Published in final edited form as:
Fertil Steril. 2014 August ; 102(2): 496–502.e5. doi:10.1016/j.fertnstert.2014.04.015.

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

Jodie N Painter, PhD [lead analyst],
QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, Queensland, 4006, Australia

Dale R Nyholt, PhD [analysis and imputation],
QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, Queensland, 4006, Australia

Lutz Krause, Phd [sequence alignment and variant calling],
QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, Queensland, 4006, Australia

Zhen Z Zhao, PhD [genotyping],
QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, Queensland, 4006, Australia

Brett Chapman, BSc [genotyping],
QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, Queensland, 4006, Australia

Christine Zhang [familial case sequencing],
Mater Medical Research Institute, Raymond Terrace, South Brisbane, Queensland, 4101, Australia

Sarah Medland, PhD [GWAS sample imputation and analysis],
QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, Queensland, 4006, Australia

Nicholas G Martin, PhD [sample collection],
QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, Queensland, 4006, Australia

Stephen Kennedy, PhD [sample collection],
Nuffield Department of Obstetrics and Gynaecology, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DU, UK

Susan Treloar, PhD [linkage and fine-mapping],
Centre for Military and Veterans' Health, The University of Queensland, Mayne Medical School, Herston Rd, Herston, Queensland, 4006, Australia

Corresponding Author: Jodie N Painter, PhD, Molecular Cancer Epidemiology Laboratory, QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, QLD, AUSTRALIA, 4006, jodie.painter@qimrberghofer.edu.au, Phone: +61 (0)7 3362 0389, Fax: +61 (0)7 3362 0105.

Conflicts of Interest: Dr Krina Zondervan, personal fees (consulting): Bayer HealthCare LTD, AbbVie LTD

Krina Zondervan, PhD [linkage and fine-mapping], and
Genetic and Genomic Epidemiology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK

Grant W Montgomery, PhD [QIMR Berghofer endometriosis project leader]
QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, Queensland, 4006, Australia

Abstract

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)—Cases = 2,271 women with surgically confirmed endometriosis; Controls = 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 SNPs in two case-control samples.

Main outcome measure(s)—Allele frequency differences between cases and controls.

Results—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 LD 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, highlighted in our previous study.

Conclusion(s)—Functional variants in *CYP2C19* may contribute to endometriosis susceptibility in both familial and sporadic cases.

Keywords

Endometriosis; association; pooled sequencing; *CYP2C19*; rs12248560; rs4244285

Introduction

Endometriosis is a gynaecological 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 (5-7, 8). Both linkage and association studies take advantage of linkage disequilibrium (LD), where the alleles of 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 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 suggestive evidence of association to 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 = 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 to a SNP independent of rs12248560, rs4244285 (LD estimate $r^2 = 0.04$), a loss-of-function variant in *CYP2C19* exon 5 which 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 be contributing 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 inter-genic and intronic regions that may harbour 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.

Material 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 1995-2002 (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 classification system (18). Control samples comprised unrelated individuals originally recruited through QIMR Berghofer for either a twin study of gynaecological health who self-reported no endometriosis (discovery sample) (4) and the sample previously utilised 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 3 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 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) (Supplementary Table 1). PCRs were performed in 15 μl reactions containing 1X PCR buffer, 1.5 mM MgCl2, 1 U Amplitaq Gold (all Applied Biosystems, CA, USA), 200 μM each dNTP (Promega, Madison, WI, USA), 1 μM each of forward and reverse primers and 50 ng of DNA. PCR cycling conditions included initial denaturation at 95°C for 5 mins; 'touchdown' cycling at 95°C for 30 secs, 60°C (-0.5 per cycle) for 30 secs and 72°C for 30 secs for 20 cycles; 95°C for 30 secs, 50°C for 30 secs and 72°C for 30 secs for 15 cycles; 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, Burlington, Ontario,Canada) 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 gynaecological health study samples (control pool) (22). Targeted re-sequencing and analysis of the case and control DNA pools were performed by deCODE Scientific Services (Reykjavik, Iceland). 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, San Diego, CA, USA). Short read sequences were then aligned and sequence variants (SVs; termed as such as many loci were yet to be verified) called by deCODE using the Illumina programs CASAVA and Elandv2. In addition, we reanalysed the sequences using programs tailored for DNA pools: sequences were aligned using the program Novoalign (http://www.novocraft.com/main/index.php) and SVs 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 of 958 endometriosis cases and 959 unrelated gynaecological 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, San Diego, CA, USA). 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 ($P < 0.05$) allele frequency differences between the case and control pools (83 SVs across the 3 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 $<1\times10^{-4}$ 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 gynaecological health study controls passing all quality control metrics using PLINK (31). Study-wide significance at the discovery phase ($P_{adjusted} <0.05$) was determined by 10,000 permutations of the dataset, 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 utilised 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 analysed genotype data and dosage scores for SNPs previously imputed to HapMap2 (32). All imputed SNPs had genotype concordance values of between 97.1-99.7% for 560 endometriosis cases for whom both actual and imputed genotypes were available, with imputation quality scores >0.9 indicating these SNPs were imputed with a high degree of accuracy.

Association analyses were performed in the replication dataset 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 (Supplementary Fig. 1).

deCODE Genetics called 3,051 SVs using their analysis pipeline, of which 2,418 (79%) were novel (not present in 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 minor allele frequencies (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 2 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 passing our quality control. Neither genomic location (inter-genic or intronic vs. exonic, potentially functional or conserved areas) or 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 0.001-0.40 in either the case or control pools, while SVs polymorphic in individual genotyping had pool MAFs ranging from 0.001-0.17.

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 utilised 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

Following 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 ($P \leq 0.05$) was detected for 35 SNPs, 19 of which remained significant ($P_{adjusted} \leq 0.05$) following permutation to correct for multiple testing (Table 2; Supplementary Table 2). Association signal was distributed across the sequenced region (Fig. 1), with the best signal seen for rs1326837 ($P = 4.6 \times 10^{-5}$, 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 are contributing to the endometriosis risk conferred by this haploptype.

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 dataset of 1,370 mostly non-familial endometriosis cases and 1,770 population controls, with a one-sided *P*-value for rs1326837 of 3.1×10^{-3} (OR = 1.19), and rs4244285 $P = 8.0 \times 10^{-3}$ (OR=1.18). Meta-analysis of the discovery and replication datasets 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 $P = 0.05/80$ 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; Supplementary 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. rs12248560 was, however, no longer significant following permutation of the 901 case and 939 control dataset, indicating the signal for this SNP was not study-wide significant in the discovery set ($P_{adjusted} = 0.16$). However, there was evidence of

association in the replication sample (1-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 this SNP to be significant over the total study sample.

Discussion

We have found genetic variants in the *CYP2C19* gene to be associated with susceptibility to endometriosis amongst 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 following our initial fine-mapping of a linkage peak on chromosome 10 (11). While study-wide significant, the rs12248560 signal is somewhat reduced in the current study (fine-mapping sample minor allele $P=4.9\times10^{-4}$, OR = 0.78 (11) versus current discovery sample $P=6.4\times10^{-3}$, OR = 0.80), this appears 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 ultra-rapid drug metaboliser 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 inter-genic 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 inter-genic 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, non-functional protein (38) and a poor metaboliser phenotype (*CYP2C19*2*). In our previous fine-mapping study including cases from Australia and the UK we found only nominal association to rs4244285 (OR = 1.23, $P = 1.1\times10^{-2}$) 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, $P = 9.7\times10^{-4}$; meta-analysis OR = 1.35, $P = 9.4\times10^{-5}$) despite the smaller sample size may indicate that the effect of this SNP on endometriosis risk is larger in the Australian than in the UK population.

CYP2C19 participates in the conversion of 17β-estradiol (E2) to estrone (E1) (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 metaboliser phenotypes (rs12248560) are protective while SNPs causing the poor metaboliser 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 localised, tissue-specific rates of conversion of E2 to E1, and/or result in altered levels of the anti-estrogenic 2α-OHE metabolite, with activity either increased or decreased as compared to 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, due mostly to small sample sizes typically analysed 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 metabolised by this gene are cleared, either by the liver or locally at extra-hepatic 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 (Supplementary 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 datasets 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 due 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 are likely to be contributing to this risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 (grants 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 NIH, US, grant 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) schemes, 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 gynaecologist 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, gynaecologists, 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. [PubMed: 20573927]
2. Simpson JL, Bischoff FZ. Heritability and molecular genetic studies of endometriosis. Ann N Y Acad Sci. 2002; 955:239–51. [PubMed: 11949952]
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. [PubMed: 18535005]
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. [PubMed: 10203663]
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. [PubMed: 21151130]
6. Uno S, Zembutsu H, Hirasawa A, Takahashi A, Kubo M, Akahane T, et al. A genome. Nat Genet. 2012; 42:707–10. [PubMed: 20601957]
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. [PubMed: 20844546]
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. [PubMed: 23104006]
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. [PubMed: 19264985]
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. [PubMed: 16080113]
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. [PubMed: 21497341]
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. [PubMed: 12865317]
13. Cribb AE, Knight MJ, Dryer D, Guernsey J, Hender K, Tesch M, Saleh TM. Role of polymorphic human cytochrome P450 enzymes in estrone oxidation. Cancer Epidemiol Biomarkers Prev. 2006; 15:551–8. [PubMed: 16537715]
14. Desta Z, Zhao X, Shin JG, Flockhart DA. Clinical significance of the cytochrome P450 2C19 genetic polymorphism. Clin Pharmacokinet. 2002; 41:913–58. [PubMed: 12222994]
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. [PubMed: 21830868]

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. [PubMed: 19499406]
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. 2010:286–90. [PubMed: 20113968]
18. Revised American Fertility Society classification of endometriosis: 1985. Fertil Steril. 1985; 43:351–2. [PubMed: 3979573]
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. [PubMed: 9915566]
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. [PubMed: 10417291]
21. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 2000; 132:365–86. [PubMed: 10547847]
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. [PubMed: 16176720]
23. Bansal V. A statistical method for detection of variants from next-generation re-sequencing of DNA pools. Bioinformatics. 2010; 26:i318–24. [PubMed: 20529923]
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. [PubMed: 17563403]
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. [PubMed: 16973828]
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. [PubMed: 17595314]
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; 23:1661–68. [PubMed: 18285324]
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. [PubMed: 18650217]
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. [PubMed: 20935158]
30. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38:e164. [PubMed: 20601685]
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. [PubMed: 17701901]
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. [PubMed: 19896111]
33. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–1. [PubMed: 20616382]
34. Helsby NA, Burns KE. Molecular mechanisms of genetic variation and transcriptional regulation of CYP2C19. Front Genet. 2012; 3:206. [PubMed: 23087703]
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. [PubMed: 18521743]

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. [PubMed: 19706858]
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. [PubMed: 19578179]
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. [PubMed: 8195181]
39. 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. [PubMed: 18535005]
40. Rahmioglu N, Missmer SA, Montgomery GW, Zondervan KT. Insights into assessing the genetics of endometriosis. Curr Obstet Gynecol Rep. 2012; 1:124–37. [PubMed: 22924156]
41. Bellelis P, Podgaec S, Abrao MS. Environmental factors and endometriosis. Rev Assoc Med Bras. 2011; 57:448–52. [PubMed: 21876930]
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. [PubMed: 18197297]
43. Chen Y, Goldstein JA. The transcriptional regulation of the human CYP2C genes. Curr Drug Metab. 2009; 10:567–78. [PubMed: 19702536]
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. [PubMed: 20675569]
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. [PubMed: 19407859]

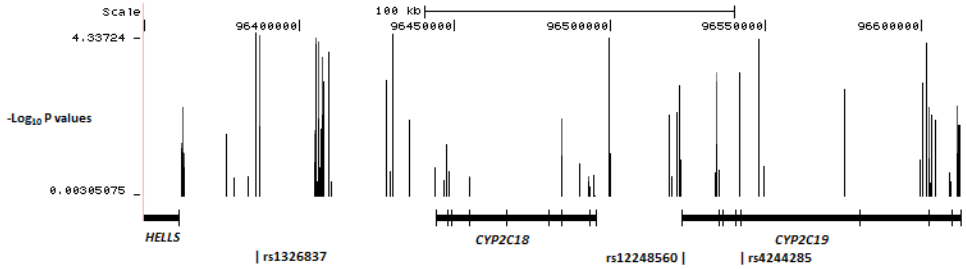

Figure 1.

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.

Table 1

Average minor allele frequencies 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%)

Table 2

Association signal to endometriosis across the region between the *HELLS* and *CYP2C19* genes on chromosome 10q26. Signal across the region is accounted for by SNPs in 2 blocks of linkage disequilbrium (LD), with the best SNP in each block indicated by *. rs4244285 is included as a known functional SNP in moderate LD (r^2 0.64) with rs1326837.

			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}	$1.6\times10\text{-}^{1}$	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)

[#]MAF = minor allele frequency