

RESEARCH PAPER



Variability of genome-wide DNA methylation and mRNA expression profiles in reproductive and endocrine disease related tissues

Nilufer Rahmioglu^{a,†}, Alexander W Drong^{a,†}, Helen Lockstone^a, Thomas Tapmeier^b, Karin Hellner^b, Merli Saare^c, Triin Laisk-Podar^c, Christine Dew^b, Emily Tough^b, George Nicholson^a, Maire Peters^c, Andrew P Morris^{a,d}, Cecilia M Lindgren^a, Christian M Becker^{b,‡}, and Krina T Zondervan^{a,b,‡}

^aWellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK; ^bEndometriosis CaRe Centre, Nuffield Department of Obstetrics & Gynaecology, John Radcliffe Hospital, University of Oxford, Oxford, OX3 7BN, UK; ^cCompetence Centre on Health Technologies, Tartu, Estonia and Women's Clinic, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia; ^dDepartment of Biostatistics, University of Liverpool, Liverpool, OX3 7BN, UK

ABSTRACT

Genome-wide association studies in the fields of reproductive medicine and endocrinology are yielding robust genetic variants associated with disease. Integrated genomic, transcriptomic, and epigenomic molecular profiling studies are common methodologies used to understand the biologic pathways perturbed by these variants. However, molecular profiling resources do not include the tissue most relevant to many female reproductive traits, the endometrium, while the parameters influencing variability of results from its molecular profiling are unclear. We investigated the sources of DNA methylation and RNA expression profile variability in endometrium ($n = 135$), endometriotic disease tissue (endometriosis), and subcutaneous abdominal fat samples from 24 women, quantifying between-individual, within-tissue (cellular heterogeneity), and technical variation. DNA samples ($n = 96$) were analyzed using Illumina HumanMethylation450 BeadChip arrays; RNA samples ($n = 39$) were analyzed using H12-expression arrays. Variance-component analyses showed that, for the top 10–50% variable DNA methylation/RNA expression sites, between-individual variation far exceeded within-tissue and technical variation. Menstrual-phase accounted for most variability in methylation/expression patterns in endometrium ($P_m = 7.8 \times 10^{-3}$, $P_e = 8.4 \times 10^{-5}$) but not in fat and endometriotic tissue; age was significantly associated with DNA methylation profile of endometrium ($P_m = 9 \times 10^{-5}$) and endometriotic disease tissue ($P_m = 2.4 \times 10^{-5}$); and smoking was significantly associated with DNA methylation in adipose tissue ($P_m = 1.8 \times 10^{-3}$). Hierarchical cluster analysis showed significantly different methylation signatures between endometrium and endometriotic tissue enriched for *WNT* signaling, angiogenesis, cadherin signaling, and gonadotropin-releasing-hormone-receptor pathways. Differential DNA methylation/expression analyses suggested detection of a limited number of sites with large fold changes ($FC > 4$), but power calculations accounting for different sources of variability showed that for robust detection >500 tissue samples are required. These results enable appropriate study design for large-scale expression and methylation tissue-based profiling relevant to many reproductive and endocrine traits.

ARTICLE HISTORY

Received 13 June 2017
Revised 1 August 2017
Accepted 9 August 2017

KEYWORDS

DNA methylation; endometrium; endocrine; endometriosis; gene expression; reproductive

Introduction

Genome-wide association studies (GWAS) are uncovering genetic variants involved in the pathogenesis of female reproductive disorders and traits.^{1,2} For instance, for endometriosis—a complex disease defined by the presence of endometriotic tissue outside the uterus associated with pelvic pain and sub-fertility—³ recent large-scale GWAS have identified 15 genetic loci robustly associated with the disease.^{4–10} Subsequent analyses demonstrating effects of endometriosis-associated variants on fat distribution (waist-to-hip ratio adjusted for body mass index, WHRadjBMI) indicate that many of these variants are likely also to impact the regulation of endocrine and metabolic processes.^{11,12} Typically, GWAS variants are located

in inter-genic regions, and understanding their functional role requires molecular profiling (e.g., gene-expression assays) of tissues relevant to the conditions.^{13,14} Such integrated analysis of genomic, transcriptomic, and epigenomic data form the premise of large-scale tissue-based molecular profiling initiatives, such as GTEx,^{15,16} and the NIH Epigenome Roadmap.¹⁷ Endometrium is not included in these genomic profiling initiatives, an omission that prevents translation of genetic findings into results that could inform biomarker and drug target discovery programs for many common endometrium-related disorders, including endometriosis, infertility and recurrent miscarriage, abnormal bleeding, and endometrial cancer.

CONTACT Krina T Zondervan  krina.zondervan@obs-gyn.ox.ac.uk,  Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK; Nilufer Rahmioglu  nilufer.rahmioglu@well.ox.ac.uk,  Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK

[†]These authors contributed equally to this work.

[‡]These authors jointly directed the work.

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Endometrium is a heterogeneous and dynamic tissue composed mainly of glandular and luminal epithelial and stromal cells that are under strong cyclic endocrine influences,¹⁸ posing important questions regarding the possible variability of molecular profiling data generated. To date, 18 studies have investigated genome-wide mRNA expression in endometrium (11 in women with endometriosis), including samples from a range of 4 to 75 cases;^{19–36} only one recent study of 123 endometrium samples from healthy women assessed genome-wide expression quantitative trait loci (eQTL).²⁵ There have been 5 genome-wide DNA methylation analyses (4 in women with endometriosis), including samples from 7 to 31 women.^{37–41} There is little overlap between differentially expressed transcripts/methylated genes identified in these studies, possibly due to: (1) variability in expression/DNA methylation patterns due to cellular heterogeneity inherent to whole-tissue analysis; (2) confounding variables unaccounted for that affect endometrial expression/DNA methylation profiles such as menstrual phase; (3) insufficient sample sizes that result in spurious or chance findings; (4) variation in tissue collection procedures and case/control definitions.

We aimed to investigate individual, tissue-derived and experimental variability of genome-wide DNA methylation and mRNA gene-expression profiles, and their implications for study design, contrasting 3 tissues: eutopic endometrium, endometriotic disease tissue (peritoneal disease, peritoneal lesions

and ovarian disease, endometriomas) and subcutaneous abdominal fat. In addition, we determined sample sizes required to detect robust and biologically meaningful results from genome-wide whole-tissue expression and DNA methylation profiles for these 3 tissues, relevant to a broad range of reproductive, endocrine, and metabolic traits.

Results

DNA and RNA Quality

The experimental design included splits and replicates of endometrium, subcutaneous abdominal fat tissues collected from 8 endometriosis cases, and 8 controls and endometriotic disease tissue from 16 cases, which were used for DNA extraction; the same endometrium tissue samples were also used for RNA extraction (Fig. 1) (See Materials and Methods). All tissues provided good yields of DNA and RNA, with fat tissue providing the lowest yields [DNA mean = 533 ng, Standard Deviation (SD) = 344] and endometrium the highest (DNA mean = 4245 ng, SD = 1942; RNA mean = 4516 ng, SD = 3100) (Fig. S1). DNA quality was modestly associated with the initial yield ($P = 0.03$) and menstrual cycle day ($P = 0.05$), but not with tissue type ($P = 0.38$) or tissue weight ($P = 0.99$). A total of 94/96 (98%) of DNA samples passed quality control

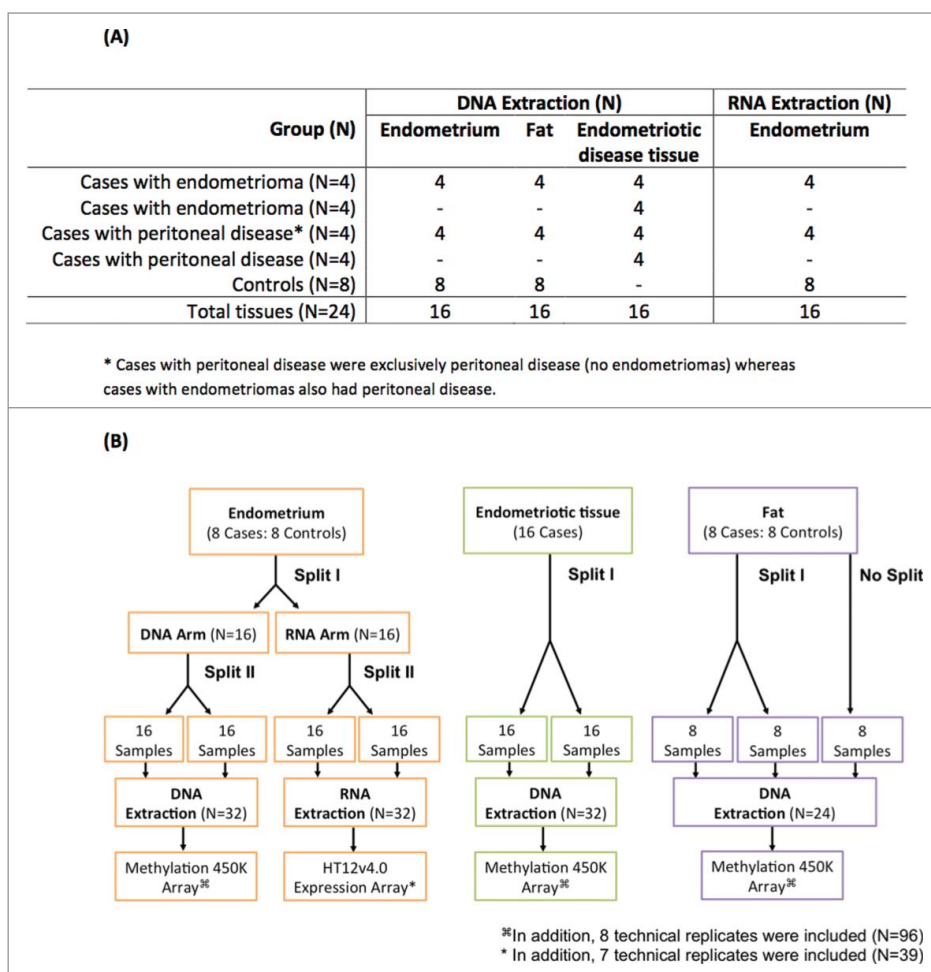


Figure 1. Experimental study design. Endometrium and subcutaneous abdominal fat from 8 cases and 8 controls, and endometriotic disease tissue from 16 cases (8 of which also contributed fat and endometrium) were used for DNA methylation analyses; the same endometrium tissue samples ($n = 8$) were also used for mRNA analyses. (A) Number of tissue samples from cases and controls processed for DNA and RNA extraction. (B) Tissue processing steps illustrating sample splits and technical replicates.

(>1000 ng DNA, >98% call rate), with probe intensities uniformly distributed for all 94 samples (Fig. S2).

RNA RIN-scores from endometrium did not correlate with initial tissue weight ($P = 0.47$) but did with RNA yield ($P = 2.0 \times 10^{-3}$), and cycle day on which the sample was collected ($P = 1.2 \times 10^{-3}$). Fourteen samples had RIN scores <7, indicating reduced quality; 10 of these (71%) were collected during the menstrual phase. These samples did not perform well in the microarray experiment (housekeeping gene signals <10,000, Fig. S3) and were excluded. The remaining 25 samples were of good quality (RIN > 7) with uniform distribution of probe intensities (Fig. S4).

Variability in DNA methylation profiles

Variability in DNA methylation was investigated between-tissues using hierarchical cluster analysis and within-tissue variability was investigated using the DNA methylation profiles from each split and technical replicate of the tissues, through principal component analysis (PCA) for each tissue type separately (see Materials and Methods).

Between-tissue variability

Genome-wide DNA methylation profiles for 94 DNA samples passing quality control (QC; see Materials and Methods) from endometrium and subcutaneous abdominal fat (cases and controls) and endometriotic disease tissue (cases) were analyzed. Clear clustering of DNA methylation profiles by tissue type was

observed (Fig. 2A). Peritoneal disease tissue and endometrioma samples formed 2 distinct clusters.

Within-tissue variability

We investigated within-tissue (cellular heterogeneity) variability, using DNA methylation profiles of same-sample splits, and contrasted replicates of splits to detect contribution of technical (experimental) variation (see Materials and Methods). The PCA of DNA methylation profiles for each tissue showed close clustering of splits and replicates of the same sample, demonstrating that between-individual variation was greater than within-tissue (cellular heterogeneity variation and technical variation). Within endometriotic disease tissue samples, peritoneal disease and endometrioma samples again clustered separately—mainly on principal component 1 (PC1) explaining 35% of the DNA methylation variance (Fig. 2B.iii)—suggesting distinctive DNA methylation signatures (see below).

Effect of covariates

We investigated the effect of specific covariates on DNA methylation profiles averaged across splits and replicates by testing for their association with the first 2 principal components that capture most of the variation in DNA methylation profiles (Table S1): age (mean = 34.8 years, range = 25–47), menstrual phase (7 proliferative, 9 secretory, 6 menstrual phase),³³ smoking (33.3% current smokers), BMI (mean = 25.4, range = 20.2–34), WHR (mean = 0.80, range = 0.72–0.89), and weight of the original tissue sample (mean = 0.39 g, range = 0.09–0.65).

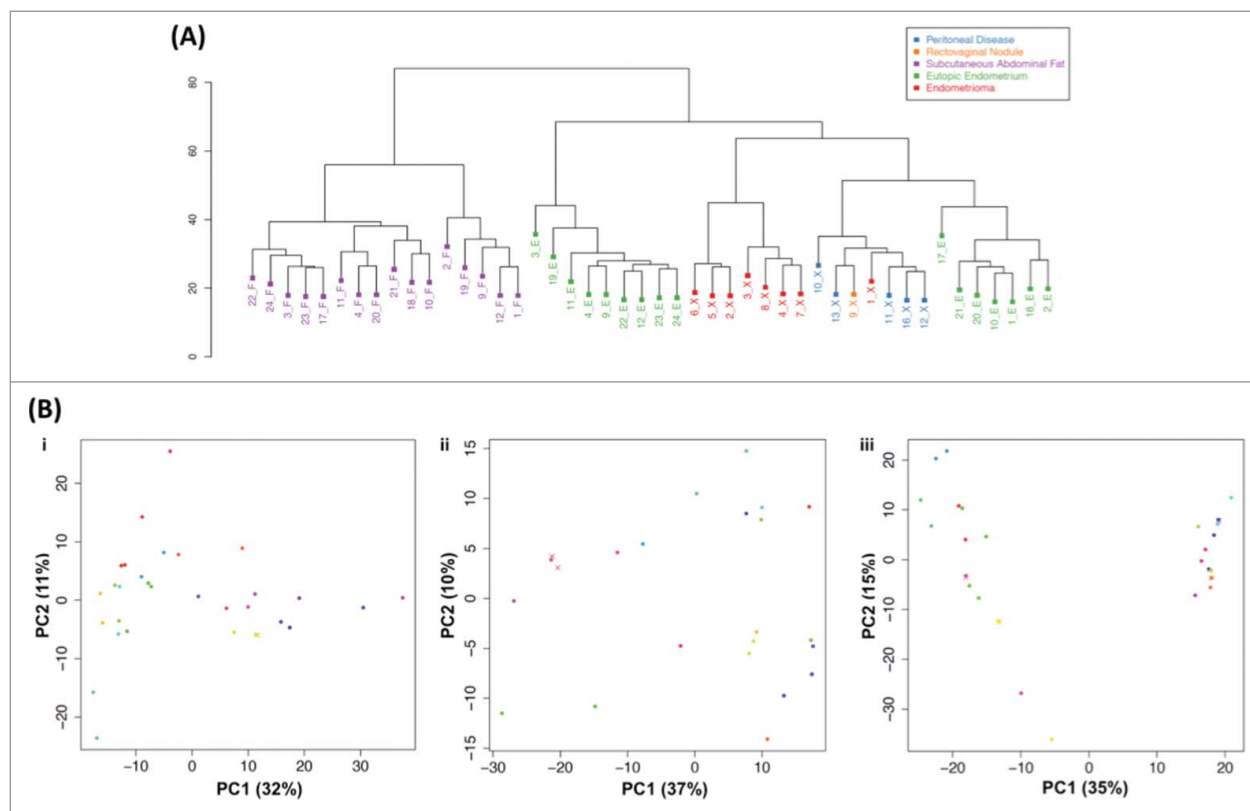


Figure 2. Variability in DNA methylation profiles from all 3 tissues. (A) Hierarchical clustering of DNA methylation profiles based on average methylation levels across splits and replicates for each sample (see Materials and Methods). Colors correspond to tissue type. (B) Principal component analysis (PCA) of genome-wide DNA methylation profiles derived from: i. endometrium; ii. subcutaneous abdominal fat; iii. endometriotic disease tissue. Each circle represents a sample; color coding designates different individuals. Same color circles designate sample splits; crosses (x) indicate technical replicates.

Menstrual phase was significantly associated with DNA methylation profiles in endometrium ($P = 7.8 \times 10^{-3}$), and in peritoneal lesions ($P = 1.3 \times 10^{-6}$) but not with endometriomas or fat (Table S1; Fig. S5). Age was also significantly associated with DNA methylation profiles in endometrium ($P = 9 \times 10^{-5}$), as well as in endometriotic disease tissue ($P = 2.4 \times 10^{-5}$). Smoking was significantly associated with DNA methylation in fat ($P = 2 \times 10^{-3}$). Notably, case/control status was only significantly associated with variability of global DNA methylation profiles in fat ($P = 4 \times 10^{-3}$).

Exploratory differential methylation analysis

We investigated if specific differentially DNA methylated sites of large effect were driving the tissue-based clustering for endometrium and endometriotic disease tissue observed in Fig. 2. We first compared DNA methylation profiles for all 14 endometriotic disease tissue and 16 endometrium samples from cases, adjusting for menstrual phase and age (see Materials and Methods), revealing 27,493 significantly differentially methylated probes, after Bonferroni multiple-testing correction, corresponding to 8,133 independent genes. The differentially methylated probes were concentrated within gene bodies (41%), inter-genic regions (31%), and regulatory regions (28%); in the context of CpG islands, they were mostly in open sea (58%) followed by shores (21%), CpG islands (10%), and shelves (11%) (Fig. S6). When we restricted the analysis to 8 paired eutopic and endometriotic disease tissue samples (linear model) from the same cases, after multiple-testing correction, 3,915 significantly differentially methylated probes in 1,860 independent genes were identified. Of the 3,915 significant probes, 3,858 were also significantly differentially methylated in the unpaired analysis. The top 30 differentially methylated probes are given in Table S4.

Pathway analysis using PANTHER⁴² on the 8,133 genes from the unpaired results showed significant enrichment in: (1) WNT signaling pathway (Bonferroni P adj = 7.49×10^{-8}); (2) Angiogenesis (P adj = 4.12×10^{-4}); (3) Alzheimer disease presenilin pathway (P adj = 4.79×10^{-4}); (4) Cadherin signaling (P adj = 9.29×10^{-4}); (5) Integrin signaling (P adj = 2.09×10^{-3}); (6) Gonadotropin-releasing-hormone-receptor pathway (P adj = 3.65×10^{-3}); (7) CCKR signaling (P adj = 6.22×10^{-3}) (Table S5). PANTHER analysis on the paired results confirmed significant enrichment in WNT signaling (P adj = 2.14×10^{-3}), Gonadotropin-releasing-hormone-receptor pathway (P adj = 1.75×10^{-3}) and CCKR signaling (P adj = 4.82×10^{-2}) (Table S6). We also tested the 27,493 differentially methylated probes for enrichment in histone modification ChIP peaks (H3K4me1, H3K4me3, and H3K27Ac, marks of open chromatin) and DNase I hypersensitivity sites in 127 different cell types in the Roadmap release 9 data sets.¹⁷ The probes were significantly enriched for H3K4me1 histone modification sites in 97 out of 127 cell types, but no other significant enrichment was observed for the other sites (Fig. S7).

Differential DNA methylation analysis of peritoneal lesions vs. endometriomas, adjusting for age (see Materials and Methods), revealed 31 differentially methylated probes significant after multiple-testing correction, 27 hypomethylated and 4 hypermethylated. Most (19/31) were located within gene bodies or regulatory regions of *MNI*, *SLCIA7*, *MYH7*, *DACT2*,

DHRS7, *EPS8L1*, *ADHFE1*, *NFYB*, *FOXD2*, *CACNA1H*, *NCOR2*, *ZBTB7A*, *LAMB3*, *AGRN*, *UBE2CBP*, *BCAR1*, *CMIP* (Table S7).

Unsurprisingly, given the sample size, differential DNA methylation analysis of endometrium between cases and controls did not show any differentially methylated probes after multiple-testing correction.

Quantification of DNA methylation variability using variance component analysis

For each tissue type, the proportion of different sources of variances contributing to total phenotypic variance was estimated using variance component analysis (See Materials and Methods). When considering all CpG methylation probes in endometrium (Fig. S8a), irrespective of their variability, the distribution of technical variance (between replicates) was the main source of variation; similar results were observed for fat and endometriotic disease tissue (Figs. S9 and S10). However, these results included many probes with no or very low variability. When we considered the 50% and 10% most variable probes, a reverse pattern was observed, with between-individual variance explaining a high percentage of the overall variance, and a low within-tissue (split) variance, suggesting a low tissue-based heterogeneity of DNA methylation profiles (Fig. S8c). Mean variance estimates for the top 10% vs. 50% most variable probes were 0.84 (SD = 0.18) vs. 0.65 (SD = 0.29) for between-individual variance; 0.18 (SD = 0.14) vs. 0.23 (SD = 0.22) for within-tissue (tissue heterogeneity); and 0.06 (SD = 0.09) vs. 0.19 (SD = 0.23) for technical variation, respectively. Patterns for endometriotic disease tissue and fat were similar (Table S8).

Variability in RNA expression profiles of endometrium

We similarly investigated tissue heterogeneity and technical variation in RNA expression profiles from 25 endometrium samples passing QC (12 individuals: 5 cases and 7 controls). Close clustering of splits and replicates of samples from the same individual was clearly observed (Fig. 3A), illustrating that between-individual variation exceeded within-tissue and technical variation.

Effects of potential covariates

Fig. 3 shows clustering of samples by the first 2 principal components, PC1 (21.4% of variation) and PC2 (15.8% of variation), color coded by covariate status. Menstrual phase clustered mainly on PC1 ($P = 8.4 \times 10^{-5}$) (Fig. 3B; Table S1), with samples from the secretory (cycle day 15+) and proliferative phases (cycle day 8–14) forming 2 distinct clusters. Clustering was also observed on the basis of case/control status on PC2 ($P = 7.4 \times 10^{-5}$) (Fig. 3C; Table S1). No other covariates were associated with RNA expression profiles (Table S1).

Differential expression analysis between endometriosis cases and controls

We performed differential expression analysis between cases and controls accounting for menstrual phase (see Materials and Methods), to identify the top contributing genes. After

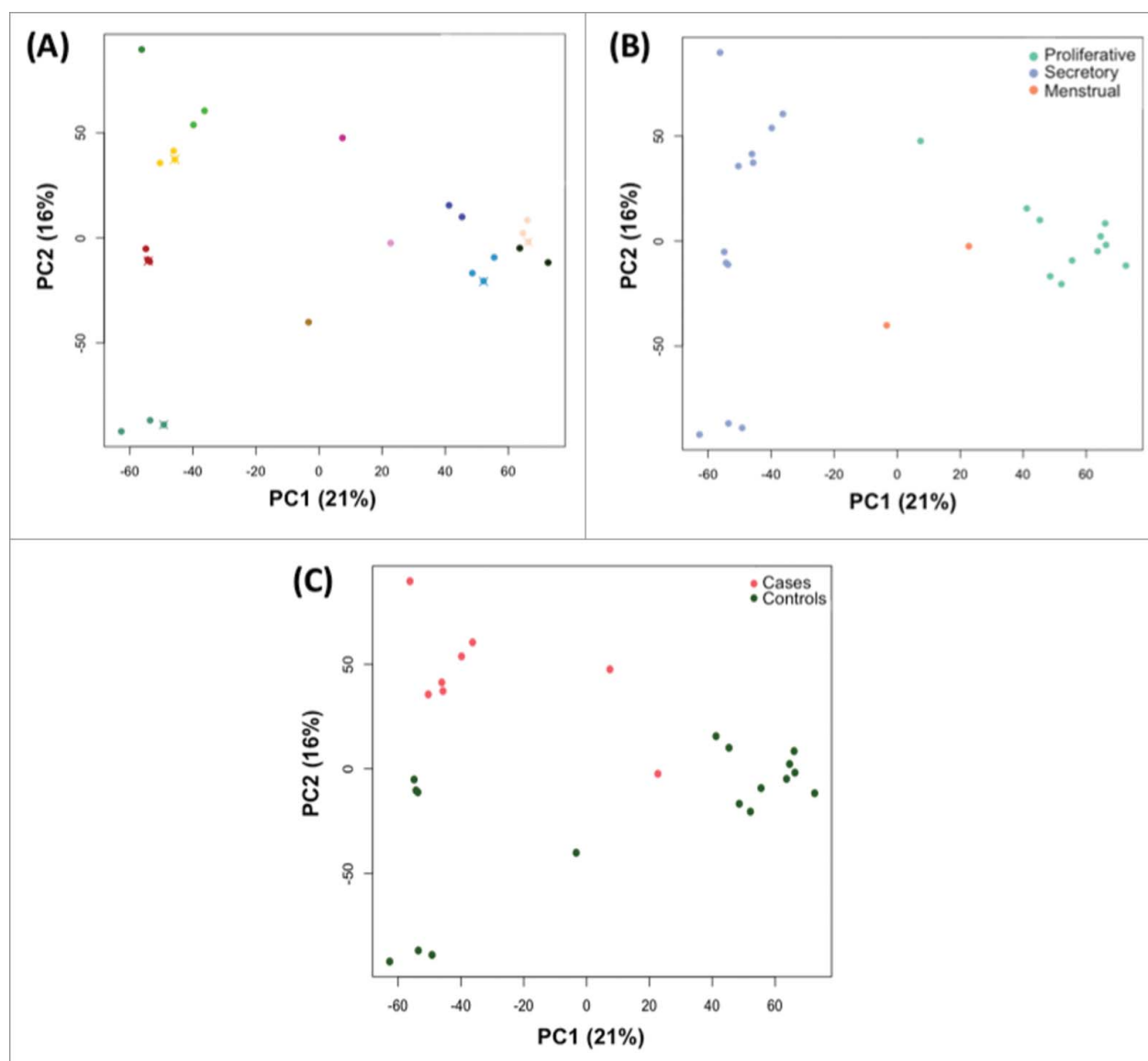


Figure 3. PCA analysis of RNA profiles in endometrium, showing total variation in profiles explained between PC1 and PC2. Each circle represents a sample; color coding designates (A) individual woman [same color circles represent the sample splits and crosses (x) indicate technical replicates]; (B) menstrual phase according to calculated cycle day based on self-reported LMP; (C) case/control status.

multiple-testing correction, 32 probes were significantly differentially expressed (Table S9), 28 upregulated and 4 downregulated. The fold-change (FC) in expression was >4 in 3 loci: *CYP26A1* (FC = 5.63, $P = 5.8 \times 10^{-6}$, $P_{adj} = 0.025$), *LRRC26* (FC = 4.49, $P = 6.6 \times 10^{-6}$, $P_{adj} = 0.026$), *LOC389816* (FC = 4.48, $P = 5.6 \times 10^{-6}$, $P_{adj} = 0.025$).

Quantification of RNA expression variability using variance component analysis

Similar to DNA methylation variance component analyses, when considering all genes probed by the RNA expression array irrespective of probe variability, the main source of variation was technical (Fig. S11a). However, when we considered the 50% vs. 10% most variable expression sites (Figs. S11b and S11c), the between-individual variance became the dominating component, although technical and within-tissue variation was larger than observed for DNA methylation patterns. Mean

proportion of variance estimates for the top 10% vs. 50% most variable probes (Table S10) were 0.48 (SD = 0.39) vs. 0.40 (0.37) for between-individual variance, 0.27 (SD = 0.24) vs. 0.25 (SD = 0.25) for within-tissue variance, and 0.30 (SD = 0.25) vs. 0.40 (SD = 0.29) for technical variation, respectively.

Power calculations

To determine the sample size needed to detect biologically relevant differential DNA methylation and expression between cases and controls, we conducted power calculations based on the 222,456 CpG-sites (80% power, $\alpha = 0.05/222,456 = \sim 2 \times 10^{-7}$) and 11,464 expression-probes (80% power, $\alpha = 0.05/11,464 = \sim 4 \times 10^{-6}$) for endometrium, where between-individual variance of DNA methylation/RNA expression exceeded within-tissue and technical variance (see Materials and Methods). Fig. 5A-C shows the results for power curves

of 3 sample sizes for differential DNA methylation in endometrium and fat, and expression in endometrium between cases and controls, and Fig. 5D for detection of differential DNA methylation between paired endometriotic lesion tissue and endometrium from cases. For example, a 2% ($\beta = 0.02$) difference in DNA methylation with 80% power between cases and controls in endometrium can be detected with a sample size of 100 in 67,968/222,456 (30%) probes, with a sample size of 500 in 172,947/222,456 (78%) probes, and with a sample size of 1000 in 204,720/222,456 (92%) probes in endometrium (Fig. 5A). Most common effect sizes observed in DNA methylation studies of complex disease/traits range between 0.01–0.10 β value differences and gene expression ranges between 1–3-fold-change.^{13,16,17,43,44}

Discussion

This is the first study that systematically investigated the different sources of variability influencing DNA methylation and RNA expression profiles in tissues relevant to many reproductive, endocrine, as well as metabolic traits. The results have direct implications for studies including these profiles in terms of experimental design and sample size requirements.

We showed that in DNA methylation profiles from endometrium, endometriotic disease tissue, and subcutaneous abdominal fat, and in RNA expression profiles from endometrium, the between-individual variation is greater than within-tissue variation (tissue heterogeneity) and experimental (technical) variation when considering moderate to highly variable sites. This is reassuring for tissue-based profiling studies, given that most disease-association studies are focused on DNA methylation or RNA expression sites that at least show moderate variability across individuals in a population. Limiting to the top 10% most variable probes, the between-individual variation was 84.2% for endometrium, 84.2% for endometriotic disease tissue, and 79% for abdominal fat tissue. In expression profiles from endometrium, the between-individual variation (47.6%) also dominated over within-tissue (27%) and experimental (30%) variation in the top 10% most variable probes, although experimental variability was clearly more pronounced in expression profiling. This difference in contribution of experimental variability (6% vs. 29%) is probably expected because of the less stable nature of RNA compared with DNA, but it highlights the importance of using standardized protocols for sample collection and tissue processing, as well as adopting good experimental design in gene expression experiments.⁴⁵

Overall, these results underline the application of whole-tissue profiling in these tissues to detect robust DNA methylation and expression differences between individuals, but they do not negate the importance of research focusing on profiling of distinct subcellular components, e.g., stromal and glandular epithelial cells in endometrium. Indeed, comparison of tissue-based with cell-specific profiles is crucial to understanding underlying regulatory biologic mechanisms. However, cell-type specific profiling is not without its own limitations (including potential for disturbance of expression profiles caused by the isolation protocols, and lack of scalability to larger studies). Our results suggest that for low-variability DNA methylation and RNA expression sites, cell-specific studies of sufficient size

are particularly important, because tissue-based (cellular) heterogeneity and technical variability are likely to hinder the detection of biologically interesting variability.

Our results showed that menstrual phase is an important covariate and potential confounder in the analysis of differential DNA methylation and RNA expression in endometrium, as previously reported,^{33,39,41} and also in DNA methylation of peritoneal lesions—a novel finding. However, its effect on variability of DNA methylation in endometriomas and fat is much less pronounced. A recent study including 17 endometriosis cases (4 proliferative, 7 early secretory, 6 mid-secretory phases) and 16 controls (6 proliferative, 5 early secretory, 5 mid-secretory phases) suggested that DNA methylation changes across the menstrual cycle are associated with changes in gene expression for many loci associated with endometriosis, and that the greatest DNA methylation differences were observed in the mid-secretory phase.³⁸ It is therefore important to investigate: (1) expression and DNA methylation profiles of endometrium in cases vs. controls; (2) DNA methylation profiles of peritoneal lesion vs. other tissues, adjusted or matched for menstrual cycle phase.

Age was a factor strongly influencing DNA methylation profiles in both endometrium and endometriotic disease tissue. There is abundant literature on DNA methylation differences associated with age, and the importance of age-related environmental influences.^{46–49} The age effect we observed could signify either population-based differences in exposures between women of different ages or the cumulative effects of lifetime exposures on DNA methylation sites. This highlights the need to ascertain in a larger study what these age-related differences constitute and how they might relate to endometrial biology and disease processes. Interestingly, smoking was an important factor influencing variability of DNA methylation in fat tissue—a novel finding. It is well established that smoking affects DNA methylation⁵⁰ and many epigenetic markers related to smoking have been identified in blood including in inflammatory diseases, such as rheumatoid arthritis.^{49,51–54} Moreover, there is some evidence, which suggest smoking as a protective factor for endometriosis⁵⁵ and endometrial cancer;⁵⁶ however, the evidence is conflicting⁵⁷ and requires further investigation. Intriguingly, endometriosis case/control status also influenced variability of DNA methylation profiles in fat tissue. We showed previously that there is significant sharing in the genetic etiology of endometriosis and fat distribution.¹² Our findings suggest differential gene regulation in fat tissue between endometriosis cases and controls that warrants follow-up in a larger study.

We have shown that DNA methylation profiles of endometrium, endometriotic disease tissue, and abdominal fat are distinctly different, and that 2 different endometriosis tissue types—endometriomas and peritoneal lesions—have different DNA methylation signatures. Despite the relatively small sample size, differential DNA methylation analysis between endometrium and endometriotic disease tissue identified a set of differentially methylated genes, which were significantly enriched for 7 pathways, the top being *WNT* signaling—a pathway involved in sex-hormone homeostasis and reproductive tract development previously associated in genome-wide association analyses with endometriosis and in shared genetic regulation between endometriosis and fat distribution.¹² Other pathways enriched for differential DNA methylation between endometrium and endometriotic disease tissue were angiogenesis, cadherin

signaling (involved in cell adhesion and inflammation), and gonadotropin-releasing-hormone-receptor pathway. This suggests that genetic pathways perturbed by genetic variation in endometriosis are also dysregulated in terms of DNA methylation profiles. To what extent these DNA methylation differences are genetically driven or reflect differences in extra-uterine milieu, and how these might influence disease processes, needs to be explored in an adequately powered study. Differential DNA methylation analysis between endometriomas and peritoneal lesions identified 31 significantly differential methylated sites, though a limitation here is the potential for peritoneal cellular contamination driving some of these differences. All our differential DNA methylation analyses were exploratory and findings need to be replicated in larger studies.

We observed 32 differentially expressed RNA probes comparing endometrium between endometriosis cases and controls, after multiple testing correction. *CYP26A1* was upregulated >5-fold in cases compared with controls and is of particular interest since it is a progesterone-regulated gene^{58,59} that has been found also to be upregulated in the proliferative phase for rAFS stage III/IV cases vs. stage I/II¹⁹, while it was found to be nominally downregulated in endometriosis cases vs. controls in the secretory phase.²² The limited number of cases and controls within each cycle phase in this study prohibited us from performing between-phases or within-phase differential expression analyses. *CYP26A1* encodes for a retinoic acid metabolizing enzyme, and maps close to a region of the genome identified to be significantly linked to endometriosis in family-based analyses.⁶⁰ Follow-up of the role of this locus in endometriosis is warranted in a larger study, in the context of menstrual phase.

A major limitation in deciphering the epigenomic and transcriptomic landscape of endometrium is that this tissue is not part of any large-scale genomic annotation projects such as ENCODE⁶¹ and NIH Epigenome Roadmap.¹⁷ For example, histone modification sites and DNase I hypersensitivity sites provide information regarding enhancer regions in the genome. Enhancer regions play a central role in driving cell type-specific gene expression and enrichment of DNA methylation marks in these regions informs us about their potential functional role. We conducted our DNA methylation marks enrichment analysis in 127 different cell types (none relevant to endometrium). H3K4me1 histone modification was enriched across the majority of cell types, suggesting this as an enriched mark that is relatively non-specific to cell type. H3K4me1 histone modification is globally linked to distal regulatory regions.⁶² Genomic annotation data for endometrium and its constituting cell types is needed to allow a more informative exploration of these results.

Our power calculations show that tissue-based differential DNA methylation and expression studies require sample sizes of at least 500 tissues to detect the most common effect sizes expected for complex disease in these tissue types. For example, with sample size of 500 tissues, we can detect 2% DNA methylation difference between cases and controls in 172,947/222,456 (78%) CpG-sites in endometrium, 209,480/255,402 (82%) sites in fat, and between endometrium and endometriotic disease tissue in 393,584/483,422 (82%) sites. Most common effect sizes observed in large-scale complex traits DNA methylation studies range between 1–10%.^{17,43,44} This sample size also allows for detection of 1.5-fold change in gene expression of 9,500/11,464

(83%) expression-probes in endometrium. Large effect sizes could be detectable with sample sizes >100; however, this is unlikely to allow the discovery of most common mechanisms robustly. Reaching these sample sizes requires collaboration between centers that collect phenotypic data and biologic samples in a standardized manner. The WERF Endometriosis Phenome and Biobanking Harmonisation Project (EPHect) has provided standard operating protocols for collection, preprocessing and storage of tissues including endometrium, along with standardized questionnaires for phenotypic data collection,^{45,63–65} which allows such data to be generated and robust studies to be performed.

In summary, our study has identified parameters affecting DNA methylation and RNA expression variability in tissues relevant to many reproductive and endocrine research studies. It provides essential guidance for studies aimed at integrated analysis of genomic, transcriptomic, and epigenomic data in endometrium, which have become staple approaches for other tissues in different diseases and will aid further understanding of the biologic mechanisms underlying endometrium-related disorders.

Materials and methods

Recruitment, tissue, and data collection

Samples and data were derived from ENDOX, a prospective study focused on endometriosis and endometrium-related conditions, in which women undergoing a laparoscopy at the Endometriosis CaRe Center, John Radcliffe Hospital, Oxford (UK), are recruited. All data collection instruments and standard operating procedures (SOPs) used in ENDOX are in accordance with WERF EPHect standards^{45,63–65} (See Supplementary Note). Ethics approval for ENDOX was obtained from the NRES Committee South Central-Oxford Research Ethics Committee (09/H0604/58).

We selected 16 endometriosis patients and 8 healthy controls with regular menstrual cycles (predictable within 1 week), who had not used hormones at least 3 months before recruitment (Table S2). Controls were symptomatic [i.e., undergoing laparoscopy either for pelvic pain ($n = 4$), or subfertility ($n = 4$)] but no endometriosis was found during laparoscopy. Four controls were diagnosed with non-endometriotic adhesions, one with a non-endometriotic cyst, and 3 had no obvious pathology. Controls were frequency-matched to cases on menstrual phase (Table S3). Menstrual cycle was categorized based on self-reported last menstrual period (LMP) and cycle day at time of surgery, adjusted for cycle length,³³ into menstrual phase (1–7 days), proliferative phase (8–14 days), secretory phase (15+ days); the menstrual phase was confirmed by RNA expression profiles (See Statistical analysis section).

Experimental design

Figure 1 shows the experimental design. A 100–120 mg section of endometrium and subcutaneous abdominal fat from 8 cases and 8 controls, and endometriotic disease tissue from 16 cases (8 of which also contributed fat and endometrium) were used for DNA methylation analyses; the same endometrium tissue samples ($n = 8$) were also used for mRNA analyses (Fig. 1A). To investigate within-tissue (cellular heterogeneity) variability

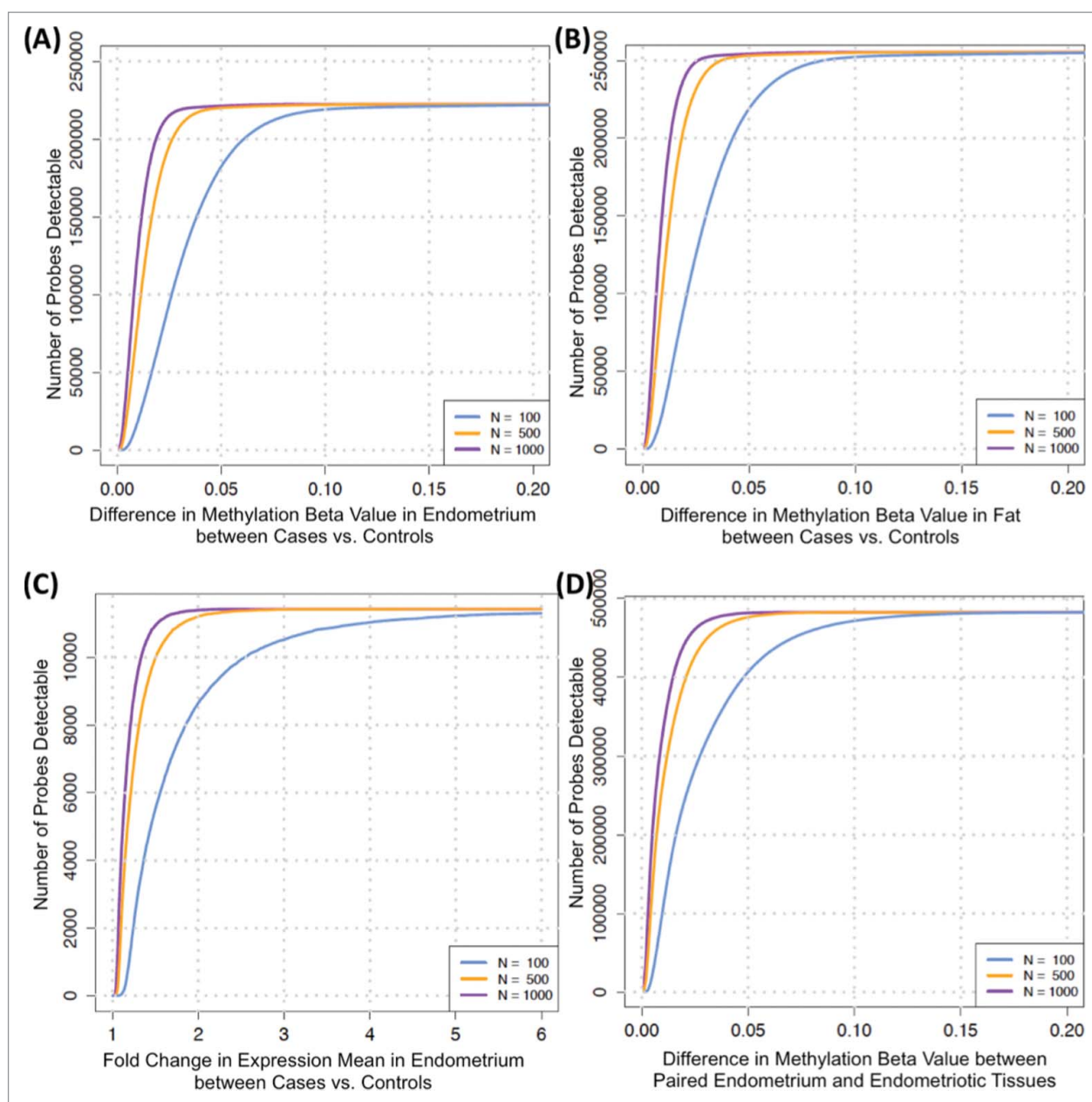


Figure 4. Power curves (80%) for 3 sample sizes (using the estimated between-individual variances) to determine the detectable (A) number of methylation probes with a difference in methylation values (β range 0–0.2) between cases and controls in endometrium; (B) number of methylation probes with a difference in methylation values (β range 0–0.2) between cases and controls in fat tissue; (C) number of expression probes with fold change in mean expression levels ranging between 1 and 6, between cases and controls in endometrium; (D) number of methylation probes with a difference in methylation values (β range 0–0.2) in cases, between endometriotic disease tissue and endometrium.

of DNA methylation and RNA expression profiles, each of the 16 endometriotic disease tissue and endometrium samples and 8 of the 16 fat samples were split in random order before extraction (Fig. 1b). DNA and RNA extractions were conducted again in random order. To further allow for investigation of technical (array-related) variability, 8 replicates of the same DNA and 7 replicates of the same RNA sample were included on the arrays. Samples were run on arrays in random order.

DNA extraction and methylation analysis

Genomic DNA was extracted from 88 samples (Fig. 1) using Qiagen DNeasy Kits and quantified by PicoGreen. Six samples

with a yield $>1 \mu\text{g}$ but concentrations lower than $16.5 \text{ ng}/\mu\text{l}$ were concentrated by SpeedVacTM (Thermo Scientific). Whole-genome DNA methylation analysis was performed for 96 samples (including 8 replicates) using Illumina Infinium HumanMethylation450 BeadChip arrays, following the Illumina HumanMethylation450 protocol,⁶⁶ and arrays were imaged using an Illumina HiScanTM scanner.

RNA extraction and expression analysis

RNA was extracted from 32 endometrium samples (Fig. 1) using Qiagen RNeasy Kit and quantified using Nanodrop. RNA integrity (RIN) scores were obtained for each sample using

Agilent Bioanalyser, before whole genome gene-expression analyses (39 samples, including 7 replicates) using Illumina Human HT12v4.0 Expression Bead arrays; arrays were imaged using an Illumina iScan Scanner.

Statistical analyses

DNA and RNA Quality

DNA quality was measured as %CpG-sites detected at the 1% detection *P* value cut-off; RNA quality was assessed using RIN-scores. DNA/RNA quality was correlated with initial yield, menstrual cycle day, and tissue weight using Pearson's correlations, and with tissue type using one-way ANOVA linear models.

Principal component analysis (PCA) and hierarchical clustering

To investigate clustering of samples based on DNA methylation or expression profiles, PCA was performed for each tissue type separately, using the *prcomp* package with default settings in R (3.2.5). Hierarchical clustering based on DNA methylation profiles was performed using the *hclust* package in R. Similarly, the hierarchical cluster analysis (complete linkage based on Euclidean distance) identified 2 sample outliers (X_14, X_15), which are likely tissue sample swaps that were excluded from any downstream analyses. The univariate association between the PCs explaining most of the variance and different parameters (e.g., menstrual phase, age, smoking, BMI, WHR) was assessed using Pearson correlations and linear models. For 3 individuals with self-reported cycle days that were close to cycle categorization cut-offs (Table S1), cycle phase was re-classified via PC-based expression profile clustering.

Variance component analysis

For each tissue type, the proportion of different sources of variances contributing to total phenotypic variance was estimated for each probe using a linear mixed-effects model fitted by maximum likelihood. The estimated variances were extracted with the *VarCorr* function implemented in *lmer* in the R package *lme4*, and were decomposed into between-individual, within-tissue, and residual (technical) variance. More detailed methods are provided in Supplementary Note 1.

Differential DNA methylation analysis

Differential DNA methylation analysis was performed on the average of probe intensities across splits and replicates of 46 independent samples (16 endometrium; 16 fat; 14 endometriotic disease tissue) at each single CpG-site (483,422 total) by fitting linear models to all DNA methylation probes in R. We have conducted the following differential DNA methylation analyses using linear regression: (1) Endometrium from cases ($n = 16$) vs. all endometriotic disease tissues ($n = 14$) adjusting for age and menstrual phase (Unpaired analysis); (2) Endometrium from cases ($n = 8$) vs. all endometriotic disease tissues ($n = 8$) (Paired analysis); (3) Ovarian (endometriomas) ($n = 8$) vs. peritoneal disease tissues ($n = 6$) adjusting for age; (4) Endometrium from cases ($n = 8$) vs. controls ($n = 8$) adjusted for menstrual phase and age; (5) Fat from cases ($n = 8$) vs. controls ($n = 8$) adjusted for menstrual phase and age. *P* values were adjusted for multiple-testing using Bonferroni correction. The significantly differentially methylated genes

were followed-up by PANTHER pathway analysis [42](#) and enrichment analysis in histone modification sites and DNaseI hypersensitivity sites from NIH Epigenome Roadmap R9 data set¹⁷ (See Supplementary note).

Differential expression analysis

Differential expression analysis was performed on the average of the probe intensities across splits and replicates of the 12 independent samples. We investigated differentially expressed genes in endometrium between cases and controls through linear model fitting adjusting for menstrual cycle phase, using *limma* in R.⁶⁷ *P* values were adjusted for multiple-testing using Benjamini and Hochberg's method at the 5% false discovery rate (FDR) level.⁶⁸

Power calculations

Power to detect differential expression or DNA methylation was assessed using *pwr.t.test* in R for 3 example sample sizes: $n = 100$; $n = 500$; $n = 1000$. The significance level ($\alpha = 0.05$) was adjusted for the number of genes and DNA methylation sites found exhibiting differential expression/DNA methylation between individuals as defined percentage change in methylation mean (β value), and the fold change in expression mean (See Supplementary Note).

Conflict of interest

CB and KZ have scientific collaborations with Bayer AG, Roche Diagnostics, Volition Ltd, and MDNA, and in addition hold research grants from the US National Institutes of Health, Wellcome Trust, and the European Commission.

Data availability

The data sets used and/or analyzed during the current study are available from the corresponding author on request.

Footnote: The power calculations to detect differential DNA methylation/RNA expression between cases and controls in endometrium and fat analysis are based on probes where between-individual variance of methylation/expression is more than within-tissue and technical variances (Fig. 4A-C). Whereas, Fig. 4D present power calculation results for detection of differential methylation between paired endometrium and endometriotic disease tissues coming from the same cases across all probes; hence, the scale of probes on y-axis includes all probes.

Acknowledgments

AWD is an MRC Skills Development Fellow (MR/N015355/1). SM, LPT, PM, NR, CML, KTZ and CMB were funded by the European Commission Horizon 2020 research and innovation program under grant agreement 692065 (project WIDENLIFE) and grant IUT34-16 from the Estonian Ministry of Education and Research. This work was supported by an MRC project grant (MR/K011480/1).

We acknowledge all the women who participated in the ENDOX study, our dedicated research nurses Carol Hubbard and Maryjane Dale for recruitment and collection of all the data in clinic, our surgeons who collect the tissue samples during laparoscopy and the Oxford Genomics Centre, Wellcome Centre for Human Genetics, for assay work.

References

- Cheng TH, Thompson DJ, O'Mara TA, Painter JN, Glubb DM, Flach S, Lewis A, French JD, Freeman-Mills L, Church D, et al. Five endometrial cancer risk loci identified through genome-wide

- association analysis. *Nat Genet.* 2016;48(6):667-74. doi:10.1038/ng.3562. PMID:27135401
2. Rahmioglu N, Nyholt DR, Morris AP, Missmer SA, Montgomery GW, Zondervan KT. Genetic variants underlying risk of endometriosis: insights from meta-analysis of eight genome-wide association and replication datasets. *Hum Reprod Update.* 2014;20(5):702-16. doi:10.1093/humupd/dmu015. PMID:24676469
 3. Giudice LC, Kao LC. Endometriosis. *Lancet.* 2004;364(9447):1789-99. doi:10.1016/S0140-6736(04)17403-5
 4. Albertsen HM, Chettier R, Farrington P, Ward K. Genome-wide association study link novel loci to endometriosis. *PLoS one.* 2013;8(3):e58257. doi:10.1371/journal.pone.0058257. PMID:23472165
 5. Nyholt DR, Low SK, Anderson CA, Painter JN, Uno S, Morris AP, MacGregor S, Gordon SD, Henders AK, Martin NG, et al. Genome-wide association meta-analysis identifies new endometriosis risk loci. *Nat Genet.* 2012;44(12):1355-9. doi:10.1038/ng.2445. PMID:23104006
 6. Pagliardini L, Gentilini D, Sanchez AM, Candiani M, Viganò P, Di Blasio AM. Replication and meta-analysis of previous genome-wide association studies confirm *vezatin* as the locus with the strongest evidence for association with endometriosis. *Hum Reprod.* 2015;30(4):987-93. doi:10.1093/humrep/dev022. PMID:25678572
 7. Painter JN, Anderson CA, Nyholt DR, Macgregor S, Lin J, Lee SH, Lambert A, Zhao ZZ, Roseman F, Guo Q, et al. Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat Genet.* 2011;43(1):51-4. doi:10.1038/ng.731. PMID:21151130
 8. Uno S, Zembutsu H, Hirasawa A, Takahashi A, Kubo M, Akahane T, Aoki D, Kamatani N, Hirata K, Nakamura Y. A genome-wide association study identifies genetic variants in the *CDKN2BAS* locus associated with endometriosis in Japanese. *Nat Genet.* 2010;42(8):707-10. doi:10.1038/ng.612. PMID:20601957
 9. Treloar SA, O'Connor DT, O'Connor VM, Martin NG. Genetic influences on endometriosis in an Australian twin sample. *Hum Reprod.* 1999;14(4):701-10. doi:10.1016/S0015-0282(98)00540-8
 10. Sapkota Y, Steinhorsdottir V, Morris AP, Fassbender A, Rahmioglu N, De Vivo I, Buring JE, Zhang F, Edwards TL, Jones S, et al. Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nat Commun.* 2017;24(8):15539. doi:10.1038/ncomms15539. PMID:28537267
 11. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinhorsdottir V, Thorleifsson G, Zillikens MC, Speliotes EK, Mägi R, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet.* 2010;42(11):949-60. doi:10.1038/ng.685. PMID:20935629
 12. Rahmioglu N, Macgregor S, Drong AW, Hedman AK, Harris HR, Randall JC, Prokopenko I, Nyholt DR, Morris AP, Montgomery GW, et al. Genome-wide enrichment analysis between endometriosis and obesity-related traits reveals novel susceptibility loci. *Human molecular genetics.* 2015;24(4):1185-99. doi:10.1093/hmg/ddu516. PMID:25296917
 13. Grunberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet.* 2012;44(10):1084-9. doi:10.1038/ng.2394. PMID:22941192
 14. Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, Dermitzakis ET. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS genetics.* 2010;6(4):e1000895. doi:10.1371/journal.pgen.1000895. PMID:20369022
 15. Consortium GT. The Genome-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-5. doi:10.1038/ng.2653. PMID:23715323
 16. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015;348(6235):648-60. doi:10.1126/science.1262110. PMID:25954001
 17. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilienky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang X, Wang J, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518(7539):317-30. doi:10.1038/nature14248. PMID:25693563
 18. Altmae S, Esteban FJ, Stavreus-Evers A, Simon C, Giudice L, Lessey BA, Horcajadas JA, Macklon NS, D'Hooghe T, Campoy C, et al. Guidelines for the design, analysis and interpretation of 'omics' data: focus on human endometrium. *Hum Reprod Update.* 2014;20(1):12-28. doi:10.1093/humupd/dmt048. PMID:24082038
 19. Aghajanova L, Giudice LC. Molecular evidence for differences in endometrium in severe versus mild endometriosis. *Reprod Sci.* 2011;18(3):229-51. doi:10.1177/1933719110386241. PMID:21063030
 20. Arimoto T, Katagiri T, Oda K, Tsunoda T, Yasugi T, Osuga Y, Yoshikawa H, Nishii O, Yano T, Taketani Y, et al. Genome-wide cDNA microarray analysis of gene-expression profiles involved in ovarian endometriosis. *International journal of oncology.* 2003;22(3):551-60. PMID:12579308
 21. Burney RO, Hamilton AE, Aghajanova L, Vo KC, Nezhat CN, Lessey BA, Giudice LC. MicroRNA expression profiling of eutopic secretory endometrium in women with versus without endometriosis. *Mol Hum Reprod.* 2009;15(10):625-31. doi:10.1093/molehr/gap068. PMID:19692421
 22. Burney RO, Talbi S, Hamilton AE, Vo KC, Nyegaard M, Nezhat CR, Lessey BA, Giudice LC. Gene expression analysis of endometrium reveals progesterone resistance and candidate susceptibility genes in women with endometriosis. *Endocrinology.* 2007;148(8):3814-26. doi:10.1210/en.2006-1692. PMID:17510236
 23. Burrows CK, Kosova G, Herman C, Patterson K, Hartmann KE, Velez Edwards DR, Stephenson MD, Lynch VJ, Ober C. Expression Quantitative Trait Locus Mapping Studies in Mid-secretory Phase Endometrial Cells Identifies HLA-F and TAP2 as Fecundability-Associated Genes. *PLoS Genet.* 2016;12(7):e1005858. doi:10.1371/journal.pgen.1005858. PMID:27447835
 24. Fassbender A, Verbeeck N, Bornigen D, Kyama CM, Bokor A, Vodolazkaia A, Peeraer K, Tomassetti C, Meuleman C, Gevaert O, et al. Combined mRNA microarray and proteomic analysis of eutopic endometrium of women with and without endometriosis. *Hum Reprod.* 2012;27(7):2020-9. doi:10.1093/humrep/des127. PMID:22556377
 25. Fung JN, Girling JE, Lukowski SW, Sapkota Y, Wallace L, Holdsworth-Carson SJ, et al. The genetic regulation of transcription in human endometrial tissue. *Hum Reprod.* 2017;32(4):893-904. doi:10.1093/humrep/dex006. PMID: 28177073
 26. Girling JE, Lockhart MG, Olshansky M, Paiva P, Woodrow N, Marino JL, Hickey M, Rogers PA. Differential Gene Expression in Menstrual Endometrium From Women With Self-Reported Heavy Menstrual Bleeding. *Reprod Sci.* 2017;24(1):28-46. doi: 10.1177/1933719116648217. PMID: 27189201
 27. Kao LC, Germeyer A, Tulac S, Lobo S, Yang JP, Taylor RN, Osteen K, Lessey BA, Giudice LC. Expression profiling of endometrium from women with endometriosis reveals candidate genes for disease-based implantation failure and infertility. *Endocrinology.* 2003;144(7):2870-81. doi:10.1210/en.2003-0043. PMID:12810542
 28. Khan MA, Sengupta J, Mittal S, Ghosh D. Genome-wide expressions in autologous eutopic and ectopic endometrium of fertile women with endometriosis. *Reproductive biology and endocrinology: RB&E.* 2012;10:84. doi:10.1186/1477-7827-10-84
 29. Ohlsson Teague EM, Van der Hoek KH, Van der Hoek MB, Perry N, Wagaarachchi P, Robertson SA, Print CG, Hull LM. MicroRNA-regulated pathways associated with endometriosis. *Mol Endocrinol.* 2009;23(2):265-75. doi:10.1210/me.2008-0387. PMID:19074548
 30. Paiva P, Lockhart MG, Girling JE, Olshansky M, Woodrow N, Marino JL, Hickey M, Rogers PA. Identification of genes differentially expressed in menstrual breakdown and repair. *Mol Hum Reprod.* 2016;22(12):898-912. doi:10.1093/molehr/gaw060. PMID:27609758
 31. Petracco RG, Kong A, Grechukhina O, Krikun G, Taylor HS. Global gene expression profiling of proliferative phase endometrium reveals distinct functional subdivisions. *Reprod Sci.* 2012;19(10):1138-45. doi:10.1177/1933719112443877. PMID:22623515
 32. Sha G, Wu D, Zhang L, Chen X, Lei M, Sun H, Lin S, Lang J. Differentially expressed genes in human endometrial endothelial cells derived from eutopic endometrium of patients with endometriosis compared with those from patients without endometriosis. *Hum Reprod.* 2007;22(12):3159-69. doi:10.1093/humrep/dem266. PMID:17956924
 33. Talbi S, Hamilton AE, Vo KC, Tulac S, Overgaard MT, Dosiou C, Le Shay N, Nezhat CN, Kempson R, Lessey BA, et al. Molecular phenotyping of human endometrium distinguishes menstrual cycle phases and underlying biological processes in normo-ovulatory women. *Endocrinology.* 2006;147(3):1097-121. doi:10.1210/en.2005-1076. PMID:16306079

34. Tamaresis JS, Irwin JC, Goldfien GA, Rabban JT, Burney RO, Nezhat C, DePaolo LV, Giudice LC. Molecular classification of endometriosis and disease stage using high-dimensional genomic data. *Endocrinology*. 2014;155(12):4986-99. doi:10.1210/en.2014-1490. PMID:25243856
35. Wong YF, Cheung TH, Lo KW, Yim SF, Siu NS, Chan SC, Ho TW, Wong KW, Yu MY, Wang VW, et al. Identification of molecular markers and signaling pathway in endometrial cancer in Hong Kong Chinese women by genome-wide gene expression profiling. *Oncogene*. 2007;26(13):1971-82. doi:10.1038/sj.onc.1209986. PMID:17043662
36. Wu Y, Kajdacsy-Balla A, Strawn E, Basir Z, Halverson G, Jailwala P, Wang Y, Wang X, Ghosh S, Guo SW.. Transcriptional characterizations of differences between eutopic and ectopic endometrium. *Endocrinology*. 2006;147(1):232-46. doi:10.1210/en.2005-0426. PMID:16195411
37. Borghese B, Barbaux S, Mondon F, Santulli P, Pierre G, Vinci G, Chapron C, Vaiman D. Research resource: genome-wide profiling of methylated promoters in endometriosis reveals a subtelomeric location of hypermethylation. *Mol Endocrinol*. 2010;24(9):1872-85. doi:10.1210/me.2010-0160. PMID:20685852
38. Houshdaran S, Nezhat CR, Vo KC, Zelenko Z, Irwin JC, Giudice LC. Aberrant Endometrial DNA Methylome and Associated Gene Expression in Women with Endometriosis. *Biol Reprod*. 2016;95(5):93. doi:10.1095/biolreprod.116.140434
39. Houshdaran S, Zelenko Z, Irwin JC, Giudice LC. Human endometrial DNA methylome is cycle-dependent and is associated with gene expression regulation. *Mol Endocrinol*. 2014;28(7):1118-35. doi:10.1210/me.2013-1340. PMID:24877562
40. Naqvi H, Ilagan Y, Krikun G, Taylor HS. Altered genome-wide methylation in endometriosis. *Reprod Sci*. 2014;21(10):1237-43. doi:10.1177/1933719114532841. PMID:24784717
41. Saare M, Modhukur V, Suhorutshenko M, Rajashekar B, Bekker K, Soritsa D, Karro H, Soplemann P, Söritsa A, Lindgren CM, et al. The influence of menstrual cycle and endometriosis on endometrial methylome. *Clin Epigenetics*. 2016;8:2. doi:10.1186/s13148-015-0168-z. PMID:26759613
42. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res*. 2016;44(D1):D336-42. doi:10.1093/nar/gkv1194. PMID:26578592
43. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, Wahl S, Meduri E, Morange PE, Gagnon F, Grallert H, et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet*. 2014;383(9933):1990-8. doi:10.1016/S0140-6736(13)62674-4. PMID:24630777
44. Kriebel J, Herder C, Rathmann W, Wahl S, Kunze S, Molnos S, Volkova N, Schramm K, Carstensen-Kirberg M, Waldenberger M, et al. Association between DNA Methylation in Whole Blood and Measures of Glucose Metabolism: KORA F4 Study. *PLoS one*. 2016;11(3):e0152314. doi:10.1371/journal.pone.0152314. PMID:27019061
45. Fassbender A, Rahmioglu N, Vitonis AF, Viganò P, Giudice LC, D'Hooghe TM, Hummelshoj L, Adamson GD, Becker CM, Missmer SA, et al. World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: IV. Tissue collection, processing, and storage in endometriosis research. *Fertility and sterility*. 2014;102(5):1244-53.
46. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS genetics*. 2012;8(4):e1002629. doi:10.1371/journal.pgen.1002629. PMID:22532803
47. Horvath S. DNA methylation age of human tissues and cell types. *Genome biology*. 2013;14(10):R115. doi:10.1186/gb-2013-14-10-r115. PMID:24138928
48. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, van den Berg LH, Ophoff RA. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome biology*. 2012;13(10):R97. doi:10.1186/gb-2012-13-10-r97. PMID:23034122
49. van Dongen J, Nivard MG, Willemsen G, Hottenga JJ, Helmer Q, Dolan CV, Ehli EA, Davies GE, van Ijzerson M, Breeze CE, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun*. 2016;7:11115. doi:10.1038/ncomms11115. PMID:27051996
50. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS one*. 2013;8(5):e63812. doi:10.1371/journal.pone.0063812. PMID:23691101
51. Guida F, Sandanger TM, Castagne R, Campanella G, Polidoro S, Palli D, Krogh V, Tumino R, Sacerdote C, Panico S, et al. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Human molecular genetics*. 2015;24(8):2349-59. doi:10.1093/hmg/ddu751. PMID:25556184
52. Svendsen AJ, Gervin K, Lyle R, Christiansen L, Kyvik K, Junker P, Nielsen C, Houen G, Tan Q. Differentially Methylated DNA Regions in Monozygotic Twin Pairs Discordant for Rheumatoid Arthritis: An Epigenome-Wide Study. *Front Immunol*. 2016;7:510. doi:10.3389/fimmu.2016.00510. PMID:27909437
53. Besingi W, Johansson A. Smoke-related DNA methylation changes in the etiology of human disease. *Human molecular genetics*. 2014;23(9):2290-7. doi:10.1093/hmg/ddt621. PMID:24334605
54. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet*. 2011;88(4):450-7. doi:10.1016/j.ajhg.2011.03.003
55. Simonelli A, Guadagni R, De Franciscis P, Colacurci N, Pieri M, Basilicata P, Pedata P, Lamberti M, Sannolo N, Miraglia N. Environmental and occupational exposure to bisphenol A and endometriosis: urinary and peritoneal fluid concentration levels. *Int Arch Occup Environ Health*. 2017;90(1):49-61. doi:10.1007/s00420-016-1171-1. PMID:27718009
56. Zhou B, Yang L, Sun Q, Cong R, Gu H, Tang N, Zhu H, Wang B. Cigarette smoking and the risk of endometrial cancer: a meta-analysis. *Am J Med*. 2008;121(6):501-8 e3. doi:10.1016/j.amjmed.2008.01.044
57. Bravi F, Parazzini F, Cipriani S, Chiaffarino F, Ricci E, Chiantera V, Viganò P, La Vecchia C. Tobacco smoking and risk of endometriosis: a systematic review and meta-analysis. *BMJ Open*. 2014;4(12):e006325. doi:10.1136/bmjopen-2014-006325. PMID:25534211
58. Deng L, Shipley GL, Loose-Mitchell DS, Stancel GM, Broaddus R, Pickar JH, Davies PJ. Coordinate regulation of the production and signaling of retinoic acid by estrogen in the human endometrium. *The Journal of clinical endocrinology and metabolism*. 2003;88(5):2157-63. doi:10.1210/jc.2002-021844. PMID:12727970
59. Jeong JW, Lee KY, Kwak I, White LD, Hilsenbeck SG, Lydon JP, DeMayo FJ. Identification of murine uterine genes regulated in a ligand-dependent manner by the progesterone receptor. *Endocrinology*. 2005;146(8):3490-505. doi:10.1210/en.2005-0016. PMID:15845616
60. Treloar SA, Wicks J, Nyholt DR, Montgomery GW, Bahlo M, Smith V, Dawson G, Mackay IJ, Weeks DE, Bennett ST, et al. Genomewide linkage study in 1,176 affected sister pair families identifies a significant susceptibility locus for endometriosis on chromosome 10q26. *Am J Hum Genet*. 2005;77(3):365-76. doi:10.1086/432960. PMID:16080113
61. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74. doi:10.1038/nature11247. PMID:22955616
62. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? *Mol Cell*. 2013;49(5):825-37. doi:10.1016/j.molcel.2013.01.038
63. Becker CM, Laufer MR, Stratton P, Hummelshoj L, Missmer SA, Zondervan KT, Adamson GD; WERF EPHeCT Working Group. World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: I. Surgical phenotype data collection in endometriosis research. *Fertility and sterility*. 2014;102(5):1213-22. doi:10.1016/j.fertnstert.2014.07.709. PMID:25150390
64. Rahmioglu N, Fassbender A, Vitonis AF, Tworoger SS, Hummelshoj L, D'Hooghe TM, Adamson GD, Giudice LC, Becker CM, Zondervan KT, et al. World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonization Project: III. Fluid biospecimen collection, processing, and storage in endometriosis research. *Fertility and sterility*. 2014;102(5):1233-43.

65. Vitonis AF, Vincent K, Rahmioglu N, Fassbender A, Buck Louis GM, Hummelshoj L, Giudice LC, Stratton P, Adamson GD, Becker CM, et al. World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonization Project: II. Clinical and covariate phenotype data collection in endometriosis research. *Fertility and sterility*. 2014;102(5):1223-32.
66. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, et al. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011;98(4):288-95. doi:10.1016/j.ygeno.2011.07.007. PMID:21839163
67. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007. PMID:25605792
68. Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*. 2003;19(3):368-75. doi:10.1093/bioinformatics/btf877. PMID:12584122