Molecular Human Reproduction Vol.13, No.8 pp. 587–594, 2007
Advance Access publication June 11, 2007 doi:10.1093/molehr/gam023

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

Susan A. Treloar[1,5], Zhen Zhen Zhao[2], Lien Le[2], Krina T. Zondervan[3], Nicholas G. Martin[1], Stephen Kennedy[4], Dale R. Nyholt[1] and Grant W. Montgomery[2]

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

[5]Correspondence address. Tel: +61-7-33620229; Fax: +61-7-33620101; E-mail: susan.treloar@qimr.edu.au

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 dysmenorrhoea (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; Stefansson *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 (λ_s) 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 (Daftary and Taylor, 2004). *EMX2* is located at 119.3 megabase pairs (Mbp), just outside the 95% confidence region for our linkage peak on chromosome 10 (119.4–127.9 Mbp).

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 case. Disease severity was assessed retrospectively from medical records using the revised American Fertility Society (rAFS) classification system (American Fertility Society, 1985; Treloar *et al*., 2002). Thirty-nine percent of cases were classified with stage B (rAFS stages III–IV) endometriosis, which almost by definition were more likely to have ovarian endometriotic cysts; the remaining cases had stage A (rAFS stages I–II) disease. A total of 645 cases (84%) were diagnosed at laparoscopy; the remaining cases were mostly diagnosed at hysterectomy, or in a small number of cases at laparotomy or at another procedure.

The controls were 768 unrelated women who had volunteered for a twin study of gynaecological health (Treloar *et al*., 1999b). Twins had been asked simply 'Have you had endometriosis?' (Treloar *et al*., 1999b). They were also asked whether they had ever had a laparoscopy and/or a hysterectomy and the reasons for each. A small number ($n = 59$) also provided relevant information in a study of twins aged over 50 (Kirk *et al*., 1999). These controls were selected as women at low risk for endometriosis because, on the basis of self-reporting and their medical records where available, they had never been diagnosed with the disease. The numbers of controls who reported having had a hysterectomy or a laparoscopy, at which no endometriosis was found, were 111 and 102, respectively.

The mean ages ($\pm$ 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, $P < .01$). 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, rs82612, rs242956, rs703409, rs703411, rs1638626 and rs2286629), 6 in the intronic region (rs385209, 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, rs11202597, rs1234213, rs1234224, rs2735343, rs17431184, rs2736627, rs926091 and rs532678), 2 at 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, Sequenom™, San Diego, CA, USA) and assembled into multiplex sets. SNPs were typed using the Sequenom™ 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 $MgCl_2$, 200 μM dNTP, forward and reverse primers to a final concentration of 50 nM for all SNPs and 0.1 U/μl 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 the 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 U/μl 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-hydroxypicolinic acid matrix, to help bind and stabilize DNA during ionization by the MassARRAY. Loading of the products was performed using a nanolitre pipetting system (MassARRAY™; Sequenom) and reaction products analysed through MALDI-TOF MassARRAY Typer 3.0.1 (Sequenom).

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 Haploview (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 (α) 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 (α) of $P = 0.00294$ 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 0.25 and 0.5, 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 (Daftary and Taylor, 2004). Expression in the endometrium generally declines around the time of implantation, but it remains elevated in women with endometriosis (Daftary 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 (Daftary and Taylor, 2004). The pathway is also implicated because *HOXA10* regulates *EMX2* expression (Daftary 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

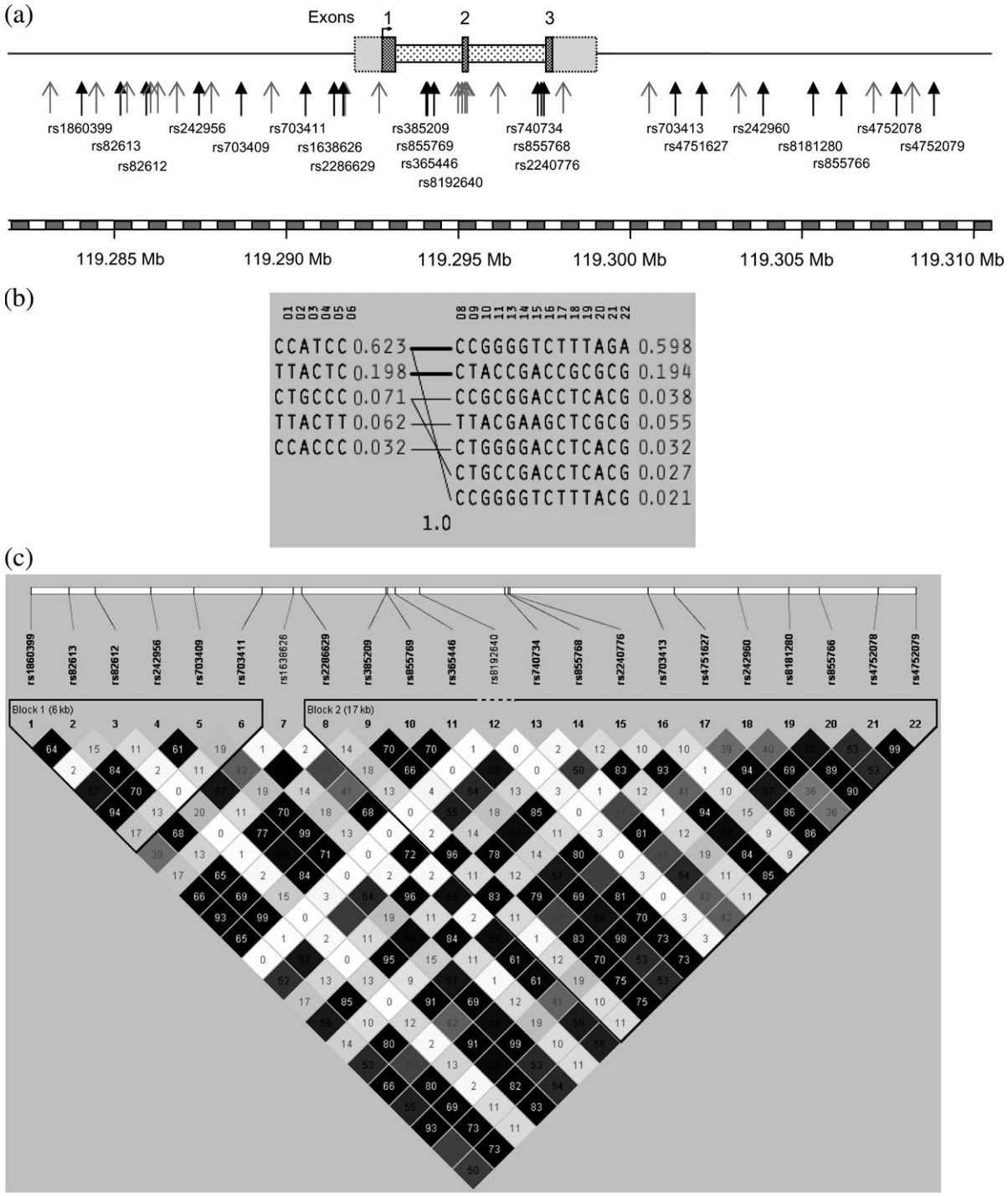

Figure 1: Variation in the human *EMX2* gene.
(**a**) The genomic structure of the *EMX2* gene showing the location of the 22 SNPs genotyped. Grey arrows represent SNPs that were excluded from final analyses because they were monomorphic, had allele frequencies $<5\%$ or results did not give us confidence 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)

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%

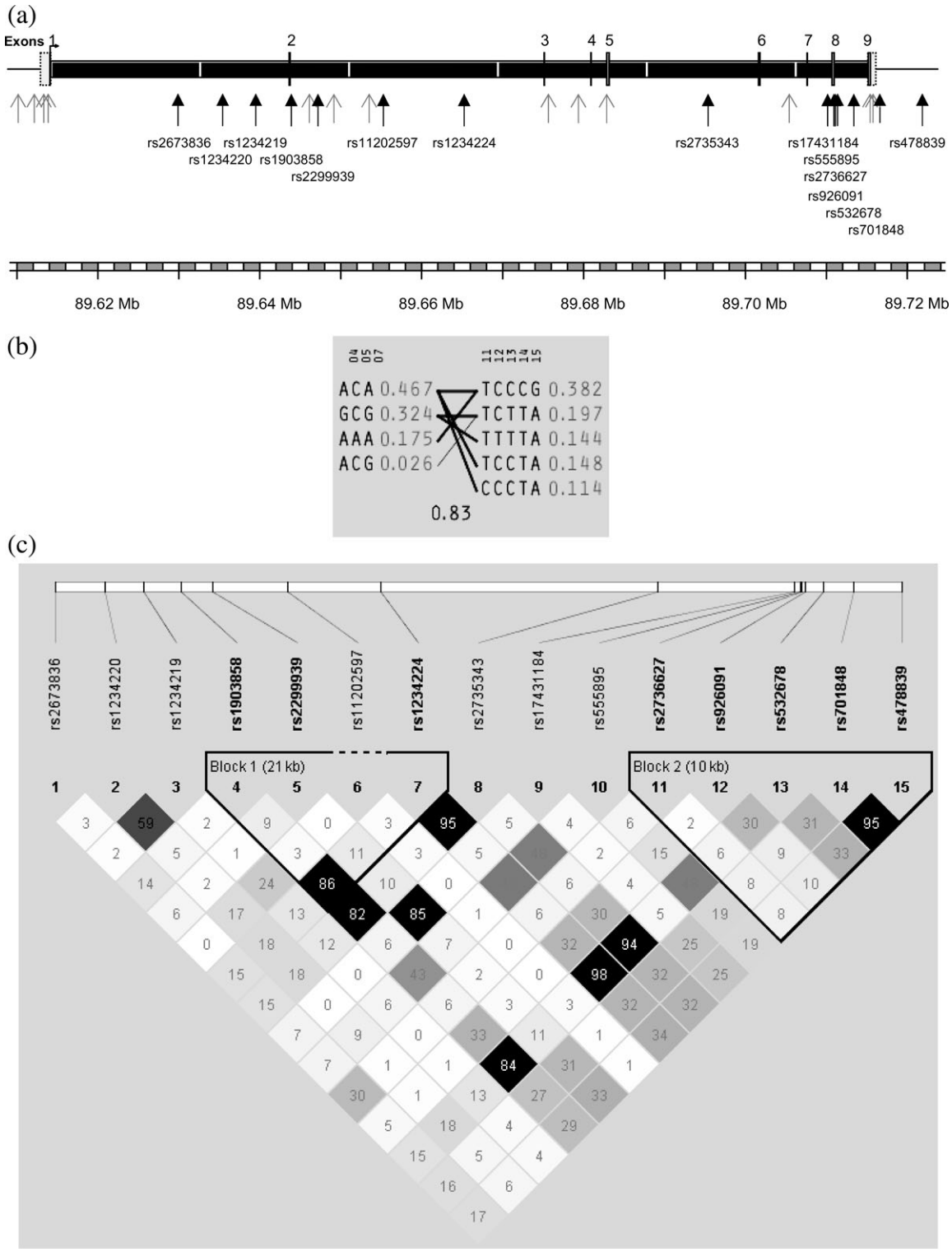

Figure 2: Variation in the human *PTEN* gene.
(**a**) The genomic structure of the *PTEN* gene showing the location of the 15 SNPs genotyped. Grey arrows represent SNPs that were excluded from final analyses because they were monomorphic, had allele frequencies $<5\%$ or results did not give us confidence in the assays. (**b**) Common haplotype blocks in *PTEN*. Dashed lines indicate locations where transition from one common haplotype to a different one is observed. (**c**) LD plot of *PTEN* SNPs; LD measured as r^2 using Haploview (Barrett et al., 2005). Grey shading represents correlation magnitudes between low r^2 (white) and high r^2 (black)

(0.05). 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

Table 1: SNPs genotyped across the *EMX2* locus

Number	dbSNP ID	Position	Relative position	Role	Alleles[a]	Frequency[b]	Association χ^{2c}	*P*-value
1	rs1860399	119284111	−8658	Promoter	C > T	0.266	0.019	0.890
2	rs82613	119285178	−7591	Promoter	C > T	0.341	1.691	0.194
3	rs82612	119285919	−6850	Promoter	A > G	0.073	3.542	0.060
4	rs242956	119287469	−5300	Promoter	T > C	0.371	1.418	0.234
5	rs703409	119288677	−4092	Promoter	C > T	0.266	0.169	0.681
6	rs703411	119290572	−2197	Promoter	C > T	0.065	2.774	0.096
7	rs1638626	119291444	−1325	Promoter	C > T	0.265	1.192	0.275
8	rs2286629	119291684	−1085	Promoter	G > T	0.065	3.309	0.069
9	rs385209	119294046	1277	Intron	C > T	0.326	1.730	0.188
10	rs855769	119294086	1317	Intron	G > A	0.267	0.147	0.702
11	rs365446	119294305	1536	Intron	G > C	0.34	2.064	0.151
12	rs8192640	119294995	2226	Intron	C > T	0.021	1.164	0.281
13	rs740734	119297368	4599	Intron	G > C	0.23	0.065	0.799
14	rs855768	119297457	4688	Intron	G > A	0.067	2.497	0.114
15	rs2240776	119297514	3–53	Intron (boundary)	T > A	0.374	1.957	0.162
16	rs703413	119301367	8598	3′ UTR	C > G	0.058	1.986	0.159
17	rs4751627	119302093	9324	3′ UTR	T > C	0.365	1.965	0.161
18	rs242960	119303897	11128	3′ UTR	A > G	0.202	0.386	0.535
19	rs8181280	119305314	12545	3′ UTR	T > C	0.365	1.665	0.197
20	rs855766	119306148	13379	3′ UTR	A > G	0.265	0.119	0.730
21	rs4752078	119307794	15025	3′ UTR	G > C	0.384	2.389	0.122
22	rs4752079	119308850	16081	3′ UTR	A > G	0.384	2.279	0.131

[a]Major and minor allele on the transcribed strand.
[b]Minor allele frequency in controls.
[c]Genotypic association χ^2 with endometriosis.

Table 2: SNPs genotyped across the *PTEN* locus

Number	dbSNP ID	Position	Relative position	Role	Alleles[a]	Frequency[b]	Association χ^{2c}	*P*-value
1	rs2673836	89629942	15735	Intron	G > A	0.241	0.86	0.354
2	rs1234220	89635453	21246	Intron	T > C	0.089	0.658	0.417
3	rs1234219	89639557	25350	Intron	T > C	0.066	0.478	0.489
4	rs1903858	89643666	2-97	Intron (boundary)	T > C	0.332	0.487	0.485
5	rs2299939	89647130	32923	Intron	A > C	0.177	0.006	0.940
6	rs11202597	89655170	40963	Intron	A > T	0.017	0.411	0.522
7	rs1234224	89665276	51069	Intron	C > T	0.349	0.3	0.584
8	r2735343	89695409	81202	Intron	G > C	0.338	0.075	0.785
9	rs17431184	89710231	96024	Intron (boundary)	T > C	0.197	1.321	0.250
10	rs555895	89710887	8+32	Intron	G > T	0.472	3.109	0.078
11	rs2736627	89711074	96867	Intron	T > C	0.114	0.076	0.782
12	rs926091	89711392	97185	Intron	T > C	0.151	0.006	0.937
13	rs532678	89713322	99115	Intron	T > C	0.344	0.044	0.834
14	rs701848	89716725	102518	3′ UTR	T > C	0.398	0.022	0.882
15	rs478839	89721850	107643	3′ UTR	G > A	0.387	0.526	0.468

[a]Major and minor allele on the transcribed strand.
[b]Minor allele frequency in controls.
[c]Association χ^2 with endometriosis.

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 endometriosis and common variants in *PTEN*. An in depth sequence approach to the *PTEN* gene would be necessary to detect allelic variants that were large in number but individually rare.

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 question respondents, 21 negative endometriosis self-reports were contradicted by medical/pathology reports, giving a kappa (κ) of 0.94 (95% CI 0.92–0.96) (Treloar *et al.*, 1999b). This suggests strong agreement, but there may well have been other undiagnosed cases in this sample, because validation was only actively sought for positive self-reports of hysterectomy and endometriosis. Clinical and/or pathology information negating a diagnosis of endometriosis was therefore available only for some twins who had had a hysterectomy. These 21 twins were excluded from our control set. Endometriosis may remain undiagnosed even if there are signs and/or symptoms. In one Australian case series, 155 (22%) patients had no symptoms, most being diagnosed coincidentally during surgical procedures for other conditions, but about half of these had shown clinical signs (O'Connor, 1987, p. 71). It is a complex question to disentangle and quantify. Although our control sample may have contained some undiagnosed cases, screening is likely to have reduced that number.

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 dysmenorrhoea 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 McNeill 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 program, 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

Abecasis GR, Ghosh D, Nichols TE. Linkage disequilibrium: ancient history drives the new genetics. *Hum Hered* 2005;**59**:118–124.

American Fertility Society. Revised American Fertility Society classification of endometriosis: 1985. *Fertil Steril* 1985;**43**:351–352.

Antoniou AC, Easton DF. Polygenic inheritance of breast cancer: implications for design of association studies. *Genet Epidemiol* 2003;**25**:190–202.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;**21**:263–265.

Bischoff F, Simpson JL. Genetics of endometriosis: heritability and candidate genes. *Best Pract Res Clin Obstet Gynaecol* 2004;**18**:219–232.

Daftary GS, Taylor HS. EMX2 gene expression in the female reproductive tract and aberrant expression in the endometrium of patients with endometriosis. *J Clin Endocrinol Metab* 2004;**89**:2390–2396.

Dinulescu DM, Ince TA, Quade BJ, Shafer SA, Crowley D, Jacks T. Role of K-ras and Pten in the development of mouse models of endometriosis and endometrioid ovarian cancer. *Nat Med* 2005;**11**:63–70.

Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* 2003;**25**:115–121.

Eskenazi B, Warner ML. Epidemiology of endometriosis. *Obstet Gynecol Clin North Am* 1997;**24**:235–258.

Flores I, Mandal DM, Bailey-Wilson JE. *Search for Endometriosis Susceptibility Genes in Puerto Rico*. Toronto: American Society of Human Genetics, 2004.

Forman RG, Robinson JN, Mehta Z, Barlow DH. Patient history as a simple predictor of pelvic pathology in subfertile women. *Hum Reprod* 1993;**8**:53–55.

Fujii H, Matsumoto T, Yoshida M, Furugen Y, Takagaki T, Iwabuchi K, Nakata Y, Takagi Y, Moriya T, Ohtsuji N *et al.* Genetics of synchronous uterine and ovarian endometrioid carcinoma: combined analyses of loss of heterozygosity, PTEN mutation, and microsatellite instability. *Hum Pathol* 2002;**33**:421–428.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M *et al.* The structure of haplotype blocks in the human genome. *Science* 2002;**296**:2225–2229.

Gangemi RM, Daga A, Marubbi D, Rosatto N, Capra MC, Corte G. Emx2 in adult neural precursor cells. *Mech Dev* 2001;**109**:323–329.

Garry R. Diagnosis of endometriosis and pelvic pain. *Fertil Steril* 2006;**86**:1307–1309.

Giudice LC, Kao LC. Endometriosis. *Lancet* 2004;**364**:1789–1799.

Guo SW, Wang Y. Sources of heterogeneities in estimating the prevalence of endometriosis in infertile and previously fertile women. *Fertil Steril* 2006;**86**:1584–1595.

Howard FM. The role of laparoscopy as a diagnostic tool in chronic pelvic pain. *Baillieres Best Pract Res Clin Obstet Gynaecol* 2000;**14**:467–494.

Kennedy S. Genetics of endometriosis: a review of the positional cloning approaches. *Semin Reprod Med* 2003;**21**:111–118.

Kennedy S, Mardon H, Barlow D. Familial endometriosis. *J Assist Reprod Genet* 1995;**12**:32–34.

Kirk KM, Hickie IB, Martin NG. Fatigue as related to anxiety and depression in a community-based sample of twins aged over 50. *Soc Psychiatry Psychiatr Epidemiol* 1999;**34**:85–90.

Kurose K, Zhou XP, Araki T, Cannistra SA, Maher ER, Eng C. Frequent loss of PTEN expression is linked to elevated phosphorylated Akt levels, but not associated with p27 and cyclin D1 expression, in primary epithelial ovarian carcinomas. *Am J Pathol* 2001;**158**:2097–2106.

Kurose K, Gilley K, Matsumoto S, Watson PH, Zhou XP, Eng C. Frequent somatic mutations in PTEN and TP53 are mutually exclusive in the stroma of breast carcinomas. *Nat Genet* 2002;**32**:355–357.

Latta E, Chapman WB. PTEN mutations and evolving concepts in endometrial neoplasia. *Curr Opin Obstet Gynecol* 2002;**14**:59–65.

Lin WM, Forgacs E, Warshal DP, Yeh IT, Martin JS, Ashfaq R, Muller CY. Loss of heterozygosity and mutational analysis of the PTEN/MMAC1 gene in synchronous endometrial and ovarian carcinomas. *Clin Cancer Res* 1998;**4**:2577–2583.

Martini M, Ciccarone M, Garganese G, Maggiore C, Evangelista A, Rahimi S, Zannoni G, Vittori G, Larocca LM. Possible involvement of hMLH1, p16(INK4a) and PTEN in the malignant transformation of endometriosis. *Int J Cancer* 2002;**102**:398–406.

Maxwell GL, Risinger JI, Gumbs C, Shaw H, Bentley RC, Barrett JC, Berchuck A, Futreal PA. Mutation of the PTEN tumor suppressor gene in endometrial hyperplasias. *Cancer Res* 1998;**58**:2500–2503.

Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucl Acids Res* 1988;**16**:1215.

Mutter GL, Lin MC, Fitzgerald JT, Kum JB, Eng C. Changes in endometrial PTEN expression throughout the human menstrual cycle. *J Clin Endocrinol Metab* 2000;**85**:2334–2338.

Noonan FC, Mutch DG, Ann Mallon M, Goodfellow PJ. Characterization of the homeodomain gene EMX2: sequence conservation, expression analysis, and a search for mutations in endometrial cancers. *Genomics* 2001;**76**:37–44.

Noonan FC, Goodfellow PJ, Staloch LJ, Mutch DG, Simon TC. Antisense transcripts at the EMX2 locus in human and mouse. *Genomics* 2003;**81**:58–66.

Nyholt DR. A simple correction for multiple testing for SNPs in linkage disequilibrium with each other. *Am J Hum Genet* 2004;**74**:765–769.

Obata K, Morland SJ, Watson RH, Hitchcock A, Chenevix-Trench G, Thomas EJ, Campbell IG. Frequent PTEN/MMAC mutations in endometrioid but not serous or mucinous epithelial ovarian tumors. *Cancer Res* 1998;**58**:2095–2097.

O'Connor DT. *Endometriosis*. Edinburgh: Churchill Livingstone, 1987.

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.

Sato N, Tsunoda H, Nishida M, Morishita Y, Takimoto Y, Kubo T, Noguchi M. Loss of heterozygosity on 10q23.3 and mutation of the tumor suppressor gene PTEN in benign endometrial cyst of the ovary: possible sequence progression from benign endometrial cyst to endometrioid carcinoma and clear cell carcinoma of the ovary. *Cancer Res* 2000;**60**:7052–7056.

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.

Simeone A, Acampora D, Gulisano M, Stornaiuolo A, Boncinelli E. Nested expression domains of four homeobox genes in developing rostral brain. *Nature* 1992a;**358**:687–690.

Simeone A, Gulisano M, Acampora D, Stornaiuolo A, Rambaldi M, Boncinelli E. Two vertebrate homeobox genes related to the Drosophila empty spiracles gene are expressed in the embryonic cerebral cortex. *EMBO J* 1992b;**11**:2541–2550.

Simpson JL, Bischoff FZ. Heritability and molecular genetic studies of endometriosis. *Ann NY Acad Sci* 2002;**955**:239–251.

Simpson JL, Bischoff FZ, Kamat A, Buster JE, Carson SA. Genetics of endometriosis. *Obstet Gynecol Clin North Am* 2003;**30**:21–40, vii.

SPSS Inc. SPSS 13.0 for Windows. Chicago, IL, 2004.

Stefansson H, Geirsson RT, Steinthorsdottir V, Jonsson H, Manolescu A, Kong A, Ingadottir G, Gulcher J, Stefansson K. Genetic factors contribute to the risk of developing endometriosis. *Hum Reprod* 2002;**17**:555–559.

Swiersz LM. Role of endometriosis in cancer and tumor development. *Ann NY Acad Sci* 2002;**955**:281–292.

Taylor HS, Fei X. Emx2 regulates mammalian reproduction by altering endometrial cell proliferation. *Mol Endocrinol* 2005;**19**:2839–2846.

The International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;**437**:1299–1320.

Treloar SA, Do KA, O'Connor VM, O'Connor DT, Yeo MA, Martin NG. Predictors of hysterectomy: an Australian study. *Am J Obstet Gynecol* 1999a;**180**:945–954.

Treloar SA, O'Connor DT, O'Connor VM, Martin NG. Genetic influences on endometriosis in an Australian twin sample. *Fertil Steril* 1999b; **71**:701–710.

Treloar S, Hadfield R, Montgomery G, Lambert A, Wicks J, Barlow DH, O'Connor DT, Kennedy S, Group IES. The International Endogene Study: a collection of families for genetic research in endometriosis. *Fertil Steril* 2002;**78**:679–685.

Treloar SA, Wicks J, Nyholt DR, Montgomery GW, Bahlo M, Smith V, Dawson G, Mackay IJ, Weeks DE, Bennett ST *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**:365–376.

Troy PJ, Daftary GS, Bagot CN, Taylor HS. Transcriptional repression of peri-implantation EMX2 expression in mammalian reproduction by HOXA10. *Mol Cell Biol* 2003;**23**:1–13.

Wicks J, Treloar SA, Martin NG. Using identity-by-descent information in affected sib pairs to increase the efficiency of genetic association studies. *Twin Res* 2004;**7**:211–216.

Wigginton JE, Abecasis GR. PEDrSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* 2005;**21**:3445–3447.

Wu Y, Halverson G, Basir Z, Strawn E, Yan P, Guo SW. Aberrant methylation at HOXA10 may be responsible for its aberrant expression in the endometrium of patients with endometriosis. *Am J Obstet Gynecol* 2005;**193**:371–380.

Zhang K, Calabrese P, Nordborg M, Sun F. Haplotype block structure and its applications to association studies: power and study designs. *Am J Hum Genet* 2002;**71**:1386–1394.

Zhou XP, Kuismanen S, Nystrom-Lahti M, Peltomaki P, Eng C. Distinct PTEN mutational spectra in hereditary non-polyposis colon cancer syndrome-related endometrial carcinomas compared to sporadic microsatellite unstable tumors. *Hum Mol Genet* 2002;**11**:445–450.

Zondervan KT, Weeks DE, Colman R, Cardon L, Hadfield RH, Scheffler J, Trainor AG, Coe CL, Kemnitz JW, Kennedy SH. Familial aggregation of endometriosis in a large pedigree of rhesus macaques. *Hum Reprod* 2004;**19**:448–455.

Submitted on January 11, 2007; resubmitted on March 4, 2007; accepted on March 8, 2007