HHS Public Access

Author manuscript

Curr Opin Endocr Metab Res. Author manuscript; available in PMC 2021 June 01.

Published in final edited form as:

Curr Opin Endocr Metab Res. 2020 June ; 12: 26–32. doi:10.1016/j.coemr.2020.02.011.

The contribution of rare genetic variants to the pathogenesis of polycystic ovary syndrome

Matthew Dapas, Ph.D., **Andrea Dunaif, M.D.**

Abstract

Polycystic ovary syndrome (PCOS) is a highly heritable disorder, but only a small proportion of the heritability can be accounted for by common genetic risk variants identified to date. It is possible that variants with lower allele frequencies that cannot be detected using genome-wide association study arrays contribute to PCOS. Here, we discuss the challenges inherent to studying rare genetic variants in complex disease and review several recent studies that have used DNA sequencing techniques to investigate whether rare variants play a role in PCOS pathogenesis. We evaluate these findings in the context of the latest literature in PCOS and complex disease genetics.

PCOS is a highly heritable complex genetic disorder (1). Genome-wide association studies (GWAS) have reproducibly mapped close to 20 susceptibility loci associated with PCOS diagnosis. The first two GWAS were completed in Han Chinese PCOS case-control cohorts fulfilling the Rotterdam diagnostic criteria. These GWAS together identified eleven PCOS association signals (2,3), several of which have since been replicated in targeted genotyping studies in European ancestry PCOS (4–6) (Table 1). The first GWAS in European ancestry PCOS included only cases fulfilling NIH diagnostic criteria (7). It identified two novel loci and replicated one of the Han Chinese loci. The second European ancestry GWAS included cases fulfilling Rotterdam diagnostic criteria so that non-NIH Rotterdam phenotypes were studied, in addition to NIH phenotype cases. A large number of cases with self-reported PCOS (8) were also studied. This GWAS replicated one signal from the first European GWAS and two from the Han Chinese GWAS, in addition to identifying three novel loci. A recent meta-analysis of European ancestry PCOS GWAS (9) identified three novel loci and replicated 11 of the previously reported loci (Table 1). The meta-analysis contained cases diagnosed by NIH and Rotterdam criteria as well as those diagnosed by self-report. However, only one locus differed significantly in its association by diagnostic criteria; otherwise, the genetic architecture was similar in NIH and non-NIH Rotterdam PCOS phenotypes as well as in self-reported PCOS cases across common variants at 13 loci.

These GWAS have significantly advanced our understanding of the pathophysiology of PCOS by implicating gonadotropin secretion (*FSHB*) and action (*LHCGR, FSHR*),

The authors have no conflicts of interest to disclose.

androgen biosynthesis (*DENND1A*), metabolic regulation (*THADA, INSR*,) and follicle development (*YAP1, HMGA2*) in PCOS pathogenesis. Nevertheless, GWAS have important limitations. GWAS are based on the hypothesis that common variants, typically defined by minor allele frequencies (MAFs) of 5% or greater, cause common, complex traits/diseases, such as PCOS (10). GWAS arrays measure single nucleotide polymorphisms (SNPs) that span the entire genome. The GWAS SNPs are located within segments of the genome where genetic variation is highly correlated or in 'linkage disequilibrium' (LD) (11). These SNPs serve as markers, so-called "tag SNPs", for the surrounding genomic regions, or haplotypes. Therefore, trait-associated alleles identified via GWAS are often not the causal variants themselves, but rather are associated or in LD with the underlying causal variant(s) within the same region.

Additional statistical methods (12–15) and DNA sequencing of GWAS loci (16,17) are required to identify potential pathogenic variants and candidate genes within a given region in a process known as "fine mapping". However, causality can only be proven with functional studies (18). The majority of GWAS variants are in noncoding portions of the genome, which makes elucidating their functional significance exceptionally challenging (19). The exact mechanisms by which most GWAS susceptibility loci contribute to disease pathogenesis remain unknown (20). Indeed, while PCOS GWAS have implicated many plausible candidate genes, the corresponding pathogenic variants within these putative causal genes have yet to be identified (21).

By definition, GWAS variants are prevalent in the population and, therefore, would not be expected to have large effects on phenotype (1). Consistent with this prediction, most of the risk variants identified from GWAS have modest effect sizes (22). In many complex traits/diseases, including PCOS, GWAS variants identified to date account for less than 10% of the estimated heritability (the phenotypic variance attributable to genetic differences) of the disorder (23,24). The prevailing hypothesis for this deficit in heritability had been that variants with lower frequency that cannot be detected using GWAS arrays contribute to complex traits (23,25,26). It was anticipated that these rare variants would be in coding portions of the genome where they would have large effects on phenotype by disrupting the encoded molecule. Consequently, elucidating the biologic relevance of such coding rare variants would be amenable to traditional molecular approaches. Support for this hypothesis came from studies that found rare genetic variants with large effects on HDL (27), adiponectin (28) and triglyceride levels (29) as well as on blood pressure (30) in the general population.

Although the precise definition varies from study to study, the term "rare variants" generally refers to variants that have a MAF on the order of less than 1-2% in a given population. Testing for rare variants has only recently become feasible because of cost-reductions in DNA sequencing technologies (31–33). Rare variants can be inferred, or "imputed", from GWAS data using large genomic databases, but this approach is only reliable down to a MAF of ~0.1% (34,35) and is not yet available for many ancestral populations (genetic variation is ancestry-dependent). Successful utilization of sequencing data to study rare variants in a complex disease framework still requires overcoming a number of unique challenges. First, rare variants occur too infrequently to perform standard case-control allele

association tests without recruiting prohibitively large cohorts. For example, having 80% power to detect an association for a variant with 0.001% MAF and a phenotypic effect size of one standard deviation ($\beta \approx 4$) would require a sample size over one million (36). Populations can be enriched for causal rare variants by studying extreme phenotypes or by incorporating families with multiple affected individuals (26), but even in enriched cohorts, most rare variant allele frequencies are still too low for individual variant associations (37), particularly because—in contrast to Mendelian disorders—pathogenic alleles in complex traits are not individually predictive of phenotypes (38,39). Collapsing methods that aggregate sets of rare variants within the same gene or genomic region into single statistics, such as burden tests (40) or sequence kernel association tests (41), can be used to test for association with a given disease or phenotype. However, aggregating variants according to presumed functional correlation presents another significant challenge because of the complex nonlinear nature of the genome (42). For example, functional variants within an intron of a gene do not necessarily act through that gene (43,44).

Second, there is a tremendous amount of variation within the human genome. Accordingly, analyzing whole exome and whole genome sequence data for functionally relevant genetic variation is computationally arduous. Efficient processing of sequencing data requires access to high-performance computing clusters and/or specially designed bioinformatics hardware (45). Many different software tools and pipelines exist for processing sequencing data and performing corresponding statistical tests, but advanced bioinformatics training is required to perform such analyses and ensure interoperability between software versions and file formats.

The power of aggregate-level tests to detect causal variants is highly dependent on *a priori* variant screening (46). Ideally, only functional alleles would be included in association testing, as inclusion of inconsequential variants introduces statistical noise. Variant effect prediction (VEP) tools exist that model either evolutionary conservation (47,48) or protein changes (49,50), but substantial differences in their design and output metrics complicate integration of their results (51). Further, thresholds based on VEP results are typically subjective (52). Functional annotation of the genome has expanded significantly in recent years (53–56) but not for all cell types and genomic regions (57). Despite the difficulties associated with studying rare variants, mounting evidence indicates that they do contribute to complex traits (58,59), and corresponding technologies and methods for studying rare variants continue to evolve at a rapid rate (60,61).

We investigated the potential role rare variants play in the pathogenesis of PCOS. We used two approaches: case-control candidate gene sequencing studies and family-based analyses. Anti-Müllerian hormone (AMH) is a highly plausible PCOS candidate gene. It is an important regulator of folliculogenesis (62), steroidogenesis (63) and neuroendocrine signaling (64). Circulating levels of AMH are elevated in women with PCOS (65,66). We performed targeted sequencing of *AMH* in a cohort of 643 PCOS cases and 153 controls and identified 18 rare (MAF≤1%) coding variants that were present in PCOS cases but not in controls (PCOS-specific variants) (67). We then measured the functional impact of these variants using AMH-mediated luciferase assays in transfected COS7 cells. Of these 18 PCOS-specific variants, 17 demonstrated a significant reduction in AMH-mediated signaling

capacity. Further, using a gene-based burden test approach (40), the functional *AMH* variants were significantly associated with PCOS using population-based controls. There were also rare coding variants that were present in both cases and controls. However, these variants had no impact on AMH signaling. This finding emphasizes the importance of functional assays to confirm the biologic relevance of genetic variants.

We subsequently investigated rare variation in the regulatory regions of *AMH* and its specific type 2 receptor, *AMHR2* (68). We identified 20 additional PCOS-specific variants in or near *AMH* and *AMHR2* that resulted in significantly reduced AMH signaling activity and were also significantly associated with PCOS. These variants included one missense variant in *AMHR2*, 16 noncoding/splicing variants in or upstream of *AMHR2*, and 3 noncoding variants upstream of *AMH*. AMH levels in cases with functional variants were significantly higher than control women (67). However, PCOS women with functional *AMH/AMHR2* variants had significantly lower AMH levels compared to other women with PCOS (68). Five of the *AMH* variants have also been found in men with persistent Müllerian duct syndrome, a rare disorder in which males retain internal Müllerian duct structures (69). AMH levels in these men are typically normal or undetectable (70). These findings suggest that AMH signaling capacity does not correlate positively with circulating AMH levels. Reduced AMH signaling could contribute to PCOS, however, by increasing androgen production due to loss of CYP17 inhibition by AMH. CYP17 is a key enzyme in androgen biosynthesis (71). The *AMH* variants with reduced AMH signaling showed a significant reduction in *CYP17a1* expression inhibition compared to wild-type *AMH* (68). Collectively, about 6.7% of PCOS-affected women from these cohorts had one or more of the *AMH/AMHR2* rare variants (67,68). Further, these studies illustrate the power of limiting sequencing to a few highly plausible candidate genes to avoid the statistical penalty of correcting for multiple testing in genome-wide analyses.

We performed whole-genome sequencing on DNA from 261 individuals from 62 families with one or more daughters with PCOS. We tested for gene-level associations of rare variants (MAF≤2%) that were predicted-to-be-deleterious (48,51) with PCOS and its concomitant hormonal traits using a quantitative trait meta-analysis (72). We found rare, primarily noncoding variants in *DENND1A* that were significantly associated with the following reproductive and metabolic traits: testosterone, dehydroepiandrosterone sulfate, sex hormone-binding globulin, insulin, luteinizing hormone (LH), and follicle-stimulating hormone (FSH). Common variants in *DENND1A* were associated with PCOS diagnosis in GWAS (2,3,9). Subsequent studies demonstrated that *DENND1A* is an important regulator of human ovarian androgen biosynthesis (73,74). However, none of the common GWAS variants were in known functionally important regions of the gene. Further, sequencing *DENND1A* in a limited number of PCOS cases did not identify pathogenic variants (73). Most of the rare variants we found were not in LD with the *DENND1A* GWAS variants (75). However, nine of the *DENND1A* variants were predicted to significantly disrupt transcription factor binding, while 23 were predicted to significantly affect RNA-binding protein motifs. Nevertheless, functional assays are needed to confirm these predictions.

Each of the rare variants in *DENND1A* was observed in only one or two families. Collectively, however, the rare variants were present in 50% of families. This finding is of

considerable interest in light of our original family studies that demonstrated a unimodal distribution for testosterone levels in control women but a bimodal distribution in sisters of women with PCOS (76) suggesting that elevated testosterone levels in PCOS were affected by specific autosomal alleles within families. Rare noncoding variants in *DENND1A* that altered gene regulation could result in increased testosterone production. Testosterone antagonizes the effects of estradiol and progesterone to slow hypothalamic GnRH contributing to enhanced LH release characteristic of PCOS (1,77). Consistent with this hypothesis, LH:FSH ratios were significantly higher in PCOS women with one or more of the *DENND1A* rare variants compared to PCOS women without any of the *DENND1A* variants (72). While no other gene associations in the family-based sequencing reached genome-wide significance, two of the top-5 associated genes were highly plausible PCOS candidate genes: *C9orf3*, which is a PCOS GWAS gene (2,3,9,78), and *BMP6*, which is a regulator of folliculogenesis in granulosa cells (79), and was more highly expressed in PCOS women compared with reproductively normal control women (80).

These studies support a model of PCOS in which pathogenic variants, while individually rare, tend to occur in genes regulating relevant disease pathways. The *DENND1A* findings support our studies implicating hyperandrogenemia as a core biologic pathway in PCOS (76). Hyperandrogenemia is a consistent reproductive phenotype in male as well as female PCOS relatives (76,81), including premenarchal daughters (82). Further, one of the earliest biomarkers in at-risk children, daughters of affected women, is evidence for global increases in 5α-reductase activity, which would enhance the conversion of T to its more potent metabolite, dihydrotestosterone (83,84).

The emerging consensus is that complex traits are primarily driven by noncoding variation (85,86), both common and rare (87,88). The idea that rare variants account for most of the heritability of complex traits was largely disproved by a series of studies that modeled cumulative contributions from all common SNPs within a population (not just those with significant trait associations), and found that common SNPs with small effect sizes collectively account for a much more significant portion of complex disease heritability than GWAS SNPs alone (59,89,90). Accordingly, as GWAS sample sizes have increased, the number of significant associations discovered per study has increased proportionally (91). However, the distribution of variant effects sizes and allele frequencies varies between different complex traits and diseases (92). Additionally, recent large-scale sequencing studies have demonstrated that remaining unexplained heritability in complex traits and diseases can be accounted for by rare variants that are not in LD with any common variation (93,94). Our studies indicate that both common and rare noncoding genetic variants contribute to PCOS pathogenesis, analogous to other complex traits/diseases. However, we also identified rare, pathogenic coding variants in *AMH* in about 3% of affected European ancestry cases (67). It is highly likely that as more candidate genes are sequenced, additional rare coding variants contributing to PCOS will be identified.

In conclusion, although the study of rare variants in PCOS remains extremely limited, results from our initial studies indicate that these variants contribute to PCOS. Candidate gene studies have indicated that rare coding and noncoding variants affecting the bioactivity of AMH or its receptor are present in a substantial minority of PCOS cases (67,68). Rare,

primarily noncoding, variants in *DENND1A* appear to play a role in the pathogenesis of familial PCOS in ~50% of families (72). These variants have implicated several plausible causal pathways, in particular androgen biosynthesis, consistent with our family studies of PCOS phenotypes (76,81,82). Functional studies are needed to determine the specific mechanisms by which these noncoding variants contribute to disease pathogenesis. More broadly studying the potential effects of rare variants represents a next step in the progression towards understanding of the genetic architecture of PCOS.

Acknowledgements

This research was supported by US National Institutes of Health grants P50 HD044405 (A.D.) and R01 HD085227 (A.D.). M.D. was supported by NRSA fellowship T32 DK007169.

References

1. Dunaif A. Perspectives in polycystic ovary syndrome: From hair to eternity. J Clin Endocrinol Metab. 2016;101(3):759–68. [PubMed: 26908109]
2. Chen ZJ, Zhao H, He L, Shi Y, Qin Y, Shi Y, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. Nat Genet. 2011 ;43(1):55–9. [PubMed: 21151128]
3. Shi Y, Zhao H, Shi Y, Cao Y, Yang D, Li Z, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. Nat Genet. 2012;44(9):1020–5. [PubMed: 22885925]
4. Goodarzi MO, Jones MR, Li X, Chua AK, Garcia OA, Chen YD, et al. Replication of association of *DENND1A* and *THADA* variants with polycystic ovary syndrome in European cohorts. J Med Genet. 2012;49(2):90–5. [PubMed: 22180642]
5. Lerchbaum E, Trummer O, Giuliani A, Gruber HJ, Pieber TR, Obermayer-Pietsch B. Susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21, and 9q33.3 in a cohort of Caucasian women. Horm Metab Res. 2011;43(11):743–7. [PubMed: 22009367]
6. Welt CK, Styrkarsdottir U, Ehrmann DA, Thorleifsson G, Arason G, Gudmundsson JA, et al. Variants in DENND1A are associated with polycystic ovary syndrome in women of European ancestry. J Clin Endocrinol Metab. 2012;97(7):E1342–7. [PubMed: 22547425]
7. Hayes MG, Urbanek M, Ehrmann DA, Armstrong LL, Lee JY, Sisk R, et al. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. Nature communications. 2015;6:7502.
8. Day FR, Hinds DA, Tung JY, Stolk L, Styrkarsdottir U, Saxena R, et al. Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. Nat Commun. 2015;6:8464. [PubMed: 26416764]
9. Day F, Karaderi T, Jones MR, Meun C, He C, Drong A, et al. Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. PLoS Genet. 2018;14(12):e1007813. [PubMed: 30566500]
10. Lander ES, Schork NJ. Genetic dissection of complex traits. Science. 1994;265(5181):2037–48. [PubMed: 8091226]
11. International HapMap Consortium. A haplotype map of the human genome Nature. 2005;437(7063): 1299–320. [PubMed: 16255080]
12. Teo YY, Ong RT, Sim X, Tai ES, Chia KS. Identifying candidate causal variants via trans-population fine-mapping. Genet Epidemiol. 2010;34(7):653–64. [PubMed: 20839287]
13. Qin H, Morris N, Kang SJ, Li M, Tayo B, Lyon H, et al. Interrogating local population structure for fine mapping in genome-wide association studies. Bioinformatics. 2010;26(23):2961–8. [PubMed: 20889494]
14. Spain SL, Barrett JC. Strategies for fine-mapping complex traits. Hum Mol Genet. 2015;24(R1):R111–9. [PubMed: 26157023]

15. van de Bunt M, Cortes A, Consortium I, Brown MA, Morris AP, McCarthy MI. Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci. PLoS Genet. 2015; 11 (9):e1005535. [PubMed: 26406328]
16. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009;27(2):182–9. [PubMed: 19182786]
17. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. Nat Methods. 2010;7(2): 111–8. [PubMed: 20111037]
18. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014;508(7497):469–76. [PubMed: 24759409]
19. Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. Nat Biotechnol. 2012;30(11): 1095–106. [PubMed: 23138309]
20. Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to Function. Am J Hum Genet. 2018;102(5):717–30. [PubMed: 29727686]
21. McAllister JM, Legro RS, Modi BP, Strauss JF 3rd. Functional genomics of PCOS: from GWAS to molecular mechanisms. Trends Endocrinol Metab. 2015;26(3): 118–24. [PubMed: 25600292]
22. Park JH, Gail MH, Weinberg CR, Carroll RJ, Chung CC, Wang Z, et al. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. Proc Natl Acad Sci U S A. 2011;108(44): 18026–31. [PubMed: 22003128]
23. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747–53. [PubMed: 19812666]
24. Vineis P, Pearce N. Missing heritability in genome-wide association study research. Nat Rev Genet. 2010;11(8):589.
25. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9(5):356–69. [PubMed: 18398418]
26. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet. 2010; 11(6):415–25. [PubMed: 20479773]
27. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science. 2004;305(5685):869–72. [PubMed: 15297675]
28. Bowden DW, An SS, Palmer ND, Brown WM, Norris JM, Haffner SM, et al. Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study. Hum Mol Genet. 2010;19(20):4112–20. [PubMed: 20688759]
29. Hegele RA, Pollex RL. Hypertriglyceridemia: phenomics and genomics. Mol Cell Biochem. 2009;326(1–2):35–43. [PubMed: 19130180]
30. Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat Genet. 2008;40(5):592–9. [PubMed: 18391953]
31. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56–65. [PubMed: 23128226]
32. Lek M, Karczewski K, Minikel E, Samocha K, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. bioRxiv. 2015.
33. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science. 2012;337(6090):64–9. [PubMed: 22604720]
34. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–9. [PubMed: 30305743]
35. Ward J, Strawbridge RJ, Bailey MES, Graham N, Ferguson A, Lyall DM, et al. Genome-wide analysis in UK Biobank identifies four loci associated with mood instability and genetic

correlation with major depressive disorder, anxiety disorder and schizophrenia. Transl Psychiatry. 2017;7(11): 1264. [PubMed: 29187730]

36. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: Biology, function, and translation. Am J Hum Genet. 2017;101(1):5–22. [PubMed: 28686856]
37. Jun G, Manning A, Almeida M, Zawistowski M, Wood AR, Teslovich TM, 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):379–84. [PubMed: 29279374]
38. Freund MK, Burch KS, Shi H, Mancuso N, Kichaev G, Garske KM, et al. Phenotype-Specific Enrichment of Mendelian Disorder Genes near GWAS Regions across 62 Complex Traits. Am J Hum Genet. 2018;103(4):535–52. [PubMed: 30290150]
39. Wright CF, West B, Tuke M, Jones SE, Patel K, Laver TW, et al. Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. Am J Hum Genet. 2019;104(2):275–86. [PubMed: 30665703]
40. 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. [PubMed: 17101154]
41. 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. [PubMed: 21737059]
42. Wu G, Zhi D. Pathway-based approaches for sequencing-based genome-wide association studies. Genet Epidemiol. 2013;37(5):478–94. [PubMed: 23650134]
43. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, et al. FTO obesity variant circuitry and adipocyte browning in humans. N Engl J Med. 2015;373(10):895–907. [PubMed: 26287746]
44. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature. 2014;507(7492):371–5. [PubMed: 24646999]
45. Miller NA, Farrow EG, Gibson M, Willig LK, Twist G, Yoo B, et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. Genome Med. 2015;7:100. [PubMed: 26419432]
46. Derkach A, Zhang H, Chatterjee N. Power Analysis for Genetic Association Test (PAGEANT) provides insights to challenges for rare variant association studies. Bioinformatics. 2018;34(9): 1506–13. [PubMed: 29194474]
47. Cooper GM, Stone EA, Asimenos G, Program NCS, Green ED, Batzoglou S, et al. Distribution and intensity of constraint in mammalian genomic sequence. Genome research. 2005;15(7):901–13. [PubMed: 15965027]
48. Huang YF, Gulko B, Siepel A. Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. Nat Genet. 2017;49(4):618–24. [PubMed: 28288115]
49. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248–9. [PubMed: 20354512]
50. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic acids research. 2003;31(13):3812–4. [PubMed: 12824425]
51. 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):310–5. [PubMed: 24487276]
52. Itan Y, Shang L, Boisson B, Ciancanelli MJ, Markle JG, Martinez-Barricarte R, et al. The mutation significance cutoff: gene-level thresholds for variant predictions. Nature methods. 2016;13(2): 109–10. [PubMed: 26820543]
53. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, et al. Genetic effects on gene expression across human tissues Nature. 2017;550(7675):204–13. [PubMed: 29022597]

54. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat Biotechnol. 2010;28(8):817–25. [PubMed: 20657582]
55. Pazin MJ. Using the ENCODE Resource for Functional Annotation of Genetic Variants. Cold Spring Harb Protoc. 2015;2015(6):522–36. [PubMed: 25762420]
56. Libbrecht MW, Rodriguez OL, Weng Z, Bilmes JA, Hoffman MM, Noble WS. A unified encyclopedia of human functional DNA elements through fully automated annotation of 164 human cell types. Genome Biol. 2019;20(1):180. [PubMed: 31462275]
57. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep. 2019;9(1):9354. [PubMed: 31249361]
58. Marouli E, Graff M, Medina-Gomez C, Lo KS, Wood AR, Kjaer TR, et al. Rare and low-frequency coding variants alter human adult height. Nature. 2017;542(7640): 186–90. [PubMed: 28146470]
59. Speed D, Cai N, Consortium U, Johnson MR, Nejentsev S, Balding DJ. Reevaluation of SNP heritability in complex human traits. Nature genetics. 2017;49(7):986–92. [PubMed: 28530675]
60. Nicolae DL. Association Tests for Rare Variants. Annu Rev Genomics Hum Genet. 2016; 17:117–30. [PubMed: 27147090]
61. Weissenkampen JD, Jiang Y, Eckert S, Jiang B, Li B, Liu DJ. Methods for the Analysis and Interpretation for Rare Variants Associated with Complex Traits. Curr Protoc Hum Genet. 2019;101(1):e83. [PubMed: 30849219]
62. Carlsson IB, Scott JE, Visser JA, Ritvos O, Themmen AP, Hovatta O. Anti-Mullerian hormone inhibits initiation of growth of human primordial ovarian follicles in vitro. Hum Reprod. 2006;21(9):2223–7. [PubMed: 16720622]
63. Teixeira J, Fynn-Thompson E, Payne AH, Donahoe PK. Mullerian-inhibiting substance regulates androgen synthesis at the transcriptional level. Endocrinology. 1999;140(10):4732–8. [PubMed: 10499532]
64. Cimino I, Casoni F, Liu X, Messina A, Parkash J, Jamin SP, et al. Novel role for anti-Mullerian hormone in the regulation of GnRH neuron excitability and hormone secretion. Nat Commun. 2016;7:10055. [PubMed: 26753790]
65. Dumont A, Robin G, Catteau-Jonard S, Dewailly D. Role of Anti-Mullerian Hormone in pathophysiology, diagnosis and treatment of Polycystic Ovary Syndrome: a review. Reprod Biol Endocrinol. 2015; 13:137. [PubMed: 26691645]
66. Pigny P, Merlen E, Robert Y, Cortet-Rudelli C, Decanter C, Jonard S, et al. Elevated serum level of anti-mullerian 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):5957–62. [PubMed: 14671196]
67. Gorsic LK, Kosova G, Werstein B, Sisk R, Legro RS, Hayes MG, et al. Pathogenic Anti-Mullerian Hormone Variants in Polycystic Ovary Syndrome. J Clin Endocrinol Metab. 2017;102(8):2862–72. [PubMed: 28505284]
68. Gorsic LK, Dapas M, Legro RS, Hayes MG, Urbanek M. Functional Genetic Variation in the Anti-Mullerian Hormone Pathway in Women with Polycystic Ovary Syndrome. J Clin Endocrinol Metab. 2019.
69. Josso N, Rey RA, Picard JY. Anti-mullerian hormone: a valuable addition to the toolbox of the pediatric endocrinologist. Int J Endocrinol. 2013;2013:674105. [PubMed: 24382961]
70. Josso N, Belville C, di Clemente N, Picard JY. AMH and AMH receptor defects in persistent Mullerian duct syndrome. Hum Reprod Update. 2005;11(4):351–6. [PubMed: 15878900]
71. Vasaitis TS, Bruno RD, Njar VC. CYP17 inhibitors for prostate cancer therapy. J Steroid Biochem Mol Biol. 2011;125(1–2):23–31. [PubMed: 21092758]
72. Dapas M, Sisk R, Legro RS, Urbanek M, Dunaif A, Hayes MG. Family-based quantitative trait meta-analysis implicates rare noncoding variants in *DENND1A* in polycystic ovary syndrome. J Clin Endocrinol Metab. 2019.
73. McAllister JM, Modi B, Miller BA, Biegler J, Bruggeman R, Legro RS, et al. Overexpression of a DENND1A isoform produces a polycystic ovary syndrome theca phenotype. Proc Natl Acad Sci U S A. 2014;111(15):E1519–27. [PubMed: 24706793]

74. Tee MK, Speek M, Legeza B, Modi B, Teves ME, McAllister JM, et al. Alternative splicing of DENND1A, a PCOS candidate gene, generates variant 2. Mol Cell Endocrinol. 2016;434:25–35. [PubMed: 27297658]

75. 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):456–76. [PubMed: 31402091]

76. 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): 14956–60. [PubMed: 9843997]

77. Burt Solorzano CM, Beller JP, Abshire MY, Collins JS, McCartney CR, Marshall JC. Neuroendocrine dysfunction in polycystic ovary syndrome. Steroids. 2012;77(4):332–7. [PubMed: 22172593]

78. Hayes MG, Urbanek M, Ehrmann DA, Armstrong LL, Lee JY, Sisk R, et al. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. Nat Commun. 2015;6:7502. [PubMed: 26284813]

79. Shi J, Yoshino O, Osuga Y, Koga K, Hirota Y, Hirata T, et al. Bone morphogenetic protein-6 stimulates gene expression of follicle-stimulating hormone receptor, inhibin/activin beta subunits, and anti-Mullerian hormone in human granulosa cells. Fertil Steril. 2009;92(5): 1794–8. [PubMed: 19539911]

80. Khalaf M, Morera J, Bourret A, Reznik Y, Denoual C, Herlicoviez M, 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):437–44. [PubMed: 23243014]

81. 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):2134–8. [PubMed: 11994353]

82. Torchen LC, Fogel NR, Brickman WJ, Paparodis R, Dunaif A. Persistent apparent pancreatic beta-cell defects in premenarchal PCOS relatives. J Clin Endocrinol Metab. 2014;99(10):3855–62. [PubMed: 25029420]

83. Wilson JD, Griffin JE, Russell DW. Steroid 5 alpha-reductase 2 deficiency. Endocr Rev. 1993; 14(5):577–93. [PubMed: 8262007]

84. Torchen LC, Idkowiak J, Fogel NR, O'Neil DM, Shackleton CH, Arlt W, et al. Evidence for increased 5alpha-reductase activity during early childhood in daughters of women with polycystic ovary syndrome. J Clin Endocrinol Metab. 2016;101(5):2069–75. [PubMed: 26990942]

85. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: From polygenic to omnigenic. Cell. 2017;169(7):1177–86. [PubMed: 28622505]

86. Ma M, Ru Y, Chuang LS, Hsu NY, Shi LS, Hakenberg J, et al. Disease-associated variants in different categories of disease located in distinct regulatory elements. BMC Genomics. 2015;16 Suppl 8:S3.

87. Igartua C, Mozaffari SV, Nicolae DL, Ober C. Rare non-coding variants are associated with plasma lipid traits in a founder population. Sci Rep. 2017;7(1): 16415. [PubMed: 29180722]

88. Zhao J, Akinsanmi I, Arafat D, Cradick TJ, Lee CM, Banskota S, et al. A burden of rare variants associated with extremes of gene expression in human peripheral blood. Am J Hum Genet. 2016;98(2):299–309. [PubMed: 26849112]

89. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nature genetics. 2010;42(7):565–9. [PubMed: 20562875]

90. Shi H, Kichaev G, Pasaniuc B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. American journal of human genetics. 2016;99(1): 139–53. [PubMed: 27346688]

91. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012;90(1):7–24. [PubMed: 22243964]

92. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB. Genetic architecture: the shape of the genetic contribution to human traits and disease. Nat Rev Genet. 2018; 19(2): 110–24. [PubMed: 29225335]

93. Wainschtein P, Jain DP, Yengo L, Zheng Z, Cupples LA, Shadyab AH, et al. Recovery of trait heritability from whole genome sequence data. bioRxiv. 2019:588020.

94. 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): 1349–55. [PubMed: 31477931]

Highlights

- Studying rare variants in complex disease presents unique challenges
- Rare variants affecting *AMH* or *AMHR2* are present in many PCOS cases
- Rare variants in *DENND1A* associated with altered hormone levels in PCOS families

Table 1.

PCOS GWAS Susceptibility Loci

Chromosome	Han Chinese[¶] Chen, 2011 (2) Shi, 2012 (3)	European 1[*], Hayes, 2015 (7)	European 2[*¶§] Day, 2015 (8)	European Meta[*¶§] Day, 2018 (9)
2p16.3	*LHCGR*			
2p16.3	*FSHR*			
2p21	*THADA*		*THADA*	*THADA*
2q34			*ERBB4*	*ERBB4*
5q13.1			*RAD50*	*RAD50*
8p32.1		*GATA4/NEIL2*		***GATA4/NEIL2***[†]
9p24.1				*PLGRKT*
9q22.32	*C9orf3*	*C9orf3*		*C9orf3*
9q33.3	*DENND1A*			*DENND1A*
11p14.1		*FSHB*	*FSHB*	*FSHB*
11q22.1	*YAP1*		*YAP1*	*YAP1*
11q23.1				*ZBTB16*
12q13.2	*RAB5B/SUOX*			*RAB5B/ERBB3*
12q14.3	*HMGA2*			
12q21.2			*KRR1*	*KRR1*
16q12.1	*TOX3*			*TOX3*
19q13.3	*INSR*			
20q11.21	*SUMO1P1*			
20q13.2				*MAPRE1*

[*] PCOS diagnosis was based on NIH criteria,

[¶] Rotterdam criteria, or

[§] self-report.

[†] PCOS association dependent on diagnostic criteria in bold; association with remaining loci was similar for PCOS diagnosed by self-report and PCOS diagnosed by NIH or non-NIH Rotterdam criteria