0021-972X/04/$15.00/0
Printed in U.S.A.

The Journal of Clinical Endocrinology & Metabolism 89(5):2408–2413

doi: 10.1210/jc.2003-031640

Large-Scale Analysis of the Relationship between *CYP11A* Promoter Variation, Polycystic Ovarian Syndrome, and Serum Testosterone

MICHELLE GAASENBEEK, BRENDA L. POWELL, ULLA SOVIO, LEMA HADDAD, NEDA GHARANI, AMANDA BENNETT, CHRISTOPHER J. GROVES, KAREN RUSH, MICAELA J. GOH, GERARD S. CONWAY, AIMO RUOKONEN, HANNU MARTIKAINEN, ANNELI POUTA, SAARA TAPONEN, ANNA-LIISA HARTIKAINEN, STEPHANIE HALFORD, MARJO-RIITTA JÄRVELIN, STEVE FRANKS, AND MARK I. MCCARTHY

Genomic Medicine, Faculty of Medicine (M.G., L.H., N.G., A.B., S.H., M.I.M.) and Institute of Reproductive and Developmental Biology (M.G., S.F.), Imperial College, Hammersmith Campus, London W12 0NN, United Kingdom; Departments of Epidemiology and Public Health (U.S., M.-R.J.) and Reproductive Endocrinology (K.R., M.J.G., S.F.), Imperial College, St. Mary's Campus, London W2 1PG, United Kingdom; Wellcome Trust Centre for Human Genetics (B.L.P., C.J.G., M.I.M.) and Oxford Centre for Diabetes, Endocrinology and Metabolism (B.L.P., A.B., C.J.G., M.I.M.), Churchill Hospital, Oxford OX3 7LJ, United Kingdom; Department of Obstetrics and Gynaecology (G.S.C.), Bloomsbury Site, University College, London WC1E 6BT, United Kingdom; and Departments of Clinical Chemistry (A.R., S.T.), Obstetrics and Gynecology (S.T., H.M., A.-L.H.), and Public Health Science and General Practice (M.-R.J., S.T., A.P.), University of Oulu, Oulu, Finland 90014

***CYP11A*, the gene encoding P450scc, a key enzyme in steroid biosynthesis, is a strong biological candidate for polycystic ovary syndrome (PCOS) susceptibility. Four of the five published studies that have examined *CYP11A* for evidence of linkage and/or association have reported significant relationships with polycystic ovary (PCO) status and/or serum testosterone levels. However, study sizes have been modest, and the current study aimed to reevaluate these findings using significantly larger clinical resources. A pair of *CYP11A* promoter microsatellites, including the pentanucleotide (D15S520) previously implicated in trait susceptibility, were genotyped in 371 PCOS patients of United Kingdom origin, using both case-control and family-based association methods, and in 1589 women from a population-based birth cohort from Finland characterized for PCO symptomatology and testosterone levels. Although nominally significant differences in allele and genotype frequencies at both loci were observed in the United Kingdom case-control study (for example, an excess of the pentanucleotide four-repeat allele in cases, $P = 0.005$), these findings were not substantiated in the other analyses, and no discernable relationship was seen between variation at these loci and serum testosterone levels. These studies indicate that the strength of, and indeed the existence of, associations between *CYP11A* promoter variation and androgen-related phenotypes has been substantially overestimated in previous studies. (*J Clin Endocrinol Metab* 89: 2408–2413, 2004)**

POLYCYSTIC OVARY SYNDROME (PCOS) is a heterogeneous endocrine disorder of premenopausal women and the commonest cause of anovulatory infertility and hirsutism (1, 2). A cardinal feature of this condition is a disturbance of ovarian steroidogenesis, with excessive androgen production responsible for many of the principal symptomatic manifestations. In addition, individuals with PCOS display metabolic abnormalities that include prominent defects in both insulin action (3) and β-cell function (4).

PCOS displays strong familial aggregation (5), and the etiology is likely to be multifactorial, individual susceptibility being determined by the action of multiple genetic and environmental risk factors. A wide range of potential candidate genes have been studied in the search for PCOS-susceptibility variants, but few of these have been studied in multiple populations, and even fewer have shown convincing replication (6, 7).

Some of the strongest evidence presented to date has related to a pentanucleotide repeat 5′ to the *CYP11A* gene. The *CYP11A* gene is a credible biological candidate because it encodes P450scc (also known as cytochrome P450scc or cholesterol side-chain cleavage), the enzyme catalyzing the initial and rate-limiting step in adrenal and ovarian androgen biosynthesis. Thecal cells cultured from PCOS subjects show increased androgen production (8) and, in most studies, increased *CYP11A* mRNA expression (9–11) compared with samples from control women. Maintenance of this cellular phenotype on long-term culture (12) strengthens the hypothesis that *CYP11A* overexpression is a stable property of PCOS theca cells that contributes to the excess androgen production that characterizes this condition.

In the search for genetic variants that might be associated with *CYP11A* overexpression, attention has focused on a pentanucleotide repeat approximately 500 bp 5′ to the translation start site (D15S520). Gharani *et al.* (13) reported evidence for linkage to this region in 20 multiplex pedigrees

Abbreviations: PCO, Polycystic ovary; PCOS, PCO syndrome; USS, ultrasound scan.

JCEM is published monthly by The Endocrine Society (http://www.endo-society.org), the foremost professional society serving the endocrine community.

segregating PCOS (LOD, ~1.74), and in an analysis of 148 PCOS subjects and 59 controls, associations between pentanucleotide length (specifically absence of the 216-bp, four-repeat allele) and both symptomatic PCOS and raised testosterone levels were found. Subsequent studies have, to varying extents, corroborated these findings. Urbanek *et al.* (14) found nominal evidence for excess allele sharing in the region in an analysis of 28 multiplex sibships, as well as a nonsignificant reduction in transmission (47%) of the four-repeat allele to PCOS offspring in 163 parent-offspring trios. Diamanti-Kandarakis *et al.* (15) studied 170 Greek subjects (80 with PCOS) and found a significant increase in the frequency of four-repeat-lacking genotypes in PCOS subjects (26% *vs.* 13% in controls) and reproduced the previously observed association with testosterone. Most recently, Daneshmand *et al.* (10) also reported a reduction in the four-repeat allele frequency in PCOS (14% in 51 PCOS women *vs.* 44% in 280 controls), which was associated with a striking compensatory difference in the frequency of the nine-repeat allele (56% in PCOS women *vs.* 7% in controls). The only published study that has failed to produce any evidence for a relationship between pentanucleotide genotype and phenotype, a Spanish study of 92 hirsute women (only 34 of whom had PCOS) and 33 controls, was also the smallest and, consequently, likely to have been underpowered (16).

On the basis of these data, *CYP11A* is generally considered one of the few well-supported examples of a PCOS susceptibility effect (6, 7). However, the sample sizes included in the studies described have been rather modest. Increasing concern over the poor performance of association studies in the identification and characterization of complex trait susceptibility effects (17–19) prompted us to reevaluate the relationship between *CYP11A* promoter variation and PCOS disease status, polycystic ovary (PCO)-related symptoms, and serum testosterone levels.

In this study, we have extended genotyping of the *CYP11A* pentanucleotide, together with a neighboring tetranucleotide repeat element not previously studied, to almost 2500 individuals, which is around twice the combined size of the previous studies. Our analyses of these adequately sized and well-characterized populations indicate that these variants do not contribute significantly to observed phenotypic variation in either serum testosterone levels or PCOS disease status.

Subjects and Methods

United Kingdom (UK) subjects

Women with PCOS (n = 583) were ascertained from infertility and endocrine clinics, predominantly at St. Mary's and the Middlesex Hospitals in London. The criteria for diagnosis of PCOS were as previously described (5, 20) and, in line with the revised (2003) consensus on the diagnostic criteria for this syndrome (21), rely on the combination of clinical symptoms, ultrasonographic examination, and biochemical data. The essential criterion for the diagnosis of PCOS in this study was the presence of polycystic ovaries on ultrasound in a patient presenting with anovulation (oligomenorrhea or amenorrhea) and/or hyperandrogenism. A diagnosis of hyperandrogenism required clinical (presence of hirsutism or acne) and/or biochemical [serum total testosterone > 0.78 ng/ml (2.7 nmol/liter)] features. Other potential endocrine and neoplastic causes of hyperandrogenemia were excluded by appropriate tests (21). From this wider patient cohort of diverse ethnic origin, all of whom were typed for the variants of interest, we restrict the analyses reported here to the 371 women with exclusively British/Irish origin (based on three generations of reported ethnicity and birthplace). For 141 of these cases (designated probands), DNA was available from both parents, such that parent-offspring trios could be generated; the remaining 230 patients without parental DNA are designated PCOS cases. Family relationships in the trios were confirmed by genotyping with a panel of highly polymorphic microsatellite markers (combined heterozygosity, >99%). Quantitative trait analyses were performed in all PCOS subjects (*i.e.* probands and cases combined) but excluded any woman taking oral contraceptive treatment, pregnant at the time of sampling, or with previous oophorectomy. A small number of women (4%, 14 of 371 women) who had received metformin were included in these analyses; their exclusion did not alter the findings.

As population control subjects for case-control analyses, we used DNA (n = 350) from a random UK population of blood donors available from the Centre for Applied Microbiology and Research (Salisbury, UK), used extensively for such purposes. All control subjects were of British UK origin. Metabolic and hormonal status was not known for these individuals. Use of population controls (rather than subjects known not to have PCOS) is generally associated with only a modest reduction in power, which can readily be compensated by increasing the number of control subjects typed (19).

Finnish subjects

For the cohort study, samples were available from the Northern Finland Birth Cohort of 1966, which is based on 12,058 live-born babies born in the northernmost two provinces of Finland arising from pregnancies with expected dates of delivery during calendar year 1966. Of 5926 cohort members (3088 women) who had undergone clinical review and DNA sampling at age 31, we genotyped 1589 for this study (22). These included 527 women who had reported symptoms of hirsutism and/or oligo/amenorrhea on the 31-year review questionnaire (symptomatic women) and 1062 randomly selected women (cohort controls) from the same cohort without such symptoms (generating an approximate 1:2 ratio of cases to controls, all matched for age given the birth cohort design). No woman who was receiving oral contraceptives at the time of metabolic sampling was included in either group. The cohort cases have been shown as a group to have biochemical features consistent with PCOS (*e.g.* hyperandrogenemia and insulin resistance) (22), but only a proportion have proven PCOS. Ovarian ultrasound scan (USS) examination of these symptomatic women is ongoing, with 67 of 178 so far examined having PCO status confirmed on USS. These 67 individuals are referred to as USS-confirmed cohort cases. For the quantitative trait analyses, which were the main emphasis of analysis in the Finnish subjects, the group size was reduced to 1435 (475 cases and 960 controls) by exclusion of 138 women who were pregnant at the time of the blood sampling and 16 in whom no testosterone measures were available. The single subject who was taking metformin at the time of recruitment is included.

Clinical features of these sample sets are provided in Table 1. All clinical investigations were conducted in accordance with the guidelines in the Declaration of Helsinki, and the study was approved by the relevant ethics committees in UK and Finland. All subjects provided fully informed consent.

Laboratory assays

In the UK women with PCOS and all the Finnish subjects, fasting samples were used for the analysis of testosterone levels. For the UK subjects, testosterone was measured with an in-house RIA using ether extraction and dextran-coated charcoal separation. Genomic DNA was extracted from whole blood using the PUREGENE kit (Gentra Systems, Minneapolis, MN) as described previously (23). For the Finnish subjects, serum testosterone was determined by automated chemiluminescence system (Ciba-Corning ACS-180; Diamond Diagnostics, Holliston, MA).

Genotyping

The pentanucleotide D15S520 was genotyped by PCR amplification in a 15-μl volume containing 30 ng genomic DNA, 10 pmol of deoxynucleotide triphosphates, 1.0 mM $MgCl_2$, 1U *Taq* polymerase (Bioline, London, UK), and 8 μM of each of the forward (FAM-labeled

TABLE 1. Clinical characteristics of study subjects

	UK subjects			Finnish subjects	
	PCOS cases	PCOS probands	Population controls	Symptomatic women	Cohort controls
n	230	141	331	527	1062
% female	100	100	55.5	100	100
Age (yr)	33.8 (6.9)[a]	31.4 (6.1)[a]	38.6 (8.1)[b]	31	31
BMI (kg/m^2)	27.4 (7.8)	27.0 (7.4)		25.2 (5.4)	24.2 (4.6)
Waist to hip ratio	0.80 (0.07)	0.80 (0.09)		0.82 (0.08)	0.81 (0.08)
Testosterone (ng/ml)[c]	0.64 (0.26)	0.68 (0.34)		0.67 (0.29)	0.59 (0.25)
Glucose (mg/dl)[d]	86 (10)	85 (18)		90 (18)	88 (12)
Symptoms					
Hirsutes	105 (46%)	87 (62%)		295 (56%)	0
Oligomenorrhea	145 (63%)	115 (82%)		312 (60%)	0
Ovarian status	USS confirmed PCOS	USS confirmed PCOS		USS underway, 67/178 confirmed	Not known but asymptomatic

Quantitative data are presented as mean (SD).

[a] Ranges for subject ages were as follows: UK PCOS cases, 19.3–57.6 yr; UK PCOS probands, 16.5–47.7 yr. All Finnish women were aged 31 yr.

[b] Age information is available from only 38% of these control subjects.

[c] To convert to SI units (nanomoles per liter) multiply by 3.467.

[d] To convert to SI units (millimoles per liter) multiply by 0.05551.

5′-CCTGGGCAACAAGAATGAAA for the UK subjects and FAM-labeled 5′-GGTGAAACTGTGCCATTGC for the Finns) and reverse (5′-GCCCAAGAGAAGGTCAGAGC for the UK subjects and 5′-GTTT-GGGGGAAATGAGGGGC for the Finns) primers. Details of PCR conditions are available from the authors. Sequence analysis of the *CYP11A* gene revealed a second (tetranucleotide) microsatellite lying in intron 1 approximately 18.4 kb upstream of exon 2. This tetranucleotide lies immediately upstream of a novel exon that is expressed as the 5′ end of an alternate *CYP11A* transcript (Gaasenbeek, M., M. I. McCarthy, S. Halford, and S. Franks, unpublished observations). We elected to type this microsatellite, which lies approximately 1500 bp from D15S520, because sequence context suggested a potential functional role analogous to that of the pentanucleotide with respect to the usual promoter. This microsatellite was amplified in a separate reaction (primers: forward, HEX-labeled 5′-AGTAGCTGGGATTATAGACCC; reverse, 5′-GGAGACTGGTGAGGCTAAGT; 1.5 mM $MgCl_2$). Resulting PCR products were diluted 1:4 in water, and 0.5 μl of each PCR reaction was mixed with formamide and 450 bp ROX size standards (Applied Biosystems, Warrington, UK), followed by separation and sizing on an ABI PRISM 3700 DNA sequencer (Applied Biosystems). Genotypes were derived using Genotyper software (Applied Biosystems). When possible, Mendel-checking methods were used to confirm genotyping accuracy. In addition, approximately 10% of samples were re-genotyped. From these data, we estimate that the overall genotyping error is less than 1%. For convenience, alleles are designated by prefix P or T, reflecting pentanucleotide or tetranucleotide locus, respectively, and a number reflecting the number of repeat units (*e.g.* P4 includes four repeats at the pentanucleotide). We did not genotype either D15S169 or D15S519 [markers typed in some of the previous studies (13, 14)] either because of distance from the gene (D15S169) or uncertain location (D15S519 does not map to the current genome assembly).

Statistical analysis

Linkage disequilibrium relationships between the two microsatellites were determined in UK and Finnish control subjects using the LDMAX option within the GOLD (Graphical Overview of Linkage Disequilibrium) package (24), which implements the expectation-maximization (EM) algorithm. Between-group comparisons of genotype and allele frequencies used standard contingency table methods implemented in STATXACT (Cytel, Cambridge, MA), which allows exact probabilities (or Monte-Carlo estimates thereof) to be generated. Haplotype frequencies were derived by standard maximum likelihood methods implemented in TRANSMIT (25), with group comparisons assessed by likelihood ratio testing and permutation methods (at least 1000 permutations). In the UK trios, family-based association tests (25, 26) were undertaken using TRANSMIT; alleles and haplotypes with a frequency less than 2% were pooled for their respective analyses. Quantitative trait analyses of serum testosterone levels were performed by ANOVA on log-transformed values, using SPSS (version 10; SPSS Inc., Chicago, IL) and SAS (version 8.2; SAS Institute, Inc., Cary, NC), with and without adjustment for age. For analyses of testosterone levels, all women taking oral contraceptive medication at the time of sampling were excluded, as were women who were pregnant or had undergone oophorectomy.

Power calculations

The UK samples (case control + trios combined) provide approximately 85% power (at $P = 0.05$) to detect effect sizes (odds ratios for PCOS, ~1.4–1.5) significantly smaller than those implied by some earlier reports (odds ratio, >2) (13, 15). The largest group studied for the relationship between *CYP11A* variation and testosterone levels (the Finnish cohort controls) provided more than 90% power to detect (at $P = 0.05$) differences between P4-containing and non-P4-containing genotypes of approximately 0.33 of a SD. Previous studies of the pentanucleotide (13, 15) have suggested effect sizes at least twice this.

Results

In the UK samples, both microsatellites were confirmed as polymorphic. Variation at the pentanucleotide generated seven alleles ranging from 169 bp (P4) to 199 bp (P10), with the P4 allele being most common (frequency, ~58% in controls). The 169-bp (P4) allele is equivalent to the 216-bp allele designated in some previous papers, the difference reflecting use of differing PCR primers (13). Seven alleles were observed at the tetranucleotide ranging from 276 bp (T6) to 304 bp (T13), with the 284-bp (T8) modal (~82% in controls). In the Finnish controls, allele distributions were similar, with the modal alleles again being P4 (69%) and T8 (81%).

In the UK population controls, the two microsatellites were in strong linkage disequilibrium (D′, ~0.86; $P < 0.001$); the most marked departures from equilibrium were positive associations between the P4 and T8 alleles (56.5% observed *vs.* 47.6% expected) and between P6 and T12 (11.3% *vs.* 3.1%) and the complementary negative associations between P6 and T8 (10.8% *vs.* 21.7%) and P4 and T12 (0.4% *vs.* 6.8%). In the Finnish controls, the extent and direction of linkage disequilibrium relationships between alleles at the two loci were similar (D′, ~0.87; $P < 0.0001$).

At the pentanucleotide, comparisons between UK PCOS

cases and population controls revealed significant differences in genotype frequencies (Monte-Carlo estimate, $P = 0.029$; 99% confidence interval, 0.027–0.030), which were largely attributable to an excess of P4 homozygotes in the cases (47.2% *vs.* 34.8% in controls; Table 2). The same was true of allele frequencies (exact $P = 0.005$, reflecting P4 allele frequencies of 67.6% in cases and 58.2% in controls; Table 2). At the tetranucleotide, there were no significant differences in genotype frequencies ($P = 0.139$), but allele frequency comparisons indicated an excess of T8 alleles in cases ($P = 0.0097$; 88.4% *vs.* 82.4% in controls). However, global comparison of haplotype frequencies (by maximum likelihood testing) revealed no significant differences between cases and controls (exact $P = 0.15$), although, as expected given the allele frequencies, the P4-T8 haplotype was more frequent in cases (65.5% *vs.* 56.5% in controls, $P = 0.005$ uncorrected).

We sought to replicate these associations in the trios (the probands from which had not been included in the previous analyses) by seeking evidence for transmission disequilibrium. No suggestion of excess transmission was seen for the P4 or any other allele at the pentanucleotide (global analysis, $\chi^2 = 4.15$, df = 4, $P = 0.39$; P4 transmissions, $\chi^2 = 0.25$, df = 1, $P = 0.62$), for the T8 or any other allele at the tetranucleotide (global analysis, $\chi^2 = 5.26$, df = 5, $P = 0.39$; T8 transmissions, $\chi^2 = 0.007$, df = 1, $P = 0.93$), or for estimates of haplotype transmission (global analysis, $\chi^2 = 4.41$, df = 5, $P = 0.49$; P4-T8 haplotype, $\chi^2 = 0.017$, df = 1, $P = 0.90$).

We also analyzed the relationship between microsatellite genotypes and testosterone levels. In the UK dataset, intermediate trait information was unavailable for the population controls, so these analyses were performed in all PCOS subjects (*i.e.* a combination of the cases and the trio probands) not receiving oral contraception. There was no association between genotype at either locus and testosterone levels (pentanucleotide, $P = 0.98$, Table 3; tetranucleotide, $P = 0.87$). In particular, given previous findings with regard to the P4 allele (13), we compared testosterone levels in subjects with one or more P4 allele [P4/−, equivalent to 216+ in previous papers (13, 15)] with all other subjects and found no appreciable difference [P4/−: geometric mean, 0.61 ng/ml (2.10 nmol/liter); SD range, 0.38–0.96 ng/ml (1.33–3.32 nmol/liter); non-P4: geometric mean, 0.55 ng/ml (1.92 nmol/liter); SD range, 0.37–0.82 ng/ml (1.29–2.85 nmol/liter); $P = 0.29$). Reanalysis after adjusting for subject age had no significant impact on these findings.

In the Finnish cohort, the distribution of pentanucleotide genotypes did not differ significantly between the symptomatic women and controls, whether genotypes were considered globally (exact $P = 0.075$) or grouped according to presence/absence of the P4 allele ($P = 0.55$). There were no differences between the groups in terms of tetranucleotide genotypes ($P = 0.18$), allele counts (pentanucleotide, $P = 0.27$; tetranucleotide, $P = 0.29$) or estimated haplotype frequencies ($P = 0.27$). These findings were unchanged when the case sample was limited to women with USS-confirmed PCO (data not shown). In neither the group of symptomatic women nor the cohort controls was any significant relationship seen between pentanucleotide (controls, $P = 0.86$; symptomatic women, $P = 0.89$) or tetranucleotide genotype (controls, $P = 0.56$; symptomatic women, $P = 0.27$) and testosterone levels (Table 3). Specifically, testosterone levels in P4/− individuals were no different from those in individuals with other genotypes [controls: P4/−, geometric mean, 0.53 ng/ml (1.83 nmol/liter); SD range, 0.35–0.81 ng/ml (1.20–2.79 nmol/liter) *vs.* non-P4, geometric mean, 0.55 ng/ml (1.89 nmol/liter); SD range, 0.38–0.80 ng/ml (1.30–2.77 nmol/liter), $P = 0.39$; symptomatic women: P4/−, geometric mean, 0.59 ng/ml (2.04 nmol/liter), SD range, 0.39–0.89 ng/ml (1.35–3.10 nmol/liter) *vs.* non-P4, geometric mean, 0.61 ng/ml (2.13 nmol/liter); SD range, 0.40–0.94 ng/ml (1.39–3.25 nmol/liter), $P = 0.49$].

Discussion

The main conclusion of our analysis is that we have failed to detect any consistent association between genomic variation at the two *CYP11A* promoter microsatellites and either PCOS status or serum testosterone despite analyses of well-sized data sets. It is worth pointing out that the nominally significant association that we found at the pentanucleotide

TABLE 2. Allele and genotype frequencies in the UK subjects (PCOS cases and population controls only)

Locus	Pentanucleotide	No. of alleles or genotypes (%)		Tetranucleotide	No. of alleles or genotypes	
		PCOS cases	Population controls		PCOS cases	Population controls
Allele counts	P4 (169 bp)[a]	261 (67.6)	391 (58.2)	T6 (276 bp)	1 (0.3)	0
	P6 (179 bp)	68 (17.6)	176 (26.2)	T8 (284 bp)	352 (88.4)	567 (82.4)
	P7 (184 bp)	0	2 (0.3)	T9 (288 bp)	0	9 (1.3)
	P8 (189 bp)	16 (4.2)	43 (6.4)	T10 (292 bp)	3 (0.8)	4 (0.6)
	P9 (194 bp)	40 (10.4)	59 (8.8)	T11 (296 bp)	3 (0.8)	12 (17.4)
	P10 (199 bp)	1 (0.3)	1 (0.2)	T12 (300 bp)	35 (8.8)	82 (11.9)
				T13 (304 bp)	4 (1.0)	14 (2.0)
Genotype counts	P4/P4 (169/169)	91 (47.2)	117 (34.8)	T8/T8 (284/284)	155 (77.9)	233 (67.7)
	P4/P6 (169/179)	47 (24.4)	97 (28.9)	T8/T9 (284/288)	0	6 (1.7)
	P4/P8 (169/189)	9 (4.7)	23 (6.8)	T8/T11 (284/296)	3 (1.5)	9 (2.6)
	P4/P9 (169/194)	23 (11.9)	34 (10.1)	T8/T12 (284/300)	33 (16.6)	72 (20.9)
	P6/P6 (179/179)	4 (2.1)	22 (6.5)	T8/T13 (284/304)	4 (2.0)	12 (3.5)
	P6/P8 (179/189)	4 (2.1)	16 (4.8)	Other genotypes[b]	4 (2.0)	12 (3.5)
	P6/P9 (179/194)	8 (4.1)	19 (5.7)			
	Other genotypes[b]	7 (3.6)	8 (2.4)			

[a] The P5 allele was not seen in the UK cases and controls.
[b] All genotypes less than 1% frequency are pooled for the purposes of this table; contingency table analyses were performed on unpooled counts.

TABLE 3. Pentanucleotide genotypes *vs.* testosterone in the Finnish birth cohort and UK cases

Genotype	Finnish subjects				UK subjects	
	Cohort controls		Symptomatic women		PCOS subjects[a]	
	n[b]	Testosterone (ng/ml)[c]	n[b]	Testosterone (ng/ml)[c]	n[b]	Testosterone (ng/ml)[c]
P4/P4	426	0.52 (0.34–0.80)	195	0.60 (0.40–0.92)	112	0.59 (0.39–0.91)
P4/P6	299	0.54 (0.35–0.83)	150	0.57 (0.37–0.88)	60	0.63 (0.40–1.01)
P4/P8	64	0.53 (0.36–0.76)	50	0.59 (0.40–0.86)	13	0.60 (0.44–0.83)
P4/P9	38	0.52 (0.34–0.81)	13	0.59 (0.45–0.79)	30	0.61 (0.35–1.08)
P6/P6	48	0.53 (0.36–0.80)	29	0.61 (0.40–0.92)	8	0.59 (0.42–0.82)
P6/P8	39	0.54 (0.38–0.76)	16	0.66 (0.44–0.96)	6	0.59 (0.40–0.87)
The rest	30	0.59 (0.40–0.85)	14	0.59 (0.36–0.97)	16	0.53 (0.34–0.81)
Significance tests						
Global	944	$P = 0.86$	467	$P = 0.89$	245	$P = 0.98$
P4/− *vs.* rest	944	$P = 0.39$	467	$P = 0.49$	245	$P = 0.29$

Testosterone values are expressed as geometric mean (SD range).
[a] PCOS subjects includes both case and trio probands.
[b] These analyses only include women not receiving oral contraceptive therapy at the time of sampling.
[c] To convert to SI units (nanomoles per liter) multiply by 3.467.

in UK cases and control subjects in this study (see *Results*) lies in the opposite direction to the previous findings and was, in any event, not substantiated in the analyses of the UK trios or the Finnish subjects.

These findings are clearly in conflict with several previous studies (10, 14, 15) that have appeared to corroborate the initial findings (13) that variation at the *CYP11A* pentanucleotide is both linked to and (with specific reference to the absence of the four-repeat allele) associated with both symptomatic PCOS and raised testosterone levels.

We can discount several possible explanations for these discrepant findings. First, we are confident that our failure to detect association was not due to genotyping error; the low rate of discrepancies observed on cryptic duplicate genotyping and the consistent pattern of linkage disequilibrium relationships between the microsatellites across multiple datasets provide strong reassurance in this regard. Second, we can be confident, given the size of the resources typed, that the failure to detect associations with effect sizes similar to those previously reported was not due to lack of power. Third, the explanation is unlikely to lie in between-cohort differences in linkage disequilibrium relationships (for example, between the pentanucleotide and some as yet unidentified etiological variant); the close similarities in haplotype structure observed between the Finnish and UK data sets suggest that there are no profound differences in linkage disequilibrium structure, at least among populations of European origin (in whom the majority of the previous studies were conducted).

Our data suggest that the strength of, indeed the existence of, any association between *CYP11A* promoter variation and androgen-related phenotypes has been overestimated in previous studies through a combination of small sample size and data exploration conducted without due allowance for the associated inflation of the nominal type 1 error rate. This may have been compounded by the effects of any publication bias that has favored studies generating positive results. These factors are increasingly being recognized as contributing to the inconsistency that has plagued association studies, particularly in the context of multifactorial traits (17–19).

A further explanation for the discrepant findings may lie in the possibility of ethnic stratification (26). In this article, we have restricted the reported analyses to individuals of European origin. However, in PCOS cases of non-European origin, we have found marked differences in microsatellite allele frequencies. For example, in both Afro-Caribbean and South Asian subjects, the P6, rather than the P4, allele was the commonest allele at the pentanucleotide repeat, with frequencies of 46.6% and 48.6%, respectively. This suggests that studies of this locus may be particularly susceptible to spurious associations arising from the failure to match cases and control subjects adequately for ethnic composition.

Concerns over the rather poor performance of association studies have led to promulgation of recommendations designed to improve study design (17–19, 27, 28). Several of these have been incorporated in the current study, including: use of well-sized data sets; attempts to seek replication of any nominally significant findings, particularly those arising from data exploration; use of both case-control and family-based association methods; and analyses based on both intermediate phenotypes and discrete disease categories.

It is important to point out, of course, that the present study does not allow us to conclude that these variants have absolutely no effect on PCOS status or on testosterone levels. However, the size of the study means that any such effect must be extremely small and certainly much smaller than the effect size implied by the previous positive findings. We are currently undertaking functional studies of these microsatellites to determine whether they have any impact on *CYP11A* expression. Equally, the present study does not allow us to exclude the possibility that *CYP11A* variants other than those studied contribute to PCOS susceptibility and population variation in testosterone levels. Extensive resequencing and genotyping will be needed to test this hypothesis. However, as discussed above, it seems unlikely that the previously reported associations at the pentanucleotide repeat were a reflection of linkage disequilibrium with etiological variants elsewhere in the gene. Otherwise, given the apparent similarities in haplotype structure, at least as judged from the relationship between the microsatellites in these two quite distinct European populations, we should have expected to detect the same effects in this study.

The aim of this study was to test whether the previously reported associations with the *CYP11A* pentanucleotide held

up to reanalysis in larger datasets. We conclude that they do not and that the major sites of genomic variation underlying susceptibility to PCOS and population variation in testosterone levels remain to be identified.

Acknowledgments

We acknowledge the many patients, relatives, nurses, and physicians who have contributed to the ascertainment of the various clinical samples used in this study.

Received September 18, 2003. Accepted February 1, 2004.

Address all correspondence and requests for reprints to: Prof. Mark McCarthy, Robert Turner Professor of Diabetes, Oxford Centre for Diabetes, Endocrinology & Metabolism, Churchill Hospital Site, Old Road, Headington, Oxford OX3 7LJ, United Kingdom. E-mail: Mark.mccarthy@drl.ox.ac.uk.

This work was supported by the United Kingdom Medical Research Council (Program Grant G9700120), the Academy of Finland, the European Commission (Framework 5 Award QLG1-CT-2000-01643), and the Wellcome Trust (Project Grant GR069224MA).

References

1. **Adams J, Polson DW, Franks S** 1986 Prevalence of polycystic ovaries in women with anovulation and idiopathic hirsutism. Br Med J 293:355–359
2. **Hull MG** 1987 Epidemiology of infertility and polycystic ovarian disease: endocrinological and demographic studies. Gynecol Endocrinol 1:235–245
3. **Book CB, Dunaif A** 1999 Selective insulin resistance in the polycystic ovary syndrome. J Clin Endocrinol Metab 84:3110–3116
4. **Holte J** 1996 Disturbances in insulin secretion and sensitivity in women with the polycystic ovary syndrome. Baillieres Clin Endocrinol Metab 10:221–247
5. **Franks S** 1989 Polycystic ovary syndrome: a changing perspective. Clin Endocrinol (Oxf) 31:87–120
6. **Franks S, Gharani N, McCarthy M** 2001 Candidate genes in polycystic ovary syndrome. Hum Reprod Update 7:405–410
7. **Dunaif A, Thomas A** 2001 Current concepts in the polycystic ovary syndrome. Annu Rev Med 52:401–419
8. **Gilling-Smith C, Story H, Rogers V, Franks S** 1997 Evidence for a primary abnormality of thecal cell steroidogenesis in the polycystic ovary syndrome. Clin Endocrinol (Oxf) 47:93–99
9. **Jakimiuk AJ, Weitsman SR, Navab A, Magoffin DA** 2001 Luteinizing hormone receptor, steroidogenesis acute regulatory protein, and steroidogenic enzyme messenger ribonucleic acids are overexpressed in thecal and granulosa cells from polycystic ovaries. J Clin Endocrinol Metab 86:1318–1323
10. **Daneshmand S, Weitsman SR, Navab A, Jakimiuk AJ, Magoffin DA** 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
11. **Nelson VL, Legro RS, Strauss III JF, McAllister JM** 1999 Augmented androgen production is a stable steroidogenic phenotype of propagated theca cells from polycystic ovaries. Mol Endocrinol 13:946–957
12. **Wood JR, Nelson VL, Ho C, Jansen E, Wang CY, Urbanek M, McAllister JM, Mosselman S, Strauss JF** 2003 The molecular phenotype of polycystic ovary syndrome (PCOS) theca cells and new candidate PCOS genes defined by microarray analysis. J Biol Chem 278:26380–26390
13. **Gharani N, Waterworth DM, Batty S, White D, Gilling-Smith C, Conway GS, McCarthy M, Franks S, Williamson R** 1997 Association of the steroid synthesis gene CYP11a with polycystic ovary syndrome and hyperandrogenism. Hum Mol Genet 6:397–402
14. **Urbanek M, Legro RS, Driscoll DA, Azziz R, Ehrmann DA, Norman RJ, Strauss JF, Spielman RS, Dunaif A** 1999 Thirty-seven candidate genes for polycystic ovary syndrome: strongest evidence for linkage is with follistatin. Proc Natl Acad Sci USA 96:8573–8578
15. **Diamanti-Kandarakis E, Batzis MI, Bergiele AT, Tsianateli TC, Kouli CR** 2000 Microsatellite polymorphism (tttta) at −528 base pairs of gene CYP11a influences hyperandrogenemia in patients with polycystic ovary syndrome. Fertil Steril 73:735–741
16. **San Millan JL, Sancho J, Calvo RM, Escobar-Morreale HF** 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
17. **Ioannidis JPA, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis JG** 2001 Replication validity of genetic association. Nat Genet 29:306–309
18. **Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN** 2003 Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33:177–182
19. **Colhoun HM, McKeigue PM, Davey Smith G** 2003 Problems of reporting genetic associations with complex outcomes. Lancet 361:865–872
20. **Haddad L, Evans JC, Gharani N, Robertson C, Rush K, Wiltshire S, Frayling TM, Wilkin TJ, Demaine A, Millward A, Hattersley AT, Conway G, Cox NJ, Bell GI, Franks S, McCarthy MI** 2002 Variation within the type 2 diabetes susceptibility gene calpain-10 (CAPN10) and polycystic ovary syndrome. J Clin Endocrinol Metab 87:2606–2610
21. **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
22. **Taponen S, Martikainen H, Jarvelin M-R, Laitinen J, Pouta A, Hartikainen A-L, Sovio U, McCarthy MI, Franks S, Ruokonen A** 2003 The hormonal profile of women with self-reported symptoms of oligomenorrhea and/or hirsutism: Northern Finland Birth Cohort 1966 Study. J Clin Endocrinol Metab 88:141–147
23. **Fahle GA, Fischer SH** 2000 Comparison of six commercial DNA extraction kits for recovery of cytomegalovirus DNA from spiked human specimens. J Clin Microbiol 38:3860–3863
24. **Abecasis GR, Cookson WO** 2000 GOLD—graphical overview of linkage disequilibrium. Bioinformatics 16:182–183
25. **Clayton D** 1999 A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. Am J Hum Genet 65:1170–1177
26. **Spielman RS, Ewens WJ** 1996 The TDT and other family-based tests for linkage disequilibrium and association. Am J Hum Genet 59:983–989
27. 1999 Freely associating. Nat Genet 22:1–2
28. **Cardon LR, Bell JI** 2001 Association study designs for complex disease. Nat Rev Genet 2:91–99

JCEM is published monthly by The Endocrine Society (http://www.endo-society.org), the foremost professional society serving the endocrine community.