Global Endometrial DNA Multi-omics Analysis Reveals Insights into mQTL Regulation and Associated Endometriosis Disease Risk

Sally Mortlock[1]*, Sahar Houshdaran[2]*, Idit Kosti[3]*, Nilufer Rahmioglu[4,5]*, Camran Nezhat[6,7,8], Allison F. Vitonis[9], Shan V. Andrews[3], Parker Grosjean[3], Manish Paranjpe[3], Andrew W. Horne[11], Alison Jacoby[2], Jeannette Lager[2], Jessica Opoku-Anane[2], Kim Chi Vo[2], Evelina Manvelyan[2], Sushmita Sen[2], Zhanna Ghukasyan[2], Frances Collins[11], Xavier Santamaria[10,16], Philippa Saunders[12], Kord Kober[3,13], Allan F. McRae[1], Kathryn L. Terry[9,14,15], Júlia Vallvé-Juanico[2,16], Christian Becker[5], Peter A.W. Rogers[17], Juan C. Irwin[2], Krina Zondervan[4,5]**, Grant W. Montgomery[1]**, Stacey Missmer[14,15,18,19]**, Marina Sirota[3,20]**, Linda Giudice[2]**

1. The Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia.
2. Center for Reproductive Sciences, Department of Obstetrics, Gynecology & Reproductive Sciences, University of California San Francisco, San Francisco, CA, USA.
3. Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, CA, USA.
4. Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK.
5. Oxford Endometriosis CaRe Centre, Nuffield Department of Women's and Reproductive Health, John Radcliffe Hospital, University of Oxford, Oxford, UK.
6. Stanford University Medical Center, Palo Alto, California, USA.
7. University of California San Francisco, San Francisco, California, USA.
8. Camran Nezhat Institute, Center for Special Minimally Invasive and Robotic Surgery, Woodside, California, USA.
9. Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.
10. Carlos Simon Foundation, Health Research Institute, Valencia, Spain.
11. MRC Centre for Reproductive Health, University of Edinburgh, QMRI, Edinburgh, UK.

12. Centre for Inflammation Research, Institute for Regeneration and Repair University of Edinburgh, Edinburgh, UK.
13. Department of Physiological Nursing, University of California San Francisco, San Francisco, CA, USA.
14. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
15. Boston Center for Endometriosis, Boston Children's Hospital and Brigham and Women's Hospital, Boston, MA, USA.
16. Group of Biomedical Research in Gynecology, Vall d'Hebron Research Institute, Barcelona, Spain.
17. University of Melbourne Department of Obstetrics and Gynaecology, Royal Women's Hospital, Melbourne, Australia.
18. Division of Adolescent and Young Adult Medicine, Department of Medicine, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA.
19. Department of Obstetrics, Gynecology, and Reproductive Biology, College of Human Medicine, Michigan State University, Grand Rapids, MI, USA.
20. Department of Pediatrics, Division of Neonatology, University of California San Francisco, San Francisco, CA, USA.

* These authors contributed equally to this work as co-first authors

** These authors contributed equally to this work as co-senior authors

Abstract

Endometriosis is a leading cause of pain and infertility affecting millions of women globally. Identifying biologic and genetic effects on DNA methylation (DNAm) in endometrium increases understanding of mechanisms that influence gene regulation predisposing to endometriosis and offers an opportunity for novel therapeutic target discovery. Herein, we characterize variation in endometrial DNAm and its association with menstrual cycle phase, endometriosis, and genetic variants through analysis of genome-wide genotype data and methylation at 759,345 DNAm sites in endometrial samples from 984 deeply-phenotyped participants. We identify significant differences in DNAm profiles

between menstrual cycle phases and at four DNAm sites between stage III/IV endometriosis and controls. We estimate that 15.4% of the variation in endometriosis is captured by DNAm, and identify DNAm networks associated with endometriosis. DNAm quantitative trait locus (mQTL) analysis identified 118,185 independent *cis*-mQTL including some tissue-specific effects. We find significant differences in DNAm profiles between endometriosis sub-phenotypes and a significant association between genetic regulation of methylation in endometrium and disease risk, providing functional evidence for genomic targets contributing to endometriosis risk and pathogenesis.

Introduction

Endometriosis is a common, estrogen-dependent, inflammatory disorder defined by the presence of endometrial-like tissue at extra-uterine (ectopic) sites, commonly within the pelvis[1]. It is associated with pain and infertility, affects an estimates 190 million women and contributes hundreds of billions of dollars to healthcare costs globally[1-4]. Pelvic disease likely derives from retrograde menstruation of steroid hormone-sensitive endometrial cells that elicit inflammation, neuroangiogenesis, fibrosis, and scarring[5-7]. Disease classification is based primarily on lesion appearance and location without correlation to symptoms or responses to surgical and hormonal therapies, and molecular classifications of disease subtypes are wanting. Although mechanisms driving the pathogenesis and pathophysiology of this heterogeneous disorder are largely unknown, studies suggest that diverse endometrial cellular components of women with disease respond abnormally to cyclic steroid hormones and exhibit aberrant signaling pathways, gene expression, decreased apoptosis, enhanced proliferation, and an inflammatory phenotype as compared to controls[8,9]. A recent genome-wide association meta-analysis identified disease-associated genomic loci in women with versus without endometriosis

involving sex steroid and hormone signaling pathways, inflammatory pathways, oncogenesis and angiogenesis, with greater effect sizes observed in those with more advanced disease stage[10]. Genetic risk factors for endometriosis have also been associated with other reproductive traits including irregular menses, uterine fibroids, and ovarian cancer[10,11]

While genetic variation underpinning complex disease etiology is actively studied, epigenetic processes such as acquired alterations in DNA methylation (DNAm) are increasingly recognized as providing an important biological link between individual exposures and disease-specific phenotypes[12]. Endometrial whole methylome and candidate gene DNAm profiles in bulk tissue and select cell populations identify genes involved with steroid hormone dependence and abnormalities in women with versus without endometriosis. However, most results fail to be replicated due to limited sample size, cellular heterogeneity in bulk tissue specimens, poor cycle phase assignments, and limited clinical metadata.[15-20] Notably, aberrant candidate gene promoter DNAm includes *HOX-A10*[13], steroid hormone receptors (*PR* and *ESR1*)[14,15], *CYP19*/aromatase[16], *SF-1*[17,18], and *COX-2*[19] and altered expression of *DNMT*s[20]. DNA methylome-wide studies that use a small number of samples have suggested DNAm profiles differ in patients with endometriosis - both between endometriosis lesion and eutopic endometrial tissues, and compared to eutopic endometrium of women without disease[8,21-23]. One study of 16 patients with and without endometriosis was designed not to test the hypothesis of differential DNAm between cases and controls but to quantify intra-tissue variation and determine sample sizes necessary for large-scale discovery[21]; a second study evaluated the DNAm in 17 patients with endometriosis and 16 controls, across the menstrual cycle and noted marked phase- and disease-dependent signatures[9]. Both studies determined that menstrual cycle phase accounted for the majority of variability in

DNAm patterns within the endometrium. Estradiol, progesterone, and their combination were found to regulate genome-wide DNAm signatures and gene transcription in endometrial stromal fibroblasts isolated from 8 women with endometriosis (4 with stage I/II and 4 with stage III/IV disease) and 7 women without endometriosis, with marked aberrances in endometriosis patients[8]. Other recent small studies on genetic regulation of endometrial gene expression (expression quantitative trait loci; eQTL)[24] and DNAm (methylation quantitative trait loci; mQTL)[25] demonstrate associations between specific genotypes and expression of genes in signaling pathways and between DNAm sites near *GREB1* and *KDR*, genes involved in endometriosis pathogenesis.

This study reports global DNAm profiles and signaling pathways associated with cell adhesion, proliferation and metabolism across the menstrual cycle in endometrium collected from a large sample of patients with surgically- and histologically-confirmed endometriosis (n=637) and women without endometriosis (n=347). Importantly, the dataset is based on patients with well-documented clinical annotation, demographic characteristics, and symptomatology[26-28]. Our robust analysis reveals novel factors affecting epigenetic regulation in endometrium associated with endometriosis risk and disease heterogeneity.

Results

Sample characteristics

Eutopic endometrial samples from 1,074 women were included. Following sample quality control (QC) filtering, 984 participants (one sample each from 637 endometriosis cases, 347 controls (201 “No Uterine or Pelvic Pathology” (NUPP) controls)) contributed to the analyses (Table 1; Supplementary Table 1). The overall study design is shown in Figure 1.

Clinical and demographic factors characterized across samples were: contributing site/institution, menstrual cycle phase, endometriosis case:control status, sample methylation array plate, sample methylation assay batch, revised American Society for Reproductive Medicine (rASRM) endometriosis disease stage[29], lesion type and pain variables (see Methods).

Single site and regional analysis identify distinct menstrual cycle phase DNAm profiles

To investigate if variation in DNAm in endometrium was associated with menstrual cycle stage or endometriosis, single site (each individual DNAm site/probe) and regional association analyses were performed using a set of linear models. Genome-wide DNAm was measured across 759,345 DNAm sites in 984 samples using the Illumina Infinium MethylationEPIC Beadchip (Figure 1, Table 1). Genetic ancestry was determined using genotype data and principal component analysis (PCA); most of the participants were of European ancestry (68%) (Supplementary Figure 1A and Table 1). DNAm data for all 984 samples before and after correction for covariates using surrogate variable analysis (SVA) are shown in Supplementary Figures 1B&C and 2. Including the surrogate variables (SVs) in the linear models reduced the association of the PCs with batch and site and identified a strong association with menstrual cycle phase (Supplementary Figure 3). The largest number of differentially methylated sites was observed between samples collected within the secretory phase versus the proliferative phase (9,654 DNAm sites) and subsequent comparisons involving the aggregated phases (Figure 2, Supplementary Figures 4 and 5, Table 2, Supplementary Tables 2 & 3) regardless of disease status. Subtle differences were also observed across secretory (progesterone-dominant) sub-phase (early (ESE), mid (MSE) and late (LSE)) comparisons (Supplementary Tables 2 & 3). The inflation lambda factors and the QQ plots are shown in Supplementary Figure 6. Clear clustering across cycle phase groups

was observed (Figure 2 and Supplementary Figures 4 and 5). The distribution of the genomic locations of the significantly differentially methylated sites is shown in Supplementary Table 4. Genes annotated to DNAm sites with differential DNAm between PE and SE were enriched in extracellular matrix (ECM)-cell interaction (adherens junctions, focal adhesion, regulation of actin cytoskeleton, Rho and Rap1 signaling) and cell proliferation and metabolism signaling (phospholipase D, PI3K-Akt, Ras signaling) pathways, consistent with known biological processes in the estrogen-dominant PE and progesterone-dominant SE[30] (Figure 2C**,** Table 3). Several pathways were also enriched for genes annotated to significant DNAm sites from the stratified analysis which included ESE, MSE and LSE phases, these included pathways unique to the ESE vs MSE comparison (eg. CRMPs in Sema3A signaling) and ESE vs LSE comparison (eg. apoptosis; Supplementary Table 5). Results from the differential methylation analysis using this SVA approach were consistent with an alternate Mixed Linear Model-based Omic Association (MOA) approach (Supplementary Data 1; Supplementary Figure 7).

Single site and regional analysis does not differentiate between women with and without endometriosis

Overall, we observed no significant difference (at genome-wide threshold) in single site or regional methylation signals between cases and controls. As effect sizes for many genetic risk factors are consistently greater in patients with rASRM stage III/IV disease[31], we analyzed results for stage III/IV cases compared with all controls. Two signals (cg02623400 in the 5'UTR region of *ELAVL4* on chromosome 1 and cg02011723 in the promoter region of *TNPO2* on chromosome 19) had higher DNAm in stage III/IV cases and passed the threshold for Bonferroni-corrected significance for a single test ($P<6.58x10^{-8}$). We also conducted an analysis of cases compared to the restricted NUPP control group. Only one site (cg26912870;

intergenic) with lower DNAm on chromosome 11 in cases, and one site on chromosome 19 (cg18305031; intron of *EEF2*) with higher DNAm in stage III/IV cases, passed the threshold for Bonferroni-corrected significance for a single test. Differences in DNAm at cg02623400 and cg0201172 were only nominally significant when restricted to NUPP controls.

DNAm networks are associated with menstrual cycle phase and endometriosis

Weighted correlation network analysis (WGCNA)[32] was used to identify clusters (modules) of highly correlated DNAm sites that are associated with menstrual cycle phase or endometriosis case:control status. Thirty five DNAm modules were identified based on the WGCNA clustering of the DNAm profiles (Supplementary Figure 8), eight of which were significantly associated ($p<0.05$) with menstrual cycle phase and four with endometriosis case:control status (Figure 3). Genes annotated to probes within clusters associated with stage III/IV case:control status were overrepresented in 74 KEGG pathways (Supplementary Table 6). The top pathways grouped into cellular proliferation, neuronal-related, ECM-cell interaction, and cancer pathways, including MAPK, Wnt, calcium and Hippo signaling, focal adhesion, axon guidance, and breast and gastric cancer (Supplementary Table 6; Figure 3). The genomic locations for the module sites are provided in Supplementary Table 7.

Genetic regulation of DNAm in endometrium

To assess the effect of genetic variation on DNAm in endometrium, and identify endometrial mQTLs, we integrated DNAm data at the 759,345 sites with genotype data for 5,290,992 SNPs from 658 samples with European ancestry (Table 1; Figure 1). A total of 12,242,158 significant *cis*-mQTLs were identified in endometrium across 107,627 DNAm sites following Bonferroni correction (p-value < 1.7×10^{-11}). Stepwise regression identified 118,185 independent *cis*-mQTL signals, evidence that some sites are regulated by multiple

independent genetic variants. In total, mQTLs for 90,856 DNAm sites were novel and were not reported in endometrium previously[25]. There was evidence of mild genomic inflation in test statistics (lambda=1.16). This is, however, consistent with inflation observed in differential DNAm models and is likely the result of residual variation from unknown mediators of variation in DNAm in endometrium. We were able to replicate 98% of endometrial mQTLs previously reported in a small preliminary study (n=66)[25].

Shared genetic regulation between endometrium and other tissues

To investigate shared vs. tissue-specific genetic regulation, we compared *cis*-mQTLs found in endometrium with those reported in publicly available mQTL datasets including, blood[33,34], skeletal muscle[35], adipose tissue[36], and brain[37]. 78% of mQTLs identified in endometrium were also significant in blood and had the same direction of effect. In total, endometrial mQTLs for 3,480 DNAm sites were not observed in blood (Min, et al. [34] = $p<1x10^{-8}$; Hannon, et al. [33]= $p<5x10^{-8}$), brain ($p<5x10^{-8}$), muscle ($p<5x10^{-8}$), or adipose tissue (corrected p-value<0.05) and may represent tissue-specific genetic regulation (Supplementary Table 8). An additional 19,317 DNAm sites with significant mQTLs in endometrium on the EPIC array were not significant in blood. Genes annotated to the 3,480 endometrium-specific DNAm sites were enriched in several Hallmark pathways including epithelial mesenchymal transition, androgen, and estrogen response, and KEGG pathways including ECM interactions and cell adhesion (Supplementary Table 9).

Effect of genetic regulation is dependent on case versus control status and menstrual cycle phase

To investigate if genetic effects on methylation differ between biological states in endometrium, we conducted context-specific mQTL analyses testing for differences in the

independent endometrial cis-mQTL effects between women with and without endometriosis and between proliferative and secretory cycle phases. Context-specific analyses across all 118,185 mQTLs identified two with effects that were significantly ($p<4.2x10^{-7}$) associated with endometriosis status. Similarly, we found evidence of significant association between the effects of 12 mQTLs and menstrual cycle phase (Supplementary Figure 8; Supplementary Table 10). mQTLs significant in the context-specific analysis suggests that the effect was larger and/or more variable in either cases or controls or proliferative or secretory samples. Larger sample sizes would be required to validate these context-specific effects.

Regulation of DNAm and transcription associated with endometriosis

To identify any associations between genetic regulation of DNAm in endometrium and risk of endometriosis, we integrated summary statistics from both endometrial mQTLs and an endometriosis genome-wide association (GWA) meta-analysis including 31,021 cases and 524,419 controls, using Summary data-based Mendelian Randomization (SMR)[38]. We identified 41 unique genetic variants that were associated with methylation at 45 DNAm sites and risk of endometriosis (Supplementary Table 11). Of the 45 mQTLs associated with endometriosis, 15 remained associated when cases were restricted to those with stage III/IV disease along with 6 mQTLs not associated with overall endometriosis (Supplementary Table 11). We were able to replicate published associations between methylation at DNAm sites near *GREB1* (Figure 4) and *KDR* and endometriosis[25]. Out of the 51 mQTLs associated with endometriosis, 31 were also associated with the expression of 18 genes in various GTEx tissues, blood, and endometrium (Supplementary Tables 11 and 12). Expression of several of these genes (*LINC00339, CDC42, GDAP1, FGD6, SRP14*) has been associated with risk of endometriosis previously[24,39-41]. Expression of four of the genes (*EEFSEC, GDAP1, ADK* & *SKAP1*) was also significantly associated with endometriosis risk when SMR analyses were

conducted using the eQTL and GWAS summary statistics (Supplementary Table 13). Note it was not a requirement that the lead mQTL SNP and eQTL SNP associated with endometriosis be the same; however, they were in high linkage disequilibrium (LD) in European populations ($r^2 > 0.8$).

EpiMap and H3K27Ac HiChIP libraries were used to functionally annotate SNPs and DNAm sites significantly associated with endometriosis. Evidence of enhancers was found for 19 SNPs and 40 DNAm sites (Supplementary Table 11). Similarly, evidence of promoters was found for 6 SNPs and 15 DNAm sites (Supplementary Table 11). Based on EpiMap predicted tissue-specific enhancer-gene links and location of predicted promoters, we identified 27 target genes with links to the enhancers and promoters (Supplementary Table 11). Target genes identified include those previously implicated in endometriosis risk on chr1p36.12. mQTLs in this region were in predicted enhancer regions and promoter regions of *WNT4* in uterus and were associated with expression of nearby *CDC42* and *LINC00339* (Supplementary Figure 10). Other novel examples include *EEFSEC* on chr3q21.3 and a cluster of HOXC genes on chr12q13.13 (Supplementary Figure 11). mQTLs in the chr3q21.3 region associated with expression of *EEFSEC* were located in a predicted enhancer region in uterus, and one SNP (rs2999046) was located in an active chromatin area predicted to interact with *EEFSEC* and *RUVBL1* in immortalized endometrial cells (Figure 5; Supplementary Table 11).

In total, 45 genes were annotated to endometriosis-associated mQTLs using epigenetic data and eQTLs, 16 of these genes are in loci not previously associated with both expression/ methylation and endometriosis risk (Supplementary Table 11). These 45 genes had similar expression across reproductive tissues (ovary, fallopian tube, uterus, cervix), shown by gene

expression clustering in FUMA (Supplementary Figure 12). Genes were enriched for androgen receptor targets and in several GO adhesion related pathways (Supplementary Table 14) and multiple protein-protein interactions were identified in STRING, the network having significantly more interactions than expected ($p=7.71x10^{-3}$) (Supplementary Figure 13).

Variants significantly associated with both DNAm and risk of endometriosis are also associated with other traits and diseases providing insight into possible shared underlying pathways. Other reproductive traits associated with these SNPs include uterine fibroids, female genital prolapse, ovarian cancer, age at menarche and bilateral oophorectomy (Supplementary Table 11).

Proportion of disease variance captured by endometrial methylation

Using residual maximum likelihood analyses we estimated the proportion of variation in endometriosis case-control status captured by all genome-wide DNAm sites and compared this to the variation captured by genetics (Table 4). Estimates of the variance captured by common SNPs (26.2% on the liability scale) was consistent with previously reported SNP-based heritability estimates[42]. The variance captured by SNPs can be interpreted as associated through causality however, variance captured by DNAm can reflect both causes and consequences of disease and therefore is dependent on the proportion of cases in the sample. As such, estimates for DNAm are read on the observed scale. We estimated that 24.2% of the variance was captured by all genome-wide DNAm sites. When including the genetic relationship matrix in the model, to account for variance captured by SNPs, the amount of variance explained by DNAm dropped to 16.1%. The proportion of variance captured by DNAm sites can be further driven by genetics given the presence of endometrial mQTLs. To

estimate of the proportion of variance captured by DNAm, independent of known genetic regulation, DNAm sites with mQTLs were removed and the variance captured by DNAm was re-calculated. In the absence of genetically regulated DNAm sites 15.4% of the variance was captured by DNAm. In total 36.9% of the variance in endometriosis case-control status was captured by a combination of both common genetic variants (21.5%) and DNAm in endometrium (15.4%).

Differential DNAm by endometriosis-related surgical and pain sub-phenotypes

We investigated differential DNAm signatures associated with endometriosis subphenotypes: disease stage (I/II and III/IV), lesion type (endometrioma, superficial, deep disease) and three different pelvic pain symptoms: dyspareunia (pain with intercourse), acyclic pain, and dyschezia (bowel movement pain). To minimize the number of tests conducted and maximize the a-priori chance of detecting signals, we focused the analysis on 11,698 sites within 500Kb± of 44 SNPs previously found to be genome-wide significantly associated with endometriosis[40] (Supplementary Table 15).

Adopting a genome-wide Bonferroni-corrected threshold ($p<4.3x10^{-6}$) across all comparisons, only one DNAm site - located in *ADAMTSL2* on chromosome 9 (cg13469396) - was significantly different, comparing cases with acyclical pain only vs. NUPP controls. When a less stringent threshold was applied ($p < 0.05/$ N of DNAm sites per GWAS locus) a total of 66/11,698 DNAm sites showed differential DNAm with at least one of the sub-phenotypes (Figure 6, Supplementary Table 16). These included five differentially methylated sites in stage III/IV endometriosis vs. controls, one in stage I/II endometriosis vs. controls (Figure 6: Blue panel, Supplementary Table 16). No sites reached statistical significance in stage III/IV vs. stage I/II. When considering more detailed surgical features of

endometriosis (e.g., endometrioma, superficial, deep), 29 sites were differentially methylated (Figure 6: Purple, Green and Red panels). Fourteen of these had mQTLs; however, no mQTL-regulating SNPs were in high LD (r^2>0.8) with a lead endometriosis-associated risk variants in these regions.

When considering pain subphenotypes, 23 DNAm sites were associated with dyspareunia and acyclical pain, including 18 with mQTLs (Figure 6: Dark blue panel & pink panel; Supplementary Table 16). One DNAm site associated with acyclical pain was located in a predicted enhancer in uterus with links to *WT1*, a transcription factor that plays an important role in gonad development and cellular development and survival and has been linked to endometriosis[43,44] (Figure 6: Pink panel, Supplementary Table 16).

Lastly, we focused on cases with endometriosis-associated painful bowel movement (dyschezia), as a symptom previously shown to be predictive of endometriosis[45] and hypothesized as a potential indicator of rectovaginal deep disease. Analysis of all dyschezia vs. controls revealed the largest number of differentially methylated sites – 11 in total (significantly more than expected by chance: binomial test p-value=1.41×10^{-5}) (Figure 6: Yellow panel, Supplementary Table 16). One of the differentially methylated sites, cg01340163, located in the *KDR* gene promoter on 4q12, has an associated mQTL in endometrium where the mQTL SNP (rs12331597) is in high LD (r^2=0.97) with an endometriosis-associated variant in this locus (rs1903068) (Supplementary Figure 14).

There was no evidence for systematic differences in effect sizes across all 66 DNAm sites comparing endometrioma or deep endometriosis with stage I/II vs. stage III/IV

(Supplementary Figure 15a,b); or comparing effect sizes for dyschezia, dyspareunia, or acyclic pain with any of the surgical subtypes (Supplementary Figure 16a-c).

Discussion

Epigenetic regulation in endometrium has the potential to influence vital aspects of endometrial function, female fertility, and disease. This study identifies significant differences in DNAm in endometrium across the menstrual cycle and in association with genetic variation. Integration with genetic data identified 118,185 independent *cis*-mQTLs including >3,000 potential tissue-specific mQTLs and numerous mQTLs associated with endometriosis risk. Differences in DNAm detected between clinically defined endometriosis subtypes also highlight the potential influence of disease heterogeneity on DNAm profiles in endometrium. To our knowledge this study is the largest collection of DNAm data from human endometrium with associated detailed clinical information, a powerful resource for advancing reproductive medicine.

Numerous DNAm sites distributed across the genome are differentially methylated in human endometrium across the menstrual cycle, with variation potentially driven by cyclicity of cell differentiation and turnover and tissue-specific response to circulating steroid hormones. Differences observed in this study are consistent with those reported in smaller studies on endometrium with the addition of many more differentially methylated sites as a result of increased power[21,25,46-48]. Evidence suggests that differences in DNAm profiles across the menstrual cycle is a phenomenon that may be specific to endometrium and is not observed in blood[25]. Pathways enriched for genes annotated to DNAm sites differentially methylated between the proliferative and secretory phases reflect changes in endometrial function including cell cycle progression (RHO and CDC42 GTPase cycle) and tissue remodeling

(regulation of actin cytoskeleton and ECM organization). Twenty one genes annotated to DNAm sites differentially methylated across the menstrual cycle are present on the Endometrial Receptivity Array (ERA)[49], which contains 238 genes in total. One of these genes (*ABCC3*) was annotated to a DNAm site differentially methylated between early and mid-secretory phases, suggesting that changes in DNAm may also influence endometrial receptivity to embryo implantation and/or the accuracy of estimating the window of implantation using these genes. Single-cell transcriptomics supports the potential role of *ABCC3* in endometrial receptivity showing it has peak expression during the WOI in endometrial epithelial cells [50]. Consideration of differences in DNAm across the menstrual cycle is vital when assessing effects of regulation of molecular mechanisms on endometrial pathologies and fertility traits.

Epigenome-wide association studies (EWAS) test associations between differentially methylated DNAm sites and complex diseases[51,52]. The biological relevance of the tissue under analysis is crucial for understanding the mechanisms involved in endometriosis. We conducted this study in endometrium as a strong candidate tissue for the origin of cells that initiate endometriosis lesions[53,54]. While single site analysis did not reveal statistically significant differences in DNAm in endometrium from women with and without endometriosis, consistent with a previously study in 66 women[25], differences in methylation at four DNAm sites was detected when cases were restricted to stage III/IV disease. Larger molecular effects in severe disease are consistent with genetic studies that show evidence of larger genetic effects in stage III/IV endometriosis. One of the genes annotated to these sites, *EEF2*, is downregulated in ectopic endometriosis lesions[55]. Differences in methylation at previously identified candidate genes, including *HOX-A10*[13], *PR*, *ESR1*[14,15], *CYP19*[16], *SF-1*[17,18], *COX-2*[19] and *DNMT*s[20], were not replicated in this study. Interestingly, gene

network analysis identified modules of DNAm sites associated with presence or absence of endometriosis. Genes annotated to these modules were enriched in WNT and MAPK signaling, adhesion and cancer pathways. Both the WNT and MAPK signaling pathways have been associated with endometriosis in previous expression studies and have been identified as potential treatment targets[56-59].

Variation in DNAm between individuals and populations can arise as consequences of environmental exposures, stochastic and genetic perturbations[60,61] and both causes and consequences of disease [62-64]. Previous studies have estimated the amount of variation in complex traits captured by methylation, estimates ranging from 0% for height, 6.5% for BMI[65] and 15-31% for amyotrophic lateral sclerosis[66], to almost 100% for age. DNAm in endometrium was estimated to capture 15.4% of the variance in the presence or absence of endometriosis in this study. This suggests that although we do not see large differences in DNAm between cases and controls, many small effects of methylation across the genome, or within pathways or networks of genes, may contribute to this variation. Consequently, larger sample sizes and/or network-based and genome-wide approaches are likely needed to accurately estimate these effects.

Characterizing endometrial specific genetic regulation is important when investigating functional effects of gynecological disorders and fertility traits. The majority of mQTLs we identified in endometrium have also been reported in blood, suggesting the genetic regulation of these DNAm sites is shared between tissues, consistent with tissue overlap reported in previous studies[25,37]. However, we did identify genetic regulation for 3,480 DNAm sites that have not been previously reported in other tissues. Genes annotated to these potential endometrium-specific mQTLs were enriched in pathways known to be important in

endometrial biology including estrogen and androgen response, regulation of epithelial to mesenchymal transition (EMT), cell adhesion, calcium signaling and ECM receptor interaction, important for endometrial tissue integrity and differentiation, decidualization, embryo implantation, and establishment of early pregnancy[67-71].

Information across omic datasets (GWAS and mQTLs) enabled identification of 51 mQTLs, distributed in 21 genomic loci, that were significantly associated with endometriosis risk from the SMR analysis. Associations between endometriosis and QTLs at 10 loci are novel and have not previously been reported[24,25,39,40]. An atlas of genetic effects on DNAm in blood generated from 32,851 individuals demonstrated that although many genetic signals are shared between DNAm sites and complex traits, altered DNAm signals are not necessarily on a causal path from genotype to phenotype[34]. The complex relationship may instead be driven by genetic effects on other regulatory factors that affect DNAm and the trait through diverging pathways (horizontal pleiotropy), reverse causation, cell-type or context specific effects or a non-linear causal path involving several interactions of regulatory features[34]. Context-specific genetic effects on DNAm in endometrium for several mQTLs varied between menstrual cycle stages and endometriosis cases and controls, warranting replication in independent datasets. Although we are unable to distinguish causal and pleiotropic pathways, we identified several endometrial mQTLs that were statistically significantly associated with endometriosis risk, providing further functional evidence for endometriosis target genes than previous studies in eQTLs[24,39,40].

Examples of new evidence for known endometriosis target genes that warrant further discussion include *GREB1*, and genes on chromosomes 1, 6, 12 and 17. *GREB1* is an estrogen- and androgen-regulated gene which mediates cell proliferation and migration, and

induces EMT in ovarian cancer[72] and hormone-stimulated proliferation in breast cancer[73]. mQTLs significantly associated with endometriosis featured DNAm sites and SNPs located in predicted enhancers that interact with *GREB1* in uterus and ovary including rs1865573 that is associated with a splicing (s)QTL for *GREB1* in ovary reported in GTEx[74], suggesting a possible role of alternative splicing of *GREB1* in endometriosis susceptibility. Signals for mQTLs significantly associated with endometriosis on chromosome 1p36.12, including two potentially endometrial-specific mQTLs, were located in regulatory regions and associated with expression of known genes of interest *LINC00339*, *CDC42* and *WNT*[24,39-41]. Similarly, potential endometrial-specific endometriosis-associated mQTLs lie in a predicted promoter for *CCDC170* and enhancer regions that may mediate regulatory effects on multiple genes in the chromosome 6q25.1–6q25.2 risk region. Expression of genes in this region are known to be highly correlated with nearby *ESR1* expression in endometrium[40,75]. Finally, endometriosis-associated mQTLs associated with expression of *GDAP1*, *SRP14*, and *HOXB9* are in moderate LD (r^2>0.6) with SNPs significantly associated with endometriosis and with eQTLs for these genes in the recent endometriosis meta-analysis[40].

Genetic regulation of DNAm in endometrium may also influence other reproductive pathologies including cancer. Genetic variants significantly associated with both endometriosis risk and DNAm in endometrium on chromosomes 10p12 and 12q22 were in LD with rs7084454 (r^2=0.89) and rs6538618 (r^2=0.99) respectively, risk SNPs shared between endometriosis and epithelial ovarian cancer. These mQTLs are located in predicted promoter regions for *MLLT10* and a bidirectional promoter for *VEZT* and *FGD6* and the latter is associated with expression of both genes in endometrium[76]. Epidemiological studies report an increased risk of ovarian cancer in women with endometriosis and recent genetic

studies report a strong genetic correlation and causal relationship between endometriosis and specific epithelial ovarian cancer subtypes[76-78].

New endometriosis target genes were detected, including *EEFSEC* (eukaryotic elongation factor, selenocysteine-TRNA specific) on chromosome 3q21.3, *SRD5A3* (steroid 5 α-reductase 3) on chromosome 4q12, *ADK* (adenosine kinase) on chromosome 10q22.2, and HOXC cluster on chromosome 12. Endometriosis associated mQTL SNPs on chromosome 3q21.3 in *EEFSEC* are in LD with variants previously associated with risk factors for endometriosis patients (i.e., gestational length, spontaneous preterm birth[79] and age at menarche[80,81]). *E*pigenetic marks and chromatin interactions suggest these mQTLs fall in *EEFSEC* promotor and enhancer regions in uterus and ovary, where the gene is highly expressed[82], and immortalized endometrial cells[83,84]. In addition, methylation at DNAm sites located near the *KDR* promoter on chromosome 4 are associated with endometriosis risk and expression of nearby *SRD5A3* in blood. *KDR* encodes vascular endothelial growth factor receptor 2 (*VEGFR-2*), a major mediator of angiogenesis, proliferation, migration, and differentiation of endothelial cells. Dysregulation of angiogenic activity in endometriotic lesions and the eutopic endometrium of women with endometriosis and is mediated by VEGF signalling[85,86].

Several mQTLs specifically associated with severe (rASRM stage III/IV) endometriosis including five on chromosome 10q22.2 are also associated with eQTLs for *ADK*, an enzyme that regulates concentrations of extracellular adenosine and intracellular adenine nucleotides. Evidence from epigenetic marks suggest that one of these DNAm sites is located in the promoter region of *ADK* in fibroblasts, uterus and ovary and is in an anchor point for chromatin interaction with *ADK* in immortalized endometrial cells [83,84]. Notably, adenosine

nucleotides as targets for endometriosis pain management are currently under investigation[87]. Whether endometrial mQTLs or eQTLs for *ADK* influence eutopic endometrial tissue-specific inflammation in women with endometriosis is an attractive hypothesis and warrants further study[87].

Detailed differential DNAm sub-phenotype analyses, based on surgically visualized disease phenotypes, endometriosis-associated pain symptomatology, and rASRM staging, could facilitate understanding mechanisms underlying particular biological sub-types of endometriosis and lead to stratified diagnosis and therapeutic target discovery. To maximize power of detection, given the relatively small sample size of sub-phenotype groups ranging from 50-600 cases, we focused sub-phenotype analyses on differential DNAm signatures linked to 44 endometriosis-associated GWAS loci[40]. The results showed multiple differentially methylated sites associated with surgical phenotypes in particular with rASRM stage III/IV and endometrioma, suggesting different genetic regulatory mechanism for this particular sub-type of endometriosis. Several of the associated DNAm sites were located in enhancer and promoter regions in uterus including enhancers predicted to interact with *WT1* on 11q14, important in gonad development, and promoters for genes associated with genitourinary anomalies (*MMPED2*). Immunostaining evidence suggests *WT1* is downregulated in the eutopic endometrium from women with endometriosis but is expressed in neurons of deep endometriosis[43,44]. Moreover, we identified that variation of DNAm at cg24360069, in *STK3* on 8q22, was associated with a greater extent of endometriotic pelvic disease, presence of dyspareunia, and had an mQTL in high LD with the lead GWAS variant at this locus. *STK3* encodes Serine/threonine-protein kinase 3, which is part of the Hippo pathway that plays an important role in tumour suppression by reducing cellular growth and promoting apoptosis. Notably, for the pain subphenotypes, dyschezia had the largest number

of DNAm sites. One of these, located in the promoter of *KDR*, had an mQTL in LD with the lead GWAS variant. The single genome-wide significant DNAm site, associated with acyclical pain, was located within an exon of *ADAMTSL2* on chromosome 9, a disintegrin and metalloproteinase with thrombospondin motifs and is suspected to interact with latent transforming growth factor beta binding protein 1.

The ability to detect differentially methylated sites associated with disease may be attributed to amount, duration, and/or time of critical exposures and the expected effect size of these exposures. DNA methylation is a process operating from early development and continuing throughout the lifespan of an individual. The critical window to quantify etiologically relevant differential methylation at DNAm sites in endometrium associated with risk may be years prior to overt symptoms or diagnosis – commonly during adolescence or prior to thelarche[88] and could be mediated over time. Endometriosis case vs. control differences may also be modified by endometriosis disease heterogeneity or mediated by previous use of medications or behavioral changes to reduce symptoms[89,90]. Without independent validation, results from this study are not generalizable to patients currently receiving hormone treatment, as exposure to hormonal treatments may contribute to variation in DNAm profiles. DNAm differences between cases and controls, limited to a single cell type within the endometrium, cannot be excluded by the present study. Future investigations into cell type-specific effects may be facilitated by emerging technologies in single-cell epigenetics.

At a genome-wide level, we did not observe statistically significant differences in DNAm in endometrium from women with and without endometriosis. Recent studies have identified few DNAm changes associated with complex diseases[52]. Effects of nominally significant differentially methylated sites between endometriosis cases and controls are relatively small

(max delta-beta=0.06, mean=0.01). Although our study is the largest of its kind to date, an estimated 4,000 samples would be needed to achieve 80% power at >80% DNAm sites to identify true genome-wide case:control differences given our results[91]. The ability to detect differences at different informative points along the life course, with larger sample sizes, particularly within case sub-phenotype strata[92] and among NUPP controls[93], and with data that allow for fine adjustment for potential confounders defined proximal to endometrial sampling, remains to be tested.

This study presents the largest and most comprehensive analysis of regulation of DNAm in the human endometrium. Menstrual cycle stage is a major source of variation in DNAm in endometrium – providing strong support for hormonally-driven differences in DNAm across the cycle being associated with known changes in gene expression and pathways responsible for endometrial physiology, function, and dysfunction. We found no evidence for large site-specific DNAm differences between endometriosis cases and controls. However, correlated network analysis identified networks of DNAm signals relevant to disease pathophysiology that differed between disease and control. We also report novel and disease-associated genetic effects on DNAm in endometrium that provide new insights into epigenetic regulation of pathways important for endometrial biology and identify target genes with a potential role in the causal pathway between genetic variation and endometriosis pathogenesis. Functional validation of these genomic targets may yield new therapeutic targets with the potential to disrupt pathogenic processes. Possible differences in DNAm associated with disease sub-phenotypes highlights the need to generate larger datasets with sample collection timed with menstrual cycle phase and comprehensive clinical information. Validation of methylation profiles associated with sub-phenotypes of disease may also allow molecular subtyping of disease that can be associated with patient outcomes to guide

personalized disease management. Findings from this study will direct future endometriosis research and datasets generated will be a valuable resource for subsequent investigations into tissue-specific effects of methylation on endometrial biology and disease, and development of potential novel, targeted therapeutics.

Methods

Sample collection and clinical data standardization

Study participants

Endometrial tissue from 679 surgically-diagnosed endometriosis patients (cases), 389 controls without endometriosis and six participants with unconfirmed endometriosis status were recruited through the University of California San Francisco, California (UCSF, n=480 samples), University of Melbourne, Melbourne, Australia (UM, n=315 samples), Endometriosis CaRe Centre in Oxford, Oxford, UK (ENDOX, n=193 samples), and EXPPECT Centre, The University of Edinburgh (EDIN), Edinburgh, Scotland, UK (n=86 samples), with collection at all sites using the World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonization Project (WERF EPHect) standardized protocols for tissue collection and processing, and participant characteristics and clinical annotation (Supplementary Table 1)[26-28]. Participants were restricted to those who had not been on contraceptive steroids or gonadotropin releasing hormone analogues for 3 months or more prior to endometrial sampling, had regular cycles (defined as 24-35 days in length) and no evidence of endometrial hyperplasia or cancer. Women without visualized endometriosis at the time of surgery or without a history of endometriosis were defined as controls: some of these had pelvic pain and documented uterine fibroids or other non-malignant gynecologic conditions including endometrial polyps, ovarian cysts, cervical abnormalities, and pelvic organ prolapse; a subset was defined as “no uterine or pelvic

pathology" (NUPP) controls (n=201) if they had no visualized or documented uterine or pelvic pathology. After quality control described in detail below, the study included DNA from a total of 984 endometrial tissue samples.

All participants provided the site-specific study investigator with informed consent. All patient data were de-identified and followed HIPAA and the Convention of the Declaration of Helsinki. This study was approved by the institutional review boards of UCSF (Administrative Multi-Principal Investigator site), Michigan State University (Multi-Principal Investigator site), University of Oxford, University of Melbourne and University of Edinburgh.

Tissue Collection and Processing

Each participant contributed a single endometrial tissue sample. Primary tissue samples were stored as fresh frozen (FF) specimens in liquid nitrogen or at -80^{o}C. Samples from the contributing sites were combined at UCSF into two batches, prepared consistently and randomized among plates. Batch I consisted of 767 samples (474 cases; 2 unknown; 291 controls) from UCSF, UM, and ENDOX. Batch II consisted of samples from additional recruitment and included 307 samples (205 cases; 4 unknown; 98 controls) from all four sites (Figure 1). The total number of cases and controls (all types) was 679 and 389 respectively. Samples in both batches were analyzed for genotyping and DNAm using the same platforms. All DNA samples were adjusted to the same volume and concentration for sodium bisulfite conversion followed by DNAm analysis on the Illumina Infinium MethylationEPIC Beadchip (Illumina, San Diego, CA) at the University of Southern California Epigenome Center Core Facilities, Los Angeles, CA. They were assessed for amount and completeness of sodium bisulfite conversion using a panel of MethyLight reactions as described previously[9]. In brief,

a MethyLight reaction for a genome-wide distributed multicopy *ALU* sequence that is bisulfite-dependent but DNAm-independent[94] assessed the integrity and quantity of the DNA samples. The completeness of bisulfite conversion was assessed by three bisulfite-dependent reactions measuring 0%, 50%, and 100% bisulfite conversion for each sample[94]. Samples were required to pass all quality control (QC) metrics [high DNA integrity (*ALU* CT<25), 100% bisulfite conversion with no amplifications at 0% or 50% conversion] before quantitative assessment of DNAm on the Illumina Infinium platform.

Cycle phase was assigned using the criteria of Noyes, et al. [95]: menstrual, early proliferative (EP), mid-proliferative (MP), late proliferative (LP), early secretory (ESE), mid-secretory (MSE), and late secretory (LSE) for all specimens. Not all cohorts had assigned proliferative substages and as such all proliferative samples were consolidated as "PE". A small number of samples (n=23) were not assigned a secretory substage and were assigned "SE". If two phases were found in reports or on review of histology, the later phase was selected (e.g. LSE/Menstrual → Menstrual; PE/ESE → ESE, Interval → ESE). Serum estradiol (E_2) and progesterone (P_4) facilitated phase assignments, and sometimes two or more pathologists re-reviewed the histology. Unsuitable samples (inactive, atrophic, PE/SE, progestin effect, dyssynchronous) were excluded. "Benign" histology descriptor and "unknown" were assigned as "unknown" in the absence of last menstrual period (LMP) and/or serum E_2 and P_4 levels (assayed at University of Virginia NIH National Institute for Child Health and Human Development (NICHD) Ligand Core). Endometrial samples from women with a history of endometriosis but no disease identified at surgery were not considered as controls and were not suitable for DNAm quantification, and thus were excluded from that analysis. All four sites completed histology assessments and/or cycle phase determinations for their respective sample specimens.

Participant Characterization

Variables across all site-specific datasets were combined and harmonized (listed in Table 1). The following woman-level covariates were included in statistical modeling. *Site*: A categorical variable with a unique value for each of the five sample contributing sites/institutions (UCSF, ENDOX, UM, EDIN). *Cycle phase*: A categorical variable with a unique value for each of the six menstrual cycle phases defined previously (Menstrual, PE, SE, ESE, MSE, LSE). *Endometriosis case:control status:* A binary variable assigning samples as either a case or control. *Sample plate:* A categorical variable with a unique value for each of the 12 sample plates used during processing. *Batch:* A binary variable assigning samples as either Batch I or II.

Endometriosis sub-phenotyping characteristics included: *rASRM endometriosis disease stage:* visualized at surgery most proximal to endometrial biopsy collection and defined by the rASRM endometriosis scoring system[29]. Stage data were used to create variables with three structures – continuous rASRM score, ordinal stages I, II, III, IV, and dichotomized as stage I+II and stage III+IV. For case patients for whom surgical documentation was noted as the rASRM stage category only, the variables were categorized as documented. However, four patients were defined in their surgical record as having visualized stage II-III and were assigned to the I-II dichotomized rASRM category. *Lesion type*: categorized according to the presence of at least one superficial peritoneal lesion, endometrioma, or deep lesion. Lesion types were binary variables coded as "any" peritoneal lesion, endometrioma, or deep lesion, regardless of co-occurrence of another lesion type and not mutually exclusive variables. *Pain:* binary variables for the presence or absence of dyspareunia, acyclic pelvic pain and dyschezia (see Supplementary Table 15 for detailed descriptions).

Data quality control and processing

DNAm data quality control

DNAm data were generated using the Illumina Infinium MethylationEPIC Beadchip (Illumina, San Diego, CA) at USC Epigenome Center Core Facilities, with data on 865,859 DNAm sites for 767 samples in batch I and 307 in batch II (total n=1,074). DNA from only one endometrial tissue sample per participant was included. DNAm QC and processing were conducted on the dataset through a series of steps (Figure 1). The *openSesame* function from the *SeSAME*[96] R package, a pipeline for Illumina Infinium human methylation processing, was used to perform background correction with dye-bias and signal intensity normalization starting from the idat files, and Beta value generation. The following filtering steps for samples and DNAm sites/probes were carried out. Samples with a low overall intensity signal, defined as a median unmethylated or methylated signal <9, were removed from the dataset. In addition to this, samples were also filtered out if they had a detection p-value >0.05 in more than 1% of DNAm sites and probe level quality control and processing were done by filtering out probes with a detection p-value >0.05 in more than 10% of samples (n=1,727 probes in batch 1, n=1,461 probes in batch II). Probes were masked based on the *SeSAME* package. These included the 59 Illumina tagged probes and three classes of probes that could be potentially problematic or ambiguously mapped (n=104,671 probes in batch I and n=104,752 probes in batch II) as well as chromosome Y probes (n=26 probes*)*. Finally, the *minfi*[97] function '*getSex*' was used to generate a predicted sex for each sample, in order to ensure all samples were confirmed as genetically female based on the median values of measurements on the X and Y chromosomes (n=0 samples were removed). The next QC stage involved probe removal of non-overlapping probes in the data between batch I and II (n=275 probes). These sample and probe level QC steps resulted in a processed dataset of

759,345 DNAm sites for n=707 and n=277 endometrial samples in batch I and II, respectively, with a final total of 984 samples (from 637 cases and 347 controls)(Table 1).

Genotyping Data Quality Control

Batch I and II samples that passed the DNAm QC were genotyped using the Axiom Precision Medicine Research array. Quality control of the data was conducted in three stages; (1) batch I samples (n=707 individuals), (2) batch II samples (n=277 individuals), (3) Merged batch I and II samples. In each stage (i) per-individual QC included identification and filtering of samples with genotype call rate < 95%, heterozygosity rate > 3 standard deviations away from the mean heterozygosity rate and removal of related (IBD > 0.200) samples, (ii) per-variant QC included filtering variants with MAF < 1%, call-rate >95% and HWE p-value < $1x10^{-5}$. Post-QC combined datasets included 953 individuals (614 cases, 339 controls) and 621,613 variants. Post-QC, data from all ancestries was pre-phased using SHAPEIT2 and imputed using the 1000 Genomes reference (1000G P3v5) all together.

Genetic ancestry determination

Genetic ancestry of each study participant was determined using 1000 Genomes P3v5 reference data. In brief, principal component analysis was conducted with common markers between the study samples and the 1000G P3v5 reference data. Then, the first 10 principal components were plotted against each other to identify population clusters in our data against the 1000G P3v5 reference data with 5 super populations including European, Eastern Asian, American, African, Southern Asian. Any samples that did not cluster with the 1000G P3v5 data were assigned to the admix category. This revealed that in the post-QC combined dataset, there were 658 European, 76 East Asian, 46 South Asian, 51 Admixed American, 47 African and 75 Admix individuals (Supplementary Figure 1A). For the purpose of

establishing genetic ancestry only, we also conducted a principal component analysis (PCA) of the genotypes of 31 samples that had failed genotype QC but did have DNAm data passing QC, bringing the total of women with genetic ancestry information to 677 European, 78 East Asian, 52 Admixed America, 48 African, 47 South Asian and 82 Admix individuals (Table 1).

Analysis of DNAm data

After QC and processing, principal component analysis (PCA) was performed using the M-values from the DNAm dataset and PCA plots were generated (Supplementary Figure 1B&C and 2). These plots were analyzed for any potential batch effects and differential clustering of the samples according to covariates. Covariates included the sample processing batch (batch I=707 samples; batch II=277 samples), genetic ancestry, cycle phase, endometriosis case:control status, site, and plate. Furthermore, we performed correlation analysis between covariates and principal components (PCs) (Supplementary Figure 3) to evaluate the importance of covariates in the analysis moving forward. Statistically significant associations of the top 10 PCs with batch, site and plate were identified. There was a partial association with cycle phase, and thus batch, site, plate, and cycle phase were included *a priori* as covariates in the downstream analyses.

Batch correction

In order to remove batch effects coming from covariates such as batch, institute, plate, and genetic ancestry, we used the SVA algorithm using R package *SmartSVA*[98]. First, the number of surrogate variables (SVs) was evaluated, and then the SVs were calculated while protecting contrasts for two variables - endometriosis case:control status and menstrual cycle

phase at endometrial sample collection - as those were the two variables we were interested in exploring biologically. SVA was run separately for the stratified analyses.

Single site analysis

In order to elucidate the association of DNAm at a single DNAm with endometriosis case:control status and menstrual cycle phase, we applied a linear model using the Limma[99] R package with the significance cutoff of Benjamini Hochberg corrected p-value of 0.05. The following model was used: DNAm values ~ endometriosis case:control status + Cycle phase + SVs and interrogated via contrasts to capture case:control and cycle phase differences. All 347 women without endometriosis were included as controls. Both cases and controls were included in the cycle phase analysis. Data were visualized using heatmaps with the *Complex Heatmap*[100] package in R. Endometriosis case vs. control and cycle phase differences were also tested using a conservative mixed linear model (MLM) based approach, MLM-based omic association (MOA)[65] (Supplementary Note 1).

Cycle phase comparisons were based on aggregated cycle definitions – secretory (SE) vs. proliferative (PE) vs. menstrual, where secretory consists of early (ESE), mid (MSE) and late (LSE) secretory phases, and also on more finely categorized cycle phase definitions using the secretory sub-phases (ESE, MSE and LSE). A sensitivity analysis was carried out for a endometriosis case vs. control comparison using the 201 pathology-free controls. Given that larger genetic effects are observed in rASRM stage III-IV disease compared to stage I-II endometriosis[31], we repeated the analysis restricted to stage III-IV cases compared to pathology-free controls.

Pathway analysis

Pathway analysis was carried out on all significant DNAm sites (FDR < 0.05) as input into the analysis and significance cutoff on the pathway overrepresentation of FDR < 0.05. Statistically significant DNAm sites were mapped to genes using the *IlluminaHumanMethylationEPICanno.ilm10b4.hg19 annotation*[101] R package. The DNAm sites were mapped to genes if they were upstream of the gene body (TS1500, TS200, 5'UTR, 1stExon) or within the gene body. The resulting gene sets were analyzed for pathway overrepresentation using the *enrichKEGG* and *enrichPathway* functions from the *clusterProfiler* R package[102]. The over-representation analyses were also run using stratified gene mapping schemes based on the genomic features (upstream or gene body) from which DNAm sites were mapped. The analysis was carried out using the KEGG[103] and Reactome databases[104]. The top 5 statistically significant pathway gene sets were visualized using the *emapplot* function from the *enrichplot*[105] R package.

Regional analysis

Comparisons made in the single site analysis were repeated using regional analysis. Regional analysis was carried out with an R package *DMRCate*[106]. A nominal p-value cutoff of 0.05 was used for the input DNAm sites into the analysis and Fisher p-value of < 0.1 was used to identify significant regions.

Weighted Correlation Network Analysis (WGCNA)

Weighted correlation network analysis (WGCNA)[32] was applied to a reduced dataset using 50% of most variable DNAm sites resulting in a dataset of 379,672 DNAm sites from the 984 samples. First, co-methylated modules were identified using the '*blockwiseModules*' command[32] with minimum block size of 30 and maximum block size of 40,000 selected. Associations between the two traits of interest (endometriosis case:control status (all 347

controls) and menstrual cycle phase) and the modules were identified through use of the *Limma*[99] R package linear models using the SVs from SVA as covariates. Statistical significance was defined as a Benjamini-Hochberg adjusted p-value of < 0.05. Significant modules were aggregated into gene sets. More specifically, for each linear model, all DNAm sites in modules with positive direction of effect were mapped to genes using the *IlluminaHumanMethylationEPICanno.ilm10b4.hg19 annotation*[101] and aggregated into a gene set, and all DNAm sites in modules with negative direction of effect were mapped to genes and aggregated into a gene set. These gene sets were then tested for overrepresentation using the same method as described above in the pathway analysis methods description.

Methylation Quantitative Trait Loci (mQTL) analysis

DNAm and genotype data from 658 European samples was used to test for associations between genotype and DNAm in endometrium (Table 2). DNAm M-values at a total of 759,345 DNAm sites, as determined in prior analyses, and genotype data for 5,290,992 single nucleotide polymorphisms (SNPs) with imputation quality info scores > 0.8 and MAF > 0.05, were included in the analysis. Associations between genotype and methylation at DNAm sites within a *cis* distance of 1Mb was carried out using a linear regression model in the *MatrixeQTL*[107] R package. Menstrual cycle phase, endometriosis case:control status and 39 SVA components were included as covariates in the model to adjust for known biological and technical variation. Women without endometriosis were classified as controls (n=221). A Bonferroni threshold of p-value < 1.7×10^{-11} was applied to account for multiple testing. Independent mQTL signals were identified using a stepwise model selection procedure in GCTA[108] for Bonferroni significant DNAm sites. Due to extensive storage requirements to compute the genomic inflation factor (λ), λ was calculated based on associations of 759,345

DNAm sites with a random subsample of 100,000 SNPs selected from the 1000 Genomes Project panel (EUR phase 3).

Overlap with mQTLs in other tissues

We compared mQTLs identified in endometrium to those that have been reported as significant in blood[33,34], skeletal muscle[35], adipose tissue[36], and brain[37]. All datasets were previously generated using the Illumina HumanMethylation450 array with the exception of significant blood mQTLs reported by Hannon *et al*. who used the Illumina EPIC array allowing comparison of the additional DNAm sites on this array. Full summary statistics were only available for blood mQTLs published by Min *et al*[46], allowing us to match *cis*-mQTLs between the tissues based on the same eSNP and DNAm site associations and direction of effect. We also limited the analysis to those DNAm sites and SNPs present in both datasets to calculate the proportion of endometrial mQTLs shared with blood. For the remaining tissues where the full SNP information was not available, we matched *cis*-mQTLs between the tissues based on DNAm site. Genes annotated to DNAm sites with mQTLs unique to endometrium were included in a pathway analysis in FUMA.

Context specific mQTLs

To investigate if the effects of mQTLs differed between women with and without endometriosis and between proliferative and secretory menstrual cycle phases, we conducted context specific mQTL analyses. A total of 118,185 independent Bonferroni significant mQTLs were included in the analysis. Anova was used to compare the following two models:

H0: DNAm ~ genotype + condition + covariates

H1: DNAm ~ genotype + condition + condition*genotype + covariates

Conditions tested included endometriosis case vs. control and proliferative (PE) versus secretory (ES+MS+LS) menstrual cycle phases. SVA components and either menstrual cycle phase or endometriosis, depending on the condition tested, were included as covariates. H1 differed from H0 by the inclusion of an interaction term between genotype and the condition being tested. To account for multiple testing, statistically significant context specific mQTLs for each condition were defined as those with a p-value < $4.2x10^{-7}$ (0.05/118,185). Results were filtered to only include comparisons that had at least 10 samples homozygous for the minor allele in each group.

Association between genetic regulation of DNAm, transcription and endometriosis

Summary data-based Mendelian Randomization (SMR)[38] was used to test the association between genetic variants, DNAm levels and endometriosis risk by integrating mQTL summary statistics with endometriosis GWAS summary statistics. Summary statistics used were generated from a subset of European cohorts from the largest endometriosis GWA meta-analysis to date conducted by Rahmioglu, et al. [10] with the addition of a FinnGen endometriosis cohort (Supplementary Table 17). Datasets that were used in generation of the mQTL analysis (ENDOX, Melbourne, UCSF) were not included in the meta-analysis to ensure independent datasets. SMR was run for overall endometriosis (31,021 cases and 524,422 controls) and stage III/IV endometriosis (8,765 cases and 373,626 controls) separately. Of note, controls in the GWA meta-analysis were population controls who did not have documentation of endometriosis diagnosis. The HEIDI (Heterogeneity In Dependent Instruments) test was also conducted to distinguish independent overlapping signals and pleiotropy/causal associations. Associations with an SMR p-value of P_{SMR} <0.05/(number of

DNAm sites tested) and a HEIDI p-value of P_{HEIDI} >0.05/(number of genes passing the SMR test) was applied as the threshold for statistical significance.

Consistent association signals at genomic loci across multiple omics layers can help identify functionally relevant genes and regulatory elements. As such we also used SMR to test the association between mQTLs and eQTLs in endometrium[24,39], blood[109] and 49 GTEx Tissues[82] using the top associated mQTL in each dataset, and the association between eQTLs and endometriosis using the top associated eQTL. We combined results from pleiotropic associations between DNAm, gene expression and endometriosis using the stringent criterion that both the DNAm site and gene of each pair were statistically significantly associated with endometriosis at a genome-wide threshold with no association rejected by the HEIDI test. The gene–endometriosis association analysis was not dependent on the DNAm–endometriosis association analysis to account for potential discrepancies in SNPs tagging the causal SNP between the datasets.

Association between SNPs significant in the SMR analyses and other traits and diseases was investigated using PhenoScanner[110] and GWAS Catalog[111]. EpiMap[83] was also used to functionally annotate SMR significant SNPs and DNAm sites using epigenome maps from relevant tissues (uterus, ovary, vagina, mammary/breast epithelium, mammary/breast fibroblasts, skin epithelium, skin fibroblasts, Ishikawa cells, T47D cells) to define chromatin states, high-resolution enhancers, enhancer modules, upstream regulators and downstream target genes. Valid promoter associated chromatin loops generated from H3K27Ac HiChIP libraries from a normal immortalized endometrial cell line (E6E7hTERT)[84] were also used to annotate SNPs and DNAm sites located in chromatin interaction anchor points.

Estimation of variance captured by genetics and DNAm in endometrium

Genome-based restricted maximum likelihood (GREML), as implemented in the Genome-wide Complex Trait Analysis (GCTA) software[112], was used to estimate the variation in endometriosis case-control status captured by common genetic variants (SNPs), also known as the SNP-based heritability. The genetic relationship matrix (GRM) used in the GREML analysis was calculated using genotype data for 5,358,309 SNPs in 953 individuals in GCTA. Similarly, omics residual maximum likelihood analyses (OREML), as implemented in the Omic-data-based Complex Trait Analysis (OSCA) software[65], was used to estimate the proportion of variance in endometriosis case-control status captured by DNAm sites and SNPs. Three different OREML models were compared, one including all DNAm sites (n=759,345) for 984 individuals in the form of an omics relationship matrix (ORM), one including the ORM and the GRM for 953 individuals with matched genetic and DNAm data and one including the ORM, calculated following exclusion of DNAm sites with mQTLs (n=651,718), and the GRM for the 953 individuals.

Targeted differential DNAm analysis investigating associations with endometriosis surgically visualized and endometriosis-associated symptom sub-phenotypes

Differential DNAm analysis for 11,698 DNAm sites within 500Kb± of 44 lead SNPs genome-wide statistically significantly associated with endometriosis[10] was conducted for 11 comparisons of endometriosis-case-specific surgical sub-phenotypes (rASRM stage I/II disease, stage III/IV disease; presence of at least one superficial peritoneal lesion, endometrioma, or deep lesion; presence of at least one each of superficial peritoneal, endometrioma, and deep lesions) and 6 comparisons of common endometriosis-associated pain symptom sub-phenotypes (dyspareunia, acyclic pelvic pain, dyschezia) (see Supplementary Table 15 for detailed descriptions). These comparisons were restricted to

NUPP controls. Only 10% of cases and NUPP controls reported no dysmenorrhea (cyclic pain with menses), and therefore we could not investigate presence or absence of dysmenorrhea as an independent sub-phenotype. In the differential DNAm analysis, post-QC M-values from the DNAm data were utilized as described in the overall endometriosis genome-wide differential DNAm analysis methods above. Differential DNAm analysis was conducted using the Limma[99] R package. For each sub-phenotype analysis, an independent set of SVs was generated and included in the models. Both a genome-wide Bonferroni multiple testing correction ($p < 4.27x10^{-6}$) and a less stringent locus-specific Bonferroni multiple testing correction (0.05/N of DNAm sites per GWAS locus) was applied. The statistically significantly differentially methylated sites were checked in the endometrium mQTL map and SMR results generated as described in the previous section. If an mQTL was identified for a differentially methylated probe, whether the DNAm-associated SNP was in LD with an endometriosis-associated SNP in the respective region was determined. Associated DNAm sites were also annotated using EpiMap as described in the previous section. Correlation between the effect sizes of differentially methylated sites between subtypes were calculated utilizing Spearman's rank correlations.

Data Availability

Methylation data used in this study has been deposited and is available from GEO (accession number on acceptance). Genotype data generated in this study is available upon approval from dbGAP (accession number on acceptance). Code used to run the analyses is available on github (link).

Author contributions

Concept and design: LG, MS, SMi, GWM, KZ, PAWR, SH, SMo, CB; Sample procurement, clinical metadata review, sub-phenotyping CN, PAWR, KCV, EM, ZG, SS, AJ, JL, AWH, JOA, FC, PS, CB, XSM, JVJ; SH, JCI, AFV, KLT, SMi, MS, LG; Sample processing, DNA preparation, QA/QC, raw data generation: SH; Data QC for analysis: NR, SH, IK, SVA, MP, SMo, AFV, MS; Data analysis and interpretation: SMo, SH, IK, NR, AFV, SVA, PG, MP, AWH, CB, JCI, AFM, KK, PAWR, KZ, GWM, SMi, MS, LG; Human subjects assurance: KCV, JCI, LG, CB, PAWR, AWH, XSM, CN; Drafting of manuscript: SMo, SH, IK, NR, KZ, GWM, PAWR, KLT, MS, SMi, LG; All authors have read and approved the final manuscript.

Competing interests

Authors declare they have no competing interests.

Funding

This work has been supported by the National Institutes of Health (NIH) NICHD R01 HD089511. It was also supported, in part, by funding from Wellbeing of Women (through sponsorship from PwC) (R42533) and the Medical Research Council (MR/N024524/1 and MR/N022556/1) and NIH HD094842 (Harvard/MSU). KK was supported by NIH NCI R37 CA233774. AFM was supported by an Australian Research Council Future Fellowship (FT200100837). GWM was supported by NHMRC Fellowship (GNT1177194).

References

1 Rowlands, I. J. *et al.* Prevalence and incidence of endometriosis in Australian women: a data linkage cohort study. *BJOG: An International Journal of Obstetrics & Gynaecology* **128**, 657-665 (2021). https://doi.org:https://doi.org/10.1111/1471-0528.16447

2 Shafrir, A. L. *et al.* Risk for and consequences of endometriosis: A critical epidemiologic review. *Best Practice & Research Clinical Obstetrics & Gynaecology* **51**, 1-15 (2018). https://doi.org:https://doi.org/10.1016/j.bpobgyn.2018.06.001

3 Armour, M., Lawson, K., Wood, A., Smith, C. A. & Abbott, J. The cost of illness and economic burden of endometriosis and chronic pelvic pain in Australia: A national online survey. *PloS one* **14**, e0223316-e0223316 (2019). https://doi.org:10.1371/journal.pone.0223316

4 Simoens, S. *et al.* The burden of endometriosis: costs and quality of life of women with endometriosis and treated in referral centres. *Human Reproduction* **27**, 1292-1299 (2012). https://doi.org:10.1093/humrep/des073

5 Burney, R. O. & Giudice, L. C. Pathogenesis and pathophysiology of endometriosis. *Fertil Steril* **98**, 511-519 (2012). https://doi.org:10.1016/j.fertnstert.2012.06.029

6 Saunders, P. T. K. & Horne, A. W. Endometriosis: Etiology, pathobiology, and therapeutic prospects. *Cell* **184**, 2807-2824 (2021). https://doi.org:https://doi.org/10.1016/j.cell.2021.04.041

7 Zondervan, K. T., Becker, C. M. & Missmer, S. A. Endometriosis. *New England Journal of Medicine* **382**, 1244-1256 (2020). https://doi.org:10.1056/NEJMra1810764

8 Houshdaran, S. *et al.* Steroid hormones regulate genome-wide epigenetic programming and gene transcription in human endometrial cells with marked aberrancies in endometriosis. *PLOS Genetics* **16**, e1008601 (2020). https://doi.org:10.1371/journal.pgen.1008601

9 Houshdaran, S. *et al.* Aberrant Endometrial DNA Methylome and Associated Gene Expression in Women with Endometriosis. *Biology of Reproduction* **95**, 93 (2016). https://doi.org:10.1095/biolreprod.116.140434

10 Rahmioglu, N. *et al.* The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. *Nature Genetics* (Accepted Oct 2022).

11 Gallagher, C. S. *et al.* Genome-wide association and epidemiological analyses reveal common genetic origins between uterine leiomyomata and endometriosis. *Nature Communications* **10**, 4857 (2019). https://doi.org:10.1038/s41467-019-12536-4

12 Leenen, F. A. D., Muller, C. P. & Turner, J. D. DNA methylation: conducting the orchestra from exposure to phenotype? *Clinical Epigenetics* **8**, 92 (2016). https://doi.org:10.1186/s13148-016-0256-8

13 Wu, Y. *et al.* Aberrant methylation at HOXA10 may be responsible for its aberrant expression in the endometrium of patients with endometriosis. *Am J Obstet Gynecol* **193**, 371-380 (2005). https://doi.org:10.1016/j.ajog.2005.01.034

14 Wu, Y., Strawn, E., Basir, Z., Halverson, G. & Guo, S.-W. Promoter Hypermethylation of Progesterone Receptor Isoform B (PR-B) in Endometriosis. *Epigenetics* **1**, 106-111 (2006). https://doi.org:10.4161/epi.1.2.2766

15 Xue, Q. *et al.* Promoter methylation regulates estrogen receptor 2 in human endometrium and endometriosis. *Biol Reprod* **77**, 681-687 (2007). https://doi.org:10.1095/biolreprod.107.061804

16 Izawa, M. *et al.* An epigenetic disorder may cause aberrant expression of aromatase gene in endometriotic stromal cells. *Fertil Steril* **89**, 1390-1396 (2008). https://doi.org:10.1016/j.fertnstert.2007.03.078

17 Xue, Q. *et al.* Transcriptional activation of steroidogenic factor-1 by hypomethylation of the 5' CpG island in endometriosis. *J Clin Endocrinol Metab* **92**, 3261-3267 (2007). https://doi.org:10.1210/jc.2007-0494

18 Xue, Q. *et al.* Methylation of a novel CpG island of intron 1 is associated with steroidogenic factor 1 expression in endometriotic stromal cells. *Reprod Sci* **21**, 395-400 (2014). https://doi.org:10.1177/1933719113497283

19 Zidan, H. E., Rezk, N. A., Alnemr, A. A. & Abd El Ghany, A. M. COX-2 gene promoter DNA methylation status in eutopic and ectopic endometrium of Egyptian women with endometriosis. *J Reprod Immunol* **112**, 63-67 (2015). https://doi.org:10.1016/j.jri.2015.06.093

20 Wu, Y., Strawn, E., Basir, Z., Halverson, G. & Guo, S.-W. Aberrant expression of deoxyribonucleic acid methyltransferases DNMT1, DNMT3A, and DNMT3B in women with endometriosis. *Fertility and Sterility* **87**, 24-32 (2007). https://doi.org:https://doi.org/10.1016/j.fertnstert.2006.05.077

21 Rahmioglu, N. *et al.* Variability of genome-wide DNA methylation and mRNA expression profiles in reproductive and endocrine disease related tissues. *Epigenetics* **12**, 897-908 (2017). https://doi.org:10.1080/15592294.2017.1367475

22 Borghese, B. *et al.* Research resource: genome-wide profiling of methylated promoters in endometriosis reveals a subtelomeric location of hypermethylation. *Mol Endocrinol* **24**, 1872-1885 (2010). https://doi.org:10.1210/me.2010-0160

23 Dyson, M. T. *et al.* Genome-wide DNA methylation analysis predicts an epigenetic switch for GATA factor expression in endometriosis. *PLoS Genet* **10**, e1004158 (2014). https://doi.org:10.1371/journal.pgen.1004158

24 Mortlock, S. *et al.* Tissue specific regulation of transcription in endometrium and association with disease. *Human Reproduction* **35**, 377-393 (2020). https://doi.org:10.1093/humrep/dez279

25 Mortlock, S. *et al.* Genetic regulation of methylation in human endometrium and blood and gene targets for reproductive diseases. *Clinical Epigenetics* **11**, 49 (2019). https://doi.org:10.1186/s13148-019-0648-7

26 Fassbender, A. *et al.* World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: IV. Tissue collection, processing, and storage in endometriosis research. *Fertil Steril* **102**, 1244-1253 (2014). https://doi.org:10.1016/j.fertnstert.2014.07.1209

27 Vitonis, A. F. *et al.* World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonization Project: II. Clinical and covariate phenotype data collection in endometriosis research. *Fertil Steril* **102**, 1223-1232 (2014). https://doi.org:10.1016/j.fertnstert.2014.07.1244

28 Becker, C. M. *et al.* World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: I. Surgical phenotype data collection in endometriosis research. *Fertil Steril* **102**, 1213-1222 (2014). https://doi.org:10.1016/j.fertnstert.2014.07.709

29 Revised American Society for Reproductive Medicine classification of endometriosis: 1996. *Fertil Steril* **67**, 817-821 (1997). https://doi.org:10.1016/s0015-0282(97)81391-x

30 Tamaresis, J. S. *et al.* Molecular Classification of Endometriosis and Disease Stage Using High-Dimensional Genomic Data. *Endocrinology* **155**, 4986-4999 (2014). https://doi.org:10.1210/en.2014-1490

31 Sapkota, Y. *et al.* Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nature Communications* **8**, 15539 (2017). https://doi.org:10.1038/ncomms15539

32 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). https://doi.org:10.1186/1471-2105-9-559

33 Hannon, E. *et al.* Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the Relationship between Methylomic Variation, Gene Expression, and

Complex Traits. *The American Journal of Human Genetics* **103**, 654-665 (2018). https://doi.org:https://doi.org/10.1016/j.ajhg.2018.09.007

34 Min, J. L. *et al.* Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nature Genetics* **53**, 1311-1321 (2021). https://doi.org:10.1038/s41588-021-00923-x

35 Taylor, D. L. *et al.* Integrative analysis of gene expression, DNA methylation, physiological traits, and genetic variation in human skeletal muscle. *Proceedings of the National Academy of Sciences* **116**, 10883 (2019). https://doi.org:10.1073/pnas.1814263116

36 Volkov, P. *et al.* A Genome-Wide mQTL Analysis in Human Adipose Tissue Identifies Genetic Variants Associated with DNA Methylation, Gene Expression and Metabolic Traits. *PloS one* **11**, e0157776 (2016). https://doi.org:10.1371/journal.pone.0157776

37 Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nature Communications* **9**, 2282 (2018). https://doi.org:10.1038/s41467-018-04558-1

38 Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics* **48**, 481-487 (2016). https://doi.org:10.1038/ng.3538

39 Fung, J. N. *et al.* Genetic regulation of disease risk and endometrial gene expression highlights potential target genes for endometriosis and polycystic ovarian syndrome. *Scientific Reports* **8**, 11424 (2018). https://doi.org:10.1038/s41598-018-29462-y

40 Rahmioglu, N. *et al.* Large-scale genome-wide association meta-analysis of endometriosis reveals 13 novel loci and genetically-associated comorbidity with other pain conditions. *bioRxiv* **DOI:10.1101/406967** (2018).

41 Powell, J. E. *et al.* Endometriosis risk alleles at 1p36.12 act through inverse regulation of CDC42 and LINC00339. *Human Molecular Genetics* **25**, 5046-5058 (2016). https://doi.org:10.1093/hmg/ddw320

42 Lee, S. H. *et al.* Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Human molecular genetics* **22**, 832-841 (2013). https://doi.org:10.1093/hmg/dds491

43 Coosemans, A. *et al.* Wilms' tumor gene 1 (WT1) overexpression in neurons in deep endometriosis: a pilot study. *Fertility and Sterility* **91**, 1441-1444 (2009). https://doi.org:https://doi.org/10.1016/j.fertnstert.2008.06.042

44 Matsuzaki, S. *et al.* Expression of WT1 is down-regulated in eutopic endometrium obtained during the midsecretory phase from patients with endometriosis. *Fertil Steril* **86**, 554-558 (2006). https://doi.org:10.1016/j.fertnstert.2006.02.101

45 Nnoaham, K. E. *et al.* Multi-Centre Studies of the Global Impact of Endometriosis and the Predictive Value of Associated Symptoms. *J Endometr* **1**, 36-45 (2009).

46 Houshdaran, S., Zelenko, Z., Irwin, J. C. & Giudice, L. C. Human Endometrial DNA Methylome Is Cycle-Dependent and Is Associated With Gene Expression Regulation. *Molecular Endocrinology* **28**, 1118-1135 (2014). https://doi.org:10.1210/me.2013-1340

47 Kukushkina, V. *et al.* DNA methylation changes in endometrium and correlation with gene expression during the transition from pre-receptive to receptive phase. *Scientific Reports* **7**, 3916 (2017). https://doi.org:10.1038/s41598-017-03682-0

48 Saare, M. *et al.* The influence of menstrual cycle and endometriosis on endometrial methylome. *Clinical Epigenetics* **8**, 2 (2016). https://doi.org:10.1186/s13148-015-0168-z

49 Díaz-Gimeno, P. *et al.* A genomic diagnostic tool for human endometrial receptivity based on the transcriptomic signature. *Fertility and Sterility* **95**, 50-60.e15 (2011). https://doi.org:https://doi.org/10.1016/j.fertnstert.2010.04.063

50 Wang, W. *et al.* Single-cell transcriptomic atlas of the human endometrium during the menstrual cycle. *Nature Medicine* **26**, 1644-1653 (2020). https://doi.org:10.1038/s41591-020-1040-z

51 McCartney, D. L. *et al.* Epigenetic prediction of complex traits and death. *Genome Biology* **19**, 136 (2018). https://doi.org:10.1186/s13059-018-1514-1

52 Nabais, M. F. *et al.* Meta-analysis of genome-wide DNA methylation identifies shared associations across neurodegenerative disorders. *Genome Biology* **22**, 90 (2021). https://doi.org:10.1186/s13059-021-02275-5

53 Suda, K. *et al.* Clonal Expansion and Diversification of Cancer-Associated Mutations in Endometriosis and Normal Endometrium. *Cell Rep* **24**, 1777-1789 (2018). https://doi.org:10.1016/j.celrep.2018.07.037

54 Montgomery, G. W., Mortlock, S. & Giudice, L. C. Should Genetics Now Be Considered the Pre-eminent Etiologic Factor in Endometriosis? *J Minim Invasive Gynecol* **27**, 280-286 (2020). https://doi.org:10.1016/j.jmig.2019.10.020

55 Meola, J. *et al.* Differentially expressed genes in eutopic and ectopic endometrium of women with endometriosis. *Fertility and Sterility* **93**, 1750-1773 (2010). https://doi.org:10.1016/j.fertnstert.2008.12.058

56 Pazhohan, A. *et al.* The Wnt/β-catenin signaling in endometriosis, the expression of total and active forms of β-catenin, total and inactive forms of glycogen synthase kinase-3β, WNT7a and DICKKOPF-1. *Eur J Obstet Gynecol Reprod Biol* **220**, 1-5 (2018). https://doi.org:10.1016/j.ejogrb.2017.10.025

57 Matsuzaki, S., Botchorishvili, R., Pouly, J. L. & Canis, M. Targeting the Wnt/β-catenin pathway in endometriosis: a potentially effective approach for treatment and prevention. *Mol Cell Ther* **2**, 36-36 (2014). https://doi.org:10.1186/s40591-014-0036-9

58 McKinnon, B. D., Kocbek, V., Nirgianakis, K., Bersinger, N. A. & Mueller, M. D. Kinase signalling pathways in endometriosis: potential targets for non-hormonal therapeutics. *Human reproduction update* **22**, 382-403 (2016). https://doi.org:10.1093/humupd/dmv060

59 Bora, G. & Yaba, A. The role of mitogen-activated protein kinase signaling pathway in endometriosis. *J Obstet Gynaecol Res* **47**, 1610-1623 (2021). https://doi.org:10.1111/jog.14710

60 McRae, A. F. *et al.* Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biology* **15**, R73 (2014). https://doi.org:10.1186/gb-2014-15-5-r73

61 Gaunt, T. R. *et al.* Systematic identification of genetic influences on methylation across the human life course. *Genome Biology* **17**, 61 (2016). https://doi.org:10.1186/s13059-016-0926-z

62 Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature Biotechnology* **31**, 142-147 (2013). https://doi.org:10.1038/nbt.2487

63 Rakyan, V. K. *et al.* Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet* **7**, e1002300 (2011). https://doi.org:10.1371/journal.pgen.1002300

64 Teschendorff, A. E. *et al.* An Epigenetic Signature in Peripheral Blood Predicts Active Ovarian Cancer. *PloS one* **4**, e8274 (2009). https://doi.org:10.1371/journal.pone.0008274

65 Zhang, F. *et al.* OSCA: a tool for omic-data-based complex trait analysis. *Genome Biology* **20**, 107 (2019). https://doi.org:10.1186/s13059-019-1718-z

66 Nabais, M. F. *et al.* Significant out-of-sample classification from methylation profile scoring for amyotrophic lateral sclerosis. *npj Genomic Medicine* **5**, 10 (2020). https://doi.org:10.1038/s41525-020-0118-3

67 Sillem, M., Prifti, S., Neher, M. & Runnebaum, B. Extracellular matrix remodelling in the endometrium and its possible relevance to the pathogenesis of endometriosis. *Human reproduction update* **4**, 730-735 (1998). https://doi.org:10.1093/humupd/4.5.730

68 Kusama, K. *et al.* Regulatory Action of Calcium Ion on Cyclic AMP-Enhanced Expression of Implantation-Related Factors in Human Endometrial Cells. *PloS one* **10**, e0132017 (2015). https://doi.org:10.1371/journal.pone.0132017

69 Sohn, J. O. *et al.* Alterations in intracellular Ca2+ levels in human endometrial stromal cells after decidualization. *Biochemical and Biophysical Research Communications* **515**, 318-324 (2019). https://doi.org:https://doi.org/10.1016/j.bbrc.2019.05.153

70 Singh, H. & Aplin, J. D. Adhesion molecules in endometrial epithelium: tissue integrity and embryo implantation. *Journal of Anatomy* **215**, 3-13 (2009). https://doi.org:https://doi.org/10.1111/j.1469-7580.2008.01034.x

71 Owusu-Akyaw, A., Krishnamoorthy, K., Goldsmith, L. T. & Morelli, S. S. The role of mesenchymal–epithelial transition in endometrial function. *Human reproduction update* **25**, 114-133 (2019). https://doi.org:10.1093/humupd/dmy035

72 Hodgkinson, K. *et al.* GREB1 is an estrogen receptor-regulated tumour promoter that is frequently expressed in ovarian cancer. *Oncogene* **37**, 5873-5886 (2018). https://doi.org:10.1038/s41388-018-0377-y

73 Rae, J. M. *et al.* GREB1 is a critical regulator of hormone dependent breast cancer growth. *Breast Cancer Research and Treatment* **92**, 141-149 (2005). https://doi.org:10.1007/s10549-005-1483-4

74 Garrido-Martín, D., Borsari, B., Calvo, M., Reverter, F. & Guigó, R. Identification and analysis of splicing quantitative trait loci across multiple tissues in the human genome. *Nature Communications* **12**, 727 (2021). https://doi.org:10.1038/s41467-020-20578-2

75 Marla, S. *et al.* Genetic risk factors for endometriosis near estrogen receptor 1 and coexpression of genes in this region in endometrium. *Molecular Human Reproduction* **27** (2021). https://doi.org:10.1093/molehr/gaaa082

76 Mortlock, S. *et al.* A multi-level investigation of the genetic relationship between endometriosis and ovarian cancer histotypes. *medRxiv*, 2021.2006.2028.21259290 (2021). https://doi.org:10.1101/2021.06.28.21259290

77 Lu, Y. *et al.* Shared genetics underlying epidemiological association between endometriosis and ovarian cancer. *Human molecular genetics* **24**, 5955-5964 (2015). https://doi.org:10.1093/hmg/ddv306

78 Kvaskoff, M. *et al.* Endometriosis and cancer: a systematic review and meta-analysis. *Human reproduction update* **27**, 393-420 (2021). https://doi.org:10.1093/humupd/dmaa045

79 Zhang, G. *et al.* Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. *New England Journal of Medicine* **377**, 1156-1167 (2017). https://doi.org:10.1056/NEJMoa1612665

80 Perry, J. R. B. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92-97 (2014). https://doi.org:10.1038/nature13545

81 Day, F. R. *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nature Genetics* **49**, 834-841 (2017). https://doi.org:10.1038/ng.3841

82 The, G. C. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318 (2020). https://doi.org:10.1126/science.aaz1776

83 Hoon, D. S. B., Rahimzadeh, N. & Bustos, M. A. EpiMap: Fine-tuning integrative epigenomics maps to understand complex human regulatory genomic circuitry. *Signal Transduct Target Ther* **6**, 179 (2021). https://doi.org:10.1038/s41392-021-00620-5

84 O'Mara, T. A., Spurdle, A. B., Glubb, D. M. & Endometrial Cancer Association, C. Analysis of Promoter-Associated Chromatin Interactions Reveals Biologically Relevant Candidate Target Genes at Endometrial Cancer Risk Loci. *Cancers* **11** (2019). https://doi.org:10.3390/cancers11101440

85 Bourlev, V. *et al.* The relationship between microvessel density, proliferative activity and expression of vascular endothelial growth factor-A and its receptors in eutopic endometrium and endometriotic lesions. *Reproduction* **132**, 501-509 (2006). https://doi.org:10.1530/rep.1.01110

86 Machado, D. E., Abrao, M. S., Berardo, P. T., Takiya, C. M. & Nasciutti, L. E. Vascular density and distribution of vascular endothelial growth factor (VEGF) and its receptor VEGFR-2 (Flk-1) are significantly higher in patients with deeply infiltrating endometriosis affecting the rectum. *Fertil Steril* **90**, 148-155 (2008). https://doi.org:10.1016/j.fertnstert.2007.05.076

87 Trapero, C. & Martín-Satué, M. Purinergic Signaling in Endometriosis-Associated Pain. *International journal of molecular sciences* **21**, 8512 (2020). https://doi.org:10.3390/ijms21228512

88 Shafrir, A. L. *et al.* Risk for and consequences of endometriosis: A critical epidemiologic review. *Best Pract Res Clin Obstet Gynaecol* **51**, 1-15 (2018). https://doi.org:10.1016/j.bpobgyn.2018.06.001

89 Correia, K. F. *et al.* Confounding and effect measure modification in reproductive medicine research. *Human reproduction (Oxford, England)* **35**, 1013-1018 (2020). https://doi.org:10.1093/humrep/deaa051

90 Farland, L. V. *et al.* The importance of mediation in reproductive health studies. *Human reproduction (Oxford, England)* **35**, 1262-1266 (2020). https://doi.org:10.1093/humrep/deaa064

91 Mansell, G. *et al.* Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genomics* **20**, 366 (2019). https://doi.org:10.1186/s12864-019-5761-7

92 Shafrir, A. L. & Missmer, S. A. Towards subtypes - deep endometriosis oestrogen receptor-α expression. *Nat Rev Endocrinol* **16**, 541-542 (2020). https://doi.org:10.1038/s41574-020-0394-0

93 Missmer, S. A. Why so null? Methodologic necessities to advance endometriosis discovery. *Paediatric and perinatal epidemiology* **33**, 26-27 (2019). https://doi.org:10.1111/ppe.12540

94 Campan, M., Weisenberger, D. J., Trinh, B. & Laird, P. W. MethyLight. *Methods Mol Biol* **507**, 325-337 (2009). https://doi.org:10.1007/978-1-59745-522-0_23

95 Noyes, R. W., Hertig, A. T. & Rock, J. Dating the Endometrial Biopsy. *Fertility and Sterility* **1**, 3-25 (1950). https://doi.org:https://doi.org/10.1016/S0015-0282(16)30062-0

96 Zhou, W., Triche, T. J., Jr., Laird, P. W. & Shen, H. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Research* **46**, e123-e123 (2018). https://doi.org:10.1093/nar/gky691

97 Fortin, J.-P., Triche, T. J., Jr. & Hansen, K. D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**, 558-560 (2017). https://doi.org:10.1093/bioinformatics/btw691

98 Chen, J. *et al.* Fast and robust adjustment of cell mixtures in epigenome-wide association studies with SmartSVA. *BMC Genomics* **18**, 413 (2017). https://doi.org:10.1186/s12864-017-3808-1

99 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015). https://doi.org:10.1093/nar/gkv007

100 Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016). https://doi.org:10.1093/bioinformatics/btw313

101 IlluminaHumanMethylationEPICanno.ilm10b4.hg19: Annotation for Illumina's EPIC methylation arrays. R package version 0.6.0. (2017).

102 Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* **16**, 284-287 (2012). https://doi.org:10.1089/omi.2011.0118

103 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research* **45**, D353-D361 (2017). https://doi.org:10.1093/nar/gkw1092

104 Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res* **48**, D498-d503 (2020). https://doi.org:10.1093/nar/gkz1031

105 Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021). https://doi.org:https://doi.org/10.1016/j.xinn.2021.100141

106 Peters, T. J. *et al.* De novo identification of differentially methylated regions in the human genome. *Epigenetics & Chromatin* **8**, 6 (2015). https://doi.org:10.1186/1756-8935-8-6

107 Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-1358 (2012). https://doi.org:10.1093/bioinformatics/bts163

108 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011). https://doi.org:10.1016/j.ajhg.2010.11.011

109 Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *bioRxiv* **DOI:10.1101/447367** (2018).

110 Kamat, M. A. *et al.* PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* **35**, 4851-4853 (2019). https://doi.org:10.1093/bioinformatics/btz469

111 Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-d1012 (2019). https://doi.org:10.1093/nar/gky1120

112 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics* **88**, 76-82 (2011). https://doi.org:https://doi.org/10.1016/j.ajhg.2010.11.011

Tables

Table 1. DNAm sample information. The number and percentage of samples included in the DNA methylation (DNAm) analysis from each institute, at each cycle phase, from each genetic ancestry and with each disease stage.

Variable	Classification	Differential DNAm Analysis				mQTL Analysis			
		Control		Endometriosis		Control		Endometriosis	
		n	%	n	%	n	%	n	%
Institute	University of Edinburgh	31	9%	52	8%	29	13%	50	11%
	University of Melbourne	87	25%	225	35%	64	29%	163	37%
	University of Oxford	52	15%	126	20%	43	19%	103	24%
	University of California San Francisco	177	51%	234	37%	85	39%	121	28%
Cycle Phase	Proliferative	170	49%	303	48%	98	44%	198	45%
	Secretory (undefined sub-phase)	7	2%	15	2%	3	1%	9	2%
	Early Secretory	44	13%	78	12%	35	16%	57	13%
	Mid Secretory	78	22%	131	21%	52	24%	96	22%
	Late Secretory	35	10%	73	11%	22	10%	50	11%
	Menstrual	13	4%	37	6%	11	5%	27	6%
Genetic Ancestry	Admixed	27	8%	55	9%	n/a	n/a	n/a	n/a
	African	35	10%	13	2%	n/a	n/a	n/a	n/a
	Admixed American	22	6%	30	5%	n/a	n/a	n/a	n/a
	East Asian	28	8%	50	8%	n/a	n/a	n/a	n/a
	European	226	65%	451	71%	221	100%	437	100%
	South Asian	9	2%	38	6%	n/a	n/a	n/a	n/a
Disease Stage	I-II	n/a	n/a	344	54%	n/a	n/a	261	60%
	III-IV	n/a	n/a	286	45%	n/a	n/a	171	39%
	Unknown	n/a	n/a	7	1%	n/a	n/a	5	1%
Additional Characteristics	Age (Mean ± SD)	37.0 ± 8.4 (n=342)		34.2 ± 7.1 (n=634)		36.1 ± 8.6 (n=220)		33.7 ± 7.1 (n=436)	
	BMI (Mean ± SD)	26.6 ± 6.1 (n=301)		24.8 ± 5.0 (n=591)		26.4 ± 6.1 (n=196)		24.9 ± 5.1 (n=414)	
	Parity (Mean ± SD)	1.0 ± 1.35 (n=325)		0.6 ± 1.0 (n=613)		0.9 ± 1.4 (n=206)		0.5 ± 1.0 (n=422)	

	Age at Menarche (Mean ± SD)	12.8 ± 1.6 (n=146)	12.7 ± 1.7 (n=486)	12.9 ± 1.6 (n=105)	12.8 ± 1.6 (n=339)

Table 2. DNAm sites significantly associated with menstrual cycle phase. The number and direction of effect of DNA methylation (DNAm) sites significantly differentially methylated between menstrual cycle phases across all 984 samples regardless of endometriosis case:control status. Cycle phases include early secretory (ESE), mid secretory (MSE), late secretory (LSE), proliferative (PE), aggregated secretory phases (SE) and menstrual. The direction of effect between the cycle phases are annotated as "Down" if the methylation at the DNAm site is lower in the first phase and "Up" if the methylation is higher.

Comparison	Direction	Number of DNAm sites (p-adj < 0.05)	DMRs (FDR < 0.1)
ESE_LSE	Down	56	85
	Up	292	
ESE_MSE	Down	24	86
	Up	291	
ESE_PE	Down	1579	997
	Up	1821	
LSE_MSE	Down	1	0
	Up	4	
LSE_PE	Down	1878	897
	Up	1495	
MSE_PE	Down	4515	2371
	Up	3293	
MSE_Menstrual	Down	108	47
	Up	118	
ESE_Menstrual	Down	19	14
	Up	38	
LSE_Menstrual	Down	23	10
	Up	50	
PE_Menstrual	Down	23	5
	Up	42	

PE_SE	Down	5157	3005
	Up	4497	
SE_Menstrual	Down	69	30
	Up	94	

Table 3. Significantly enriched pathways. The top 20 pathways significantly enriched for genes annotated to DNA methylation (DNAm) differentially methylated between the proliferative and secretory phase. N is the total number of genes in the pathway and DM is the number of genes from the pathway that were annotated to the differentially methylated DNAm sites.

Pathway ID	Description	N	DM	P-value	FDR
R-HSA-9012999	RHO GTPase cycle	445	152	2.62E-10	3.94E-07
hsa04510	Focal adhesion	201	80	6.18E-10	1.19E-07
R-HSA-1474244	Extracellular matrix organization	299	110	7.55E-10	5.67E-07
hsa04015	Rap1 signaling pathway	210	82	1.12E-09	1.19E-07
hsa04810	Regulation of actin cytoskeleton	217	84	1.16E-09	1.19E-07
hsa04520	Adherens junction	71	38	1.43E-09	1.19E-07
R-HSA-9013148	CDC42 GTPase cycle	157	67	1.80E-09	9.02E-07
hsa04360	Axon guidance	182	73	2.36E-09	1.58E-07
hsa05165	Human papillomavirus infection	326	111	1.60E-08	8.93E-07
hsa04010	MAPK signaling pathway	294	102	2.10E-08	1.00E-06
hsa04151	PI3K-Akt signaling pathway	344	113	1.08E-07	4.28E-06
hsa05205	Proteoglycans in cancer	204	75	1.15E-07	4.28E-06
R-HSA-9013106	RHOC GTPase cycle	73	36	1.37E-07	5.15E-05
R-HSA-9013149	RAC1 GTPase cycle	184	70	2.14E-07	6.42E-05
hsa04750	Inflammatory mediator regulation of TRP channels	98	43	2.50E-07	8.34E-06
R-HSA-8980692	RHOA GTPase cycle	146	58	4.35E-07	1.09E-04
hsa04072	Phospholipase D signaling pathway	147	57	5.19E-07	1.49E-05
hsa04310	Wnt signaling pathway	168	63	5.35E-07	1.49E-05
hsa04611	Platelet activation	124	50	6.54E-07	1.68E-05
hsa04390	Hippo signaling pathway	157	59	1.13E-06	2.47E-05

Table 4. Phenotypic variance captured by endometrial DNAm. Proportion of variance in endometriosis case-control status captured by common genetic variants and genome-wide DNA methylation (DNAm) in endometrium estimated using different GREML and OREML models. GRM is genetic relationship matrix and ORM is omic-relationship matrix.

Model	Variance Captured (s.e.)		Phenotypic Variance (s.e.)
GRM	26.2%* (0.15)		0.225 (0.01)
ORM (all sites n=759,345)	24.2% (0.07)		0.255 (0.02)
ORM (all sites n=759,345) + GRM	16.1% (0.07) 20.9% (0.14)	37% (0.15)	0.242 (0.02)
ORM (non-mQTL sites n=651,718) + GRM	15.4% (0.07) 21.5% (0.14)	36.9% (0.15)	0.242 (0.02)
*liability scale			

Figures

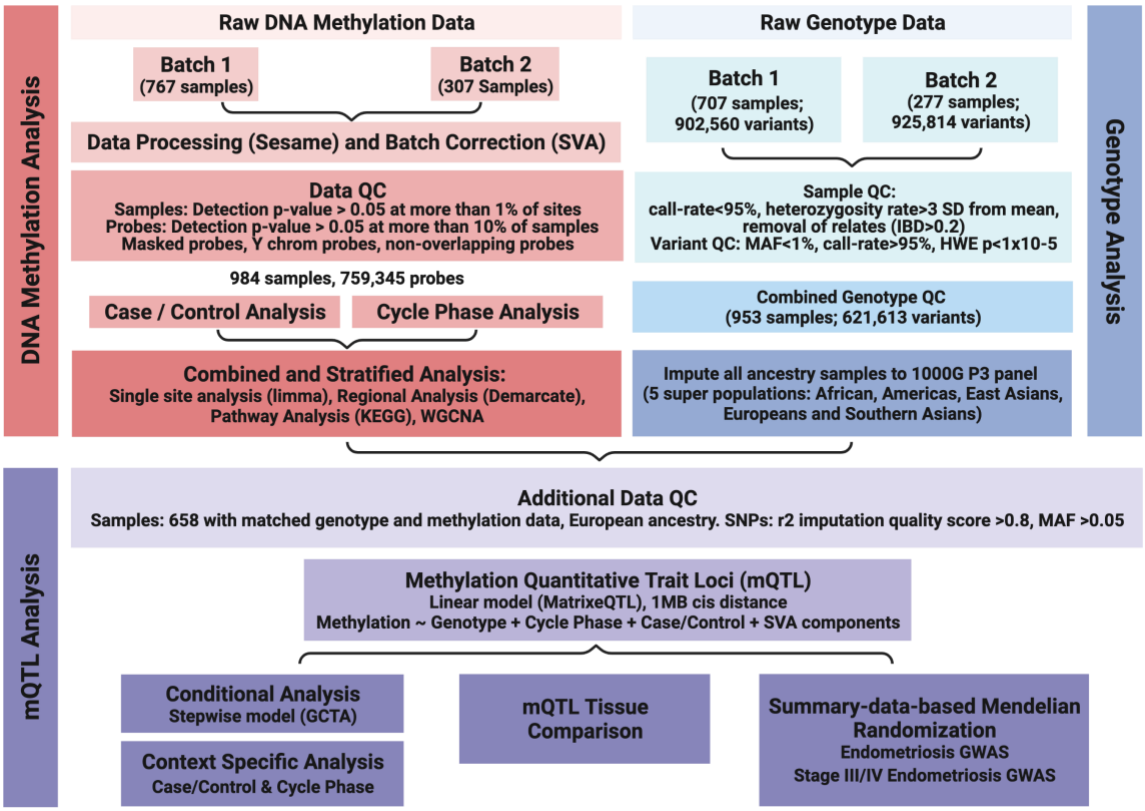

Figure 1. Data QC and study overview. Flow diagram of the quality control (QC) and analysis steps conducted for the DNA methylation (DNAm) data (red), genotype data (blue) and integration of both DNAm and genetic data (purple).

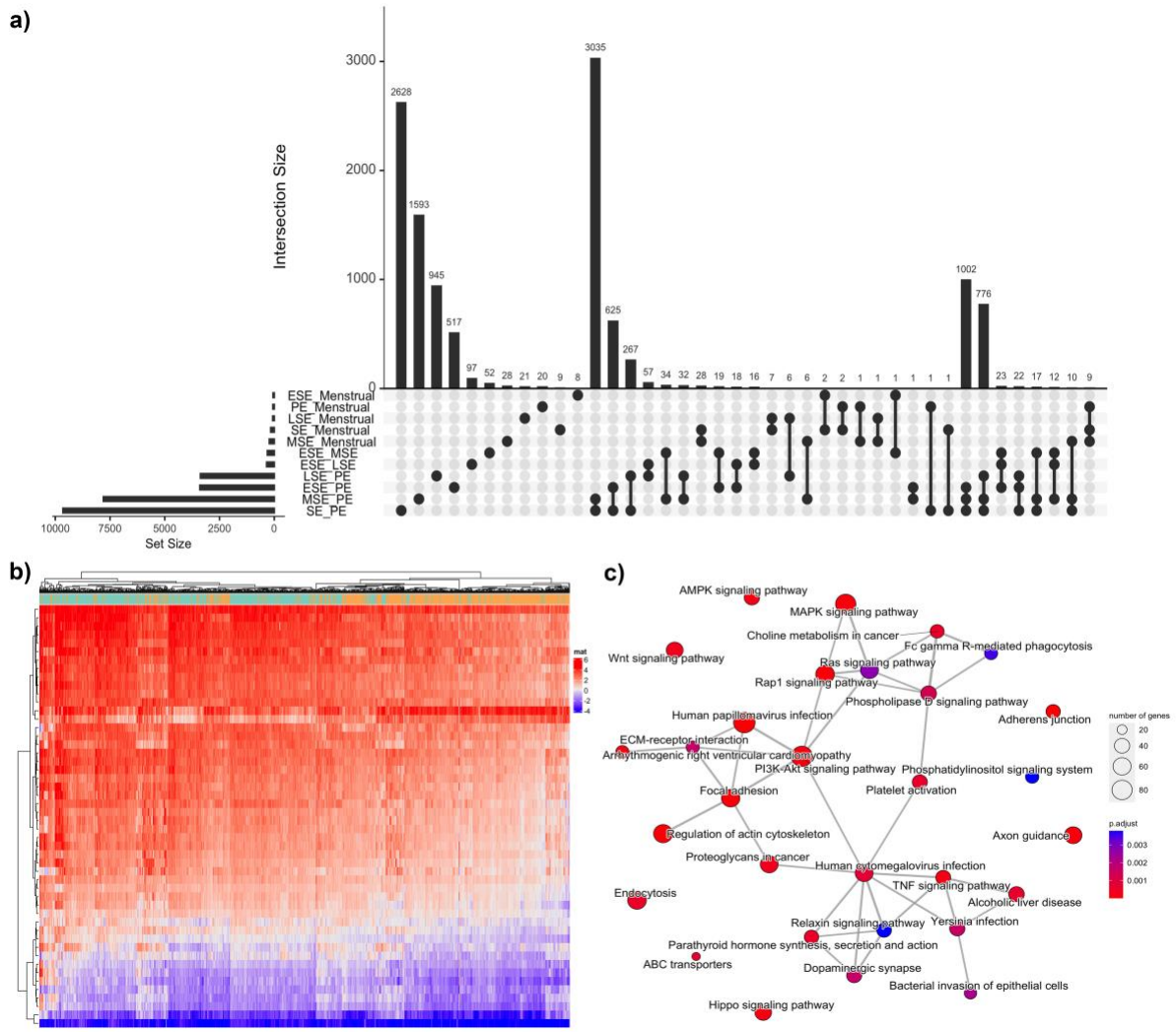

Figure 2. Cycle Phase Differences. a) Upset plot showing the number of significantly differently methylated sites between each comparison on the y-axis bar plot and the number of intersecting DNA methylation (DNAm) sites on the x-axis bar plot. The comparisons included in the intersection are depicted by dots and lines adjoining those dots. b) Heatmap of the top 50 most differentially methylated sites between the secretory (orange) and proliferative phase (green) of the menstrual cycle as annotated on the bar above the heatmap. Red denotes positive methylation M-values and blue denotes negative methylation M-values. c) Pathways significantly enriched for genes annotated to differentially methylated sites between the secretory and proliferative phase.

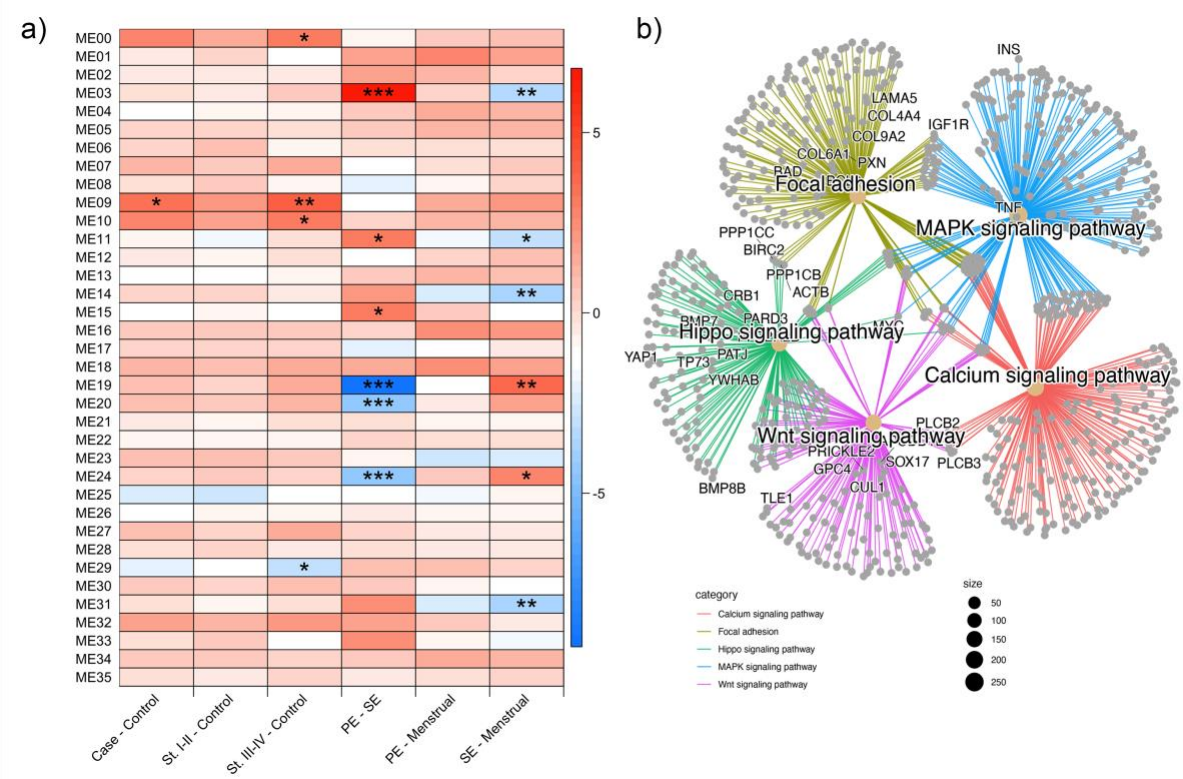

Figure 3. Network analysis. a) Modules of DNA methylation (DNAm) sites defined using WGCNA and their association with endometriosis status and menstrual cycle stage. * represents the degree of significance (*** p-value=0-0.001; ** p-value=0.001-0.01; * p-value=0.01-0.05). b) Network map of the top 5 most significantly enriched pathways for genes annotated to modules associated with endometriosis stage III/IV case:control.

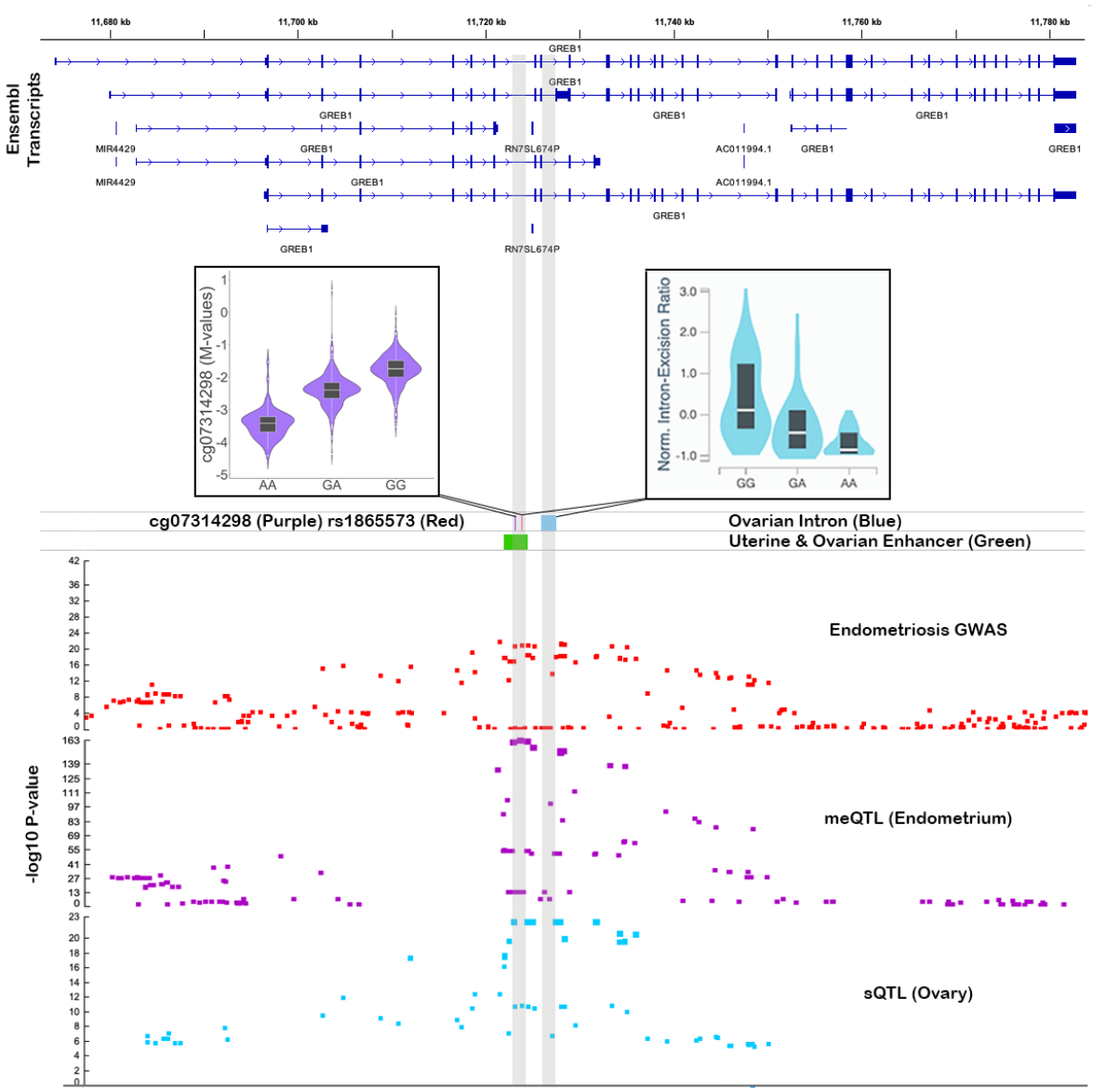

Figure 4. *GREB1* mQTL associated with endometriosis risk. The top panel shows ensemble transcripts present in the locus. The bottom panel consists of association plots, each point is a SNP plotted according to its genomic position on the x-axis and -log10 p-value for its association with endometriosis (red) DNA methylation (DNAm) at cg07314298 (purple) and *GREB1* splicing in ovary (blue) on the y-axis. The position of the associated spliced intron (blue) is featured in the middle panel alongside the position of the significant SMR mQTL SNP (red) and DNAm site (purple) and the position of a predicted enhancer in uterus and ovary

(green). Boxplots show the difference in DNAm and intron excision according to the genotype at rs1865573.

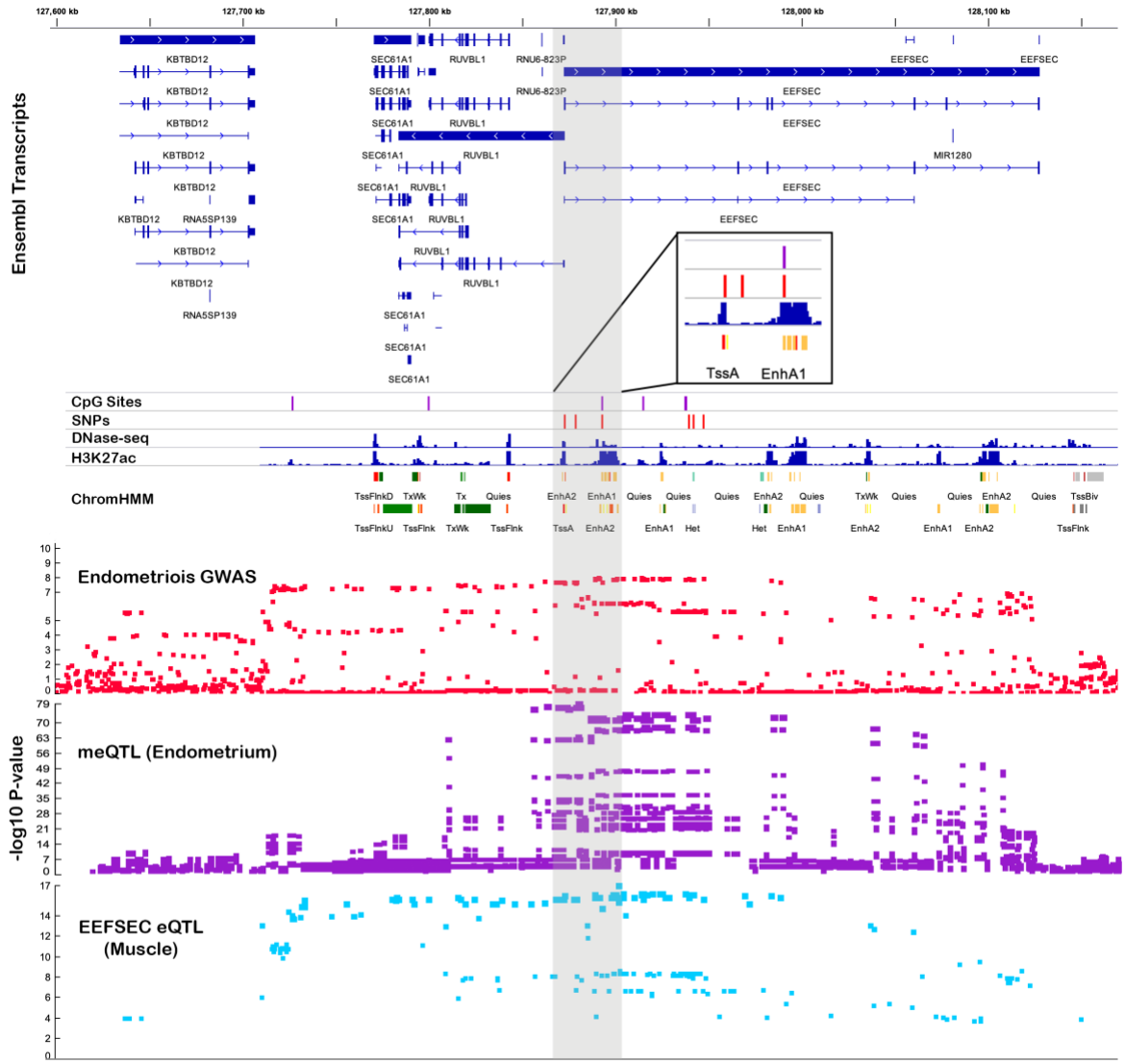

Figure 5. *EEFSEC* mQTL associated with endometriosis risk. The top panel shows ensemble transcripts present in the locus. The bottom panel consists of association plots, each point is a SNP plotted according to its genomic position on the x-axis and -log10 p-value for its association with endometriosis (red) DNA methylation (DNAm) at six SMR significant DNAm sites (purple) and *EEFSEC* expression in muscle (blue) on the y-axis. The position of

the significant SMR mQTL SNPs (red) and DNAm sites (purple) is featured in the middle panel above DNase-seq peaks, H3K27ac peaks and predicted chromatin marks in uterus.

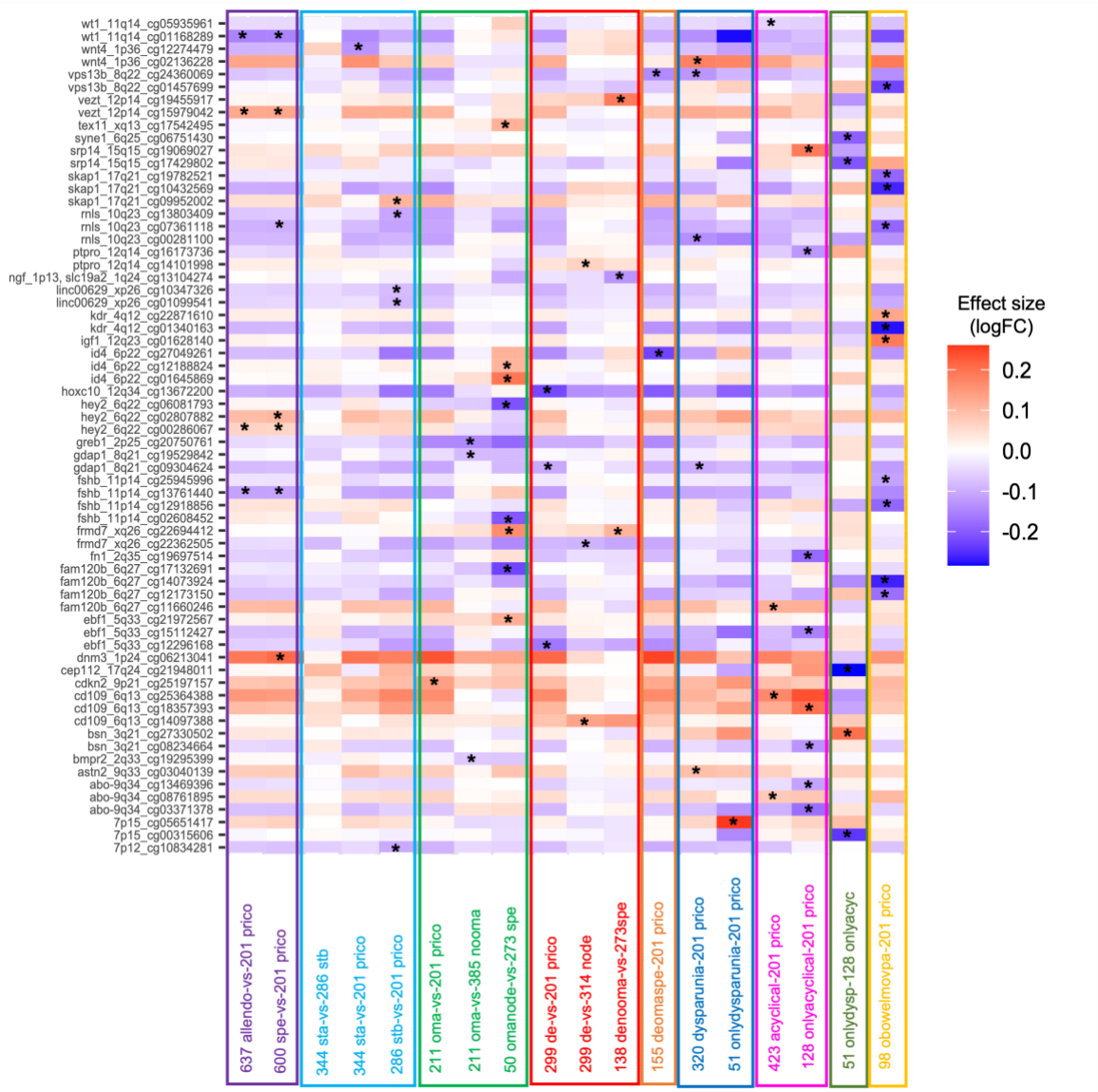

Figure 6. Heatmap for effect sizes of 66 differentially methylated sites across 17 sub-phenotype comparisons. DNA methylation (DNAm) sites are presented as rows and differential DNAm analysis for each phenotype is presented as columns. * denotes statistically significant DNAm sites passing the locus specific Bonferroni-based multiple-testing correction ($p<0.05/N$ of DNAm sites per GWAS locus). Abbreviations: allendo: all endometriosis cases (n=637), prico: NUPP controls (n=201), spe: superficial lesions (n=600), sta: rASRM stage I/II

disease (n=344), stb: rASRM stage III/IV disease (n=286), oma: endometrioma (n=211), omanode: cases with endometriomas but no deep lesions (n=50), de: deep lesions (n=299), denooma: cases with deep lesions but no endometriomas (n=138), deomaspe: cases with co-occurrence of superficial lesions, endometriomas and deep lesions (n=155), dyspareunia: cases with dyspareunia (n=320), onlydysparunia: cases with dyspareunia but no acyclic or dychezia (n=51), acyclic: cases with acyclic pelvic pain (n=423), onlyacyclical: cases with acyclic pelvic pain but no dyspareunia or dyschezia (n=128), obowelmovpa: cases with only dyschezia (n=98).