The genetic basis of polycystic ovary syndrome
N Xita, I Georgiou and A Tsatsoulis
Division of Endocrinology, Department of Medicine and Genetics Unit, Department of Obstetrics and Gynaecology, University of Ioannina, Ioannina 45110, Greece
(Correspondence should be addressed to A Tsatsoulis; Email: atsatsou@cc.uoi.gr)

Abstract
Polycystic ovary syndrome (PCOS) is a common endocrine disorder in women of reproductive age. The disorder is characterized by clinical features of hyperandrogenism, menstrual irregularities and often central obesity and hyperinsulinaemia. PCOS may increase the risk for infertility, type 2 diabetes mellitus, dyslipidaemia, cardiovascular disease and endometrial cancer, emphasizing the need for early diagnosis of the syndrome.

The genetic basis of PCOS is unknown. There is a strong familial component but the mode of inheritance is uncertain and several candidate genes have been proposed to contribute to susceptibility. Not only genes involved in steroid hormone biosynthesis have been studied but also genes associated with the regulation of insulin secretion and action since hyperinsulinaemia is a characteristic of PCOS. So far there is evidence that INS VNTR (insulin variable number of tandem repeats) or CYP11a (cholesterol side chain cleavage) genes are associated with this syndrome. PCOS appears, however, to be an oligogenic disorder and more studies are necessary to define the genetic basis.

European Journal of Endocrinology 147 717–725

Introduction
Polycystic ovary syndrome (PCOS) is the most common endocrine disorder affecting women of reproductive age. It is considered to be a syndrome not a disease although the term polycystic ovarian disease is often used synonymously, but incorrectly, to describe the syndrome which is manifested by heterogeneous clinical features. The most common features of PCOS are irregular menstrual cycles (oligomenorrhoea or amenorrhoea), signs of androgen excess (hirsutism, acne, alopecia) and often obesity. However, only 5–10% of women with PCOS express the typical clinical features of the syndrome (1, 2). At present the diagnosis of PCOS is usually based on the criteria derived from the 1990 NIH-NICHHD (National Institutes of Health–National Institute of Child Health and Human Development) conference, which are ovulatory dysfunction, clinical evidence of hyperandrogenism and/or hyperandrogenaemia and exclusion of related disorders such as congenital adrenal hyperplasia, hyperprolactinaemia or Cushing’s syndrome (3). The finding of polycystic ovaries on ultrasound, which is defined by the presence of eight or more subcapsular follicular cysts ≤10 mm in diameter and increased ovarian stroma, is not a prerequisite for the diagnosis of PCOS according to these criteria (4).

Data on the epidemiology of PCOS are variable, because of the lack of well-accepted criteria for diagnosis. Studies using ultrasonography to identify the presence of polycystic ovaries have reported prevalence rates of 21–22% among randomly selected women (5). Although menstrual cycle disturbance and hyperandrogenism are more common in women with polycystic ovaries it is generally accepted that many women with polycystic ovaries are endocrinologically normal. Two recent studies have assessed the prevalence of PCOS as defined by both oligomenorrhoea and hyperandrogenism at 4.6% with a possible range of 3.4–11.2% (6). Ethnic differences in the prevalence of PCOS have not been well explored. Knochenhauer et al. indicated no significant difference in prevalence between white and black women in the USA, reporting an overall 4.0% prevalence of PCOS (7). Diamanti-Kandarakis et al. reported a 6.8% prevalence of PCOS in the Greek island of Lesbos (8) and a similar prevalence (6.5%) was found in Caucasian women from Madrid, Spain (9). There are also significant ethnic and racial variations in the clinical presentation of PCOS, the frequency of obesity, insulin resistance and the incidence of diabetes mellitus. A study comparing women with PCOS from the USA, Italy and Japan reported less obesity and hirsutism among Japanese women, yet comparable rates of biochemical androgen excess and insulin resistance (10). Dunaif et al. (11) studied the
effect of PCOS and ethnicity on insulin action comparing Caribbean-Hispanic women with non-Hispanic Caucasian women and they found that Caribbean-Hispanics with PCOS tended to be the most insulin resistant. The effect of ethnicity on the prevalence and presentation of PCOS was also examined in another study comparing European, Maori and Pacific Island women and it was reported that Maori and Pacific Island women with PCOS were more obese, more insulin resistant and more likely to present lipid abnormalities and infertility than European women (12).

Present data strongly support an association between PCOS and several long-term disease risks. Conditions that have been linked to PCOS include type 2 diabetes, hypercholesterolaemia, hypertension, cardiovascular disease, gestational diabetes mellitus, pregnancy-induced hypertension and endometrial cancer, and recently associations between PCOS and breast cancer and ovarian cancer have been reported (6, 13). This emphasizes the need for early diagnosis of the syndrome and close follow-up of women with PCOS.

Several reports have stressed that PCOS is a familial disorder and various features of the syndrome may be differentially inherited (14). However, the genetic basis of the syndrome remains controversial. It is difficult to determine the mode of inheritance of this syndrome because its presentation is heterogeneous and there are no universally accepted diagnostic criteria. In addition there is an absence of an equivalent male phenotype and PCOS primarily affects women of reproductive age. However, a recent study of 150 subjects revealed an autosomal dominant mode of inheritance considering premature balding in men as the primary male phenotype (15, 16). On the other hand, there are studies of families with high prevalence of PCOS in which the Mendelian autosomal dominant mode of inheritance cannot explain the mode of inheritance of the syndrome (17) while in another study an X-linked model was postulated (18). As a result, the mode of inheritance remains unclear and more than one gene defect seems to participate in the pathogenesis of the syndrome.

PCOS appears to be an oligogenic disorder. Several genes involved in reproduction, genes affecting the secretion or action of insulin and those involved in obesity and energy regulation have been tested as candidate genes. In particular, attention has been focused on genes coding for steroidogenic enzymes in

Figure 1 Suggested pathogenic mechanisms in the development of PCOS.
the androgen biosynthetic pathway and those involved in the secretion and the action of insulin (Fig. 1: Table 1).

The aim of this review is to present a critical analysis of the recent data on the genetics of PCOS.

**Genes involved in secretion and action of insulin**

Many studies have revealed abnormalities of both the secretion and action of insulin in women with PCOS. More than a decade ago, convincing evidence showed that most women with PCOS, both obese and lean, have a degree of insulin resistance (resistance to the action of insulin in skeletal muscles and liver) and compensatory hyperinsulinaemia in comparison with normal women of the same age and weight (19). This led to the assumption that genes involved in the secretion and action of insulin may play a role in the pathogenesis of PCOS.

**The insulin receptor gene**

The impaired sensitivity to insulin action both in vivo and in vitro led to the hypothesis that genetic abnormalities of the insulin receptor gene or the post-receptor signalling may contribute to the pathogenesis of PCOS.

**Table 1** Candidate genes investigated for their possible association with PCOS.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steroidogenic acute regulatory protein (StAR)</td>
<td>Encodes a key enzyme in the biosynthesis of androgens</td>
</tr>
<tr>
<td>CYP11α-cytochrome P450 side-chain cleavage enzyme</td>
<td>Encoding a key enzyme in the biosynthesis of androgens</td>
</tr>
<tr>
<td>CYP17-cytochrome P450 17α-hydroxylase/17,20 lyase</td>
<td>Encoding a key enzyme in the biosynthesis of androgens</td>
</tr>
<tr>
<td>CYP19-cytochrome P450 aromatase</td>
<td>Encoding a key enzyme in the biosynthesis of estrogen</td>
</tr>
<tr>
<td>17β-hydroxysteroid dehydrogenase type I, II, III</td>
<td>Encoding enzymes involved in the biosynthesis of androgens</td>
</tr>
<tr>
<td>3β-hydroxysteroid dehydrogenase type I, II</td>
<td>Encoding enzymes involved in the biosynthesis of androgens</td>
</tr>
<tr>
<td>Insulin receptor</td>
<td>Key component in the action of insulin</td>
</tr>
<tr>
<td>Insulin gene VNTR</td>
<td>Variants exist that may alter the function of the insulin receptor</td>
</tr>
<tr>
<td>Insulin receptor substrate 1</td>
<td>Variants exist that may alter the function of the insulin receptor</td>
</tr>
<tr>
<td>Insulin-like growth factor-I receptor</td>
<td>Variants exist that may alter the function of the insulin receptor</td>
</tr>
<tr>
<td>Insulin-like growth factor binding protein 1 + 3</td>
<td>Variants exist that may alter the function of the insulin receptor</td>
</tr>
<tr>
<td>Peroxisome proliferator-activated receptor-gamma</td>
<td>Variants exist that may alter the function of the insulin receptor</td>
</tr>
</tbody>
</table>

**The insulin gene**

The observation that although insulin resistance is largely reversible by weight reduction in obese women with PCOS, an abnormality of first-phase insulin secretion from β-cells of the pancreas still exists, led to the hypothesis that genetic abnormalities of the insulin gene may contribute to the pathogenesis of PCOS.
to investigation of the role of the insulin gene in the pathogenesis of PCOS (24).

The insulin gene has been localized between genes for tyrosine hydroxylase and insulin-like growth factor-II (IGF-II) on 11p15.5 (25) (Fig. 2). The minisatellite of insulin gene INS VNTR (insulin gene variable number of tandem repeats), which lies 5′ to the insulin gene, has been investigated, since this region is directly implicated in the regulation of insulin secretion. This minisatellite is a polymorphic region composed of tandemly repeated 14–15 bp sequences related to the most common repeat ACAGGGGTGTGGGG and positioned 596 bp from the insulin translation starting codon (ATG) in the promoter region. The number of the repeats of the INS VNTR ranges from about 26 to 200 (26).

There are three related to the size classes of INS VNTR alleles. Class I are the shorter polymorphic regions and are composed of alleles with 26–63 repeat units (average 40 repeats). Class II alleles are rare in Caucasians and have an average length of about 80 repeat units. Class III are the longest alleles having 141–209 repeat units (average 157 repeats) (26).

The INS VNTR is a functional polymorphism, so it regulates the transcription of the insulin gene and probably the expression of the IGF-II gene which is adjacent to the insulin gene (27). It does not encode any known gene product and it has been shown that an insulin promoter containing the long INS VNTR allele has greater transcriptional activity than one containing a short INS VNTR allele (28). The INS VNTR might transduce transcriptional signals by binding to one or more transcriptional factors, as is usually the case with transcriptional control regions. Kennedy et al. (28) have shown that the INS VNTR contains high-affinity binding sites for the transcriptional factor Pur-1, which has the capacity after binding to INS VNTR to raise basal transcription from the insulin promoter even in cells that do not normally express that gene.

The INS VNTR and especially class III alleles and III/III genotypes have been associated with insulin resistance (29) and diabetic hypertriglyceridaemia (30). Class III alleles have also been associated with central obesity (29) and increased birth weight (31). A meta-analysis supports an association of INS VNTR III/III with type 2 diabetes. Furthermore the type 1 diabetes locus IDDM2 has been mapped to INS VNTR and class I allele homozygosity is associated with an elevated risk of developing type 1 diabetes (32, 33).

Waterworth et al. (34) reported an association between PCOS and allelic variation at the INS VNTR locus in three different populations. They showed that class III alleles and especially III/III genotypes are associated with PCOS and are most strongly associated with anovulatory PCOS. The group of women with one or two class III alleles had significantly higher fasting insulin levels and higher mean body mass index than women with I/I genotype. This is in agreement with the observation that hyperinsulinaemia is a more prominent feature in women with polycystic ovaries who have anovulatory menses than in equally hyperandrogenaemic subjects with regular menses (35). It has also been observed that the homozygous genotype III/III is more frequent among women with polycystic ovaries and symptoms than among women with polycystic ovaries without symptoms (34). Waterworth

Figure 3 The major steroid biosynthetic pathways in adrenal cortex and theca cells of ovaries. P450scc, cholesterol side chain cleavage; 3β-HSD, 3β-hydroxysteroid dehydrogenase and Δ5-isomerase; 17β-HSD, 17β-hydroxysteroid dehydrogenase, 17α-hydroxylase activity and 17,20 lyase activity is mediated by P450c17α.
and colleagues (36) also reported that class III alleles are transmitted more often than class I alleles from I/III heterozygous parents to girls with anovulatory PCOS. The transmission of class III alleles from mothers is found to be 47% while the transmission from fathers is more common. These findings were confirmed in other studies (37, 38). If we take into consideration that INS VNTR class III alleles have also been associated with type 2 diabetes (33) we can conclude that the INS VNTR genotype might not only determine the predisposition to anovulatory PCOS but also the concomitant risk for development of type 2 diabetes. This assumption remains to be confirmed, however.

Genes encoding steroidogenic enzymes

The 17-hydroxylase/17,20-lyase gene (CYP17)

Recent studies have shown that PCOS may be the result of overfunction of the enzyme which catalyses androgen production (cytochrome P450c17α). Cytochrome P450c17α is an enzyme with two functions since it has both 17α-hydroxylase and 17,20-lyase activities. In the thecal cells P450c17α converts progesterone to 17α-hydroxyprogesterone through its 17α-hydroxylase activity and then it converts 17α-hydroxyprogesterone to androstenedione through its 17,20-lyase activity (39) (Fig. 3).

Clinical studies have shown an abnormality in the regulation of 17α-hydroxylase/17,20-lyase (the rate-limiting step in androgen biosynthesis in the ovaries and the adrenals) in women presenting with PCOS, as evidenced by increased 17α-hydroxylase and to a lesser extent 17,20-lyase activity since in these women there is an exaggerated serum 17α-hydroxyprogesterone response to stimulation by gonadotrophin-releasing hormone agonists (40). Since the relative ratio of 17α-hydroxylase to 17,20-lyase activity of P450c17α represents a key control locus for androgen biosynthesis, a shift in this ratio could account for aberrations in ovarian and adrenal androgen production. As a result the initial investigations focused on the possible role of CYP17, the gene that codes cytochrome P450c17α and which is located on chromosome 10q24.3 (41). A polymorphism has been found in the regulatory region of CYP17 gene. This was a single base change, a T to C substitution 34 bp from the translation initiation point in the promoter region. It has been proposed that this change may up-regulate the expression of CYP17 resulting in an increased synthesis of androgens. Initial investigations associated this polymorphism with PCOS but included only a small number of subjects (42). An extension of the initial study to a larger case-control study (96 persons) showed no statistically significant difference in the frequency of this polymorphism between patients with PCOS and the control group and also no association between this polymorphism and the levels of testosterone (43). Therefore this base change identified in the promoter region of the CYP17 gene is a common polymorphism with no obvious role in the aetiology of PCOS or hyperandrogenaemia.

It has recently been shown that serine phosphorylation is also involved in the post-translational regulation of 17,20-lyase activity and therefore androgen secretion. The serine residues that are phosphorylated and the kinase that mediates the phosphorylation remain to be identified. However, this observation raises the possibility that there may be a common aetiology for both the insulin resistance and hyperandrogenism in PCOS (44).

The cholesterol side chain cleavage gene (CYP11α)

It has been shown that theca cells from polycystic ovaries produce an excess of progesterone, 17α-hydroxyprogesterone and androstenedione in vitro when compared with normal theca cells (45). It was assumed that the defect in steroidogenesis occurs at or above the level of progesterone biosynthesis. The enzyme for cholesterol side chain cleavage (cytochrome P450scC) catalyses the conversion of cholesterol to pregnenolone, which is the initial and rate-limiting step at the start of the steroid hormone biosynthetic pathway (Fig. 3). It has been hypothesized that up-regulation of this enzyme could lead to increased androgen production (41).

This observation led to studies of the cholesterol side chain cleavage gene (CYP11α), which encodes cytochrome P450scC. This gene is located on the long arm of chromosome 15 at position 15q24 (41). In the promoter region of CYP11α a polymorphic region has been identified. This polymorphism is a pentanucleotide repeat (tttta)n ~ 528 bp 5' from the ATG initiation of the translation site. It is postulated that this polymorphism, being a variable number tandem-repeat (VNTR), might play an important role in the regulation of CYP11α expression, possibly by increasing the rate of transcription. Four-, six-, eight- and nine-repeat-units have been found in the normal population, with the allele carrying four-repeat-units being the most common. PCOS appears to be associated with the absence of the four-repeat-units allele. It has been found that this CYP11α allele is also associated with serum testosterone levels (46). The association of the (tttta)n polymorphism and PCOS has been confirmed by Diamanti-Kandarakis et al. in Greek patients (47). This suggests that CYP11α may be a major genetic susceptibility locus for PCOS. However, a more recent study failed to confirm the linkage between the CYP11α polymorphism and PCOS. Results were initially significant for association with PCOS but after correction for multiple testing no association was
found between any of the alleles of the CYP11A and the presence of PCOS (48). Another study in Spain found no influence of this polymorphism in any aspect related to hyperandrogenism (49). The difference between these results and the previous ones may be due to ethnic or racial differences between the studied populations.

**The aromatase gene (CYP19)**

Aromatase is an enzyme complex (cytochrome P450arom) that catalyses the conversion of the C19 steroids (androgens) to C18 (oestrogens) (50). Some cases of hyperandrogenism have been reported in patients with aromatase deficiency (51, 52). Immunohistochemical studies of polycystic ovaries could not detect aromatase in antral follicles of various sizes (53). On the other hand it has been found that granulosa cells from anovulatory polycystic ovaries are hyperresponsive to follicle-stimulating hormone (FSH) in vitro, displaying significantly greater oestradiol production than granulosa cells from normal ovaries, suggesting no functional deficiency of aromatase in polycystic ovaries (54).

These results led to the study of the aromatase gene (CYP19) located on chromosome 15 at position 15q21.1 (50). Thus far, there is no evidence of any association of alleles of this gene with PCOS (46).

**The luteinizing hormone (LH) β-subunit gene**

Another gene studied is the LH β-subunit gene, since 40% of women with PCOS have hypersecretion of LH and this is associated with anovulation possibly through an adverse effect of LH on oocyte maturation (55).

Two molecular variants of LH have been recently identified. One is an immunologically anomalous LH, caused by two point mutations in exon 2 of the LH β-subunit gene, which are changes of residues Trp (TGG) to Arg (CGG) at position 8 and Ile (ATC) to Thr (ACC) at position 15. These mutations lead to structural changes in the LH molecules (56, 57). This first genetic variant is a universally common polymorphism with a prevalence of 18.5%. The highest frequencies were found between any of the alleles of the CYP11A and the presence of PCOS (48). Another study in Spain found no influence of this polymorphism in any aspect related to hyperandrogenism (49). The difference between these results and the previous ones may be due to ethnic or racial differences between the studied populations.

**The androgen receptor (AR) gene**

All androgens act through the AR. The AR belongs to a family of nuclear transcription factors. AR is encoded by a gene located on the X chromosome at Xq11-12 (64, 65). The AR, like other members of the steroid receptor family, has three functional domains: the transactivation domain, the DNA-binding domain and the ligand-binding domain. The AR gene contains a polymorphic CAG repeat within exon 1, which encodes a polyglutamine chain in the N-terminal transactivation domain (66). In healthy populations the number of CAG repeats ranges from 11 to 31, with the most common allele having 20 repeats (67). The length of the polymorphic CAG repeat sequence is inversely correlated to the AR transcriptional activity. The longer the CAG repeat the lower the transcriptional activity of the AR (68).

Clinical studies support the functional importance of the CAG repeat sequence of the AR gene. Pathological elongation of this CAG repeat polymorphism (>40 CAG repeats) causes Kennedy’s syndrome, an X-linked syndrome of progressive spinal and bulbar muscular atrophy that is associated with variable degrees of androgen insensitivity (69). Relatively few (<22) CAG repeats are associated with a higher risk of prostate cancer (70) and a lower risk of infertility in men (71).

An association between increased hirsutism and decreased CAG repeat length has been demonstrated in a group of Hispanic women with normal testosterone levels (72) and an association between short CAG repeat length and androgenic skin disorders has also been demonstrated in both men and women (73). However, other investigators were unable to demonstrate differences in the number of repeats between hirsute patients and normal controls and support the theory that hirsutism is associated with skewing of the X-chromosome inactivation which leads to the longer
of the two AR alleles being preferentially methylated to silence, thus allowing the shorter, and more functional, AR allele to be expressed on the active X-chromosome (74).

Recently the hypothesis that CAG repeat polymorphism of the AR influences the disease process leading to PCOS has been put forward. It has been shown that an association exists between short CAG repeat length and low serum androgens in the subset of anovulatory patients, suggesting that the pathogenic mechanism of polycystic ovaries in these patients could be due to the increased intrinsic androgenic activity associated with short AR alleles (75).

However, further studies need to be conducted to investigate the role of AR polymorphisms in the pathogenesis of PCOS.

The gene of follistatin

Recently Urbanek et al. (48) investigated 37 candidate genes involved in reproduction, insulin secretion and action and energy metabolism and found statistically significant evidence with a marker near the follistatin gene, showing a possible association between the gene of follistatin and PCOS. Follistatin is a binding protein of activin and acts by neutralizing the biological activity of activin both in vivo and in vitro (76). Both follistatin and activin are expressed in several tissues including the ovaries, pituitary, adrenal cortex and pancreas (77). Activin enhances the development of ovarian follicles, inhibits androgen production by ovarian thecal cells, and increases pituitary FSH release and pancreatic β-cell insulin secretion (77, 78). An increase in the activity of follistatin has the opposite effect and this has been confirmed in animals with elevated levels of follistatin which developed phenotypes similar to PCOS (79).

However, sequencing of the follistatin gene has failed to detect any variants associated with PCOS (80, 81). Seventeen genetic variants of the gene were studied of which 16 were very rare and the only common genetic variant was a substitute of a single base in the last exon, a region that is untranslated. So the contribution of follistatin gene in PCOS, if any, is only modest.

Conclusions

The genetic aetiology of PCOS remains unknown. A family history of PCOS is relatively common among PCOS patients, suggesting a significant genetic component, although the mode of inheritance has not been firmly established.

Genetic studies face certain difficulties in their performance, for several reasons, but PCOS is considered to be an oligogenic disorder in which the interaction of a number of genetic and environmental factors (notably dietary) seem to determine the typically heterogeneous, clinical and biochemical phenotype. A potential approach to deal with heterogeneity would be the implementation of strict clinical criteria facilitating the grouping of PCOS patients according to (biochemical and/or clinical parameters) differences or similarities of the phenotype, dietary factors and family history.

Several candidate genes have been tested as the underlying genes for PCOS, but evidence that any of these play a role in PCOS has not been convincingly replicated. The INS VNTR and the CYP11A, which encodes P450 side chain cleavage, appear to be promising candidate genes at least in some populations. More candidates genes, especially those involved in the metabolic and transport pathways of hormones and accompanying molecules, would be appropriate for further investigations. However, further studies are necessary to determine the genetic associations with certain genes. Our knowledge of the human genome is expanding rapidly. Novel polymorphisms have been identified which may allow detailed mapping of the disease locus for PCOS in the future. This will permit family screening, early diagnosis of women with PCOS and definition of those who are at risk of developing complications.

References

10 Carmina E, Koyama T, Chang L, Stanczyk FZ & Lobo RA. Does ethnicity influence the prevalence of adrenal hyperandrogenism
39 Rosenfeld RL, Barnes RB, Cara JF & Lucky AW. Dysregulation of cystochrome P450c17a as the cause of polycystic ovarian syndrome. Fertility and Sterility 1990 53 785–791.
47 Diamantopoulou-Chaouakis E, Bartazzi MI, Bergiele AT, Tsianateli TC & Kouli CR. Microsatellite polymorphism (t(ita)ta) at 5q28 base pairs of gene CYP11a influences hyperandrogenemia in patients with polycystic ovary syndrome. Fertility and Sterility 2000 73 735–741.
50 Nebert DW, Nelson DR, Adesnik M, Coon MJ, Estabrook RW,