The genetics of the polycystic ovary syndrome

Margrit Urbanek

SUMMARY

Polycystic ovary syndrome (PCOS) is a very common endocrine disorder that has a strong genetic component and is characterized by polycystic ovaries, hyperandrogenemia, and menstrual irregularity. During the past decade, the roles of more than 70 candidate genes have been evaluated for a causal role in PCOS; however, because of genetic and phenotypic heterogeneity and underpowered studies, the results of many of these studies remain inconclusive. Here, the results of the genetic analysis of several candidate genes and gene regions—*CYP11A* (encoding cytochrome P450, family 11, subfamily A polypeptides), *CAPN10* (encoding calpain 10), the insulin gene VNTR (variable number of tandem repeats), and D19S884 (a dinucleotide repeat marker mapping to chromosome 19p13.2)—are discussed in detail. Although past genetic studies of PCOS have yielded only modest results, resources and techniques have been assembled to remedy the major deficits of these early studies, promising that the next few years will be a very exciting and rewarding era for the genetic analysis of PCOS.

KEYWORDS ***CAPN10*, *CYP11A*, D19S884, genetics of complex diseases, polycystic ovary syndrome**

REVIEW CRITERIA

I searched for original articles focusing on PCOS in MEDLINE published between 2002 and present. The search term I used was "PCOS". I then selected all papers that characterized genetic variation in PCOS. Data predating 2002 were obtained from a similar literature search that was carried out in the summer of 2002. The vast majority of papers were English-language full-text papers. Manuscripts examining related phenotypes such as hyperandrogenism were also examined, although in a much less rigorous manner.

M Urbanek is an Assistant Professor in the Division of Endocrinology, Metabolism, and Molecular Medicine and the Center for Genetic Medicine, Northwestern University Medical School, Chicago, IL, USA.

Correspondence
Division of Endocrinology, Metabolism, and Molecular Medicine, Northwestern University School of Medicine, 303 East Chicago Avenue, Tarry 15-717, Chicago, IL 60611, USA
m-urbanek@northwestern.edu

Received 14 April 2006 **Accepted** 20 October 2006

www.nature.com/clinicalpractice
doi:10.1038/ncpendmet0400

INTRODUCTION

Polycystic ovary syndrome (PCOS) is a genetically complex disorder that is characterized by hyperandrogenemia with amenhorrea or oligomenhorrea. It is the most common endocrinopathy among women of reproductive age.[1,2] In addition to its reproductive features, PCOS is associated with an increased risk of obesity, insulin resistance, and type 2 diabetes mellitus.[3] Although PCOS can be diagnosed only in women of reproductive age, first-degree relatives of women with PCOS also have an increased incidence of obesity, insulin resistance, and elevated triglyceride levels.[3–7] As a consequence, PCOS and its associated pathologies contribute significantly to the health burden of Western societies. The heterogeneous disorder has been defined according to various criteria, ranging from purely morphological (polycystic ovaries) to predominantly biochemical (hyperandrogenemia). This variability has contributed to the complexity of the genetic studies of PCOS.

Here, I give a brief overview of genetic studies of PCOS done so far with discussion of some of the difficulties that have plagued the field, followed by a more detailed discussion of the four candidate genes and gene regions—*CYP11A* (encoding cytochrome P450, family 11, subfamily A polypeptides), *CAPN10* (encoding calpain 10), the insulin gene VNTR (variable number of tandem repeats), and D19S884 (a dinucleotide repeat marker mapping to chromosome 19p13.2)—for which the most conclusive data are available. Finally, I conclude with a discussion of the future of genetic analysis of PCOS.

FAMILIAL BASIS OF PCOS

Environmental and genetic factors contribute to the etiology of PCOS. For example, Kiddy *et al.*[8] have shown that weight loss can reverse the symptoms of PCOS, and Norman and colleagues[9–12] have shown that lifestyle modifications such as moderate weight loss and exercise significantly improve the hormonal profiles of both the reproductive and metabolic features of PCOS.

Box 1 Selection of candidate genes.[77]

Positional cloning can be used to identify candidate regions within the genome even in the absence of functional candidate genes. Genes mapping within these regions are promising candidate genes for a disorder. Mapping techniques include the following:

- Cytogenetic abnormalities like translocations, insertion, and deletions can be used to localize candidate gene regions.
- Linkage analysis is a family-based genetic test that allows the screening of an entire genome for the candidate gene, using a relatively small number of markers (~400 short tandem repeat polymorphisms). This is a very powerful method for Mendelian traits (i.e. cystic fibrosis) but has not been very successful for complex genetic traits.
- Whole-genome association studies use association mapping, which is more powerful for mapping complex genetic disorders than are linkage studies. It requires the study of many markers (300,000–500,000 single-nucleotide polymorphisms), which is expensive and requires sophisticated analytical techniques. These studies are only now becoming relatively routine.
- Whole-genome sequencing is a technology that is currently being developed.[78] Whole-genome sequencing will provide a complete and direct assessment of the genetic variation of an individual. It is currently not known when this technology will become available and how expensive it will be.

Functional information can be used to identify disease candidate genes. Potentially functional genes include:

- A gene that encodes a known protein that has an altered function in the disease state.
- A gene that encodes a protein that might be altered according to the phenotype of the disorder.

Alternatively, a large familial component to the etiology of PCOS has long been recognized;[13–16] both dominant and multigenic modes of inheritance have been proposed.[13–22] Vink *et al.*[22] used a twin-study paradigm to demonstrate that there is a strong heritable component to the etiology of PCOS. These authors characterized 1,332 monozygotic (share ~100% of genetic material) and 1,873 dizygotic or singleton sister pairs (share ~50% of genetic material) from the Netherlands Twin Registry. The overall heritability for PCOS was 0.72 with a correlation of 0.71 for monozygotic sister pairs and 0.38 for dizygotic twin or nontwin sister pairs.

The strong correlation between dizygotic twins and between nontwin sister pairs indicates that the genetic component of PCOS is probably due to a few genes with a significant impact rather than to many genes with minimal effects. If this is the case, it lends support to the feasibility of searching for PCOS susceptibility genes.

CURRENT STATUS OF PCOS GENETIC STUDIES

During the past decade, more than 60 studies have investigated the genetic contribution of more than 70 genes to the etiology of PCOS (see details from these studies in Supplementary Table 1 online). All these studies have been candidate gene analyses, an approach that is very powerful when the biology of a disorder is sufficiently understood to provide legitimate functional candidates (Box 1). One obvious drawback of the candidate gene approach is that genes that are not *a priori* reasoned to play a role in the disorder will not be evaluated. Several metabolic and/or biochemical pathways are implicated in the etiology of PCOS, including those that regulate gonadotropin secretion, affect androgen production and action, and influence insulin signaling (Box 2).[23,24] As the connection between PCOS and the metabolic syndrome becomes more clearly established, genes implicated in the development of diabetes and obesity also become ideal candidates.

To date, the roles of more than 70 genes in the etiology of PCOS have been evaluated (see Supplementary Table 1 for a summary of relevant PCOS genetic studies). With a few notable exceptions (discussed below), the findings of these studies have, however, been ambiguous or lacked replication. This lack of reproducibility is endemic to association studies of common traits[25–28] and can be attributed to multiple factors that plague the genetic analysis of both complex diseases in general and PCOS in specific. These factors include an absence of consistent phenotype criteria across studies, limited sample sizes, and incomplete characterization of candidate genes.

Phenotyping

PCOS is a complex syndrome, and multiple features (hyperandrogenism, menstrual irregularity, ovarian morphology, and gonadotropin dysregulation) contribute to the overall phenotype. Different aspects of the disease

might be under the control of unique genes; therefore, the characteristics that are emphasized in the assignment of the phenotype may affect the outcome of the genetic analysis.

Although, for clinical purposes, it may be useful to apply as broad a phenotype as possible, for genetic analyses, it is sometimes more powerful to restrict the phenotype with the expectation that this phenotypically more homogeneous subset of patients is also genetically more homogeneous. For instance, we limit our analyses to patients with biochemical hyperandrogenemia and fewer than six menses per year. We realize that these criteria do not include all possible patients who have PCOS but instead identify a more severely affected subset of family members who may have the same underlying defect. This approach has allowed us to identify a PCOS susceptibility locus (D19S884; see below) mapping to chromosome 19p13.2.

In 1990 and in 2003, guidelines were established to systematically define PCOS (Box 3);[29,30] however, these guidelines still allow for multiple designations of the phenotype. Although a given set of criteria for the phenotype of PCOS is not necessarily better or more appropriate than another set of criteria, it is critical that investigators and readers of the literature are aware of the potential differences in phenotypic characterizations and the impact of such variability on the outcome of studies.

Box 2 Functional candidate genes in polycystic ovary syndrome.

Genes affecting gonadotropin secretion
ACTR1 (activin receptor 1), *ACTR2A* (activin receptor 2A), *ACTR2B* (activin receptor 2B), *FSHR* (follicle-stimulating hormone receptor), *FST* (follistatin), *INHA* (inhibin subunit A), *INHBA* (inhibin subunit βA), *INHBB* (inhibin subunit βB), *INHC* (inhibin subunit C), *LHB* (luteinizing hormone β subunit), *LHCGR* (luteinizing hormone/choriogonadotropin receptor), *MIS* (Müllerian inhibiting substance), *SHBG* (sex hormone-binding globulin), *SMAD4* (mothers against DPP homolog 4)

Genes affecting androgen production and secretion
AR (androgen receptor), *BMP15* (bone morphogenic protein), *GDF9* (growth differentiation factor 9), *CYP11A* (cytochrome P450, family 11, subfamily A polypeptides), *CYP17A1* (cytochrome P450, family 17, subfamily A, polypeptide 1), *CYP19A1* (cytochrome P450, family 19, subfamily A, polypeptide 1), *HSD3B1* (hydroxyl-δ-5-steroid dehydrogenase, 3β- and steroid δ-isomerase 1), *HSD3B2* (hydroxyl-δ-5-steroid dehydrogenase, 3β- and steroid δ-isomerase 2), *HSD17B1* (hydroxysteroid (17-β)-dehydrogenase 1), *HSD17B2* (hydroxysteroid (17β)-dehydrogenase 2), *HSD17B3* (hydroxysteroid (17β)-dehydrogenase 3), *STAR* (steroidogenic acute regulator)

Genes influencing insulin secretion and obesity
ADIPOQ (adiponectin), *CAPN10* (calpain 10), insulin gene VNTR (variable number of tandem repeats), *INSR* (insulin receptor), *IRS1* (insulin receptor substrate 1), *IRS2* (insulin receptor substrate 2), *IGF1* (insulin-like growth factor 1), *IGF2* (insulin-like growth factor 2), *IGF1R* (insulin-like growth factor 1 receptor), *IGF2R* (insulin-like growth factor 2 receptor), *IGFBP1* (insulin-like growth factor binding protein 1) plus *IGFBP3* (insulin-like growth factor binding protein 3), *PPARG* (peroxisome proliferator-activated receptor γ), *INSL3* (Leydig insulin-like protein), *LEPR* (leptin receptor), *OB* (leptin), *MC4R* (melanocortin-4 receptor), *TNF* (tumor necrosis factor), *UCP2* (uncoupling protein 2), *UCP3* (uncoupling protein 3), *POMC* (propiomelanocortin)

Sample sizes

Most PCOS studies reported so far have been based on very small sample sizes (Supplementary Table 1). Based on the results of successful studies of other genetically complex diseases, it is now clear that most susceptibility alleles have relatively small effect sizes (i.e. the magnitude of the impact that a single genetic variant has on a given phenotype [reviewed by Hirschhorn[26] and by Hattersley and McCarthy[28]]) and as such require sample sizes of several hundred, if not thousand, patients. For instance, in the case of the peroxisome proliferator-activated receptor γ Pro12Ala polymorphism—one of the few unambiguous susceptibility alleles for type 2 diabetes identified to date—the more common proline residue is associated with a ~1.25-fold increased risk of developing type 2 diabetes and a population attributable risk for diabetes of 25%.[31] However, because of the small relative risk associated with this variant, it required a meta-analysis of around 3,000 individuals to definitively establish its role in the pathology of type 2 diabetes.

Similarly, Grant *et al.*[32] were able to identify transcription factor 7-like 2 (*TCF7L2*), a very promising susceptibility gene for type 2 diabetes associated with a relative risk of 1.5, in an association study of 1,185 patients and 931 controls. Given the large sample sizes required to detect these loci, the majority of PCOS genetic

Box 3 Diagnostic criteria for polycystic ovary syndrome.

1990 National Institutes of Health criteria—the two criteria below need to be fulfilled for the diagnosis of PCOS[29]

- Clinical (acne or hirsuitism) and/or biochemical hyperandrogenemia (measured elevated androgen levels)
- Menstrual irregularity

2003 Rotterdam criteria—two of the three criteria below need to be fulfilled for the diagnosis of PCOS[30]

- Clinical (acne or hirsuitism) and/or biochemical hyperandrogenemia (measured elevated androgen levels)
- Menstrual irregularity
- Polycystic ovarian morphology on ultrasonography

studies have been vastly underpowered. To date, studies from only a few groups have used sample sizes with more than 300 probands (see Supplementary Table 1 online).

One reason for the small sample sizes used in the genetic studies of PCOS is that although PCOS is a very common disorder, it has been difficult to collect sufficiently large cohorts of patients because of several inherent characteristics of the disorder. One major limitation is that PCOS can be accurately assessed only in women of reproductive age who are not taking several common medications (oral contraceptives, insulin-sensitizing medication, etc.). This automatically excludes men, severely limits the pool from which women with PCOS can be recruited, and is even more detrimental to the recruitment of the affected family members critical for linkage studies. Although it is clear that male relatives of women with PCOS are at an increased risk of developing features of the metabolic syndrome,[6,7] it is currently impossible to unambiguously assign a phenotype to them. Finally, because PCOS is a disorder with reduced fertility, large families are uncommon and families with multiple eligible sisters are rare.

Variant choice

For Mendelian traits, a single mutation within one gene often accounts for most of, if not all, the genetic variation contributing to the phenotype of the disease. By definition, genetically complex diseases like PCOS have multiple variants that contribute to the etiology of a disease as a result of genetic heterogeneity (variants in multiple genes) and/or allelic heterogeneity (multiple variants within one gene). In addition, susceptibility alleles for complex traits are not constrained to missense or nonsense mutations. Because variants contributing to complex traits are expected to be common and therefore cannot be under strong selective pressure,[26] they are expected to have more subtle consequences that are likely to affect transcript expression levels or processing. Such variants are much more difficult to recognize.

It is therefore critical to characterize the genetic variation of an entire candidate gene, including the regulatory regions. For most genetic studies of PCOS, including most of the studies by our group, only one or a few variants per gene have been tested and very few studies have characterized the entire gene directly using single-nucleotide polymorphisms or indirectly using haplotype analysis.

PCOS CANDIDATE GENES

Given the limitations of current PCOS genetic studies, there are several loci for which there is convincing evidence regarding their role in PCOS and there are a few additional genes for which a consensus may be reached in the near future. These loci include *CYP11A* (five studies), the VNTR of the insulin gene (six studies), *CAPN10* (three studies), and D19S884 (three studies).

CYP11A

CYP11A is an ideal functional candidate gene for PCOS because it encodes the enzyme that cleaves the side-chain of p450, the rate-limiting step in androgen biosynthesis. In 1997, Gharani *et al.*[33] showed evidence for linkage between *CYP11A* and PCOS. Subsequently, Urbanek *et al.*[23] (and my own unpublished data) found modest evidence for linkage between PCOS and *CYP11A*. Gharani *et al.*[33] also found an association with allele 5 of D15S520, which is located in the promoter region of *CYP11A*, although Urbanek *et al.*[23] could not provide replication for association at D15S520. Two additional studies, by Diamanti-Kandarakis *et al.*[34] and Daneshmand *et al.*[35], did show evidence for association between *CYP11A* promoter alleles and PCOS; however, a study by San Millan *et al.*[36] in a Spanish cohort failed to show

evidence for association between the D15S520 pentanucelotide repeat and PCOS.

In 2004, Gaasenbeek *et al.*[37] assessed the role of *CYP11A* in the etiology of PCOS using a large-scale family-based and population-based association study including 371 PCOS patients and 331 population control subjects from the UK and 527 symptomatic women and 1,062 cohort control subjects from the Northern Finland Birth Cohort of 1966. The UK cohort included 141 parent–offspring trios. Although the authors did detect a statistical difference in allele frequencies for both markers in the patients and control subjects from the UK, the association was with a different allele (allele 4) than the one previously reported (allele 5)[33,34]; the association with allele 5 was not replicated in the trios or the Finnish cohort. There were no significant associations between the haplotypes of the two markers and PCOS in these samples.

Taken as a whole, the *CYP11A* genetic association studies indicate that there is no evidence for association between genetic variants in the promoter of the gene and PCOS in the Caucasian population and there is inconclusive evidence for linkage. In addition, variants in *CYP11A* that are not in linkage disequilibrium with the promoter variants could play a role in the etiology of PCOS.

The insulin gene VNTR

The VNTR is a series of 14 bp or 15 bp repeats that are located in the 5' regulatory element of the insulin gene. The insulin gene VNTR alleles are grouped into three classes based on repeat size: class I with an average length of 26–63 repeats, class II with an average of 64–140 repeats, and class III with an average of 141–209 repeats.[38] The VNTR regulates transcription of the insulin gene and is associated with hyperinsulinemia, susceptibility to type 2 diabetes, birth weight, fasting insulin levels, and the development of childhood and juvenile obesity (reviewed by Pugliese and Miceli[39]). In 1997, Waterworth *et al.*[40] showed evidence for modest linkage between PCOS and the insulin gene VNTR in 17 families and preferential transmission of the class III alleles of the insulin gene VNTR from fathers of PCOS patients. This positive association between paternally inherited class III alleles was confirmed by Michelmore *et al.*[41]

By contrast, Urbanek *et al.*[23] found no evidence for linkage between PCOS and the insulin gene VNTR in 28 multiplex families; nor was there evidence of association between the insulin gene VNTR class III allele in general or the paternally inherited allele in a much larger sample (150 nuclear families) or in an independent follow-up study of similar size by the same authors (M Urbanek *et al.*, unpublished data). In a case–control study, Calvo *et al.*[42] failed to detect an association between the insulin gene VNTR and hyperandrogenemia in Spanish women.

Finally, in a large multicohort study that included 255 parent–proband trios, 185 PCOS patients and 1,062 controls from the UK, and 1,599 women from the Northern Finland Birth Cohort of 1966, Powell *et al.*[43] failed to find evidence for association between the insulin gene VNTR class III alleles and PCOS or testosterone levels. It is therefore unlikely that this locus plays an important role in the etiology of PCOS.

CAPN10

The locus containing *CAPN10*, which encodes the cysteine protease calpain 10, was first identified as a susceptibility locus for type 2 diabetes through positional cloning.[44,45] Association and linkage analysis identified two single-nucleotide polymorphism haplotypes (designated 112 and 121) defined by closely linked noncoding single-nucleotide polymorphisms that conferred an increased risk for diabetes to individuals who were doubly heterozygous for the haplotypes (i.e. had the genotype 112/121).[45] These findings were replicated in populations of northern European descent[45] but not in a study of British patients with diabetes[46] or in Samoans with type 2 diabetes.[47] Functional studies have shown that the G allele at UCSNP-43 is associated with reduced muscle levels of calpain 10 messenger RNA and insulin resistance in nondiabetic Pima Indians. Calpain 10 has also been found to play a role in insulin secretion and action. Because type 2 diabetes and insulin resistance are often associated with PCOS, *CAPN10* is also a plausible candidate gene for PCOS.

Three studies (Ehrmann *et al.*,[48] Haddad *et al.*,[49] and M Urbanek *et al.*, unpublished data) have examined the contribution of calpain 10 to the etiology of PCOS. Ehrmann *et al.*[48] tested 124 PCOS patients of European descent and 57 African American patients who had PCOS for association between the type 2 diabetes-associated DNA polymorphism and a series of phenotypic characteristics of PCOS and type 2

diabetes. They found no evidence for association between either the individual single-nucleotide polymorphisms or the haplotype combinations and any of the phenotypic characteristics in PCOS patients of European descent; however, the 112/121 haplogenotype was found to be significantly associated with higher insulin levels in response to a glucose challenge in the African American PCOS population.

Haddad *et al.*[49] also found no evidence for association between the 112/121 genotype and PCOS or any intermediate traits in a sample of 331 PCOS patients. Our group has tested for association between individual single-nucleotide polymorphisms (UCSNPs 43, 44, 19, and 63) and haplotypes within *CAPN10* and PCOS in 390 PCOS trios of predominantly European ancestry (unpublished data). We did not detect evidence for association between any *CAPN10* single-nucleotide polymorphisms or haplotypes and PCOS. Taken together, these studies indicate that calpain 10 is unlikely to play a major role in the etiology of PCOS in Europeans.

Chromosome 19p13.2 PCOS susceptibility locus (D19S884)

In a screen of 37 PCOS candidate genes in 150 families, our group found that the strongest evidence for association occurred on chromosome 19p13.2.[23]

Further characterization of this region with 18 additional markers and 217 additional families replicated the original findings and found that the strongest evidence for association is still with allele 8 of D19S884.[50] D19S884 is a dinucleotide-repeat polymorphism that maps 800 kilobases centromeric to *INSR* (which encodes the insulin receptor gene) and was originally selected to assess linkage between the PCOS candidate gene (*INSR*) and PCOS. Because 800 kb exceeds the usual distance over which linkage disequilibrium (the biological basis for allelic association) is maintained, it is unlikely that this variant is directly associated with genetic variation within the candidate gene *INSR* itself. In the complete cohort of 367 families, the region of chromosome 19p13.2 containing D19S884 also has the strongest evidence for linkage to PCOS of any of the 33 candidate gene regions in the families we tested.[50] Two relatively small case–control studies have tested for association between D19S884 and PCOS; one replicated our results[51] and one did not.[52]

D19S884 maps 105 bp 3' to exon 55 of the fibrillin 3 gene (*FBN3*), the third member of the fibrillin extracellular matrix protein family. The functional role of D19S884 remains to be determined. Variations in dinucleotide-repeat polymorphisms like D19S884 have, however, been shown to play a role in both transcriptional[53–55] and splicing enhancer[56,57] activity. Whether D19S884 acts as a distal enhancer element for the *INSR* (the candidate gene that directed our attention to this area of the genome) or affects the expression and/or splicing pattern of fibrillin 3 represents areas of active research in our laboratory and others.

EPIGENETIC EFFECTS

In addition to the genetic studies described, there is increasing interest in the prenatal origins of PCOS that are predicted to be due to genetic variation in the mother and/or the fetus or to environmental factors[58–65] and in epigenetic contributions to the etiology of PCOS.[66]

One very interesting area of research in PCOS is the investigation of the role of genetic variation in the androgen receptor gene (*AR*) and the role of X-inactivation in PCOS. Because PCOS is a disorder of androgen excess, genes involved in androgen signaling are also believed to be important in the etiology of PCOS and related disorders. *In vitro* studies demonstrating an inverse relationship between the length of the CAG repeat encoding the polyglutamine tract in the N-terminal transactivation domain of the androgen receptor (AR) and receptor activity[67,68] suggest that the length of the CAG repeat could be expected to affect the degree of androgen sensitivity. Genetic studies of the role of variation in the CAG repeat have, however, yielded ambiguous results.[69–74]

Because the *AR* receptor gene is X-linked and one copy of the X chromosome is inactivated in women, the pattern of X-inactivation could influence *AR* activity and PCOS. In an analysis of X-inactivation of 88 sisters of women with PCOS, Hickey *et al.*[66] found that sisters with the same *AR* CAG repeat genotype and the same clinical presentation (unaffected–unaffected or PCOS–PCOS pairs) more frequently showed the same pattern of X-inactivation than did sisters with different clinical presentations (85% versus 16%) but the same genotype. In other words, sister pairs in whom the same copy of the *AR* is transcriptionally active also have the same phenotype, whereas sister pairs with

the same genotype but different transcriptionally active X-chromosomes do not have the same phenotype.

These findings support the hypothesis that the *AR* CAG repeat number plays an important role in the etiology of PCOS. This study also demonstrates that epigenetic factors like X-inactivation can have a significant effect on the expression of complex genetic diseases such as PCOS and may complicate the interpretation of genetic results. Because of the relatively small sample of sisters analyzed in this study, it is critical that this study be replicated in an independent cohort.

CONCLUSIONS

Although the above-described summary of the current state of genetic studies of PCOS might seem rather discouraging, there are many reasons for optimism. We are currently at the brink of a very exciting and potentially extremely rewarding era for the identification of genetic determinants of PCOS, because several critical factors are now converging. Within the past few years, new reagents and tools have been assembled to make successful analysis of genetically complex disorders eminently feasible: these tools include a nearly complete catalogue of common human genetic variation by the International HapMap Project,[75,76] efficient and relatively inexpensive high-volume genotyping technologies, development of easily accessible analysis software, and, most important, the assembly of sufficiently large PCOS patient cohorts[37,43,50] to detect genetic variants with effect sizes as observed in other complex diseases.

When applied to candidate genes, these tools make it possible to fully explore the genetic relevance of these genes to the etiology of PCOS and may help to reconcile some of the discrepant results observed in studies of different variants within the same gene. Finally, it is now possible to carry out genome-wide association studies of PCOS that will identify potentially novel and unexpected genes and variants contributing to the etiology of PCOS. The next 10 years promise to be a very exciting and productive era in the genetic analysis of PCOS.

Supplementary information in the form of a table is available on the *Nature Clinical Practice Endocrinology & Metabolism* website. This provides a compilation of studies examining candidate genes for PCOS that have been evaluated at the DNA level.

KEY POINTS

- Polycystic ovary syndrome (PCOS) is a very common endocrine disorder that has a large impact on the health burden of Western societies and that is believed to have a strong genetic basis
- Dissection of the genetic basis of PCOS is currently an area of intensive investigation, with more than 70 genes that have been evaluated for their impact on the etiology of PCOS; however, to date, most studies have been underpowered and have yielded inconclusive, if not contradictory, results
- The appropriate tools, including appropriately sized PCOS cohorts and genetic reagents, have now been assembled to carry out more suitable studies, promising that the next 10 years will be a very exciting and productive era in analysis of the genetic basis of PCOS

References

1 Diamanti-Kandarakis E *et al*. (1999) A survey of the polycystic ovary syndrome in the Greek island of Lesbos: hormonal and metabolic profile. *J Clin Endocrinol Metab* **84:** 4006–4011
2 Knochenhauer ES *et al*. (1998) Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States: a prospective study. *J Clin Endocrinol Metab* **83:** 3078–3082
3 Sam S and Dunaif A (2003) Polycystic ovary syndrome: syndrome XX? *Trends Endocrinol Metab* **14:** 365–370
4 Yildiz BO *et al*. (2003) Glucose intolerance, insulin resistance, and hyperandrogenemia in first degree relatives of women with polycystic ovary syndrome. *J Clin Endocrinol Metab* **88:** 2031–2036
5 Sam S *et al*. (2005) Dyslipidemia and metabolic syndrome in the sisters of women with polycystic ovary syndrome. *J Clin Endocrinol Metab* **90:** 4797–4802
6 Sir-Petermann T *et al.* (2002) Prevalence of type II diabetes mellitus and insulin resistance in parents of women with polycystic ovary syndrome. *Diabetologia* **45:** 959–964
7 Yilmaz M *et al*. (2005) Glucose intolerance, insulin resistance and cardiovascular risk factors in first degree relatives of women with polycystic ovary syndrome. *Hum Reprod (Oxf)* **20:** 2414–2420
8 Kiddy DS *et al*. (1992) Improvement in endocrine and ovarian function during dietary treatment of obese women with polycystic ovary syndrome. *Clin Endocrinol* **36:** 105–111
9 Norman RJ *et al*. (2004) Improving reproductive performance in overweight/obese women with effective weight management. *Hum Reprod Update* **10:** 267–280
10 Moran LJ *et al*. (2004) Short term energy restriction (using meal replacements) improves reproductive parameters in polycystic ovary syndrome. *Asia Pacific J Clin Nutr* **13 (Suppl):** S88
11 Moran L and Norman RJ (2004) Understanding and managing disturbances in insulin metabolism and body weight in women with polycystic ovary syndrome. *Best Pract Res Clin Obstet Gynaecol* **18:** 719–736

12 Norman RJ *et al*. (2002) The role of lifestyle modification in polycystic ovary syndrome. *Trends Endocrinol Metab* **13:** 251–257
13 Cooper HE *et al*. (1968) Hereditary factors in Stein-Leventhal syndrome. *Am J Obstet Gynecol* **100:** 371–387
14 Givens JR (1988) Familial polycystic ovarian disease. *Endocrinol Metab Clin N Am* **17:** 771–783
15 Hague W *et al*. (1988) Familial polycystic ovaries: a genetic disease. *Clin Endocrinol* **29:** 593–605
16 Ferriman D and Purdie AW (1979) The inheritance of polycystic ovarian disease and a possible relationship to premature balding. *Clin Endocrinol* **11:** 291–300
17 Carey AH *et al*. (1993) Evidence for a single gene effect causing polycystic ovaries and male pattern baldness. *Clin Endocrinol* **38:** 653–658
18 Legro RS *et al*. (1998) Evidence for a genetic basis for hyperandrogenemia in polycystic ovary syndrome. *Proc Natl Acad Sci U S A* **95:** 14956–14960
19 Kahsar-Miller M and Azziz R (1999) Heritability and the risk of developing androgen excess. *J Steroid Biochem Mol Biol* **69:** 261–268
20 Jahanfar S *et al*. (1997) A twin study of polycystic ovary syndrome and lipids. *Gynecol Endocrinol* **11:** 111–117
21 Kahsar-Miller MD *et al*. (2001) Prevalence of polycystic ovary syndrome (PCOS) in first degree relatives of patients with PCOS. *Fertil Steril* **75:** 53–58
22 Vink J *et al*. (2005) Heritability of polycystic ovary syndrome in a Dutch twin-family study. *J Clin Endocrinol Metab* **91:** 2100–2104
23 Urbanek M *et al*. (1999) Thirty-seven candidate genes for polycystic ovary syndrome: strongest evidence for linkage is with follistatin. *Proc Natl Acad Sci USA* **96:** 8573–8578
24 Escobar-Morreale HF *et al*. (2005) The molecular-genetic basis of functional hyperandrogenism and the polycystic ovary syndrome. *Endocr Rev* **26:** 251–282
25 Newton-Cheh C and Hirschhorn JN (2005) Genetic association studies of complex traits: design and analysis issues. *Mutat Res* **573:** 54–69
26 Hirschhorn JN (2005) Genetic approaches to studying common diseases and complex traits. *Pediatr Res* **57:** 74R–77R
27 Hirschhorn JN and Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* **6:** 95–108
28 Hattersley AT and McCarthy MI (2005) What makes a good genetic association study? *Lancet* **366:** 1315–1323
29 Zawadzki JK and Dunaif A (1992) Diagnostic criteria for polycystic ovary syndrome: towards a rational approach. In *Polycystic Ovary Syndrome,* 377–384 (Eds Dunaif A *et al.*) Boston: Blackwell Scientific
30 The Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group (2004) Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod* **19:** 41–47
31 Altshuler D *et al*. (2000) The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* **26:** 76–80
32 Grant SF *et al*. (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* **38:** 320–323
33 Gharani N *et al*. (1997) Association of the steroid synthesis gene CYP11a with polycystic ovary syndrome and hyperandrogenism. *Hum Mol Genet* **6:** 397–402
34 Diamanti-Kandarakis E *et al*. (2000) Microsatellite polymorphism (tttta)(n) at -528 base pairs of gene CYP11α influences hyperandrogenemia in patients with polycystic ovary syndrome. *Fertil Steril* **73:** 735–741
35 Daneshmand S *et al*. (2002) Overexpression of theca-cell messenger RNA in polycystic ovary syndrome does not correlate with polymorphisms in the cholesterol side-chain cleavage and 17α-hydroxylase/C(17-20) lyase promoters. *Fertil Steril* **77:** 274–280
36 San Millan JL *et al*. (2001) Role of the pentanucleotide (tttta)(n) polymorphism in the promoter of the CYP11a gene in the pathogenesis of hirsutism. *Fertil Steril* **75:** 797–802
37 Gaasenbeek M *et al.* (2004) Large-scale analysis of the relationship between CYP11A promoter variation, polycystic ovarian syndrome, and serum testosterone. *J Clin Endocrinol Metab* **89:** 2408–2413
38 Bell GI *et al*. (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature* **295:** 31–35
39 Pugliese A and Miceli D (2002) The insulin gene in diabetes. *Diabetes Metab Res Rev* **18:** 13–25
40 Waterworth DM *et al*. (1997) Linkage and association of insulin gene VNTR regulatory polymorphism with polycystic ovary syndrome. *Lancet* **349:** 986–990
41 Michelmore K *et al*. (2001) Clinical features in women with polycystic ovaries: relationships to insulin sensitivity, insulin gene VNTR and birth weight. *Clin Endocrinol (Oxf)* **55:** 439–446
42 Calvo RM *et al*. (2002) Insulin gene variable number of tandem repeats regulatory polymorphism is not associated with hyperandrogenism in Spanish women. *Fertil Steril* **77:** 666–668
43 Powell BL *et al*. (2005) Analysis of multiple data sets reveals no association between the insulin gene variable number tandem repeat element and polycystic ovary syndrome or related traits. *J Clin Endocrinol Metab* **90:** 2988–2993
44 Hanis CL *et al*. (1996) A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nat Genet* **13:** 161–166
45 Horikawa Y *et al*. (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* **26:** 163–175
46 Evans JC *et al*. (2001) Studies of association between the gene for calpain-10 and type 2 diabetes mellitus in the United Kingdom. *Am J Hum Genet* **69:** 544–552
47 Tsai HJ *et al*. (2001) Type 2 diabetes and three calpain-10 gene polymorphisms in Samoans: no evidence of association. *Am J Hum Genet* **69:** 1236–1244
48 Ehrmann DA *et al*. (2002) Relationship of calpain-10 genotype to phenotypic features of polycystic ovary syndrome. *J Clin Endocrinol Metab* **87:** 1669–1673
49 Haddad L *et al*. (2002) Variation within the type 2 diabetes susceptibility gene calpain-10 and polycystic ovary syndrome. *J Clin Endocrinol Metab* **87:** 2606–2610
50 Urbanek M *et al*. (2005) Candidate gene region for polycystic ovary syndrome on chromosome 19p13.2. *J Clin Endocrinol Metab* **90:** 6623–6629
51 Tucci S *et al*. (2001) Evidence for association of polycystic ovary syndrome in Caucasian women with a marker at the insulin receptor locus. *J Clin Endocrinol Metab* **86:** 446–449
52 Villuendas G *et al*. (2003) Association between the D19S884 marker at the insulin receptor gene locus and polycystic ovary syndrome. *Fertil Steril* **79:** 219–220
53 Hata R *et al*. (2000) Association of functional microsatellites in the human type I collagen α2 chain (COL1A2) gene with systemic sclerosis. *Biochem Biophys Res Commun* **272:** 36–40
54 Fenech AG *et al*. (2004) Novel polymorphisms influencing transcription of the human CHRM2 gene in airway smooth muscle. *Am J Respir Cell Mol Biol* **30:** 678–686

55 Huang TS *et al*. (2003) Shortening of microsatellite deoxy(CA) repeats involved in GL331-induced down-regulation of matrix metalloproteinase-9 gene expression. *Biochem Biophys Res Commun* **300:** 901–907
56 Gabellini N (2001) A polymorphic GT repeat from the human cardiac Na^+Ca^{2+} exchanger intron 2 activates splicing. *Eur J Biochem* **268:** 1076–1083
57 Hui J *et al*. (2003) HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat Struct Biol* **10:** 33–37
58 Abbott DH *et al*. (2002) Developmental origin of polycystic ovary syndrome—a hypothesis. *J Endocrinol* **174:** 1–5
59 Abbott DH *et al*. (2005) Androgen excess fetal programming of female reproduction: a developmental aetiology for polycystic ovary syndrome? *Hum Reprod Update* **11:** 357–374
60 Foecking EM *et al*. (2005) Neuroendocrine consequences of prenatal androgen exposure in the female rat: absence of luteinizing hormone surges, suppression of progesterone receptor gene expression, and acceleration of the gonadotropin-releasing hormone pulse generator. *Biol Reprod* **72:** 1475–1483
61 de Zegher F *et al*. (2006) Prenatal growth restraint followed by catch-up of weight: a hyperinsulinemic pathway to polycystic ovary syndrome. *Fertil Steril* **86:** S4
62 Ibanez L *et al*. (2002) Reduced ovulation rate in adolescent girls born small for gestational age. *J Clin Endocrinol Metab* **87:** 3391–3393
63 Ibanez L *et al*. (2002) Anovulation in eumenorrheic, nonobese adolescent girls born small for gestational age: insulin sensitization induces ovulation, increases lean body mass, and reduces abdominal fat excess, dyslipidemia, and subclinical hyperandrogenism. *J Clin Endocrinol Metab* **87:** 5702–5705
64 Ibanez L *et al*. (2001) Polycystic ovary syndrome after precocious pubarche: ontogeny of the low birthweight effect. *Clin Endocrinol* **55:** 667–672
65 Ibanez L *et al*. (2006) Early development of adiposity and insulin resistance after catch-up weight gain in small-for-gestational age children. *J Clin Endocrinol Metab* **91:** 2153–2158
66 Hickey TE *et al*. (2006) Epigenetic modification of the X chromosome influences susceptibility to polycystic ovary syndrome. *J Clin Endocrinol Metab* **91:** 2789–2791
67 Mhatre AN *et al*. (1993) Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. *Nat Genet* **5:** 184–188
68 Tut TG *et al.* (1997) Long polyglutamine tracts in the androgen receptor are associated with reduced trans-activation, impaired sperm production, and male infertility. *J Clin Endocrinol Metab* **82:** 3777–3782
69 Jakubiczka S *et al*. (1997) Mutations of the androgen receptor gene in patients with complete androgen insensitivity. *Hum Mutat* **9:** 57–61
70 Jaaskelainen J *et al*. (2005) Androgen receptor gene CAG length polymorphism in women with polycystic ovary syndrome. *Fertil Steril* **83:** 1724–1728
71 Legro R *et al*. (1994) Size polymorphisms of the androgen receptor among female Hispanics and correlation with androgenic characteristics. *Obstet Gynecol* **83:** 701–706
72 Mifsud A *et al*. (2000) Androgen receptor gene CAG trinucleotide repeats in anovulatory infertility and polycystic ovaries. *J Clin Endocrinol Metab* **85:** 3484–3488
73 Mohlig M *et al*. (2006) The androgen receptor CAG repeat modifies the impact of testosterone on insulin resistance in women with polycystic ovary syndrome. *Eur J Endocrinol* **155:** 127–130
74 Hickey T *et al*. (2002) The androgen receptor CAG repeat polymorphism and X-chromosome inactivation in Australian Caucasian women with infertility related to polycystic ovary syndrome. *J Clin Endocrinol Metab* **87:** 161–165
75 The International HapMap Consortium (2003) The International HapMap Project. *Nature* **426:** 789–796
76 Altshuler D *et al*. (2005) A haplotype map of the human genome. *Nature* **437:** 1299–1320
77 Pasternack JJ (1999) Discovering human disease genes. In *An Introduction to Human Molecular Genetics*, 218–219 (Ed. Pasternack JJ) Bethesda, MD: Fitzgerald Science Press
78 Metzker ML (2005) Emerging technologies in DNA sequencing. *Genome Res* **15:** 1767–1776

Competing interests
The author declared she has no competing interests.