Variants in *EMX2* and *PTEN* do not contribute to risk of endometriosis

Susan A. Treloar¹,5, Zhen Zhen Zhao², Lien Le², Krina T. Zondervan³, Nicholas G. Martin¹, Stephen Kennedy⁴, Dale R. Nyholt¹ and Grant W. Montgomery²

¹Genetic Epidemiology Laboratory, Queensland Institute of Medical Research, PO Royal Brisbane Hospital, Brisbane, Queensland 4029, Australia; ²Molecular Epidemiology Laboratory, Queensland Institute of Medical Research, PO Royal Brisbane Hospital, Brisbane, Queensland 4029, Australia; ³Wellcome Trust Centre for Human Genetics, Oxford OX4 1SX, UK; ⁴Nuffield Department of Obstetrics and Gynaecology, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

Endometriosis has a genetic component, and significant linkage has been found to a region on chromosome 10q. Two candidate genes, *EMX2* and *PTEN*, implicated in both endometriosis and endometrial cancer, lie on chromosome 10q. We hypothesized that variation in *EMX2* and/or *PTEN* could contribute to the risk of endometriosis and may account for some of the linkage signal on 10q. We genotyped single nucleotide polymorphisms (SNPs) in a case–control design to evaluate association between endometriosis and common variations in these two genes. The genotyping and statistical analysis were based on samples collected from Australian volunteers. The cases were 768 unrelated women with surgically confirmed endometriosis selected from affected sister pair (ASP) families participating in the Australian Genes behind Endometriosis Study. The controls were 768 female participants in twin studies who, based on screening questions, did not have a diagnosis of endometriosis. Genotypes of 22 SNPs in the *EMX2* gene and 15 SNPs in the *PTEN* gene were the main outcome measures. Statistical analysis provided measures of linkage disequilibrium and association. Permutation testing showed no globally significant association between any SNPs or haplotypes and endometriosis for either gene. It is unlikely that the *EMX2* or *PTEN* gene variants investigated contribute to risk for initiation and/or development of endometriosis.

**Keywords:** association test; *EMX2*; endometriosis; polymorphism; *PTEN*

**Introduction**

Endometriosis (MIM 131200) is a common gynaecological disease that causes pelvic pain, severe dysmenorrhea (painful periods) and sub-fertility. It is defined as the presence of tissue resembling endometrium in sites outside the uterus, most commonly the pelvic peritoneum, ovaries and rectovaginal septum (Giudice and Kao, 2004). The population prevalence is difficult to measure, but the best estimates indicate that endometriosis affects 8–10% of women in their reproductive years (Eskenazi and Warner, 1997; Treloar et al., 1999a). The disease is inherited as a complex genetic trait (Simpson and Bischoff, 2002; Kennedy, 2003; Giudice and Kao, 2004) but its precise aetiology is uncertain. Familial aggregation has been reported in humans (Kennedy et al., 1995; Steffansson et al., 2002) and in non-human primates with spontaneous disease (Zondervan et al., 2004). Genetic factors account for 52% of the variation in liability to endometriosis (Treloar et al., 1999a), and the genetic relative-recurrence risk for sibs (λ₁) was estimated to be 2.34 in a sample of Australian twin families (Treloar et al., 1999a).

Finding disease-predisposition genes will help to define the mechanisms responsible for the initiation and development of endometriosis. To that end, we recruited a large cohort of families in Australia for genetic studies of endometriosis, in collaboration with a UK group (Treloar et al., 2002). A combined linkage scan in 1176 Australian and UK families containing sisters with surgically diagnosed disease identified a region of significant linkage on chromosome 10q26 with a maximum LOD score of 3.09 (Treloar et al., 2005). The peak linkage signal was located at 148.75 cM between markers D10S587 and D10S1656 and the 95% confidence interval (CI) spans a region of 8.5 Mb (Treloar et al., 2005).

Identifying specific positional candidates within a large linkage region is problematic for most complex diseases. Many genes could predispose to endometriosis as a number of pathways may be involved. Two candidates on chromosome 10q, which have previously been implicated in endometriosis and endometrial cancer, are homologue of Drosophila, empty spiracles, 2 (*EMX2*, MIM #600035) (Daftary and Taylor, 2004) and phosphatase and tensin homologue (*PTEN*, MIM #601728) (Sato et al., 2000; Kurose et al., 2001, 2002; Fujii et al., 2002; Latta and Chapman, 2002; Martini et al., 2002; Swiersz, 2002; Zhou et al., 2002; Simpson et al., 2003; Bischoff and Simpson, 2004; Dinulescu et al., 2005).

*EMX2* is a transcription factor essential for reproductive tract development. The gene is expressed in the developing vertebrate brain and in the epithelium of the developing urogenital tract (Simeone et al., 1992a,b). Mice homozygous for a knockout mutation in *EMX2* have defects in Müllerian duct development and renal function (Gangemi et al., 2001). *EMX2* is also expressed in the adult uterine endometrium with decreased expression during the luteal phase of the menstrual cycle (Troy et al., 2003; Daftary and Taylor, 2004). *EMX2* may...
have anti-proliferative effects in the endometrium as expression of the gene is decreased in endometrial tumours and increases after the menopause (Noonan et al., 2001, 2003). In patients with endometriosis, the marked decrease in EMX2 expression usually seen in the luteal phase fails to occur (Dalfy and Taylor, 2004). EMX2 is located at 119.3 megabase pairs (Mb), just outside the 95% confidence region for our linkage peak on chromosome 10 (119.4–127.9 Mb).

Phosphatidylinositol-3,4,5-triphosphate 3-phosphatase (PTEN) promotes cell survival and proliferation and plays an important role in endometrial tumorigenesis (Lin et al., 1998; Obata et al., 1998; Kurose et al., 2001; Latta and Chapman, 2002; Zhou et al., 2002). The data suggest that inactivation of PTEN is an early event in endometrial hyperplasia and the development of endometrial and ovarian cancers (Maxwell et al., 1998). PTEN was first reported in relation to synchronous endometrial and ovarian cancers (Lin et al., 1998), as well as separate endometrial (Maxwell et al., 1998) and ovarian cancers (Obata et al., 1998). PTEN gene expression changes under the influence of steroid hormones during the menstrual cycle (Mutter et al., 2000) and reduced expression of PTEN has been reported in some cases of endometriosis (Martini et al., 2002). PTEN lies at 89.6 Mb, more centromeric than our region of significant linkage. However, the linkage peak is broad and there is evidence for linkage and association with endometriosis in Puerto Rican families at marker D10S677 (Flores et al., 2004), which is located at 113.34 cm (95.95 Mb), close to the PTEN locus.

We hypothesized that variation in EMX2 and/or PTEN could contribute to the risk of endometriosis and may account for some of the linkage signal on chromosome 10q. We genotyped single nucleotide polymorphisms (SNPs) in EMX2 and PTEN in a case–control design, drawing cases from the families showing linkage to the chromosome 10 region so as to evaluate association between endometriosis and common variation in these two candidate genes.

Materials and Methods

Participants and sample collection

The project was approved by the Human Research Ethics Committee of the Queensland Institute of Medical Research and phenotypic data, medical records and DNA samples were collected after obtaining informed, written consent. Unrelated cases with surgically confirmed endometriosis were selected, one from each of the 768 Australian families with an affected sister pair (ASP) participating in our Australian Genes behind Endometriosis Study (Treloar et al., 2002, 2005). We selected the sister with the most severe stage of disease and if sisters had the same stage, we selected the younger pair (ASP) participating in our Australian Genes behind Endometriosis Study (Maxwell et al., 1998). Thirty-nine percent of the EMX2 SNPs were either monomorphic or had low frequencies in our sample and were excluded from genotyping. Four further assays subsequently provided unreliable genotype data, and these were also excluded. The remaining 22 EMX2 SNPs were selected for genotyping: 8 polymorphisms in the promoter region (rs1860399, rs82613, rs703409, rs700411, rs1638626 and rs2286629), 6 in the intronic region (rs853209, rs855769, rs365446, rs8192640, rs740734 and rs855768), 1 at an intron/exon boundary (rs2240776) and 7 in the 3′ UTR region (rs703413, rs4751627, rs242960, rs8181280, rs855766, rs4752078 and rs4752079).

Seventeen PTEN SNPs were polymorphic in our sample and selected for genotyping: 12 in the intron (rs2673836, rs1234220, rs1234219, rs2299939, rs1102597, rs1234231, rs1234224, rs2735343, rs17431184, rs2376627, rs926091 and rs532678), 2 at an intron/exon boundaries (rs1903858 and rs555895), 1 in exon 5 (rs9651492) and 2 in the 3′ UTR (rs701848 and rs478839).

Genotyping

Forward and reverse PCR primers and an extension primer for each assay were designed using SpectroDESIGNER software (version 2.1, SequenomTM, San Diego, CA, USA) and assembled into multiplex sets. SNPs were typed using the SequenomTM h-ME™ protocol on matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) MassARRAY platform. The PCR volume of 2.5 μl contained 10 ng DNA, 1 × PCR buffer, 1 mM MgCl2, 200 μM dNTP, forward and reverse primers to a final concentration of 50 nM for all SNPs and 0.1 μM of Taq polymerase. PCR cycling conditions were an initial denaturation step of 15 min at 95°C, followed by 45 cycles of 20 s at 95°C, 30 s at 56°C, 1 min at 72°C and a final extension step of 3 min at 72°C. To neutralize and remove residual unincorporated dNTPs from PCR, shrimp alkaline phosphatase was added and the reaction incubated at 37°C for 20 min and then 85°C for 5 min. The MassEXTEND reaction mix contained allele-specific extension primers at 600 nM each, h-ME extension mix and 0.08 μM of ThermoSequenase. PCR cycling conditions were a denaturation step of 2 min at 94°C, followed by 55 cycles of 5 s at 94°C, 5 s at 52°C and 5 s at 72°C. Purification of h-ME reaction products was performed by the addition of resin. Primer products were spotted onto 384-well chips, precoated with a 3-hydroxypropionic acid matrix, to help bind and stabilize DNA during ionization by the MassARRAY. Loading of the products was performed using a nanolitre pipetting system (MassARRAYTM, Sequenom) and reaction products analysed through MALDI-TOF MassARRAY Typer 3.0.1 (Sequenom).

The mean ages (± SD) of the cases and controls at the time of data collection were 35.6 ± 9.1 years (range = 17–65) and 45.7 ± 12.2 (range = 29–90) years, respectively. Cases and controls differed significantly in terms of mean number of live births (1.13 ± 1.25 and 2.19 ± 1.48 births, P < .001), but not subfertility, although the ‘problems conceiving’ question was more stringent for controls, involving a medical consultation, than cases. Cases reported a history of severe pelvic pain more than controls (79.5% versus 9.2%, P < .001), and age at onset was significantly younger (mean 21.3 ± 7.6 years compared with 25.9 ± 9.2 years, P < .001). Likewise, a history of dysmenorrhoea was more common in cases than controls (89.7% versus 18.0%, P < .001), and age at onset was younger (17.6 ± 6.3 compared with 21.0 ± 8.1 years, P < .001). Hysterectomy was more common in the cases (26.7%) than controls (13.8%) (P < .001), and mean age at hysterectomy was significantly younger (37.4 ± 5.9 years compared with 41.8 ± 8.4 years, P < .001). Ancestry reported for the four grandparents of cases and controls was comparable, with 93.2% of cases and 89.5% of controls reporting at least 50% Northern European origins.

DNA was extracted from peripheral blood lymphocytes (Miller et al., 1988) or from buccal swabs with the use of Microcon Centrifugal Filter Devices (Amicons) and was stored at 4°C at a concentration of 30–50 ng/μl. Aliquots of genomic DNA were adjusted to final concentrations of 2 ng/μl for Sequenom® MassARRAY analysis.

SNP selection

We selected 40 SNPs in the EMX2 gene and 28 SNPs in the PTEN gene on the basis of frequency information and distribution across the gene. The chosen EMX2 list comprised 16 promoter, 7 intronic, 2 intron/exon boundary, 4 exonic and 11 in the 3′ untranslated region (UTR) SNPs. Two promoter, 17 intronic, 2 intron/exon boundary, 5 exonic and 2 3′ UTR SNPs were chosen in the PTEN gene. Fourteen EMX2 SNPs were either monomorphic or had low frequencies in our sample and were excluded from genotyping. Four further assays subsequently provided unreliable genotype data, and these were also excluded. The remaining 22 EMX2 SNPs were selected for genotyping: 8 polymorphisms in the promoter region (rs1860399, rs82613, rs242956, rs703409, rs703411, rs1638626 and rs2286629), 6 in the intronic region (rs853209, rs855769, rs365446, rs8192640, rs740734 and rs855768), 1 at an intron/exon boundary (rs2240776) and 7 in the 3′ UTR region (rs703413, rs4751627, rs242960, rs8181280, rs855766, rs4752078 and rs4752079).

Seventeen PTEN SNPs were polymorphic in our sample and selected for genotyping: 12 in the intron (rs2673836, rs1234220, rs1234219, rs2299939, rs1102597, rs1234231, rs1234224, rs2735343, rs17431184, rs2376627, rs926091 and rs532678), 2 at an intron/exon boundaries (rs1903858 and rs555895), 1 in exon 5 (rs9651492) and 2 in the 3′ UTR (rs701848 and rs478839).
Statistical analysis

Phenotypic data analyses were performed using SPSS Version 13.0 (SPSS Inc., 2004). Marker genotypes were inspected and results tested for departures from Hardy–Weinberg equilibrium separately for cases and controls using the PEDSTATS program (Wigginton and Abecasis, 2005). The Haplo.stat (Schaid et al., 2002) and UNPHASED/COCAPHASE (Dudbridge, 2003) programs were used to test for association between endometriosis and individual markers or combinations of markers (haplotypes) (Dudbridge, 2003). In addition to obtaining nominal P-values, 10,000 permutation tests were performed to obtain a region-wide empirical P-value for each marker. This maintained the individual genotype as a whole while the individual’s disease status was shuffled. The method preserves the relation between SNPs (linkage disequilibrium, LD) while breaking the relation between status and the genotypes. For each replicate or permutation, each SNP was tested for association and the most significant P-value was stored. The global significance level was derived from these permutation tests. Haplotype blocks were determined by Haplovie (Barrett et al., 2005) using the default method of Gabriel et al. (2002).

We performed power calculations for our case–control study using the Genetic Power Calculator (Purcell et al., 2003), assuming a disease prevalence of 10%. Power calculations for our total sample assumed loci with dominant, recessive and multiplicative modes of inheritance (MOI) and were based on 768 unrelated cases and 768 unrelated controls. Power calculations used a significance threshold (\(P = 0.040\)) which corrects for the number of independent SNPs estimated via Nyholt’s Single Nucleotide Polymorphism Spectral Decomposition (SNPSpD) approach (Nyholt, 2004). Using this SNPSpD approach (Nyholt, 2004), the 22 EMX2 and 15 PTEN SNPs equated to approximately 8.01 and 9.00 effectively independent SNPs in our population, respectively. Therefore, all power calculations used a significance threshold (\(P = 0.00294\)) providing correction for testing a total of 17.01 independent SNPs. For our total sample of 768 cases and 768 controls, there is over 80% power to detect dominant disease-predisposing alleles of frequency 0.05, 0.25 and 0.5 contributing a genotype relative risk (GRR) of 1.72, 1.48 and 1.79, respectively. For recessive alleles, there is >80% power to detect alleles of frequency 0.05 and 0.25 with a GRR of 2.28 and 1.57, respectively. For a multiplicative MOI, there is >80% power to detect alleles of frequency 0.05, 0.25 and 0.5 contributing a GRR of 1.66, 1.32 and 1.29, respectively.

Results

A total of 22 polymorphic SNPs spanning a region of 7.1 Kb across EMX2 were included in the final analyses (Fig. 1). Figure 1a shows the position of the SNPs genotyped in the EMX2 gene. Figure 1b shows common haplotype blocks and Fig. 1c shows a LD plot of SNPs in EMX2.

In total, 15 PTEN SNPs spanning a region of 108.8 Kb were included in the final analyses (Fig. 2). Figure 2a shows the position of the SNPs genotyped in the PTEN gene. Figure 2b shows common haplotype blocks, and Fig. 2c shows a LD plot of SNPs in PTEN.

The minor allele frequencies of the EMX2 SNPs ranged from 0.021 to 0.384 (Table 1). The minor allele frequencies of the PTEN SNPs ranged from 0.017 to 0.472 (Table 2). We found no evidence for association between endometriosis and individual SNPs in either EMX2 or PTEN for either the allelic or the genotypic association tests. As 62 controls were older than 65 years at the time of data collection and, at ages above 65, selection starts to play a role in survival and could potentially introduce population stratification, we also compared minor allele frequencies between controls under and over age 65 and ran analyses excluding controls over 65. Although the EMX2 SNP rs82612 gave a lower P-value of 0.040, we conclude that in the multiple testing context, this was not significant. No SNPs in PTEN showed a \(P < 0.05\) using the age-restricted control set. Stratification of cases into the 293 cases who were never pregnant and the 475 cases with a history of at least one pregnancy also showed no significant differences from controls for any SNPs in EMX2 or PTEN, given the multiple testing context.

Haplotype analyses using two through five SNP sliding windows of contiguous SNPs and haplotype block analyses showed no significant associations for haplotypes between either EMX2 or PTEN and endometriosis.

Permutation test results (running 10,000) for EMX2 gave a best \(P\)-value of 0.060, but a global significance of 0.348 (SE 0.005). To maximize power (Wicks et al., 2004), restricting the cases to a set of 384 in families showing linkage at the Chromosome 10q peak and to a set of 384 controls resulted in a best \(P\)-value of 0.007 and a global significance of 0.069 (SE 0.003). In non-linked families (384 cases) and 384 controls, permutations showed the best \(P\)-value was 0.098, but the global result was non-significant (\(P = 0.444\), SE 0.005). Stratification of cases according to stage of disease (469 stage A cases and 768 controls) gave a best \(P\)-value of 0.031, but the global result was non-significant (\(P = 0.196\), SE 0.004). In the permutations for 296 cases diagnosed with stage B and 768 controls, the best \(P\)-value was 0.076, and global significance was 0.412 (SE 0.005).

Permutation tests (10,000) for PTEN gave a best \(P\)-value of 0.078, but a global significance of 0.515 (SE 0.005). Restricting the cases to a set of 384 women in the families showing linkage at the Chromosome 10q peak and a set of 384 controls resulted in a best \(P\)-value of 0.044 and a global significance of 0.336 (SE 0.005). In non-linked families (384 cases) and 384 controls, permutations showed the best \(P\)-value was 0.022 but the global result was non-significant (\(P = 0.213\), SE 0.004). Stratification of cases according to stage of disease (469 stage A cases and 768 controls) gave a best \(P\)-value of 0.071, but the global result was non-significant (\(P = 0.478\), SE 0.005). In the permutations for 296 cases diagnosed with stage B disease and 768 controls, the best \(P\)-value was 0.071, and global significance was 0.471 (SE 0.005).

Discussion

Our results do not support an association between endometriosis and variation in either EMX2 or PTEN. Although EMX2 is located just outside our linkage CI on chromosome 10q, we selected it for association testing because of its role in endometrial cell proliferation (Taylor and Fei, 2005) and the evidence of altered EMX2 expression in endometriosis (Daftry and Taylor, 2004). Expression in the endometrium generally declines around the time of implantation, but it remains elevated in women with endometriosis (Daftry and Taylor, 2004) which, on the basis of murine and human studies (Taylor and Fei, 2005), would lead to reduced endometrial proliferation. Hence, altered EMX2 expression in the endometrium may represent a marker of endometriosis-associated infertility (Daftry and Taylor, 2004). The pathway is also implicated because HOXA10 regulates EMX2 expression (Daftry and Taylor, 2004) and aberrant HOXA10 methylation has been reported in the endometrium of women with endometriosis (Wu et al., 2005). An antisense RNA transcript (EMX2OS) at the EMX2 locus may also have a regulatory function in EMX2 expression (Noonan et al., 2003). However, the results of our association study show that any effects of common variants in EMX2 itself that contribute to the risk of endometriosis, if present, must be small.

PTEN also lies outside our linkage region but is located close to a peak reported by Flores et al. (2004). Reduced expression of PTEN is thought to play a role in the malignant evolution of endometriosis, but the effect is also not in the expected direction (Martini et al., 2002). We saw no evidence for common variation in PTEN influencing the risk of endometriosis, but somatic rather than germ-line changes may be the basis of the association between PTEN and malignant transformation, given data suggesting that inactivation of the
**PTEN** tumour suppressor gene is an early event in the development of some endometrial cancers (Maxwell et al., 1998).

We chose SNPs that gave good coverage of both genes before information was available on allele frequencies from the HapMap project (The International HapMap Consortium, 2005); hence, we were unable to select sets of tag SNPs. Of the 104 known SNPs in **EMX2** and the 226 known SNPs in **PTEN**, SNPs were chosen on the basis of available frequency information, i.e. with frequencies above 5%.

---

**Figure 1:** Variation in the human **EMX2** gene.

(a) The genomic structure of the **EMX2** gene showing the location of the 22 SNPs genotyped. Grey arrows represent SNPs that were excluded from final analyses because they were monomorphic, had allele frequencies <5% or results did not give us confidence in the assays. (b) Common haplotype blocks in **EMX2**. Dashed lines indicate locations where transition from one common haplotype to a different one is observed. Haplotypes are combinations of genetic variants that are near each other and tend to be inherited together, i.e. they are regions of linked variants (http://www.hapmap.org/whatishapmap.html). The human genome has a haplotype block structure, in that it can be divided into discrete blocks of limited haplotype diversity (Zhang et al., 2002). Haplotype block structure is a summary measure describing the linkage disequilibrium structure across a region (Abecasis et al., 2005). (c) LD plot of **EMX2** SNPs; LD measured as the correlation coefficient $r^2$ using Haploview (Barrett et al., 2005). LD causes tightly linked genetic variants to be highly correlated (Abecasis et al., 2005). Grey shading represents correlation magnitudes between low $r^2$ (white) and high $r^2$ (black).
All coding SNPs, intronic/exonic boundary SNPs and a large proportion of promoter SNPs were chosen. Further, extra SNPs were selected to maximize even distribution across the genes. Many SNPs were not polymorphic in our sample and genotyping information was available for fewer SNPs, especially in the case of *EMX2*. Our data for *EMX2* show strong LD in two blocks covering most of the gene (Fig. 1c). Although we acknowledge the subjectiveness of haplotype ‘blocks’, we are confident that an association, if present, would have been found using the SNPs genotyped and haplotype analyses of window-sizes up to five SNPs. There was no evidence for
association in either of these genes that would warrant a more detailed screen. For PTEN, the two haplotype blocks included a smaller proportion of SNPs genotyped. Nevertheless, many correlation coefficients between SNPs were still of substantial magnitude across the gene (Fig. 2c). Greater numbers of SNPs that identify all common alleles would need to be typed to exclude association completely between PTEN and common variants in EMX2 locus.

One potential limitation of our control group is that it may have included undiagnosed cases of endometriosis. It is impossible to know how many undiagnosed cases exist either in the population or in our control sample. Any prevalence estimate for endometriosis is not only age-dependent but also is likely to covary with age at surgery and referral patterns (Guo and Wang, 2006). In 1993 data for the twin sample from which controls were drawn, out of 2973

<table>
<thead>
<tr>
<th>Number</th>
<th>dbSNP ID</th>
<th>Position</th>
<th>Relative position</th>
<th>Role</th>
<th>Alleles</th>
<th>Frequency</th>
<th>Association χ²</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs1860399</td>
<td>11928411</td>
<td>−8658</td>
<td>Promoter</td>
<td>C &gt; T</td>
<td>0.266</td>
<td>0.019</td>
<td>0.194</td>
</tr>
<tr>
<td>2</td>
<td>rs82613</td>
<td>11928517</td>
<td>−7591</td>
<td>Promoter</td>
<td>C &gt; T</td>
<td>0.341</td>
<td>1.691</td>
<td>0.060</td>
</tr>
<tr>
<td>3</td>
<td>rs82612</td>
<td>11928519</td>
<td>−6850</td>
<td>Promoter</td>
<td>A &gt; G</td>
<td>0.073</td>
<td>3.542</td>
<td>0.060</td>
</tr>
<tr>
<td>4</td>
<td>rs242956</td>
<td>11928746</td>
<td>−5300</td>
<td>Promoter</td>
<td>T &gt; C</td>
<td>0.371</td>
<td>1.418</td>
<td>0.234</td>
</tr>
<tr>
<td>5</td>
<td>rs703409</td>
<td>11928867</td>
<td>−4092</td>
<td>Promoter</td>
<td>C &gt; T</td>
<td>0.266</td>
<td>0.160</td>
<td>0.681</td>
</tr>
<tr>
<td>6</td>
<td>rs703411</td>
<td>11929057</td>
<td>−2197</td>
<td>Promoter</td>
<td>C &gt; T</td>
<td>0.065</td>
<td>2.774</td>
<td>0.096</td>
</tr>
<tr>
<td>7</td>
<td>rs1638626</td>
<td>11929144</td>
<td>−1325</td>
<td>Promoter</td>
<td>C &gt; T</td>
<td>0.265</td>
<td>1.972</td>
<td>0.275</td>
</tr>
<tr>
<td>8</td>
<td>rs2286629</td>
<td>11929168</td>
<td>−1085</td>
<td>Promoter</td>
<td>G &gt; T</td>
<td>0.065</td>
<td>3.309</td>
<td>0.069</td>
</tr>
<tr>
<td>9</td>
<td>rs385209</td>
<td>11929404</td>
<td>1277</td>
<td>Intron</td>
<td>C &gt; T</td>
<td>0.326</td>
<td>1.730</td>
<td>0.188</td>
</tr>
<tr>
<td>10</td>
<td>rs855769</td>
<td>11929408</td>
<td>1317</td>
<td>Intron</td>
<td>G &gt; A</td>
<td>0.267</td>
<td>0.147</td>
<td>0.702</td>
</tr>
<tr>
<td>11</td>
<td>rs365446</td>
<td>11929430</td>
<td>1536</td>
<td>Intron</td>
<td>G &gt; C</td>
<td>0.34</td>
<td>2.064</td>
<td>0.151</td>
</tr>
<tr>
<td>12</td>
<td>rs1929640</td>
<td>11929499</td>
<td>2226</td>
<td>Intron</td>
<td>C &gt; T</td>
<td>0.021</td>
<td>1.164</td>
<td>0.281</td>
</tr>
<tr>
<td>13</td>
<td>rs740734</td>
<td>11929736</td>
<td>4599</td>
<td>Intron</td>
<td>G &gt; C</td>
<td>0.23</td>
<td>0.065</td>
<td>0.799</td>
</tr>
<tr>
<td>14</td>
<td>rs555768</td>
<td>11929745</td>
<td>4688</td>
<td>Intron</td>
<td>G &gt; A</td>
<td>0.067</td>
<td>2.497</td>
<td>0.114</td>
</tr>
<tr>
<td>15</td>
<td>rs2240776</td>
<td>11929751</td>
<td>3–53</td>
<td>Intron (boundary)</td>
<td>T &gt; A</td>
<td>0.374</td>
<td>1.957</td>
<td>0.162</td>
</tr>
<tr>
<td>16</td>
<td>rs703413</td>
<td>11930136</td>
<td>8598</td>
<td>3’ UTR</td>
<td>C &gt; G</td>
<td>0.058</td>
<td>1.986</td>
<td>0.159</td>
</tr>
<tr>
<td>17</td>
<td>rs4751627</td>
<td>11930209</td>
<td>9324</td>
<td>3’ UTR</td>
<td>T &gt; C</td>
<td>0.365</td>
<td>1.965</td>
<td>0.161</td>
</tr>
<tr>
<td>18</td>
<td>rs242960</td>
<td>11930389</td>
<td>11128</td>
<td>3’ UTR</td>
<td>A &gt; G</td>
<td>0.202</td>
<td>0.386</td>
<td>0.353</td>
</tr>
<tr>
<td>19</td>
<td>rs8111280</td>
<td>11930514</td>
<td>12545</td>
<td>3’ UTR</td>
<td>T &gt; C</td>
<td>0.365</td>
<td>1.665</td>
<td>0.197</td>
</tr>
<tr>
<td>20</td>
<td>rs555766</td>
<td>11930618</td>
<td>13379</td>
<td>3’ UTR</td>
<td>A &gt; G</td>
<td>0.265</td>
<td>0.119</td>
<td>0.730</td>
</tr>
<tr>
<td>21</td>
<td>rs4752078</td>
<td>11930794</td>
<td>15025</td>
<td>3’ UTR</td>
<td>G &gt; C</td>
<td>0.384</td>
<td>2.389</td>
<td>0.122</td>
</tr>
<tr>
<td>22</td>
<td>rs4752079</td>
<td>11930885</td>
<td>16081</td>
<td>3’ UTR</td>
<td>A &gt; G</td>
<td>0.384</td>
<td>2.279</td>
<td>0.131</td>
</tr>
</tbody>
</table>

aMajor and minor allele on the transcribed strand.
bMinor allele frequency in controls.
cGenotypic association χ² with endometriosis.

<table>
<thead>
<tr>
<th>Number</th>
<th>dbSNP ID</th>
<th>Position</th>
<th>Relative position</th>
<th>Role</th>
<th>Alleles</th>
<th>Frequency</th>
<th>Association χ²</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs2673386</td>
<td>89629942</td>
<td>15735</td>
<td>Intron</td>
<td>G &gt; A</td>
<td>0.241</td>
<td>0.86</td>
<td>0.354</td>
</tr>
<tr>
<td>2</td>
<td>rs1234220</td>
<td>89635453</td>
<td>21246</td>
<td>Intron</td>
<td>T &gt; C</td>
<td>0.089</td>
<td>0.658</td>
<td>0.417</td>
</tr>
<tr>
<td>3</td>
<td>rs1234219</td>
<td>89635557</td>
<td>25350</td>
<td>Intron</td>
<td>T &gt; C</td>
<td>0.066</td>
<td>0.478</td>
<td>0.489</td>
</tr>
<tr>
<td>4</td>
<td>rs1909585</td>
<td>89643666</td>
<td>2-97</td>
<td>Intron (boundary)</td>
<td>T &gt; C</td>
<td>0.332</td>
<td>0.487</td>
<td>0.485</td>
</tr>
<tr>
<td>5</td>
<td>rs2299939</td>
<td>89647130</td>
<td>32923</td>
<td>Intron</td>
<td>A &gt; C</td>
<td>0.177</td>
<td>0.006</td>
<td>0.940</td>
</tr>
<tr>
<td>6</td>
<td>rs11202597</td>
<td>89655170</td>
<td>40963</td>
<td>Intron</td>
<td>A &gt; T</td>
<td>0.017</td>
<td>0.411</td>
<td>0.522</td>
</tr>
<tr>
<td>7</td>
<td>rs1234224</td>
<td>89665276</td>
<td>51069</td>
<td>Intron</td>
<td>C &gt; T</td>
<td>0.349</td>
<td>0.3</td>
<td>0.584</td>
</tr>
<tr>
<td>8</td>
<td>rs7353433</td>
<td>89695409</td>
<td>81202</td>
<td>Intron (boundary)</td>
<td>G &gt; C</td>
<td>0.338</td>
<td>0.075</td>
<td>0.785</td>
</tr>
<tr>
<td>9</td>
<td>rs17431184</td>
<td>89710231</td>
<td>96024</td>
<td>Intron (boundary)</td>
<td>T &gt; C</td>
<td>0.197</td>
<td>1.321</td>
<td>0.250</td>
</tr>
<tr>
<td>10</td>
<td>rs555895</td>
<td>89710887</td>
<td>8 + 32</td>
<td>Intron</td>
<td>G &gt; T</td>
<td>0.472</td>
<td>3.109</td>
<td>0.078</td>
</tr>
<tr>
<td>11</td>
<td>rs2736627</td>
<td>89711074</td>
<td>96867</td>
<td>Intron</td>
<td>T &gt; C</td>
<td>0.114</td>
<td>0.076</td>
<td>0.782</td>
</tr>
<tr>
<td>12</td>
<td>rs926961</td>
<td>89711392</td>
<td>97185</td>
<td>Intron</td>
<td>T &gt; C</td>
<td>0.151</td>
<td>0.006</td>
<td>0.937</td>
</tr>
<tr>
<td>13</td>
<td>rs532678</td>
<td>89713322</td>
<td>99115</td>
<td>Intron</td>
<td>T &gt; C</td>
<td>0.344</td>
<td>0.044</td>
<td>0.834</td>
</tr>
<tr>
<td>14</td>
<td>rs701848</td>
<td>89716725</td>
<td>102518</td>
<td>3’ UTR</td>
<td>T &gt; C</td>
<td>0.398</td>
<td>0.022</td>
<td>0.882</td>
</tr>
<tr>
<td>15</td>
<td>rs478839</td>
<td>89721850</td>
<td>107643</td>
<td>3’ UTR</td>
<td>G &gt; A</td>
<td>0.387</td>
<td>0.526</td>
<td>0.468</td>
</tr>
</tbody>
</table>

aMajor and minor allele on the transcribed strand.
bMinor allele frequency in controls.
cAssociation χ² with endometriosis.
We selected controls based on non-diagnosis, rather than on non-reporting symptoms that could have been indicative of endometriosis. This was for two main reasons. First, not everyone with symptoms like severe dysmenorrhea will have endometriosis, as it can be predictive for unrelated conditions such as pelvic adhesions (Forman et al., 1993). It is estimated that two-thirds of all patients who undergo diagnostic laparoscopy will not have endometriosis (Garry, 2006), as one-third will have no visible pathology and one-third other gynaecological conditions (Howard, 2000). Secondly, because the case set was unselected for symptoms, selecting controls using symptom criteria might have had unintended biasing effects in relation to the phenotypic characteristics of the case set. In summary, our control selection was the best compromise available to us on the ideal, but infeasible, means of estimating true prevalence by universal laparoscopic investigation of the reproductive-aged female population.

We estimate that we had much higher power to detect novel gene associations of small to moderate effect than a standard case–control study would have provided because our cases had a family history. Indeed, Antoniou and Easton (2003) showed that for multiplicative and dominant models, cases selected in terms of family history can provide approximately twice the power to detect association compared to unselected cases. For recessive models, cases with two affected sibs and cases with an affected parent and sib, respectively, provide ~100–150% and 18–43% more power to detect association compared with unselected cases. Although methylation studies in tissue samples may shed further light on the epigenetic influences of EMX2 or PTEN in endometriosis, we conclude it is unlikely that variants in these genes are responsible for initiation and/or development of endometriosis in women or account for the linkage signal on chromosome 10q.

Acknowledgements
We thank Dr Daniel T. O’Connor for confirmation of diagnosis and staging of disease from clinical records of 251 of cases; Barbara Haddon for co-ordination of family recruitment, blood and phenotype collection; Anjali Henders and Megan Campbell for managing sample processing and DNA preparation; Renée Mayne for DNA extraction and quantification; Leanne McNiell and Patricia Keith for blood processing; Olivia Zheng, David Smyth and Xiaping Lin for computing and database management. This study was supported by the Cooperative Research Centre for Discovery of Genes for Common Human Diseases, Australia (1997–2004) under the Australian Government’s Cooperative Research Centres programs, by Cerylid Biosciences Ltd and the National Health and Medical Research Council (NHMRC) of Australia (339430). G.W.M. and D.R.N. are supported by NHMRC Fellowships.

References
Purcell S, Cherny SS, Sham PC. Genetic Power Calculator; design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003;19:149–150.

Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002;70:425–434.


