

Functional evaluation of genetic variants associated with endometriosis near *GREB1*

Jenny N. Fung^{1,*†}, Sarah J. Holdsworth-Carson^{†,2}, Yadav Sapkota¹, Zhen Zhen Zhao¹, Lincoln Jones¹, Jane E. Girling², Premila Paiva², Martin Healey², Dale R. Nyholt¹, Peter A. W. Rogers², and Grant W. Montgomery¹

¹Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane QLD 4029, Australia ²Department of Obstetrics and Gynaecology, Gynaecology Research Centre, University of Melbourne, Royal Women's Hospital, Parkville VIC 3052, Australia

*Correspondence address. QIMR Berghofer Medical Research Institute, Locked Bag 2000, Royal Brisbane Hospital, Herston, Queensland 4029, Australia. Tel: +61-7-3362-0186; E-mail: jenny.fung@qimrberghofer.edu.au

Submitted on November 25, 2014; resubmitted on February 5, 2015; accepted on February 12, 2015

STUDY QUESTION: Do DNA variants in the growth regulation by estrogen in breast cancer 1 (*GREB1*) region regulate endometrial *GREB1* expression and increase the risk of developing endometriosis in women?

SUMMARY ANSWER: We identified new single nucleotide polymorphisms (SNPs) with strong association with endometriosis at the *GREB1* locus although we did not detect altered *GREB1* expression in endometriosis patients with defined genotypes.

WHAT IS ALREADY KNOWN: Genome-wide association studies have identified the *GREB1* region on chromosome 2p25.1 for increasing endometriosis risk. The differential expression of *GREB1* has also been reported by others in association with endometriosis disease phenotype.

STUDY DESIGN, SIZE, DURATION: Fine mapping studies comprehensively evaluated SNPs within the *GREB1* region in a large-scale data set (>2500 cases and >4000 controls). Publicly available bioinformatics tools were employed to functionally annotate SNPs showing the strongest association signal with endometriosis risk. Endometrial *GREB1* mRNA and protein expression was studied with respect to phases of the menstrual cycle ($n = 2-45$ per cycle stage) and expression quantitative trait loci (eQTL) analysis for significant SNPs were undertaken for *GREB1* [mRNA ($n = 94$) and protein ($n = 44$) in endometrium].

PARTICIPANTS/MATERIALS, SETTING, METHODS: Participants in this study are females who provided blood and/or endometrial tissue samples in a hospital setting. The key SNPs were genotyped using Sequenom MassARRAY. The functional roles and regulatory annotations for identified SNPs are predicted by various publicly available bioinformatics tools. Endometrial *GREB1* expression work employed qRT-PCR, western blotting and immunohistochemistry studies.

MAIN RESULTS AND THE ROLE OF CHANCE: Fine mapping results identified a number of SNPs showing stronger association ($0.004 < P < 0.032$) with endometriosis risk than the original GWAS SNP (rs13394619) ($P = 0.034$). Some of these SNPs were predicted to have functional roles, for example, interaction with transcription factor motifs. The haplotype (a combination of alleles) formed by the risk alleles from two common SNPs showed significant association ($P = 0.026$) with endometriosis and epistasis analysis showed no evidence for interaction between the two SNPs, suggesting an additive effect of SNPs on endometriosis risk. In normal human endometrium, *GREB1* protein expression was altered depending on the cycle stage (significantly different in late proliferative versus late secretory, $P < 0.05$) and cell type (glandular epithelium, not stromal cells). However, *GREB1* expression in endometriosis cases versus controls and eQTL analyses did not reveal any significant changes.

LIMITATIONS, REASONS FOR CAUTION: In silico prediction tools are generally based on cell lines different to our tissue and disease of interest. Functional annotations drawn from these analyses should be considered with this limitation in mind. We identified cell-specific and hormone-specific changes in *GREB1* protein expression. The lack of a significant difference observed following our *GREB1* expression studies may be the result of moderate power on mixed cell populations in the endometrial tissue samples.

[†] The authors consider that the first two authors should be regarded as joint First Authors.

WIDER IMPLICATIONS OF THE FINDINGS: This study further implicates the *GREB1* region on chromosome 2p25.1 and the *GREB1* gene with involvement in endometriosis risk. More detailed functional studies are required to determine the role of the novel *GREB1* transcripts in endometriosis pathophysiology.

STUDY FUNDING/COMPETING INTEREST(S): Funding for this work was provided by NHMRC Project Grants APP1012245, APP1026033, APP1049472 and APP1046880. There are no competing interests.

Key words: endometriosis / *GREB1* / eQTLs / GWAS

Introduction

Endometriosis is a common gynaecological disease affecting 6–10% of reproductive-aged women (Burney and Giudice, 2012) and the most common cause of chronic pelvic pain (Hickey et al., 2014). Other symptoms include dysmenorrhea, dyspareunia, irregular uterine bleeding and reduced fertility. Given the breadth of symptoms and varying degrees of severity, endometriosis is under-diagnosed and associated with significant diagnostic delays, ranging from 5 to 9 years between symptom onset and definitive diagnosis (Culley et al., 2013). The financial burden of endometriosis on the healthcare system is substantial, with estimates of \$AU6 billion per annum (Bush et al., 2011). There is currently no cure for endometriosis. Following surgical management, symptomatic recurrence ranges from 20 to 40% often requiring additional surgery at some stage (Vercellini et al., 2009).

There is an urgent need for a better understanding of endometriosis pathogenesis and pathophysiology. Endometriosis is a complex multifactorial disease; however, several studies report a genetic basis to the disease (Moen and Magnus, 1993; Kennedy et al., 1995; Treloar et al., 1999). Genome-wide association studies (GWAS) for endometriosis to date have identified 10 genomic regions at a genome-wide significant level ($P < 5 \times 10^{-8}$) (Uno et al., 2010; Painter et al., 2011; Nyholt et al., 2012; Rahmioglu et al., 2014; Sapkota et al., 2015). The next critical step to enable translation of these results is to identify the specific genes and pathways contributing to endometriosis risk from these regions and to characterize their functional effects. The GWAS meta-analysis published by Nyholt et al. (2012), identified rs13394619 as the sentinel SNP (SNP with the strongest association signal). This SNP is located in an intronic region between exon 9 and exon 10 in the growth regulation by estrogen in breast cancer 1 (*GREB1*) gene. *GREB1* is an estrogen-regulated gene first identified in breast cancer cell lines and tumours (Ghosh et al., 2000). Knock down of *GREB1* inhibits estrogen induced growth in MCF-7 breast cancer cell lines (Rae et al., 2005). Estrogen receptor α (ER α) binds to and induces *GREB1* transcription via three estrogen response elements (EREs) located ~20 kb upstream of the gene (Deschenes et al., 2007; Sun et al., 2007). Overall ER-regulated transcription is highly complex and involves over 100 ER-associated proteins. *GREB1* is the most estrogen-dependent ER interactor where it functions as an essential component of the ER transcription complex (Mohammed et al., 2013). In ER-positive breast tumours, *GREB1* expression correlates closely with estrogen levels across the menstrual cycle (Haynes et al., 2013). Evidence for a relationship between *GREB1* and estrogen (ER) is well established; therefore, it is highly plausible that *GREB1* plays a role in an estrogen-dependent disease such as endometriosis.

The majority of SNPs associated with complex diseases identified by GWAS are located in non-coding regions of the genome and are likely to exert their effect by influencing transcriptional output through multiple

mechanisms (such as transcript levels and splicing) (Freedman et al., 2011). Gene expression is highly heritable and many SNPs have functional impacts on gene expression (Morley et al., 2004; Lappalainen et al., 2013). Differential expression of *GREB1* (mRNA and protein) has been reported in endometriosis with increased expression in ectopic endometriotic lesions and eutopic endometrium from cases when compared with eutopic endometrium from control women (Pellegrini et al., 2012). Consequently, *GREB1* is a strong candidate for the target gene in this region and genetic variants at chromosome 2p25.1 associated with endometriosis could function by changing the expression of *GREB1* transcripts.

The aim of this study was to evaluate SNP variation and the role of *GREB1* in increased risk of endometriosis. We first evaluated the role of genetic variants within the coding region of *GREB1*, and performed a high-resolution fine-mapping analysis to comprehensively evaluate the association of the genetic variations across the *GREB1* locus at 2p25.1. We then analysed gene and protein expression for *GREB1* in samples of endometrium from women with and without endometriosis and the effects of SNP variation on expression of *GREB1*.

Materials and Methods

Datasets for fine mapping

Fine mapping in the region of *GREB1* was conducted in overlapping sample sets drawn from the Australian Endometriosis Study recruited by The Queensland Institute of Medical Research (Treloar et al., 2002; Zhao et al., 2006; Painter et al., 2011). All participants provided written informed consent for inclusion in this study. Study protocols were reviewed and approved by the QIMR Berghofer Human Research Ethics Committee and the Australian Twin Registry.

Human endometrial tissues (Endometrial tissue project)

The study included 102 women of European descent who were recruited from the Royal Women's Hospital in Melbourne. Women undergoing laparoscopic surgery for pelvic pain provided written consent prior to the operation. Only premenopausal women who were free from hormone treatment (in the 3 months prior to surgery) were included in this study. Detailed patient questionnaires, past and present clinical histories, pathology findings and surgical notes were recorded for each participant. Endometrial cycle stage was determined following histological assessment at pathology. Endometrial tissue samples were collected by curettage from women with surgically diagnosed endometriosis ($n = 66$) and controls with no known history of endometriosis ($n = 36$). Of the 66 cases; $n = 33$ were Stage 1, $n = 6$ were Stage 2, $n = 9$ were Stage 3, $n = 5$ were Stage 4 and $n = 13$ cases remained undetermined. Samples were stored in RNAlater (Life Technologies, Grand Island, NY, USA) at -80°C until RNA and/or protein extraction. Paraffin

blocks of curettes were recalled from pathology following completion of diagnostic studies for immunohistochemistry (IHC). Whole blood was also collected to investigate the effect of SNPs located within the chromosome 2 locus on expression levels of *GREB1* transcripts in endometrial tissues. The study was approved by the Human Ethics Committees of the Royal Women's Hospital in Melbourne and the QIMR Berghofer Medical Research Institute.

RNA extraction and cDNA synthesis

Total RNA was extracted from homogenized endometrial tissues using RNA lysis solution (RLT buffer) and RNeasy Plus Mini Kit according to the manufacturer's instructions (QIAGEN, Valencia, CA, USA). RNA quality was assessed with the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA), and concentrations were determined using the NanoDropND-6000. Samples of 800 ng of RNA were converted to cDNA using a Qiagen kit and preamplified using Taqman preamplification kit (Life Technologies). The primers for the six targeted transcripts of *GREB1* were designed using Primer3, version 0.4.0 (<http://bioinfo.ut.ee/primer3-0.4.0/>, 2 October 2013, date last accessed). Refer to Fig. 1 for primer locations and Supplementary data, Table S1 for primer sequences.

Fluidigm qPCR

The quantitative reverse-transcriptase polymerase chain reaction (qRT-PCR) was performed using the Fluidigm BioMark HD platform (Fluidigm Corporation, San Francisco, CA, USA). Preamplified cDNA samples were loaded into the DynamicArray 48.48 chips. Evagreen gene expression assays were diluted and pipetted out into the DynamicArray chips. The chips were then placed into the NanoFlex controller. All qRT-PCR reactions were performed in the BioMark Real-Time PCR system, immediately followed by melting curve analysis. To minimize potential batch effect in this high-throughput experiment, three repeated reference genes (*GAPDH*, *PPIA* and *ACTB*) were targeted in each run to serve as the internal controls.

Genotyping

For the fine-mapping study, additional genotyping was conducted using HumanCoreExome genotyping chips (Illumina Inc, San Diego), including ~240 000 common markers and 240 000 coding variants available. These variants were genotyped on DNA samples from the Australian study including 2213 surgically confirmed endometriosis cases and 2044 controls (Painter *et al.*, 2011). We also genotyped 76 variants (the top genotyped SNP, the 7 key imputed SNPs identified from the GWAS meta-analysis, 47 coding and 21 potential functional variants) across the 2p25.1 locus in a subset of samples, including 960 of the surgically confirmed cases with a family history of endometriosis and a control group of 960 unrelated women with self-reports of no endometriosis (Zhao *et al.*, 2006).

For the expression quantitative trait loci (eQTLs) study, a total of five variants (the original GWAS SNP, the new sentinel non-coding SNP from the fine mapping, two coding variants and a SNP with strong evidence for functional roles) were genotyped in blood DNA samples from individuals in the endometrial tissue project. The genotyping was conducted with iPLEX chemistry using Sequenom MassARRAY technology (Sequenom Inc., San Diego, CA, USA).

In silico analysis

Potential functional roles and regulatory annotations for identified SNPs were investigated using HaploReg v2 (Broad Institute), UCSC Genome Browser and the ENCODE database. The RegulomeDB program was employed to rank the potential functional roles for SNPs based on their location across the entire 107 kb region. The scoring system for RegulomeDB ranges from 1 to 6 with 1 being the strongest evidence for functional roles (Boyle *et al.*,

2012). Linkage disequilibrium (LD) for the genotyped SNPs was examined using SNAP Pairwise LD (Broad Institute).

Immunohistochemistry

Initial menstrual cycle stage of *GREB1* expression were performed on full thickness endometrium and myometrial tissues which were obtained following consent from a separate group of premenopausal women undergoing hysterectomy (for prolapse, fibroids or adenomyosis) ($n = 35$) at the Royal Women's Hospital (Melbourne). These women had no history of endometriosis. Full thickness tissues were immediately fixed in formalin for IHC.

IHC was performed as per our previously published protocols (Holdsworth-Carson *et al.*, 2014). Briefly, paraffin sections (3 μ m) were dewaxed and dehydrated. Sections then underwent antigen retrieval (boiling citrate buffer pH6.0, 15 min). Incubation with primary antibody, mouse monoclonal *GREB1* (clone *GREB1Ab*) (Millipore, Temecula, CA USA), occurred overnight at 4°C (at 1 μ g/ml). A mouse IgG1 isotype control matched for concentration was also performed (Dako). Sections were incubated with anti-mouse EnVision+ System-Horse Radish Peroxidase (HRP) (Dako) (30 min at RT), followed by 3,3'-diaminobenzidine (Sigma-Aldrich) for 5 min at RT. Sections were counterstained briefly with haematoxylin. Images were acquired on an Aperio ScanScopeXT system (Aperio, CA, USA) at the Melbourne Health Tissue Bank (Royal Melbourne Hospital).

Protein extraction

A minimum of 10 mg of frozen endometrium was used per patient for whole cell protein extraction. To eliminate red blood cells, endometrium was incubated with 250 μ l of ammonium-chloride-potassium (ACK) lysing buffer (Life Technologies) with phosphatase inhibitor cocktail 2 (at a 1:200 dilution) and protease inhibitor cocktail (at a 1:1000 dilution) (both from Sigma-Aldrich, NSW, Australia) for 5 min at RT. Endometrium was homogenized in radioimmuno-precipitation assay (RIPA) buffer (Cell Signaling, QLD, Australia) as per our previously established protocols (Zaitseva *et al.*, 2013). Protein concentration was determined using the Pierce BCA Protein Assay following the manufacturer's instructions (Thermo Fisher Scientific, Victoria, Australia).

Western blotting

Protein samples (50 μ g) were resolved on 7% NuPAGE Tris-Acetate (TA) polyacrylamide gels and transferred onto 0.45 μ m PVDF membrane (Life Technologies). Membranes were blocked with 5% (w/v) skim milk powder in tris-buffered saline plus 0.1% v/v tween-20 (TBS-T) for 1 h at RT. Primary antibody incubations occurred at 4°C overnight in 2% (w/v) skim milk powder in TBS-T with mouse monoclonal *GREB1* (clone *GREB1Ab*) (Millipore) (1:1000), or mouse monoclonal β -actin (Sigma-Aldrich) (1:10 000). HRP-conjugated goat anti-mouse secondary antibody (Bio-Rad Laboratories, CA, USA) was incubated for 1 h at RT (1:2000 for *GREB1* blots and 1:5000 for β -actin blots). ECL Prime reagent (GE Healthcare, Uppsala, Sweden) was used for signal detection and captured using an ImageQuant LAS 4000 (GE Healthcare). Densitometry was measured using Multi Gauge V3.0 software (Fujifilm, Brookvale, NSW, Australia).

The *GREB1* antibody selected for the above work was raised against a 119 amino acid peptide specific to human *GREB1* and aligns from 1098 to 1217 aa of the *GREB1a* isoform (UniProtKB: Q4ZG55) (Hnatyszyn *et al.*, 2010 and personal communication Dr James Hnatyszyn). Refer to Fig. 1.

Data analysis

Genotyping quality control

Standard quality control (QC) procedures were applied to individual datasets as outlined previously (Luong *et al.*, 2013). Briefly, SNPs with >5% missing



Figure 1 A diagram illustrating the genomic context of the *GREB1* gene region and the genomic region detected by the *GREB1* primers used for qRT-PCR experiments and protein sequence of *GREB1* antibody used for western blotting and IHC studies. F = Forward primers and R = Reverse primers for qRT-PCR. Arrowhead = exons and continuous line = introns of RefSeq Gene and transcripts. The diagram was generated from Geneious software platform.

rate, Hardy-Weinberg Equilibrium $P < 10^{-6}$ in controls and minor allele frequency $< 1\%$ were excluded. Following these QC measures, there were 51 genotyped SNPs within the *GREB1* gene. The samples overlapping between datasets, from non-European ancestry and with low call rates ($< 95\%$) were excluded from the downstream analyses. The final combined Australian dataset for the fine mapping analyses consisted of 2594 cases and 4496 controls. Of the total 7090 individuals in the combined dataset, 6503 are unrelated while 587 are related to some degree.

The final merged data were then imputed using the MACH program (Li et al., 2009, 2010) to impute missing genotypes. Quality of the imputed genotypes was assessed by R^2 metric, which estimates the squared correlation between true and imputed genotypes. All SNPs passed standard imputation quality control threshold ($R^2 > 0.3$). Association analysis of imputed genotype dosage scores with endometriosis was performed using PLINK software (<http://pngu.mgh.harvard.edu/purcell/plink/>, 22 July 2014, date last accessed) (Purcell et al., 2007). To account for relatedness in the dataset, the analysis was conducted using a robust variance estimation approach (Barlow, 1994; Williams, 2000) available in PLINK.

Gene expression analysis

Gene expression analysis was performed on the three assays showing specific amplification (single peak) on the melting curve. Data were analysed using Ct values obtained from the BioMark Gene Expression Data Analysis software, version 3.0.2 and statistical analyses were carried out in R version 2.15.1 (R Core Team). Missing expression measurements were defined by Ct values beyond detection limits (30 Ct). The Ct values of all the transcripts were converted into relative quantities using the deltaCt method implemented in *SLqPCR* package (Vandesompele et al., 2002; Kohl, 2007) and normalized by geometric mean of relative quantities of the three reference genes. The normalized data were rank transformed using the rank normalization method implemented in *GenABEL* R package (GenABEL project developers,

2013). Differential gene expression between cases and controls was evaluated using logistic regression, with and without adjusting for phases of menstrual cycle. Similarly, one-way analysis of variance (ANOVA) followed by Tukey's *post hoc* test was used to assess for differential expression across phases of menstrual cycle, with and without adjusting for case and control status. An interaction term was also included in the model to assess for possible interaction between phases of the menstrual cycle and case and control status.

eQTLs analysis

For each of the four risk variants examined, a cis-eQTL analysis was performed to investigate putative association between the variant and expression levels of nearby transcripts. The eQTL analysis was performed on the total of 94 tissue samples [with endometriosis ($n = 58$), without endometriosis ($n = 35$) and undetermined case/control ($n = 1$)] with recoded SNP genotypes based on minor allele dosage and fitted linear regression models, with phases of the menstrual cycle included as a covariate in the model. The Bonferroni method ($0.05/\text{number of tests performed}$) was used to correct for multiple comparisons with the significant threshold of $P < 0.003$.

Protein expression analysis

GREB1 immunostaining was semi-quantitatively assessed across the menstrual cycle by three independent observers (early, mid and late proliferative phase and early, mid and late secretory phase). Staining was graded from 0 to 3, where 0 was no staining and 3 was an intense staining. All protein data were analysed using Graph Pad Prism™ software (version 5, Graph Pad software, CA, USA). IHC scoring and western blot densitometry was analysed by one-way ANOVA using Tukey's *post hoc* test (with significance at $P < 0.05$).

Results

Common variants within the GREB1 region

To fine map this locus, we conducted additional genotyping in Australian case and control samples. Then we combined the data with previous GWAS results and re-analysed the association signals for endometriosis risk in the GREB1 region. We genotyped 51 common variants, including 11 coding variants, within the GREB1 gene (chr2:11675472-11782785; snp138/hg19) in a dataset of 2594 cases and 4496 controls of European origin. There were 16 SNPs including the initial GWAS SNP (rs13394619) and three coding variants showing nominally significant association ($P < 0.05$) with endometriosis risk. Results for non-coding SNPs showing LD ($r^2 > 0.25$) with the initial GWAS SNP (rs13394619) are listed in Table 1. Four SNPs (rs1898003, rs11674184, rs1865574 and rs2884374) showed stronger association with endometriosis risk than rs13394619 (Table 1).

Coding variants within GREB1 region

Amongst the 11 common coding variants within the GREB1 gene genotyped, association signals for three common non-synonymous coding variants were nominally significant for endometriosis risk (Table II). SNP rs2304402 showed the strongest association with endometriosis risk within the GREB1 gene ($P = 4.60 \times 10^{-3}$, OR = 1.118). The other SNPs rs142882892 ($P = 1.76 \times 10^{-2}$, OR = 1.475) and rs10929757 ($P = 2.18 \times 10^{-2}$, OR = 1.088) also showed evidence of association with endometriosis risk. The SNP rs2304402 is a non-synonymous A to G change in exon 29 and is predicted to change the amino acid from Asparagine (Asn) to Aspartic acid (Asp) at position 1687. SNP rs142882892 is a non-synonymous coding variant that changes C to T in exon 4 and is predicted to change the amino acid from Alanine (Ala) to Valine (Val) at position 115. The SNP rs10929757 is a non-synonymous coding variant that changes A to C in exon 3. This is predicted to change the amino acid from Asparagine (Asn) to Threonine (Thr) at position 77.

Conditional analyses

In order to investigate the presence of potential secondary and independent signals at the GREB1 locus, we performed a conditional analysis on the lead SNP from GWAS (rs13394619). Of the 15 common and 3 coding SNPs that showed significant associations with endometriosis at $P < 0.05$, signals for only two coding SNPs rs2304402 and rs142882892 persisted ($P = 4.83 \times 10^{-3}$ and $P = 2.40 \times 10^{-2}$, respectively) after conditioning on rs13394619. Coding SNPs rs2304402 and rs142882892 are in very weak LD with each other ($r^2 = 0.003$) and with the lead GWAS SNP rs13394619 ($r^2 = 0.001$ and $r^2 = 0.011$, respectively). Furthermore, when we conditioned on the new sentinel coding SNP (rs2304402), association signal of rs142882892 remained significant ($P = 2.28 \times 10^{-2}$), indicating that both of the coding SNPs may harbour independent signals at the GREB1 locus, in addition to the lead GWAS SNP rs13394619.

Haplotype analysis and pairwise SNP-SNP interactions

Considering results from the conditional analyses, which suggested multiple independent association signals at the GREB1 locus, we performed a haplotype analysis and pairwise epistasis analysis for rs13394619 and rs2304402, using observed genotype data available for 2213 endometriosis cases and 2044 controls. We excluded SNP rs142882892 from haplotype analysis as SNP rs142882892 is a low-frequency variant and is less likely to form a haplotype with the other two common variants with similar allele frequencies. As shown in Supplementary data, Table SII, results indicated that the haplotype formed by the risk alleles (GG) also showed significant association ($P = 0.026$) with endometriosis, but did not survive multiple testing corrections ($0.05/4 = 0.0125$).

We then used the Plink option ‘-epistasis’ and fitted logistic regression models based on allele dosage for each SNP to assess pairwise interaction. Results provided no evidence ($P = 0.13$) of pairwise interactions between the two SNPs rs13394619 and rs2304402.

Table 1 Endometriosis association information for common SNPs with LD ($r^2 > 0.25$) with the lead SNP (rs13394619) and nominally significant association with endometriosis risk ($P < 5 \times 10^{-2}$).

SNPs	Position (hg19)	LD (r^2)	RA	OA	RAF ^a case	RAF ^a control	OR ^b (95% CIs)	P
rs1898003	11704836	0.26	A	C	0.5901	0.5672	1.100 (1.024–1.182)	9.86×10^{-3}
rs11674184	11721535	0.62	A	C	0.6312	0.6112	1.094 (1.015–1.178)	1.91×10^{-2}
rs1865574	11681438	0.28	G	A	0.7272	0.7098	1.106 (1.016–1.204)	2.06×10^{-2}
rs2884374	11731846	1	A	G	0.5358	0.5168	1.081 (1.007–1.161)	3.26×10^{-2}
rs13394619	11727507	1	G	A	0.5363	0.5175	1.080 (1.006–1.159)	3.45×10^{-2}
rs17529680	11727087	0.39	C	A	0.7181	0.7014	1.088 (1.005–1.177)	3.68×10^{-2}
rs2358040	11731774	1	T	G	0.5369	0.5195	1.082 (1.004–1.166)	4.06×10^{-2}
rs7578132	11718858	0.66	C	T	0.5702	0.5526	1.080 (1.003–1.162)	4.22×10^{-2}
rs16857668	11723110	0.44	G	A	0.6664	0.6501	1.085 (1.003–1.173)	4.39×10^{-2}
rs7576826	11723021	0.88	C	T	0.5363	0.5196	1.078 (1.001–1.162)	4.97×10^{-2}

The bold SNP is to highlight the lead SNP and show that there are more stronger signals than this SNP.

^aRisk allele frequency.

^bOdd ratios were calculated for the risk allele.

Table II Endometriosis association information for common non-synonymous coding SNPs nominally significant association with endometriosis risk ($P < 5 \times 10^{-2}$).

SNPs	Position (hg19)	RA	OA	RAF ^a case	RAF ^a control	OR ^b (95% CIs)	P	Codon change	Amino acid change
rs2304402	11774324	G	A	0.4972	0.4725	1.118 (1.035–1.207)	4.60×10^{-3}	Aac/Gac	Asn/Asp
rs142882892	11706672	T	C	0.0206	0.0161	1.475 (1.072–2.030)	1.76×10^{-2}	gCg/gTg	Ala/Val
rs10929757	11702661	C	A	0.5956	0.5752	1.088 (1.013–1.168)	2.18×10^{-2}	aAc/aCc	Asn/Thr

^aRisk allele frequency.

^bOdd ratios were calculated for the risk allele.

Functional annotation for GREB1 common variants

We used publically available databases and bioinformatics tools to predict the possible functional roles of the SNPs with the strongest association with endometriosis risk. The RegulomeDB score, functional prediction and regulatory annotations of all the top SNPs are listed in Table III. Most of the SNPs identified from the fine mapping (Table I) are intronic and seven have regulatory roles in at least one cell type from the ENCODE project. There is limited functional prediction and regulatory annotation for the initial GWAS SNP (rs13394619) and the sentinel non-coding SNP (rs1898003). The regulatory chromatin state from the ENCODE project suggested a long-range regulatory role of the new sentinel SNP (rs1898003), which is located in transcriptional transition (initiation/elongation) sequences and associated with DNase hypersensitivity and histone modifications. The SNP (rs7576826), with the best score of 2a from RegulomeDB, is in strong LD ($r^2 = 0.87$) with the GWAS SNP rs13394619 (Table I) and in LD ($r^2 = 0.38$) with the SNP rs1898003. This SNP showed evidence of regulatory potential in multiple cells and is associated with transcription enhancer sequences, altered regulatory motifs for transcription factor(s) and is located in a binding site for the transcription factor CCCTC-binding factor (CTCF).

Expression of GREB1 protein and transcripts and protein across the menstrual cycle

Since GREB1 functionally interacts with the estrogen receptor, it was considered important to establish if GREB1 mRNA and protein expression were regulated across the menstrual cycle in endometrial tissues. Immunoreactive GREB1 was detected in both the endometrial stroma and glands, within both the nuclei and cytoplasm (Fig. 2a). However, when staining intensity was high, subcellular localization tended to be nuclear. We observed that GREB1 positive cells were more associated with spiral arterioles within the endometrial basalis during the secretory phase compared with the proliferative phase. Overall, staining intensity was largely equal between the endometrial functionalis and basalis. In endometrial glands, GREB1 increased from early to late proliferative phase then declined in expression until late secretory phase (Fig. 2a and c). In particular, GREB1 glandular staining was significantly decreased in late secretory relative to late proliferative phase endometrium (Fig. 2c). Changes in GREB1 expression across the menstrual cycle in the stroma did not reach statistical significance (Fig. 2b).

We designed primers for three assays targeting transcripts for GREB1; one assay amplified the transcript at the 5' end of the GREB1 gene

(GenBank: AF245390); the other two assays, including one set of primers to repeat a previous study targeting the sequence between exons 19 and 20 of the GREB1 reference gene (Pellegrini et al., 2012), amplified multiple transcripts at the 3' end of the GREB1 gene (GenBank: AB011147, BC172757, BC054502 and AB385351). There was no suitable assay to detect individual expression of the transcripts AB011147, BC172757, BC054502 and AB385351 (GenBank). All three assays showed a single peak from the melting curve analysis and passed all quality control steps. The positions of each primer set targeting GREB1 are summarized in Fig. 1. The Fluidigm real-time qPCR results confirmed that GREB1 transcripts are expressed in endometrial tissues (cases and controls). There was no significant difference in GREB1 mRNA expression across menstrual cycles (Fig. 2d).

We also performed an expression analysis (protein and transcripts) of cases versus controls across the menstrual cycle. The two-way ANOVA summary showed that there was no significant change in GREB1 mRNA expression in endometriosis cases ($n = 63$) compared with controls ($n = 36$) (nor for GREB1 protein expression by western blot (cases ($n = 28$) and controls ($n = 16$)) (data not shown).

Genetic regulation of RNA transcription in endometrial tissue with defined genotypes (eQTLs)

We genotyped the initial GWAS SNP (rs13394619), the new sentinel non-coding SNP (rs1898003), two coding variants (rs2304402 and rs142882892) and a SNP (rs7576826) with score 2a from Regulome DB on chromosome 2 in DNA samples extracted from blood of the same patients from the endometrial tissue project. Women were categorized as homozygous for the risk allele, homozygous for the alternative allele or heterozygous for each of the variants. An eQTL analysis was then conducted to examine the effect of genotypes at these SNPs on the expression of the GREB1 transcripts. After adjusting for menstrual cycle stage, none of the genotypes showed significant association with the expression of GREB1 transcripts (Fig. 3a–d).

Expression of GREB1 protein in endometrial tissue with defined genotypes

An analysis of GREB1 protein expression was performed on endometrial tissue (curette) from women genotyped for rs13394619 [homozygous for the GREB1 risk allele (GG), homozygous for the alternative allele (AA) or heterozygous (GA)]. GREB1 western blots detected a single high-

Table III Functional regulation annotation for common SNPs nominally significant association with endometriosis risk ($P < 5 \times 10^{-2}$).

SNPs	Position (hg19)	RegulomeDB score ^a	Functional prediction ^b	Regulatory annotation ^c
rs7576826	11723021	2a	Intronic/regulatory region variant	TF binding + matched TF motif + matched DNase Footprint + DNase peak
rs16857668	11723110	4	Intronic/Regulatory region variant	Transcription factor binding + DNase peak
rs7578132	11718858	4	Intronic/regulatory region variant	Transcription factor binding + DNase peak
rs1865574	11681438	5	Intronic/regulatory region variant	Transcription factor binding or DNase peak
rs1898003	11704836	5	Intronic/regulatory region variant	Transcription factor binding or DNase peak
rs11674184	11721535	5	Intronic/regulatory region variant	Transcription factor binding or DNase peak
rs17529680	11727087	5	Intronic/regulatory region variant	Transcription factor binding or DNase peak
rs2358040	11731774	5	Intronic/3'UTR variant	Transcription factor binding or DNase peak
rs2884374	11731846	5	Intronic/3'UTR variant	Transcription factor binding or DNase peak
rs13394619	11727507	—	Intronic/splice acceptor	Regulatory motifs altered

^aScores from RegulomeDB. Prediction for SNP with score.^bFunctional prediction from HaploReg v2 and UCSC Genome Browser.^cRegulatory annotation from ENCODE project in at least one cell type.

molecular-weight band of ~190 kDa in all AA, GA and GG women (Fig. 4a). A trend was observed where GG individuals had increased GREB1 protein expression relative to GA and AA women (Fig. 4b), however, this did not reach significance. Subdividing endometrium into proliferative or secretory phase did not alter the result (and there was no significant difference between cycle stages for AA, GA or GG women) (data not shown). In addition to rs13394619, further analyses were undertaken for GREB1 protein expression for the following SNPs of interest; rs1898003, rs2304402 and rs7576826. Like, rs13394619, these new variants did not demonstrate any effect on GREB1 protein expression (Fig. 4c–e).

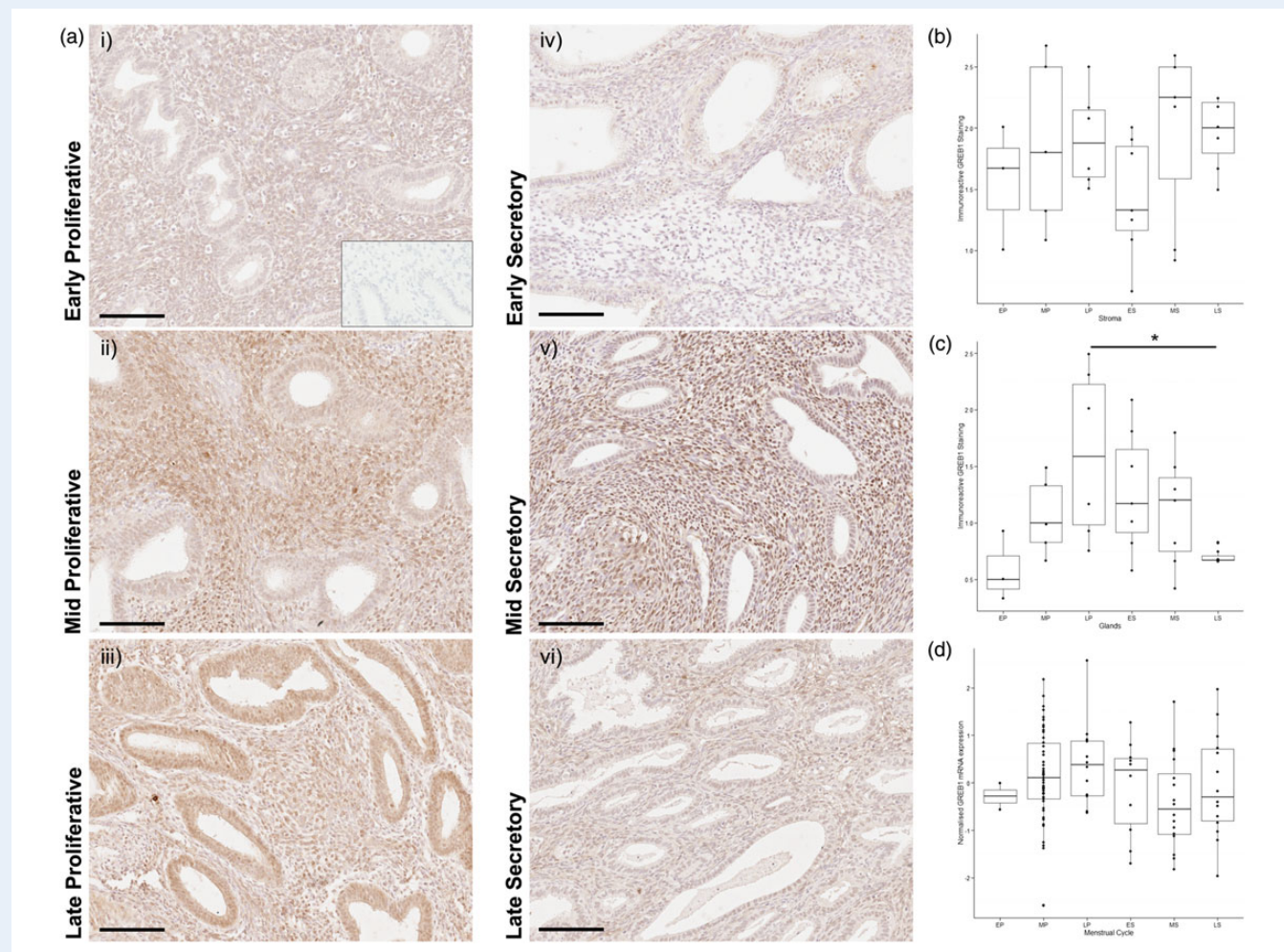
IHC was able to demonstrate altered expression profiles of GREB1 protein relative to phases the menstrual cycle (Fig. 2b and d). Therefore, IHC was used to identify any cell-specific differential GREB1 expression between women carrying the at risk allele for rs13394619 (GG) compared with women not-at-risk (AA) from curettes collected at the time of laparoscopy. Figure 5 shows representative micrographs of GREB1 staining in GG and AA endometrium (note that while mid-proliferative and mid-secretory phases were selected for the figure, all phases of the cycle stage where included in the analysis). Patterns of GREB1 staining from curettes matched those observed from full thickness endometrial sections (Fig. 2a). We did not observe any noticeable differences in GREB1 staining intensity, subcellular localization or specific cell phenotype when comparing GG and AA individuals (Fig. 5).

Discussion

GWAS for endometriosis report a strong association for disease risk with markers at chromosome 2p.25.1 and identified a genome-wide significant SNP rs13394619, which is located within the *GREB1* gene (Nyholt et al., 2012). GREB1 stands out as an excellent candidate for functional input in endometriosis pathogenesis because of its known estrogen-dependent properties (Mohammed et al., 2013) and previous

reports of altered expression in endometriosis (Pellegrini et al., 2012). Fine mapping results in this study identified a number of SNPs showing stronger association with endometriosis risk than the original GWAS SNP (rs13394619). There was evidence for multiple SNPs associated with disease risk and two independent signals, rs2304402 and rs142882892. There were common signals with the sentinel non-coding SNP rs1898003 located in intron 3 of *GREB1*, and two non-synonymous coding SNPs rs2304402 and rs142882892. SNP rs2304402 is common with a risk allele frequency of 0.49 and SNP rs142882892 is a low-frequency SNP with a risk allele frequency of 0.02. These SNPs are located in highly conserved sequences across species. Haplotype analysis of the common SNPs rs13394619 and rs2304402 showed similar signals for the haplotypes and individual SNPs and there was no evidence for interaction from epistasis analysis, suggesting an additive effect of the two SNPs on disease risk. Some SNPs showed potential functional roles following evaluation in publicly available databases. The majority of the SNPs associated with complex diseases are found within introns and intergenic regions. The causal variants are likely to have functional roles in regulating transcriptional output through multiple mechanisms (such as transcript levels and splicing) (Maurano et al., 2012; Schaub et al., 2012). The original GWAS SNP (rs13394619) is located in an intronic region between exon 9 and exon 10 of the *GREB1* gene. Using in silico tools, we did not find evidence for likely functional roles for this SNP. High-resolution fine mapping of the 2p25.1 locus identified additional SNPs, with stronger evidence for association with endometriosis susceptibility and predicted functional roles for the altered DNA sequences.

The SNP with the best predicted functional regulation was rs7576826, which also shows association with endometriosis risk ($P = 4.97 \times 10^{-2}$) and is in strong LD ($r^2 = 0.88$) with rs13394619. SNP rs7576826 has a Regulome DB score of 2a, suggesting effects of this SNP on multiple regulatory elements within the region. This SNP is located at the intronic region between exon 7 and exon 8 of *GREB1*. More specifically, from the ENCODE data, rs7576826 is aligned at the position with a predicted



TF binding site for CTCF in a number of cell types, including the MCF-7 breast cancer cell line, HeLa-S3 cervical carcinoma cells and human mammary epithelia cells (ENCODE_Project_Consortium, 2012). CTCF is a conserved zinc finger transcription factor and has been identified to be a transcriptional activator, repressor and/or silencer, depending on the DNA context of different gene loci (Phillips and Corces, 2009). In breast cancer cells, CTCF has been shown to act upstream of FOXA1 in determining the genomic response to estrogen (Zhang et al., 2010) and influence ER binding to chromatin (Ross-Innes et al., 2011), suggesting that CTCF motifs are involved in modulating estrogen-mediated gene expression changes. Given endometriosis is an estrogen-dependent disease, the functional annotation of the rs7576826 SNP to a CTCF TF binding site is intriguing and suggests that ENCODE data from breast cancer cell lines may be relevant to

functional annotation of genomic regions associated with endometriosis. To date, there is no information regarding a role for CTCF in endometriosis risk, however, a mutation in the *CTCF* gene has been linked to endometrial cancer (Zigheboim et al., 2014).

A large component of this study utilized bioinformatic tools and publicly available databases, which can provide some functional predictions of SNPs. The most abundant of these regulatory sequences are enhancers, but other regulators such as promoters, insulators and silencers may also be susceptibility targets. This regulation can be *cis*- (nearby) or *trans*- (distant) acting. Unlike promoters at the transcription start sites of the genes, distal regulation sequences such as enhancers are often cell type and disease specific and therefore imputed functional predictions identified in this investigation may not be specific to our disease or tissue of interest. Moreover, the ENCODE project currently has

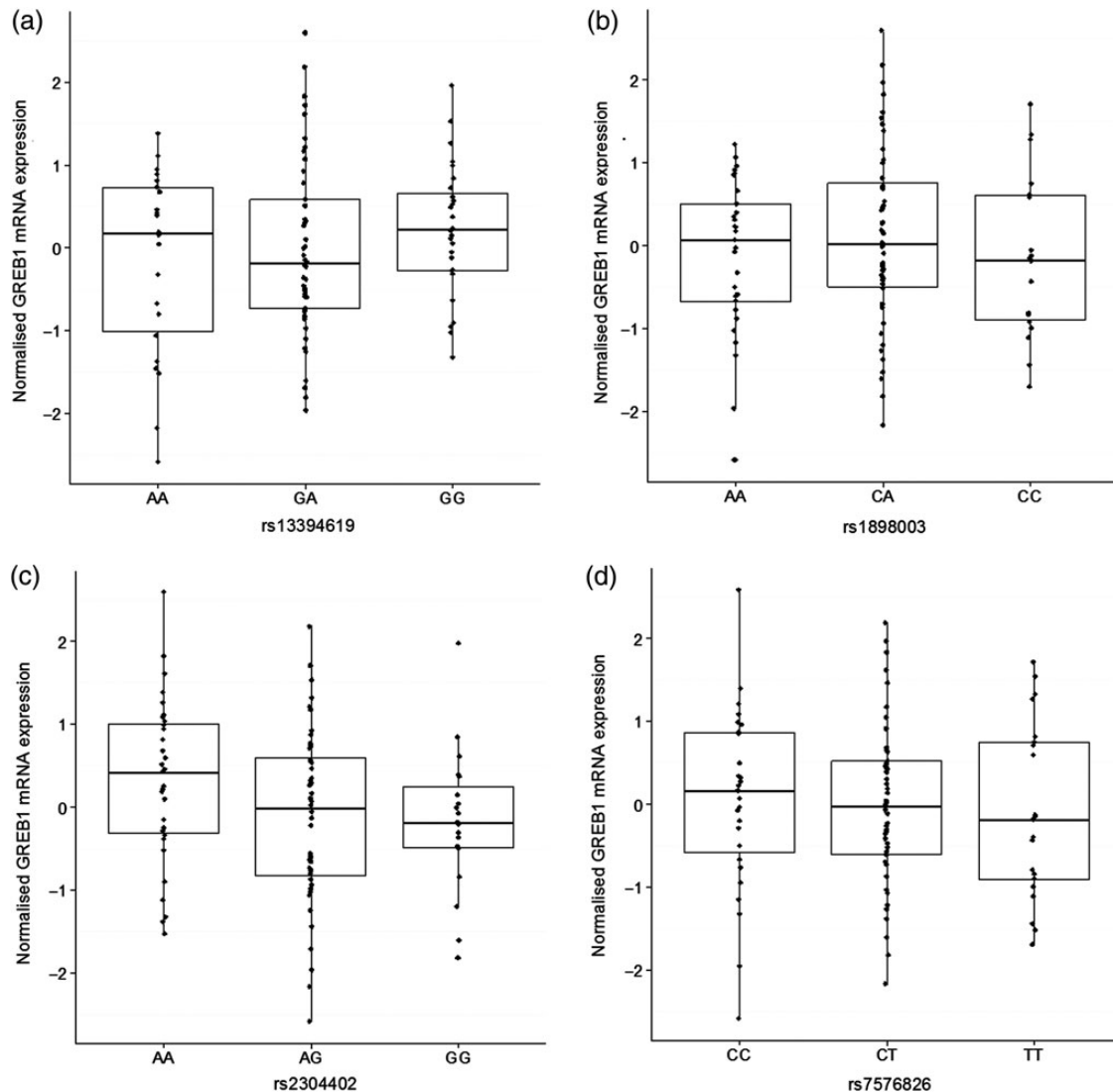


Figure 3 Representative eQTLs of the normalized expression level of GREB1 transcripts at 3' end (GenBank: BC172757, BC654502, AB011147, AB385351) amplified from assay 'primer exon 19–20' (note that while only the GREB1 mRNA expression from assay 'primer exon 19–20' was selected for the figures, all assays were included in the analyses). (a) The original GWAS SNP (rs13394619) [$n = 22$ (AA), $n = 45$ (AG) and $n = 27$ (GG)], (b) the new sentinel non-coding SNP (rs1898003) [$n = 29$ (AA), $n = 47$ (CA) and $n = 18$ (CC)], (c) top coding variant (rs2304402) [$n = 31$ (AA), $n = 44$ (AG) and $n = 19$ (GG)] and (d) the functional SNP (rs7576826) [$n = 27$ (CC), $n = 46$ (CT) and $n = 21$ (TT)]. The horizontal lines represent mean with the box representing the 25th and 75th percentiles and the individual sample represented by dots.

limited data on relevant cell lines specific for reproductive tissues. Therefore, we initiated our endometrial tissue project to investigate the effect of genetic variants on *GREB1* gene and protein expression. This approach allowed us to evaluate the effects of SNPs in our tissue (endometrium) and disease (endometriosis) of interest.

As endometrial biology and *GREB1* expression are both regulated by estrogen, we first examined *GREB1* gene and protein expression in endometrial tissues across different stages of the menstrual cycle. We did not observe any significant difference in *GREB1* transcript expression when comparing proliferative phase with the secretory phase. However, our IHC results showed a similar expression pattern of *GREB1* protein in stromal and glandular endometrial cells as *GREB1* transcripts, but this

time with a significant change in *GREB1* protein expression between late proliferative and late secretory phases in glands but not in stroma. The result provides evidence for tissue and cell-specific estrogen regulation of *GREB1* in endometrium during the menstrual cycle. The peak in *GREB1* glandular staining during the late proliferative phase correlates with the peak in estrogen levels in the menstrual cycle. A previous study also showed that endometrial glandular epithelial cells bind more estrogen in the proliferative phase of the cycle compared with stromal cells (Bergqvist et al., 1985), suggesting that *GREB1* levels are mirroring estrogen-regulated effects in human endometrium (particularly in the glands). Similar cycle stage-specific regulation of *GREB1* expression has been identified previously in breast cancer studies (Haynes et al.,

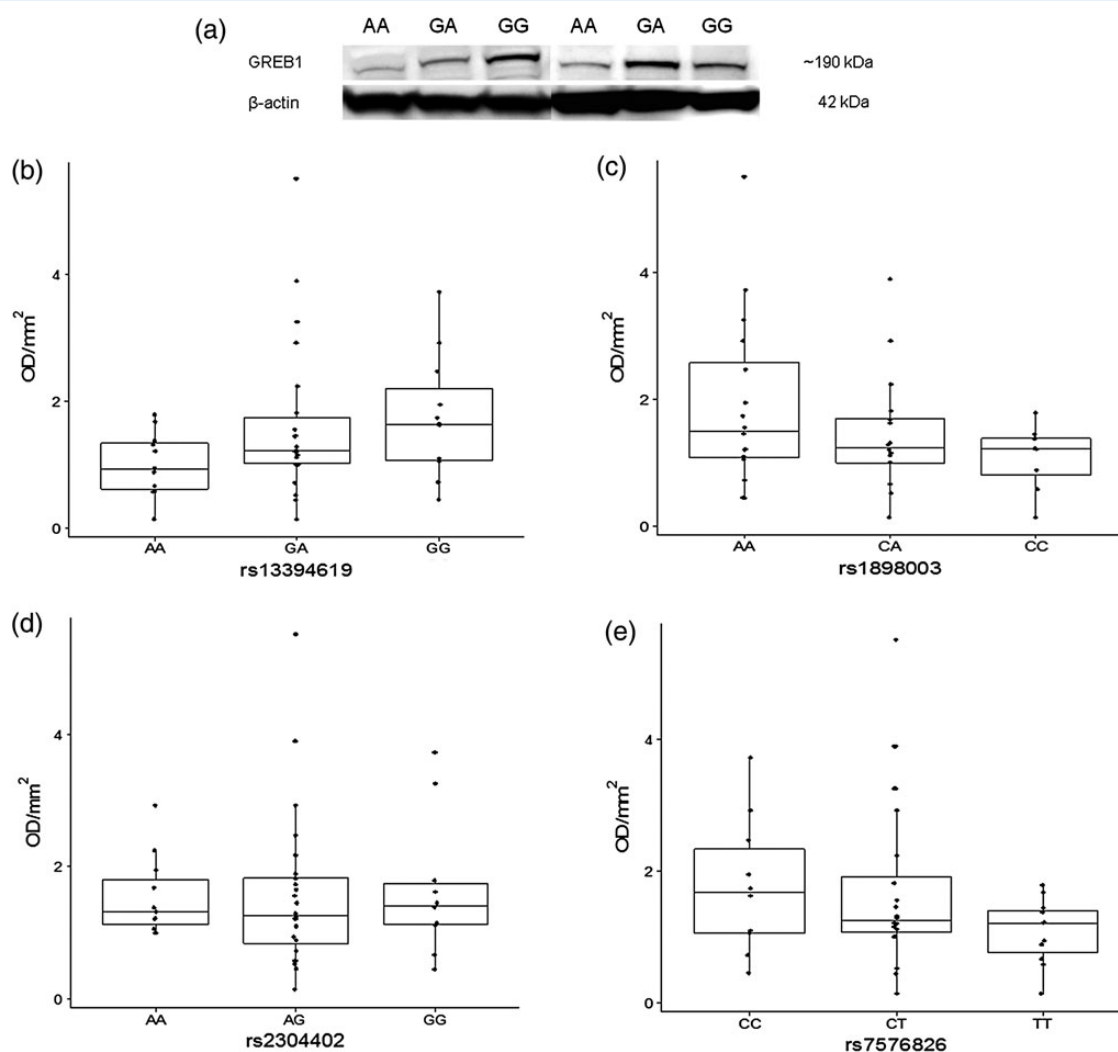


Figure 4 Endometrial expression of GREB1 protein in AA, GA and GG individuals for SNP rs13394619. (a) Representative western blots are displayed showing the detected GREB1 band at ~190 kDa. (b) GREB1 protein expression is graphed as the optical density (OD/mm²) for SNP rs13394619 [AA (*n* = 11), GA (*n* = 22) and GG (*n* = 11) individuals]. (c) GREB1 protein expression is graphed as OD/mm² for SNP rs1898003 [AA (*n* = 8), CA (*n* = 16) and CC (*n* = 16) individuals]. (d) GREB1 protein expression is graphed as OD/mm² for SNP rs2304402 [AA (*n* = 10), AG (*n* = 20) and GG (*n* = 11) individuals]. (e) GREB1 protein expression is graphed as OD/mm² for SNP rs7576826 [CC (*n* = 11), CT (*n* = 24) and TT (*n* = 10) individuals]. For the boxplots, the horizontal lines represent mean with the box representing the 25th and 75th percentiles and the individual sample represented by dots.

2013). However, the lack of a significant difference observed at the mRNA level may be the result of mixed cell populations in the tissue samples and studies using whole tissue may not be sufficiently sensitive to detect changes restricted to particular cell types of *GREB1* transcripts.

We then compared the *GREB1* gene and protein expression between endometriosis cases and controls, but we did not observe any significant difference in expression levels between the two groups. Our results did not replicate the previous study by Pellegrini and coworkers, which showed an increase in *GREB1* expression in eutopic endometrium from cases when compared with eutopic endometrium from controls (Pellegrini et al., 2012). Our qRT-PCR assays were designed to target *GREB1* expression at exon 19–20 as in the previous study (Pellegrini et al., 2012) and also a broader range of transcripts for the *GREB1* gene. However, there was no suitable assay that we could use to

detect the mRNA expression of individual transcripts at the 3' end of the *GREB1* gene in this study. Further studies investigating the individual transcripts using systemic whole-genome assays such as RNA-sequencing are needed. Pellegrini et al. (2012) reported a qualitative change in *GREB1* protein expression using IHC; we quantified *GREB1* protein following western blotting and did not identify a significant change in *GREB1* protein from whole tissue protein extracts.

We next turned our attention to the effects of individual SNPs on the expression of *GREB1* transcripts. Gene expression is highly heritable. Considering the SNPs identified from the fine mapping may have functional impacts on *GREB1* gene and protein expression, an eQTL approach was used in this study to examine the effects of specific genetic markers on *GREB1* transcripts and protein expression. Our results did not detect significant differences in expression levels between genotypes

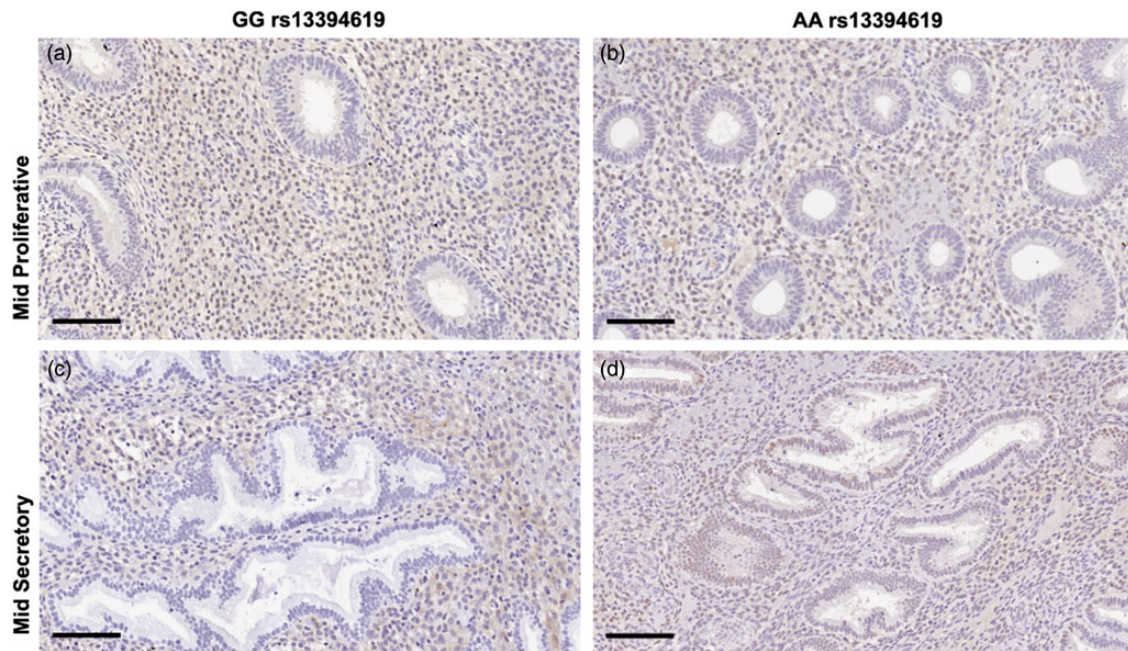


Figure 5 Endometrial (curette) expression of immunoreactive GREB1 protein from women homozygous for the risk allele (GG) and homozygous for the other allele (AA) of SNP rs13394619. Representative IHC micrographs with GREB1 staining in brown (**a** and **b**) mid-proliferative phase and (**c** and **d**) mid-secretory phase. IHC was performed on $n = 19$ (GG) and $n = 25$ (AA) women from matched stages of the menstrual cycle. Scale bar is equal to 100 μm .

for any SNPs tested (rs13394619, rs1898003, rs75768262, rs2304402 or rs142882892). There are multiple predictions for functional effects of SNP rs7576826 from ENCODE project. However, genetic variants can also affect the expression of genes that reside further away (*trans*) or are on different chromosomes by formation of chromatin loops (Westra *et al.*, 2013). A recent analysis by the ENCODE project shows that only 27% of the distal regulatory elements have an interaction with the nearest promoter, suggesting that the nearest gene is often not the target of a given GWAS association (Sanyal *et al.*, 2012).

In conclusion, we have identified new SNPs that have stronger association with endometriosis risk than the initial GWAS SNP (rs13394619) at the *GREB1* locus on chromosome 2p25.1. Some of these SNPs are predicted to lie in regions of functional regulation of the *GREB1* gene based on information from publicly available databases. Our combined SNP data now show independent signals for two non-synonymous coding variants in *GREB1*, providing further evidence that GREB1 may be involved in endometriosis pathophysiology. We did not find statistically significant evidence that any of these SNPs altered GREB1 gene or protein expression. However, our results show that GREB1 is modulated in a cell-specific and hormone-dependent manner in human endometrium and it is possible that whole tissue mRNA eQTLs and protein results are not detecting cell-specific effects. We did not detect eQTL effects for the sentinel SNP, although this might be a function of power. We estimated that with a sample size of 100, we have 80% power to detect an eQTL explaining 10% of the variation in *GREB1* expression. A substantial increase in sample size will be needed to detect more subtle effects. Given the gene expression of endometrium varies throughout the menstrual cycle, even larger sample sizes with well-defined stage of menstrual cycle will be required to understand gene regulation in the endometrium

and effects of the stage of the menstrual cycle. More detailed studies are required to determine the role of the two independent signals at the *GREB1* locus in increasing risk of developing endometriosis and whether they alter cell-specific regulation and function of GREB1 protein or act through some other mechanism.

Supplementary data

Supplementary data are available at <http://humrep.oxfordjournals.org/>.

Acknowledgements

The authors would like to acknowledge all the study participants and Ranita Charitra (Research Nurse) for her hard work and dedication in recruiting the women involved in the Endometrial tissue project. Research nurses Tracy Middleton and Irene Bell are also acknowledged for assistance with recruitment. We would like to acknowledge Leonie Cann for her technical assistance with IHC. Anatomical Pathology staff members from both the Royal Women's and the Royal Children's Hospitals (Parkville) are thanked for their assistance in recalling patient samples for the Endometrial tissue project. We also thank Leanne Wallace, Lisa Bowdler and Ashleigh Henders for sample processing, DNA extraction and gene expression analysis.

Authors' roles

J.N.F., S.J.H.-C., Y.S., Z.Z.Z., L.J., J.E.G., P.P., M.H., D.R.N., P.A.W.R. and G.W.M. all (i) played a role in the conception/design of the study

or acquisition and or analysis/interpretation of data, (ii) drafted or critically revised the article and (iii) gave their final approval of the article.

Funding

Funding for this work was provided by National Health and Medical Research Council of Australia (NHMRC) Project Grants APP1012245, APP1026033, APP1049472 and APP1046880. P.P. was supported by an NHMRC Fellowship (1013774). D.R.N. was supported by an NHMRC Fellowship (613674) and ARC Future Fellowship (FT0991022) schemes. G.W.M. was supported by NHMRC Fellowship Schemes (619667, 1078399).

Conflict of interest

The authors have nothing to declare.

References

- Barlow WE. Robust variance estimation for the case-cohort design. *Biometrics* 1994;**50**:1064–1072.
- Bergqvist A, Ekman R, Ljungberg O. Binding of estrogen and progesterone to human endometrium in the different phases of the menstrual cycle. A histochemical study. *Am J Clin Pathol* 1985;**83**:444–449.
- Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;**22**:1790–1797.
- Burney R, Giudice L. Pathogenesis and pathophysiology of endometriosis. *Fertil Steril* 2012;**98**:511–519.
- Bush D, Evans S, Vancaillie T. The Pelvic Pain Report. Australia: The Pelvic Pain Steering Committee, 2011.
- Culley L, Law C, Hudson N, Denny E, Mitchell H, Baumgarten M, Raine-Fenning N. The social and psychological impact of endometriosis on women's lives: a critical narrative review. *Hum Reprod Update* 2013;**19**:625–639.
- Deschenes J, Bourdeau V, White JH, Mader S. Regulation of GREB1 transcription by estrogen receptor alpha through a multipartite enhancer spread over 20 kb of upstream flanking sequences. *J Biol Chem* 2007;**282**:17335–17339.
- Encode_project_consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**:57–74.
- Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, Casey G, De Biasi M, Carlson C, Duggan D et al. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet* 2011;**43**:513–518.
- Ghosh MG, Thompson DA, Weigel RJ. PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer1, 2. *Cancer Res* 2000;**60**:6367–6375.
- Haynes BP, Viale G, Galimberti V, Rotmensz N, Gibelli B, A'hern R, Smith IE, Dowsett M. Expression of key oestrogen-regulated genes differs substantially across the menstrual cycle in oestrogen receptor-positive primary breast cancer. *Breast Cancer Res Treat* 2013;**138**:157–165.
- Hickey M, Ballard K, Farquhar C. Endometriosis. *Br Med J* 2014;**348**:1–9.
- Hnatyszyn HJ, Liu M, Hilger A, Herbert L, Gomez-Fernandez CR, Jorda M, Thomas D, Rae JM, El-Ashry D, Lippman ME. Correlation of GREB1 mRNA with protein expression in breast cancer: validation of a novel GREB1 monoclonal antibody. *Breast Cancer Res Treat* 2010;**122**:371–380.
- Holdsworth-Carson SJ, Zaitseva M, Vollenhoven BJ, Rogers PA. Clonality of smooth muscle and fibroblast cell populations isolated from human fibroid and myometrial tissues. *Mol Hum Reprod* 2014;**20**:250–259.
- Kennedy S, Mardon H, Barlow D. Familial endometriosis. *J Assist Reprod Genet* 1995;**12**:32–34.
- Kohl M. *SLqPCR: Functions for Analysis of Real-Time Quantitative PCR Data at SIRS-Lab GmbH*. Jena, Germany: SIRS-Lab GmbH, 2007.
- Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013;**501**:506–511.
- Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet* 2009;**10**:387–406.
- Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2010;**34**:816–834.
- Luong HT, Painter JN, Shakhbazov K, Chapman B, Henders AK, Powell JE, Nyholt DR, Montgomery GW. Fine mapping of variants associated with endometriosis in the WNT4 region on chromosome 1p36. *Int J Mol Epidemiol Genet* 2013;**4**:193–206.
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012;**337**:1190–1195.
- Moen MH, Magnus P. The familial risk of endometriosis. *Acta Obstet Gynecol Scand* 1993;**72**:560–564.
- Mohammed H, D Santos C, Serandour AA, Ali HR, Brown GD, Atkins A, Rueda OM, Holmes KA, Theodorou V, Robinson JL et al. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Rep* 2013;**3**:342–349.
- Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. Genetic analysis of genome-wide variation in human gene expression. *Nature* 2004;**430**:743–747.
- Nyholt DR, Low SK, Anderson CA, Painter JN, Uno S, Morris AP, Macgregor S, Gordon SD, Henders AK, Martin NG et al. Genome-wide association meta-analysis identifies new endometriosis risk loci. *Nat Genet* 2012;**44**:1355–1359.
- Painter JN, Anderson CA, Nyholt DR, Macgregor S, Lin J, Lee SH, Lambert A, Zhao ZZ, Roseman F, Guo Q et al. Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat Genet* 2011;**43**:51–54.
- Pellegrini C, Gori I, Achtari C, Hornung D, Chardonens E, Wunder D, Fiche M, Canny GO. The expression of estrogen receptors as well as GREB1, c-MYC, and cyclin D1, estrogen-regulated genes implicated in proliferation, is increased in peritoneal endometriosis. *Fertil Steril* 2012;**98**:1200–1208.
- Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell* 2009;**137**:1194–1211.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559–575.
- Rae JM, Johnson MD, Scheys JO, Cordero KE, Larios JM, Lippman ME. GREB1 is a critical regulator of hormone dependent breast cancer growth. *Breast Cancer Res Treat* 2005;**92**:141–149.
- Rahmioglu N, Nyholt DR, Morris AP, Missmer SA, Montgomery GW, Zondervan KT. Genetic variants underlying risk of endometriosis: insights from meta-analysis of eight genome-wide association and replication datasets. *Hum Reprod Update* 2014;**20**:702–716.
- Ross-Innes CS, Brown GD, Carroll JS. A co-ordinated interaction between CTCF and ER in breast cancer cells. *BMC Genomics* 2011;**12**:593.
- Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature* 2012;**489**:109–113.

- Sapkota Y, Low SK, Attia J, Gordon SD, Henders AK, Holliday EG, Macgregor S, Martin NG, Mcevoy M, Morris AP et al. Association between endometriosis and the interleukin 1A (IL1A) locus. *Hum Reprod* 2015;**30**:239–248.
- Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res* 2012;**22**:1748–1759.
- Sun J, Nawaz Z, Slingerland JM. Long-range activation of GREB1 by estrogen receptor via three distal consensus estrogen-responsive elements in breast cancer cells. *Mol Endocrinol* 2007;**21**:2651–2662.
- Treloar SA, O'connor DT, O'connor VM, Martin NG. Genetic influences on endometriosis in an Australian twin sample. *Fertil Steril* 1999;**71**:701–710.
- Treloar S, Hadfield R, Montgomery G, Lambert A, Wicks J, Barlow DH, O'connor DT, Kennedy S. The International Endogene Study: a collection of families for genetic research in endometriosis. *Fertil Steril* 2002;**78**:679–685.
- Uno S, Zembutsu H, Hirasawa A, Takahashi A, Kubo M, Akahane T, Aoki D, Kamatani N, Hirata K, Nakamura Y. A genome-wide association study identifies genetic variants in the CDKN2BAS locus associated with endometriosis in Japanese. *Nat Genet* 2010;**42**:707–710.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;**3**:RESEARCH0034.
- Vercellini P, Barbara G, Abbiati A, Somigliana E, Vigano P, Fedele L. Repetitive surgery for recurrent symptomatic endometriosis: What to do? *Eur J Obstet Gynecol Reprod Biol* 2009;**146**:15–21.
- Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013;**45**:1238–1243.
- Williams RL. A note on robust variance estimation for cluster-correlated data. *Biometrics* 2000;**56**:645–646.
- Zaitseva M, Holdsworth-Carson SJ, Waldrip L, Nevzorova J, Martelotto L, Vollenhoven BJ, Rogers PA. Aberrant expression and regulation of NR2F2 and CTNNB1 in uterine fibroids. *Reproduction* 2013;**146**:91–102.
- Zhang Y, Liang J, Li Y, Xuan C, Wang F, Wang D, Shi L, Zhang D, Shang Y. CCCTC-binding factor acts upstream of FOXA1 and demarcates the genomic response to estrogen. *J Biol Chem* 2010;**285**:28604–28613.
- Zhao ZZ, Nyholt DR, Le L, Martin NG, James MR, Treloar SA, Montgomery GW. KRAS variation and risk of endometriosis. *Mol Hum Reprod* 2006;**12**:671–676.
- Zigheboim I, Mutch DG, Knapp A, Ding L, Xie M, Cohn DE, Goodfellow PJ. High frequency strand slippage mutations in CTCF in MSI-positive endometrial cancers. *Hum Mutat* 2014;**35**:63–65.