAGE) Study reveals evidence for multiple signals at established loci. *Hum Genet.* 2017;136(6):771-800.
554. Ke W, Rand KA, Conti DV, *et al.* Evaluation of 71 coronary artery disease risk variants in a multiethnic cohort. *Front Cardiovasc Med.* 2018;5:19.
555. Asimit JL, Hatzikotoulas K, McCarthy M, Morris AP, Zeggini E. Trans-ethnic study design approaches for fine-mapping. *Eur J Hum Genet.* 2016;24(9):1330-1336.
556. Lonjou C, Zhang W, Collins A, *et al.* Linkage disequilibrium in human populations. *Proc Natl Acad Sci U S A.* 2003;100(10):6069-6074.
557. Rees E, Carrera N, Morgan J, *et al.*; GROUP Investigators. Targeted sequencing of 10 198 samples confirms abnormalities in neuronal activity and implicates voltage-gated sodium channels in schizophrenia pathogenesis. *Biol Psychiatry.* 2019;85(7):554-562.
558. Davis CA, Hitz BC, Sloan CA, *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* 2018;46(D1):D794-D801.
559. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57-74.
560. Zuo C, Shin S, Keleş S. atSNP: transcription factor binding affinity testing for regulatory SNP detection. *Bioinformatics.* 2015;31(20):3353-3355.
561. Marbach D, Lamparter D, Quon G, Kellis M, Kutalik Z, Bergmann S. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat Methods.* 2016;13(4):366-370.
562. Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* 2017;9(1):75.
563. Udler MS, McCarthy MI, Florez JC, Mahajan A. Genetic risk scores for diabetes diagnosis and precision medicine. *Endocr Rev.* 2019;40(6):1500-1520.
564. Hughes E, Tshiaba P, Wagner S, *et al.* Integrating clinical and polygenic factors to predict breast cancer risk in women undergoing genetic testing. *JCO Precis Oncol.* 2021;5:307-316.
565. Weale ME, Riveros-Mckay F, Selzam S, *et al.* Validation of an integrated risk tool, including polygenic risk score, for atherosclerotic cardiovascular disease in multiple ethnicities and ancestries. *Am J Cardiol.* 2021;148:157-164.
566. Riveros-Mckay F, Weale ME, Moore R, *et al.* Integrated polygenic tool substantially enhances coronary artery disease prediction. *Circ Genom Precis Med.* 2021;14(2):e003304.
567. Makrinou E, Drong AW, Christopoulos G, *et al.* Genome-wide methylation profiling in granulosa lutein cells of women with polycystic ovary syndrome (PCOS). *Mol Cell Endocrinol.* 2020;500:110611.
568. Gibson-Helm M, Teede H, Dunaif A, Dokras A. Delayed diagnosis and a lack of information associated with dissatisfaction in women with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2017;102(2):604-612.
569. Lin AW, Bergomi EJ, Dollahite JS, Sobal J, Hoeger KM, Lujan ME. Trust in physicians and medical experience beliefs differ between women with and without polycystic ovary syndrome. *J Endocr Soc.* 2018;2(9):1001-1009.