Common variants in the *CYP2C19* gene are associated with susceptibility to endometriosis

Jodie N. Painter, Ph.D.,[a] Dale R. Nyholt, Ph.D.,[a] Lutz Krause, Ph.D.,[a] Zhen Z. Zhao, Ph.D.,[a] Brett Chapman, B.Sc.,[a] Christine Zhang, B.Bio.Med.Sc.(Hons),[b] Sarah Medland, Ph.D.,[a] Nicholas G. Martin, Ph.D.,[a] Stephen Kennedy, Ph.D.,[c] Susan Treloar, Ph.D.,[d] Krina Zondervan, Ph.D.,[e] and Grant W. Montgomery, Ph.D.[a]

[a] QIMR Berghofer Medical Research Institute and [b] Mater Medical Research Institute, Brisbane, Queensland, Australia; [c] Nuffield Department of Obstetrics and Gynaecology, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom; [d] Centre for Military and Veterans' Health, University of Queensland, Mayne Medical School, Queensland, Australia; and [e] Genetic and Genomic Epidemiology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

Objective: To follow-up previous studies highlighting a possible role for cytochrome P450, family 2, subfamily C, 19 (CYP2C19) in susceptibility to endometriosis by searching for additional variants in the *CYP2C19* gene that may be associated with the disease.
Design: Case-control study.
Setting: Academic research.
Subject(s): The cases comprised 2,271 women with surgically confirmed endometriosis; the controls comprised 939 women with self-report of no endometriosis and 1,770 unscreened population samples.
Intervention(s): Sequencing of the *CYP2C19* region and follow-up of 80 single nucleotide polymorphisms (SNPs) in two case-control samples.
Main Outcome Measure(s): Allele frequency differences between cases and controls.
Result(s): Sequencing of the *CYP2C19* gene region resulted in the detection of a large number of known and novel SNPs. Genotyping of 80 polymorphic SNPs in 901 endometriosis cases and 939 controls resulted in study-wide significant association signals for SNPs in moderate or complete linkage disequilibrium with rs4244285, a functional SNP in exon 5 that abrogates *CYP2C19* function through the creation of an alternative splice site. Evidence of association was also detected for another functional SNP in the *CYP2C19* promoter, rs12248560, which was highlighted in our previous study.
Conclusion(s): Functional variants in *CYP2C19* may contribute to endometriosis susceptibility in both familial and sporadic cases. (Fertil Steril® 2014;102:496–502. ©2014 by American Society for Reproductive Medicine.)
Key Words: Endometriosis, association, pooled sequencing, *CYP2C19*, rs12248560, rs4244285

Discuss: You can discuss this article with its authors and with other ASRM members at http://fertstertforum.com/painterj-variants-cyp2c19-gene-endometriosis/

Use your smartphone to scan this QR code and connect to the discussion forum for this article now.*

* Download a free QR code scanner by searching for "QR scanner" in your smartphone's app store or app marketplace.

Endometriosis is a gynecological disease that affects 6%–10% of women of reproductive age and causes a variety of symptoms including severe menstrual pain, chronic pelvic pain, and subfertility (1). The disease is inherited as a complex trait (1–3), with up to 52% of the variation in liability accounted for by genetic factors (4). Recent genome-wide association studies (GWAS) have revealed a number of common genetic variants associated with susceptibility to endometriosis (5–8). Both linkage and association studies take advantage of linkage disequilibrium (LD), where the alleles of single nucleotide polymorphisms (SNPs) located within short distances of each other are correlated. As has been the case in other complex diseases, the causal variants underlying endometriosis risk may be uncommon or novel SNPs in LD with the common markers genotyped in the original studies. Such SNPs can be detected by sequencing of regions surrounding association and linkage signals and

Received October 16, 2013; revised and accepted April 10, 2014; published online May 3, 2014.
J.N.P. has nothing to disclose. D.R.N. has nothing to disclose. L.K. has nothing to disclose. Z.Z.Z. has nothing to disclose. B.C. has nothing to disclose. C.Z. has nothing to disclose. S.M. has nothing to disclose. N.G.M. has nothing to disclose. S.K. has nothing to disclose. S.T. has nothing to disclose. K.Z. reports consulting fees from Bayer HealthCare LTD and AbbVie LTD. G.W.M. has nothing to disclose.
Reprint requests: Jodie N. Painter, Ph.D., Molecular Cancer Epidemiology Laboratory, QIMR Berghofer Medical Research Institute, 300 Herston Road, Herston, QLD 4006, Australia (E-mail: jodie.painter@qimrberghofer.edu.au).

Fertility and Sterility® Vol. 102, No. 2, August 2014 0015-0282/$36.00

http://dx.doi.org/10.1016/j.fertnstert.2014.04.015

may contribute more to disease risk in the population than the common variants originally detected (9).

We recently fine-mapped a linkage peak for endometriosis on chromosome 10 (10), finding evidence suggestive of association with the cytochrome P450 family 2, subfamily C, 19 (*CYP2C19*) gene (11). The CYP2C19 protein participates in the metabolism of estrogen (12, 13) and up to 10% of clinically administered medications (14) including the anti-estrogenic drug tamoxifen (15). The gene has a number of functional variants that influence drug metabolism, one of which, rs12248560, we found to be associated with endometriosis risk (11). As this common SNP (minor allele frequency [MAF] = 0.21) did not fully account for the linkage signal, it is possible that additional SNPs contribute to the endometriosis susceptibility associated with this chromosomal region. Interestingly, we also found nominal association with a SNP independent of rs12248560, rs4244285 (LD estimate $r^2 = 0.04$), a loss-of-function variant in *CYP2C19* exon 5 that had been previously associated with endometriosis risk in a small candidate gene study (16), although this result was not replicated in an even smaller sample (17).

The previously detected fine-mapping association signal extended across 252.3 kilobases (kb) of chromosome 10 from the 3' end of the helicase, lymphoid-specific (*HELLS*) gene to the 3' end of *CYP2C19* (11). This region includes another *CYP2C* gene family member, *CYP2C18*. In the current study, we aimed to search for additional variants in this region that may contribute to endometriosis risk. We screened endometriosis cases for such risk variants using two strategies. First, to search for variants that may have a direct effect on gene activity we screened the 5' and 3' untranslated regions (UTRs), exons and intron-exon boundaries of *CYP2C18* and *CYP2C19* in 20 unrelated endometriosis cases with a strong family history of disease. Next, to broaden the search to include intergenic and intronic regions that may harbor variants with regulatory effects on gene activity, we sequenced the entire 252.3 kb association region in DNA pools of 384 endometriosis cases and 384 controls. Case-control association analyses were then performed on a subset of the variants we detected by sequencing and provided further evidence that functional SNPs in *CYP2C19* may contribute to endometriosis risk.

MATERIALS AND METHODS

Samples

Case samples were taken from the set of 3,908 endometriosis cases with and without a family history of disease recruited by the QIMR Berghofer Medical Research Institute between 1995 and 2002 (5). All women had completed a questionnaire and provided a blood sample. A surgical diagnosis was confirmed by retrospective examination of medical records; disease severity was determined using the revised American Fertility Society (rAFS) classification system (18). Control samples comprised unrelated individuals originally recruited through QIMR Berghofer for either a twin study of gynecological health who self-reported no endometriosis (discovery sample) (4) or the sample previously used for our endometriosis GWAS (replication sample) (5) recruited through the Brisbane Adolescent Twin Study (19, 20). Approval for this study and to obtain medical records, for collection of blood for DNA extraction and all questionnaires and interview schedules and for the inclusion of twin individuals recruited through the Australian Twin Registry, was obtained from the QIMR Berghofer Human Research Ethics Committee. All participants gave written informed consent.

Familial Sequencing

DNA samples from 20 unrelated women from 20 of the most case-dense endometriosis families included in our linkage study (10) were subjected to Sanger sequencing to search for potentially high-risk rare sequence variants contributing to endometriosis risk. Each family had three or more affected first- or second-degree relatives, with the sample chosen for sequencing having either the more severe disease (according to rAFS stages I–IV) (18) or the earliest age of onset if affected family members had been assigned the same stage.

Primers for the polymerase chain reaction (PCR) amplification and sequencing of the promoter and 3' UTR regions, intron-exon boundaries, and all exons of the *CYP2C18* and *CYP2C19* gene were designed using Primer 3.0 (21) (Supplemental Table 1). PCRs were performed in 15 μL reactions containing 1× PCR buffer, 1.5 mM $MgCl_2$, 1 U Amplitaq Gold (all Applied Biosystems), 200 μM each deoxynucleotide triphosphate (Promega), and 1 μM each of forward and reverse primers and 50 ng of DNA. PCR cycling conditions included initial denaturation at 95°C for 5 minutes; touchdown cycling at 95°C for 30 seconds, 60°C (−0.5 per cycle) for 30 seconds, and 72°C for 30 seconds for 20 cycles; 95°C for 30 seconds, 50°C for 30 seconds, and 72°C for 30 seconds for 15 cycles; and final extension 72°C for 10 minutes. Products were verified by electrophoresis through 2% agarose gels, cleaned with Exonuclease I and Shrimp Alkaline Phosphotase (Fermentas), and sequenced using BigDye 3.0 terminator chemistry (Applied Biosystems).

Pooled Sequencing of the *HELLS-CYP2C19* Region

DNA samples had previously been combined into two pools comprising 384 endometriosis cases (case pool) and 384 unrelated gynecological health study samples (control pool) (22). Targeted resequencing and analysis of the case and control DNA pools were performed by deCODE Scientific Services. Briefly, for each pool, DNA within the ~252.3-kb region extending from the 3' end of *HELLS* to the 3' end of *CYP2C19* (covering chromosome 10 bases 96,361,710–96,704,010 [genome build GRCh37/hg19]) was amplified in sections by long-range PCR (average length, 8,021 bases) and sequenced on an Illumina GAIIx platform (Illumina). Short read sequences were then aligned, and sequence variants (SVs; termed as such as many loci were yet to be verified) were called by deCODE using the Illumina programs CASAVA and Elandv2. In addition, we reanalyzed the sequences using programs tailored for DNA pools: sequences were aligned using the program Novoalign (http://www.novocraft.com/main/index.php), and SVs were called using CRISP (23).

Individual Genotyping and Association Analyses

Genotyping of SVs detected by individual and pooled sequencing was conducted in a discovery sample of 958 endometriosis cases and 959 unrelated gynecological health study controls examined in previous genetic association studies for endometriosis conducted at QIMR Berghofer (24–29). This sample includes all 768 individuals (384 cases and 384 controls) included in the DNA pools described above. SVs were genotyped in multiplex assays using the Sequenom MassARRAY Genomics Platform (Sequenom). We selected 197 SVs for individual genotyping including two SNPs detected only by familial-case sequencing, 60 SVs called by deCODE only, 60 SVs called by QIMR only, and 75 SVs called by both analyses. SVs detected by targeted sequencing were included according to the following criteria: [1] nominally significant ($P<.05$) allele frequency differences between the case and control pools (83 SVs across the three groups), and [2] location in potentially functional areas (e.g., all exonic SVs in addition to SVs located in 5' and 3' UTRs, regions of conserved sequence or within putative transcription factor binding sites) as determined by the program ANNOVAR (112 SVs) (30).

Genotyping quality control was performed for each multiplex separately, where SVs with >5% missing genotypes and Hardy-Weinberg P values $<1 \times 10^{-4}$ were excluded from further analyses. In addition, individuals missing >5% of data, or for whom we have recent genetic evidence of non-Caucasian ancestry detected during our GWAS analysis (5), were also removed. Single SNP and haplotype association analyses were performed including 901 of the 958 endometriosis cases and 939 of the 959 gynecological health study controls passing all quality control metrics using PLINK (31). Study-wide significance at the discovery phase ($P_{adjusted}<.05$) was determined by 10,000 permutations of the data set, where case-control status was randomly swapped to break the genotype-phenotype relationship while preserving the LD structure between the SNPs, to generate empirical significance levels corrected for multiple testing resulting from the large number of SNPs included in the study.

Replication

As only Australian samples were genotyped in the discovery phase, we performed in silico replication on Australian cases and controls only, using a sample comprising 1,370 unrelated endometriosis cases and 1,770 unrelated Brisbane Adolescent Twin study controls drawn from our recent GWAS for endometriosis (5). Briefly, all Australian GWAS individuals not included as, or related to, an individual from the 958 case sample were used as cases or controls for the replication sample. As the best SNPs from the discovery phase analysis were not present on the Illumina 610K genotyping chips used to genotype the GWAS samples, we analyzed genotype data and dosage scores for SNPs previously imputed to HapMap2 (32). All imputed SNPs had genotype concordance values between 97.1% and 99.7% for 560 endometriosis cases for whom both actual and imputed genotypes were available, with imputation quality scores >0.9, indicating that these SNPs were imputed with a high degree of accuracy.

Association analyses were performed in the replication data set for three SNPs, rs1326837, rs4244285, and rs12248560, as described above. Since only associations in the same direction as in the discovery sample can be considered to be replicated, one-sided P values were obtained by halving the standard (two-sided) P values (8). Unadjusted results obtained for both the discovery and replication sets were then included in a meta-analysis performed using METAL (8, 33), which converts P values to Z-scores that are then weighted by the square root of the sample size for each sample set, to determine the total evidence of association for each SNP.

RESULTS

Sequencing Familial Endometriosis Cases

Sequencing of 20 cases from 20 unrelated case-dense endometriosis families for both the *CYP2C18* and *CYP2C19* genes revealed the presence of a number of known SNPs in addition to two novel intronic variants, one in intron 7 of *CYP2C18* (chr10:96492949) and the second in intron 7 of *CYP2C19* (chr10:96602941). As both SNPs were located at least 50 bases from the respective intron-exon boundaries, neither were predicted to have an effect on gene splicing (ESEfinder3.0: http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home). A number of affected women carried the known functional *CYP2C19* SNPs rs12248560 (*CYP2C19*17*, see http://www.cypalleles.ki.se/cyp2c19.htm for allele nomenclature) in the promoter region (associated with endometriosis risk in our previous fine-mapping study; (11)), rs17878459 (p.Glu92Asp: *CYP2C19*3*) in exon 2, and rs4244285 (p.Pro271Pro: *CYP2C19*2*) in exon 5.

Sequencing DNA Pools of Endometriosis Cases and Controls

Sequencing of the 252.3 kb between the end of the *HELLS* gene to the end of *CYP2C19* in our two DNA pools of 384 endometriosis cases and 384 controls resulted in adequate sequence data (>500-fold coverage) for 96.8% and 94.0% of the region in each pool, respectively. Average sequence coverage was 1,261-fold for the case pool and 1,372-fold for the control pool. The largest gap of missing sequence (~20 kb) was in the region of 96.38–96.40 Mb for both pools (Supplemental Fig. 1).

DeCODE Genetics called 3,051 SVs using their analysis pipeline, of which 2,418 (79%) were novel (not present in

TABLE 1

Average MAFs in the endometriosis case and control DNA pools by SNP calling method.

MAF range	deCODE (%)	QIMR (%)
<0.001	26 (0.9)	58 (0.04)
0.001–0.01	2281 (74.7)	651 (45)
0.01–0.05	441 (14.5)	323 (22.3)
0.05–0.1	109 (3.5)	152 (10.5)
0.1–0.5	194 (6.4)	263 (18.2)

Painter. CYP2C19 and endometriosis risk. Fertil Steril 2014.

TABLE 2

Association signal to endometriosis across the region between the *HELLS* and *CYP2C19* genes on chromosome 10q26.

				Fine-mapping			Replication		Meta-analysis	
LD block	SNP	Location	MAF Ca/Co	*P* value	$P_{adjusted}$	OR (95% CIs)	*P* value	OR (95% CIs)	*P* value	OR (95% CIs)
1	rs1326837*	*HELLS-CYP2C18*	0.25/0.19	4.6×10^{-5}	1.2×10^{-3}	1.38 (1.18–1.61)	3.1×10^{-3}	1.19 (1.07–1.30)	1.3×10^{-6}	1.26 (1.07–1.42)
1	rs4244285	*CYP2C19* exon 5	0.17/0.13	9.7×10^{-4}	2.5×10^{-2}	1.35 (1.12–1.61)	8.0×10^{-3}	1.19 (1.05–1.33)	9.4×10^{-5}	1.25 (1.06–1.41)
2	rs12248560*	*CYP2C19* promoter	0.80/0.76	6.4×10^{-3}	1.6×10^{-1}	0.80 (0.69–0.94)	2.5×10^{-2}	0.88 (0.78–0.97)	5.8×10^{-4}	0.85 (0.75–0.95)

Note: Signal across the region is accounted for by SNPs in two blocks of LD, with the best SNP in each block indicated by asterisk (*). The SNP rs4244285 is included as a known functional SNP in moderate LD ($r^2 = 0.64$) with rs1326837. CI = confidence interval.

Painter. CYP2C19 and endometriosis risk. Fertil Steril 2014.

SNP databases such as dbSNP, HapMap or the July 2010 [Phase2] release of the 1000 Genomes data). Most of the SVs were rare: 2,319 (76%) had pool MAFs of <1% (range, <0.0001–0.499; Table 1). At QIMR we called 1,447 SVs using the CRISP program, of which 958 (66%) were novel and 706 (49%) had pool MAFs <1% (range <0.0001–0.497; Table 1). A total of 629 SVs were called by both deCODE and QIMR using the two variant-calling approaches. Of these, 148 (23.5%) were novel and 158 (25%) had pool MAFs <1% (range, 0.0004–0.50).

Sequence Variant Polymorphism in Individual Genotyping

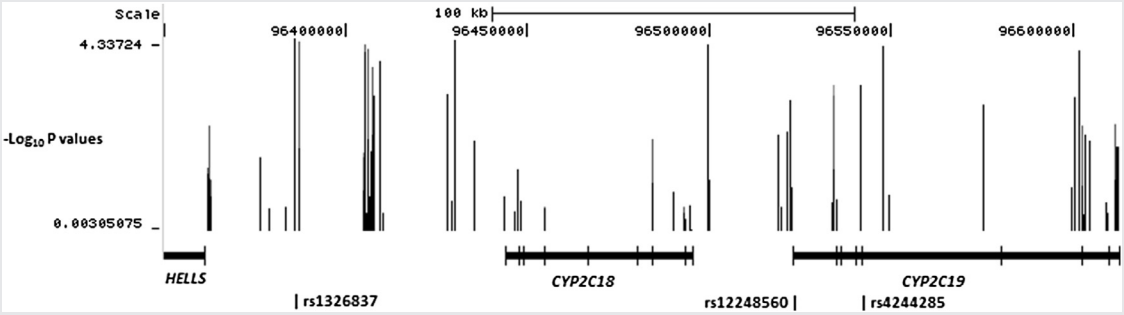
We selected 197 SVs for individual genotyping (of which 63 were known SNPs), including two SVs detected only by familial-case sequencing, 60 SVs called by deCODE only, 60 SVs called by QIMR only, and 75 SVs called by both analyses. Genotyping of the 197 SVs in the discovery sample revealed that 119 were in fact monomorphic (e.g., no alternative allele was detected) in the 384 case and 384 control samples used to make up the DNA pools, although two of these SVs were polymorphic in the larger sample of 901 cases and 939 controls that passed our quality control. Neither genomic location (intergenic or intronic vs. exonic, potentially functional or conserved areas) nor MAF in the pools were reliable predictors of whether SVs would be polymorphic or monomorphic when individually genotyped: SVs monomorphic in the individual genotyping had pool MAFs ranging from 0.001 to 0.40 in either the case or control pools, while SVs polymorphic in individual genotyping had pool MAFs ranging from 0.001 to 0.17.

SVs were instead more likely to be polymorphic if they were known SNPs (i.e., they had previously been assigned an "rs" number) and had been called by both variant-calling programs used in this study. Only 17/134 (12.7%) novel SVs were polymorphic in individual genotyping. Of the 63 known SNPs, 61 (97%) were polymorphic when individually genotyped, although these had pool MAFs within the range of the SVs shown to be monomorphic upon individual genotyping (0.001–0.5).

Fine-mapping Across the *CYP2C19* Region

After quality control, 80 SNPs were polymorphic in the discovery sample of 901 endometriosis cases and 939 controls included in the association analysis. A nominal association signal ($P \leq .05$) was detected for 35 SNPs, 19 of which remained significant ($P_{adjusted} \leq .05$) after permutation to correct for multiple testing (Table 2; Supplemental Table 2). Association signal was distributed across the sequenced region (Fig. 1), with the best signal seen for rs1326837 ($P=4.6 \times 10^{-5}$, odds ratio [OR] = 1.38; $P_{adjusted}=1.2 \times 10^{-3}$). All 19 significant SNPs were in moderate to high LD with each other, with analysis conditioning on rs1326837 removing all signal from SNPs in high LD ($r^2 > 0.8$) and reducing but not eliminating the significance of SNPs in moderate LD ($r^2 = 0.5$–0.8), indicating that SNPs in both high and moderate LD with rs1326837 contribute to the endometriosis risk conferred by this haploptype.

FIGURE 1



Association signal across the 252.3-kb region between the *HELLS* and *CYP2C19* genes on chromosome 10. The genomic positions of the top SNP rs1326837 and the two functional SNPs rs4244285 and rs12248560 are indicated by dashes next to the SNP names.

Painter. CYP2C19 *and endometriosis risk. Fertil Steril 2014.*

The corresponding *P* value for the 19-SNP haplotype was 8.19×10^{-5}, and there was no evidence of an independent effect for rs1326837. These 19 SNPs include rs4244285, representing the subgroup of SNPs in moderate LD with rs1326837, which achieved an unadjusted $P=9.2 \times 10^{-4}$ ($P_{adjusted}=2.5 \times 10^{-2}$) and retained an unadjusted $P=2.7 \times 10^{-2}$ in the conditional analyses.

The association signal for both rs1326837 and rs4244285 was replicated in the independent data set of 1,370 mostly nonfamilial endometriosis cases and 1,770 population controls, with a one-sided *P* value for rs1326837 of 3.1×10^{-3} (OR = 1.19) and for rs4244285 of $P=8.0 \times 10^{-3}$ (OR = 1.18). Meta-analysis of the discovery and replication data sets produced a combined *P* value for rs1326837 of 1.3×10^{-6} and for rs4244285 of 9.4×10^{-5}, surpassing our threshold of total study-wide significance calculated as 6.2×10^{-4} (where $P=.05/80$ SNPs).

Evidence of association was also detected for rs12248560 ($P=6.4 \times 10^{-3}$, OR = 0.80), the functional promoter SNP highlighted in our previous fine-mapping study (Table 2; Supplemental Table 2). LD between rs12248560 and SNPs of the rs1326837 haplotype is low (average $r^2 = 0.04$), hence the association signal was not affected by the conditional analysis described above, indicating that this SNP represents an independent effect on endometriosis risk. However, rs12248560 was no longer significant after permutation of the 901 case and 939 controls data set, indicating the signal for this SNP was not study-wide significant in the discovery set ($P_{adjusted}=.16$). However, there was evidence of association in the replication sample (one-sided $P=2.5 \times 10^{-2}$; OR = 0.88), and meta-analysis of the unadjusted discovery and replication data sets produced a combined *P* value for rs12248560 of 5.8×10^{-4}, indicating that this SNP is significant over the total study sample.

DISCUSSION

We have found genetic variants in the *CYP2C19* gene to be associated with susceptibility to endometriosis among women with and without a family history of the disease. Previously, we found the rs12248560 functional promoter SNP to be associated with a decreased risk of endometriosis after our initial fine-mapping of a linkage peak on chromosome 10 (11). While it is study-wide significant, the rs12248560 signal is somewhat reduced in the current study (fine-mapping sample minor allele $P=4.9 \times 10^{-4}$; OR = 0.78 [11] vs. current discovery sample $P=6.4 \times 10^{-3}$, OR = 0.80); this appears to be mostly due to the smaller sample size as only Australian cases and controls were included here. Although the effect of rs12248560 on gene transcription requires clarification (*CYP2C19*17* was initially suggested to cause an ultrarapid drug metabolizer phenotype, but this has recently been questioned; [34]), this SNP has also been associated with a decreased risk of breast cancer, particularly in women treated with hormone replacement therapy for 10 years or more (35), indicating that functional *CYP2C19* variants can influence the risk of estrogen-dependent conditions.

The finding of additional independent association signals provides evidence for the possibility that multiple variants contributed to the original linkage peak in this region (10). Of particular interest is the top SNP from this study, rs1326837. Located in the intergenic region between *HELLS* and *CYP2C18*, there is currently no evidence for a regulatory role for this particular SNP. However, previous GWAS to find variants influencing pharmacological drug response have detected SNPs in this intergenic region (e.g., rs12772169 and rs12777823) that are in complete and moderate LD with rs1326837 and are also linked to functional SNPs in various *CYP2C* genes (36, 37). Such SNPs may be contributing to the risk of endometriosis detected here.

SNPs in moderate LD with rs1326837 include the functional SNP rs4244285 ($r^2 = 0.64$). This synonymous SNP in *CYP2C19* exon 5 causes no change to the amino acid (p.Pro227Pro) but creates an alternative splice site 40 bases downstream, resulting in a truncated, nonfunctional protein (38) and a poor metabolizer phenotype (*CYP2C19*2*). In our previous fine-mapping study including cases from Australia and the United Kingdom, we found only nominal association with rs4244285 (OR = 1.23, $P=1.1 \times 10^{-2}$) in 1,158 familial and sporadic cases (11), indicating that association with this SNP was not driving the association signal detected in that sample. Given that the signal for rs4244285 is more significant in the current study (discovery sample OR = 1.35, $P=9.7 \times 10^{-4}$; meta-analysis OR = 1.35, $P=9.4 \times 10^{-5}$),

despite the smaller sample size, this may indicate that the effect of this SNP on endometriosis risk is larger in the Australian than in the United Kingdom population.

CYP2C19 participates in the conversion of 17β-estradiol (E_2) to estrone (E_1) (12) and in the production of 2α-hydroxy estrogen (2α-OHE) (12, 13). This gene may exert its disease risk altering-effects through two mechanisms, either independently or in concert. That SNPs causing the increased metabolizer phenotypes (rs12248560) are protective, while SNPs causing the poor metabolizer phenotype (rs4244285) increase the risk of disease, supports the hypothesis that *CYP2C19* SNPs may alter endometriosis risk through an effect on local and/or central estrogen metabolism. Altered CYP2C19 activity may affect localized, tissue-specific rates of conversion of E_2 to E_1 and/or result in altered levels of the antiestrogenic 2α-OHE metabolite, with activity either increased or decreased as compared with the wild-type isoform depending on the true causal variant(s). Evidence for the role of other genes involved in sex-steroid metabolism in endometriosis is currently equivocal at best, owing mostly to the small sample sizes that are typically analyzed in candidate gene studies (39, 40).

Alternatively, CYP2C19 may affect endometriosis risk through altered metabolism of exogenous substances, which may increase the risk of developing this disease (41, 42). *CYP2C19* expression and activity is altered in response to xenobiotic as well as steroid hormone exposure (43, 44). As seen for drug metabolism, it may be that *CYP2C19* SNPs influence the rates at which other substances metabolized by this gene are cleared, either by the liver or locally at extrahepatic sites, and this may moderate endometriosis risk. As the use of plastics and associated chemicals has increased over the past decades (45), we examined effect sizes for rs12248560 and rs4244285 by period of endometriosis diagnosis among our cases (pre- and from 1994, the median year of diagnosis in our sample), and while larger effects were seen for both SNPs for cases diagnosed from 1994, these results were not significantly different (Supplemental Table 3).

In this study we sought to detect rare and possibly novel sequence variants associated with an increased risk of developing endometriosis through DNA sequencing of women with a family history and subsequent association analyses including case samples with and without a family history of disease. Instead, the best association signals detected were for common variants linked to known functional SNPs in the *CYP2C19* gene. The effect sizes of these *CYP2C19* variants are small but similar to variants associated with endometriosis in recent GWAS conducted by us (5) and others (6, 7). Although our current results are not significant at a genome-wide level, they are significant at a study-wide level and replicated in a large independent sample. Further studies to validate our findings in additional case and control data sets or functional experiments such as determining the association of SNPs with gene expression in relevant tissues will help to clarify the role of *CYP2C19* in estrogen-dependent diseases. *CYP2C19* is an important candidate owing to its role in the metabolism of estrogen and possibly also other substances that might be associated with the risk of developing endometriosis, and our study indicates that multiple SNPs in this gene region likely contribute to this risk.

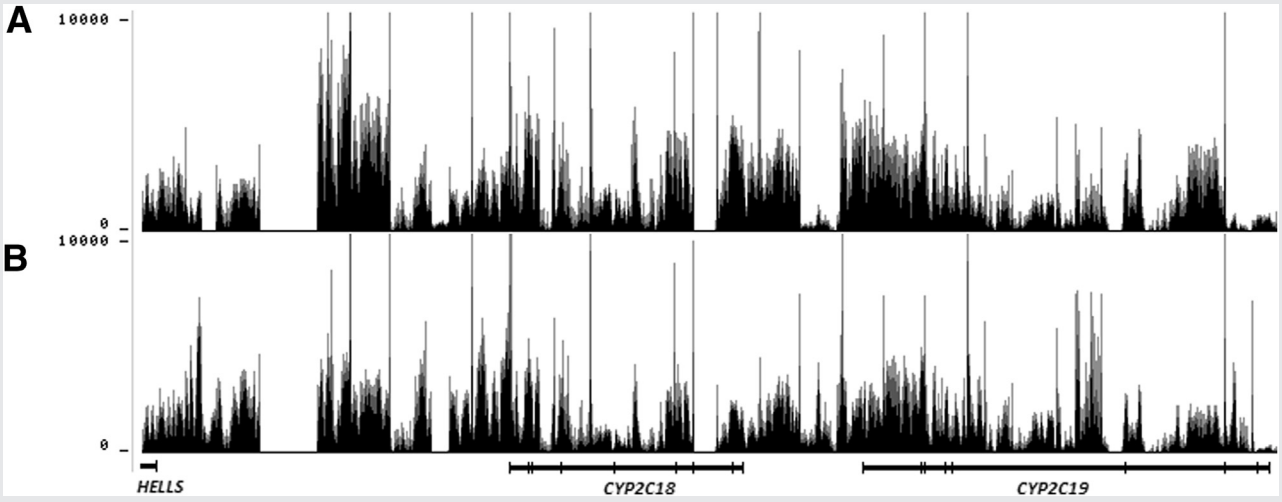
Acknowledgments: We thank the women who participated in the QIMR Berghofer Medical Research Institute study and Endometriosis Associations for supporting recruitment. The study was supported by National Health and Medical Research Council (NHMRC) of Australia (grant nos. 241944, 339462, 389927,389875, 389891, 389892, 389939, 443036, 442915, 442981, 496610, 496739, 552485, 552498), the Cooperative Research Centre for Discovery of Genes for Common Human Diseases, Cerylid Biosciences (Melbourne), and donations from Neville and Shirley Hawkins, the Endometriosis Associations of Queensland and Western Australia, and the family of the late Kim Goodwin. The fine-mapping data on which this study is based were generated by funding from the National Institutes of Health, grant no. R01HD50537. The GWAS data were generated as part of a study supported by the Wellcome Trust (WT084766/Z/08/Z). DRN is supported by the NHMRC Fellowship (339462 and 613674) and ARC Future Fellowship (FT0991022), KTZ by a Wellcome Trust Research Career Development Fellowship (WT085235/Z/08/Z), and GWM by the NHMRC Fellowships Scheme (339446, 619667). We thank B. Haddon, D. Smyth, H. Beeby, O. Zheng, A. Henders, and additional research assistants and interviewers for database management and sample processing and Brisbane gynecologist Dr. Daniel T. O'Connor for confirmation of diagnosis and staging of disease for some Australian patients. We also thank the many hospital directors and staff, gynecologists, general practitioners, and pathology services in Australia who provided assistance with confirmation of diagnoses and blood collection and Sullivan Nicolaides and Queensland Medical Laboratory for pro bono collection and delivery of blood samples.

REFERENCES

1. Giudice LC. Endometriosis. N Engl J Med 2010;362:2389–98.
2. Simpson JL, Bischoff FZ. Heritability and molecular genetic studies of endometriosis. Ann N Y Acad Sci 2002;955:239–51.
3. Montgomery GW, Nyholt DR, Zhao ZZ, Treloar SA, Painter JN, Missmer SA, et al. The search for genes contributing to endometriosis risk. Hum Reprod Update 2008;14:447–57.
4. Treloar SA, Do KA, O'Connor VM, O'Connor DT, Yeo MA, Martin NG. Predictors of hysterectomy: an Australian study. Am J Obstet Gynecol 1999; 180:945–54.
5. Painter JN, Anderson CA, Nyholt DR, Macgregor S, Lin JH, Lee SH, et al. Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. Nat Genet 2011;43:51–4.
6. Uno S, Zembutsu H, Hirasawa A, Takahashi A, Kubo M, Akahane T, et al. A genome. Nat Genet 2012;42:707–10.
7. Adachi S, Tajima A, Quan J, Haino K, Yoshihara K, Masuzaki H, et al. Meta-analysis of genome-wide association scans for genetic susceptibility to endometriosis in Japanese population. J Hum Genet 2010;55:816–21.
8. Nyholt DR, Low SK, Anderson CA, Painter JN, Uno S, Morris AP, et al. Genome-wide association meta-analysis identifies new endometriosis risk loci. Nat Genet 2012;44:1355–9.
9. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 2009;324:387–9.
10. Treloar SA, Wicks J, Nyholt DR, Montgomery GW, Bahlo M, Smith V, et al. Genomewide linkage study in 1,176 affected sister pair families identifies

a significant susceptibility locus for endometriosis on chromosome 10q26. Am J Hum Genet 2005;77:365–76.
11. Painter JN, Nyholt DR, Morris A, Zhao ZZ, Henders AK, Lambert A, et al. High-density fine-mapping of a chromosome 10q26 linkage peak suggests association between endometriosis and variants close to CYP2C19. Fertil Steril 2001;95:2236–40.
12. Lee AJ, Cai MX, Thomas PE, Conney AH, Zhu BT. Characterization of the oxidative metabolites of 17beta-estradiol and estrone formed by 15 selectively expressed human cytochrome p450 isoforms. Endocrinology 2003; 144:3382–98.
13. Cribb AE, Knight MJ, Dryer D, Guernsey J, Hender K, Tesch M, et al. Role of polymorphic human cytochrome P450 enzymes in estrone oxidation. Cancer Epidemiol Biomarkers Prev 2006;15:551–8.
14. Desta Z, Zhao X, Shin JG, Flockhart DA. Clinical significance of the cytochrome P450 2C19 genetic polymorphism. Clin Pharmacokinet 2002;41: 913–58.
15. van Schaik RH, Kok M, Sweep FC, van Vliet M, van Fessem M, Meijer-van Gelder ME, et al. The CYP2C19*2 genotype predicts tamoxifen treatment outcome in advanced breast cancer patients. Pharmacogenomics 2001; 12:1137–46.
16. Cayan F, Ayaz L, Aban M, Dilek S, Gumus LT. Role of CYP2C19 polymorphisms in patients with endometriosis. Gynecol Endocrinol 2009;25:530–5.
17. Bozdag G, Alp A, Saribas Z, Tuncer S, Aksu T, Gurgan T. CYP17 and CYP2C19 gene polymorphisms in patients with endometriosis. Reprod Biomed Online 2010;20:286–90.
18. Revised American Fertility Society classification of endometriosis: 1985. Fertil Steril 1985;43:351–2.
19. McGregor B, Pfitzner J, Zhu G, Grace M, Eldridge A, Pearson J, et al. Genetic and environmental contributions to size, color, shape, and other characteristics of melanocytic naevi in a sample of adolescent twins. Genet Epidemiol 1999;16:40–53.
20. Zhu G, Duffy DL, Eldridge A, Grace M, Mayne C, O'Gorman L, et al. A major quantitative-trait locus for mole density is linked to the familial melanoma gene CDKN2A: a maximum-likelihood combined linkage and association analysis in twins and their sibs. Am J Hum Genet 1999;65:483–92.
21. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 2000;132:365–86.
22. Zhao ZZ, Nyholt DR, James MR, Mayne R, Treloar SA, Montgomery GW. A comparison of DNA pools constructed following whole genome amplification for two-stage SNP genotyping designs. Twin Res Hum Genet 2005; 8:353–61.
23. Bansal V. A statistical method for detection of variants from next-generation re-sequencing of DNA pools. Bioinformatics 2010;26:i318–24.
24. Treloar SA, Zhao ZZ, Le L, Zondervan KT, Martin NG, Kennedy S, et al. Variants in EMX2 and PTEN do not contribute to risk of endometriosis. Mol Hum Reprod 2007;13:587–94.
25. Zhao ZZ, Nyholt DR, Le L, Martin NG, James MR, Treloar SA, et al. KRAS variation and risk of endometriosis. Mol Hum Reprod 2006;12:671–6.
26. Zhao ZZ, Nyholt DR, Le L, Thomas S, Engwerda C, Randall L, et al. Genetic variation in tumour necrosis factor and lymphotoxin is not associated with endometriosis in an Australian sample. Hum Reprod 2007; 22:2389–97.
27. Zhao ZZ, Pollock PM, Thomas S, Treloar SA, Nyholt DR, Montgomery GW. Common variation in the fibroblast growth factor receptor 2 gene is not associated with endometriosis risk. Hum Reprod 2008;23:1661–8.
28. Zhao ZZ, Nyholt DR, Thomas S, Treloar SA, Montgomery GW. Polymorphisms in the vascular endothelial growth factor gene and the risk of familial endometriosis. Mol Hum Reprod 2008;14:531–8.
29. Zhao ZZ, Croft L, Nyholt DR, Chapman B, Treloar SA, Hull ML, et al. Evaluation of polymorphisms in predicted target sites for micro RNAs differentially expressed in endometriosis. Mol Hum Reprod 2011;17: 92–103.
30. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38: e164.
31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.
32. Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G, et al. Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet 2009;85:750–5.
33. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–1.
34. Helsby NA, Burns KE. Molecular mechanisms of genetic variation and transcriptional regulation of CYP2C19. Front Genet 2012;3:206.
35. Justenhoven C, Hamann U, Pierl CB, Baisch C, Harth V, Rabstein S, et al. CYP2C19*17 is associated with decreased breast cancer risk. Breast Cancer Res Treat 2009;115:391–6.
36. Shuldiner AR, O'Connell JR, Bliden KP, Gandhi A, Ryan K, Horenstein RB, et al. Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. JAMA 2009;302: 849–57.
37. Teichert M, Eijgelsheim M, Rivadeneira F, Uitterlinden AG, van Schaik RH, Hofman A, et al. A genome-wide association study of acenocoumarol maintenance dosage. Hum Mol Genet 2009;18:3758–68.
38. de Morais SM, Wilkinson GR, Blaisdell J, Nakamura K, Meyer UA, Goldstein JA. The major genetic defect responsible for the polymorphism of S-mephenytoin metabolism in humans. J Biol Chem 1994;269: 15419–22.
39. Painter JN, Zondervan KT, Missmer SA, Montgomery GW. Molecular genetics of endometriosis. In: eLS. Chichester: John Wiley & Sons Ltd, 011. Available at: http://www.els.net. http://dx.doi.org/10.1002/9780470015902.a0021473. Last accessed April 28, 2014.
40. Rahmioglu N, Missmer SA, Montgomery GW, Zondervan KT. Insights into assessing the genetics of endometriosis. Curr Obstet Gynecol Rep 2012;1: 124–37.
41. Bellelis P, Podgaec S, Abrao MS. Environmental factors and endometriosis. Rev Assoc Med Bras 2011;57:448–52.
42. Calafat AM, Ye X, Wong LY, Reidy JA, Needham LL. Exposure of the U.S. population to bisphenol A and 4-tertiary-octylphenol: 2003–2004. Environ Health Perspect 2008;116:39–44.
43. Chen Y, Goldstein JA. The transcriptional regulation of the human CYP2C genes. Curr Drug Metab 2009;10:567–78.
44. Mwinyi J, Cavaco I, Pedersen RS, Persson A, Burkhardt S, Mkrtchian S, et al. Regulation of CYP2C19 expression by estrogen receptor alpha: implications for estrogen-dependent inhibition of drug metabolism. Mol Pharmacol 2010;78:886–94.
45. Shelby MD. NTP-CERHR Monograph on the potential human reproductive and developmental effects of bisphenol A. NTP CERHR MON 2008. v, vii–ix, 1–64 passim.

SUPPLEMENTAL FIGURE 1



Sequence coverage in the (A) endometriosis case and (B) control DNA pools, scaled to 10,000-fold coverage.

Painter. CYP2C19 *and endometriosis risk. Fertil Steril 2014.*

SUPPLEMENTAL TABLE 1

Primers for the PCR amplification and sequencing of *CYP2C18* and *CYP2C19*.

Primer name	Primer sequence	Amplicon	Product size, bp
18P1F	gccaaatttgtggttagatagga	CYP2C18promoter_Amplicon1	850
18P1R	gcccgctgtggtatattctc		
18P2F	gagaggggagaaaccctagc	CYP2C18promoter_Amplicon2	827
18P2R	tttaacacatgcgcaggaag		
18P3F	tgaaagcaggttgcagattg	CYP2C18promoter_Amplicon3	847
18P3R	agcccaggattttgttgttg		
18E1F	ccagcacttagggagaccag	CYP2C18Amplicon_exon1	817
18E1R	ccctgatatgatcaattcacaa		
18E2&3F	acttgcttggcacattgaag	CYP2C18Amplicon_exon2&3	847
18E2&3R	ttgatcattacacattctgtgcat		
18E4F	aaacaagctagggtggaatga	CYP2C18Amplicon_exon4	582
18E4R	gggtaaagggagggatagca		
18E5F	tcatctttgatcccttgttcag	CYP2C18Amplicon_exon5	533
18E5R	tcaattataaattctccatggtgct		
18E6F	cacaccgtgcaattgttgta	CYP2C18Amplicon_exon6	597
18E6R	tgtgcctgctacaaatcgtc		
18E7F	ttagcttgtaggagagttggatg	CYP2C18Amplicon_exon7	760
18E7R	gattcaacctgtcaggacatca		
18E8F	gctcaggacttcaagtgcaa	CYP2C18Amplicon_exon8	698
18E8R	tctcatggcagatgagaaatg		
18E9F	tgttgagtgaagggtgccta	CYP2C18Amplicon_exon9	842
18E9R	cagcacagcctcttcacatt		
18UF	tgctgtcacctgcaattctc	CYP2C18UTR_Amplicon1	850
18UR	aattaccattgcatttatgttcag		
19P1F	ggaggagcagaactggaaca	CYP2C19promoter_Amplicon1	838
19P1R	taaatggacccacagcacaa		
19P2F	tgaccagtgaaacattgtgc	CYP2C19promoter_Amplicon2	844
19P2R	ttcaaatgggaaaagggaga		
19P3F	aagagataatgcgccacgat	CYP2C19promoter_Amplicon3	823
19P3R	cagagcacaaggaccacaaa		
19E1F	aatgtacagagtgggcactgg	CYP2C19_Amplicon_exon1	561
19E1R	ttcattccactatttctgacactga		
19E2&3F	ttgtctgaccattgccttga	CYP2C19_Amplicon_exon2&3	833
19E2&3R	tctcagcttcaaaccctgct		
19E4F	aaatttagcatttgagcaacca	CYP2C19_Amplicon_exon4	583
19E4R	tgggatattcatttcctgtgc		
19E5F	tcatctttgattctcttgtcagaa	CYP2C19_Amplicon_exon5	679
19E5R	cctgtgctgatctcactgga		
19E6F	ctggaagcattccctttgaa	CYP2C19_Amplicon_exon6	391
19E6R	gcaccttcccaacatctttc		
19E7F	catttcttcctgccttccttt	CYP2C19_Amplicon_exon7	682
19E7R	gccagaacaatcttgaagaaaga		
19E8F	ggccttaagctcatgcctct	CYP2C19_Amplicon_exon8	696
19E8R	agggacaccatgtggaagac		
19E9F	cacccaaccacccatctatc	CYP2C19_Amplicon_exon9	754
19E9R	catggcctcctggaactcta		

Note: Numbers refer to the amplicon (exonic primer names refer to the respective exon amplified). P = promoter region primer; E = exonic primer; F = forward primer; R = reverse primer.

Painter. CYP2C19 *and endometriosis risk. Fertil Steril 2014.*

SUPPLEMENTAL TABLE 2

Effect allele frequencies and significance of association for 80 polymorphic SNPs in endometriosis cases and controls.

					DNA pools[b]		384 individuals[c]		901 cases, 939 controls[d]				
Location	Position GRCh37/hg19	Name[a]	Amino acid change	Effect allele	EAF Ca	EAF Co	EAF Ca	EAF Co	EAF Ca	EAF Co	Unadjusted *P* value	OR	95% confidence intervals
HELLS													
Downstream	96362188	rs4918417		A	0.425	0.383	0.405	0.395	0.420	0.392	7.7×10^{-1}	1.13	0.98–1.28
Downstream	96362206	rs4918418		G	0.458	0.447	0.416	0.397	0.425	0.393	4.0×10^{-2}	1.14	1.00–1.30
Downstream	96362251	rs4918419		G	0.504	0.467	0.416	0.398	0.426	0.393	4.4×10^{-2}	1.14	1.00–1.30
Downstream	96362681	–		T	0.011	2.0×10^{-9}	0.011	0.001	0.006	0.001	9.8×10^{-3}	5.78	1.28–26.10
Downstream	96362748	rs75771163		C	0.026	0.033	0.031	0.033	0.039	0.034	4.3×10^{-1}	1.15	0.81–1.61
Downstream	96362792	rs11594492		T	0.213	0.145	0.216	0.210	0.226	0.203	9.4×10^{-2}	1.14	0.97–1.33
Downstream	96362813	–		A	0.008	0.013	0.005	0.012	0.010	0.014	2.9×10^{-1}	0.72	0.39–1.32
Intergenic	96376189	rs7100415		C	–	0.166	0.447	0.433	0.466	0.428	2.2×10^{-2}	1.16	1.02–1.32
Intergenic	96378763	–		T	1.1×10^{-8}	0.013	–	0.001	–	0.001	3.2×10^{-1}	0.00	–
Intergenic	96383383	rs113869604		G	0.025	0.002	0.001	–	0.001	–	3.1×10^{-1}	–	–
Intergenic	96383524	rs76541158		G	0.135	0.079	0.057	0.055	0.065	0.056	2.9×10^{-1}	1.16	0.88–1.51
Intergenic	96386206	rs1326837[+]		C	0.225	0.110	0.236	0.201	0.251	0.195	4.6×10^{-5}	1.38	1.18–1.61
Intergenic	96387079	rs11188044		G	0.264	0.173	0.172	0.177	0.171	0.176	7.4×10^{-1}	0.97	0.82–1.15
Intergenic	96387178	rs34195517[+]		T	0.191	0.110	0.203	0.168	0.214	0.167	2.8×10^{-4}	1.36	1.15–1.59
Intergenic	96404786	–		G	0.001	0.011	0.008	–	0.001	–	3.1×10^{-1}	–	–
Intergenic	96405072	–		C	0.009	0.002	0.007	0.009	0.004	0.006	3.8×10^{-1}	0.65	0.25–0.17
Intergenic	96405146	rs117092153		C	0.004	4.7×10^{-9}	0.009	0.001	0.006	0.001	1.7×10^{-2}	5.22	1.14–23.87
Intergenic	96405329	rs12772169[+]		T	0.216	0.165	0.238	0.202	0.250	0.196	7.4×10^{-5}	1.37	1.17–1.59
Intergenic	96405502	rs12777823[+]		A	0.163	0.113	0.180	0.146	0.184	0.138	1.2×10^{-4}	1.41	1.18–1.68
Intergenic	96405806	rs111261647		T	0.018	0.048	0.034	0.040	0.030	0.035	3.9×10^{-1}	0.85	0.59–1.22
Intergenic	96406166	rs74150870		A	0.052	0.055	0.063	0.056	0.067	0.058	2.4×10^{-1}	1.17	0.89–1.15
Intergenic	96406408	rs56655020[+]		C	0.170	0.172	0.222	0.173	0.214	0.168	3.2×10^{-4}	1.35	1.14–1.59
Intergenic	96406777	–		A	0.001	0.005	–	0.003	–	0.001	1.7×10^{-1}	0.00	0.00
Intergenic	96406961	rs117389245		G	0.006	0.002	0.009	0.003	0.006	0.001	1.7×10^{-2}	5.23	1.14–23.92
Intergenic	96407020	–		A	0.006	7.1×10^{-9}	0.009	0.001	0.006	0.001	1.7×10^{-2}	5.24	1.14–23.92
Intergenic	96407301	–		G	0.006	0.003	0.009	0.001	0.006	0.001	1.7×10^{-2}	5.23	1.14–23.92
Intergenic	96407313	rs34582766[+]		A	0.166	0.139	0.182	0.143	0.182	0.138	2.7×10^{-4}	1.39	1.16–1.66
Intergenic	96407352	rs7089349[+]		C	0.143	0.105	0.168	0.143	0.178	0.137	6.8×10^{-4}	1.36	1.13–1.62
Intergenic	96407511	rs10882487		T	0.174	0.184	0.165	0.179	0.169	0.175	6.5×10^{-1}	0.96	0.81–1.14
Intergenic	96407978	rs1998591[+]		A	0.227	0.193	0.167	0.138	0.251	0.196	8.7×10^{-4}	1.35	1.13–1.61
Intergenic	96409503	rs12782132[+]		C	0.181	0.135	0.179	0.147	0.185	0.139	1.5×10^{-4}	1.40	1.17–1.67
Intergenic	96410281	rs112454374		C	0.014	0.043	0.034	0.040	0.031	0.035	4.1×10^{-1}	0.86	0.60–1.22
Intergenic	96427951	rs12784245[+]		A	0.117	0.333	0.166	0.142	0.176	0.137	8.3×10^{-4}	1.35	1.13–1.61
Intergenic	96429130	rs79950930		G	0.066	–	0.057	0.054	0.066	0.056	2.1×10^{-1}	1.19	0.90–1.55
Intergenic	96429971	rs2104543[+]		T	0.219	0.129	0.232	0.198	0.249	0.194	5.2×10^{-5}	1.38	1.17–1.60
Intergenic	96435053	rs112104288		C	0.011	0.035	0.018	0.026	0.198	0.233	9.7×10^{-3}	0.81	0.53–1.34
CYP2C18													
5' UTR	96443501	–		A	0.004	0.005	–	0.003	–	0.001	1.7×10^{-1}	0.00	0.00
Intron	96446632	rs1010570		A	0.278	0.258	0.033	0.032	0.277	0.229	3.6×10^{-1}	1.17	0.83–1.64
Exon 2	96447562	rs41291550	Tyr68	A	0.006	0.006	0.011	0.003	0.006	0.002	4.3×10^{-2}	3.49	0.95–12.69
Exon 3	96447920	rs117111102	Arg124Trp	T	0.017	0.026	0.011	0.017	0.014	0.019	2.2×10^{-1}	0.73	0.44–1.21
Exon 4	96454755	rs41303835	Asp188Gly	G	1.4×10^{-8}	0.005	0.008	0.001	–	0.001	3.3×10^{-1}	0.00	0.00
Exon 4	96454820		Ser210Pro	C	0.002	0.005	–	–	0.001	–	3.1×10^{-1}	–	–

Painter. CYP2C19 *and endometriosis risk. Fertil Steril 2014.*

SUPPLEMENTAL TABLE 2

Continued.

					DNA pools[b]		384 individuals[c]		901 cases, 939 controls[d]				
Location	Position GRCh37/hg19	Name[a]	Amino acid change	Effect allele	EAF Ca	EAF Co	EAF Ca	EAF Co	EAF Ca	EAF Co	Unadjusted *P* value	OR	95% confidence intervals
Exon 7	96484145	rs41286880	Arg335Gln	A	0.030	0.017	0.032	0.020	0.026	0.021	3.9×10^{-1}	1.20	0.78–1.84
Exon 7	96484153	rs41286882	Cys338Arg	C	0.009	0.003	0.015	0.001	0.006	0.001	1.7×10^{-2}	5.23	1.14–23.92
Intron	96490176	–		G	0.007	4.7×10^{-13}	0.007	0.003	0.007	0.003	1.3×10^{-1}	2.09	0.78–5.58
Intron	96492949	–		G	–	–	–	–	0.001	–	3.2×10^{-1}	–	–
Exon 8	96493058	rs2281891	Thr385Met	T	0.163	0.126	0.182	0.163	0.172	0.135	5.3×10^{-1}	1.06	0.89–1.25
Intron	96493332	rs17110236		T	0.057	0.043	0.049	0.052	0.057	0.054	5.4×10^{-1}	1.09	0.82–1.44
Intron	96494912	rs79429965		T	0.066	0.061	0.057	0.055	0.064	0.057	2.7×10^{-1}	1.16	0.88–1.51
3' UTR	96495232	rs2860840		T	0.348	0.366	0.387	0.386	0.354	0.362	9.9×10^{-1}	1.00	0.87–1.14
Intergenic	96499512	rs7917617+		T	0.063	0.039	0.241	0.201	0.070	0.057	6.4×10^{-5}	1.37	1.17–1.59
Intergenic	96500010	rs74152352		A	0.047	0.020	0.067	0.057	0.071	0.056	7.4×10^{-2}	1.27	0.97–1.65
Intergenic	96519061	rs11188072		T	0.177	0.173	0.211	0.227	0.197	0.235	7.0×10^{-3}	0.81	0.68–0.94
Intergenic	96519691	rs77957608		A	0.050	0.033	0.057	0.055	0.065	0.056	2.9×10^{-1}	1.16	0.88–1.51
CYP2C19													
Upstream	96521657	rs12248560		T	0.198	0.209	0.212	0.228	0.198	0.236	6.0×10^{-3}	0.81	0.68–0.94
Upstream	96522365	rs4986894+		C	0.165	0.128	0.165	0.140	0.173	0.135	1.2×10^{-3}	1.34	1.12–1.60
Exon 1	96522561	rs17885098	Pro33Pro	C	0.045	0.037	0.071	0.058	0.071	0.058	1.0×10^{-1}	1.24	0.95–1.61
Intron	96534043	rs55805503		A	0.006	0.008	0.059	0.055	0.066	0.057	2.3×10^{-1}	1.18	0.90–1.54
Intron	96534200	rs17885958		A	0.015	0.021	0.019	0.018	0.016	0.018	6.3×10^{-1}	0.88	0.53–1.46
Intron	96534234	rs17885348		G	0.014	0.010	0.065	0.058	0.066	0.057	2.3×10^{-1}	1.18	0.90–1.54
Intron	96534263	rs6583954+		T	0.125	0.094	0.170	0.137	0.175	0.134	5.4×10^{-4}	1.37	1.14–1.64
Exon 2	96534922	rs17878459	Glu92Asp	C	0.015	0.020	0.033	0.034	0.035	0.028	1.9×10^{-1}	1.27	0.88–1.84
Exon 5	96541616	rs4244285+	Pro227Pro	A	0.157	0.120	0.165	0.136	0.171	0.133	5.3×10^{-4}	1.37	1.14–1.63
Intron	96547463	rs12767583+		T	0.128	0.089	0.238	0.197	0.174	0.135	7.3×10^{-5}	1.37	1.17–1.59
Intron	96549421	–		T	0.277	0.382	0.008	0.007	0.008	0.004	1.6×10^{-1}	1.84	0.76–4.38
Intron	96575242	rs1322179+		T	0.130	0.113	0.165	0.139	0.174	0.134	1.5×10^{-3}	1.33	1.11–1.59
Intron	96599689	rs11188093		A	0.393	0.318	0.442	0.409	0.433	0.407	1.1×10^{-1}	1.11	0.97–1.26
Intron	96600655	rs12775283+		T	0.147	0.089	0.166	0.139	0.173	0.134	1.0×10^{-3}	1.35	1.12–1.63
Intron	96601794	rs66883721+		G	0.329	0.236	0.234	0.199	0.247	0.193	9.0×10^{-5}	1.37	1.16–1.55
Exon 7	96602623	rs3758581	Ile331Val	A	0.080	0.052	0.059	0.053	0.066	0.056	1.1×10^{-2}	5.64	1.24–25.40
Exon 7	96602691	rs17882744	His353His	T	0.009	2.3×10^{-8}	0.004	0.001	0.002	0.002	9.5×10^{-1}	1.04	0.26–4.17
Intron	96602941	–		T	–	–	0.007	0.007	0.003	0.005	4.4×10^{-1}	0.67	0.23–1.87
Intron	96603414	rs11592737		G	0.093	0.210	0.213	0.229	0.198	0.236	7.0×10^{-3}	0.81	0.69–0.94
Intron	96604273	rs11597626		G	0.040	0.209	0.034	0.041	0.200	0.232	9.7×10^{-3}	0.81	0.69–0.95
Intron	96609064	rs1322181		A	0.334	0.462	0.448	0.423	0.445	0.425	2.3×10^{-1}	1.08	0.95–1.23
Intron	96609568	rs4917623		T	0.337	0.500	0.482	0.465	0.475	0.461	4.1×10^{-1}	1.06	0.92–1.20
Intron	96611628	rs35739901		C	0.047	0.200	0.218	0.227	0.198	0.233	9.7×10^{-3}	0.81	0.69–0.95
Intron	96611834	rs4617515		A	0.290	0.405	0.347	0.357	0.445	0.425	5.6×10^{-1}	0.96	0.84–1.09
Intron	96612040	rs12779363		A	0.039	0.155	0.221	0.226	0.199	0.232	1.3×10^{-2}	0.82	0.70–1.96
Intron	96612371	rs12268020		T	0.071	0.210	0.221	0.226	0.199	0.233	1.3×10^{-2}	0.82	0.70–0.96

[a] SNPs remaining nominally significant after permutation to correct for multiple testing.
[b] Effect allele frequencies (EAFs) in two DNA pools of 384 cases and 384 controls.
[c] EAFs in the 384 cases and 384 controls making up each pool.
[d] EAF and significance of association in the larger sample of 901 endometriosis cases and 939 controls analyzed in the discovery phase of the study.

Painter. CYP2C19 *and endometriosis risk. Fertil Steril 2014.*

SUPPLEMENTAL TABLE 3

Meta-analysis of the *CYP2C19* association signal in the discovery and replication data sets with endometriosis by year of diagnosis.

	1965–1993[a]		1994–2002[a]		
SNP	**Unadjusted *P***	**OR (95% CI)**	**Unadjusted *P***	**OR (95% CI)**	***P* difference[b]**
rs1326837	2.9×10^{-4}	1.25	6.4×10^{-5}	1.26	.67
rs4244285	9.5×10^{-3}	1.21	1.8×10^{-4}	1.27	.23
rs12248560	2.3×10^{-2}	0.86	8.6×10^{-4}	0.82	.27

Note: Our endometriosis cases were diagnosed between 1965 and 2002, corresponding to a period during which potential exposure to environmental contaminants that may increase endometriosis risk (33) may have increased through, for instance, greater use of plastics and associated chemicals (34). CI = confidence interval.

[a] Cases were divided into two groups by the median year of diagnosis in our data set, 1994, corresponding to a period after which exposure to environmental toxins as measured by bis-phenol-A are reported to have increased. Data from the U.S. National Health and Nutrition Examination Survey suggest a possible doubling of urinary bisphenol A levels between study periods encompassing 1988–1994 and 2003–2004 (34). Overall case N (1965–1993) = 966, N (1994–2002) = 1,285, and control N = 2,709 for both analyses.

[b] The significance between the effects of each SNP on endometriosis and the 2 years of diagnosis groups were tested by polytomous logistic regression (in STATA), appropriate for analyses comparing different sets of cases and the same set of controls.

Painter. CYP2C19 *and endometriosis risk. Fertil Steril 2014.*