Elucidating the Heterogeneity of Endometriosis using Multi-omics



Sze Kan Cecilia CHEUK

Linacre College

University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2020

Declaration

I declare that this thesis is entirely my own work, and except where otherwise stated, describes my own research.

Sze Kan Cecilia CHEUK

Linacre College

Dedication

This thesis is dedicated to my family, May and Jason,
for their unconditional love, support, and encouragement.

'Behold the fowls of the air: for they sow not,
neither do they reap, nor gather into barns;
yet your heavenly Father feedeth them.
Are ye not much better than they?'

Matthew 6:26

Abstract

Endometriosis is a common chronic inflammatory gynaecological condition associated with a spectrum of clinical presentations including different types of pelvic pain and infertility. It is defined as the presence of endometrial-like tissue outside the uterus. Diagnosis requires invasive surgery, resulting in a typical diagnostic delay of 7-10 years, whilst treatment options are limited to repeated surgeries or hormonal medication. Although the disease has a moderate (50%) heritability, the genetic and epigenetic causes of the disease are largely unknown; where genetic variants have been identified, the biological pathways that are perturbed remains unknown. Despite the evidence that endometriosis is a clinically and aetiologically heterogeneous condition, most studies to date are studying endometriosis as binary condition, without regard for the wide spectrum of sub-phenotypes that patients are experiencing. Understanding biological differences between sub-phenotypes of endometriosis, both surgical and symptomatic, is important for the development of better targeted treatment methods as well as non-invasive methods of diagnosis. This thesis aims to examine for the first time differences in genomic, transcriptomic, and epigenomic characteristics of different sub-phenotypes of endometriosis, to aid understanding of their pathogenesis.

To improve understanding of 46 established genetic variants associated with endometriosis in genome-wide association studies (GWAS) at a sub-phenotypic level, a genetic association study of detailed phenotypic data from patients recruited in the Oxford ENDOX study was conducted (Chapter 3). Out of 46 GWAS variants for endometriosis, 16 showed nominal association with at least one sub-phenotype: rs11674184/2p25.1/*GREB1* and rs1903068/4q12/*KDR* were associated with ovarian endometriosis (endometrioma), rs71575922/6q25.1/*SYNE1* with ASRM stage I/II, and rs12441483/15q15.1/*BMF*, rs71575922/6q25.1/*SYNE1* and rs1903068/4q12/*KDR* with multiple pain sub-phenotypes.

Next, prior to conducting transcriptome and DNA methylome analyses of eutopic endometrium — the tissue of origin for endometriotic lesions – differences in RNA-seq gene expression profiles of eutopic endometrium of ENDOX recruits between major menstrual cycle phases was investigated (Chapter 4) to allow for appropriate adjustment of menstrual cycle phase in subsequent analyses. The analyses confirmed the literature reports of a significant cycle phase effect on gene expression patterns, and highlighted the complement system as one of the most robustly enriched and up-regulated pathways in the secretory phase of the endometrium

compared to proliferative phase, which may aid future building of model for cycle phase prediction through transcriptomic profiling.

Transcriptomic differences in eutopic endometrium comparing endometriosis cases, by surgical sub-phenotypes, and endometriosis-free controls, were analysed adjusting for cycle phase (Chapter 5). The differential expression analyses in proliferative and secretory phase endometrium highlighted the genome region of 21q11.2 for its potential involvement in inflammatory and oestrogen pathways for ASRM surgical stage I/II and III/IV cases. Five functionally similar clusters of differential expressed genes were proposed: innate immunity (*BPIFB1*, *MUC5AC*, *MUC5B*, *EDN2*, *THBS1*), inflammation (*SAA1*, *SAA2*), neuro-proliferation (*FAT2*, *FAT3*), haemoglobin (*HBG2*, *ALAS2*), and lipid metabolism (*PLIN1*, *PLIN4*, *GPD1*).

Lastly, differential DNA methylation of eutopic endometrium was explored, comparing endometriosis cases with endometriosis-free controls, in the context of GWAS loci and expression QTL SNPs nearby (*in cis*) (Chapter 6). For the first time, these analyses considered surgical sub-phenotypes (ASRM disease stage) of endometriosis. The analysis revealed loci at chromosome 4q35.1 (*FAM149A* & *TLR3*) that may be responsible for normal regulation of menstrual cycle. Nearby genes of nominally significant EWAS loci ($p<1x10^{-5}$) for endometriosis suggested five biological mechanisms in common between ASRM surgical stage I/II and III/IV, involving cell cycle (*CCND1*, *RAD52*, *PAGR1*, *STEAP3*), immunity (*PIANP*, *SPP1*, *TLR3*, *SHE*), neural regulation (*PIANP*, *C1QL2*, *GPR158*), iron (*LYRM4*, *BOLA2*, *STEAP3*), and oestrogen pathways (*PAGR1*, *ABCG2*, *SPP1*, *TLR3*, *TDRD10*). Further validation of the results is needed in an independent dataset before drawing further conclusions.

In conclusion, the thesis outlined potential differences between endometriosis sub-phenotypes in genetic, transcriptomic, and epigenetic signatures, although the relatively modest sample sizes mean all results require replication. Nevertheless, they encourage future research to continue exploring sub-phenotypes of endometriosis comprehensively, at much larger scale, using standardised phenotype definitions.

Acknowledgements

This work would not be possible without the E.P.A. Cephalosporin Scholarship from Linacre College, University of Oxford, and the Pay It Forward Scholarship from China Oxford Scholarship Fund.

I would like to express my gratitude to Professor Krina T Zondervan, Dr Nilufer Rahmioglu, Dr Christian M Becker, and Dr Karin Hellner for their supervision and support over the course of my study. I am grateful to colleagues from the Oxford Endometriosis CaRe Centre to support the research, and women who participated in our study.

Table of Contents

Chapter 1 Introduction 1
1.1 Overview 1
1.2 Hypothesis for endometriosis histogenesis 2
1.3 Current diagnosis and treatments 3
1.4 Sub-phenotypes of endometriosis 5
1.5 Surgical staging system for endometriosis 7
1.6 Pain and endometriosis 9
1.7 The Endometrium and menstrual cycle 11
1.8 Hormones and hormone receptor modulation in endometriosis 17
1.9 Genetics of endometriosis 19
1.10 Transcriptome for endometriosis 22
1.11 DNA methylation and detection technologies 25
1.12 DNA methylome of endometriosis 27
1.13 Thesis Aims 30
Chapter 2 General methodology 32
2.1 Overview 32
2.2 Resources 33
2.2.i Endometriosis Oxford (ENDOX) Study 33
2.2.ii Fibroids and Endometriosis Oxford (FENOX) Study 34
2.2.iii International Endometriosis Genome Consortium (IEGC) 34
2.2.iv NIH Epigenetics of Endometriosis Consortium 35
2.3 General data/ sample selection and preparation 35
2.3.i Phenotypic data records 35
2.3.ii Sample selection and preparation for genotyping for IEGC meta-analysis 36
2.3.iii Genotyping QC, imputation, and preliminary GWAS for NIH epigenetic discovery dataset 37
Chapter 3 Sub-phenotype analysis of endometriosis associated loci 38
3.1 Introduction 38
3.2 Materials and methods 42
3.2.i Selection of genetic variants to test for association with endometriosis sub-phenotypes 42
3.2.ii Genetic data from the ENDOX study 44
3.2.iii Sub-phenotypic data from the ENDOX study 46

3.2.iv Association testing for sub-phenotypes of endometriosis in ENDOX phase I & II, and meta-analysis 52
3.2.v Correlation of effect sizes of genetic variants between sub-phenotypes in endometriosis 53
3.2.vi Dyspareunia missing value analysis 54
3.3 Results 55
3.3.i Phenotypic data structure among endometriosis patients 55
3.3.ii Results of 46 endometriosis SNP associations with endometriosis sub-phenotypes 60
3.3.iii Investigation of missingness in dyspareunia severity reporting by participants ... 84
3.4 Discussion 85
3.5 Conclusion 99
Chapter 4 The endometrium transcriptome across menstrual cycle phases 100
4.1 Introduction 100
4.2 Materials and methods 102
4.2.i Sample selection, sample QC, and preparation of RNA-seq libraries 102
4.2.ii Cycle phase determination of eutopic endometrium 103
4.2.iii Differential gene expression of endometrium between cycle phases 105
4.2.iv Multivariate model for cycle phase prediction 106
4.2.v Pathway enrichment analysis on differentially expressed genes between cycle phases 107
4.3 Results 109
4.3.i Comparison of histology, self-reported data, and transcriptomic discrimination of cycle phase in eutopic endometrium 109
4.3.ii Differential gene expression between cycle phases in endometrium 114
4.3.iii Multivariate model for cycle phase prediction 123
4.3.iv Pathway enrichment analysis on differentially expressed genes between cycle phases 125
4.4 Discussion 130
4.5 Conclusion 143
Chapter 5 The endometrial transcriptome in endometriosis 144
5.1 Introduction 144
5.2 Materials and methods 147
5.2.i Sample grouping by cycle phase on RNA-seq data 147
5.2.ii Further sample quality control 148
5.2.iii Differential gene expression in eutopic endometrium between endometriosis cases and clinic controls 148

5.2.iv Comparison of differentially expressed genes by cycle phase, literature reports, and endometrium eQTL map 149
5.2.v Differential expression of genes around GWAS loci for endometriosis sub-phenotypes 151
5.3 Results 151
5.3.i Cluster analysis of samples by proliferative and secretory phase groups and model determination for DE analysis 151
5.3.ii Differential gene expression in eutopic endometrium between endometriosis cases and clinic controls 154
5.3.iii Further investigations of significantly differentially expressed genes 163
5.3.iv Differential expression of genes around GWAS loci associated with endometriosis sub-phenotypes 165
5.4 Discussion 166
5.5 Conclusion 180
Chapter 6 Epigenome-wide association study of endometrium in endometriosis 181
6.1 Introduction 181
6.2 Materials and methods 183
6.2.i Criteria for sample selection and arrays 183
6.2.ii DNA extraction and histological cycle phase dating of eutopic endometrium 184
6.2.iii Quality control of DNA methylation data 184
6.2.iv Genome-wide differential methylation analysis (EWAS) 186
6.2.v EWAS loci in relation to human eutopic endometrium eQTL map 187
6.2.vi EWAS loci near GWAS variants associated with endometriosis sub-phenotypes 187
6.2.vii Visualisation of DNA methylation status of nearby genes of EWAS loci 187
6.3 Results 188
6.3.i Sample data structure before QC 188
6.3.ii Quality control of DNA methylation data and determination of covariates 189
6.3.iii EWAS on human eutopic endometrium 192
6.4 Discussion 201
6.5 Conclusion 211
Chapter 7 Conclusion and future work 212
Supplementary materials 215
A. List of publications and presentations 215
B List of supplementary figures 220
C List of supplementary tables 224

D. Supplementary figures 226
E. Supplementary tables 277
F. Permission of use for copy-righted figures 313
G. Abbreviations 325
H. Software packages (versions) 332
I. Reference 333

List of Figures

Figure 1.1 Illustration of the human menstrual cycle with three phases: menstrual, proliferative and secretory phases. 13
Figure 1.2 Schematic diagram of a flanking region of a CpG island. 25
Figure 3.1 Timeline for discovery of genetic loci associated with endometriosis through GWAS (2010 to 2020). 40
Figure 3.2 Venn diagrams depicting overlap in surgical sub-phenotype diagnoses. 56
Figure 3.3 Venn diagrams for pain sub-phenotypes and their overlap with self-reported infertility. a. Overlap of five pain sub-phenotypes in endometriosis cases in ENDOX. b. Overlap of three main pain sub-phenotypes with self-reports of infertility (trying to conceive ≥ 12 months consecutively) in ENDOX. 57
Figure 3.4 Correlation heat map of 16 major sub-phenotypes based on 211 endometriosis patients with full details on surgical and pain sub-phenotypes, coloured coded by Pearson's correlation coefficient. 59
Figure 3.5 Correlation plot of effect estimates for association of 46 genetic variants with ASRM stage III & IV versus stage I & II sub-phenotype of endometriosis; Pearson's coefficient r^2=0.26. 66
Figure 3.6 Correlation plot of effect estimates of associations for 46 genetic variants with dysmenorrhea onset at/after age of 18 versus before 18 in endometriosis; Pearson's coefficient r^2=0.27. 67
Figure 3.7 Forest plot for effect estimates of the T allele of rs11674184 on 2p25.1, in the intron of *GREB1*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. 70
Figure 3.8 Forest plot for effect estimates of the C allele of rs7924571 on 11p13, downstream of *WT1* & *FSHB*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. 71
Figure 3.9 Forest plot for effect estimates of the G allele of rs71575922 on 6q25.1, in the intron of *SYNE1*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.. 73
Figure 3.10 Forest plot for effect estimates of the A allele of rs1903068 on 4q12, upstream of *KDR*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. 76
Figure 3.11 Forest plot for effect estimates of the T allele of rs10983311 on 9q33.1, in the intron of *ASTN2*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.. 77
Figure 3.12 Forest plot for effect estimates of the C allele of rs12441483 on 15q15.1, downstream of *BMF*, for major surgical, infertility, and pain sub-phenotypes of endometriosis 78
Figure 3.13 Forest plot for effect estimates of the A allele of rs1903068 on 4q12, upstream of *KDR*, for sensitivity testing of dysmenorrhea sub-phenotype associations. 80
Figure 3.14 Forest plot for effect estimates of the T allele of rs10983311 on 9q33.1, in the intron of *ASTN2*, for sensitivity testing of dysmenorrhea, dyspareunia, and acyclical pelvic pain sub-phenotypes associations. 81
Figure 3.15 Forest plot for effect estimates of the C allele of rs12441483 on 15q15.1, downstream of *BMF*, for sensitivity testing of infertility, dysmenorrhea, dyspareunia, and acyclical pelvic pain sub-phenotypes associations. 82
Figure 3.16 Forest plot for effect estimates of the G allele of rs71575922 on 6q25.1, in the intron of *SYNE1*, for sensitivity testing of infertility, dysmenorrhea, dyspareunia, and acyclical pelvic pain sub-phenotypes associations. 83

Figure 4.1 Comparison of histological-confirmed and self-reported cycle phases in all cases and controls. 110
Figure 4.2 Comparison of histological-confirmed and self-reported within sub-secretory phases in all cases and controls. 111
Figure 4.3 Principal component 1 (PC1) vs. principal component 2 (PC2) plot of PCA analysis for samples with histology-confirmed cycle phase (n=80) in three phases. 112
Figure 4.4 Principal component 1 (PC1) vs. principal component 2 (PC2) plot of PCA analysis for samples with histology-confirmed secretory phase only (n=53) in three phases. 113
Figure 4.5 Venn diagram depicting the number of differentially expressed genes between histological-confirmed proliferative and secretory phase samples from edgeR and DESeq2 pipelines. 115
Figure 4.6 Scatter plot of $\log_2$ fold change of differentially expressed genes between proliferative and secretory phases (adj.$p \leq 0.05$) common to edgeR and DESeq2 analyses. 116
Figure 4.7 Scatter plot of ranking by adjusted *p*-value of differentially expressed genes between proliferative and secretory phases common to edgeR and DESeq2 analyses. 116
Figure 4.8 Volcano plot for differentially expressed genes between proliferative and secretory phase samples by edgeR analysis. 117
Figure 4.9 Volcano plot for differentially expressed genes between proliferative and secretory phase samples by DESeq2 analysis. 118
Figure 4.10 Venn diagram for differentially expressed (DE) genes between in-house (proliferative vs. all secretory) RNA-seq analysis, Fung's[86] and Talbi's[82] published DE gene lists (proliferative vs. early-secretory) from microarray studies. 120
Figure 4.11 Scatter plot of $\log_2$ fold change of differentially expressed genes (adj.$p \leq 0.05$) common to in-house edgeR (proliferative vs. all secretory) and Fung et al. (proliferative vs. early-secretory) analyses. 121
Figure 4.12 Volcano plot for differentially expressed genes between mid-secretory and late-secretory samples by edgeR analysis. 122
Figure 4.13 The receiver operating characteristic (ROC) curve from the prediction of proliferative vs. secretory phase based on the 18-gene model. 124
Figure 4.14 Top 10 enriched Ingenuity canonical pathways for differentially expressed genes between proliferative and secretory phase endometrium by IPA. 125
Figure 4.15 Top 10 enriched diseases and bio-function annotations for differentially expressed genes between proliferative and secretory phase endometrium using IPA. 126
Figure 4.16 Complement system pathway highlighting molecules up-regulated in secretory than in proliferative phase based on differential gene expression analysis on endometrium. 129
Figure 5.1 MDS plot for secretory samples (n=91) for subsequent differential expression analysis on endometriosis disease status. 152
Figure 5.2 Volcano plot for differentially expressed genes between endometriosis stage I/II cases (n=18) and clinic controls (n=10) in the proliferative phase of the endometrium. 155
Figure 5.3 Volcano plot for differentially expressed genes between endometriosis stage I/II (n=18) and stage III/IV cases (n=16) in the proliferative phase of the endometrium. 157
Figure 5.4 Volcano plot for differentially expressed genes between endometriosis stage III/IV cases (n=26) and clinic controls (n=31) in the secretory phase of the endometrium. 160

Figure 5.5 Volcano plot for differentially expressed genes between endometriosis stage I/II (n=34) and stage III/IV cases (n=26) in the secretory phase of the endometrium. 162
Figure 6.1 Density plots showing (top panel) raw beta value and (bottom panel) pre-processed normalised beta value of type I green probes in eutopic endometrium samples from discovery (red curves) and replication (blue curves) batches. 189
Figure 6.2 PCA plots of samples before 3rd round of QC (n=159) colour-coded by five cycle phases: PC1 vs PC2 (left), PC1 vs PC3 (middle), PC2 vs PC3 (right). 190
Figure 6.3 PCA plots of samples after QC (n=145) colour-coded by five cycle phases: PC1 vs PC2 (left), PC1 vs PC3 (middle), PC2 vs PC3 (right). 191
Figure 6.4 Manhattan plot of eutopic endometrium EWAS for endometriosis cases (n=100) vs. clinic controls (n=38) by the SVA model. 193
Figure 6.5 Manhattan plot of eutopic endometrium EWAS for endometriosis ASRM stage I/II cases (n=68) vs. clinic controls (n=38) by the SVA model. 195
Figure 6.6 Manhattan plot of eutopic endometrium EWAS for endometriosis ASRM stage III/IV cases (n=32) vs. clinic controls (n=38) by the SVA model. 196
Figure 6.7 Manhattan plot of eutopic endometrium EWAS for endometriosis ASRM stage III/IV cases (n=32) vs. clinic controls (n=38) by the model adjusting only for user-provided covariate (i.e. cycle phase). 197
Figure 6.8 Average DNA methylation level of CpG probes around 38 genes mapped to 14 nominal EWAS loci 200

List of Tables

Table 2.1 Data resources description and utilization in respective experimental chapters..... 32
Table 3.1 List of 46 genetic variants tested for association with sub-phenotypic endometriosis. 43
Table 3.2 Definition of surgical sub-phenotypes for endometriosis cases..... 48
Table 3.3 Definition of infertility and pain sub-phenotypes for endometriosis cases 49
Table 3.4 Loci passing the multiple-testing corrected *p*-value threshold for three to 32 independent sub-phenotype groups with respective multiple-testing correction thresholds... 61
Table 3.5 Summary of significant associations of 16 genetic variants (out of 46 tested) with surgical, infertility, and pain sub-phenotypes of endometriosis. 62
Table 4.1 Assigning self-reported cycle phase to samples based on varying cycle-length... 104
Table 4.2 List of significantly enriched KEGG canonical pathways (Benjamini-Hochberg $p \leq 0.05$, n=12) for genes differentially expressed between proliferative and secretory phase in endometrium, from DAVID analysis. 127
Table 5.1 Publications with differential expression gene lists published on human eutopic endometrium between women with and without endometriosis and their comparison criteria. 146
Table 5.2 Number of endometriosis cases and clinic controls across three sub-secretory phases in the secretory phase group (n=91). 153
Table 5.3 List of protein-coding genes in the same gene family with *ANKRD20A11P*, *CYP4F29P* & *SNX18P13* pseudogenes..... 165
Table 6.1 Number of eutopic endometrium samples with and without endometriosis in each of five cycle phases for Oxford European dataset (n=163)..... 188
Table 6.2 Table of 14 epigenome-wide nominal EWAS loci for endometriosis, ASRM surgical stage I/II, and stage III/IV cases 194
Table 6.3 Gene sets of nearby genes of 14 nominal EWAS loci for endometriosis and its ASRM surgical sub-phenotypes 198

Chapter 1 Introduction

1.1 Overview

Endometriosis is a chronic inflammatory gynaecological condition with an estimated prevalence of 5-10% of women within the reproductive age span[1,2]. The condition is defined by the presence of endometrial-like tissue outside the uterus. However, it has a heterogeneous presentation with a wide range of potential symptoms such as pelvic pain, heavy menstrual bleeding, and reduced fertility, although it can also be present in asymptomatic women. Surgically, endometriotic lesions can be broadly classified into three types: superficial peritoneal, deep, and ovarian cysts arising from endometrial implants, i.e. endometrioma. The gold standard for diagnosis of endometriosis is laparoscopy[3,4], although some endometrioma and deep lesions can be visualised via ultrasonography and magnetic resonance imaging (MRI). There are no validated non-invasive diagnostic biomarkers available for clinical use as yet. Besides surgical removal of disease lesions, hormone-suppressing medication (e.g. oral contraceptives, progestins, GnRH agonists) and painkillers are currently the only treatments to control symptoms of endometriosis[3,4].

Symptoms and the limited management options of endometriosis have significant impact on quality of life[5], both physically and psychologically, of patients and their families. The average direct health care and indirect costs (e.g. productivity loss) were estimated at €9579 per patient-year in 2012, equivalent to the substantial economic burden related to diabetes mellitus and Crohn's disease[6]. Endometriosis is a multifactorial disease with genetic, epigenetic, and environmental factors[7] contributing to its pathogenesis. In this chapter, I summarise the current aetiological understanding of endometriosis from genomic, transcriptomic and epigenetic studies, with a focus on knowledge surrounding sub-phenotypes of disease.

1.2 Hypothesis for endometriosis histogenesis

Endometriosis was first described in 1860 by von Rokitansky[8] as he observed endometrium-like structures in uterine polyps dissected at autopsy. He identified them as 'phyllodes tumour', which later became known as endometriosis. Multiple hypotheses have been proposed over the years for the histogenesis of endometriosis; the most widely accepted is the oldest one — the transplantation theory involving retrograde menstruation — first described by Sampson[9] in 1927. Retrograde menstruation refers to the regurgitation of endometrial cells from the uterus lining (i.e. endometrium) via fallopian tubes into the pelvic cavity. This theory explained the origin of endometriotic lesions in the abdomen for endometriosis patients; however, the regurgitation is observed in 90% of healthy women with cyclically active endometrium and patent tubes[10]. The theory alone is also unable to explain some rare forms of endometriosis with lesions found outside of the pelvis e.g. thoracic cavity[11,12].

The coelomic metaplastic theory proposed by Ferguson et al.[13] in 1969 to explain extra-pelvic endometriosis suggested that serial metaplasia of peritoneal mesothelial cells and ovarian epithelial cells gives rise to endometriotic lesions within the peritoneum and on the ovarian surfaces, respectively. This theory is also extended to propose the differentiation of stem cells to form glandular endometrial-like lesions. These stem cells may originate from bone marrow[14] or the endometrium itself, such as endometrial mesenchymal stem-like cells[15] and progenitor cells (also called endometrial side population cells[16]). They have the ability to differentiate into glandular epithelial, stromal, and endothelial cell lineages[15] upon hormonal stimulation during reproductive age.

Apart from the retrograde menstruation as the main mode in the transplantation theory, Sampson also proposed the transportation of normal endometrial tissues from the uterine cavity (i.e. eutopic endometrium) to other ectopic sites via vein and lymph vessels[17]. This theory was recently supported by the observation of endometrial cells in iliac lymph nodes of majority of patients suffering from endometriotic deep lesions[18].

None of the above theories can explain all forms of endometriosis, according to different anatomical positions or time of occurrence (e.g. rare cases of pre-menarchal and post-menopausal endometriosis). The three hypotheses are not competing, but it is likely that all are involved to a greater or lesser extent in the histogenesis of the disease. Other factors beyond the establishment of lesions are required for its maintenance and growth at ectopic sites; hormonal and immunological factors[19-21] are some to account for the pathogenesis of the condition.

1.3 Current diagnosis and treatments

The current gold standard of diagnosing endometriosis relies on histological confirmation of endometriotic implants obtained during diagnostic laparoscopic surgery[22]. A confirmation of endometriosis can be established when endometrial glands/epithelial cells and stromal cells ($CD10^{+}$)[23] are found[4]; the presence of hemosiderin-laden macrophages is extra evidence[24] but optional. However, negative findings in histology cannot rule out endometriosis as sampling may have missed an area with characteristic cellular presentation. Although surgical visualisation of lesions with histology evidence is the robust diagnostic method, imaging can play a role in assessing the extent of implants, e.g. transvaginal ultrasonography[25] and pelvic MRI[26] particularly for ovarian endometrioma and deep nodular/infiltrating lesions. Peritoneal superficial lesions, however, are not detectable with imaging techniques.

There are no serum reliable biomarkers for diagnostic purposes; a tumour marker primarily for detection of ovarian cancer in high risk women[27,28], cancer antigen (CA) 125, has previously been proposed to correlate with endometriosis, especially ovarian endometrioma and deep implants[29]. However, it is not recommended as a test indicative for endometriosis according to National Institute for Health and Care Excellence (NICE) guidelines[30] as its normal level did not rule out the condition.

Current development of serology biomarkers for endometriosis tend to focus on immunological and angiogenic characteristics of the disease; however, none have reached phase III of biomarker discovery[31] including clinical validation and utility determination. The requirement of an invasive procedure for diagnosis, as well as the lack of biomarkers for early detection contribute to the delayed diagnosis, especially for non-ovarian sub-phenotypes of endometriosis, which remains on average 8 years in the United Kingdom as well as globally[32]. The low awareness of the condition among women and their families, heterogeneous presentation at medical consultation, and hesitancy to refer from primary to secondary, specialist care also add burden to the global delay[33].

First-line treatment for endometriosis typically involves medication, with surgical intervention secondary due to its invasiveness[34]. Usually neither represent a cure. Surgical removal or ablation of implants and adhesions may help restoring a distorted anatomical pelvic structure to improve fertility outcome and pain management; however, around 11-23% of cases undergoing conservative surgery relapse[35]. A study focused on deep endometriosis found that the proportion of women with symptom recurrence, such as pelvic pain (21%), is higher than those with a relapse confirmed by surgery (9%)[36].

Most medical therapies for endometriosis aim at pain relief; the most commonly used, over-the-counter combined oral contraceptives (COCs), is the first-line of medication which aims at lowering overall oestrogen level in circulation while promoting progestogenic environment for atrophy of endometriotic lesions[34]. Second and third line medication includes progestin-only pills (POPs), GnRH agonists, and danazol — a synthetic androgen to prevents the surge of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) during menstrual cycle[37]. As all of these drugs suppress ovarian function through their action on the hypothalamic-pituitary-gonadal (HPG) axis, they are not the preferred long-term treatment for women with infertility complains. Aromatase inhibitors is an emerging treatment class[38] that bypasses the HPG axis, instead targeting abnormal extra-ovarian conversion of androgens to oestrogens in ectopic sites[39,40]; its clinical use, however, remains under development.

1.4 Sub-phenotypes of endometriosis

Endometriosis has long been postulated to be a heterogeneous condition, not only in presentation but also in origin and pathogenesis[41]. Similar to e.g. cancer, endometriosis is likely to comprise of different subtypes that — once identified — would open up the potential of improved diagnostic accuracy and better targeted treatments. Specifically, identifying sub-phenotypes could help elucidate sub-phenotype-specific pathogenic mechanisms. This concept is the core aim of my thesis.

Clinical heterogeneity in disease presentation is typically characterised in three categories: surgical characteristics, self-reported pain profiles, and infertility status. Surgical sub-phenotypes describe lesion types based on their anatomical locations and visual characteristics. The most common staging system used to capture surgical presentation is the American Society for Reproductive Medicine (ASRM) revised classification system (details in Section 1.5). Pelvic pain sub-phenotypes are typically grouped into dysmenorrhea (pain with periods), dyspareunia (pain during/after sexual intercourse), and acyclical pelvic pain, the three most frequently reported pain types associated with endometriosis[42]. In addition, bowel movement or irritable bowel syndrome (IBS)-associated pain, and bladder pain are also associated with endometriosis.

Infertility is the third class of endometriosis-associated sub-phenotypes, typically defined as failing to conceive for more than 12 months with regular sex and without the use of contraceptive methods. Women with endometriosis have a two-fold increased risk of incident infertility when compared to whom without endometriosis[43] after adjustment for age and parity. A study[44] involving a thousand women with symptoms suspected of endometriosis conducted in the United Kingdom showed that 26% identified infertility as one of the symptoms leading to medical attention and subsequent diagnosis.

The importance of recording detailed clinical symptoms and sub-phenotypes through standardised protocols and definitions was highlighted as one of the research priorities by the World Endometriosis Society in 2017[45]. However, it is important to note that the above-mentioned sub-phenotypes for endometriosis are not discrete groups; patients mostly report multiple symptoms, making stratification for investigation of sub-phenotype-specific biological pathways difficult. Moreover, symptoms for endometriosis are mostly non-specific, and common to other pelvic conditions such as uterine fibroids, pelvic inflammatory disease, ovarian cysts, IBS, and interstitial cystitis. As most pain symptoms, e.g. dysmenorrhea and acyclical pelvic pain, in endometriosis are common to other pelvic conditions, studies tended to focus on reporting these conditions in the context of exploring possible comorbidity of endometriosis[46], instead of looking into the potential molecular differences between various pain sub-phenotypes within endometriosis patients. Among all pain sub-phenotypes, IBS-associated pain is probably most commonly investigated[47,48], highlighting a higher odds (five-fold) of endometriosis patients suffering from IBS.

The majority of the literature to date exploring endometriosis pathogenesis have not addressed particular sub-phenotypes of endometriosis. Among the small proportion that did, especially genomic[49,50] and some recent transcriptomic studies[51], focus of analyses was mostly on differences in surgical staging, and less on pain sub-phenotypes. These studies[49,50] showed clear genetic heterogeneity of endometriosis according to different surgical sub-phenotypes (Section 1.9). Also, distinct immunological pathways and profiles, such as polarization of M1/M2 types of macrophages, have been shown to be enriched in eutopic endometrium from different surgical stages of the disease in transcriptomic analysis[51]. However, there is only weak correlation between surgical sub-phenotypes and symptomatic (pain and infertility) presentation[52].

Apart from the above sub-phenotypes that are informed by existing clinical understanding of the disease (i.e. knowledge-based sub-phenotypes), sub-phenotyping by unsupervised machine learning has also been described recently[53] with the use of extensive surgical[54] and patient clinical questionnaires[55] developed by the World Endometriosis Research Foundation – Endometriosis Phenome and Biobanking Harmonisation Project (WERF–EPHect). The study showed that some clusters correlated with clinically defined sub-phenotypes, and some provide new insights, reflecting on patient experiences such as classes of medication (e.g. pain, hormonal) used for treatment of endometriosis[53]. The WERF-EPHect published a set of standardised clinical questionnaires and surgical information sheets in 2014[54-57], aiming at establishment of international consensus on standard operating protocols (SOPs) for detailed data collection and biobanking of samples for endometriosis research. There are currently 32 centres across the globe employing these criteria available in 13 languages for endometriosis research.

1.5 Surgical staging system for endometriosis

There are three main surgical descriptions of endometriotic tissues found in the pelvis: superficial lesions, deep lesions, and ovarian endometrioma. Superficial lesions can be of any size with less than 5mm depth from the peritoneum surface. Deep lesions include deeply infiltrating nodules (typically found in pouch of Douglas) and endometriotic tissues with more than 5mm depth from the peritoneum surface[58,59]. The term endometrioma refers to an endometriotic cyst embedded in the ovary, but not superficial lesions attached to the outside of the ovary (i.e. extra-ovarian lesions).

To systematically describe the various surgical presentations of lesions, there is a widely utilized[60] staging system for endometriosis; it was first suggested by the American Fertility Society (AFS) in 1979 and revised by the same body renamed as American Society for Reproductive Medicine (ASRM) in 1996[61]. The staging system is referred to in the literature as rAFS/ASRM/rASRM staging. In my thesis, I will refer to it as ASRM staging.

In ASRM staging, the extent of disease is classified into four stages according to the number, size, location, and morphology of endometrial implants, as well as associated pelvic adhesions: stage I (minimal), stage II (mild), stage III (moderate), and stage IV (severe). Presence of these features observed during surgery is allocated with a point value to arbitrarily represent the extent of the disease. The sum of these points inform the 'stages' of endometriosis: 1-5 points for stage I, 6-15 points for stage II, 16-40 points for stage III, and >40 points for stage IV. ASRM stage I and II are commonly grouped together (stage I/II), as are stage III and IV (stage III/IV). This is because often studies used retrospectively collected surgical records to define staging, and distinction between stages I vs. II and III vs. IV are less accurate. In addition, previous studies[49,50] showed a clear difference in genetic risk between stage I/II vs. stage III/IV.

It is important to note that the ASRM staging protocol is an arbitrary algorithm with a continuous point system; the four stages are not biologically discrete variables, but rather a crude classification of the extent of the disease. There are some patterns in the coding that related to underlying surgical characteristics. For example, if a woman has an endometrioma of size >1cm, which scores 16 to 20 points alone, she would typically be classified as stage III or IV disease even if she has no evidence of other endometriotic disease. However, being diagnosed with stage III or higher does not necessarily imply the presence of endometrioma, for example, a case can be classified as stage III with multiple deep lesions (if size >3cm, 6 points), and extensive adhesions on ovaries and tubes (enclosing two-thirds, 16 points) with a total score of 22.

Despite the popularity of the ASRM staging protocol and recommendation as first-line classification tool for endometriosis to date[60], the disease stages of this algorithm do not correlate with severity nor location of the disease[34], neither with pain symptomatology[62], infertility[63], nor disease prognosis[64]. It should also not necessarily be interpreted as a progressive staging system, like classifications in the cancer field, as there is a lack of knowledge of the natural history of endometriosis.

Two other classification systems have been recommended to supplement the use of ASRM[60]: the endometriosis fertility index (EFI)[63] and Enzian classification for deep endometriosis (ENZIAN-Score)[65]. The EFI is developed to predict the natural conception rate for endometriosis patients experiencing fertility issues according to both historical and surgical factors, and with this different goal was not intended to replace the ASRM system. Rather, it takes into account the ASRM points for endometriosis lesions, and total ASRM score. The ENZIAN-Score aimed at addressing the distinction between anatomical locations and morphologies for deep endometriotic lesions only; it is used as a supplement to ASRM system mostly in German speaking countries[64].

Although the ASRM classification system has limited clinical or prognostic utility (at best, it is used to describe the complexity of surgical intervention), however, it is still the most widely accepted method to describe endometriosis in the clinical and scientific communities. In the thesis, I will focus on ASRM staging as one of the surgical sub-phenotypes of endometriosis.

1.6 Pain and endometriosis

There is a very high prevalence of pelvic pain among women suffering from endometriosis as it is one of the main clinical presentations; depending on different pain sub-phenotypes, rates vary from 40 to 80%[66], with dysmenorrhea being the most common pain experienced by patients, followed by dyspareunia and acyclical pelvic pain. One study[44] showed that dysmenorrhoea constitute 79% of cases as one of the symptoms leading to seek medical advice and thus receive an endometriosis diagnosis; these rates were 45%, 69%, and 29% for dyspareunia, general pelvic pain, and bowel movement pain respectively.

Apart from the conventional stratification of pain sub-phenotypes according to time of occurrence of pain episodes, classification according to its underlying mechanisms is also recommended as top priority for endometriosis research[45]. However, a limited number of studies has achieved this due to lack of standardization in definitions. Recently, Yong et al.[67] proposed a new pain phenotyping framework coded as types I-IV with an attempt to unify description of peripheral and/or central pain mechanisms.

Peripheral mechanisms combined with central nervous system (CNS) changes ('sensitisation') are part of a major theory to explain endometriosis-associated pain. Altered neurological innervation and immunological/inflammatory environment of the pelvis are two examples of peripheral factors. One study[68] showed that the density of sympathetic nerve fibres, which exhibit pro-inflammatory features through neurotransmitters in the beginning of an inflammation episode[69], is significantly lower in lesion-free peritoneum of endometriosis patients than in healthy controls. It suggested a potential role of unaffected peritoneum in the disease state priming for, or changed by, nearby endometriotic lesions. In contrast, higher density of pro-inflammatory sensory neurons were observed near lesions in peritoneum of patients than in controls. In addition, cytokine profiles in peritoneal fluid of endometriosis patients with pain symptoms were found to be different from those without pain or in healthy controls[70]. Tumour necrosis factor α (TNFα), transforming growth factor β1[71], osteoprotegerin (in TNF receptor superfamily), and nerve growth factors (although with conflicting evidence[72,73]) are some examples of factors that may be involved in activation and sensitisation of peripheral nerves or the CNS[70], thus affecting pain signal transmission.

In addition to peripheral factors, a key study[74] has also demonstrated structural changes in the CNS of endometriosis patients with pain symptoms, such as differences in volume of brain areas responsible for pain perception and their connectivity. As-Sanie et al.[74] showed the reduction of grey matter volume in areas of pain perception from women with pain regardless of endometriosis status, compared to those without pain (i.e. pain-free endometriosis patients and healthy controls); it highlighted that the structural changes may depend on pain status but not on endometriosis status. Furthermore, the team[74] proposed that women with pain-free endometriosis might have a higher threshold in pain induction, as a primer and/or consequence of the pain-free status. Their suggestion is supported by the observation of an increased volume of the periaqueductal grey region, which is part of the descending (i.e. from brain to spinal cord) pain inhibitory pathways, in pain-free patients. In addition, pain-processing brain regions (e.g. anterior insula) of endometriosis patients with pain showed greater connectivity to other brain areas in resting state, with a higher concentration of excitatory neurotransmitters than pain-free patients and healthy controls[75].

Research into endometriosis pain sub-phenotypes by underlying mechanisms is highly complex; most general studies are unable to incorporate brain-related phenotypic data into their phenotyping effort, and therefore are limited to exploring conventional, clinically defined pain sub-phenotypes (i.e. stratified by time and location of pain triggers).

1.7 The Endometrium and menstrual cycle

The endometrium is one of the most dynamic tissues in the human female body, cyclically influenced by and responding to endocrines throughout reproductive years. It is located over the myometrium, the uterine smooth muscle layer, and borders the lumen of the uterus. It consists of two layers: 1) the stratum functionalis (also called spongiosum) with stromal and glandular tissues, facing the lumen; and 2) the stratum basalis, which consists mostly of stromal tissue, and lies between the functionalis and myometrium. The functionalis is the layer most responsive to hormonal changes and exhibits highly differential morphology throughout the menstrual cycle. The basalis is responsible for the regeneration of the superficial functionalis after menstrual breakdown. The discussion of endometrium in this thesis focuses on the superficial functionalis layer, which is the layer that is sampled in studies.

The typical menstrual cycle has a median length of 28 days; clinically, the 'normal range' of cycle length extends from 21-35 days, with most falling between 25 to 30-days[76]. The menstrual cycle consists of three phases (Figure 1.1). Starting at the onset of menses is the menstrual phase characterised by extensive cellular apoptosis and tissue breakdown, usually lasting 3 to 7 days; this phase gives a histological signature[77] of apoptotic bodies, lymphocytic infiltration, and haemorrhage. Next is the proliferative phase, corresponding to the follicular phase in the ovary, characterised by repair and regeneration of the endometrium under the influence of oestrogen, involving mitogenesis and angiogenesis; the length of this phase varies but typically lasts from day 8 to 14. Histological characteristics include mitotic figures within pseudostratified epithelium of small developing circular glands[77].

Ovulation is triggered after a surge of gonadotropic hormones (FSH and LH) mid-cycle and this day is usually referred to as the transition day. At that time, the endometrium is in its interval state, meaning it manifests both features from a pre and post-ovulatory state; typically, mitotic events can still be detected, and glands are relatively round with some subnuclear vacuoles[77] present. Progesterone production by the corpus luteum starts to increase post-ovulation, and dominates the third phase of the cycle including a continuous oestrogenic effect — the secretory phase; this is also called the luteal phase when referring to corresponding ovarian changes. This phase typically lasts 14 days[78], and its duration is relatively constant when compared to proliferative phase. The end of the secretory phase is signalled by the first day of menses of the next cycle. Decidualization is a key process underpinning the secretory phase, preparing the endometrium for embryo implantation around 5 to 10 days after ovulation (the window of implantation)[79-81]. Elongated and coiled-up glands, as well as increased stromal volume (edema[77]) are part of the histological signature of a decidualizing endometrium. Informed by the transcriptomic signature of secretory samples[82], three sub-phases are identified: early (day 15-19), mid (20-22), and late-secretory phases (23-28); the five phases became a standard method for characterising endometrial changes in the transcriptomic and other omics fields.

Cycle length differs between women and between cycles. Recently, a study[76] observing 612k cycles from mobile phone application with estimated day of ovaulation based on users' body temperature suggested that, on average, cycle length decreased by 10% (2.9 days) when age increases from 19 to 45, where most of the reduction is explained by shortening of the proliferative phase, with secretory phase staying relatively constant. The biological machnisms behind this is not clear yet. Irregularity of menstrual cycle can occur in women without apparent gynaecological disorders; however, it is uncertain if there is a higher proportion of endometriosis patients reporting irregular periods[21].

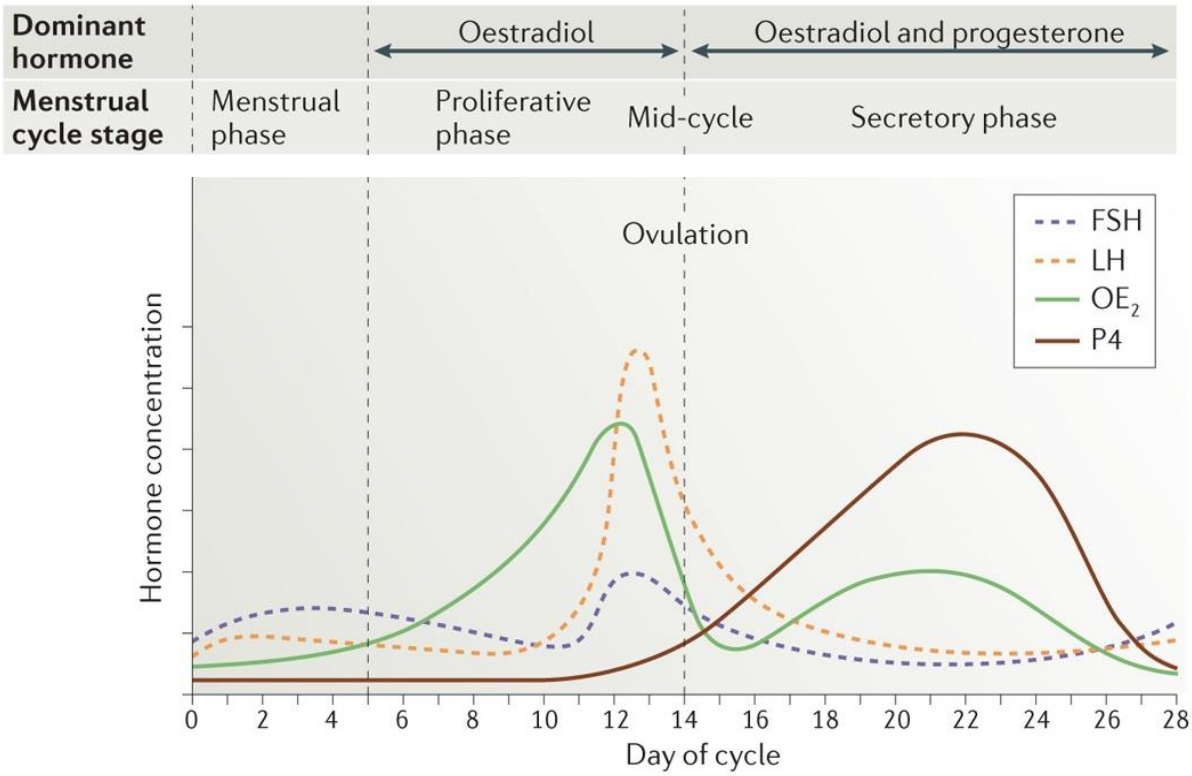


Figure 1.1 Illustration of the human menstrual cycle with three phases: menstrual, proliferative and secretory phases.

Oestradiol dominates proliferative and secretory phases; where progesterone increases post-ovulation. Key: purple dotted line = follicle stimulating hormone (FSH), orange dotted line = luteinizing hormone (LH), green line = oestradiol (OE_2), red line = progesterone (P4).

Adapted by permission granted on 13 Sep 2020 from Springer Nature Customer Service Centre GmbH: Nature Rev Immunol, The role of sex hormones in immune protection of the female reproductive tract, Wira, C., Rodriguez-Garcia, M. & Patel, M., copyrighted (2015) [83] (permission letter in supplementary materials F.i).

As endometrium is a highly dynamic tissue, it is not surprising that the transcriptome[82] and DNA methylome[84] of endometrium differ extensively between menstrual cycle phases. A number of studies have considered transcriptomic differences of endometrium across the menstrual cycle; the three most methodologically robust will be discussed here, and they are also referred to in analyses presented in Chapter 4. Ponnampalam et al.[85] tested gene expression profiles of 43 endometrium samples across different phases of the cycle with two-colour 10.5k cDNA glass slide microarrays and correlated profiles to histological findings. They were the first to report that unsupervised clustering based on the transcriptomic signature of endometrium can give insight into a sample's cycle phase: 1.5k genes were reported to be differentially expressed across the menstrual cycle but comparison between consecutive phases (e.g. proliferative versus secretory) was not provided. Around 30% of those differentially expressed genes had a fold-change >2 across the cycle; however, these results were difficult to interpret without knowledge of comparisons made and therefore direction of effects. The investigators also attempted to identify a panel consisting of expression levels of six genes to predict menstrual phase, but no internal cross-validation or external validation data were presented.

Another study by Talbi et al.[82] looked into the expression profiles of 22 endometrium samples from healthy women with a normal cycle, not using hormones, using a microarray (Affymetrix HG U133 Plus2.0 chip) covering 54.6k genes. Similarly, they conducted unsupervised clustering of samples by various methods (e.g. principal component analysis (PCA), K-means clustering) and revealed four distinct clusters: proliferative, early-secretory, mid-secretory, and late-secretory clusters. They were the first to report differential expression patterns of endometrium between consecutive cycle phases using gene ontology (GO) analysis: 3.0k, 2.9k, and 2.2k genes were found to be differentially expressed between proliferative vs. early-secretory, early-secretory vs. mid-secretory, and mid-secretory vs. late-secretory endometrium, respectively. In the proliferative phase, cell adhesion, cell cycle regulation, and mitosis were highlight biological processes. Metabolism of lipids and steroids, as well as ion and peptide transports were among gene functions characteristic of the early-secretory phase. In mid-secretory phase, genes for immune, wound and stress response, as well as amino acid and polysaccharides synthesis were reported; whereas genes that involved phagocytosis, coagulation, as well as immune and wound response featured in the late-secretory phase. These highlighted biological functions of differentially expressed genes across the cycle relate very well to our traditional understanding of how the menstrual cycle works: the growth (mitosis and cell adhesion) of endometrium during proliferative phase, the immunological shift and mucin production in mid-secretory phase especially important for the window of implantation, and immuno-infiltration and endometrium breakdown in the late-secretory phase.

The last transcriptomic study is from Fung and colleagues[86]; they profiled the endometrial transcriptome by Illumina Human HT-12 v4.0 Expression Beadchip from 123 women of European ancestry, not using hormones, >90% of whom had an endometriosis diagnosis or other gynaecologic pathologies. A total of 12k probes were assessed and mapped post-quality control (QC), with differential expression analyses performed on adjacent phase-pairs similar to Talbi et al.'s[82] approach. They provided detailed annotated gene lists to make comparisons across studies possible (see Chapter 4). The results they presented partly similar to Talbi et al.'s[82]. Between proliferative and early-secretory phases, 1.6k genes showed differential expression (fold change >|1.5|); their Ingenuity pathway analysis (IPA) highlighted oestrogen-dependent cell cycle and signalling pathways. Different from previous findings, the number of genes identified as differentially expressed within sub-secretory phases with a fold-change threshold (>|1.5|) was much lower, with only 118 and 253 genes reported respectively in early-secretory vs. mid-secretory, and mid-secretory vs. late-secretory comparisons. The most significantly enriched pathway among sub-secretory phases was natural killer cell signalling, in line with the featured immune responses highlighted before.

Saare and the team[84] were the first to explore genome-wide DNA methylome changes within endometrium across the cycle. They performed DNA methylation profiling on 55 endometrium samples from women with (56%) and without endometriosis by the Illumina Infinium HumanMethylation 450K BeadChip array. Using unsupervised hierarchical cluster analysis, they showed that samples clustered by menstrual cycle phase rather than by disease state. In addition, almost half of the differentially methylated regions (DMRs) detected between endometriosis status were also detected between cycle phases, illustrating the significance of addressing methylation pattern changes throughout the cycle. The number of DMRs detected between menstrual cycle phases differ significantly by paired comparison; they reported the highest number of DMRs (~200-4,000) whenever menstrual or late-secretory phase samples were involved in the comparison. Minimal DMRs were reported (0-5) for comparisons between proliferative, early-secretory, and mid-secretory phases. Kyoto Encyclopedia of Genes and Genomes (KEGG)[87] analysis of genes harbouring aberrant methylation marks between menstrual and proliferative phases suggested enrichment for cell and extracellular matrix interaction, and cell cycle signalling pathways (i.e. PI3K-Akt).

The above studies help us not only to appreciate the complexity of the endometrium from cyclical physiological observations, but also its molecular mechanisms underpinning continuous changes of the tissue that need to be accounted for in studies aimed at understanding the role of the endometrium in reproductive health and disease.

1.8 Hormones and hormone receptor modulation in endometriosis

Although endometriosis is an oestrogen-dependent condition, the circulating reproductive hormone levels, both oestrogen and progesterone, of endometriosis patients do not appear to be significantly different from whom those without the disease[88,89]. Rather, it is believed that local (paracrine) changes of the hormonal milieu surrounding endometriotic lesions is key[34,88]. Most literature has described hormonal changes in two areas, pro-oestrogenic[90] and anti-progesterone[91] signalling of abnormal tissues, where some of the proposed pathways are common to both epithelial and stromal cells and some are distinct[34,92,93].

Apart from the mitogenic and pro-inflammatory effect of circulating oestrogen on maintenance of endometriotic lesions at ectopic sites, stromal cells in ectopic tissues also facilitate oestradiol synthesis in their microenvironment; this works through prostaglandin E_2 and steroidogenic factor 1 (SF-1) signalling, which leads to up-regulation of aromatase[94] — a catalytic enzyme for transforming androgens into oestrogens[95,96] in the cytochrome P450 superfamily. Normal endometrium, however, does not express aromatase[97] and thus, does not produce oestradiol like endometriotic lesions. On the other hand, epithelial cells in ectopic lesions also play a role in the accumulation of oestradiol locally with altered oestradiol metabolism. The enzyme 17β-hydroxysteroid dehydrogenase type 2 (HSD17B2) is under-expressed[98] in ectopic epithelial cells which hinders normal conversion of oestradiol to its less potent form, i.e. oestrone.

The amount of oestradiol present around disease lesions, the sensing and uptake of the oestrogen and its influence on growth of ectopic cells[99] are not the only important hormonally-related factors. Oestrogen receptors α and β, encoded by oestrogen receptor 1 (*ESR1*) and 2 (*ESR2*) genes, are both present in stromal and epithelial cells[39] of normal eutopic and abnormal ectopic endometrium. Reports of a disrupted ratio of the two within endometriotic lesions[100,101] (with up-regulated ERβ than ERα) has suggested a role in aberrant oestrogen reception; nevertheless, there are a few conflicting reports[88,102] in which receptor subtype is the dominant one in diseased tissue. Xue et al.[101] have proposed that the hypomethylation of *ESR2* promoter initiates the imbalance between ERα and ERβ in endometriotic cells.

In addition to the post-oestrogenic signalling, resistance to progesterone has also been observed in endometriotic tissue. The progesterone milieu of ectopic lesions appears to display insensitivity to progestins due to a down-regulation of progesterone receptors (PRA and PRB) when compared to the normal endometrium. Similar to oestrogen receptors, it is proposed that the ratio of PRA to PRB is altered in diseased tissue[103], although there is some inconsistent evidence from the literature[104] to fully support the claim. Most literature described the reduction of PRB[34,105] in endometriotic lesions, but findings differ among lesion types (e.g. endometrioma, superficial peritoneal)[106]. One study[107] described hypermethylation of a promoter sequence specific to the B isoform of the progesterone receptor gene (*PGR*), which may explain the suppression of PRB observed.

In normal endometrium, HSD17B2 serves both the oestrogen metabolism and downstream of progesterone signalling pathways; progesterone induces HSD17B2 expression in epithelial cells[98], and thus facilitates the conversion of oestradiol to oestrone. In endometriosis, ectopic tissues lack these mechanisms[98], possibly due to the reduction of progesterone receptors, and thus contributing to the specific hormonal microenvironment that promotes lesion development and maintenance.

1.9 Genetics of endometriosis

Twin and family based studies have shown that endometriosis is a heritable condition. Treloar et al.[108] and Saha et al.[109] reported an increased concordance in monozygotic twins compared to dizygotic twins based on self-reported data from a thousand twin pairs in Australia and Sweden, respectively. They suggested that 47%[109]-51%[108] of variance is explained by additive genetic factors underlying endometriosis pathogenesis; while environmental factors account for the remaining proportion for the disease risk.

Several studies have shown a familial tendency for endometriosis[110,111], with higher prevalence[112] of the condition in first-degree relatives of endometriosis sufferers; Moen et al.[113] reported a seven-fold increase in relative risk for first-degree relatives, however, some reported a less dramatic increase in familial risk for the disease[111].

Endometriosis is a common polygenic disease, meaning that multiple common and rare genetic variants underlie the disease. Common variants will each have a small effect size, which when manifested together, could explain part of the disease risk — the common disease-common variant hypothesis (CDCV)[114-116]. Common genetic variants are defined as single nucleotide polymorphisms (SNPs) typically with a minor allele frequency (MAF) of more than 1% in a given population[117]; these variations are widely abundant in human genome and are not necessarily disease-causing. For endometriosis, the contribution of common SNPs to disease variance has been estimated at 26%[118].

Another hypothesis, the common disease-rare variant hypothesis, proposes that many rare variants in the population contribute to the susceptibility of a common disease. These variants could confer large or modest effect sizes. In the context of endometriosis, a family-based linkage analysis by Zondervan et al.[119] suggested one or more loci on chromosome 7p13-15 with moderate to high penetrance for familial cases, but the variants underlying this signal have not yet been identified. Other studies[120,121] suggested linkage to cytoband 10q26, which harbours gene like *CYP2C19* involved in steroidogenesis, although the contribution of variants associated with this gene has not been firmly established[122]. Family and population-based sequencing studies involving these two chromosomal areas are ongoing.

Most genetic studies of endometriosis to date have been association studies, either candidate gene association studies or genome-wide association studies (GWAS). In hypothesis-driven candidate gene studies, most candidates tested for endometriosis targeted hypothesised pathways involved in its potential pathogenesis, e.g. regulation of cell cycle, proliferation, differentiation, and adhesion[34]. Genes related to immune and inflammatory responses (e.g. *IL16*[123,124]), as well as angiogenesis (e.g. *VEGF*[125,126]) have also been tested. Similar to candidate gene association studies as a whole[117,127], results from most of these studies were not replicated[128].

GWAS analysis is a tool proposed and developed for gene mapping since 2000s[129]. It tests differences in frequencies of common single base pair changes across the genome, i.e. SNPs, between cases and controls. GWAS has the power to detect either common variants with small effect sizes or low-frequency variants with moderate effect sizes[130], thus suitable to study common and complex disease like endometriosis. GWAS studies have the benefit of being hypothesis-free, and able to generate new hypotheses agnostically[131,132]. The CDCV hypothesis is the scientific paradigm for GWAS[133,134], and this statistical and analytical framework has also been applied to study endometriosis. The first GWAS for endometriosis was conducted in 2010. In total, ten GWAS have been published in the past ten years suggesting novel loci associated with the disease; details of the timeline for locus discovery can be found in Chapter 3 (Figure 3.1).

The largest meta-analysis of GWAS study to date is ongoing, and based on >60k cases and >700k controls from 24 independent datasets by the International Endometriosis Genomics Consortium (IEGC). It includes samples mainly from European and Japanese ancestries, and has revealed 42 lead SNPs, with 59 index SNPs in total. An interim analysis was released in 2018[135], and the full publication is currently under preparation. Novel SNPs from the interim analysis include variants nearby *PDLIM5* (rs2510770), *IGF1* (rs17727841), *BMF* (rs4923850) loci, etc.

Sub-phenotypic characterisation of genetic signals to date has been limited. Most GWAS efforts[49,135-141] have reported loci associated with endometriosis by ASRM surgical sub-phenotype as this is the most commonly available information. Other sub-phenotypes such as infertility status[135] and pain experiences are less accessible for such analysis due to lack of SOPs for recording these phenotypes and the typical use of retrospective patient records. WERF-EPHect SOPs published in 2014 are helping to address this problem. GWAS[49,50] have consistently suggested that the genetic effects from endometriosis associated loci are larger for ASRM stage III/IV than stage I/II disease. Painter et al.[49] showed that the estimated proportion of disease variation explained by common variants for stage III/IV was twice that of stage I/II disease (34% vs. 15% respectively). As ASRM staging is an arbitrary surgically defined algorithm, not a classification system informed by biological mechanisms; the meaning behind these genetic differences in genetic effects remains unknown. One potential reason for the differences in association pattern with stage I/II vs. III/IV disease could be differential association with other sub-phenotypes, such as more detailed surgical and symptomatic sub-phenotypes — a topic of investigation in this thesis.

A large proportion of the endometriosis risk estimated to relate to common SNPs (26%[118]) remains undiscovered. In the most recent unpublished IEGC meta-analysis, the total variance explained by 42 lead SNPs for stage III/IV endometriosis (4.10%) is higher than that for all cases (3.99%), and stage I/II (2.06%) when considering cases confirmed through medical records. In addition, most of these SNPs locate in intergenic and intronic regions for which the function is still unclear. Further investigation beyond the genomic level, integrating information from transcriptomic and epigenetic signatures in relevant tissues, is needed to further explain the functionality of these GWAS loci.

1.10 Transcriptome for endometriosis

Studying the eutopic endometrium is the key to address endometriosis pathology; apart from the support from Sampson's theory of it being the origin of the disease lesions, the endometrium exhibits varying biological properties from women with or without endometriosis[39], from the ability to adhere[142] and invade to ectopic sites[143], to its aberrant expression of steroid receptors, etc. The transcriptome is dynamic and temporal, as well as being tissue and cell-type specific. Most expression profiling of tissues in endometriosis research utilised the microarray technology[144,145], either with targeted mRNA sequence detection by probes designed specifically for genes of interest in the field, or with commercially available microarrays, which probe sets aimed at covering sequence transcriptome-wide. Microarray involves a targeted quantification of transcripts of interest. A drawback is that if the probe design is poor, it can prompt problems such as cross-reactivity and non-specific hybridization.

A newer, more expensive, alternative to transcriptome profiling is RNA-sequencing (RNA-seq). Currently, only one transcriptomic study of eutopic endometrium has been published for endometriosis using this technology[146]. As RNA-seq does not rely on the use of probes complementary to sequences of interest, it can detect novel transcripts that microarrays may not cover. In RNA-seq, RNA from samples are first reverse transcribed to create stable cDNA libraries. The sequencing of cDNA libraries is competitive in nature, which may results in a bias towards high abundance transcripts that needs to be addressed through normalisation; nevertheless, the technique had been shown to better detect lowly expressed gene transcripts compared to microarrays[147].

RNA-seq has the ability to allow identification of splice isoforms of transcripts (i.e. transcript-based mapping), as well as determination of 5' and 3' boundaries of genes of interest. Unlike microarrays, RNA-seq is able to catalogue all species of transcripts, ranging from messenger RNA (mRNA), long non-coding RNA (lncRNA) to micro RNA (miRNA), and ribosomal RNA (rRNA), although a focus on small vs. large RNA molecules needs to be specified at the time of analysis as different RNA processing protocols are required. In samples that are more prone to degradation, like fresh-frozen or preserved clinical samples, a ribosomal depletion method is employed to improve signals for functionally relevant transcript species (e.g. mRNA & lncRNA) by removing rRNA that constitutes 90% of total extracted RNAs[148].

Studies published to date have focused on two main comparisons in transcriptomic studies related to endometriosis: 1) ectopic lesions to paired (or unpaired) eutopic endometrium from women with endometriosis; and 2) eutopic endometrium from women with vs. without the disease. Due to the huge transcriptomic variation of eutopic endometrium across different cycle phases (Section 1.7), only studies with sample sizes >10 in each comparison group, and having addressed or accounted for cycle phase differences, are summarised here.

In general, and not surprisingly, more transcriptomic differences have been observed between ectopic lesion and eutopic endometrium, than when comparing eutopic endometrium between women with and without the disease. One study[149] involving 18 fertile patients with ASRM stage III-IV endometriosis and ovarian lesions hypothesized that dysregulation of the *CLOCK* (Circadian Locomotor Output Cycles Kaput), *ESR1*, and *MYC* transcription factor networks are involved in the pathogenesis of endometrioma. In addition, upregulation of complement pathway[150] was highlighted in gene expression of epithelial cells in endometrioma. A study by Crispi and colleagues[151] focused on comparison of expression profile of superficial/deep ectopic lesions and paired eutopic endometrium in 27 women showing deregulation of *BMP4*, *GREM1*, and *LEFTY2* in ectopic sites, which are involved in embryonic mammary, mesenchymal, and Müllerian duct development respectively. Their gene expression levels were shown not to be associated with cycle phase. The issue with these comparisons is that we cannot conclude if the observed perturbed expressions are a result of aberrant endometrial cells that promote the establishment of lesions ectopically when they end up in the pelvis through retrograde menstruation, or an adaptation of 'normal' endometrium to ectopic sites in response to e.g. the peritoneum/ovarian environment, or a change of cytokine and hormonal micro-milieu. Comparison of eutopic endometrium profiles between women with and without endometriosis can help to distinguish between these questions.

To date, less than ten studies focused on eutopic endometrial profiles in endometriosis vs. controls; most showing no differential expressed genes after multiple testing correction. Only Burnet et al.[144] suggested lists of differential expressed genes with extremely small sample groups (n=3 to 9). In 2012, Fassbender et al.[152] reported no significantly differentially expressed genes in 31 women with endometriosis vs. 18 controls, irrespective of ASRM stage or cycle phase (menstrual and early secretory phases only). Similarly, Fung et al.[145] and Mortlock et al.'s[146] recent studies using microarray and RNA-seq methods respectively, and with larger sample sizes (27-143 cases vs. 14-89 controls), accounting for menstrual cycle phase, showed no significant differences. Detailed comparison of these studies are further discussed in Chapter 5, Table 5.1.

One microarray study by Aghajanova et al.[153] looked at the transcriptional differences between eutopic endometrium from 19 ASRM stage I/II versus 44 III/IV cases across three cycle phases with multiple-testing correction; they reported upregulation of EGFR (epidermal growth factor receptor) and VCAN (extracellular matrix proteoglycan versican) signalling pathways in stage III/IV disease. The cyclic adenosine monophosphate (cAMP)-regulated genes, progesterone, and thyroid metabolic pathways were also found to be altered between the two.

The analyses to be presented in this thesis (Chapter 5) make a welcome addition to existing knowledge of differential expression patterns in eutopic endometrium from endometriosis in relation to menstrual cycle phases and in particular ASRM sub-phenotypes of disease, in the context of local associated genetic variants (expression quantitative trait loci eQTL SNPs).

1.11 DNA methylation and detection technologies

DNA methylation refers to the addition of a methyl group to the cytosine residues of CG dinucleotides by a group of DNA methyltransferases; where a CpG island is defined by a region, of at least 200 base pairs, with 50% or above of cytosine (C) followed by a guanine (G). The pattern of DNA methylation marks is tissue and time specific; it can help to explore the complex regulation of gene transcriptions, which is one of many levels in epigenetic regulation (e.g. histone modification, chromosomal organisation). Hypermethylation in promoter sequences is associated with suppression of downstream gene transcription; while hypermethylation of flanking sequence around CpG islands (±4kb) is typically associated with higher gene expression[154]. The flanking sequences of CpG islands are called shores and shelves when they are within ±0-2kb and ±2-4kb respectively from the CpG region (Figure 1.2). Sequences that are ±4kb away from CpG islands are referred to as open sea.

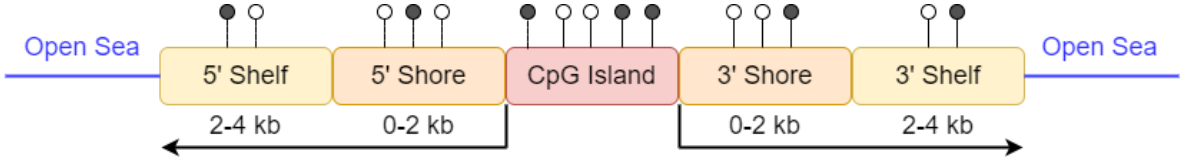


Figure 1.2 Schematic diagram of a flanking region of a CpG island.

A shore is defined as 0-2 kilo-bases (kb) around CpG island; a shelf as 2-4 kb around a CpG island. Graphic elements are not to scale.

Key: blue line = open sea, i.e. DNA sequence out of ±4 kb from a CpG island; black lollipop = methylated cytosines, hollow lollipop = unmethylated cytosines. The density of lollipop is an arbitrary illustration of relative density of methylation sites in each region, with the CpG island harbouring the largest number.

Before the introduction of bisulfite-based amplicon sequencing and high throughput beadchip array technology for DNA methylation detection at genome-wide scale, past experiments were only able to detect candidate methylation/CpG sites, often informed by nearby differential expressed genes, using methylation-specific polymerase chain reaction (PCR).

There are more than 28 million CpG sites in the human genome[155]. The Illumina Infinium Human Methylation EPIC BeadChip system (EPIC array) is the fourth generation array developed and launched by Illumina in 2016, detecting 850k methylation sites across different positions in the genome. Detection sites expanded upon previous array versions i.e. the 27k and 450k arrays, with the same probe technology; most of its additional probes are designed at regulatory and enhancer regions predicted by the Encyclopedia of DNA Elements (ENCODE) and Functional Annotation of the Mammalian Genome (FANTOM5) consortia[156]. As part of the protocol, sodium bisulfite conversion prior to profiling is crucial, which involves deamination of unmethylated C to uracils (U), leaving out the methylated cytosines (5-mC and 5-hmC). The Us are then converted to thymines (T) during DNA amplification.

The Illumina DNA methylation arrays utilised two types of probes for signal detection: Infinium I (type I) and Infinium II (type II). Type I probes are made up of two separate sequences, one of which is complementary to methylated CpG, and the other to the unmethylated CpG. Signals detected by these probes are read by the same colour channel. Type II probes comprise single sequences only, and the colour detection is based on which colour-tagged nucleotide incorporates immediately after the probe complementary to the target sequence: adenine (A) and T are red-tagged; C and G are green-tagged. Both probe types work together to optimise detection of methylation pattern with different CpG content/percentages (e.g. type I probe performs better in measuring methylation at dense CpG regions[157]). A β-value is then calculated as: the intensity of methylation/(intensity of unmethylation + intensity of methylation + 100). It ranges from 0 to 1 and equates to the percentage of methylation. The $\log_2$ transformed ratio, M-value, is calculated as the $\log_2$ [β-value/ (1- β-value)].

However, not all probes work at the same efficiency; a well-known and long-standing problem of some of these probe designs is cross-reactivity[156]. It is recommended that for each generation of arrays, a unique list of probes under-performing due to sequence polymorphisms or non-specific hybridization[156] should be removed in the QC steps.

1.12 DNA methylome of endometriosis

The first study of DNA methylation in endometriosis was performed by Wu et al.[158] in 2005 in which he observed a hypermethylated 5' region in the first intron of *HOXA10* — a down-regulated gene in eutopic endometrium of endometriosis patients — when contrasting eutopic endometrium from ASRM stage III/IV cases (n=6) to controls (n=4).

As mentioned in previous section (1.7), the DNA methylome in eutopic endometrium changes throughout the menstrual cycle[84,159]. Similar to the transcriptomic investigations, studies have compared DNA methylation signals of 1) ectopic lesions to paired (or unpaired) eutopic endometrium from endometriosis patients or from controls; and 2) eutopic endometrium from women with and without the disease; the latter less frequently.

There are a number of candidate CpG studies showing differential methylation patterns for ectopic lesions compared with eutopic endometrium. Xue et al.[160] suggested hypomethylation of CpG islands in the promoter and first exon of *SF-1* in ovarian lesions (n=8) compared with eutopic endometrium from controls (n=8). *SF-1* encodes for transcriptional factor targeting steroidogenic genes for oestrogen synthesis[161], and has been found to be highly expressed in endometrioma. When comparing both ectopic and eutopic endometrium from endometriosis patients (n=60) to eutopic endometrium from clinical controls (n=30, all in secretory phase), Zidan et al.[162] reported hypomethylation in the promoter region of prostaglandin-endoperoxide synthase 2 (*PTGS2*, previously named *COX-2*), which was negatively correlated with increased expression of *PTGS2* in diseased tissue. The encoded protein is an enzyme for biosynthesis of prostaglandin H_2, a lipid mediator for inflammation and vasocontraction[163].

The first use of genome-wide DNA methylation arrays in endometriosis was published in 2014. In a carefully designed study, Dyson and colleagues[164] reported 403 genes mapped to differentially methylated CpGs in stromal cells isolated from endometrioma vs. eutopic endometrium from clinical controls (all in proliferative phase) by the Infinium 450k array. They highlighted significantly differentially methylated signatures not only in the *HOX* gene family, but also in nuclear receptor genes and the *GATA* transcription factors family, especially the hypomethylation of *GATA6*, along with its high expression in endometrioma contributing to progestrone resistance downstream.

A small-scale pilot study using Infinium 450k array by Rahmioglu and her team[159] aimed at assessing DNA methylation variability within tissues considered differentially methylated regions between all types of ectopic endometriotic lesions (n=14) with unpaired eutopic endometrium (n=16) through hierarchical clustering, with adjustment of age and menstrual phase. They reported seven enriched pathways from the PANTHER database[165]: angiogenesis, Alzheimer disease presenilin, GnRH receptor, WNT, cadherin, integrin, and cholecystokinin receptor (CCKR) signalling. In the same study, they also compared endometrioma tissue (n=8) to peritoneal lesions (n=6), which revealed 31 differentially methylated probes. Tissue-based cellular heterogeneity in particular for the endometrioma vs. peritoneal lesion comparison is likely to have influenced the results.

Several very small studies contrasted signatures of eutopic endometrium between endometriosis cases and controls. A small study by Naqvi et al.[166] used the Infinium HumanMethylation 27k array to compare eutopic endometrium from women with (n=7) and without endometriosis (n=6). They reported over a hundred differentially methylated genes, however, without consideration of cycle effects, multiple-testing correction, and with limited sample size. Houshdaran et al.[167] investigated differences in methylation signals between eutopic endometrium from ASRM stage IV patients (n=17) and women without endometriosis (n=16) independent of cycle phases by 27k array; the study suggested three loci: hypomethylation of ribosome production factor 2 homolog (*RPF2*), and hypermethylation of period circadian regulator 1 (*PER1*) and *FAM181A* in the disease group. Within cycle phase, mid-secretory phase (n=6 vs. n=5) showed the most number of differentially methylated loci (125); whereas early-secretory phase (n=7 vs. n=5) showed the least (36 loci). The team also considered the action of steroid hormone on the DNA methylome of cultured stromal fibroblast of eutopic endometrium taken from ASRM stage I or IV patients and compared this to tissues from clinical controls[168]. Their findings suggested that variations observed in the diseased group, i.e. hormone-induced differential methylation, were the consequence of existing DNA methylation marks in the tissue, highlighting the importance of studying these marks in endometriosis pathophysiology.

However, a larger study by Saare et al.[84] using 450k array on eutopic endometrium from endometriosis patients (n=31) and clinical controls (n=28) revealed minimal differences between their methylome. They reported top signals 13kb upstream from *HOXA* gene cluster, and in eight genes after addressing the cycle phase effect: *AHRR*, *AKR1B1*, *APEH*, *CST11*, *ELOVL4*, *CLCF1*, *HBE1* and *NEGR1*. However, when they compared only mid-secretory samples from their dataset (cases n=8 vs. controls n=17), no differentially methylated regions were detected.

Another very small study by Yotova et al.[169] focused on the methylome of cultured stromal cells from eutopic endometrium of cases (n=5) and clinical controls (n=5) assayed on the Illumina 450k array. They reported 68 differentially methylated CpG sites that mapped to 43 genes, which mainly encoded for transcription factors (e.g. *GATA*, *HOXA*, *HAND2*, *NR5A1*, *OSR2*, *TBX3*) and cell signalling components (e.g. *BDNF*, *DAPK1*, *ROR1*, *WNT5A*). This is one of only two studies that are documented in the epigenome-wide association study (EWAS) atlas/ data hub[170] at the time of writing, however, the interpretation of results is problematic due to the small study size and the fact that cycle phase effect was not addressed in analyses. The second study included in the 'EWAS atlas' focused on comparison between methylome of endometrioma lesions vs. eutopic endometrium from endometriosis patients, and from clinical controls, by Wang and colleagues[171]. They reported only one differentially methylated CpG site between eutopic endometrium from endometriosis cases (n=6) and controls (n=6, all in secretory phase) using the Illimina 450k array. More than 12,000 differentially methylated CpG sites mapped to 4,700 genes were observed in the comparison of endometrioma (n=6) vs. paired eutopic endometrium (n=6). GO analysis of these genes featured immune/inflammation, steroid reception, cell adhesion, and mitogen-activated protein kinases networks, however, results are difficult to interpret due to the likely non-endometrial tissue contamination in endometrioma.

In summary, the understanding of the endometriosis-associated DNA methylome so far is based on very small-scale studies; similar to the transcriptome, more differences are observed in the methylome when comparing ectopic lesions with eutopic endometrium than when comparing eutopic endometrium from cases and controls. There is no genome-wide differential methylation study that has considered DNA methylation of eutopic endometrium by sub-phenotype of endometriosis, whilst simultaneously considering the effect of cycle phase. The analyses in the thesis will try to address this research gap.

1.13 Thesis Aims

The central aim of this thesis is to understand the differential genetic, transcriptomic, and DNA methylation signatures of endometriosis clinical sub-phenotypes. Specific objectives/research questions are to:

1. characterise the effect of established genetic variants associated with endometriosis for association with WERF-EPHect compliant surgical and symptomatic sub-phenotypes;
2. consider the transcriptome of eutopic endometrium across the menstrual phases;
3. investigate differential expression between surgical sub-phenotypes of endometriosis and clinical controls; and
4. explore epigenome-wide DNA methylation pattern in eutopic endometrium associated with surgical sub-phenotypes of the disease and clinical controls.

The work addressing these objectives/questions is described in six chapters. Chapter 2 provides an overview of the data resources for all analyses presented in subsequent chapters. In Chapter 3, the association of established endometriosis-associated genetic variants, from the latest GWAS analyses, with a range of WERF-EPHect compliant surgical and symptomatic sub-phenotypes of endometriosis is investigated. In Chapter 4, transcriptomic signatures of 80 human eutopic endometrium samples across menstrual cycle phases are analysed using RNA-seq data, along with the utility of transcriptomic profiling to adjust for cycle phase effect in subsequent omics analyses. In Chapter 5, the RNA-seq transcriptome of eutopic endometrium from 94 endometriosis patients with different surgical sub-phenotypes are compared with endometrium from 41 clinical (endometriosis-free) controls, stratified by menstrual phase, and results are contrasted with the two main studies published to date. In Chapter 6, results are presented from an EWAS analysis of eutopic endometrium for 100 endometriosis cases, and surgical sub-phenotypes, vs. 38 controls, adjusting for cycle phase. Lastly, in Chapter 7, thesis results are summarised and recommended future work outlined that aims to further improve our understanding of molecular signatures of potential sub-phenotypes of endometriosis.

Chapter 2 General methodology

2.1 Overview

In this chapter, I will introduce the data resources for genomic, transcriptomic, and epigenomic analyses presented in the thesis. There are three main resources that supported this work (Table 2.1): 1) ENDOX and FENOX studies conducted by the Oxford Endometriosis CaRe Centre in the Nuffield Department of Women's and Reproductive Health, University of Oxford, 2) The International Endometriosis Genomics Consortium (IEGC), and 3) The National Institute of Health (NIH) Epigenetics of Endometriosis Consortium.

Table 2.1 Data resources description and utilization in respective experimental chapters

Data Sources	Data Types	Data Included in Chapters	Remarks
ENDOX & FENOX (Oxford)	Clinical Samples: a) Phenotypic data b) Genotyping data c) Transcriptomic data (RNA-seq)	a) Chapter 3-6 b) Chapter 3 c) Chapter 4-5	—
International Endometriosis Genomics Consortium (IEGC)	GWAS loci for endometriosis	Chapter 3	ENDOX genotyping contributed to the data source
Epigenetics of Endometriosis Consortium (NIH, US)	a) DNA Methylation Profiling (Illumina EPIC array) b) Genotyping data	Chapter 6	ENDOX & FENOX (phenotypic data & DNA) contributed to the data source

2.2 Resources

2.2.i Endometriosis Oxford (ENDOX) Study

The ENDOX study is a prospective study aimed at understanding pathogenic mechanisms of endometriosis and biomarker discovery through extensive biological sample, surgical, clinical, and patient-based data collection; recruitment ran from 2012 to 2018 at the Oxford Endometriosis CaRe Centre. Ethical approval (REC: 09/H0604/58; IRAS: 8282) was obtained from the Central University Research Ethics Committee, University of Oxford.

The study recruited 799 women, aged between 18 to 50 (pre-menopausal), scheduled to undergo a laparoscopy for symptoms suspected of endometriosis, such as infertility, abnormal bleeding, and pelvic pain. Patients with previous cancer diagnosis were excluded from recruitment. Patients were invited to participate, with full informed consent in place, and asked to complete a detailed questionnaire, and provided blood, urine and saliva samples, prior to surgery. During surgery, an eutopic endometrium sample was taken, as well as endometriosis lesions where present, and subcutaneous abdominal and visceral adipose tissues, as well as a sample of peritoneal fluid. All biological samples were collected according to the World Endometriosis Research Foundation's (WERF) Endometriosis Phenome and Biobanking Harmonisation Project (EPHect) standard operating protocols (SOPs) [54-57]. Patients were diagnosed visually during surgery with endometriosis and/or other gynaecological conditions, such as uterine fibroid and adenomyosis.

Phenotypic data collection consisted of a detailed patient questionnaire including items such as menstrual history, medical history (e.g. precancerous lesion), use of medication, pain experience, infertility, health and physical features, mental health and quality of life; and a detailed questionnaire completed by the surgeon including variables comprising ASRM staging, lesion types, location, and size, as well as adhesion extensiveness. Both included questions based on the WERF-EPHect standardised questionnaires. Information from patient questionnaires was initially stored in hand-written hard copies, and later changed to digital recording on a structured query language (SQL) database powered by research electronic data capture (REDCap)[172,173].

All samples for transcriptomic analyses presented in this thesis were selected from the ENDOX study. For the genomic and epigenomic analyses, majority of the samples were also from ENDOX; minority were from University of Liverpool, University of Edinburgh (details in Section 2.3.ii & 3.2.ii.a), and Oxford FENOX study (Section 2.2.ii) respectively.

Endometriosis cases included in my data analyses were those with surgically confirmed diagnosis regardless of the present of other comorbid conditions. For the purpose of genetic analyses, women with a history of surgically diagnosed endometriosis were included in genetic studies, but they were excluded from transcriptomic or epigenetic analyses. Women without an endometriosis diagnosis, irrespective of the presence of other gynaecological conditions, were grouped as the clinical controls group.

2.2.ii Fibroids and Endometriosis Oxford (FENOX) Study

The FENOX study[174] is an ongoing (2018-2022) prospective study and successor of ENDOX. Ethical approval (REC: 17/SC/0664; IRAS: 227921) was obtained from the Central University Research Ethics Committee, University of Oxford. FENOX was built based on ENDOX, sharing the same study design, criteria for patient recruitment, sample/data collection and storage procedures for endometriosis research, but with an extension to allow the study of uterine fibroids as a separate arm of the project. As a relatively new study at the time of sample analysis, only eight samples from the endometriosis arm of FENOX study were selected, for the epigenetic study (Chapter 6).

2.2.iii International Endometriosis Genome Consortium (IEGC)

The IEGC is a global collaborative effort comprising 25 international teams to conduct the largest meta-analysis of genome-wide association studies (GWAS) for endometriosis to date. Oxford Endometriosis CaRe is one of the leading groups of this consortium, coordinating and contributing to the analysis, as well as sharing various datasets including ENDOX for meta-analysis purposes. The latest meta-analysis involved 61k endometriosis cases and 711k controls (manuscript under preparation, with the preliminary data published in a preprint[135]). The GWAS loci reported by the IEGC form the basis of Chapter 3 (Table 2.1), which explores genetic variants associated with sub-phenotypes of endometriosis in the Oxford dataset.

2.2.iv NIH Epigenetics of Endometriosis Consortium

The NIH Epigenetics of Endometriosis consortium aims to conduct parallel DNA methylation and genetic profiling of eutopic endometrium from surgically-confirmed endometriosis cases and clinical controls in an effort to identify markers for disease. Five centres were involved in the discovery phase: University of California, San Francisco (UCSF), Boston Centre for Endometriosis (BCE) from Harvard Medical School, University of Melbourne, QIMR Berghofer Medical Research Institute, and the Oxford Endometriosis CaRe Centre; this was later expanded to include additional teams from Barcelona and Edinburgh for the replication phase (data not included in this thesis). I was responsible for Oxford sample preparation and was an integral member of the data analysis team. We contributed samples from both ENDOX (n=216) and FENOX (n=8) studies to the discovery (2017-18) and replication (2018-19) phases of the project. DNA methylation and genetic profiles of samples were generated by the Centre for Reproductive Sciences, UCSF using Illumina® Methylation EPIC Beadchips array and Affymetrix Axiom™ Precision Medicine Genotyping array, respectively. I was responsible for quality control of DNA methylation data from Oxford samples, which formed the basis of Chapter 6 to explore epigenome-wide association study (EWAS) for endometriosis.

2.3 General data/ sample selection and preparation

2.3.i Phenotypic data records

With help from Dr Kurtis Garbutt, data scientist from the team, I made a data dictionary template for ENDOX and FENOX patient questionnaires adaptable to our local questionnaires' formats, based on templates from the Boston group, and generated a REDCap database with the support of Medical Science Division (MSD) at Oxford. Data were entered by temporary clerks working for Oxford Endometriosis CaRe Centre. I extracted surgical, infertility, and pain experience data from cases and controls in the study, and performed data cleaning for the sub-phenotypic association (Chapter 3), transcriptomic (Chapter 5), and epigenomic analyses (Chapter 6).

2.3.ii Sample selection and preparation for genotyping for IEGC meta-analysis

Genotyping within the ENDOX study was conducted in two phases. Phase I consisted of 150 endometriosis cases and 77 clinical controls genotyped on the Affymetrix Axiom™ Human UK Biobank array prior to my time in Oxford. For Phase II, we selected 181 buffy coat samples from endometriosis cases for genotyping. An additional 44 whole blood samples collected by the Women's Centre at University of Liverpool, and 26 by Excellence in Pelvic Pain and Endometriosis Care and Treatment (EXPPECT) Centre at University of Edinburgh under their respective local ethics approval and material transfer agreements were also shipped to our laboratory for DNA extraction. I extracted DNA from these samples with the DNeasy® Blood and Tissue Kit (Qiagen, ID: 69504). NanoDrop (NanoDrop Technologies, Inc.) and Qubit fluorometric quantitation (Thermo Fisher Scientific) were conducted for quality control (QC) before plating. The DNA samples were genotyped on the Illumina Infinium® Global Screening array by the Oxford Genomics Centre, Wellcome Centre for Human Genetics, University of Oxford.

Dr Nilufer Rahmioglu led the QC for genotype raw data and imputation of European samples (defined by HapMap v3) to the Haplotype Reference Consortium (HRC) v1.1 reference panel. The data was then contributed to the IEGC preprint published in 2018[135]. I analysed this dataset in parallel with Dr Rahmioglu to train in genotyping QC and imputation pipelines, in preparation for my analysis of the NIH Epigenetic study genetic data.

2.3.iii Genotyping QC, imputation, and preliminary GWAS for NIH epigenetic discovery dataset

Genotyping of DNA samples for the NIH Epigenetic study was conducted on the Affymetrix Axiom™ Precision Medicine Genotyping array. I conducted sample and variant QC for genotyping data from samples contributed by three centres (n=772) for the discovery phase of the NIH Epigenetic study using `PLINK` (flow chart in Supplementary Figure S 2.1). With 505 European samples and 325k variants passing QC, strand-check and pre-phasing (by `SHAPEIT`) were performed before genetic imputation up to the HRC v1.1 reference panel. Post-imputation quality was checked with the `ic` programme developed by Dr William Rayner at Wellcome Centre for Human Genetics. Preliminary genetic association test for endometriosis status was performed with `snptest` (Supplementary Figure S 2.2-3). The post-QC dataset and result of the association testing were shared with the NIH and IEGC consortia.

Chapter 3 Sub-phenotype analysis of endometriosis associated loci

3.1 Introduction

Endometriosis is a complex and heterogeneous condition with both environmental and genetic factors involved in its pathogenic mechanism[175]. Endometriosis often presents with a wide spectrum of clinical symptoms, such as reduced fertility and various types of pelvic pain, that do not correlate well with the varied extent of disease in the peritoneal cavity[176]. To better understand heterogeneity of the disease, there is an urgent need to perform standardised deep-phenotypic characterisation of the cases[45,50] with an aim to develop further disease subtype classifications and diagnostic guidelines.

The aim of this chapter is to characterise the effect of established genetic variants associated with endometriosis to date, on various surgical and symptomatic sub-phenotypes of the disease. The analysis could help to reveal potential similarities and differences in biological pathways involved in sub-phenotypes of this currently coarsely classified disease. The most popular staging algorithm for endometriosis — the American Society for Reproductive Medicine (ASRM) staging (details in Section 1.5) — classify the condition into four stages: stage I (minimal), stage II (mild), stage III (moderate), and stage IV (severe).

Genome-wide association study (GWAS) tests differences in frequencies of common single base pair changes across the genome, i.e. single nucleotide polymorphisms (SNPs), between cases and controls. GWAS has the power to detect either common variants (>1% population allele frequency) with small effect sizes or low-frequency variants with moderate effect sizes[130], thus suitable to study common and complex disease like endometriosis according to the common disease-common variant hypothesis[117].

The GWAS for endometriosis was conducted only in 2010, three years later than the first GWAS for type 2 diabetes mellitus[177] (T2D), a common disease with similar health and economic burden as endometriosis[6]. To date, the ongoing International Endometriosis Genomics Consortium (IEGC) GWAS meta-analysis efforts including 25 studies with 60,674 cases and 701,926 controls have now identified 49 primary loci proposed to be associated with endometriosis (right-most of timeline in Figure 3.1). Despite the relatively few lead SNPs that have been identified in endometriosis compared to other common diseases like T2D, the recent IEGC efforts present a huge leap in the loci discovery for endometriosis since 2017[141]. Novel SNPs from the IEGC update include rs6970537 in cytoband 7p15.2, harbouring *HOXA10* — encoding a transcription factor important for Müllerian duct development and endometrial receptivity[178]; and SNPs nearby *HOXC10* (rs3803042), nerve growth factor (*NGF*, rs12030576), *CD109* (rs4540228) loci, etc.

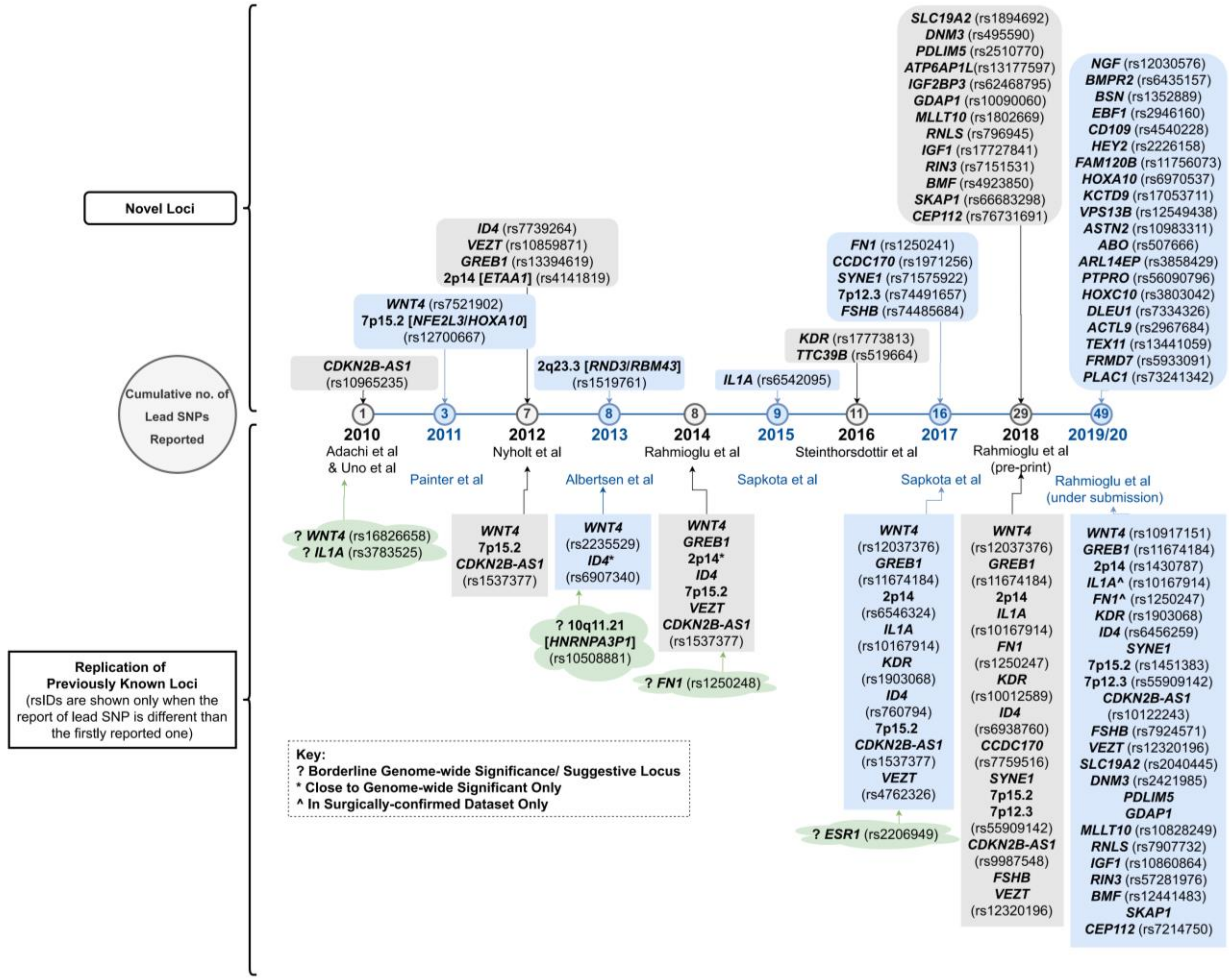


Figure 3.1 Timeline for discovery of genetic loci associated with endometriosis through GWAS (2010 to 2020).

The timeline is denoted with first author of the study, cumulative number of primary loci discovered for endometriosis in circle by that year, and with gray-blue colour alternating between consecutive years. Only loci with reported minor allele frequency (MAF) >1% are shown. The gene names shown are the nearest to the reported loci (rsIDs), or else their cytobands; the upper part of the timeline shows novel loci reported by the study, where the lower part shows validation/ replication of previously reported loci in rectangle boxes, and suggestive loci (with borderline genome-wide significance) in green bubbles. In the lower panel, rsIDs are only shown when the reported lead SNP in that study is different from the firstly reported one in the same locus.

In the past decade, GWAS studies have only focussed on potential differences in effects of endometriosis associated variants between different ASRM surgical stages of the disease. As the staging information is easier to access and standardised across centres than other sub-phenotypes such as infertility status and pain experiences. Stages are typically collapsed into stage I/II and stage III/IV, with stage I/II capturing mostly superficial peritoneal disease, and stage III/IV for more extensive disease including endometrioma. Since the early phase of endometriosis GWAS discovery, various studies have reported that the endometriosis risk collectively explained by discovered loci was higher in stage III/IV than stage I/II endometriosis. This pattern is also true for most loci when considering per-SNP effect sizes. For example, the SNPs rs13394619 and rs11674184 in *GREB1* (Growth Regulating Oestrogen Receptor Binding 1) locus first associated with endometriosis in 2012[136] have been shown in multiple studies to have a higher effect size in stage III/IV (odds ratio (OR): 1.17) than in overall endometriosis (OR: 1.11) [135,136,138,141] and stage I/II diseases (OR: 1.09)[50,135]. Of the 42 lead SNPs in the meta-analysis update by IEGC (publication under preparation), 38 loci showed larger ORs in stage III/IV analysis compared to the all cases analysis. These loci are: *WNT4* (rs10917151), *GREB1* (rs11674184), *ETAA1* (rs1430787), *KDR* (rs1903068), *ID4* (rs6456259), *SYNE1* (rs71575922), 7p15.2 (rs1451383), 7p12.3 (rs55909142), *CDKN2B-AS1* (rs10122243), *FSHB* (rs7924571), *VEZT* (rs12320196), *SLC19A2* (rs2040445), *DNM3* (rs2421985), *BMPR2* (rs6435157), *PDLIM5* (rs2510770), *EBF1* (rs2946160), *CD109* (rs4540228), *HEY2* (rs2226158), *FAM120B* (rs11756073), *HOXA10* (rs6970537), *GDAP1* (rs10090060), *VPS13B* (rs12549438), *ASTN2* (rs10983311), *ABO* (rs507666), *MLLT10* (rs10828249), *RNLS* (rs7907732), *ARL14EP*/*WT1* (rs3858429), *PTPRO* (rs56090796), *HOXC10* (rs3803042), *IGF1* (rs10860864), *DLEU1* (rs7334326), *RIN3* (rs57281976), *BMF* (rs12441483), *SKAP1* (rs66683298), *CEP112* (rs7214750), *ACTL9* (rs2967684), *FRMD7* (rs5933091), and *PLAC1*/*LINC00629* (rs73241342).

Painter et al[49] showed in 2011 that the estimated proportion of disease variation explained by common variants for stage III/IV, overall endometriosis, and stage I/II disease were 34%, 27%, and 15% respectively. However, to date, only 4.10% of disease variance is explained by the 42 SNPs reported in the latest IEGC study for stage III/IV (unpublished results), up from 3.83% by 27 variants reported in 2018 IEGC preprint[135].

The ASRM endometriosis classification system is based on surgical observation. Epidemiological studies have shown a weak correlation between the ASRM surgical stages and pain symptomatology reported by the patients[60]. It is likely that the genetic differences between stage I/II and III/IV indicate an as yet undefined difference in molecular characteristics. It is also not certain whether certain clinical sub-phenotypes, for instance, endometriosis associated with infertility or with a certain pain type, contribute to the higher genetic effect observed in stage III/IV disease.

With the availability of detailed phenotypic data collected by WERF-EPHect-compliant surveys from the ENDOX study (Section 2.2.i), the stratifications of sub-phenotypes beyond ASRM staging is now possible. In this chapter, I aimed to test the association of 46 genome-wide significant endometriosis associated genetic variants with extensive sub-phenotypic data available in the ENDOX dataset. In addition to ASRM surgical stages, I investigated infertility associated endometriosis status, surgical sub-phenotypes (superficial lesion, deep lesion, and endometrioma), and five pain symptomatology: presence and intensity of dysmenorrhea, dyspareunia, acyclical pelvic pain, bowel movement pain, bladder pain, and number of pain symptoms.

3.2 Materials and methods

3.2.i Selection of genetic variants to test for association with endometriosis sub-phenotypes

The 42 lead SNP from the latest IEGC GWAS meta-analysis, along with two independent secondary signals (rs12631337 in *SEMA3F* and rs7740302 in *CCDC170* loci) near the lead SNPs, and two SNPs (rs10167914 near *IL1A* loci and rs1250247 in *FN1* loci) that were genome-wide significant in previous publications[135,141] and borderline significant in the latest IEGC GWAS meta-analysis, were chosen to test for association with endometriosis sub-phenotypes. The list of these 46 genetic variants is given in Table 3.1.

Table 3.1 List of 46 genetic variants tested for association with sub-phenotypic endometriosis.

Key: Chr = Chromosome number; Position based on human Ensemble GRCh37/hg19; RA = Risk allele for overall endometriosis based on IEGC GWAS meta-analysis in 2019; RA Freq = Risk allele frequency in European population based on 1000G Phase3. * Secondary SNPs in IEGC 2019 study; ^ Additional primary SNPs reported in IEGC 2018 preprint[135]; ** no genes within 50 kb.

rsID	Chr	Position	RA	RA Freq	Nearest Gene(s)	Cytoband
rs10917151	1	22422721	A	0.17	*WNT4* / *CDC42*	1p36.12
rs12030576	1	115817221	G	0.65	*NGF*	1p13.2
rs2040445	1	169216412	G	0.97	*SLC19A2*	1q24.2
rs2421985	1	172099136	C	0.49	*DNM3*	1q24.3
rs11674184	2	11721535	T	0.61	*GREB1*	2p25.1
rs1430787	2	67868498	A	0.31	*ETAA1*	2p14
rs10167914^	2	113529183	C	0.32	*IL1A*	2q13
rs6435157	2	203441224	T	0.77	*BMPR2*	2q33.1
rs1250247^	2	216299629	C	0.28	*FN1*	2q35
rs1352889	3	49652148	T	0.17	*BSN*	3p21.31
rs12631337*	3	50198537	A	0.44	*SEMA3F*	3p21.31
rs1903068	4	56008477	A	0.68	*KDR*	4q12
rs2510770	4	95479372	A	0.32	*PDLIM5*	4q22.3
rs2946160	5	157891878	A	0.74	*EBF1*	5q33.3
rs6456259	6	19761718	G	0.16	*ID4*	6p22.3
rs4540228	6	74613060	C	0.66	*CD109*	6q13
rs2226158	6	125995467	G	0.46	*HEY2*	6q22.31
rs7740302*	6	151826249	C	0.19	*CCDC170*	6q25.1
rs71575922	6	152554014	G	0.15	*SYNE1*	6q25.1
rs11756073	6	170363580	A	0.17	*FAM120B*	6q27
rs1451383	7	25891809	G	0.74	—**	7p15.2
rs6970537	7	27217359	G	0.57	*HOXA10*	7p15.2
rs55909142	7	46673774	C	0.59	—**	7p12.3
rs17053711	8	25311269	G	0.73	*KCTD9*	8p21.2
rs10090060	8	75257608	A	0.57	*GDAP1*	8q21.11
rs12549438	8	100150371	T	0.21	*VPS13B*	8q22.2
rs10122243	9	22158924	T	0.41	*CDKN2-BAS1*	9p21.3
rs10983311	9	119470373	T	0.77	*ASTN2*	9q33.1
rs507666	9	136149399	A	0.20	*ABO*	9q34.2
rs10828249	10	21824727	A	0.35	*MLLT10*	10p12.31
rs7907732	10	90182764	T	0.39	*RNLS*	10q23.31
rs3858429	11	30343757	C	0.84	*ARL14EP*	11p14.1
rs7924571	11	32350027	C	0.76	*FSHB* / *WT1*	11p14.1
rs56090796	12	15406228	A	0.33	*PTPRO*	12p12.3
rs3803042	12	54387947	A	0.42	*HOXC10*	12p13.13
rs12320196	12	95645385	G	0.47	*VEZT*	12q22
rs10860864	12	102822862	C	0.82	*IGF1*	12q23.2
rs7334326	13	50967732	C	0.16	*DLEU1*	13q14.2
rs57281976	14	93112974	G	0.77	*RIN3*	14q32.12
rs12441483	15	40353416	C	0.47	*BMF*	15q15.1

rs66683298	17	46277748	C	0.60	*SKAP1*	17q21.32
rs7214750	17	63903119	C	0.93	*CEP112*	17q24.1
rs2967684	19	8786624	A	0.16	*ACTL9*	19p13.2
rs13441059	X	70108889	A	0.36	*TEX11*	Xq13.1
rs5933091	X	131278107	T	0.69	*FRMD7*	Xq26.2
rs73241342	X	133691738	A	0.04	*PLAC1*	Xq26.3

3.2.ii Genetic data from the ENDOX study

3.2.ii.a Overview

The details of the in-house ENDOX study, including ethics approval, study design, patient recruitment criteria, and sample preparation, are described in Chapter 2 General Methodology.

The genotyping effort was conducted in two phases (I and II). The samples genotyped were mainly from the ENDOX collection, supplemented with samples collected in the Women's Centre at the University of Liverpool (Phase I & II), and the Excellence in Pelvic Pain and Endometriosis Care and Treatment (EXPPECT) Centre at University of Edinburgh (Phase II only). The disease status of samples was confirmed surgically through laparoscopy at all sites.

3.2.ii.b ENDOX genotyping phase I

In phase I, samples were prepared and genotyped using Affymetrix Axiom^{TM} Human UK Biobank (UKBB) array (typing 776,520 SNPs) at the Oxford Genomics Centre, Wellcome Centre for Human Genetics, before the start of this DPhil project. Standard sample and variant QC was performed by Dr Nilufer Rahmioglu: samples with genotyping call-rate <0.97, outlying heterozygosity rate (3 standard deviation around mean heterozygosity rate), and ethnic outliers (non-European based clustering of samples with HapMap phase3[179] reference data) were excluded; variants with call rates <97%, Hardy-Weinberg equilibrium (HWE) *p* value of $<1\times10^{-6}$ or MAF <0.01 were filtered out. In total, 150 endometriosis cases (136 from ENDOX, 14 from Liverpool collection) and 77 controls (61 from ENDOX, 16 from Liverpool) passed QC with 655,448 SNPs for further analysis. The control set for Phase I (n=4,930) consisted of in-house clinical controls (symptomatic women, free of endometriosis at laparoscopy, n=77), UKBB clinical (likely symptomatic, n=2,420), and population controls (random female; n=2,433). Clinical controls in the UKBB (accessed in 08/2018) were defined as women who had undergone gynaecological or laparoscopic surgery without the report of endometriosis in hospital record (International Classification of Diseases, 10th revision, Clinical Modification (ICD-10-CM) code: N80) and who did not report an endometriosis diagnosis. The ENDOX and UKBB data was merged revealing 626,941 shared SNPs. After strand check, removal of AT/CG SNPs and SNPs with MAF<0.01 in the merged dataset, 551,798 SNPs remained in the dataset. The combined genotyped data was pre-phased using SHAPEIT v2 and imputed to Haplotype Reference Consortium (HRC) reference panel v1.1 (2016) using Minimac4 on the Michigan imputation server (https://imputationserver.sph.umich.edu/).

3.2.ii.c ENDOX genotyping phase II

In phase II, samples were prepared (see Chapter 2, Section 2.3.ii) and genotyped using the Illumina Infinium® Global Screening array (700,024 SNPs) by the Oxford Genomics Centre. After standard sample QC (for European) and variant QC, 211 endometriosis cases (155 from ENDOX, 30 from Liverpool, 26 from EXPPECT) and two controls (both from ENDOX) passed QC with 497,530 SNPs for further analysis.

The control set for Phase II (n=349) consisted of only two in-house (symptomatic) controls and population controls (all female; n=347) selected from a UK-based intracerebral haemorrhage study conducted in University College London, genotyped using the same array. Quality control of the control set genotype data (n=484,343 SNPs) was conducted separately. Then the genotype data passing QC from cases and controls was merged (n=480,766), and genotype QC was performed again on the merged set (n=461,736) before pre-phasing and imputation up to HRC v1.1.

3.2.iii Sub-phenotypic data from the ENDOX study

3.2.iii.a Utilisation of WERF-EPHect compliant questionnaires in ENDOX

There are two forms we used in ENDOX for collecting sub-phenotypic data from study participants; a surgical form (filled in by surgeons) and a patient baseline (pre-operation, self-report) questionnaire, adapted from the WERF-EPHect standard surgical form (SSF) and clinical questionnaire (standard) (EPQ-S v1.0), respectively. The in-house forms were developed in 2012 when ENDOX started, and were subsequently used as the backbone for the WERF EPHect guidelines finalised in 2014[54,55]. The in-house surgical form is identical to EPHect SSF. The different patient baseline questionnaire versions have slight modifications, mainly extended questions, to suit our centre's research needs and cater for post-operational follow-ups. The patients with genotypes available were recruited throughout 2012 to 2016, and had filled-in three different versions of the in-house questionnaires available at the time: ENDOX Health Questionnaire v1 (from 2012 to Oct 2013), ENDOX Baseline v3 (from Oct 2013 to Apr 2014), and ENDOX Baseline v4.0 (from Apr 2014 to 2016). The question numbers in the latter two are the same (Table 3.3, 4th column). For samples from Liverpool and Edinburgh, only ASRM staging information is available.

3.2.iii.b Endometriosis case & sub-phenotype definitions

Cases for endometriosis in this dataset (Phase I & II combined n_T=361; Table 3.2) were defined based on visual confirmation through laparoscopy (recorded in SSF), as well as histological reference to endometriosis (e.g. presence of hemosiderin-laden macrophages or glandular epithelial and stromal cells resembling endometrial cells) whenever hospital pathology reports were available for biopsies obtained during surgery. Biopsies were usually taken from deep and ovarian lesions for pathology confirmation by surgeons; however small superficial lesions were not typically sent for histology confirmation (Table 3.2, 3rd column). Endometriosis patients with pre-malignant conditions such as cervical intra-epithelial neoplasia (CIN) 2/3, and self-reported pre-malignant history were excluded from the study (n=5). Suspected cases with endometriotic-like lesion, where subsquent histology suggested endosalpingiosis were removed (n=2). Endometriosis cases with lesions only present on Caesarean-section scars were also excluded (n=1).

Sub-phenotypes among endometriosis cases were explored in three main categories: 1) surgical (Table 3.2), 2) infertility, and 3) pain (Table 3.3). Among each category, several sub-phenotypes were investigated. Taking the pain category as an example: dysmenorrhea is one of five pain sub-phenotypes (Table 3.3, 1st column) investigated, for which there are six correlated variables which were subsequently tested characterising the nature and intensity of dysmenorrhea (sensitivity analysis). For pain variables defined by the numerical rating scale (NRS) (Table 3.3, 3rd column), the cut-off for mild, moderate, and severe pain were set at 1-3, 4-6, and 7-10 respectively[180]. An additional variable explored was the total number of pain types experienced by patients — a proxy for hyperalgesia in the pelvic region. The overlap between key sub-phenotypes were visualised by Venn diagrams using online tools (Lucidchart and http://bioinformatics.psb.ugent.be/webtools/Venn/). Correlation between 16 main sub-phenotypes reported by endometriosis patients (with both surgical and pain details reported, n=211 out of 361) was calculated by Pearson's correlation and visualized in a heat map. Similar analysis based solely on patients reporting age of onset for dysmenorrhea (n=162) was performed. The samples size of each sub-phenotype for individual genotyping phases (n_1, n_2) and combined (n_T) are shown in tables below.

Table 3.2 Definition of surgical sub-phenotypes for endometriosis cases

Descriptive definitions for surgical variables; responses with reference to section and item numbers in WERF-EPHect Standard Surgical Form (SSF v3.0); and sample size (n_1, n_2) for ENDOX Phase I, II for sub-phenotypic genetic association testing, and combined (n_T) for meta-analysis.

Key: CIN=Cervical Intra-epithelial Neoplasia; CGIN= Cervical Glandular Intra-epithelial Neoplasia; ASRM Stage = American Society for Reproductive Medicine/ revised American Fertility Society

Endometriosis Surgical sub-phenotype	**Sub-phenotype Definition**	**Consider Hospital Pathology Report**	**Section No. in WERF-EPHect Standard Surgical Form (SSF)**	**Items on WERF-EPHect SSF v3.0, Dec 2014** *Responses*	**ENDOX Genotyping Sample size for Analysis**		
					Phase 1 (n_1)	**Phase 2 (n_2)**	**Meta-analysis Total (n_T)**
All Endometriosis	All endometriosis cases except for pre-cancer (e.g. CIN2, CIN3, CGIN), known cancer history, C-section endometriosis only, histologically-confirmed endosalpingiosis	Yes	III	Previous surgical diagnosis of endometriosis *Yes*	150	211	361
			VIII	Visual diagnosis of endometriosis *Yes*; ASRM scoring table			
ASRM Stage III & IV	Score 16+ points[61]	No	VIII	ASRM scoring table – total score	79	101	180
ASRM stage I & II	Score 1 to 15 points				70	109	179
Superficial Lesion	Peritoneal superficial lesions – any size	No	VIII	ASRM scoring table – peritoneum superficial <1, 1-3, >3 cm	78	108	186
Deep Lesion	Peritoneal deep lesions (>5mm depth from the peritoneum[58,59]) – any size; deeply infiltrative nodules	Yes	VIII	ASRM scoring table – peritoneum deep <1, 1-3, >3 cm; Deeply infiltrative endometriosis *Yes*	54	55	109
			IX	Location of endometriosis: bladder/colon – deep infiltrating			
			XI	Endometriotic nodule *Yes*			
Endometrioma	Ovarian deep lesions – any size; endometrioma	Yes	VIII	ASRM scoring table – ovary deep <1, 1-3, >3 cm	54	35	89
			X	Endometrioma *Yes*			

Table 3.3 Definition of infertility and pain sub-phenotypes for endometriosis cases

Descriptive definitions for infertility and pain variables; responses with reference to question numbers in ENDOX Health (v1), ENDOX Baseline (v3&4.0), and WERF-EPHect Clinical Standard Questionnaire (EPQ-S v1.0); and sample size (n_1, n_2) for ENDOX Phase I, II for sub-phenotypic genetic association testing, and combined (n_T) for meta-analysis. Differences in wordings of questions in ENDOX Health v1 from the shown ENDOX Baseline v 3 & 4.0 questions are noted in square brackets.

Endometriosis with Sub-phenotypes	**Variables (for Sensitivity Testing)**	**Sub-phenotypes Definition**	**Question No. in Questionnaire**			**Questions on Clinical Form ENDOX Baseline v4.0 Apr 2014** *Responses*	**ENDOX Genotyping Sample size for Sub-phenotype**		
			ENDOX		**WERF-EPHect**				
			Baseline v3 & 4.0 Oct 2013 Apr 2014	**Health v1 Jan 2013**	**Clinical Standard (EPQ-S) v1.0 Mar 2014**		**Phase 1 (n_1)**	**Phase 2 (n_2)**	**Meta-analysis Total (n_T)**
Infertility	Infertility	Tried to conceive ≥ 12 months	D3.1	28	B2.1	What was the longest amount of time that you tried, whether or not you actually got pregnant?	39	35	74
	Subfertility	Tried to conceive ≥ 6 months; infertility as a reason for surgery; nurses noted as infertility cases	B1	12	—	What symptom led to your planned surgery? *Infertility*	83	57	140
			D3	—	B2	Have you ever tried to get pregnant for more than 6 months in a row without succeeding?			
Dysmenorrhea	All	Reports of dysmenorrhea (moderate to severe cramps), regardless of temporality	E2	31 [last 3 months]	C2	Have you ever had pain during your periods? *Moderate cramps, severe cramps*	60	94	213
	Severe; recent	Pain Numeric Rating Scale (NRS) 7-10[180]; recent 1-3 cycles	E8	38 [last 3 months]	C8	Please rate how severe your period pain was at its worst during your last period using a scale from 0 to 10 where 0=no pain and 10=worst imaginable pain.	64	69	133
	Moderate to Severe; recent	Pain NRS 4-10; recent 1-3 cycles					79	88	167
	Mild to Moderate; recent	Pain NRS 1-6; recent 1-3 cycles					20	27	47
	Age of onset < 18 years old	Reports of dysmenorrhea, with onset age under 18 years old, regardless of severity	B1	12	—	What symptom led to your planned surgery? Symptom started at age (years) for painful periods	40	79	119
			E2.1	—	C2.1	At what age did you start having period pain?			
	Age of onset ≥ 18 years old	Reports of dysmenorrhea, with onset age at/over 18 years old, regardless of severity	Same as above				23	26	49

Dyspareunia	All	All reports of pain during OR in 24 hours of vaginal sexual intercourse/penetration, regardless of severity and temporality	E16	40 [last 3 months]	C16	Have you ever had pain during intercourse or in the 24 hours following vaginal sexual intercourse/penetration? *NRS>=1*	60	94	154
	Severe; recent	Pain NRS 7-10; during vaginal intercourse/penetration; last 1-3 times	E20	46 [last 3 months]	C20	Please rate how severe your pain was at its worst during the last time you had vaginal intercourse/penetration using a scale from 0 to 10.	18	32	50
	Moderate to Severe; recent	Pain NRS 4-10; during vaginal intercourse/penetration; last 1-3 times					42	61	103
	Mild to Moderate; recent	Pain NRS 1-6; during vaginal intercourse/penetration; last 1-3 times					31	42	73
	Severe; at worst in lifetime	Pain NRS 7-10; during vaginal intercourse/penetration; at its worst in lifetime	E27	—	C27	The following question is about the time in your life when your pain with vaginal intercourse/penetration was at its worst. Please rate how severe your pain with vaginal intercourse/penetration was when it was at its worst using a scale from 0 to 10.	23	54	77
	Mild to Moderate; at worst in lifetime	Pain NRS 1-6; during vaginal intercourse/penetration; at its worst in lifetime					16	27	43
Acyclical Pelvic Pain	All	All reports of pelvic pain not related to periods, intercourse, surgery & pregnancy; regardless of types of pain (cramping, shooting, stabbing), severity & temporality	E29	48 [last 3 months]	C29	Have you ever experienced pelvic pain in general? Do not count: pain caused by menstrual cramps, intercourse, surgery, pregnancy, childbirth, sports-related or other injury, food poisoning, or stomach flu. *Yes*	71	96	167
	Severe; recent	Pain NRS 7-10; last 3 months	E34	56	C33	Please rate how severe your pelvic pain was at its worst in the last 3 months using a scale from 0 to 10.	53	71	124
	Moderate to Severe; recent	Pain NRS 4-10; last 3 months					68	83	151
	Mild to Moderate; recent	Pain NRS 1-6; last 3 months					18	18	36
	Severe; at worst in lifetime	Pain NRS 7-10; at its worst in lifetime	E44	55	C39	The following question is about the time in your life when your pelvic/lower abdominal pain was at its worst. Please rate how severe your pelvic/lower abdominal pain was when it was at its worst using a scale from 0 to 10.	33	63	96
	Mild to Moderate; at worst in lifetime	Pain NRS 1-6; at its worst in lifetime					5	18	23

Bowel Movement/ Irritable Bowel Syndrome (IBS) Pain	—	Pain with bowel movement OR Pain fulfilling IBS Rome IV criteria[181]	E48	12 [symptom lead to surgery]	C43	Please indicate whether you have (had) the following other types of pain in the last 12 months: Pain with bowel movement *Yes, in last month* or *Yes, more than 1 month ago*	42	64	106
						According to revised 2016 Rome IV IBS criteria, fulfilling ALL the following three conditions:			
			E29	48	C29	a) Have you ever experience pelvic pain in general? *Yes* & When did you last have this pain? *≤ 3 months*			
			E31	51	C31	b) Approximately how long in total did you have this pain for in the last 3 months? *≥one day per week*			
			E40	58	C37	c) Two out of three of the following answers as ≥ *Often*: 1) How often did your pelvic pain in the last 3 months get better/worse after you had a bowel movement? 2) How often did you have more/less frequent bowel movements when the pain started? 3) How often were your stools looser/harder when the pain started?			
Bladder Pain	—	All reports of bladder pain OR pain during urination; regardless of severity and temporality	E42	62	—	Please rate on the following scale your worst bladder pain in the last 7 days, where 0=no pain and 10=worst imaginable pain. *NRS>= 1*	33	74	107
			E48	12 [symptom lead to surgery]	C43	Please indicate whether you have (had) the following other types of pain in the last 12 months: Pain with urination *Yes, in last month* or *Yes, more than 1 month ago*			
Report of Above Pain Types (max. 5)	Pain Types >=4	Report of more than or equal to 4 pain types		—		i.e. dysmenorrhea (moderate to severe), dyspareunia, acyclical pelvic pain, bowel movement/IBS pain, and bladder pain as described above	42	67	109
	Pain Types >=3	Report of more than or equal to 3 pain types					62	93	155
	Pain Types >=2	Report of more than or equal to 2 pain types					82	116	198

3.2.iv Association testing for sub-phenotypes of endometriosis in ENDOX phase I & II, and meta-analysis

The association of 46 SNPs (3.2.i) with all endometriosis cases was tested against the control set within each genotyping phase using `snptest`[182]. Expected genotype counts of the input `(-method expected)` were used for the Cochran-Armitage trend test, with the additive model as the assumption for effect size (beta) estimates `(-frequentist 1)`. No covariates were adjusted for in the analyses of phase I and II data, of which the inflation factor lambda (λ) were at 1.02 and 1.01 receptivity.

Preliminary association testing of SNPs for endometriosis cases against three tiers of controls: a) in-house clinical controls only (n=77), b) both in-house and biobank clinical controls (n=2,497), and c) all clinical and population controls (n=4,930) in Phase I was performed. Owning to limited number of controls in a) and b), only statistics from genetic associations with the 30 sub-phenotypes of endometriosis listed in section 3.2.iii.b against the largest control set c) were reported.

As most sub-phenotype characteristics were not mutually exclusive in individuals (e.g. a woman could be diagnosed with both an endometrioma and superficial peritoneal disease), formal statistical testing of effect size differences between sub-phenotypes was not possible due to lack of independence between groups. Only the differential effect of SNPs on mutually exclusive groups ASRM stage III/IV vs stage I/II was tested statistically through frequentist association test.

The direction of test statistics from both genotyping phases were flipped to designated risk alleles' reported in IEGC GWAS meta-analysis (paper under submission); data were then meta-analysed using `METAL`[183] (v Mar 2011) based on their standard errors `(SCHEME STDERR)`, correcting using genomic control `(GENOMICCONTROL ON)`, and allowing estimation for heterogeneity (I^2 index with P_{Het}) between phases `(ANALYZE HETEROGENEITY)`. The final meta-analysed, locus-specific association statistics for all endometriosis cases, sub-phenotypes of endometriosis (representative variables), and ASRM stage III/IV versus I/II were visualised in forest plots with `forestplot` in R.

The focus of analysis interpretation rested with the main sub-phenotype categories. Direct comparison of statistics, such as effect sizes, from different sub-phenotypes was not statistically feasible, owing to the fact that most sub-phenotypes were not independent variables (i.e. non-mutually exclusive sub-phenotypes). Detailed results of variables under the same sub-phenotype (e.g. trying to conceive ≥ 6 and ≥ 12 months within the infertility sub-phenotype) served as sensitivity tests to explore consistency in results (e.g. flag up for outlying signals arising by chance in this relatively small dataset). Relevant results for sensitivity testing were shown in separate forest plots and tables.

Statistics from meta-analysis, including OR, *p* value, and heterogeneity index (I^2) were reported. Significance of tests should be corrected with multiple-testing; most conservatively, with 46 loci tested for 32 different sub-phenotypic definitions (Table 3.4), only $p < 0.05/(46x32) = 3.40x10^{-5}$ should be considered passing the Bonferroni's adjustment threshold. However, those 32 sub-phenotypes are not mutually exclusive variables; with possible correlations between certain sub-phenotypes (e.g. ASRM stage III/IV & endometrioma, or, infertility & subfertility), it is difficult to assess the number of independent variables tested in the association analysis. If taking only main sub-phenotypes into account, 12 of them (all items in the first column of Table 3.2 & Table 3.3, except for cumulative pain types reported), *p* value less than $0.05/(46x12) = 9.06x10^{-5}$ could be a less stringent threshold to consider. The more relaxed multiple correction would be taking only three main sub-phenotypic categories into account, i.e. surgical, infertility, and pain. The significant threshold would then be $p < 0.05/(46x3) = 3.62x10^{-4}$. Table 3.4 shows the number of test statistics passing different tiers of multiple-correction thresholds as outlined above.

3.2.v Correlation of effect sizes of genetic variants between sub-phenotypes in endometriosis

As mentioned, formal statistical comparison of effect sizes between most sub-phenotype groups was not possible due to non-independence of observations. Instead, differential effect estimates of genetic variant associations with one sub-phenotype versus another were plotted to visualise results. To maintain clarity in the correlation plots, 95% confidence intervals (95% CI) of data points were not shown; and only odd ratios ranging from 0.5 to 2.0 were plotted, on a fixed scale, to allow consistent comparison. Pearson's correlation coefficient (r^2) between effects sizes of sub-phenotype paired comparisons were also determined.

3.2.vi Dyspareunia missing value analysis

Sensitivity analysis for dyspareunia indicated inconsistent variant effect directions. Questions in the clinical questionnaires capturing self-reported dyspareunia data were explored, and analysis of missingness patterns was conducted among participants who experienced dyspareunia. Differential missingness of pain NRS for this sub-phenotype was tested against concurrent reports of severity in dysmenorrhea, acyclical pelvic pain, number of pain types reported, and age group by the Yates's chi-squared test.

3.3 Results

3.3.i Phenotypic data structure among endometriosis patients

3.3.i.a Overlap between sub-phenotypes

In 24 out of 32 sub-phenotypes included, sample size was larger than one-fifth of all cases (n=72). The other 6 sub-phenotypes (Table 3.3, n_T= 23-50 < 72) had more limited case sample sizes. Four of these reflected mild to moderate pain types; this reflects the left-skewed pattern for pain NRS reports among women with endometriosis recruited in our study. The other two sub-phenotypes with limited sample numbers are 'age of dysmenorrhea onset at/over 18 years old' (showing that most women first had period pain in their teens) and 'recent severe dyspareunia'. Missing value analysis for the latter was performed; details are in Section 3.3.iii. Among women with surgical subtype information (n=251, Figure 3.2), 9.2% (n=23) vs. 35% (n=87) reported all three vs. two surgical diagnoses, respectively. The largest group of women (42%, n=106) with only one type of lesion found was those with superficial lesions. Most women who presented with endometrioma, 89% (n=79/89) had another form of lesion; as did those with deep lesions (77%, n=84/109), whereas this percentage was lower for those with superficial lesions (43%, n=80/186).

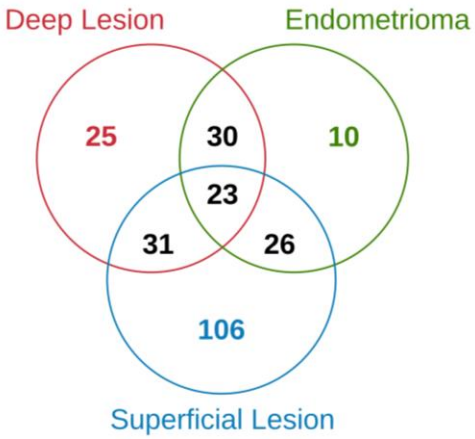
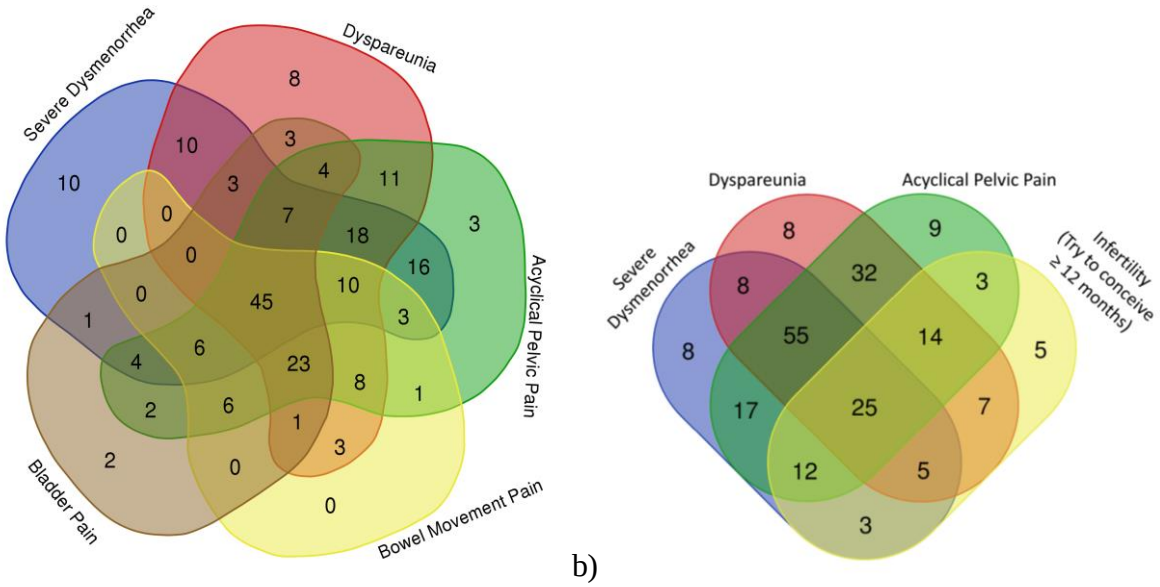


Figure 3.2 Venn diagrams depicting overlap in surgical sub-phenotype diagnoses.
Numbers shown are based on the total from both genotyping phase I & II (n=251) with surgical sub-phenotype identified and recorded

Next, overlap between pain symptomatology was investigated. Given that mild dysmenorrhea is a very common phenomenon (here, 98% (n=213/217) of endometriosis patients reported certain degree of dysmenorrhea), subsequent analysis focussed on only severe dysmenorrhea cases (defined as pain NRS 7-10). Among all endometriosis patients, 58% (n=208, Figure 3.3a) reported suffering from at least one of the five main pain types: severe dysmenorrhea, dyspareunia, acyclical pelvic pain, bowel movement/ irritable bowel syndrome (IBS) pain, or bladder pain. As many as one fifth (22%, n=45) reported experiencing all five pain types, with 89% (n=185) women experiencing at least two pain types. It is worth noting that nobody reported only bowel movement pain, and a limited number of cases had only bladder pain (n=2, 1.9%) or only acyclical pelvic pain (n=3, 1.8%). With all possible 15 combinations of five pain sub-phenotypes, the second largest group (11%, n=23) was the co-occurrence of all pain types except for severe dysmenorrhea. All of these women reported dysmenorrhea as a symptom: three with mild, seven with moderate dysmenorrhea, and 13 with missing pain NRS scores. The third largest group (8.7%, n=18) had co-occurring severe dysmenorrhea, dyspareunia, and acyclical pelvic pain, considered as the classic pain 'triad' associated with endometriosis. To visualise the overlap between different pain types and self-reported infertility (Figure 3.3b), only the pain triad is shown for clarity. Most women suffering all three pain types had no self-reported infertility (n=55/80), suggesting that pain vs. infertility groups are of interest to explore further in terms of differential genetic effects.



a) b)

Figure 3.3 Venn diagrams for pain sub-phenotypes and their overlap with self-reported infertility. a. Overlap of five pain sub-phenotypes in endometriosis cases in ENDOX. b. Overlap of three main pain sub-phenotypes with self-reports of infertility (trying to conceive ≥ 12 months consecutively) in ENDOX.

Numbers shown are based on the total from both genotyping phase I & II (n=208) with self-reported pain types.

3.3.i.b Correlation between sub-phenotypes

Among 16 major sub-phenotypes, apart from mutually exclusive sub-phenotypes, the most negatively correlated pair was ASRM stage I/II with endometrioma (r = -0.92, Figure 3.4), followed by superficial lesion with deep lesion (r = -0.49); the first observation was in line with the scoring algorithm of the ASRM staging protocol. The most positively correlated pair, apart from cumulative number of pain types reported, was ASRM stage III/IV with endometrioma (r = +0.92), followed by bowel movement/IBS pain with bladder pain (r = +0.54). Magnitude of Pearson's correlation coefficients across main categories of sub-phenotypes, i.e. surgical, infertility, and pain, were low. One of the pairs with highest absolute correlation, for example, between deep lesion with bowel movement/IBS pain was only with r = +0.21. There were no correlation between infertility and surgical/ASRM sub-phenotype. Similarly, no correlation were observed for general pain-based symptoms with surgical/ASRM sub-phenotype. However, age of dysmenorrhea onset ≥18 showed mild correlation with endometrioma and stage III/IV disease with r = +0.21 and +0.24 respectively; whereas, age of dysmenorrhea onset <18 was mildly correlated to bowel movement/IBS pain (r = +0.22) and bladder pain (r = +0.26). When limiting the dataset, based solely on patients with detailed surgical information and reported age of onset for dysmenorrhea, to control for phenotypic record missingness, a coherent pattern of correlations was observed between the larger (n=211, Figure 3.4) and smaller matrices (n=162, Supplementary Figure S 3.1).

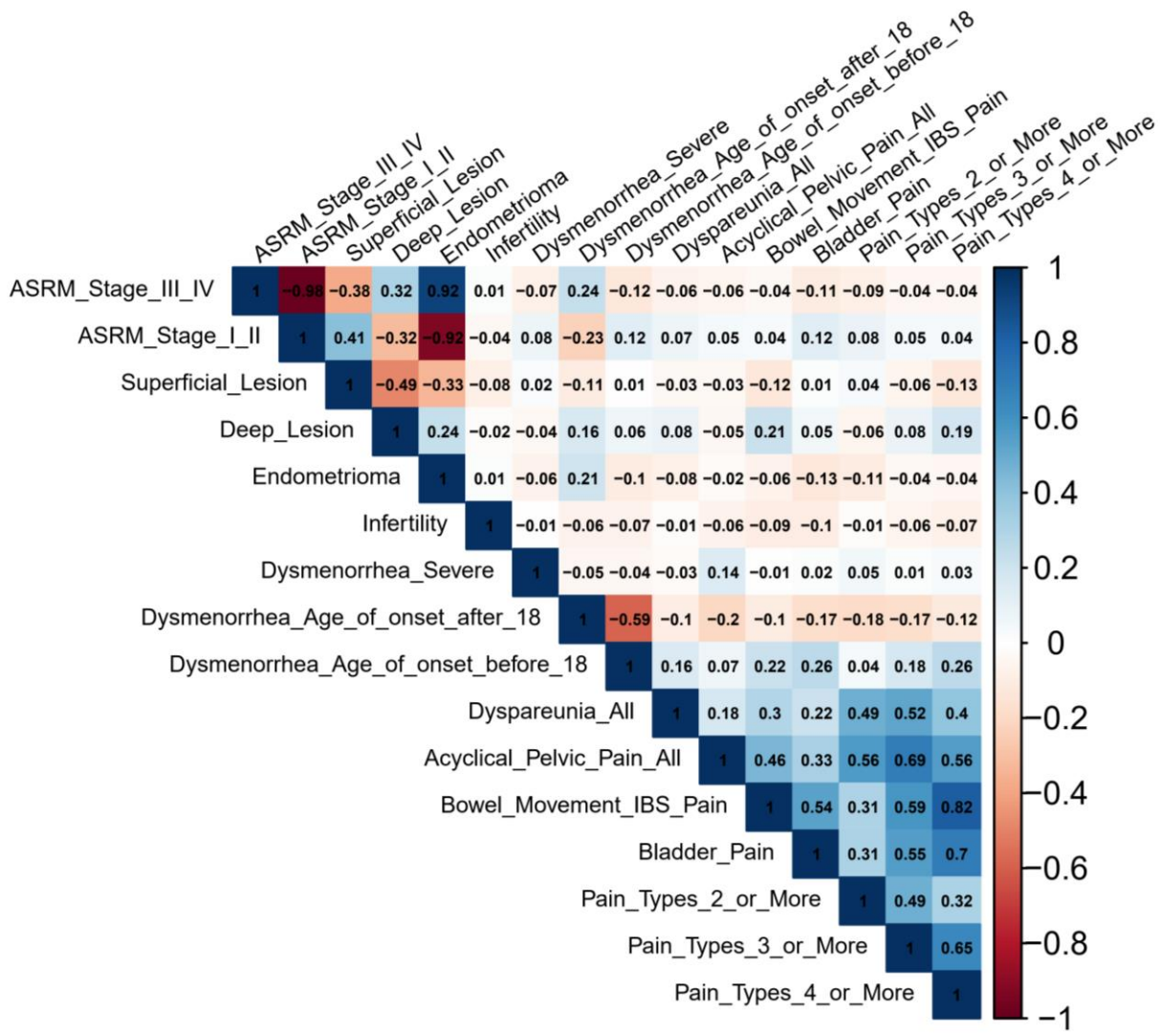


Figure 3.4 Correlation heat map of 16 major sub-phenotypes based on 211 endometriosis patients with full details on surgical and pain sub-phenotypes, coloured coded by Pearson's correlation coefficient.

Key: shades of blue = positive correlations; shades of red = negative correlations.

3.3.ii Results of 46 endometriosis SNP associations with endometriosis sub-phenotypes

3.3.ii.a Overview

For Phase I data, consistency in effect sizes using only clinical controls (n=77) and UKBB controls (n=4853) was examined first. Effect sizes were concordant in both magnitudes and directions (Supplementary Table S 3.1). For all subsequent analyses, the two control groups were combined. For phase II data, only population controls were available, genotyped on the Illumina Global gene chip array. As discussed in the Methods, association analyses were conducted separately for Phase I and Phase II, and results meta-analysed comprising a total of 361 cases and 5,279 controls. From 1,472 total number of tests carried out (i.e. 46 loci tested on 32 sub-phenotypes), 83 showed nominally significant results as shown in Table 3.4. Majority of these (96%) did not pass the least stringent multiple-correction threshold depicted by 46 loci x 3 main sub-phenotypic categories tested for (i.e. $p < 3.62 \times 10^{-4}$). Only signals from the rs71575922 locus (6q25.1/ *SYNE1*) for some sub-phenotypes passed the above threshold. Other nominal results would also be discussed throughout the chapter; and the *p* value reported were before multiple correction.

Of the 46 genetic variants tested, 16 showed nominal significant association (i.e. 95% CI of effect sizes not crossing one) with at least one of the tested sub-phenotypes. Detailed meta-analysis results of all 46 SNPs with overall endometriosis and 30 other sub-phenotypes are listed in Supplementary Table S 3.2-9. The effect estimates of 16 variants on sub-phenotypes are shown in Table 3.5 and forest plots are presented in Figure 3.7-Figure 3.16 and in the Supplementary Figure S 3.2-11. Out of these 16 variants, there are five loci showing significant associations in more than one sub-phenotype: rs11674184 located in 2p25.1 and the intron of *GREB1* (Figure 3.7); rs71575922 in 6q25.1 and intron of *SYNE1* (Spectrin Repeat Containing Nuclear Envelope Protein 1, Figure 3.9); rs1903068 in 4q12 and upstream of *KDR* (Kinase Insert Domain Receptor, Figure 3.10); rs10983311 in 9q33.1 and intronic to *ASTN2* (Astrotactin 2, Figure 3.11); and rs12441483 in 15q15.1 and downstream of *BMF* (B-cell Lymphoma 2 (BCL2) Modifying Factor, Figure 3.12).

Table 3.4 Loci passing the multiple-testing corrected *p*-value threshold for three to 32 independent sub-phenotype groups with respective multiple-testing correction thresholds.

Key: * = nominal association; majority of test statistics did not pass the least stringent multiple-correction threshold

Number / combination of independent variables tested				Multiple-testing correction threshold for *p*	Number of loci statistics passing the threshold (out of 1,472 tests)	Details of the significant loci (Associated sub-phenotypes)
Loci	**No. of Sub-phenotypes**	**Details of Sub-phenotypes**	**Total**			
46	32	all sub-phenotypes tested	1,472	3.40×10^{-5}	0	—
	12	Case-control, 2 ASRM groups, 3 surgical, 1 infertility & 5 pain types	552	9.06×10^{-5}	1	rs71575922/ *SYNE1* (ASRM Stage I/II)
	3	Surgical, infertility & pain categories	138	3.62×10^{-4}	3	rs71575922/ *SYNE1* (ASRM Stage I/II, case-control & recent severe dyspareunia)
	1	—	46	1.09×10^{-3}	4	rs71575922/ *SYNE1* (ASRM Stage I/II; case-control; dysmenorrhea & recent severe dyspareunia)
—	—	—	1	0.05 (no adjustment)	83*	—

The aim of this chapter is not to replicate established signals for overall endometriosis cases in this relatively small dataset. Nevertheless, one SNP showed nominally significant association ($p<3.62 \times 10^{-4}$) and two at $p<0.05$ with overall endometriosis with same direction of effects as reported in the latest IEGC meta-analysis (Supplementary Table S 3.2). They are located in intron of *SYNE1* on 6q25.1 (rs71575922 OR=1.61, 95% CI: 1.25-2.07, $p=2.40 \times 10^{-4}$), upstream of *KDR* on 4q12 (rs1903068 OR=1.26, 95% CI: 1.01-1.57, $p=3.65 \times 10^{-2}$), and upstream of *ARL14EP* (ADP Ribosylation Factor Like GTPase 14 Effector Protein; rs3858429 on 11p14.1; OR=1.39, 95% CI: 1.04-1.86, $p=2.71 \times 10^{-2}$, Supplementary Figure S 3.2).

Table 3.5 Summary of nominally significant associations of 16 genetic variants (out of 46 tested) with surgical, infertility, and pain sub-phenotypes of endometriosis.

Bladder pain sub-phenotype was also tested against all variants, though not shown as there is no association observed in this dataset. The odds ratios (ORs) are shown, with 95% confidence intervals in brackets, and *p*-value in square brackets. ORs for pain types refer to report of the pain, regardless of pain severity nor temporality. Key: * = Mild-moderate dyspareunia; # = Severe dysmenorrhea; + = OR indicates protective effect.

Variants (Cytoband/ Nearby Gene)	Sub-phenotypes of Endometriosis										
	Surgical					Infertility	Pain Types				
	ASRM Stage		Superficial Lesion	Deep Lesion	Ovarian Lesion		Dysmenorrhea	Dyspareunia	Acyclical Pelvic	Bowel Movement /IBS	≥ 3
	I & II	III & IV									
rs11674184 (2p25.1/ *GREB1*)		1.48 (1.11-1.98) [7.71x10^{-3}]			1.65 (1.11-2.45) [1.26 x10^{-2}]						
rs10983311 (9q33.1/ *ASTN2*)			1.34 (1.00-1.80) [4.89x10^{-2}]		1.80 (1.09-2.97) [2.25x10^{-2}]		1.41 (1.07-1.86) [1.61x10^{-2}]	1.61* (1.01-2.59) [4.70x10^{-2}]			
rs1903068 (4q12/ *KDR*)		1.40 (1.04-1.87) [2.50x10^{-2}]	1.44 (1.11-1.87) [5.34x10^{-3}]	1.53 (1.03-2.26) [3.38x10^{-2}]	1.61 (1.07-2.42) [2.13x10^{-2}]		1.32 (1.04-1.67) [2.07x10^{-2}]				1.37 (1.04-1.80) [2.42x10^{-2}]
rs12441483 (15q15.1/ *BMF*)			1.28 (1.02-1.60) [3.31x10^{-2}]		1.72 (1.21-2.44) [2.55x10^{-3}]	1.54 (1.10-2.16) [1.15x10^{-2}]	1.37 (1.11-1.69) [3.96x10^{-3}]	1.29 (1.01-1.64) [3.90x10^{-2}]	1.33 (1.04-1.71) [2.51x10^{-2}]	1.43 (1.07-1.90) [1.54x10^{-2}]	1.36 (1.06-1.74) [1.46x10^{-2}]
rs71575922 (6q25.1/ *SYNE1*)	1.76 (1.33-2.33) [7.26x10^{-5}]	1.53 (1.09-2.13) [1.27x10^{-2}]	1.49 (1.12-1.98) [5.50x10^{-3}]	1.76 (1.17-2.65) [6.63x10^{-3}]		1.92 (1.30-2.83) [1.12x10^{-3}]	1.58 (1.21-2.06) [7.46x10^{-4}]	1.59 (1.18-2.14) [2.48x10^{-3}]	1.56 (1.15-2.13) [4.63x10^{-3}]	1.50 (1.05-2.15) [2.64x10^{-2}]	1.63 (1.20-2.20) [1.65x10^{-3}]
rs3858429 (11p14.1/ *ARL14EP*)		1.75 (1.15-2.65) [9.09x10^{-3}]									
rs10917151 (1p36.12/ *WNT4*)		1.47 (1.05-2.05) [2.45x10^{-2}]									

Variants (Cytoband/ Nearby Gene)	Sub-phenotypes of Endometriosis (Continued)										
	Surgical					Infertility	Pain Types				
	ASRM Stage		Superficial Lesion	Deep Lesion	Ovarian Lesion		Dysmenorrhea	Dyspareunia	Acyclical Pelvic	Bowel Movement /IBS	≥ 3
	I & II	III & IV									
rs6456259 (6p22.3/ *ID4*)		1.43 (1.03-1.99) [$3.12x10^{-2}$]									
rs7740302 (6q25.1/ *CCDC170*)		1.48 (1.08-2.02) [$1.36x10^{-2}$]									
rs1451383 (7p15.2)		1.41 (1.02-1.95) [$4.01x10^{-2}$]									
rs7214750 (17q24.1/ *CEP112*)		1.98 (1.05-3.74) [$3.56x10^{-2}$]									
rs7907732 (10q23.31/ *RNLS*)				1.43 (1.02-2.02) [$3.89x10^{-2}$]							
rs7924571 (11p14.1/ *FSHB*)					1.66 (1.01-2.74) [$4.64x10^{-2}$]						
rs507666 (9q34.2/ *ABO*)							1.37# (1.00-1.86) [$4.69x10^{-2}$]				
rs2967684 (19p13.2/ *ACTL9*)								1.42 (1.04-1.94) [$2.75x10^{-2}$]			
rs2510770 (4q22.3/ *PDLIM5*)								0.56+ (0.33-0.95) [$3.03x10^{-2}$]			

3.3.ii.b Correlations of effect estimates of tested genetic variants between endometriosis sub-phenotypes

Correlation of effect sizes from 46 tested SNPs between sub-phenotypes of endometriosis were analysed (Supplementary Figure S 3.12 for heat map). The heat map suggest that certain sub-phenotype groups are more genetically correlated with each other in terms of effect sizes than others. For instance, the effect sizes of SNPs associated with infertility sub-phenotypes are more correlated with the effect sizes for superficial lesions (r^2 = 0.66) than for deep lesions (r^2 = 0.43) or endometrioma (r^2 = 0.39); the phenotypic correlation, however, showed weak correlations (Figure 3.4) between infertility with superficial lesions (r = -0.08), with deep lesions (r = -0.02), and with endometrioma (r = +0.01). There is also a higher correlation of genetic variant effect sizes for dysmenorrhoea sub-phenotypes with superficial lesions (r^2=0.72-0.84) than deep lesions (r^2=0.47-0.65). The phenotypic correlations of severe dysmenorrhea with these surgical sub-phenotypes were, again, very weak (r = +0.02 for superficial lesions, and -0.04 for deep lesions).

Next, effect sizes for genetic associations with mutually exclusive sub-phenotypes were compared to investigate the relative genetic contributions to each 46 loci: ASRM stage I/II or III/IV, and age of onset for dysmenorrhea under or at/over 18 years old. As expected, genetic variant effect sizes were generally greater for ASRM stage III/IV than I/II. Figure 3.5 depicts this, showing 29/46 loci with effects greater for stage III/IV than stage I/II (above the diagonal, r^2 = 0.26). Variants rs3858429 (11p14.1/*ARL14EP*) and rs7214750 (17q24.1/*CEP112*), with MAF of 16% and 7% respectively, show the two highest effect sizes for association with stage III/IV among all tested, OR=1.75, 95% CI: 1.15-2.65, p=9.09×10^{-3}, and OR=1.98, 95% CI: 1.05-3.74, p=0.04 respectively. Similarly, SNPs rs71575922 (6q25.1/*SYNE1*, MAF: 15%) and rs73241342 (Xq26.3/*PLAC1*, MAF: 4%) loci show the two highest effect for stage I/II disease, OR=1.76, 95% CI: 1.33-2.33, p=7.26×10^{-5}, and OR=1.59, 95% CI: 0.89-2.84, p=0.12 respectively, though the estimate for the latter is not nominally significant with a wide confidence interval crossing one.

Figure 3.6 shows the effect estimates of tested SNPs comparing onset of dysmenorrhea before or after 18 years of age. A tendency towards greater effect sizes for later onset (likely secondary) dysmenorrhoea was observed (r^2=0.27). Variants rs71575922 (6q25.1/*SYNE1*, MAF: 15%) and rs7214750 (17q24.1/*CEP112*, MAF: 7%) showed the two highest effect sizes for onset of dysmenorrhea at/after than 18 years of age: $OR_{\geq 18}$=1.76, 95% CI: 1.02-3.03, p=0.04, and $OR_{\geq 18}$=1.87, 95% CI: 0.35-9.89, p=0.46 respectively. Although the estimate of rs7214750 (17q24.1/*CEP112*) for dysmenorrhea onset after 18 is not nominally significant, it showed opposite effect with that of onset before 18: $OR_{<18}$=0.85, 95% CI: 0.51-1.44, p=0.55. Similarly, SNPs rs12441483 (15q15.1/*BMF*, MAF: 47%) and rs507666 (9q34.2/*ABO*, MAF: 20%) were among the three showing highest effect sizes for dysmenorrhea onset before age of 18 with OR=1.53, 95% CI: 1.14-2.06, p=4.60×10^{-3}, and OR=1.40, 95% CI: 0.98-1.99, p=0.06 respectively.

For correlation plots of 46 effect estimates between surgical sub-phenotypes (deep vs. superficial lesions, endometrioma vs. superficial lesion) and pain sub-phenotypes (severe dysmenorrhea vs. dyspareunia, acyclical pelvic pain vs. dyspareunia, and bowel movement/IBS pain vs. bladder pain), please refer to Supplementary Figure S 3.13-14.

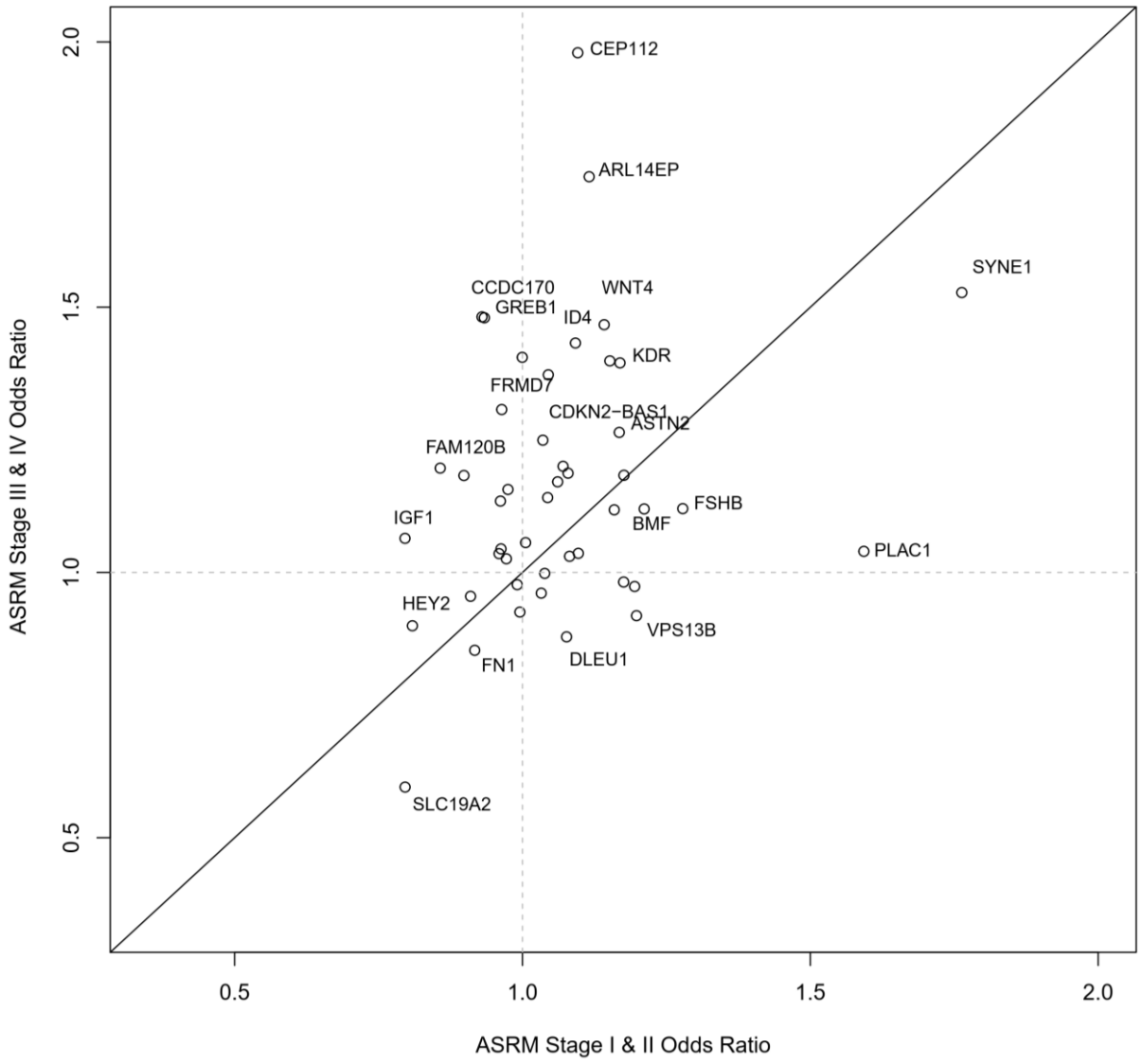


Figure 3.5 Correlation plot of effect estimates for association of 46 genetic variants with ASRM stage III & IV versus stage I & II sub-phenotype of endometriosis; Pearson's coefficient r^2=0.26.

Key: diagonal line is for reference when x=y; only odds ratios between 0.5-2.0 are displayed, 95% confidence intervals of data points are not shown.

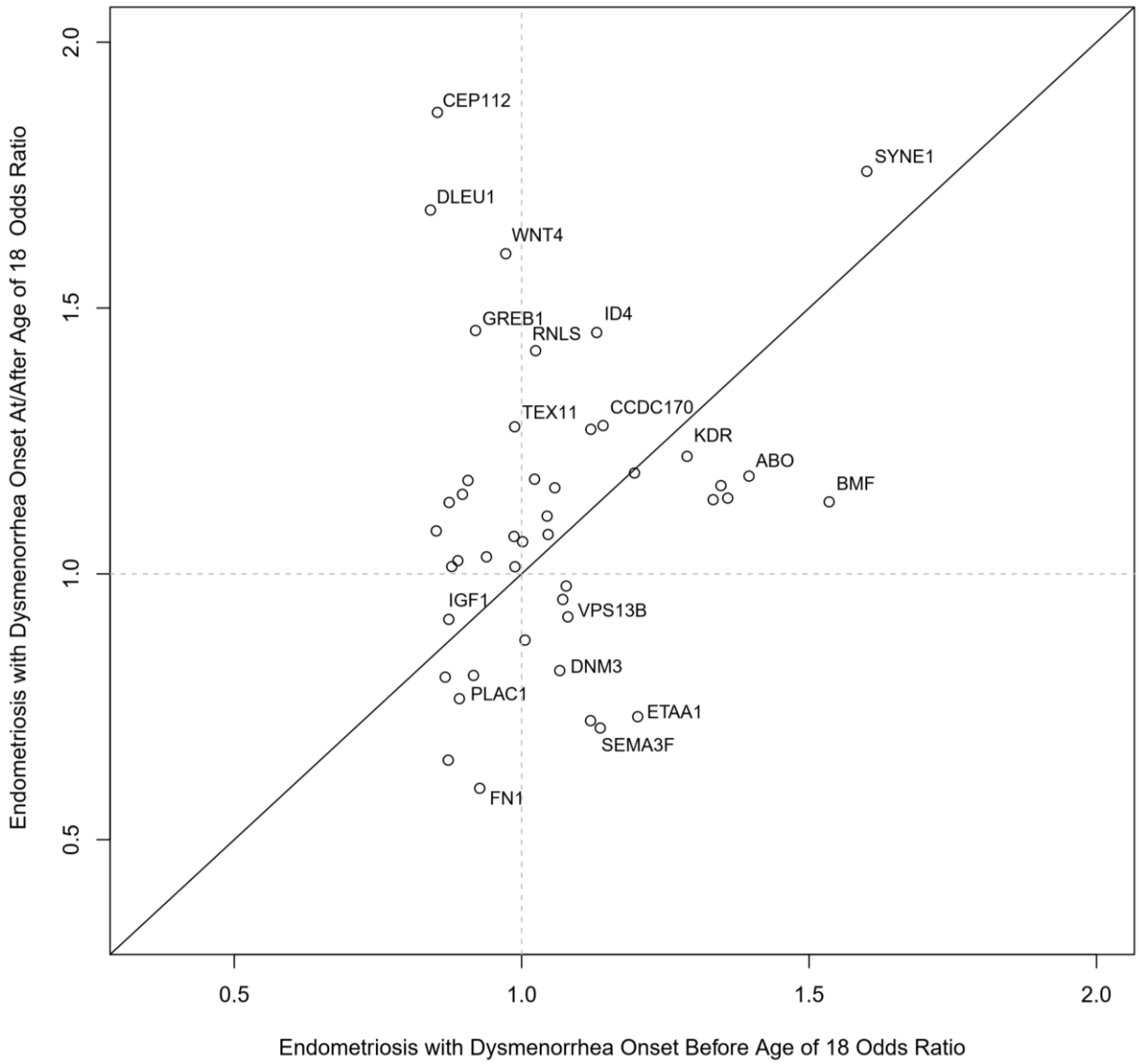


Figure 3.6 Correlation plot of effect estimates of associations for 46 genetic variants with dysmenorrhea onset at/after age of 18 versus before 18 in endometriosis; Pearson's coefficient r^2=0.27.

Key: diagonal line is for reference when x=y; only odds ratios between 0.5-2.0 are displayed, 95% confidence intervals of data points are not shown.

3.3.ii.c Association of selected genetic variants on surgical sub-phenotypes of endometriosis

Only one variant shows a nominally differential effect for stage III/IV disease when tested directly against stage I/II cases, i.e. rs11674184 in the intron of *GREB1* on 2p25.1 with T as the risk allele (RA); OR=1.55, 95% CI: 1.06-2.29, $p=2.51x10^{-2}$ (Figure 3.7, *Stages III&IV vs. I&II). When comparing stage III/IV cases with controls, in addition to the association signal observed in the *GREB1* locus (OR=1.48, 95% CI: 1.11-1.98, $p=7.71x10^{-3}$), eight other variants also nominally associate with stage III/IV (Table 3.5 & Supplementary Table S 3.2); besides rs1903068 (4q12/*KDR*, OR=1.40, 95% CI: 1.04-1.87, $p=2.50x10^{-2}$), rs71575922 (6q25.1/*SYNE1*, OR=1.53, 95% CI: 1.09-2.13, $p=4.01x10^{-2}$), and rs3858429 (11p14.1/*ARL14EP*, OR=1.75, 95% CI: 1.15-2.65, $p=9.09x10^{-3}$) loci, the other five variants are rs10917151 downstream of *WNT4* on 1p36.12 (OR=1.47, 95% CI: 1.05-2.05, $p=2.45x10^{-2}$, Supplementary Figure S 3.3), rs6456259 upstream of *ID4* on 6p22.3 (OR=1.43, 95% CI: 1.03-1.99, $p=3.12x10^{-2}$, Supplementary Figure S 3.4), rs7740302 in intron of *CCDC170* on 6q25.1 (OR=1.48, 95% CI: 1.08-2.02, $p=1.36x10^{-2}$, Supplementary Figure S 3.5), rs1451383 upstream of *NFE2L3* on 7p15.2 (OR=1.41, 95% CI: 1.02-1.95, $p=4.01x10^{-2}$, Supplementary Figure S 3.6), and lastly rs7214750 in intron of *CEP112* on 17q24.1 (OR=1.98, 95% CI: 1.05-3.74, $p=3.56x10^{-2}$, Supplementary Figure S 3.7).

For endometrioma diagnosis, nominal associations are shown for five SNPs, including rs11674184 (2p25/*GREB1*, OR=1.65, 95% CI: 1.11-2.45, p=1.26x10^{-2}) and rs1903068 (4q12/*KDR*, OR=1.61, 95% CI: 1.07-2.42, p=2.13x10^{-2}); rs10983311 in intron of *ASTN2* on 9q33.1 (OR=1.81, 95% CI: 1.09-2.97, p=2.25x10^{-2}) and rs12441483 downstream of *BMF* on 15q15.1 (OR=1.72, 95% CI: 1.21-2.44, p=2.55x10^{-3}). The variant rs7924571, 2Mb downstream of *FSHB* (Follicle Stimulating Hormone Beta Subunit) on 11p13 also shows OR of 1.66 (95% CI: 1.01-2.74, p=4.64x10^{-2}, Figure 3.8) for endometrioma sub-phenotype. Deep lesion diagnosis was nominally associated to rs1903068 (4q12/*KDR*, OR=1.53, 95% CI: 1.03-2.26, p=3.38x10^{-2}) and rs71575922 (6q25.1/*SYNE1*, OR=1.76, 95% CI: 1.17-2.65, p=6.63x10^{-3}). In addition with rs7907732 in the intron of *RNLS* (Renalase – FAD Dependent Amine Oxidase) on 10q23.31 (OR=1.43, 95% CI: 1.02-2.02, p=3.89x10^{-2}, Supplementary Figure S 3.8). Superficial lesions were nominally associated with four variants, all of which associated with other surgical sub-phenotypes mentioned above: rs71575922 (6q25.1/*SYNE1*, OR=1.49, 95% CI: 1.12-1.98, p=5.50x10^{-3}), rs1903068 (4q12/*KDR*, OR=1.44, 95% CI: 1.11-1.87, p=5.34x10^{-3}), rs10983311 (9q33.1/*ASTN2*, OR=1.34, 95% CI: 1.00-1.80, p=4.89x10^{-2}), and rs12441483 (15q15.1/*BMF*, OR=1.28, 95% CI: 1.02-1.60, p=3.31x10^{-2}).

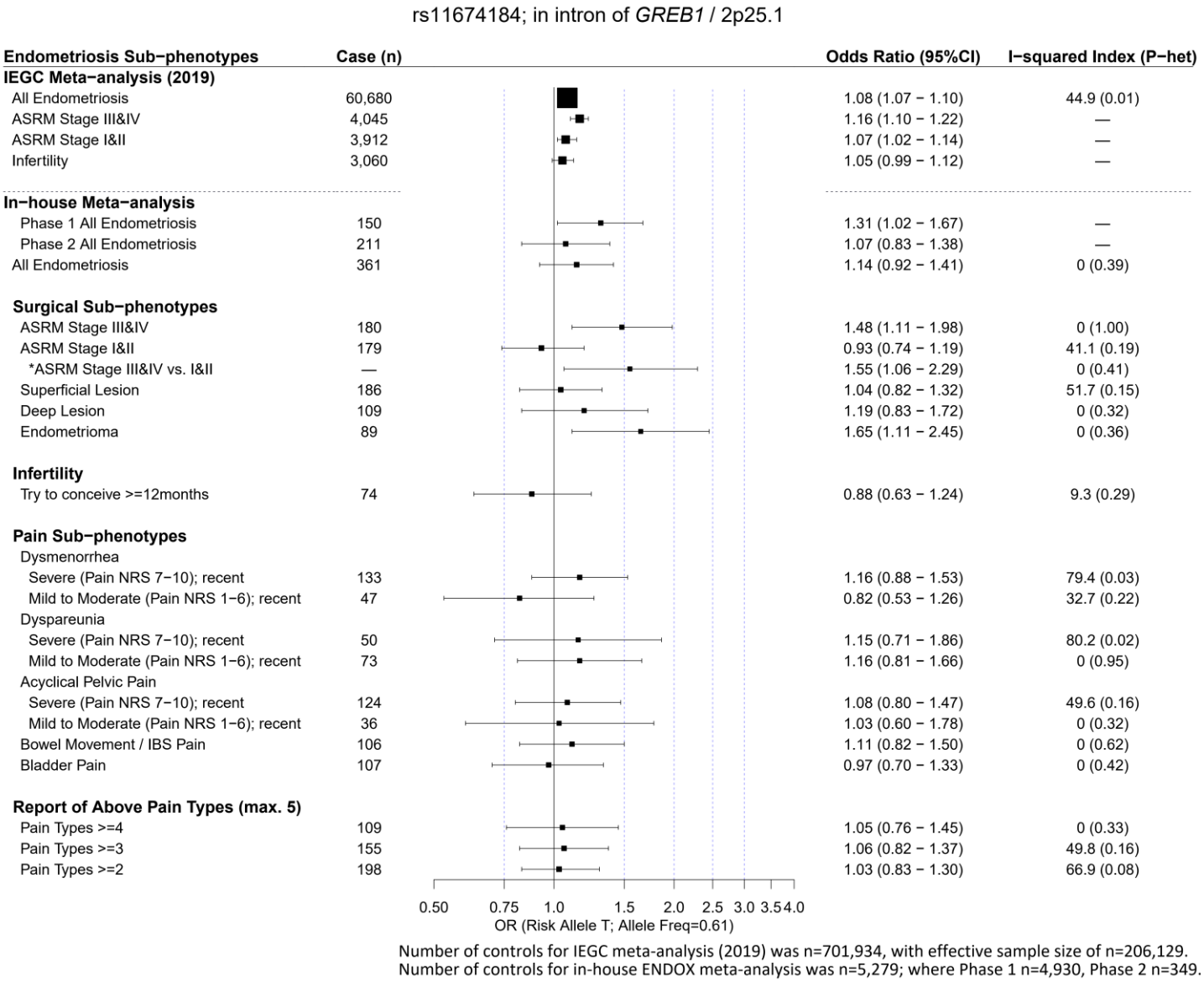


Figure 3.7 Forest plot for effect estimates of the T allele of rs11674184 on 2p25.1, in the intron of *GREB1*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. Key: I^2 index = Estimation for Heterogeneity between ENDOX Genotyping Phase I & II; P_{Het} = P-value of the Estimated Heterogeneity

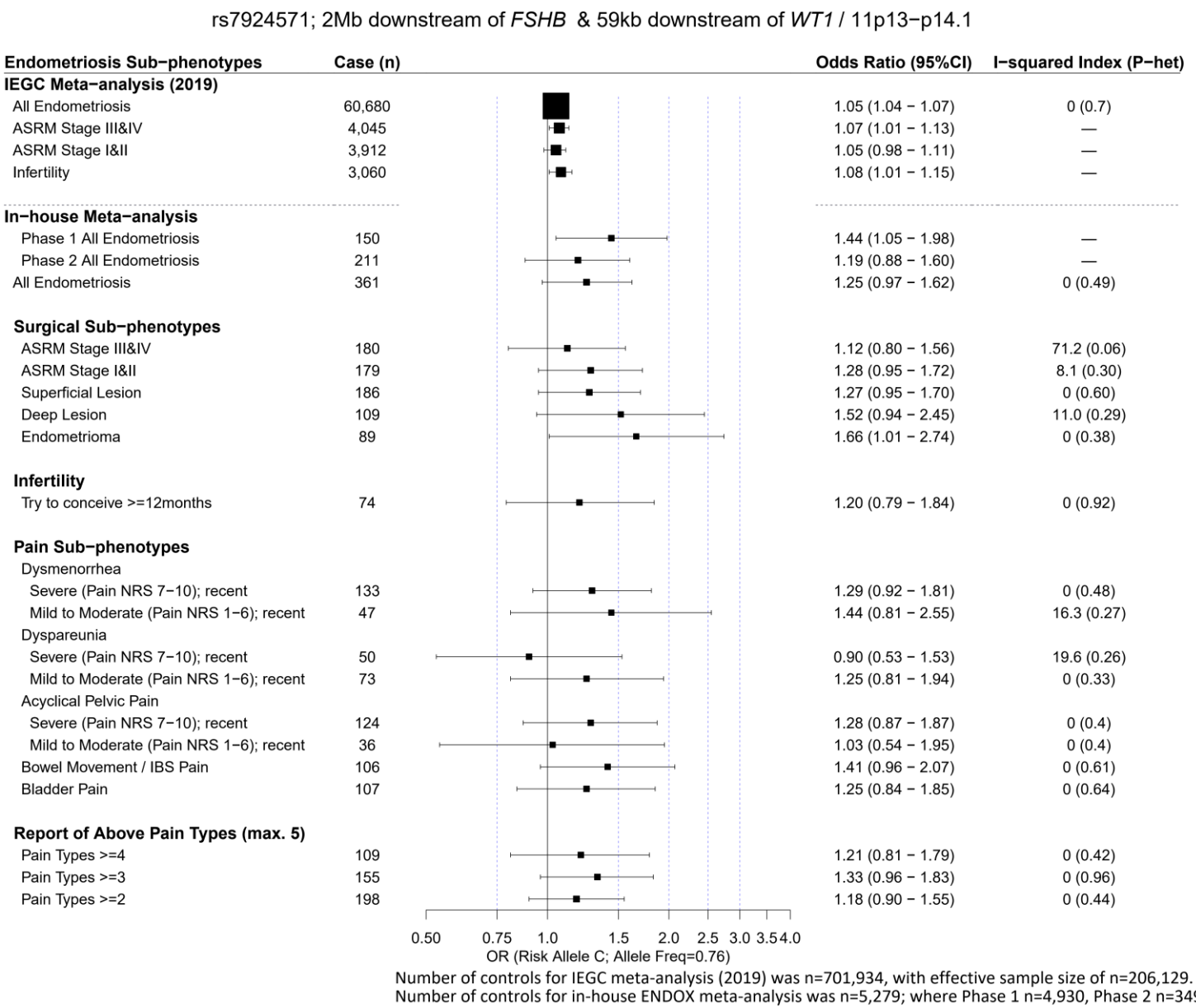


Figure 3.8 Forest plot for effect estimates of the C allele of rs7924571 on 11p13, downstream of *WT1* & *FSHB*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.

3.3.ii.d Effect of selected genetic variants on infertility sub-phenotypes of endometriosis

Out of all variants tested, two loci showed nominal association with self-reported infertility related to endometriosis: rs12441483 (15q15.1/*BMF*, OR=1.54, 95% CI: 1.10-2.16, $p=1.15x10^{-2}$, Figure 3.12) and rs71575922 (6q25.1/*SYNE1*, OR=1.92, 95% CI: 1.30-2.83, $p=1.12x10^{-3}$, forest plot below). The effect estimates for both these variants are larger (OR=1.54 and 1.92 respectively) when infertility is defined as trying to conceive consecutively for ≥ 12 months, than the less stringent definition of ≥ 6 months (OR=1.35, 95% CI: 1.02-1.80, p=0.04, and OR=1.54, 95% CI: 1.08-2.19, p=0.02 respectively, Figure 3.15-Figure 3.16 in Section 3.3.ii.f).

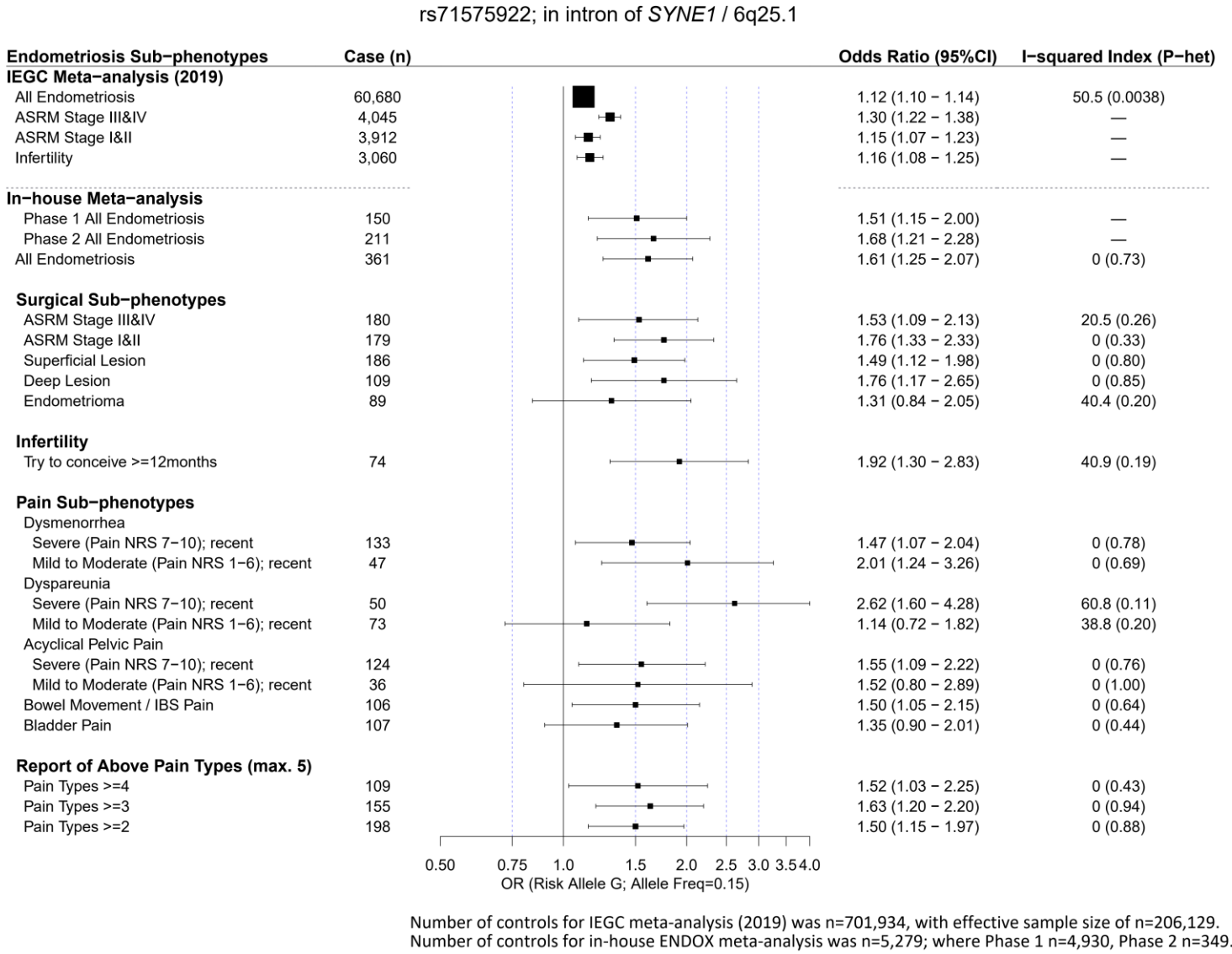


Figure 3.9 Forest plot for effect estimates of the G allele of rs71575922 on 6q25.1, in the intron of *SYNE1*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.

3.3.ii.e Effect of selected genetic variants on pain sub-phenotypes of endometriosis

For the total number of pain types reported (a proxy for hyperalgesia) in endometriosis, the results for ≥ 2, ≥ 3 and ≥ 4 pain types (out of maximum of 5) were similar. Variants rs71575922 (6q25.1/*SYNE1*), rs1903068 (4q12/*KDR*), and rs12441483 (15q15.1/*BMF*) showed association with ≥ 3 pain types with effect estimates at OR=1.63 (95% CI: 1.20-2.20, $p=1.65x10^{-3}$, Figure 3.9), OR=1.37 (95% CI: 1.04-1.80, $p=2.42x10^{-2}$, Figure 3.10), and OR=1.36 (95% CI: 1.08-1.66, $p=1.46x10^{-2}$, Figure 3.12) respectively. Among the 46 tested SNPs, there were no SNPs showing nominal association with the bladder pain sub-phenotype of endometriosis in this dataset.

For endometriosis patients who reported dysmenorrhea in recent cycles, five loci showed nominal association with symptom severity. Variants rs1903068 (4q12/*KDR*, OR=1.35, 95% CI: 1.01-1.80, $p=4.33x10^{-2}$, Figure 3.10), rs10983311 (9q33.1/*ASTN2*, OR=1.45, 95% CI: 1.03-2.06, $p=3.44x10^{-2}$, Figure 3.11), and rs12441483 (15q15.1/*BMF*, OR=1.32, 95% CI: 1.02-1.71, $p=3.40x10^{-2}$, Figure 3.12) exhibited nominal association with severe dysmenorrhea; their effects for overall, and moderate to severe dysmenorrhea (not for *BMF* locus) are also nominally significant (Figure 3.13-Figure 3.15). The rs71575922 in *SYNE1* locus is nominally associated with dysmenorrhea across all severity groupings (Figure 3.9 & Figure 3.16), but with a larger effect size in mild to moderate pain (OR=2.01, 95% CI: 1.24-3.26, $p=4.72x10^{-3}$) than in severe pain (OR=1.47, 95% CI: 1.07-2.04, $p=1.92x10^{-2}$). The variant rs507666 in intron of *ABO* (Alpha 1-3-N-Acetylgalactosaminyltransferase and Alpha 1-3-Galactosyltransferase) on 9q34.2 showed a borderline association for severe dysmenorrhea with OR=1.37, 95% CI: 1.00-1.86, $p=4.69x10^{-2}$ (Supplementary Figure S 3.9); however, its effect for overall dysmenorrhea was not nominally significant (OR=1.26, 95% CI: 0.97-1.63, $p=8.44x10^{-2}$, Supplementary Figure S 3.15).

Similar to the results for dysmenorrhea, three of the above mentioned loci (five variants) also showed significant effects for dyspareunia (in the rs71575922/*SYNE1*, rs12441483/*BMF* & rs10983311/*ASTN2* loci). The SNPs in the *SYNE1*, *BMF* loci, and rs2967684 upstream of *ACTL9* (Actin Like 9) on 19p13.2 were nominally associated with overall dyspareunia. Their respective ORs were 1.59 (95% CI: 1.18-2.14, $p=2.48x10^{-3}$, Figure 3.16), 1.29 (95% CI: 1.01-1.64, $p=3.90x10^{-2}$, Figure 3.15), and 1.42 (95% CI: 1.04-1.94, $p=2.75x10^{-2}$, Supplementary Figure S 3.16). For rs71575922/*SYNE1*, effect sizes increased with pain severity; the effect estimate for severe dyspareunia (recent) is 2.62 (95% CI: 1.60-4.28, significant $p=1.24x10^{-4}$ passing the least stringent multiple correction threshold of $p<3.62x10^{-4}$, Table 3.4) when compared to 1.14 (95% CI: 0.72-1.82, p=0.57) for mild to moderate pain.

For two other loci, rs2967684/*ACTL9* (Supplementary Figure S 3.10) and rs10983311/*ASTN2*, effect sizes appeared larger for milder pain. For instance, the effect estimate for mild to moderate dyspareunia (recent) of rs10983311/*ASTN2* was 1.61 (95% CI: 1.01-2.59, $p=4.70x10^{-2}$, Figure 3.11&Figure 3.14), whereas for moderate to severe pain it was 1.14 (95% CI: 0.79-1.62, p=0.49). The fifth SNP, rs2510770 in the intron of *PDLIM5* (PDZ And LIM Domain 5) on 4q22.3, the established risk allele A for endometriosis showed a protective effect on severe dyspareunia, both for recent pain (OR=0.56, 95% CI: 0.33-0.95, $p=3.03x10^{-2}$, n=50, Supplementary Figure S 3.11), and at worst in lifetime (OR=0.59, 95% CI: 0.40-0.88, $p=9.04x10^{-3}$, n=77, Supplementary Figure S 3.17).

As noted in Figure 3.3a, there was considerable overlap in reporting of acyclical pelvic pain and bowel movement/IBS pain; 61% of cases included in the acyclical pelvic pain group had bowel movement/IBS pain and 96% of women with bowel/IBS pain had acyclical pelvic pain. rs71575922/*SYNE1* and rs12441483/*BMF* showed nominal association with both pain conditions. The effect sizes of variants of *BMF* and *SYNE1* loci were consistent across different severity definitions of acyclical pelvic pain (Figure 3.15-Figure 3.16), with ORs for overall acyclical pelvic pain at 1.33 (95% CI: 1.04-1.71, $p=2.51x10^{-2}$) for rs12441483/*BMF* and 1.56 (95% CI: 1.15-2.13, $p=4.63x10^{-3}$) for rs71575922/*SYNE1*. The effect estimates for bowel movement/IBS pain were OR=1.43 (95% CI: 1.07-1.90, $p=1.54x10^{-2}$, Figure 3.12) for rs12441483/*BMF*, and 1.50 (95% CI: 1.05-2.15, $p=2.64x10^{-2}$, Figure 3.9) for rs71575922/*SYNE1*.

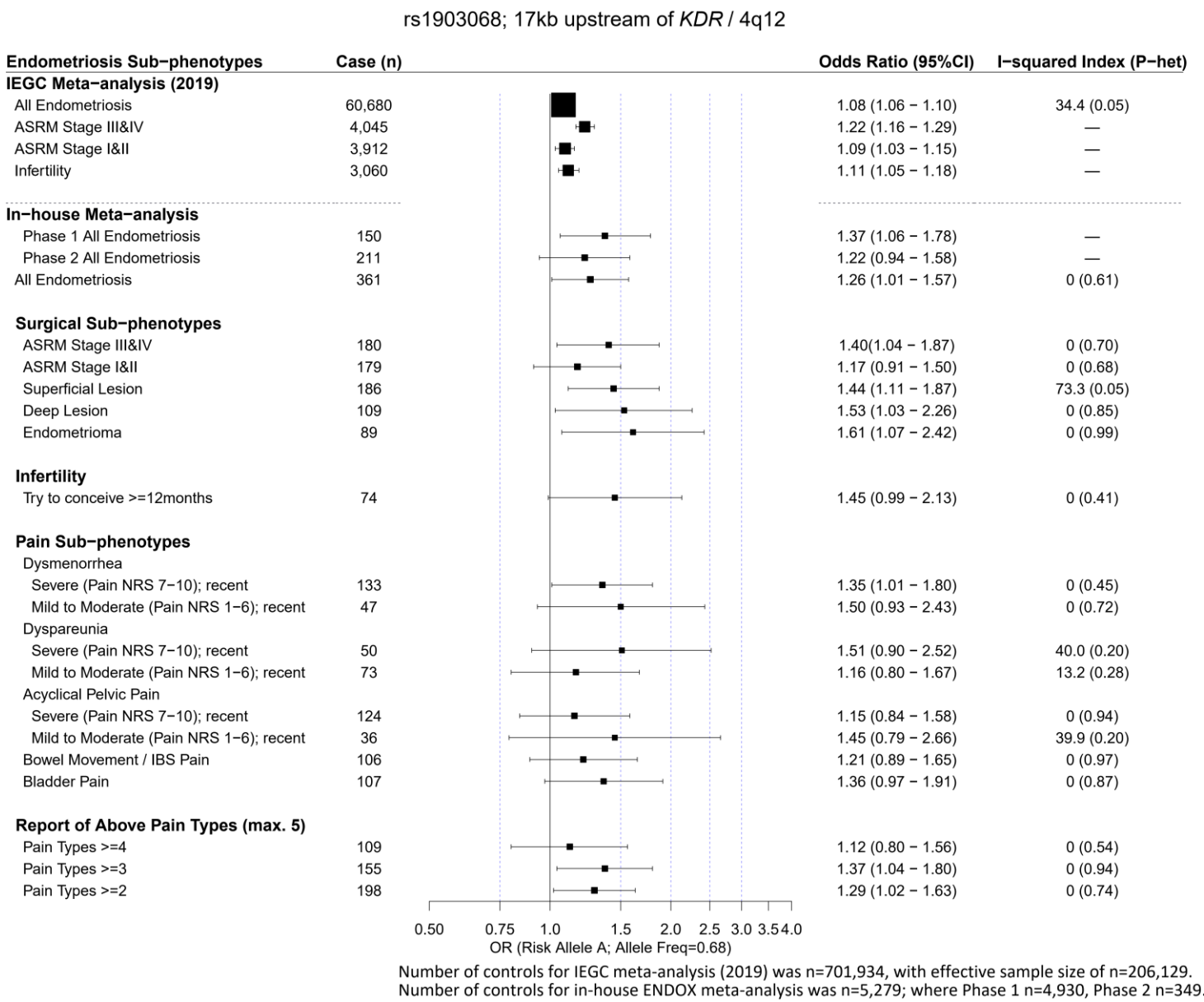


Figure 3.10 Forest plot for effect estimates of the A allele of rs1903068 on 4q12, upstream of *KDR*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.

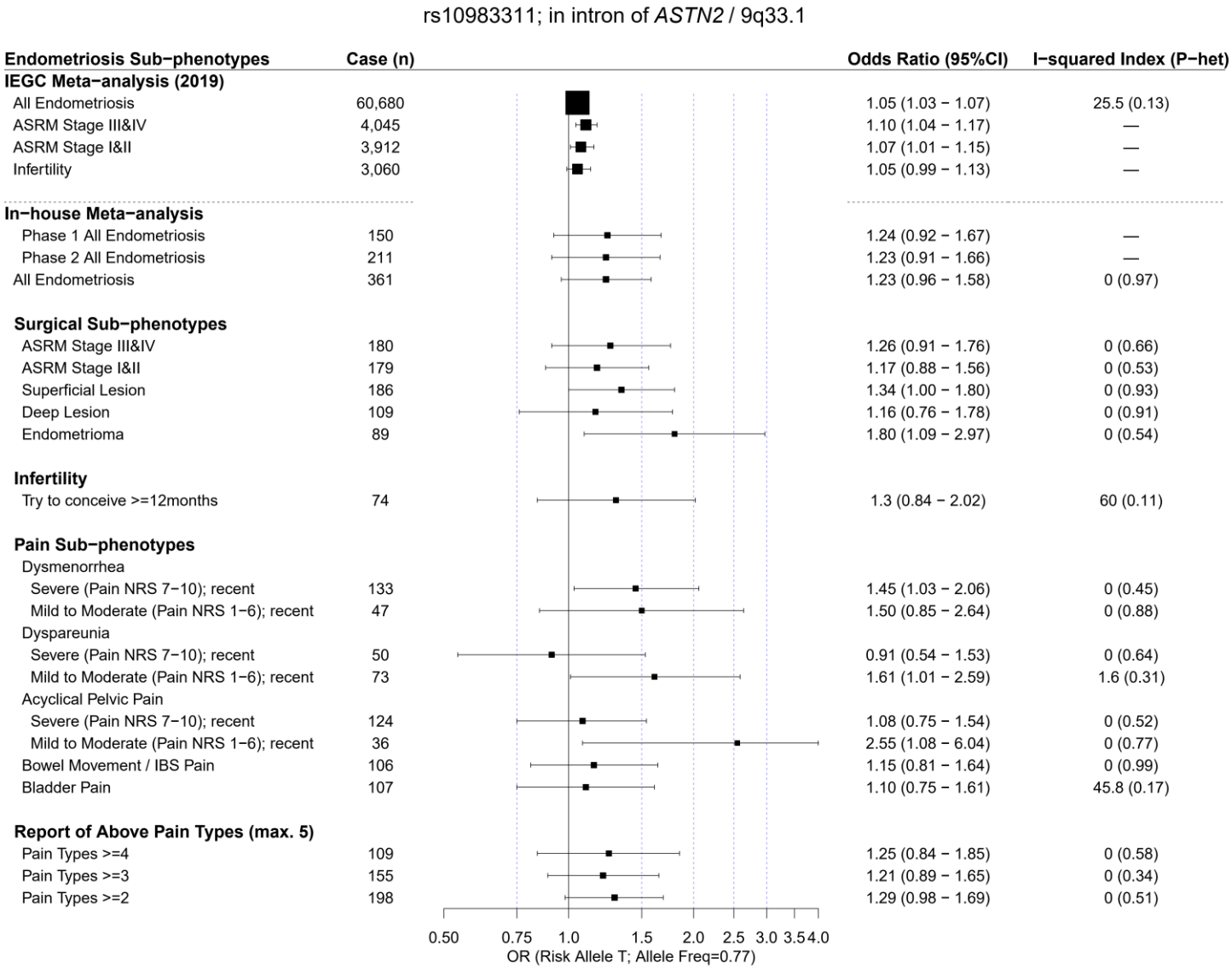


Figure 3.11 Forest plot for effect estimates of the T allele of rs10983311 on 9q33.1, in the intron of *ASTN2*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.

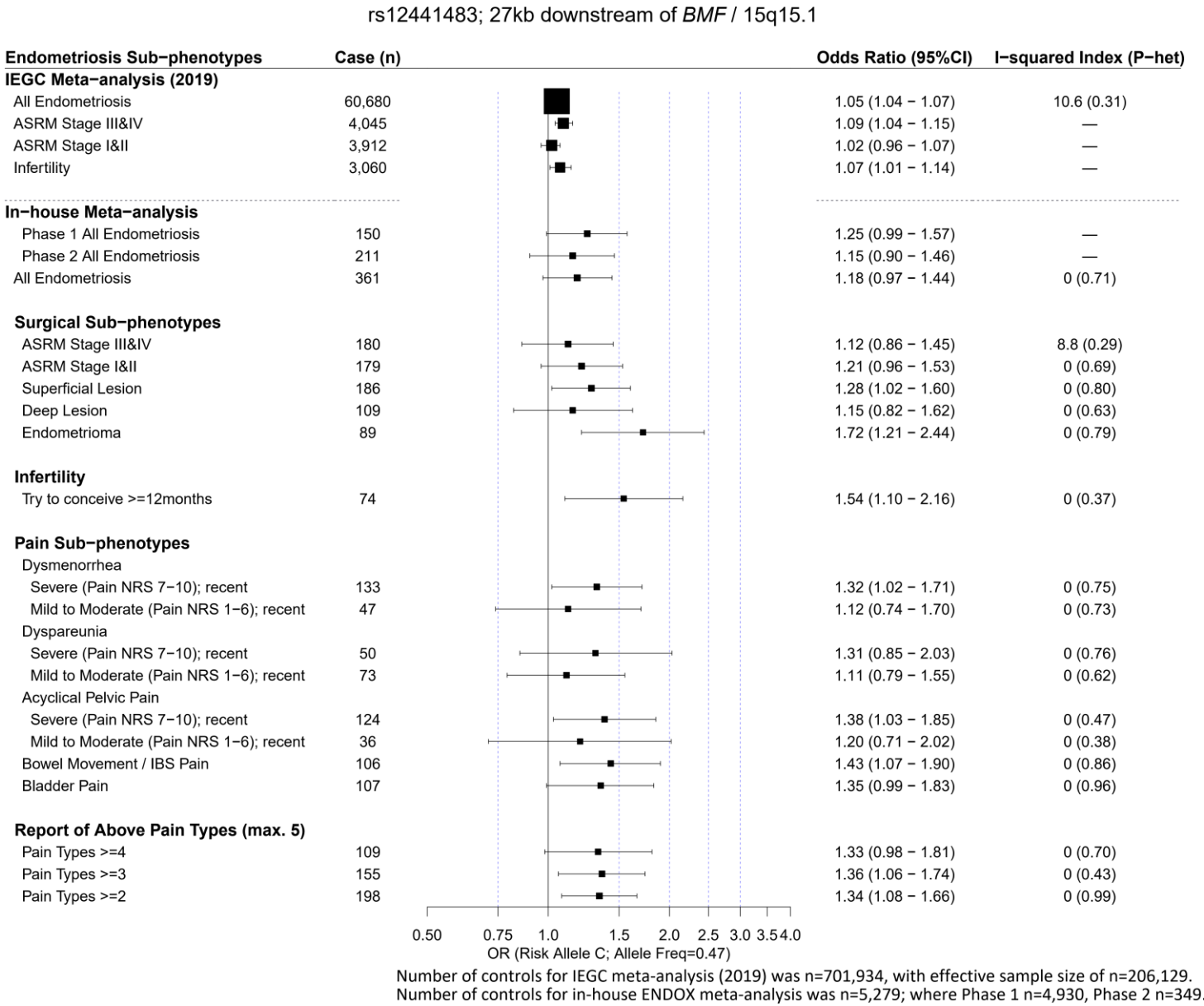


Figure 3.12 Forest plot for effect estimates of the C allele of rs12441483 on 15q15.1, downstream of *BMF*, for major surgical, infertility, and pain sub-phenotypes of endometriosis

3.3.ii.f Sensitivity testing of pain and infertility sub-phenotypes of endometriosis

Next, effects among varying definitions of severity and temporality of the three classic triad pain types for endometriosis — dysmenorrhoea, dyspareunia and acyclical pelvic pain — were explored, to investigate if better resolution and confidence in the validity of association signals could be gained. Self-reported infertility was also tested with two definition tiers (Section 3.3.ii.d). The sizes of effect estimates for genetic variants were generally consistent across varying definitions of dysmenorrhea (differences in ORs ± 0.10, i.e. <10% difference in odds ratio); taking the variant rs1903068 upstream of *KDR* as an example (Figure 3.13), the ORs ranged from 1.32 to 1.35 for severity-stratified groups with more than one hundred samples. Consistent patterns were also observed for variants in rs10983311/*ASTN2*, rs12441483/*BMF*, rs71575922/*SYNE1* (Figure 3.14-Figure 3.16), and *ABO* loci (Supplementary Figure S 3.15). Worth noting was that the rs12441483/*BMF* variant appeared to a nominally significant association with dysmenorrhea onset before 18 years of age (OR=1.53, 95% CI: 1.14-2.06, $p=4.60x10^{-3}$), but not for onset ≥ 18 years old (Figure 3.15). The opposite was observed for SNPs rs10983311/*ASTN2* ($OR_{onset\ \geq age\ 18}$=2.53, 95% CI: 1.20-5.31, $p=1.44x10^{-2}$) and rs71575922/*SYNE1* ($OR_{onset \geq age\ 18}$=1.76, 95% CI: 1.02-3.03, $p=4.30x10^{-2}$ vs. $OR_{onset <age\ 18}$=1.60, 95% CI: 1.11-2.31, $p=1.19x10^{-2}$).

For acyclical pelvic pain, the effect estimates across different definitions (with sample size near to one hundred) were consistent for some variants (e.g. rs12441483/*BMF* & rs71575922/*SYNE1*), but not others (e.g. rs10983311/*ASTN2* & rs2967684/*ACTL9*, Supplementary Figure S 3.16). It is statistically not feasible to make direct comparison of these results from multiple testing of correlated traits, hence, the uncertainty. For dyspareunia, the effect estimates across stratified groups were relatively inconsistent for most tested loci. The observation of 'the milder the pain, the larger the effect size', appeared particularly pronounced in signals in rs10983311/*ASTN2* and rs2967684/*ACTL9*; there was also an unexpected, albeit non-significant 'flipped' effect in the severe pain group who experienced recent pain (n=50, Figure 3.14 & Supplementary Figure S 3.16). Due to statistical non-independence of groups, any differences in effect sizes between groups could not be tested for significance. However, the inconsistent pattern of risk versus protective effects could reflect potential differences in responses in terms of pain temporality and/ or a biased capture of information, based on frequency of intercourse, for this variable in particular. This led to a further investigation of missingness patterns in the next section.

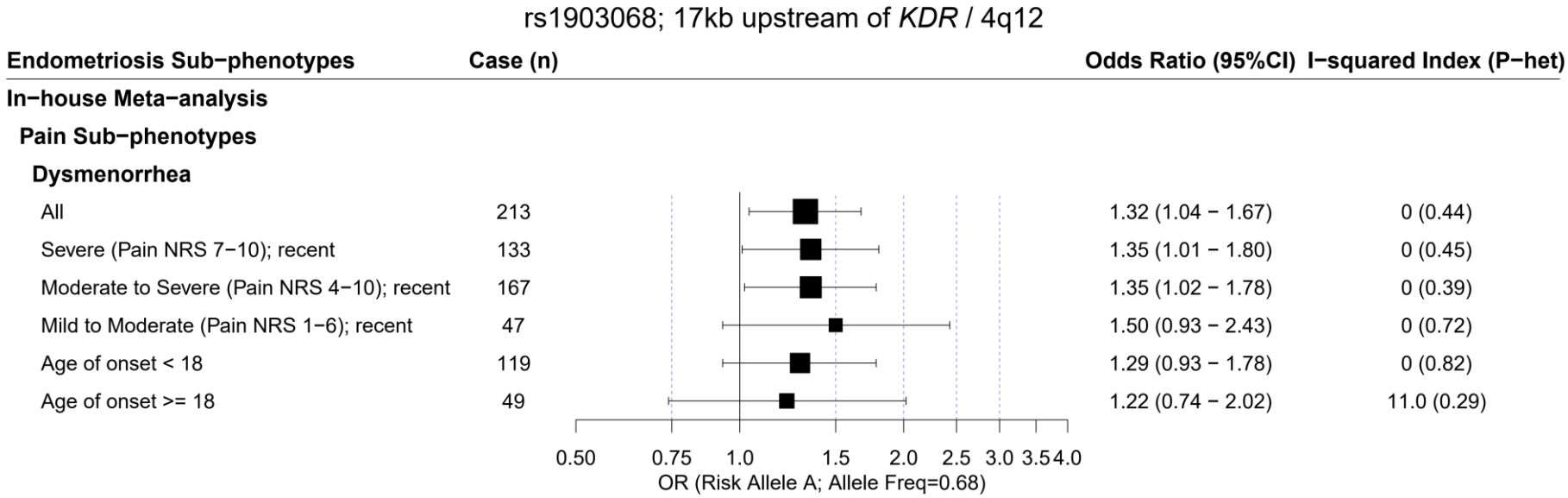


Figure 3.13 Forest plot for effect estimates of the A allele of rs1903068 on 4q12, upstream of *KDR*, for sensitivity testing of dysmenorrhea sub-phenotype associations.

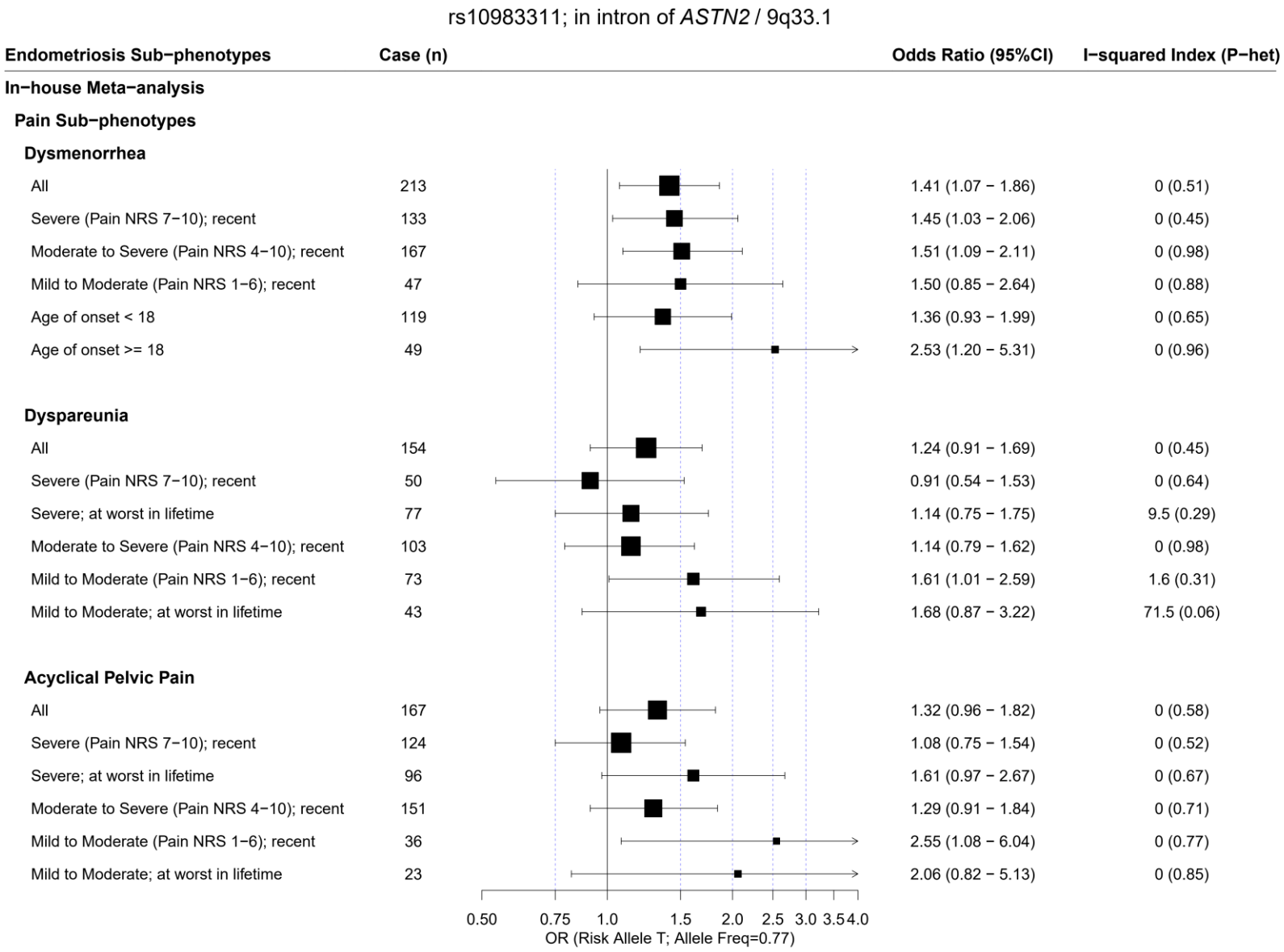


Figure 3.14 Forest plot for effect estimates of the T allele of rs10983311 on 9q33.1, in the intron of *ASTN2*, for sensitivity testing of dysmenorrhea, dyspareunia, and acyclical pelvic pain sub-phenotypes associations.

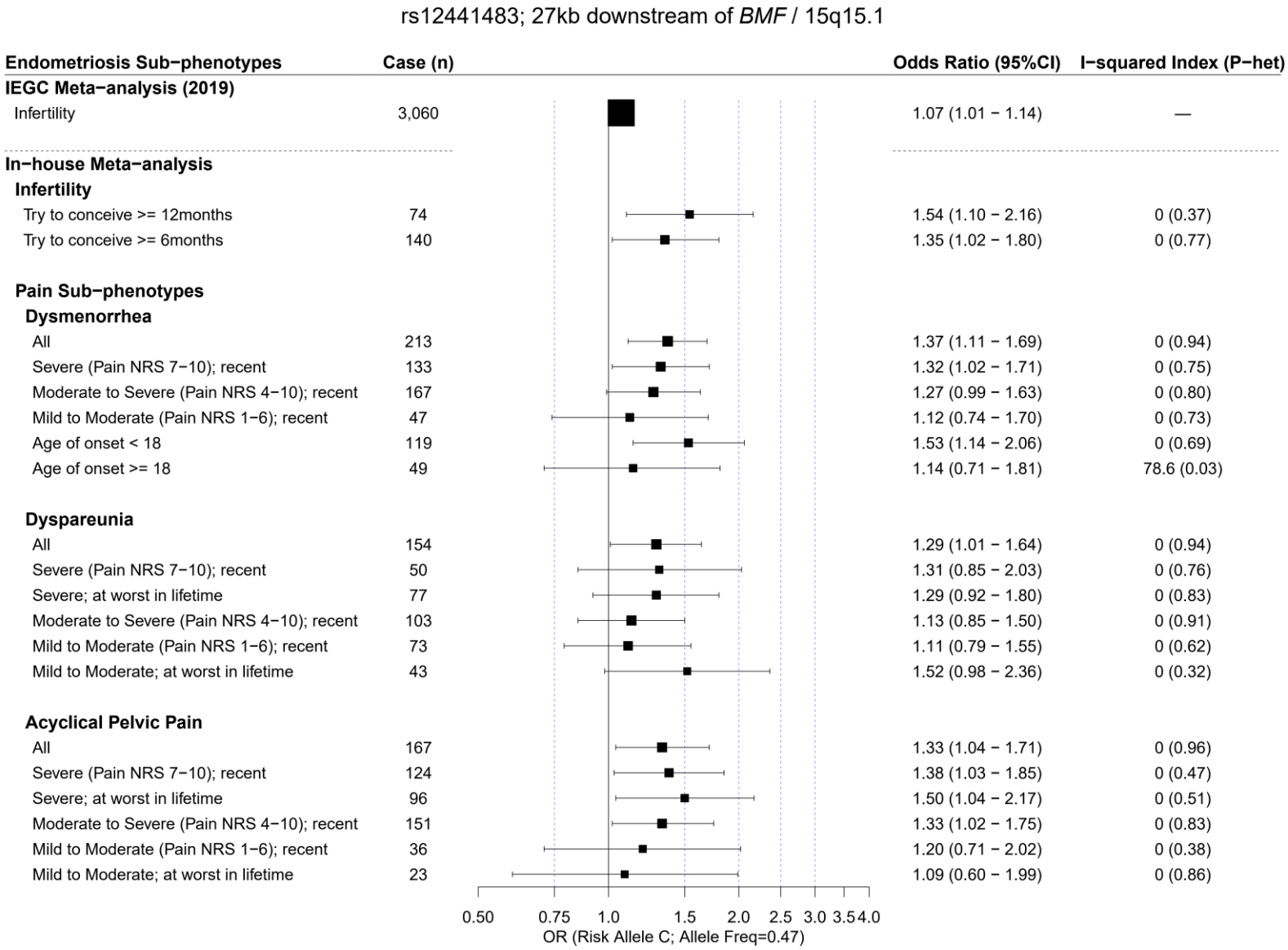


Figure 3.15 Forest plot for effect estimates of the C allele of rs12441483 on 15q15.1, downstream of *BMF*, for sensitivity testing of infertility, dysmenorrhea, dyspareunia, and acyclical pelvic pain sub-phenotypes associations.

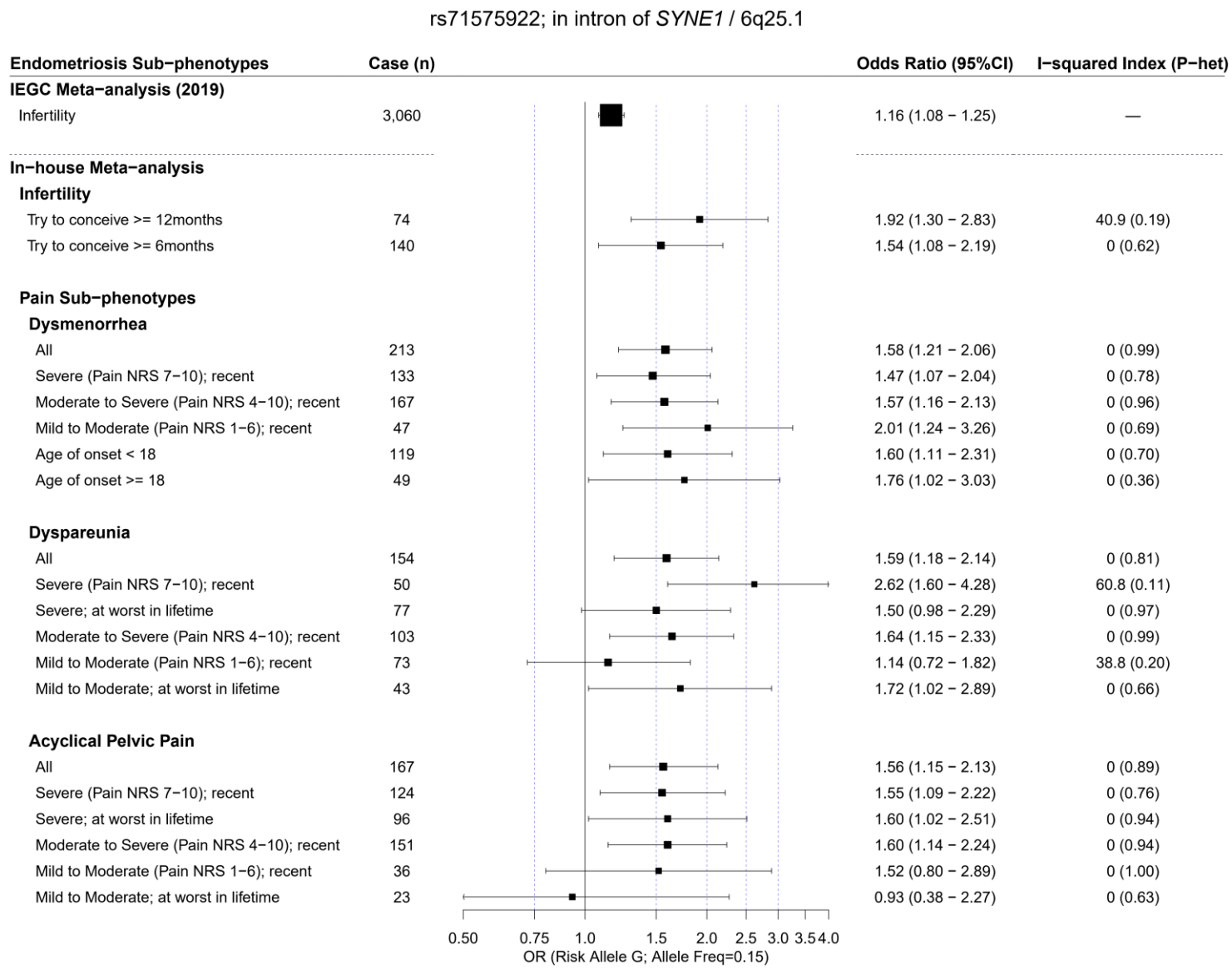


Figure 3.16 Forest plot for effect estimates of the G allele of rs71575922 on 6q25.1, in the intron of *SYNE1*, for sensitivity testing of infertility, dysmenorrhea, dyspareunia, and acyclical pelvic pain sub-phenotypes associations.

3.3.iii Investigation of missingness in dyspareunia severity reporting by participants

The variables designating dyspareunia severity appeared to show a higher level of missingness than other variables; out of 154 women who reported dyspareunia regardless of severity or temporality (all dyspareunia), 78% (n=123) gave a pain NRS for dyspareunia (experienced recently), compared to 96% for acyclical pelvic pain. The order and design of the WERF-EPHect questions around the subject of dyspareunia prevented women from reporting the severity of dyspareunia experienced during the last time of vaginal intercourse/penetration (Q. C20) if they did not have intercourse in the last 12 months (Supplementary Figure S 3.18). A very small number of participants, six women, indicated pelvic pain as a reason for intercourse avoidance in the last 12 months (Q. C17) and therefore did not have not report recent pain NRS. The data for these women could not be included in the association testing for severity of dyspareunia in the last 12 months, despite these clearly being severe cases.

Analysis of the pattern of missingness for this variable (missing of n=25) showed no statistical differences between women who experienced (overall) dyspareunia and reported a pain NRS to those who did not, in terms of their dysmenorrhoea severity (χ^2 with Yates's correction=3.5, p=0.06, Supplementary Table S 3.10), acyclical pelvic pain severity (χ^2_{Yates}=3.5, p=0.06), total number of pain types ≥ 3 (χ^2_{Yates}=1.9, p=0.17), or age group (χ^2=7.9, p=0.09). It therefore seemed unlikely that differential missingness and selection bias played a role in the observed inconsistent results.

3.4 Discussion

This study was intended as a pilot to investigate association of genetic variants established for endometriosis with specific disease sub-phenotypes. It informed and contributed to a larger sub-phenotypic meta-analysis by the IEGC, involving other WERF-EPHect datasets, the first of its kind in the field (manuscript in preparation). The utilisation of detailed sub-phenotypic data in association with genetic data was intended to provide further biological insights into the genetic associations observed with ASRM stage III/IV endometriosis, with particular emphasis on surgical and symptomatic sub-phenotypes. There were two, linked, key challenges in this analysis. First, the relatively small sample size for each sub-phenotype group; second, the large overlap between most sub-phenotypes that makes interpretation of results of differential genetic signals rather difficult. Despite this, the results give preliminary insights into several loci that may predispose more strongly to specific surgical or pain sub-phenotype, which can be explored more fully in larger datasets including that of the IEGC.

The array of symptoms experienced by endometriosis patients is very heterogenous. Moreover, previous studies[52] have shown it is difficult to correlate pain symptoms with ASRM surgical staging or infertility; the weak correlations were also observed in this dataset (Figure 3.4). The Venn diagrams with sub-phenotypic overlap in this dataset (Figure 3.2-Figure 3.3) demonstrated the complexity of the comorbidity issue. Women commonly experience >3 pain types, but those that do appear less likely to also have infertility. This could be due to selection biases at play in clinical patient groups sampled from a secondary/tertiary referral centre. The symptomatic prevalence rates mirror data reported in the literature: depending on the pain type, 45-79%[44] of endometriosis patients reported it as one of the symptoms leading to diagnosis, whereas the proportion of infertility cases was reported to be around 30-50%[184].

The correlation statistics of variants' effect sizes between sub-phenotypes shown in Section 3.3.ii.b (Supplementary Figure S 3.12) will be inflated if there is a substantial overlap and existing correlation between the sub-phenotypic pairs (Figure 3.4). The differences in the trend of correlations shown in infertility and dysmenorrhea with three different surgical sub-phenotypes beckons further investigations in mutually exclusive sample groups in future. In mutually exclusive sub-phenotypes, such as stage III/IV versus stage I/II disease (Figure 3.5, r^2=0.26), the lower the correlation of effect sizes of SNPs, the more distinctive are the biological pathways underlying the mechanisms of such sub-phenotype pairs. Outlier data points in the correlation graphs with high effect sizes mostly were not significantly associated with the sub-phenotype (with 95% CI crossing one). This is because most had a lower minor allele frequency in the tested population (e.g. MAF<5%) which can show a higher effect estimate in association testing, e.g. rs73241342 (Xq26.3/*PLAC1*) with MAF_{EUR} 4%.

The dichotomisation of age group for onset of dysmenorrhea at age of 18 applied in sensitivity testing (Figure 3.6) was based on an arbitrary cut-off, without taking into account the age of menarche. Thus, the differences shown between these two groups cannot simply infer biological meaning such as primary or secondary dysmenorrhea. To better investigate the differences in pattern of dysmenorrhoea onset shown in loci such as rs12441483 (15q15.1/*BMF*), rs10983311 (9q33.1/*ASTN2*), and rs71575922 (6q25.1/*SYNE1*) through association testing, a variable more representative of the temporal aspect of period-associated pain experience would be more informative.

The analysis of dyspareunia was hindered by higher rates of missingness. Apart from the question design (Section 3.3.iii), dyspareunia is arguably the only pain type one can avoid through abstinence of intercourse. An amendment to the order of questions is advised (i.e. Q. C20 before Q. C17) to capture the most recent dyspareunia NRS before asking if patients have skipped intercourse because of pain. An additional limitation is that other pain types such as dysmenorrhea and acyclical pelvic pain are managed by taking hormones or painkillers. Due to the cross-sectional questionnaire design, patients' pain severity before and after taking medication could not be captured; this is a potentially important effect modifier that could have impacted the results. There is also limitation in infertility assessment by capturing only the self-reported duration of trying to conceive, but without taking women's current reproductive status and child wishes into account[45].

Future study of much larger, mutually exclusive sub-phenotypic groups in endometriosis is the key to validation of the results. The difficulty would be in obtaining a relatively large dataset to consider only cases with distinct sub-phenotypes to achieve sufficient power. The mutually exclusive surgical sub-phenotypes would be easier to achieve than symptomatic sub-phenotypes due to commonness of comorbid pain experienced by endometriosis patients and potential missing information for traits such as infertility.

Previous studies[49,50] have shown that ASRM stage III/IV disease is more genetically driven than stage I/II. One of the aims of this analysis was to see whether this observation can be explained by and further dissected into associations with different surgical or other sub-phenotypes. The ASRM scoring algorithm is heavily biased towards having an ovarian lesion (with 20 points), meaning that almost all cases with endometrioma are classified as at least stage III (point 16+). The deep lesion definition used in this study refers to both the lesion under peritoneum (>5mm)[58,59] of any size or a nodule traditionally termed as 'deeply infiltrating endometriosis'; thus, the scoring for a single deep lesion can range from stage I to IV.

Despite the overlap between surgical sub-phenotypes, especially the occurrence of endometrioma and deep lesions, the variant rs11674184 in the *GREB1* locus exhibits nominal association with endometrioma (endometriosis cases with, but not exclusive to, ovarian lesion), but not with superficial or deep lesion sub-phenotypes, with a larger effect size (OR=1.65) than that of stage III/IV disease (OR=1.48). Given that only 10 cases were diagnosed with endometrioma as the only surgical finding among the endometrioma sub-phenotype group, and others had coexisting lesions or nodules, this signal could not be further investigated in this group alone. Nevertheless, a hypothesis is that the rs11674184/*GREB1* locus predisposes in particular to endometrioma with stage III/IV disease — a proxy for this ovarian subtype[10-13]. The higher effect shown in stage III/IV than in stage I/II disease at this locus is in line with previous reports[50,135]. *GREB1* encodes protein responding to the oestrogen receptor (ER)-regulated pathway.

A similar pattern can be found for another variant, rs1903068, in the 5' untranslated region (UTR) of *KDR*. Effect estimates for this variant are greatest for endometrioma (OR=1.61, Table 3.5), followed by deep lesions (OR=1.53), and then superficial lesions (OR=1.44); the latter effect size is similar to stage III/IV disease (OR=1.40). It suggests that the *KDR* locus may also predispose endometrioma over other types of lesion, and could drive the higher effects shown in stage III/IV in both IEGC meta-analysis[135] and in-house result. Conversely, the variant rs71575922 in *SYNE1* shows nominal association with all surgical sub-phenotypes apart from endometrioma. The effect observed in our data appears larger for stage I/II (OR= 1.76, p=7.26x10^{-5} passing the multiple-correction threshold of 9.06x10^{-5}, Table 3.4) compared with stage III/IV disease (OR=1.53, p=0.01), which is reverse from that suggested by previous results based on much larger sample sizes[135], but the 95% CIs of our estimates mostly overlap. The results of our study suggest that the *SYNE1* variant does not predispose to endometrioma in particular, but more clarity is needed for its effect on other surgical sub-phenotypes.

For genetic association testing of symptomatic sub-phenotypes, respective case groups were tested against a control set that included a mixture of in-house clinical (n=79), biobank clinical (n=2,420), and population controls (n=2433+347=2,780). The biobank clinical controls (46%) were initially included in the GWAS analysis to allow detection of signals driven solely by endometriosis, but not symptom-related traits. Participants included as in-house clinical controls for endometriosis (1.5%=79/5279) may have experienced symptoms similar to the case group, and thus possibly counteracting signals detected for symptom of interest. Similarly, some of the participants in the biobank as population controls may also exhibit symptoms, which further hinders signal detection for symptom-specific traits. However, no differences were observed in effect sizes depending on different tiers of control groups (Supplementary Table S 3.1).

The two variants rs12441483 in 3' UTR of *BMF* and rs71575922 in *SYNE1* associated with infertility sub-phenotypes also show nominal association with most tested pain types. The confidence in identifying these two loci as infertility-related loci is limited due to the extensive overlaps between infertility and three main pain types of endometriosis in the dataset (shown in Figure 3.3b). Only five infertility cases were free from any of the three pain types. Nevertheless, it can be argued that the effect estimates for infertility in both loci (OR_{BMF}=1.54, OR_{SYNE1}=1.92, n=74) are considerably larger than those in dysmenorrhea, dyspareunia, and acyclical pelvic pain (ORs_{BMF}: 1.29-1.37, ORs_{SYNE1}: 1.56-1.59, n=154-213); nevertheless, the accuracy of these effect estimates is influenced by the relatively small sample sizes, and caution is warranted in the interpretation of their relative magnitude.

With substantial overlap between five tested pain types (Figure 3.3a), the opportunity for dissecting signals associated with particular pain types is relatively low. However, several loci showed evidence of contributing to general pain experience across pain sub-phenotypes, such as the variants rs12441483/*BMF* and rs71575922/*SYNE1*. Both show nominal association with similar magnitude in effects across all pain types (except for bladder pain) (ORs_{BMF}: 1.29-1.43, ORs_{SYNE1}: 1.50-1.59). These two loci also show nominal associations with reporting of ≥ 3 pain types, which add evidence to their possible roles in pain hyperalgesia. Several loci showed nominal association with only one pain type, such as variant rs10983311/*ASTN2* and rs1903068/*KDR* locus for dysmenorrhea. However, due to the overlap between dysmenorrhoea with other pain types (Figure 3.3a), it remains unclear to what extent there are dyspareunia or general pelvic pain signals. The *KDR* locus is also nominally associated with the reporting of ≥ 3 pain types with similar OR (1.37) to dysmenorrhea (OR=1.32, Table 3.5), suggesting that this locus may predispose to multiple pain sub-phenotypes. *KDR* encodes one of the two receptors for vascular endothelial growth factor (VEGF) – VEGFR2.

Another variant, rs2510770 intronic to *PDLIM5*, showed a seemingly protective effect for endometriosis with dysmenorrhea. *PDLIM5* encodes a protein involved in the morphology of dendritic spines at synapses[185], and therefore a prime candidate for pain association. This locus was identified in the recent IEGC GWAS[135] with a relatively small risk effect estimate of 1.04 for overall endometriosis. In our dataset, it showed an opposite effect (OR<1) for most of the tested sub-phenotype groups, as well as overall endometriosis cases. QC pipelines were checked to make sure an error was not made in data processing, but no such error was detected. Its nominal association with dysmenorrhea within endometriosis cases therefore did not seem implausible, and could be a consequence of selection bias, or chance. Interestingly, however, the 'protective' effect of this variant is also observed in the larger IEGC meta-analysis dataset for the other two pain types (dyspareunia OR=0.86, 95% CI: 0.72-1.03, I^2=0, P_{het}=0.47, and acyclic pelvic pain OR=0.92, 95% CI: 0.77-1.10, I^2=0, P_{het}=0.83) from a dataset with 393-547 cases and 40,852 controls with consistent direction effect across all five studies), prompting future investigation of its potential neuroprotection effect in endometriosis. Interestingly, rs2510770 (RA_{ENDO}: A) associated with endometriosis is in high linkage disequilibrium (LD) (r^2_{EUR}=0.98) with a GWAS hit for uterine fibroids rs4699299 (RA_{UF}: T)[186], but with opposite direction of effects with the linked allele.

The definition of bladder pain used in this phenotypic dataset was relatively non-specific, and insufficient to distinguish interstitial cystitis from pain induced by urinary tract infection[187]. For the 107 cases of bladder pain included in the analyses, no associations were observed for any of the 46 tested genetic variants. However, the result contributed to the IEGC sub-phenotypic meta-analysis and yielded five associations with tested SNPs in the larger dataset (paper in preparation).

Having briefly highlighted the biology of some of the most interesting loci above, the following will focus on the genetic variants and the evidence for the genes and the transcription of which they may perturb. Among 16 SNPs showing at least one nominal association with tested sub-phenotypes, the six loci (including 8 SNPs) reviewed are: three are involved in sex hormone related pathways (*GREB1*, *FSHB*, *SYNE1*), one linked to angiogenesis (*KDR*), and the last two are neural related (*ASTN2*, *BMF*). All three areas are described in proposed mechanisms for histogenesis and maintenance of lesions in endometriosis[34].

GREB1/2p25.1: rs11674184

The *GREB1* locus was one of the earliest loci reported in association with endometriosis[136]. The lead SNP from the latest GWAS meta-analysis is rs11674184 which is intronic to Growth Regulating Oestrogen Receptor Binding 1 (*GREB1*). It is the only variant among all those tested that showed statistical difference in association with stage III/IV disease when compared to stage I/II nominally. It also showed nominal association with the endometrioma sub-phenotype, but no other surgical sub-phenotypes, with a larger effect size than the stage III/IV disease. This suggests contribution of the endometrioma sub-phenotype to the higher genetic loading previously observed for stage III/IV at this locus.

GREB1 encodes three protein isoforms[188] with different levels of response to the oestrogen receptor (ER)-regulated pathway. It is traditionally identified as a moderator for hormone-dependent cancer growth such as in ERα+ breast cancer[189]. The variant was hypothesised to work on *GREB1,* and with other cofactors or genes in the ER-DNA transcription complex[190,191] for its pathogenic role through increasing oestrogen sensitivity of endometriotic lesions[141]. This SNP is in weak LD with two other SNPs associated with uterine fibroids in the same locus: rs35417544 (r^2_{EUR}=0.21, RA_{UF}=T)[186] and rs10929757 (r^2_{EUR}=0.29, RA_{UF}=C)[192], both risk alleles are linked to risk allele (T) for endometriosis.

FSHB and *ARL14EP*/11p14.1: rs7924571

The intergenic variant rs7924571 is located in the 3' UTR of *WT1* (Wilms' Tumor 1) gene, and sits 2Mb downstream of *FSHB* and *ARL14EP*. There is another independent signal, rs3858429, at 1kb upstream of *ARL14EP* (Table 3.1). This intergenic region is often referred as '*FSHB*/11p14.1 region', and this reference is used throughout the chapter. The SNP near to *WT1* shows nominal association with endometrioma sub-phenotype only in this dataset.

The *FSHB* gene is known for its role in the female hormonal axis through encoding the beta subunit of follicle-stimulating hormone (FSH). *WT1* is named after a rare kidney cancer – Wilms' Tumor, encoding a transcription factor curial for urogenital development[193], including ovaries in female. The protein is also involved in embryonic and induced pluripotent stem cell (iPSC) differentiation pathway. There are reports of GWAS hits for uterine fibroids in the *WT1* locus[186,194] and for endometrial cancer in the nearby *WT1-AS* locus[195], though none of them are in LD with the lead variant for endometriosis. The endometriosis association from the *FSHB*/11p14.1 region was first reported in 2017[141] with tag SNP (rs74485684) which is in strong LD (r^2_{EUR}=1.00) with the variant near *ARL14EP* (rs3858429). It shows nominal association only with ASRM stage III/IV disease in the ENDOX dataset; the result of IEGC sub-phenotypic meta-analysis, with additional datasets from collaborators, shows nominal association with the endometrioma sub-phenotype (OR_{IEGC}=1.21, 95% CI_{IEGC}: 1.01-1.46; OR_{ENDOX}=1.72, 95% CI_{ENDOX}: 0.96-3.05), similar to the *WT1* loci mentioned above.

The *ARL14EP* encodes an effector protein that interacts with molecules in the ARL14 – MYO1E pathway to control transportation of major histocompatibility class II molecule (MHC-II) in dendritic cells via the actin architecture[196]. Currently, there is limited knowledge of the gene function directly related to the reproductive system; however, signals in the *ARL14EP* locus are commonly reported in GWAS studies for gynaecological factors, such as age at menopause[197], FSH and luteinizing hormone (LH) measurements[198], etc. Interestingly, the index SNP for endometriosis in *ARL14EP* locus is in high LD with two SNPs in the locus associated with heavy menstrual bleeding (HMB)[186] (rs11031005, r^2_{EUR}= 0.82, RA_{HMB}: T) and polycystic ovarian syndrome (PCOS)[199] (rs11031006, r^2_{EUR}=0.82, RA_{PCOS}: G), of which both risk alleles are linked to the risk allele for endometriosis (C). The first SNP is also associated to PCOS[200] with the same risk allele as HMB, where the second SNP was reported to be associated with overall endometriosis in Greek population by genotyping[201]. It suggests a possible sharing of biological pathways, perturbed by these variants, between these conditions.

SYNE1 and *CCDC170*/ 6q25.1: rs71575922 and rs7740302

Two SNPs, rs71575922 intronic to *SYNE1* (lead signal) and rs7740302 intronic to *CCDC170* (a secondary independent signal), were both included for association testing. They are 700kb apart, sitting at either side of the *ESR1* (Oestrogen Receptor 1) gene on chromosome 6. Both loci were first mentioned in 2017; the study[141] reported one association signal in the *CCDC170* locus, two (one secondary signal) in *SYNE1*, and another secondary signal for the *ESR1* locus. Similarly, in the latest IEGC GWAS meta-analysis, there are seven independent signals in total across the *RMND1--CCDC170--ESR1--SYNE1* genomic region revealed upon conditional analysis, suggesting possible complex regulatory or transcriptional interactions within the area.

The SNP in *CCDC170* locus shows only one nominal association with stage III/IV disease, whereas the variant in *SYNE1* shows multiple associations to almost all surgical, infertility and pain sub-phenotypes (some passing the least stringent multiple-correction significant threshold), except for endometrioma and bladder pain. The effect sizes of these associations with risk allele G at rs71575922 are high despite the relatively common allele frequency of 15% in the European population.

The genomic region of interest centres on *ESR1*, a well-studied gene in the female hormonal system, especially for oestrogen-dependent gynaecological cancers. Endometriosis is also an oestrogen-dependent condition; changes in oestrogen receptivity of normal and diseased tissues would alter the effect of circulating oestrogen on these tissues, with wide-range of properties on growth factors[202], inflammation[203] and pain modulations[204]. *ESR1* encodes the nuclear ERα, which not only binds oestrogen, but also DNA for transcriptional regulation[205]. The protein is also involved in the LKB1 – mTOR signalling pathway[206,207] that plays key role in cell growth and metabolism[208].

SYNE1 is a very large size gene (~500kb), which encodes a protein (with >8k amino acids) localised to nuclear membrane for maintaining spatial organisation of organelles within the cell[209]; it is expressed in many tissues, especially in Purkinje cells for neural signalling. The gene is historically associated with various neurodegenerative disorders (mainly cerebellar ataxia and muscular dystropy) due to mutations transmitted in families[209,210]. The SNP in the *SYNE1* locus for endometriosis (rs71575922) is also reported to be associated with uterine fibroids without the coexisting presence of endometriosis[192], sharing the same risk allele G. It is in high LD (r^2_{EUR}=0.91) with another variant rs58415480 associated with uterine fibroids (overall), though the report of which allele (G/C) is the risk allele for the condition is conflicting[186,192].

The function of *CCDC170*, however, is not known, nor the role of its encoded protein. Nevertheless, there is a report of increased breast cancer susceptibility through genetic alterations in *CCDC170*, but not in its neighbouring *ESR1*, confirmed by differential allele-specific expression analysis on a human mammary epithelial cell line[211]. According to the NHGRI-EBI GWAS catalog[212], there are eight GWAS variants in the gene region associated with breast carcinoma in the European population (rs2046210, rs3757318, rs3734805, rs12662670, rs11155804, rs9397437, rs3757322, and rs60954078; https://www.ebi.ac.uk/gwas/, accessed at 07 May 2020); however, none of them is in linkage disequilibrium with the endometriosis associated SNP rs7740302 (r^2_{EUR}: 0.001-0.36).

***KDR*/4q12: rs1903068**

The association of endometriosis with the *KDR* locus was first reported in an Icelandic GWAS in 2016[140]. The SNP in our study (rs1903068, 5' UTR of *KDR*) is in high LD (r^2_{EUR}=1.00) with the one in this first report, and shows nominal association with stage III/IV, all surgical sub-phenotypes (superficial, deep, and ovarian lesions), dysmenorrhoea, and ≥ 3 pain types. The gene encodes VEGFR2; The soluble and insoluble, transmembrane form of this receptor work antagonistically in the VEGF signalling pathway[140,213]. Expression of receptors are traditionally thought to be the prime event for endothelial cell differentiation, and for VEGF binding throughout[214,215]. Apart from a role in angiogenesis[216,217], VEGF also help mediating immuno-inflammation[218] as well as reorganisation of cytoskeleton and sub-cellular structure[219]. There is evidence of VEGF modulation in pain via inflammatory and neuropathic mechanisms such as in osteoarthritis[220], and in chemotherapy drug-induced[221] or diabetic neuropathy[222], respectively.

One study has reported higher expressions of both VEGFR2 (the protein encoded by *KDR* gene) and VEGF in glandular epithelial and stromal cells of endometrium in the presence of combinations of steroid hormones[223]; while another study[224] proposed that the regulation of *KDR* expression in endometrial capillaries (i.e. endothelial cells) is stimulated by oestrogen via paracrine action from VEGF. Furthermore, up-regulation of VEGFR2 is reported in endothelial cells of eutopic endometrium (in secretory phase) from endometriosis patients when compared to those without the disease[225]; the same was described in ectopic endometriotic lesions with high proliferative activity compared to lesions with lower activity[225]. Apart from the receptor itself, a higher VEGF expression in glandular epithelial cells of eutopic endometrium (secretory) has been found in endometriosis cases versus controls from the same study, highlighting its potential role in pathogenesis of endometriosis.

Among 46 SNPs studied in this analysis, the *KDR* locus is the only one showing nominal associations with all surgical sub-phenotypes; the function of this locus in establishment and maintenance of endometriotic lesions at different sites is worthy of further functional investigations.

***ASTN2*/9q33.1: rs10983311**

The tag SNP, rs10983311, in the intron of *ASTN2* gene was first reported in the latest IEGC study; it shows nominal association not only with surgical subtypes (superficial and ovarian lesions), but also with two common pain types in endometriosis, i.e. dysmenorrhea and dyspareunia (mild to moderate). The gene encodes a protein presence in neurons facilitating their migration in cerebellum with its counterpart ASTN1[226]. The lead SNP is in weak LD (r^2_{EUR}=0.19) with a popular variant (rs6478241) reported in association with migraine[227,228] – a pain condition commonly reported by endometriosis patients[229]. However, the associated alleles do not have the same direction in effect in the two conditions. The migraine associated variant rs6478241 was not associated with endometriosis in the recent conditional analysis in the IEGC study. A previous study[230] has shown shared molecular genetic evidence between endometriosis and migraine. Further conditional analysis in the IEGC dataset can be done to determine if the pair of SNPs are tagging the same locus, thus suggesting a shared underlying biological mechanism.

***BMF*/15q15.1: rs12441483**

The index SNP rs12441483 in the 3' UTR of *BMF* is intronic to a long non-coding RNA (lncRNA) called *SRP14-AS1* (Signal Recognition Particle 14 Anti-sense RNA1). The nearest gene *BMF* encodes a protein with an apoptotic activation domain, which cascade cell death signals in the event of stress[231,232]. The locus was firstly described in our IEGC preprint in 2018[135], reporting lead SNP rs4923850, which is in high LD (r^2_{EUR}=0.99) with the index SNP used in this study; it shows nominal association with superficial endometriosis, endometrioma, infertility, and almost all tested pain sub-phenotypes, except for bladder pain.

Not much is known about how *SRP14-AS1* work on biological pathways, however, the tag SNP is predicted to alter the motif of a regulatory element, Fetal Alz-50 Clone 1 (FAC1), a DNA binding protein traditionally known to be involved in neurological development[233] and neurodegenerative disease such as Alzheimer's[234]. The prediction is retrieved by HaploReg v4.1[235] and a web-based tool for regulatory motif discovery in ENCODE transcription factor binding database[236]. In addition, four variants in LD (r^2_{EUR}>0.75) with the index SNP are eQTL hits (eSNPs) for nerve (tibial) correlated to *SRP14-AS1* reported in Genotype-Tissue Expression Project (GTEx) v6[237]. They are rs2412474 (r^2=0.75, p-value for $eSNP_{nerve}$ =4.5x10^{-17}), rs7181985 (r^2=0.84, p=4.5x10^{-12}), rs4924414 (r^2=0.82, p=2.3x10^{-12}), and rs7162269 (r^2=0.89, p=2.8x10^{-12}). The above two observations suggest that this locus may play a role in neurological function.

There are two additional variants worth noting, rs7181230[238] and rs28620926[239], both in LD (r^2_{EUR}=0.48) with the endometriosis SNP and associated with serum concentration of dehydroepiandrosterone sulphate (DHEAS). DHEAS is a sulfate ester of DHEA, the most abundant neurosteroid in the body produced by adrenal glands acting as neurotrophin[240], weak androgen (role in puberty), and precursor for oestrogen production in female[241]. It works with nerve growth factor (NGF) and brain-derived neurotrophic factor (BDNF) for its neurotrophic effect in neurogenesis and differentiation[240]. Although the effect direction for endometriosis and DHEAS serum level of linked alleles are opposite, the moderate LD between the SNPs may suggest a potential role of this locus in hormonal pathways for endometriosis in addition to its neurological functionality.

3.5 Conclusion

The data structure of surgical and symptomatic sub-phenotypes of endometriosis is complex with a high degree of overlap. The sub-phenotype association testing of genetic variants revealed several SNPs in nominal association with ASRM stage I/II (rs71575922/6q25.1/*SYNE1*), stage III/IV disease (rs11674184/2p25.1/*GREB1*), endometrioma (rs11674184/2p25.1/*GREB1*, rs1903068/4q12/*KDR*), and multiple pain sub-phenotypes (rs12441483/15q15.1/*BMF*, rs71575922/6q25.1/*SYNE1*, and rs1903068/4q12/*KDR*) for endometriosis. The analysis was subsequently included in the IEGC meta-analysis of GWAS on endometriosis sub-phenotypes (manuscript in preparation).

Suggestions on modification of questions to better capture self-reported pain severity data using clinical questionnaires are also made. The association and correlation results should be validated in larger and ideally mutually exclusive sub-phenotype datasets in the future.

Chapter 4 The endometrium transcriptome across menstrual cycle phases

4.1 Introduction

The endometrium, inner lining of the uterus, is a dynamic tissue constantly responding to hormonal changes throughout different phases of the menstrual cycle during women's reproductive years. A typical menstrual cycle last for 28 days; day 1 is defined as the start of lining shedding. The first 7 days of shedding is referred as the menstrual phase. Day 8 to 13 are referred to as proliferative phase designating the endometrial status, or follicular phase, designating the status of the ovary. Day 14 is the ovulation/transition day, followed by the secretory or luteal phase (from day 15 to 28); as outlined in more detail in Chapter 1 Section 1.7. As the physiological functionality of the endometrium tissue changes in different cycle phases, from shedding, lining renewal, to preparing for embryo implantation, its transcriptome is hypothesised to change throughout the cycle.

The understanding of endometrial transcriptomic changes is crucial for the investigation of endometrium-related conditions, such as endometriosis, implantation failure (infertility), endometrial hyperplasia, and endometrial cancer. This is because differential expression in endometrium is often used to study disease vs. health states, but such analyses require adjustment for the considerable cyclic difference when comparing and matching of diseased tissues to control tissues. Previous efforts[82,85,86] to understand transcriptomic changes in normal human endometrium have shown substantial changes in transcriptomic signals throughout the cycle using microarray technology (Section 1.7).

To date, only one study by Fung et al.[86] explored the transcriptome of 123 eutopic endometrium samples from women with endometriosis and gynaecological conditions using microarray, and provided detailed annotated gene lists of their findings. In this chapter, the transcriptome of human eutopic endometrium is characterised across cycle phases using genome-wide RNA-sequencing (RNA-seq) data to see how the finding agrees with previous record. Investigating the utility of transcriptomic profiling to adjust for the cycle phase effect would also help addressing this important covariate in subsequent omics analysis for endometriosis.

The aims of the work described in this Chapter were to: 1) investigate the utility of transcriptomic profiling, histology and self-reported data in cycle phase discrimination; 2) compare two analysis pipelines for differential gene expression in RNA-seq; 3) conduct differential expression (DE) analysis between cycle phases, and compare results with those reported in previous studies[82,86]; 4) evaluate the potential for cycle phase prediction through a multivariate model based on differentially expressed genes; and 5) conduct pathway analysis to determine biological pathways enriched for the transcriptome changes involved in cycle phase physiology.

4.2 Materials and methods

4.2.i Sample selection, sample QC, and preparation of RNA-seq libraries

The samples analysed in this chapter were part of an Medical Research Council, UK (MRC)-funded study investigating genetic drivers of expression phenotypes in eutopic and ectopic endometrium. A total of 195 frozen eutopic endometrial pipelle biopies from women with or without endometriosis, who did not use hormones in the three months prior to recruitment were selected from the ENDOX study (Section 2.2.i). Samples were taken across the menstrual cycle (day 1 to 3 samples in the menstrual phase were excluded). RNA extraction was outsourced to Hologic Ltd., and samples were then sent to Oxford Genomics Centre (OGC) at the Wellcome Centre for Human Genetics (WCHG) for RNA sequencing.

Samples were QC-ed to quantify amount of RNA obtained and assess quality. RNA integrity number (RIN) values were not available for all samples, however, 81% (n=153) were reported with RIN from 2.0 to 9.6 (mean = 6.6; 10 samples with RIN ≤ 3). No RIN cut-off for samples was applied on the basis that sequencing can still sometimes work well for lower-quality samples, and poor samples could be excluded at the data analysis stage. Following ribosomal RNA depletion (theory in Section 1.10) of total RNA, libraries were prepared; sequencing (100bp paired-end reads) was performed on Illumina® HiSeq4000 machine.

The average sequencing depth was 36.5 million (paired-end) reads per sample. Reads were aligned to human reference genome GRCh37/hg19 using the splice-aware aligner `Hisat2` Gene annotation files were downloaded in GTF format from Ensembl release 76. Reads mapping to annotated gene features were quantified with `featureCounts`, part of `subread`, using default parameters. On average, more than 90% of reads were mapped by the alignment program. Only uniquely mapped reads were considered by `featureCounts`, and a raw gene count table with 195 samples and their respective gene counts for 63,794 Ensembl genes was produced as output. The steps described above were performed by Dr John Broxholme of the Bioinformatics Core at WCHG with pipelines developed and implemented by Dr John Broxholme, Dr Eshita Sharma, and Dr Ben Wright.

4.2.ii Cycle phase determination of eutopic endometrium

4.2.ii.a Histological dating/staging of cycle phase

Histological dating of menstrual stages was performed on all eutopic endometrium samples with excess frozen tissue after RNA-seq data generation. To produce histology slides, three to six frozen sections were obtained at different depth of the pipelle at 6-7 μm per sample for showing of endometrial glands (i.e. the stratum functionalis) for dating purposes. The sections were stained with haematoxylin and eosin and scanned by the Aperio slide scanner (Leica) for digital record. One of the five cycle phases (menstrual, proliferative, early-secretory, mid-secretory and late-secretory) were assigned on the digital slide records (Aperio ImageScope v12.3.2) in a double-blinded manner (i.e. without the knowledge of the self-reported menstrual cycle day/length by the patient) and independently by myself and Dr Karin Hellner, consultant gynaecologist in our research group. We both received extensive training from Dr Sanjiv Manek, gynaecological pathologist at the Oxford University Hospitals Foundation Trust for the staging skill. Dated results were compared, and for those that were discrepant, the physical slides were sent to an independent pathologist with extensive experience in cycle phase dating for final decision.

4.2.ii.b Self-reported data derived cycle phase

Self-reported menstrual history was recorded in the ENDOX patient questionnaire in compliance with WERF-EPHect guidelines[55]. The following information was considered for assigning self-reported cycle phase of the patient on the day of surgery when the endometrial pipelle was taken: 1) average cycle length in the last 3 months; 2) date of the first day of last menstrual period (LMP); 3) date of sample taken to calculate current cycle day; 4) average number of days of bleeding, and 5) regular or irregular periods in past 3 months.

Calculation of menstrual cycle phase adjusted for cycle length[78] is based on the assumption that: the first 3 days in a cycle are the menstrual phase; the proliferative phase fluctuates with shortening/lengthening cycle length/bleeding days; and that the secretory phase is kept constant to 14 days (5, 3 and 6 days for early, mid and late-secretory respectively). The typical breakdown of the cycle phases for a regular 28-day cycle is given in Table 4.1 along with assignment of phases based on varying cycle-length.

For women reporting average cycle length within a range (e.g 24 to 30 day-cycle), the phases were calculated using both the lower and upper bounds and a phase was assigned if both were the same. For those with missing data for cycle length, or regularity of cycles, the cycle was assumed to be a 28 day-cycle. For women reporting irregularity in their cycles, no cycle phase was assigned using their self-reported data. Menstrual cycle phase of the samples with both histology-confirmed and self-reported data were compared.

Table 4.1 Assigning self-reported cycle phase to samples based on varying cycle-length.

Cycle Length	**Corresponding Phases for Cycle Day (as at Surgery)**					
21 to 35 day-cycle	**Menstrual**	**Menstrual/Proliferative (Informed by Length of Bleeding Days)**	**Proliferative**	**Early-secretory**	**Mid-secretory**	**Late-secretory**
21	1-3	4-7	—	8-12	13-15	16-21
22	1-3	4-7	8	9-13	14-16	17-22
23	1-3	4-7	8-9	10-14	15-17	18-23
24	1-3	4-7	8-10	11-15	16-18	19-24
25	1-3	4-7	8-11	12-16	17-19	20-25
26	1-3	4-7	8-12	13-17	18-20	21-26
27	1-3	4-7	8-13	14-18	19-21	22-27
28	**1-3**	**4-7**	**8-14**	**15-19**	**20-22**	**23-28**
29	1-3	4-7	8-15	16-20	21-23	24-29
30	1-3	4-7	8-16	17-21	22-24	25-30
31	1-3	4-7	8-17	18-22	23-25	26-31
32	1-3	4-7	8-18	19-23	24-26	27-32
33	1-3	4-7	8-19	20-24	25-27	28-33
34	1-3	4-7	8-20	21-25	26-28	29-34
35	1-3	4-7	8-21	22-26	27-29	30-35

4.2.ii.c Data QC, gene filtering, and visualisation of transcriptomic profiles of samples

The following parameters were required for each sample to be retained for further analysis : 1) >5 million reads mapped to Ensembl genes; 2) <20% of reads mapped to top 100 most highly expressed genes (complexity measurement); 3) >20k genes with ≥ 1 transcripts per million (TPM) reads & 4) <10 million unmapped reads. Diagnostic hierarchical clustering revealed five outliers (2.6%) to be excluded, with 190 samples passing the above QC thresholds in the final dataset (Stage I/ II: n=69, Stage III/IV: n=59 & control: n=62). The raw counts table for 190 samples was passed on for gene-level filtering.

First, only genes annotated on autosomal chromosomes and X chromosome were kept (52,190 genes). Second, low expressed genes were removed by applying a filter of counts per million (CPM) > 1 (i.e. gene expressed) in at least 80% of the samples in `edgeR`. Variance stabilisation was applied to gene counts by `getVarianceStabilizedData` function in `DESeq`. Principal component analysis (PCA) was performed on all 190 samples, on 80 samples with histological cycle phases, and separately on a subset of 56 secretory samples for visualisation of five cycle phase clusters. Clustering by sample quality (e.g. RIN, RNA concentration/ yield) and sequencing batch were not observed on PCA plots, suggesting that these factors are not major sources of variation in the data. Detailed cellular composition of endometrium were not available from this experiment, nor from public databases, to allow adjustments to be made.

To verify the result from the above `DESeq` analysis, parallel analysis using `DESeq2`, which features different initial gene filtering and variance stabilisation steps was performed, with visualisation on the subset of 80 samples with histological cycle phase assignment by PCA. Genes with less than the total of five reads across all samples with histology (n=80) were filtered out manually, with library size normalisation and variance stabilisation using the `rlog` function.

4.2.iii Differential gene expression of endometrium between cycle phases

4.2.iii.a Comparison of `edgeR` (1st) and `DESeq2` (2nd) pipelines

For the purpose of comparing two analysis pipelines, differential expression analysis was conducted in a subset of the data including only histologically-confirmed proliferative (n=23) and secretory samples (n=53). The model matrix was designed according to the histology-based cycle phase dating regardless of endometriosis case-control status (`design ~ histology phase`) and the same applied to both pipelines.

For the pipeline using `edgeR`, a gene filter of CPM>1 was applied across 76 samples prior to library size normalisation with trimmed mean of M-values (TMM) using `calcNormFactors`. Overall and gene-wise dispersion were estimated, followed by fitting of a negative binomial generalised log-linear model (GLM) using `glmLRT` for differential expression analysis with multiple testing correction.

For the second pipeline using `DESeq2`, no further manual gene filter and library size normalisation were applied after the read QC step. Dispersion estimation and GLM fitting were done through `DESeq`, a function in `DESeq2`. Automated independent filtering of genes with false discovery rate (FDR) cut off set at 5% was performed with the `result` function; its filter statistics rely on the average of normalised counts on comparison groups of interest within the model matrix provided.

In both pipelines, markers were annotated using human Ensembl GRCh37.p13/hg19 with `biomaRt`. The extent of overlap between results generated from both pipelines was visualised using Venn diagrams. For genes that were identified as differentially expressed in both analysis pipelines, the effect size ($\log_2$ fold-change ($\log_2$FC)) and significance (adjusted *p*-value (adj.*p*)) were compared. For presentation of results, the most up-to-date generic gene names (from GRCh38/hg38) are reported over and supplemented with the previous HUGO Gene Nomenclature names that GRCh37/hg19 based on; for gene identification and input for subsequent analysis, Ensembl gene IDs are always used.

4.2.iii.b Differential expression analysis between cycle phases

Subsequent analyses were based on the `edgeR` pipeline. Only samples with cycle phases determined by histology were tested for differential gene expressions between proliferative versus secretory phases, and mid-secretory versus late-secretory phases. Analyses on comparison involving menstrual phase (n=4) and early-secretory samples (n=3) were not conducted due to limited data points from these two groups. The extent to which genes were differentially expressed was visualised using volcano plots.

4.2.iv Multivariate model for cycle phase prediction

A multivariate model for prediction of proliferative vs. secretory phase of endometrium samples was built based on the differentially expressed genes informed by the `edgeR` pipeline. Differentially expressed genes included in the model building were restricted to genes with $\log_2$FC>|1| that were also reported as differentially expressed between proliferative and early-secretory phases in Fung's[86] and Talbi's[82] studies to avoid including signals from chance findings or specific to our own dataset.

Raw counts of these selected genes from all in-house histology-confirmed proliferative and secretory samples were normalised for library size and variance stabilised by `rlog` function in `DESeq2`. For model building purpose, due to the limited number of samples from the two phases (n=23 &53), all samples available were assigned to the training set, without further splitting the data to allow for an internal test set.

Single-gene logistic regression, followed by the forward and backward step-wise, and all-included logistic regression for the selected gene set was performed on data in a general linear model (GLM), assuming a single dispersion parameter in the error distribution (i.e. `family = binomial`). Correlation of expression patterns of the selected genes was visualised using a heatmap.

The models suggested were compared for goodness-of-fit and parsimony using the Akaike Information Criterion (AIC) and assessed normality of residuals from logistic regression with quantile-quantile (Q-Q) and residuals-versus-fitted plots. Performance of the model in terms of sensitivity and specificity was based on prediction from the training dataset. Receiver operating characteristic (ROC) and accuracy variant signature curves were then generated for visualisation.

4.2.v Pathway enrichment analysis on differentially expressed genes between cycle phases

Pathway enrichment analysis on differentially expressed genes between proliferative and secretory phases was performed in three different analysis platforms: Ingenuity Pathway Analysis[242] (IPA®, QIAGEN Inc.), Database for Annotation, Visualization and Integrated Discovery[243] (DAVID), and Protein ANalysis THrough Evolutionary Relationships[244,245] (PANTHER).

Only genes that were differentially expressed (DE) with adj.$p \leq 0.05$ and $\log_2 FC > |1|$ were included in the pathway enrichment analyses. In the IPA analysis platform, the Ingenuity Knowledge Base was referenced for both direct and indirect relationship of genes and endogenous chemicals in the human database. Significance of the input DE gene list in terms of overlap with pathways in the database was presented as -log *p*, evaluated by the Fisher exact test with Benjamini-Hochberg correction for multiple testing.

The DAVID platform utilises KEGG[87,246] (Kyoto Encyclopedia of Genes and Genomes), BioCarta[247], and BBID[248] (Biological Biochemical Image Database) databases for functional annotation. Significance of genetic enrichment was reported using modified Fisher's exact test and Benjamini-Hochberg multiple correction method was applied. The modified test refers to the adaptation of Fisher's exact test to measure the gene-enrichment in annotation terms in DAVID algorithm. In PANTHER and Reactome databases, Mann-Whitney U rank-sum test was utilised to assess statistical enrichment. The pathways enriched for genes differentially expressed between the different cycle phases from these platforms were compared; and for pathways that were identified in multiple databases, their respective database-specific gene lists were compared.

4.3 Results

4.3.i Comparison of histology, self-reported data, and transcriptomic discrimination of cycle phase in eutopic endometrium

Out of 190 eutopic endometrium samples with RNA-seq data passing sample QC, 82 had excess frozen tissue for histological dating; four (4.9%, endometriosis case n=2) samples were dated as menstrual, 23 (28%, endometriosis case n=14) as proliferative, and 53 (65%, endometriosis case n=31) as secretory, with two (2.4%) unclassified due to poor quality of sample (e.g. mucus only). Among secretory phase samples, three (5.7%) were dated as early-secretory, 23 (43%) mid-secretory and 18 (34%) late-secretory, with nine samples (17%) without sub-secretory staging available. Distinctive characteristics of endometrial gland changes throughout different phases in the menstrual cycle are shown in Supplementary Figure S 4.1. In total, 67 samples (84% of 80 samples with histology) had cycle phases assignment from both histology and self-reported data; of these, cycle phase of 45 samples (67%) matched (Figure 4.1). 36 samples (69%) have matched cycle phase in endometriosis case group and 9 samples (60%) have matched cycle phase in clinical control group (Supplementary Table S 4.1).

The discrepancy between histology-confirmed and self-reported cycle phases was the highest in proliferative phase (histology): 45% matched wrongly to secretory phase (self-reported) and 5% matched wrongly to menstrual phase (self-reported); among those matched wrongly to secretory phase from self-reported data, half of them are on the transitional day (i.e. day 15, the day after ovulation) of the menstrual cycle (Figure 4.1). There was less agreement between histology-confirmed vs. self-reported data in sub-secretory phases. In late-secretory phase (n=14), there was 58% agreement between histology and self-reported phasing; and there was no agreement of the early-secretory phase (n=2) samples (by histology) (Figure 4.2, Supplementary Table S 4.1).

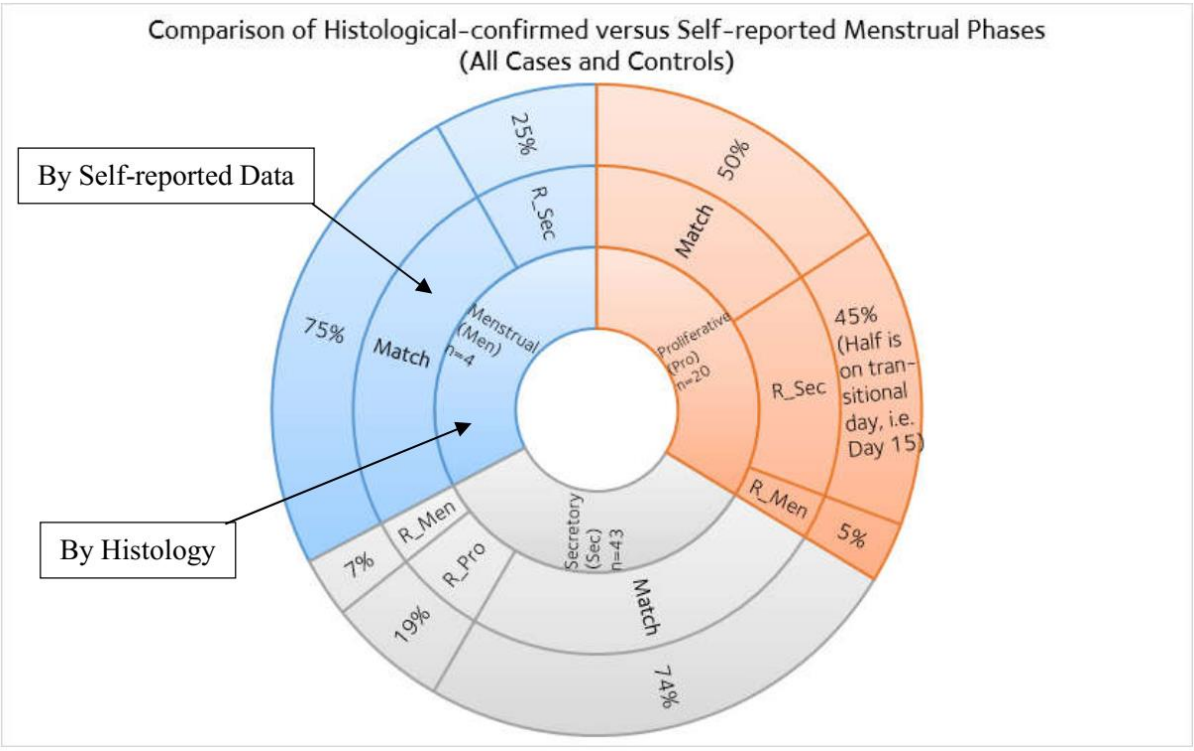


Figure 4.1 Comparison of histological-confirmed and self-reported cycle phases in all cases and controls.

Key: Inner rim = cycle phases by histology; R_Men = self-reported as menstrual, R_Pro = self-reported as proliferative, R_Sec = self-reported as secretory.

Histological-confirmed cycle phase (n=4) with matched self-reported phase (n=3) and R_Sec (n=1); histological-confirmed proliferative phase (n=20) with matched self-reported phase (n=10), R_Men (n=1), R_Sec (n=9); histological-confirmed secretory phase (n=43) with matched self-reported (n=32), R_Men (n=3) and R_Pro (n=8).

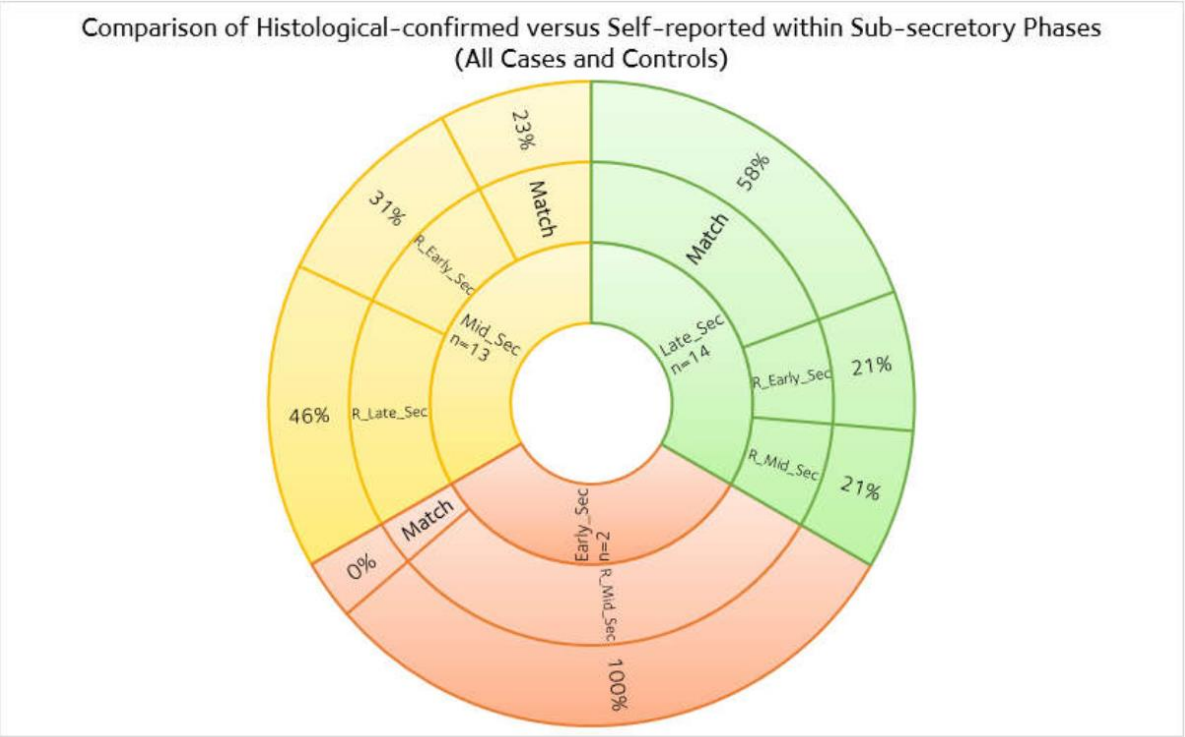


Figure 4.2 Comparison of histological-confirmed and self-reported within sub-secretory phases in all cases and controls.

Key: Inner rim = cycle phases by histology; R_Early_Sec = self-reported as early-secretory, R_Mid_Sec = self-reported as mid-secretory, R_Late_Sec = self-reported as late-secretory. Histological-confirmed early-secretory phase (n=2) with unmatched R_Mid_Sec (n=2); histological-confirmed mid-secretory phase (n=13) with matched self-reported phase (n=3), R_Early_Sec (n=4) and R_Late_Sec (n=6); histological-confirmed late-secretory phase (n=14) with matched self-reported (n=8), R_Early_Sec (n=3) and R_Mid_Sec (n=3).

To visualise the transcriptomic profiles of samples in PCA plots for cycle phase comparison, the expression of 18,030 genes across 80% of 190 samples (Section 4.2.ii.c) were included using the `DESeq` pipeline. The PCA plot of all samples with histological-confirmed cycle phases (n=80) showed clustering by phase, but the overlapping clusters meant that confident gating to allow distinction of the three cycle phases was not possible. The first two principal components (PC) explained 35.1% of the variation as shown in Figure 4.3. For PC1 vs. PC2 and PC3 vs. PC4 plots coloured with five cycle phase, please refer to Supplementary Figure S 4.2-3 respectively. The clustering did not change when the four menstrual phase samples were excluded. There was no clustering based on endometriosis case-control status of the samples.

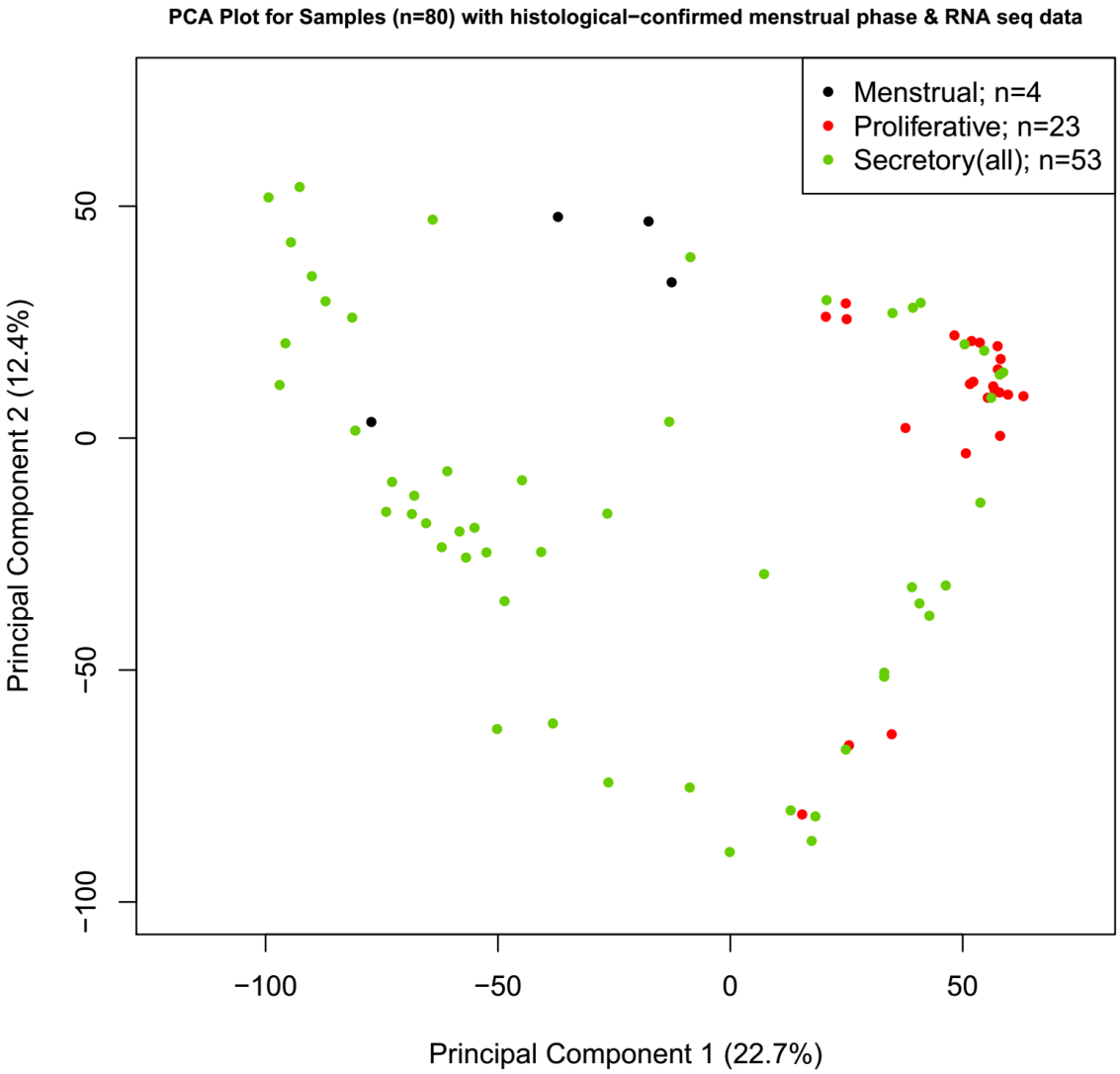


Figure 4.3 Principal component 1 (PC1) vs. principal component 2 (PC2) plot of PCA analysis for samples with histology-confirmed cycle phase (n=80) in three phases.
Percentage variation explained by each principal component is given in axis titles respectively. Key: Black dots = menstrual phase samples, red dots = proliferative phase samples, and green dots = secretory phase samples.

The clustering between sub-secretory phases in 17,910 genes across 80% of 53 histological-confirmed secretory samples showed much less clear clustering, although 44.2% of the variation was explained by the first two PCs (Figure 4.4, Supplementary Figure S 4.4). The clustering did not change when the three early-secretory phase samples were excluded. There was no clustering based on endometriosis case-control status of the sample.

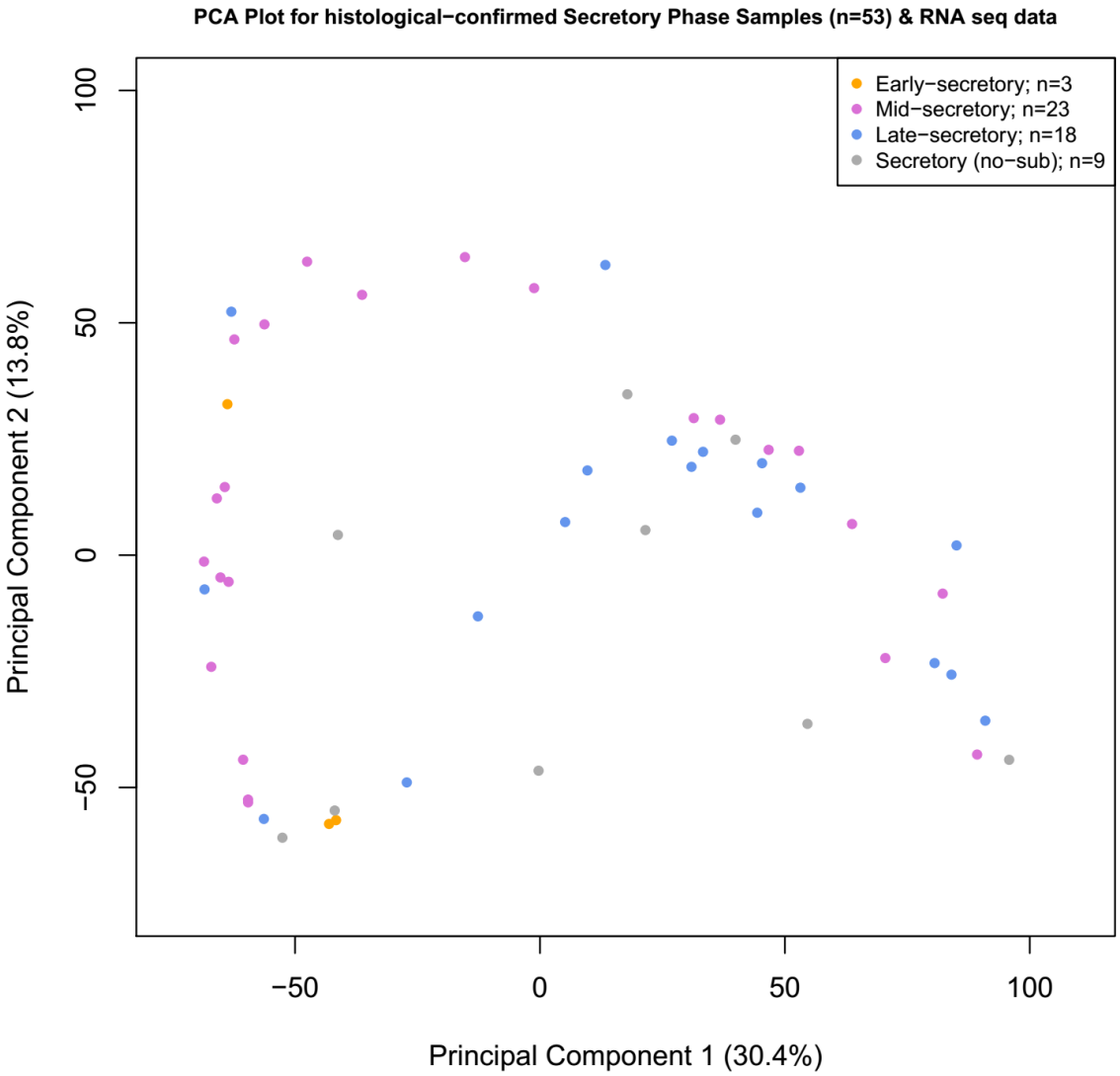


Figure 4.4 Principal component 1 (PC1) vs. principal component 2 (PC2) plot of PCA analysis for samples with histology-confirmed secretory phase only (n=53) in three phases.
Percentage variation explained by each principal component is given in axis titles respectively.
Key: Orange dots = early-secretory phase samples, purple dots = mid-secretory phase samples, blue dots = late-secretory phase samples, and grey dots = samples with no sub-secretory classification.

4.3.ii Differential gene expression between cycle phases in endometrium

4.3.ii.a Comparison of `edgeR` (1st) and `DESeq2` (2nd) pipelines

For comparison of the two pipelines, only histology-confirmed proliferative and secretory samples (n=23 & 53) were utilised for differential expression analysis. Apart from the histology menstrual cycle phase as the main covariate in the analyses, endometriosis case-control status was also tested for inclusion. Absence of clustering in the PC1 vs. PC2 plot (Supplementary Figure S 4.5, by the `DESeq2` pipeline) showed that cycle phase and endometriosis case-control status can be treated as two independent factors. The fit of two models for differential expression, one including cycle phase only, and the second including both factors (`design ~ histology phase + case-control status`) were compared by likelihood ratio test (LRT). The two-term model was only significant on 0.13% of genes (53 out of 42,348); thus, subsequent analyses were based on the simpler one-term model (`design ~ histology phase`).

In the `edgeR` (1st) pipeline, 24,277 genes were manually filtered out (not passing CPM>1 criteria) before differential expression analysis; there were 7,417 genes out of 18,071 that were differentially expressed at adj.$p \leq 0.05$. In the `DESeq2` (2nd) pipeline, 6,569 genes were independently filtered out; there were 11,463 out of 35,779 genes shown as differentially expressed at adj.$p \leq 0.05$, replicating 96.4% of those suggested by `edgeR` (Figure 4.5).

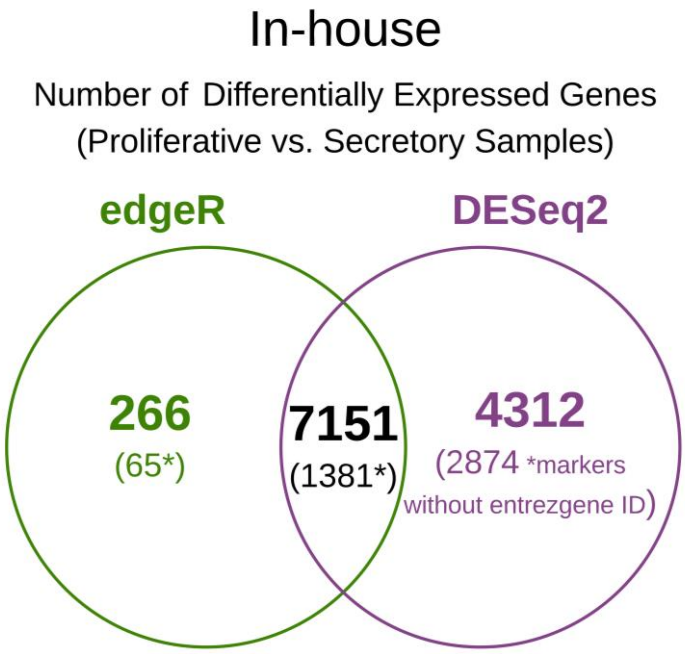


Figure 4.5 Venn diagram depicting the number of differentially expressed genes between histological-confirmed proliferative and secretory phase samples from `edgeR` and `DESeq2` pipelines.

Number shown are based on unique Ensembl ID of markers; number with asterisk are the number of markers without an entrezgene ID in that category.

The 7,151 genes commonly identified as differentially expressed in both pipelines, showed strong positive correlation between the two pipelines (Pearson correlation r = 0.95-0.99) in $\log_2$FC (Figure 4.6), ranked by adj.*p* (Figure 4.7), and log adj.*p* (Supplementary Figure S 4.6).

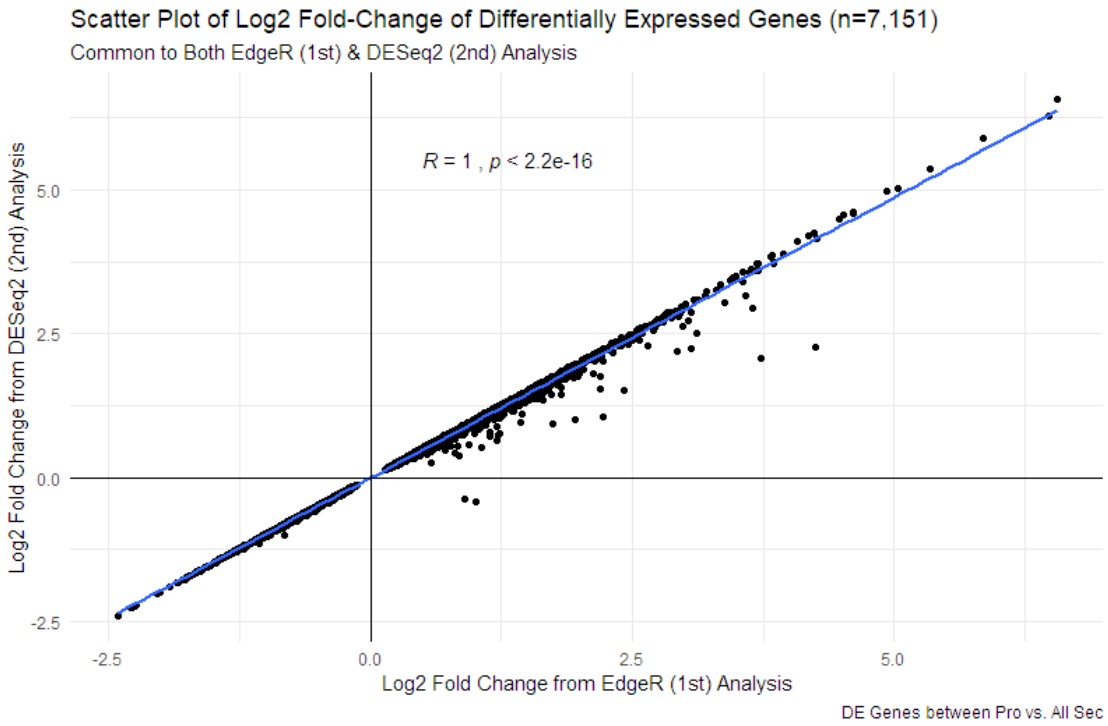


Figure 4.6 Scatter plot of $\log_2$ fold change of differentially expressed genes between proliferative and secretory phases (adj.$p \leq 0.05$) common to `edgeR` and `DESeq2` analyses.
Key: blue line = regression line by linear model.
Positive $\log_2$ fold change indicates higher expression in secretory than in proliferative phase.

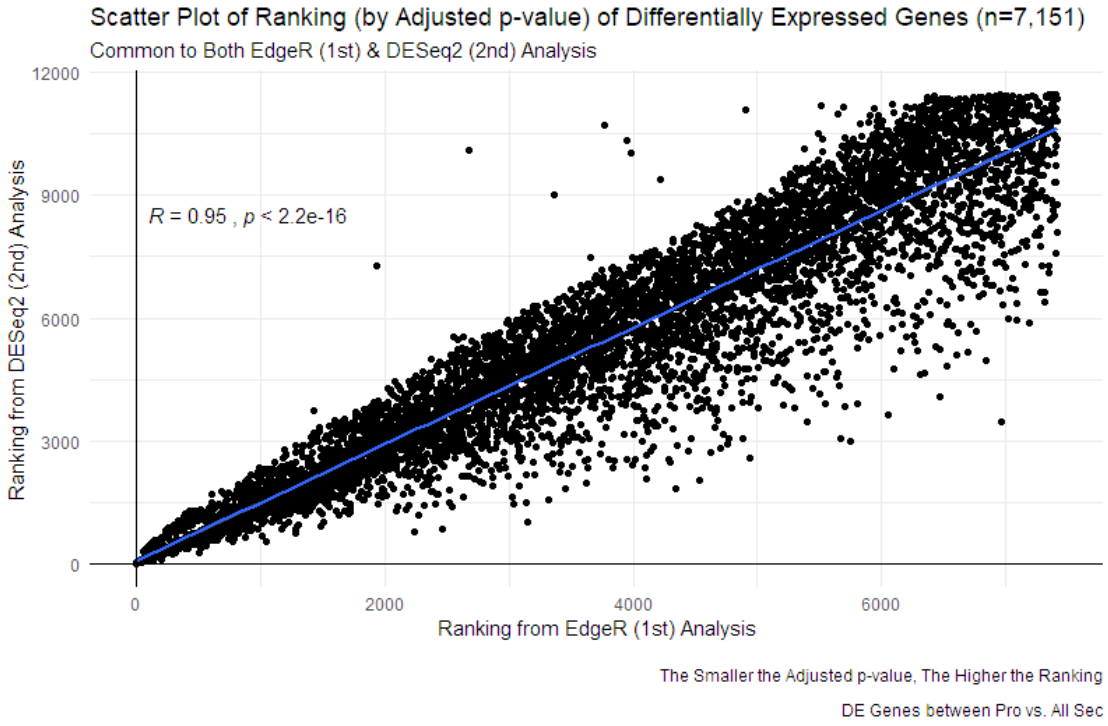


Figure 4.7 Scatter plot of ranking by adjusted *p*-value of differentially expressed genes between proliferative and secretory phases common to `edgeR` and `DESeq2` analyses.
Key: blue line = regression line by linear model; the smaller the adj.*p*, the higher the ranking.

Despite the similarity in results from both analysis pipelines, DESeq2 (2nd) suggested more unique genes (Figure 4.5, Supplementary Table S 4.2 for top 50 DE genes by adj.*p* in DESeq2 compared to edgeR) that were highly differentially expressed with larger absolute fold changes as shown in Figure 4.9 with purple-labelled gene names.

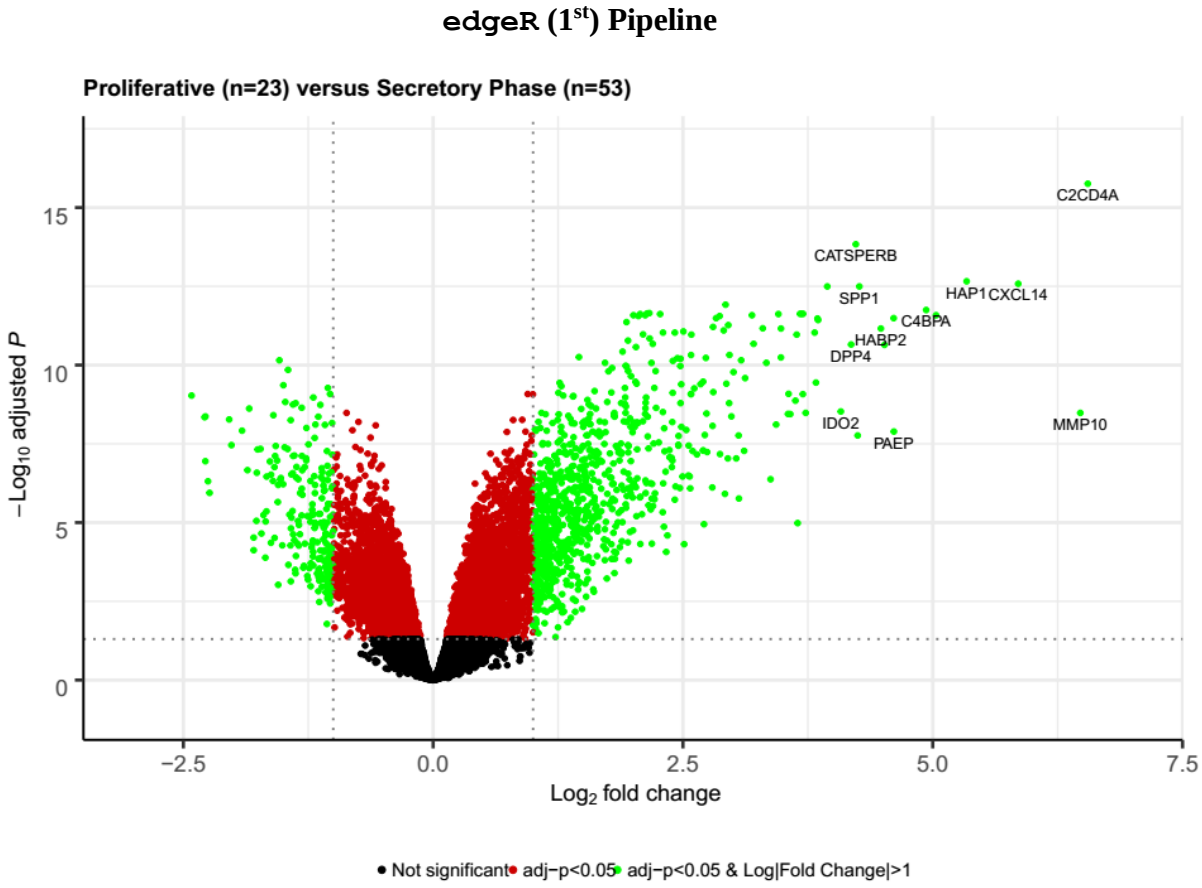


Figure 4.8 Volcano plot for differentially expressed genes between proliferative and secretory phase samples by edgeR analysis.

Key: green dots = gene with both adj.*p*<0.05 & Log_2|FC|>1, red dots = gene with adj.*p* <0.05 only, and black dots = gene with adj.*p*>0.05. Positive log_2 fold change indicates higher expression in secretory than in proliferative phase.

DESeq2 (2nd) Pipeline

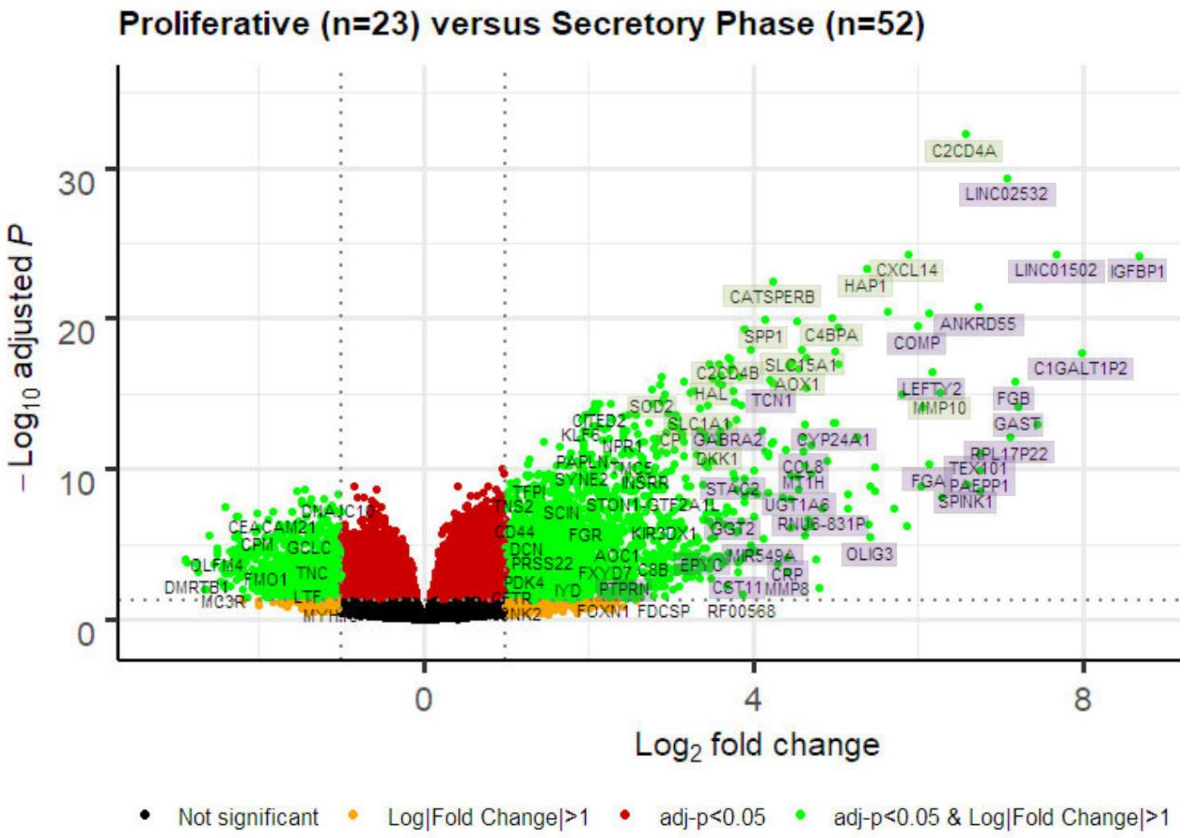


Figure 4.9 Volcano plot for differentially expressed genes between proliferative and secretory phase samples by DESeq2 analysis.

Key: green dots = gene with both adj.p<0.05 & Log_2|FC|>1, red dots = gene with adj.p <0.05 only, orange dots = gene with Log_2|FC|>1 only, and black dots = gene with adj.p>0.05.

Purple-boxed gene name = uniquely highlighted as differentially expressed (DE) gene in DESeq2 analysis; green-boxed gene name = gene highlighted as DE gene in both edgeR and DESeq2 analyses. Positive log_2 fold change indicates higher expression in secretory than in proliferative phase.

4.3.ii.b Differential gene expression between proliferative and secretory phases

Given the more conservative results from the `edgeR` (1st) pipeline, the subsequent analyses were based on the results from this pipeline. The differentially expressed genes (n=7,417) between proliferative (n=23) and secretory phase (n=53) samples (Figure 4.8, Supplementary Table S 4.3 for top 50 DE genes by adj.*p*) were compared to those published in Fung's[86] and Talbi's[82] microarray experiments for 62 and 6 proliferative versus 15 and 3 early-secretory phases samples, respectively.

My analysis replicated 53% (Figure 4.10, n=834 out of 1588 when considering unique generic gene names) of the differentially expressed genes reported in Fung et al[86]. These genes (n=839 with unique Ensembl ID) reported in common showed good correlation (r= +0.87) in $\log_2$FC from both analyses (Figure 4.11): 97.3% (n=816/839) of the data points were located in positively correlated quadrants. However, there was no correlation in ranking by adj.*p* (r= -0.042, Supplementary Figure S 4.7) and log adj.*p* (r= -0.092, Supplementary Figure S 4.8) from the two studies. Correlation of $\log_2$FC of DE genes common to Talbi's[82] and my analysis could not be plotted due to limited information available from their publication.

Among in-house DE genes, 11.3% were reported in Fung's[86], 95% of which were with matched directions in fold change. However, only 1.6% of the in-house DE genes were reported in Talbi's[82] study, 88% of which matched direction. Considering the large differences between the three experimental setups, the extent of overlap between these studies was appreciable, suggesting 115 genes in common as shown in Figure 4.10. However, out of these 115 DE genes, only 40 were reported with same direction of effects across all studies.

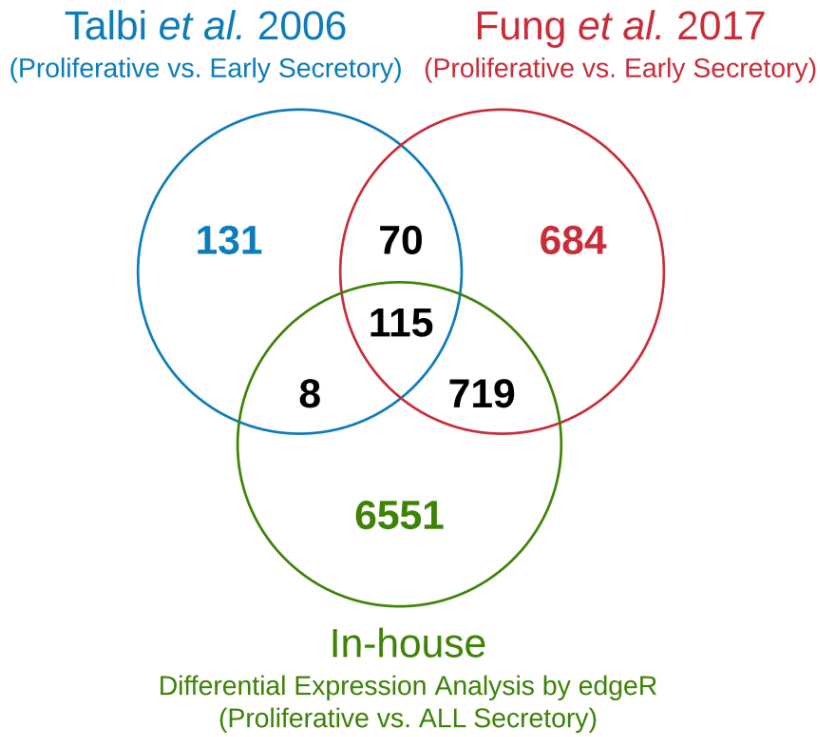


Figure 4.10 Venn diagram for differentially expressed (DE) genes between in-house (proliferative vs. all secretory) RNA-seq analysis, Fung's[86] and Talbi's[82] published DE gene lists (proliferative vs. early-secretory) from microarray studies.

Number of markers shown was based on their unique generic external gene names reported in publication; duplicated gene names in published DE lists were removed prior to plotting the Venn diagram.

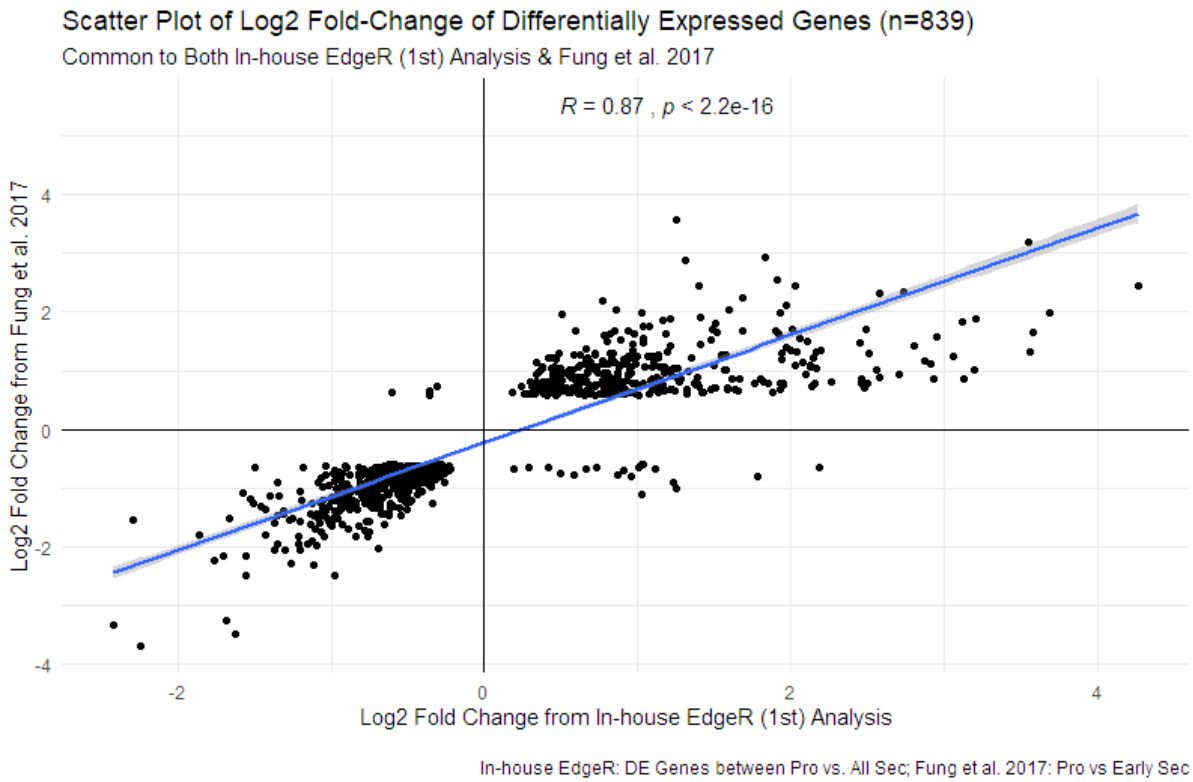


Figure 4.11 Scatter plot of $\log_2$ fold change of differentially expressed genes (adj.$p \leq 0.05$) common to in-house `edgeR` (proliferative vs. all secretory) and Fung et al. (proliferative vs. early-secretory) analyses.

Key: blue line = regression line by linear model, blue shading = 95% confidence interval (CI) of regression line.

Positive $\log_2$ fold change indicates higher expression in secretory than in proliferative phase.

4.3.ii.c Differential gene expression between mid-secretory and late-secretory phases

For the differential expression analysis between mid (n=23) and late-secretory phase samples (n=18), 21 significant differentially expressed genes were detected at adj.$p \leq 0.05$ out of 18,029 expressed genes (Figure 4.12, Supplementary Table S 4.4). None of these were reported in previous studies[82,86] of the same cycle phases comparison; only one gene, complement component 7 (*C7*), was mentioned in a study[85] with more than two-fold change in expression without specifying the cycle phase comparison.

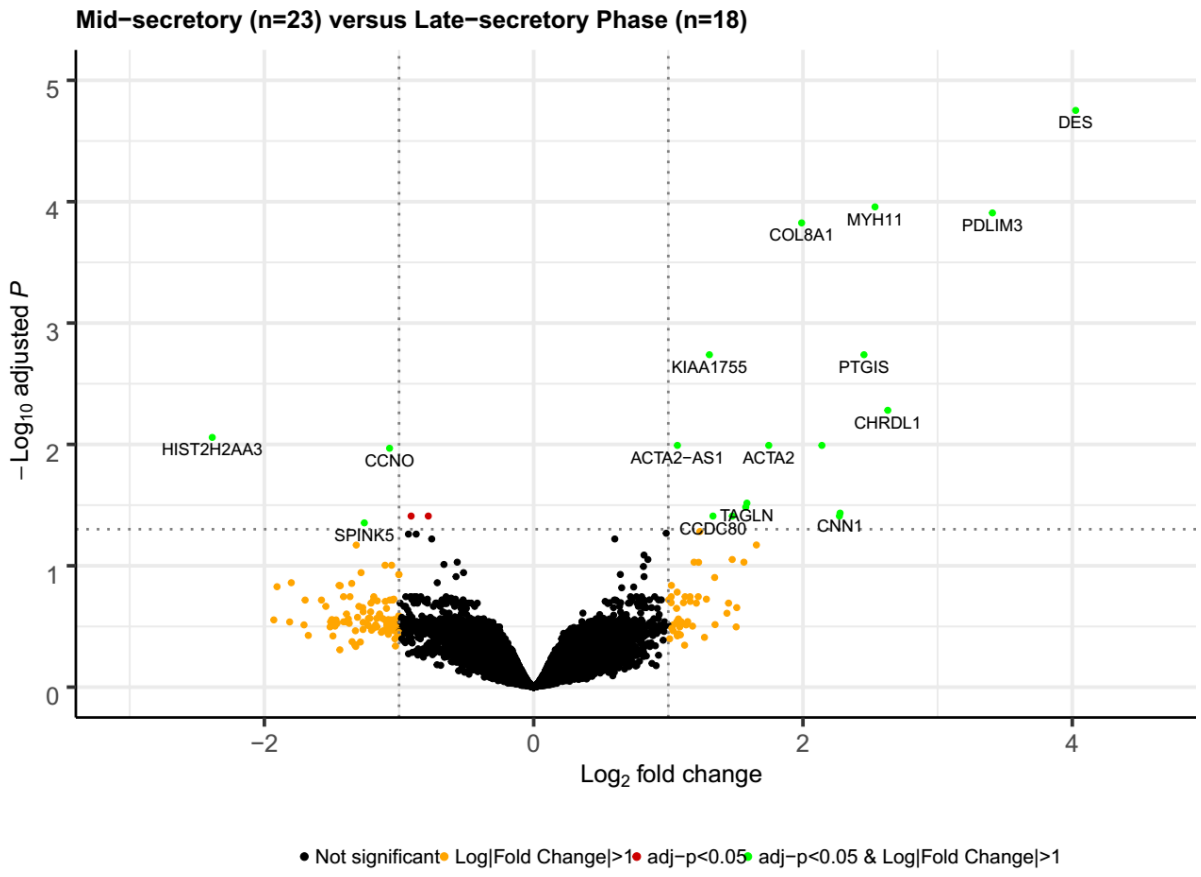


Figure 4.12 Volcano plot for differentially expressed genes between mid-secretory and late-secretory samples by `edgeR` analysis.

Key: green dots = gene with both adj.p<0.05 & Log_2|FC|>1, red dots = gene with adj.p <0.05 only, orange dots = gene with Log_2|FC|>1 only, and black dots = gene with adj.p>0.05. Positive log_2 fold change indicates higher expression in late-secretory than in mid-secretory phase.

I presented the preliminary differential expression analysis from Section 4.3.ii.b & 4.3.ii.c at 'the Genomics of Common Diseases 2017' international conference; abstract and poster can be found in supplementary materials A.

4.3.iii Multivariate model for cycle phase prediction

Out of 40 genes that were identified as differentially expressed and with the same direction of effect between proliferative and secretory phases in our in-house, Fung's[86], and Talbi's[82] studies, 63% (25 genes) of them were up-regulated in the secretory phase. The 26/40 genes with in-house $\log_2$FC >|1| were taken further for building the multivariate prediction model (Supplementary Table S 4.5). About half of gene pairs showed correlation in their expression patterns as shown in Supplementary Figure S 4.9.

In the single-gene logistic regression analysis (Supplementary Table S 4.5, Section 4.2.iv), only arginase 2 (*ARG2*) showed a significant result (p=0.044). In the multivariate forward and both-direction step-wise logistic regression analyses, a single-gene model including only *ARG2* (AIC=91.99) was also suggested to be the best fit. However, the performance of the single *ARG2* model to predict proliferative vs. secretory phase was not optimal, with accuracy of the model at around 65%; the area under ROC curve (AUC) was 0.664 (95%CI: 0.538-0.789).

An alternative, 18-gene model (AIC=72.87), was suggested by the backward step-wise logistic regression: Proliferative / Secretory phase ~ *ARG2* + *TNC* + *FGFR3* + *SEMA3A* + *PTGS1* + *SFRP1* + *TFPI2* + *DKK1* + *MSX2* + *OLFM1* + *CD36* + *SLC3A1* + *FOXO1* + *MT1E* + *CDH2* + *MT1X* + *APOD* + *MAOA*. This multi-gene model had a higher accuracy (92%, Figure 4.13) than the single-gene model, with 0.96 AUC (95% CI: 0.921-0.999) of ROC. However, there is a high risk of over-fitting in this multi-gene model as half of those 18 genes were not passing significance ($p \leq 0.05$) in the step-wise logistic regression statistics. The residuals-versus-fitted plot, Q-Q plot, and accuracy variant signature curves can be found in Supplementary Figure S 4.10-11.

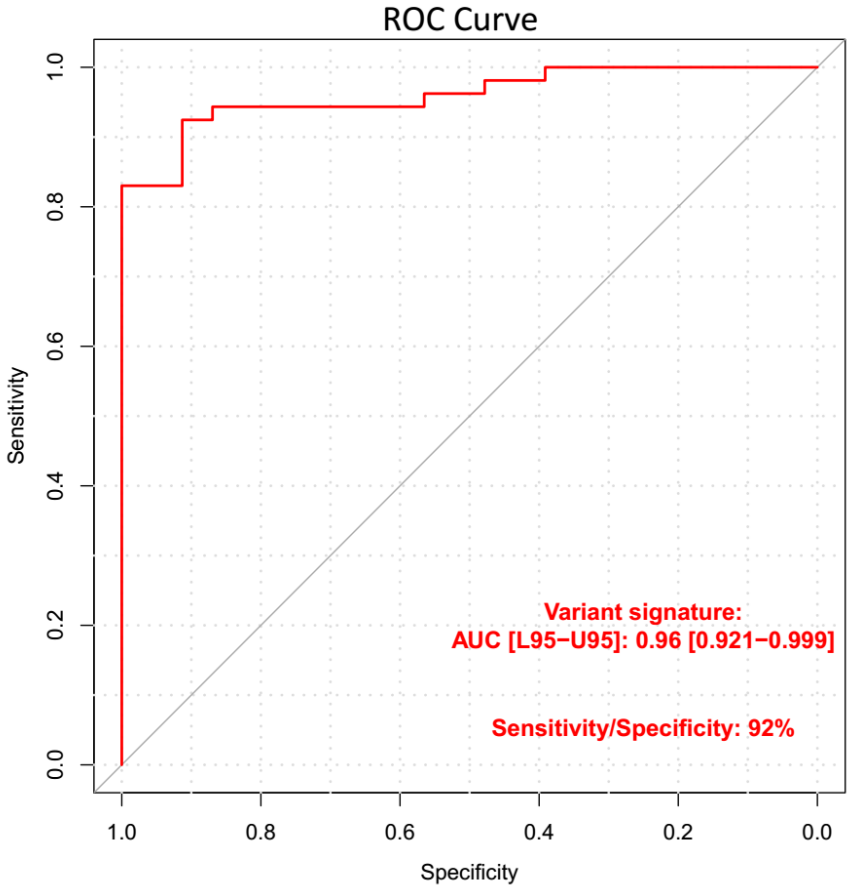


Figure 4.13 The receiver operating characteristic (ROC) curve from the prediction of proliferative vs. secretory phase based on the 18-gene model.

Key: Sensitivity/Specificity = sensitivity and specificity weighted equally (i.e. average of two terms), gives a percentage similar to the model diagnostic accuracy when an unknown-phase sample is input to the model, where the probability of it being a proliferative or a secretory phase is the same.

4.3.iv Pathway enrichment analysis on differentially expressed genes between cycle phases

A total of 1,089/7,417 differentially expressed genes between proliferative (n=23) and secretory (n=53) phases with $\log_2$FC >|1| in the in-house data (using `edgeR` pipeline), were utilised in the pathway enrichment analysis. In the IPA Ingenuity analysis, 1,087 genes were matched to the database, with 862 genes up-regulated in the secretory phase compared to the proliferative phase. In IPA Ingenuity canonical pathway analysis, 168 significantly enriched pathways were identified (Benjamini-Hochberg $p \leq 0.05$, i.e. -log $p \geq 1.3$; Supplementary Table S 4.6). Top 10 enriched canonical pathways are shown in Figure 4.14 . Positive z-score suggested that most genes involved in a particular pathway were up-regulated in the input DE gene list, and thus the pathway was considered as significantly activated (z>+2) in secretory phase.

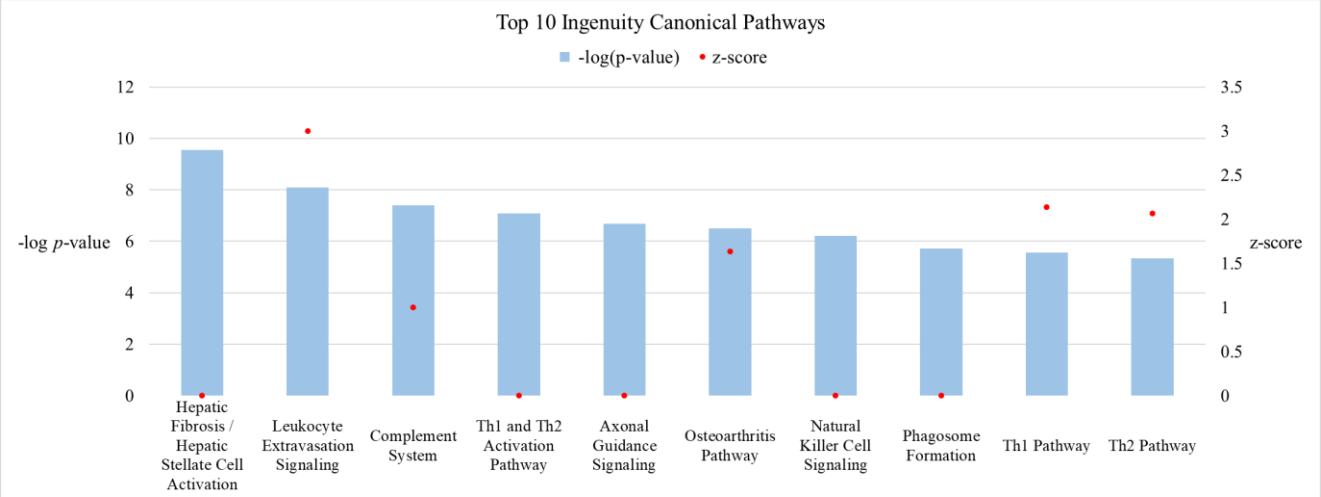


Figure 4.14 Top 10 enriched Ingenuity canonical pathways for differentially expressed genes between proliferative and secretory phase endometrium by IPA.

Key: z-score = 0 means no suggestion of direction of activation/inhibition from the database

Under the diseases and bio-function analysis by IPA platform, the enriched annotations mainly involved cellular movement, immune and blood cells trafficking. Out of 500 significantly enriched diseases and bio-function annotations (Benjamini-Hochberg $p \leq 7.94 \times 10^{-7}$, i.e. -log $p \geq 6.1$), 89 were predicted to be activated (z>+2) in the secretory phase (Supplementary Table S 4.7), the rest were with undetermined direction in prediction; none were identified as inhibited. The top 10 functional annotation by activation z-score is shown in (Figure 4.15).

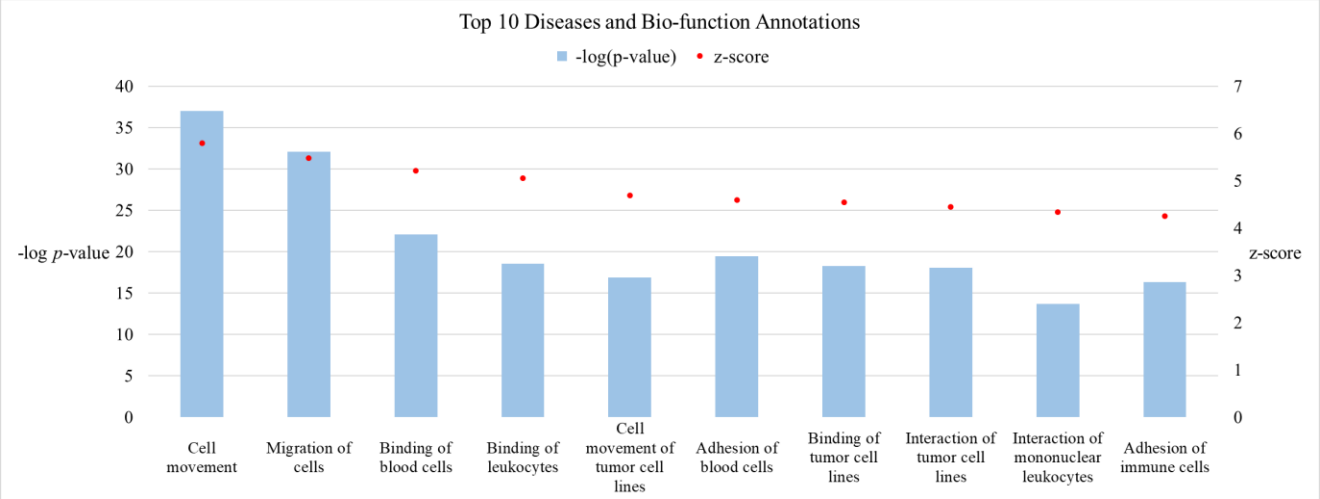


Figure 4.15 Top 10 enriched diseases and bio-function annotations for differentially expressed genes between proliferative and secretory phase endometrium using IPA.

Key: z-score = IPA activation score, >+2 the pathway is activated in secretory relative to proliferative phase.

Reassuringly, reproductive system disease (p= 5.80x10^{-7}, involving 661 genes) was one of the top five highlighted enriched 'diseases and disorders' categories. A detailed list of the top 50 diseases and functional annotations by significance, regardless of the predicted activation status, can be found in Supplementary Table S 4.8; it includes the endometriosis disease pathway at the 11th by significance.

In the analysis using the DAVID platform, 975 differentially expressed genes were matched with the Ensembl IDs in KEGG, BioCarta, and BBID databases. Fifty pathways were enriched at nominal $p \leq 0.05$, of which 12 were significantly enriched after adjustment for multiple testing (Table 4.2). All 12 enriched pathways passing multiple testing correction were from the KEGG database; among the other 38 pathways not passing the threshold, 71% were from KEGG, 24% from BioCarta, and 5% from BBID.

Table 4.2 List of significantly enriched KEGG canonical pathways (Benjamini-Hochberg $p \leq 0.05$, n=12) for genes differentially expressed between proliferative and secretory phase in endometrium, from DAVID analysis.

Key: No. of genes = total number of genes from the DE list involved in a particular pathway; % Genes in Pathway = percentage of the involved genes from the DE list to the total number of genes constituting a particular pathway from the database

KEGG Pathway Code	**Canonical Pathways**	**Benjamini-*p***	**Fold Enrichment**	**No. of Genes**	**% Genes in Pathway**
hsa04610	Complement and coagulation cascades	3.26E-05	4.39	19	1.95
hsa04151	PI3K-Akt signalling pathway	2.31E-03	1.99	43	4.41
hsa05150	Staphylococcus aureus infection	1.60E-03	4.14	14	1.44
hsa04640	Hematopoietic cell lineage	1.29E-03	3.30	18	1.85
hsa05200	Pathways in cancer	9.20E-03	1.79	44	4.51
hsa04512	ECM-receptor interaction	1.14E-02	2.94	16	1.64
hsa02010	ABC transporters	1.04E-02	3.99	11	1.13
hsa04514	Cell adhesion molecules (CAMs)	1.49E-02	2.36	21	2.15
hsa04060	Cytokine-cytokine receptor interaction	1.44E-02	1.97	30	3.08
hsa04668	TNF signalling pathway	2.23E-02	2.54	17	1.74
hsa05202	Transcriptional misregulation in cancer	3.52E-02	2.10	22	2.26
hsa05146	Amoebiasis	4.62E-02	2.41	16	1.64

In the PANTHER analysis, 916 differentially expressed genes were mapped to identifiers in PANTHER and Reactome databases. From the PANTHER database, three pathways were suggested to be enriched and up-regulated in the secretory vs. proliferative phase, though these did not pass the significance threshold after Bonferroni correction ($p \leq 0.05/714$). These pathways were Huntington disease (PANTHER pathway code: P00029, $p=1.65 \times 10^{-2}$, number of genes matched to pathway n=3), heterotrimeric G-protein signalling pathway-Gq alpha and Go alpha mediated pathway (P00027, $p=3.53 \times 10^{-2}$, n=9), and CCKR (cholecystokinin receptors) signalling map (P06959, $p=3.57 \times 10^{-2}$, n=26).

From the Reactome database, 57 pathways were highlighted ($p \leq 0.05$) though, similarly, none passed the Bonferroni-adjusted significance threshold for multiple testing correction ($p \leq 0.05/714$). Examples of pathways that were down-regulated in the secretory phase were regulation of cell cycle, DNA repair, synthesis and replication, and oestrogen-dependent pathways. Complement cascade, platelet metabolism, and cytokine signalling pathways were up-regulated in secretory compared to proliferative phase in endometrium. A detailed list of all 57 pathways is given in Supplementary Table S 4.9.

In summary, there appeared to be an enrichment of upregulated complement cascade and coagulation signalling pathways in secretory vs. proliferative phase endometrium across most databases, except for PANTHER and BBID. The gene lists that constitute the complement pathway from the Ingenuity (Figure 4.16), KEGG and BioCarta are very similar to each other; nine out of 19 genes in those pathways were a member of more than one lists. The list from the Reactome database for complement system consisted of many more genes (n=139) than the other databases, but shared the same key gene members, e.g. *MASP1*, *SERPING1*, as well as the complement component genes *C2*, *C3*, *C4A*, *C4B* and *C7*.

When looking up the expression of these pathway constituting genes expression in the in-house DE gene list, 18 out of 19 genes were consistently up-regulated in secretory vs proliferative phase; however the rankings by adj.*p* of these genes in the in-house list (with 7k genes in total) were not in the top of the list: only three genes were in the top 50 differentially expressed genes, and another two in the top 100.

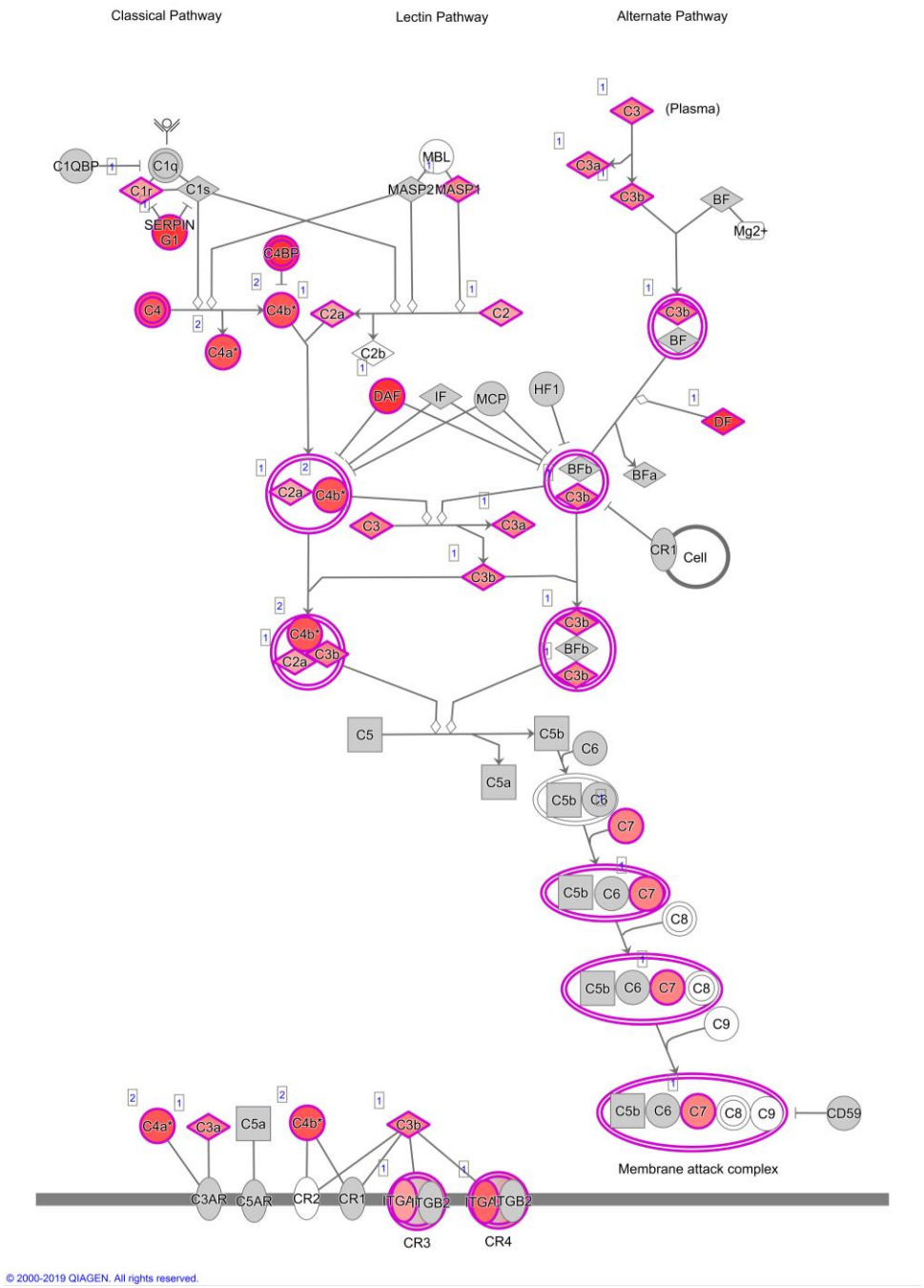


Figure 4.16 Complement system pathway highlighting molecules up-regulated in secretory than in proliferative phase based on differential gene expression analysis on endometrium.

(Permission for re-distribution (both print and digital) of copyrighted figure (open-access CC-BY 4.0) is granted on 19 May 2020 by IPA, QIAGEN Inc., permission letter in supplementary materials F.ii)

Key: Molecules marked in degree of reds = the deeper the red, the more up-regulated for that molecule in secretory phase

4.4 Discussion

In cycle phase determination, histology staging on eutopic endometrium is the gold standard and is deemed to give the most reliable result; however, there are practical limitations to obtaining histology information on every sample, such as tissue availability and the quality of the biopsy obtained. The assignment of the phase depends on good visualisation of features of the endometrial glands in the stratum functionalis (details in Section 1.7); this may not always be possible, or in cases where a sample is taken at time of transition between phases, assessment of different sections from one block may give different results.

Studying transcriptomic signatures of endometrium throughout the cycle may improve the accuracy of cycle phase determination without the need of histology, but my analyses showed there are considerable challenges. The transcriptomic profiles of three main phases, i.e. menstrual, proliferative, and secretory phase, showed some degree of clustering as shown in the PCA plot (Figure 4.3). Proliferative samples clustered with positive PC1 and spread along PC2. Secretory samples scattered along both PC1 and PC2; when these samples were labelled in their sub-secretory phases, early-secretory samples (n=3) clustered closely with proliferative samples (Supplementary Figure S 4.2). Though only four samples were identified as menstrual samples, they clustered on PC2. One menstrual and one secretory sample seems positioned in the opposite cluster in Figure 4.3; their histology-based staging was double-checked by the pathologist, which ruled out mislabelling.

Due to the wide spread of the secretory samples on PC1 and PC2 of the PCA plot, a PCA analysis sub-setting only secretory samples to examine potential sub-secretory pattern was performed (Figure 4.4). Early-secretory samples (n=4) clustered along negative PC1, together with half of the mid-secretory samples. There is a smaller proportion of late-secretory samples that cluster with the early-secretory samples, but more on positive PC1 with the other half of mid-secretory samples. The overlap between the clusters was such that cluster-based phase identification of samples with unknown histological cycle phase (n=110) could not be achieved. However, such attempt have been successful in previous literature[82] to assign samples with ambiguous histology phase (n=6) with a predicted phase based on PCA.

I investigated whether cycle phase informed by self-reported LMP can be an alternative for the unknown-phase samples (n=110); the concordance (67%) between LMP derived and histology-based phases (Figure 4.1) for 67 samples is moderate, and thus could be the backup source for phase assignment. The largest portion of samples that did not match was between self-reported early-secretory and histology-based proliferative samples. It is important to note that half of these mismatches were self-reported on day 15 of a typical 28-day cycle. It is difficult to track ovulation from self-reported LMP data when compared to histology since in histology we can easily look for distinct structure, i.e. the sub-nuclear vacuoles in glands (Supplementary Figure S 4.1) to evaluate if ovulation has occurred to stage the sample with accuracy to day 14 or 15. Thus, caution is needed when using LMP-derived early-secretory vs. proliferative phase samples in subsequent disease-based analysis in the absence of histology information.

There are more mismatches between self-reported and histology-based sub-secretory phases as shown in Figure 4.2. One of the reasons is that, histologically, it is sometimes difficult to distinguish early and mid-secretory samples, especially in the transition between the two sub-phases. This compares to the relative ease to date late-secretory samples, due to its distinctive pattern of stromal cells condensation, expansion of pre-decidua, lymphocytic infiltration, and haemorrhage. This is apparent in this study as well where we have observed the highest concordance of self-reported and histology-based for late-secretory samples in comparison to the other two sub-secretory phase samples.

A second aim was to better understand the normal physiological transcriptional differences between cycle phases through differential expression analyses. To verify the in-house results before comparing to other publications, I used two pipelines. The first one included `DESeq` for pre-processing and `edgeR` for differential expression analysis; the second utilised `DESeq2` for both steps. The two pipelines have different strengths, with different normalisation algorithms and follow-up statistical analysis frameworks; however, most recent papers in literature[249-251] mainly utilise `DESeq2`[252] for pre-processing as it is designed to be the improved version of `DESeq`. In the first pipeline, `DESeq` did not automatically filter out lowly expressed genes independently according to experimental design. This meant that stringent manual filtering needed to be in place to exclude genes (n=24,277) expressed in less than 80% of our samples, prior to DE analysis, to prevent an over-conservative multiple-testing correction to imposed on the final results. In the second pipeline, `DESeq2` applied a less stringent gene filtering independently for every experimental comparison group, with FDR > 5%. This resulted in exclusion of 6,569 genes that did not pass the FDR threshold when conducting the DE analysis, and allowed for more genes to go through the analysis pipeline. Reassuringly, despite the technical differences between two pipelines, the `DESeq2` pipeline replicated most (96%) of the results from the `edgeR` pipeline under the same model. The overlap of DE genes is huge (n=7,151) and remarkably consistent in their $\log_2$FC and ranking by adj.*p* (Figure 4.6 and Figure 4.7), which provides confidence in the results especially for those with high absolute fold change and significance value. Another similarity between results from both pipelines is that more than one third of the total genes being tested (7,417 DE genes/ 18,071 genes tested in `edgeR` & 11,463 DE genes/ 35,779 genes tested in `DESeq2`), passing respective filtering steps, have been highlighted as significant. The number of significantly differentially expressed genes is large in this dataset; a study has shown that the median number of DE genes reported by 600+ datasets, mainly testing for the effect of disease and treatment groups, is around a hundred[253]. In the context of endometrium, Talbi et al.[82] reported around 5% of all tested genes to be differentially expressed across cycle (2-3k DE genes/ 55k genes tested); while Fung et al.[86] reported around 13% (1.6k DE genes/ 12k genes tested). It is likely that the DE analysis results are a reflection of the considerable changes in normal physiological processes occurring in dynamic endometrium, with the expression of large numbers of genes changing across the cycle. The overlap in the proportion of differentially expressed genes between two pipelines confirm that the observed results are not driven by methodological issues.

The major difference in the results between the two pipelines is that there are 4,312 significant DE genes solely suggested by the DESeq2 pipeline (Figure 4.5). It is important to note that out of the total number of significant DE markers suggested through DESeq2 analysis (n=11,463), 38% (n=4,320) are non-protein coding and most of these (n=4,222) are without an entrezgene ID from NCBI database but still with an unique Ensembl ID. The main categories that include these non-protein coding transcripts are different types of RNAs (e.g. microRNAs, long non-coding RNAs (lncRNA) and pseudogenes). These markers are dispersed evenly throughout the DE gene list, i.e. not showing predominance among high or low ranked genes (Supplementary Table S 4.2). If we deduct the number of non-protein coding transcripts from the results of the DESeq2 pipeline, a similar number of DE genes (n=7k) remains as suggested by edgeR, with largely matching gene identities as well. Therefore, the major difference in total number of DE genes discovered between two methods is likely due to the ability of DESeq2 to include more non-protein coding genes in the DE analysis, which count towards the higher number of total DE genes observed. It is also possible that the edgeR pipeline, employing more stringent filtering criteria (CPM>1), may have missed the ability to detect these non-protein coding genes if they were not expressed in more than 80% of our samples; transcripts that are uniquely expressed in certain cell types in the endometrium (e.g. glandular epithelium) may also have missed if the sampling for pipelle for RNA-seq was not homogenized.

Alhough the `DESeq2` pipeline identified many non-protein coding transcripts as differentially expressed, it also suggested some differentially expressed protein-coding genes that the `edgeR` pipeline did not identify (labelled in purple in Figure 4.9). One of these is the Insulin Like Growth Factor Binding Protein 1 (*IGFBP1*) gene, showing the highest $\log_2$FC and ranked 5th by adj.*p* among all DE genes. *IGFBP1* was not expressed in 26% (n=20/76 with raw count < 5) across our samples. It was up-regulated in the secretory phase compared to the proliferative phase. The literature[254] suggested it is involved in endometrial development, decidualisation, and interaction with trophoblast. This gene is located within the 7p12.3 cytoband, which contains two independent genome-wide significant variants (rs74491657[141], rs55909142[135]) associated with endometriosis. I hypothesise that there may be a shared mechanism between regulation of secretory phase in normal endometrium and endometriosis pathophysiology. Further identification of variants associated with expression level of *IGFBP1* in endometrium (i.e. eQTL SNPs) and its correlation with the two endometriosis GWAS variants of interest can be performed to see if the regulation of *IGFBP1* expression plays a role. To allow a better comparison to DE gene lists found in the literature[82,85,86], which have tended to only have reported protein-coding genes, the results from the `edgeR` pipeline was used for the differential expression and multivariate model analyses.

In the DE analysis between proliferative and secretory phases, many more up-regulated DE genes were observed in secretory phase (Figure 4.8). This may be due to the unbalanced representation in number of samples in each phases (23 in proliferative phase versus 53 in secretory phase), and also to the fact that the secretory phase encompasses three sub-phases (early, mid and late). Ideally, the comparison of differential expression associated with transition of one phase to another should focus on comparison between proliferative and early-secretory phases as most studies in the literature have done. This was not possible in this study, as there were only three early-secretory samples with histology information.

The Fung et al.[86] and Talbi et al.[82] studies utilised microarray technology and compared gene expression between proliferative and early-secretory phases. The sample size was very small in Talbi et al.: n=5 for proliferative and n=3 for early-secretory phase; Fung et al. compared 66 proliferative with 15 early-secretory samples. In my DE analysis of endometrium samples, I replicated half of Fung's list, and 38% of Talbi's, with good correlation in directional $\log_2$FC (Figure 4.11). Given the number of technical and experimental differences of my study set-up with those by Fung et al. and Talbi et al., this overlap is perhaps surprising. The small number of DE genes with opposite directions of FCs (shown in the negative correlation quadrant in the scatter plot) had lower effect sizes ($\log_2$FCs < 1). Despite the good overlap of DE genes, their ranking by adj.*p* in respective studies did not correlate highly between studies (Supplementary Figure S 4.7).

The top five differentially up-regulated genes in secretory phase in my analysis were not reported in either Fung et al. nor Talbi et al. list for early-secretory versus proliferative phase (Supplementary Table S 4.3). However, strong evidence of up-regulation in secretory phase of all five genes has also been provided by several microarray experiments in the literature. The first gene, C2 Calcium Dependent Domain Containing 4A (*C2CD4A*), is responsible for regulating cell architecture, adhesion and inflammation process[255], which in line with the properties of maintaining the lining in early and mid-secretory phases and the lymphocytic infiltration in late-secretory phase. It is reported to be the third most up-regulated gene in mid-secretory phase in a joint analysis of published endometrial microarray data[256] and is included as one of 53 genes in a predictor panel for endometrial receptivity[257]. Interestingly, the encoded protein is predicted to be interacting with eight other proteins by the Integrated Interactions Database (IID)[258], of which the eight genes are also highlighted as top 50 DE genes between secretory and proliferative phases. These eight proteins are CXCL14, SDCBP2, C4BPA, IRX3 (suggested by the `DESeq2` pipeline), RIMKLB, MAP3K5, NABP1, and MAOA. In addition, *ARG2* — one of the 26 genes for building of multivariate model (Supplementary Table S 4.5) — its protein is also predicted to be interacting with C2CD4A. As there are only 17 proteins reported to be interacting with C2CD4A according to the database, the chance of the match (i.e. 9 out of 17 recorded as DE genes) being a random finding is relatively low, and the likelihood of these 9 genes working in the same pathway is high.

The second ranked, Cation Channel Sperm Associated Auxiliary Subunit Beta (*CATSPERB*) gene, is involved in progesterone response and hyperactivation of sperm cells[259], which appears in line with the functionality in the implantation window of a secretory endometrium. A microarray study[260] has also shown that *CATSPERB* is up-regulated in mid-secretory phase and moderately expressed at late-secretory phase. The third gene, Huntingtin Associated Protein 1(*HAP1*) is known for its association with neurological Huntington disease[261]; the electron transport chain is one of its related pathways according to the online PathCards database (https://pathcards.genecards.org/card/respiratory_electron_transport_atp_synthesis_by_chemiosmotic_coupling_and_heat_production_by_uncoupling_proteins, accessed on 18 May 2020). Two other studies[262,263] have shown that *HAP1* expression in endometrium is changed during implantation. The gene is also included in the 53-gene panel for prediction of endometrial receptivity[257].

The fourth, C-X-C Motif Chemokine Ligand 14 (*CXCL14*), encodes for an anti-microbial chemokine regulating immune and inflammatory responses, commonly expressed in epithelial cells[264,265], which may play a role in maturation of endometrial vasculature across secretory phase. It is reported to be the top up-regulated gene in the mid-secretory phase in the joint analysis of published endometrial microarray data[256]. Another study[266] has reported its up-regulation in secretory phase with mRNAs localised to the glandular epithelial cells by *in situ* hybridization; they proposed it is regulated by progesterone to recruit uterine natural killer cells for perturbation of leukocyte population in endometrium. *CXCL14* is also included in another 238-gene panel for endometrial receptivity evaluation[267]. The fifth gene is Decidual Protein Induced by Progesterone 1 Autophagy Regulator (*DEPP1*, previously known as *DEPP/C10orf10*) that was originally reported in stromal cells of decidualizing endometrium as its name suggests[268]. It is also reported in the 238-gene endometrial receptivity panel[267], and found to be up-regulated in mid-secretory compared to early-secretory phase in a microarray study[257]. The protein plays a role in autophagy, i.e. cells clean out any damaged or unnecessary components, in an increased reactive oxygen species (ROS) environment[269]. As the changing level of autophagy across the cycle may contribute to apoptosis of endometrial cells[270], this gene may be involved in the breakdown and dissociation of glands and stromal layers towards the end of a cycle. The top five genes from the DE between secretory vs. proliferative phase endometrium were all reported in previous literature with their potential roles specific to endometrium physiology, which gives confidence in the results. The DE gene list will be referenced when exploring DE genes by endometriosis status in Chapter 5 to avoid reporting genes manifested mainly due to cycle phase differences as DE genes for the disease status.

In my differential expression analysis between mid and late-secretory phases, the number of significantly DE genes was much smaller (n=21) than between the proliferative and secretory phases. The same was observed in Fung's[3] study when compared to their list for proliferative versus early-secretory samples. This may be explained by a decreased variability in gene expression in a smaller portion, i.e. day 20 to 28, of the cycle; as well as a smaller sample size that would necessarily result in lower power of detection of differential expression effects. The top three differentially up-regulated genes in late-secretory phase in my analysis (Supplementary Table S 4.4) all encode for cytoskeletal proteins: the top gene Desmin (*DES*) encodes a muscle filament and mediate protein binding[271], which, if mutated, may result in myopathy[272]. The protein was found in stromal cells around vessels in basal endometrium in the secretory phase in a previous study[273], and is believed to be regulating angiogenesis in the menstrual cycle. It is traditionally one of the immunohistochemical markers for endometrial stromal sarcoma and uterine leiomyosarcomas[274,275]. The second gene, Myosin Heavy Chain 11 (*MYH11*), encodes a subunit of smooth muscle myosin for contraction; the third, PDZ and LIM Domain 3 (*PDLIM3*), encodes protein that binds to alpha-actinin-2[276] to facilitate cytoskeletal assembly. There is no previous report of *MYH11* and *PDLIM3* expression specifically in healthy endometrium in the literature. Different levels of cytoskeleton regulation across the menstrual cycle is believed to be crucial for maintaining polarity and intercellular junctions of endometrial epithelial cells in its healthy state[277]. It may play a role in the rapid glandular and stromal degradation in late-secretory phase. Interestingly, cytoskeletal proteins, especially alpha-actinin, have been previously reported to be down-regulated in endometriotic lesions when compared to normal endometrium[278]. Thus, there may be shared mechanisms in normal physiology of regulating late-secretory endometrium and features present in endometriotic lesions.

As the attempt to gate and assign cycle phase to samples of unknown phase through visualisation of transcriptomic profile is study-specific and labour intensive for histological sample annotation, building a multivariate model to predict cycle phase based on expression level of a few markers would be beneficial to future studies. The largest available comparison groups with differential expression data was the proliferative (n=23) vs. secretory (n=53) phase groups, and arguably being able to assign at least these two phases accurately would be of greatest relevance for subsequent work. However, due to the small number of proliferative samples, model building that typically requires a training set of model building, and an independent test set for model validation was not possible; I allocated all available samples to a training set to maximise its power of prediction. An independent validation set is needed to verify the model proposed in future studies.

The selection of a panel of 26 genes for step-wise logistic regression was based on the fact that they were reported with the same direction of effect in all three studies ($\log_2$FC>|1| in-house, Fung et al.[3] and Talbi et al.[2]); this boosted confidence that these differential expressions were not incidental findings from a single study. As this selection was not based on significance levels, only 14 out of 26 genes were ranked in the top 1000s (ranked from 6th to 6,578th) of genes in the DE lists from the three studies (Supplementary Table S 4.5), and may not be the group of genes that accounted for most transcriptional differences between the two phases. Ideally, only genes that are not highly correlated in expression levels should be included for model building; the correlation of expression pattern across 26 genes (Supplementary Figure S 4.9) was not overwhelming to prevent further analysis. It is expected to have some degree of correlation based on how the genome works. The correlation heatmap can help identifying if the proposed model includes variables that are highly correlated which may need extra attention.

Although all 26 genes were significantly differentially expressed in my dataset, only one, *ARG2* (ranked 74th in the DE list by adj.*p*), showed significance in the single-gene logistic regression model comparing proliferative with secretory phase (Supplementary Table S 4.5). Despite to effort to make sure the normalisation of raw count data in two analyses were the same, the difference may arise from the statistical prior of DE pipelines modified for RNA-seq data distribution. Unsurprisingly, a single-gene model including only *ARG2* had a performance not far from a non-discriminating test (AUC=0.5), with AUC=0.66.

The 18-gene model resulting from statistical model selection had a much better predictive performance with 92% accuracy and AUC=0.96 (Figure 4.13). However, it is very important to note that the performance of the model was not based on an internal or external validation (test) set, but on the training set that would inflate the performance parameter greatly. In addition, this model was suggested by backward step-wise logistic regression and may have resulted in overfitting. A future independent validation set can assess the degree to which overfitting will have been present.

There has not been much success in the building of a prediction model based on the endometrial transcriptome in the literature[279]. There are likely a number of reasons for this. Firstly, endometrium is very dynamic and fast changing tissue; secondly, a large sample size is required to allow for enough samples in each phase for comparison, especially for sub-secretory phases; lastly, obtaining well-timed samples with histological confirmation are often time consuming. My analyses showed the difficulties in finding a few representative genes to capture transcriptomic differences in a wide window (day 8 to 28) of the menstrual cycle. With sufficient numbers of histology-annotated late-proliferative and early-secretory samples, focused on a narrower window, i.e. on pre/post ovulation, it may be possible to have a simple multi-gene model that can capture the precise physiological shift on the transition day.

In the pathway enrichment analysis I aimed at uncovering normal physiological differences between proliferative and secretory endometrium through interrogation of six databases as each has different focuses and strengths. Most include pathways focused on cellular and metabolic signalling[280]; Ingenuity knowledge base from IPA also includes gene regulation, protein to protein, and protein to compound interactions. Results obtained from these different databases should therefore be considered as complementary and are not easily compared. Consistent with the differential expression analysis (Figure 4.8), the majority of the enriched pathways were up-regulated in secretory vs. proliferative phase. The overall results from these databases collectively suggested the following pathways as enriched and up-regulated in secretory phase of the endometrium: complement and phagosome system under innate immune responses; natural killer and helper T cells signalling; coagulation signalling; and pathways for cellular movement, adhesion and proliferation. These results fit with the general physiological characteristics of a secretory endometrium which is responsible for lining maintenance and receptivity in the implantation window (involving immune cell modulation), and the subsequent glandular exhaustion, stromal cells condensation, dissociation and degradation of linings which involves granulocytes and lymphocytes infiltration, as well as controlled phagocytosis[281]. Although the IPA database was traditionally developed for studying heart and vascular conditions, I did not observe any enriched pathway relating to angiogenesis and vascular activities in secretory phase despite its known characteristic growth of coiled-up arterioles (i.e. spiral artery) supplying blood to the endometrium. From the 'disease and disorder' category in IPA, the 'endometriosis pathway' involving 89 molecules curated from literature was suggested to be enriched as one of the top 10 disease pathways (p=1.73x10^{-27}, Supplementary Table S 4.8). Even though around 70% of the sample utilised for this study were from individuals suffering from endometriosis, the proportions of endometriosis cases in both cycle phases being compared for differential expression were very similar (73% in proliferative phase, and 70% in secretory samples).

For the PANTHER pathway analyses, the pathways suggested as enriched all had a limited number of genes (e.g. n=3) that overlapped with the input DE list; none passed the threshold for multiple-testing correction. The suggested enriched 'Huntington disease' pathway was due to the highly differentially expressed *HAP1* gene, which I discussed above.

For the Reactome pathway results (Supplementary Table S 4.9), two main pathways were highlighted as down-regulated in secretory phase: 1) the 'oestrogen dependent pathway' and 2) 'DNA replication, repair and cell cycle and its checkpoints'. Both fit with expectation. Oestrogen levels fall sharply to baseline after ovulation in early-secretory phase. Furthermore, the transition from proliferative to secretory phase is characterised by the cessation of active mitosis. The secretory lining maintains its cellular content without replenishing and 'proliferating' characteristic of the proliferative phase. The up-regulated enriched pathways suggested by Reactome database were common to findings from other databases, highlighting similar complement, platelet and cytokine pathways.

The enrichment of upregulation of complement and coagulation pathways suggested by multiple databases in my analysis are backed by a similar report in Talbi et al.[2], but not discussed in Fung et al.[3] and Ponnampalam et al.[1]. The complement family molecules are extensively reported across five cycle phases in the Talbi et al.[2] study, with specific mention of their peak in the mid-secretory phase (*DAF*, Adipsin, *C1* to *C4*). The coagulation pathway was also mentioned in their comparison of proliferative versus early-secretory experiment without detailing the direction of effect.

My analyses showed up-regulation of genes in multiple levels of the complement cascade in secretory endometrium. As shown in Figure 4.16, not only the soluble complement components (e.g. *C2*, *C7*) were up-regulated, but also the membrane-bound regulators (e.g. *ITGAM*, *ITGAX*). The complement system may facilitate clearance of cellular debris and immune cells when the lining is exhausted later in the secretory phase without triggering an overwhelming inflammatory response through opsonisation. The complement family would be a good target to follow-up on for model building for cycle phase prediction given its dynamic expressions between proliferative and secretory phases from my results, and across the cycle as suggested by previous literature[82]; in addition, the complement system would be worth explored for infertility-related pathogenesis[282], one of the clinical manifestations of endometriosis.

4.5 Conclusion

My aim was to characterise the transcriptome of eutopic endometrium across the menstrual cycle phases, and investigate the utility of transcriptomic profiling to adjust for menstrual phase in omics analyses. My results showed that cycle phase determination by transcriptional profiling of eutopic endometrium was challenging for multiple reasons, as outlined. Therefore, cycle phase derived from self-reported last menstrual period will be used for subsequent disease-focused analyses in this thesis if histology assignment was not available. Results from two differential expression analysis pipelines agreed well; in accordance with current literature, the `DESeq2` pipeline will be used for subsequent analysis. My results, and those of others, showed that there is a significant cycle phase effect on gene expression patterns that could mask potential transcriptional differences between disease state and normal endometrium samples. Therefore, it is a crucial factor that needs to be adjusted for in subsequent analysis involving differential expression analysis in endometrium. Lastly, multiple pathway enrichment analyses commonly highlighted the complement system as one of the most robustly enriched and up-regulated pathways in the secretory phase of the endometrium compared to the proliferative phase. Focusing on members in the complement cascade may aid future model building for cycle phase prediction through transcriptomic profiling.

Chapter 5 The endometrial transcriptome in endometriosis

5.1 Introduction

It is proposed that endometriosis tissue may originate from the eutopic endometrium. Therefore, investigating potential differences in transcriptomic signatures of the eutopic endometrium from women with and without endometriosis is important for understanding the aetiology of the disease. The results in chapter 4 demonstrated, in line with previous studies, that the transcriptomic profile of human eutopic endometrium varies hugely across menstrual cycle phases, especially between proliferative and secretory phases. Thus, the comparison of the endometrial transcriptome from women with and without disease is confounded by the cycle phase effect. Taking account of cycle phase difference in analyses can be done by adjusting for cycle phase as a covariate in differential expression analysis, or to investigate differences within a specific cycle phase (stratified analyses).

There are a limited number of studies describing transcriptome differences in the eutopic endometrium between endometriosis and control to date (Section 1.10). In 2007, Burnet et al.[144] reported substantial differences in expression profiles of eutopic endometrium by endometriosis status within each cycle phase, using microarray technology. It was the first study with an attempt to control for transcriptomic variation between cycle phase by looking at one cycle phase at a time; however, the study had a very small sample size (n = 3 to 9) which may have influenced the interpretability of the results (Table 5.1). In contrast, other reports[145,146] to date have suggested no to minimal transcriptomic differences in eutopic endometrium from endometriosis cases compared to controls when correcting for multiple testing. The analysis by Fung et al.[145] was based on Illumina Human HT-12 v4.0 Expression Beadchip microarray and treated cycle phase as a covariate to correct for menstrual effect in their model (n=123), reporting no significantly expressed genes after multiple testing correction; Mortlock et al.[146] reported nominally significant differential expression (DE) data by RNA- sequencing (RNA-seq) only in mid-secretory phase samples (n=41), and no significant results were reported from mid-proliferative eutopic endometrium between endometriosis and control (n=72).

Whilst there remains a relative lack of data, in particular RNA-seq, on differential endometrial expression patterns in endometriosis vs. controls, no study to date has investigated differences between women with different surgical sub-phenotypes of the disease with controls despite GWAS data strongly suggesting differential regulation for ASRM stage III/IV and stage I/II disease. With the availability of extensive sub-phenotypic data, the ENDOX dataset is highly suited to the investigation of potential transcriptomic differences between ASRM stages.

In this Chapter, the transcriptome of human eutopic endometrium in endometriosis is explored within proliferative and secretory phases separately using the same RNA-seq dataset described in chapter 4. Included is: 1) differential expression analysis between endometriosis status and surgical sub-phenotypes (ASRM Stage I/II and Stage III/IV) vs. clinic controls; 2) comparison of the results from the differentially expressed genes between cycle phases from the same dataset (Chapter 4); 3) comparison of differentially expressed genes with previous reports in the literature[144-146]; 4) endometrial *cis*-expression quantitative trait loci (eQTL) analysis to investigate the presence of regulatory single nucleotide polymorphisms (eSNPs) near (within 1Mb) the identified differentially expressed genes; and 5) investigation of differentially expressed genes near 16 endometriosis genetic risk loci shown to be nominally associated with an endometriosis sub-phenotype in Chapter 3.

Table 5.1 Publications with differential expression gene lists published on human eutopic endometrium between women with and without endometriosis and their comparison criteria.

Publication	**Endometriosis Case vs. Control Comparison in Particular Cycle phases**		**Sample Size Case:Control**	**Expression Measured by**	**No. of Differentially Expressed Genes Reported (FDR-*p*<0.05: *p*<0.05)**	**Total no. of Probes Tested**
Mortlock et al. 2020[146]	All cycle phases; corrected for cycle phase effect		143:63	Illumina HiSeq 4000 & 2000 (Seq depth: 37.5 M)	0: 0	17,022
	Mid-proliferative		56:16		0: 0	
	Mid-secretory		27:14		0: 43	
Fung et al. 2018[145]	Differential expression	All cycle phases; corrected for cycle phase effect	133:89	Illumina Human HT-12 v4.0 Beadchips	0: 486	12,321
	On/ off gene				NA: 675	9,626
Burney et al. 2007[144]	Proliferative		6:5	GeneChip™ Human Genome U133 Plus 2.0 Array	289: NA	54,600
	Early-secretory		6:3		1,211: NA	
	Mid-secretory		9:8		282: NA	

5.2 Materials and methods

5.2.i Sample grouping by cycle phase on RNA-seq data

The RNA-seq dataset generated from 195 eutopic endometrial pipelle biopsies, as described in chapter 4.2.i, was used for this analysis. The same raw counts table with reads mapping to 52,190 Ensembl genes (GRCh37/hg19) was utilized. From the 190 samples passing the initial quality control (QC), samples were assigned to proliferative or secretory phase for subsequent analyses. A total of 52 samples were classified as proliferative by histological dating (40%; n=21), as described in Section 4.2.ii, and supplemented with self-reported proliferative phase (60%; n=31) when histology was not available. A total of 104 samples were classified as secretory phase, with 52% (n=54) assigned from histology and 48% (n=50) from self-reported data.

Within proliferative phase samples, 39 were from women with surgically confirmed endometriosis (ASRM Stage I/II n=22, Stage III/IV n=17) and 13 were clinic controls from women referred to our tertiary clinic for laparoscopy due to various reasons, such as pain symptoms or infertility, but without an endometriosis diagnosis. For secretory samples, 37 (36%) were classified as Stage I/II cases, 32 as Stage III/IV cases, and 35 as clinic controls.

5.2.ii Further sample quality control

Further data cleaning was performed for each cycle phase group individually. Firstly, genes with less than the total of five reads across samples in the group were manually removed; then, library sizes and variance were normalised and stabilised using the `rlog` function in `DESeq2` package. For proliferative samples (n=52), three rounds of QC were done until there were no obvious outliers, by observing multi-dimensional scaling (MDS) cluster plots, to obtain the final set with 44 samples. Supplementary Figure S 5.1 illustrates the workflow of identifying outliers in detail. An additional round of outlier removal, determined by clustering of samples in the MDS plot, was done to obtain a smaller subset with 38 samples for sensitivity testing (see bottom panel of Supplementary Figure S 5.1). For secretory samples (n=104), five rounds of sample QC were done based on MDS plots and normalised log-transformed counts until no outlier was identified (Supplementary Figure S 5.2). The final set contained 91 samples. Principal components (PC) were also calculated based on both sample groups after sample QCs to determine the proportion of variance explained by the first 10 PCs.

5.2.iii Differential gene expression in eutopic endometrium between endometriosis cases and clinic controls

5.2.iii.a Comparison made within proliferative phase

For proliferative samples, a simple model with a single term (`design1 ~ endometriosis case-control status`) was used without adjusting for any covariates. Dispersion estimation and generalised log-linear model (GLM) fitting were done through `DESeq`, a function in `DESeq2`. Comparison of all endometriosis cases versus controls, stage I/II cases only versus controls, stage III/IV cases only versus controls, and stage I/II versus stage III/IV were conducted on 44 samples, as well as on the smaller subset of 38 samples for sensitivity analysis.

Automated independent filtering of genes with false discovery rate (FDR) cut off set at 5% (reported as adj.p<0.05) was performed with the `result` function. Markers were annotated to human Ensemble GRCh37 with `biomaRt`. Top differentially expressed genes were visualised utilising volcano plots produced in `EnhancedVolcano` package.

5.2.iii.b Comparison made within secretory phase

For 91 secretory samples, apart from using the simple model with only one term for disease case-control status for analysis, another model taking potential expression differences between sub-secretory phase into account (`design2 ~ endometriosis case- control status + sub-secretory phase`) was also used. Possible interaction between two terms: case-control status and sub-secretory phase, was investigated through principal component analysis (PCA) on the variance stabilising transformed (VST) data. After dispersion estimation and GLM fitting, likelihood ratio test (LRT) was done to test for statistical differences between two models. Similar to the analysis for the proliferative group, four comparisons of all endometriosis cases, stage I/II cases only, stage III/IV cases only versus control respectively, and stage I/II versus stage III/IV were conducted, with FDR set at 5%.

5.2.iv Comparison of differentially expressed genes by cycle phase, literature reports, and endometrium eQTL map

5.2.iv.a Differentially expressed genes by cycle phase

As the cycle phase grouping was based on both histology and self-reported data, lists of DE genes from each of the above comparisons were compared against the results from Chapter 4 to identify expression patterns that may capture the cycle phase effect, rather than reflecting a difference between case-control status. The full list of DE genes between histological-confirmed proliferative and secretory samples (Supplementary Table S 4.2) was compared to the DE results for endometriosis status in the proliferative group, based on adjusted *p*-value (adj.*p*). Similarly, the list of DE genes between mid-secretory and late-secretory samples (Supplementary Table S 4.4) was compared to the results from the overall secretory phase group.

5.2.iv.b Comparison of differentially expressed genes with published endometrium-based DE datasets

For comparison of study results to those reported in the literature, genes (with $p<0.05$) mentioned in Fung et al.[145] case-control DE analysis, and on/off gene prediction in endometrium ($p<0.05$, with positive estimates denoting over-expression in cases than in controls) were compared against DE results from both proliferative and secretory groups in this study. Reports from Mortlock et al.[146] focusing on mid-secretory samples were compared to the results from the secretory group only. Similarly, for genes reported as differentially expressed by endometriosis case-control status in the proliferative samples by Burney et al[144] were compared to my results in proliferative group, and those reported in the early and mid-secretory samples were compared to my results for the secretory group. The total number of uniquely reported, nominally significant, DE genes, collectively from the above three studies for proliferative samples were 1,373, and 2,326 for secretory phase.

5.2.iv.c Comparison of differentially expressed genes with human eutopic endometrium eQTL map

A *cis*-eQTL map, with 'cis' defined as a one megabase (Mb) window of the transcriptional start site (TSS), was generated previously based on the expression of 163 human eutopic endometrium samples from the ENDOX study by Dr Nilufer Rahmioglu for an ongoing project. The preparation of expression data was the same as described in chapter 4.2.i with reads mapped to the GRCh38, and with transcript assembly and gene count data obtained using the `StringTie`[283] programme. The eQTL discovery analysis was based on 18,145 genes which are expressed in 90% of the samples with counts per million (CPM) larger than 0.22. The first three PCs were taken into account as covariates, with the enforcement of normal distribution upon expression levels by `QTLtools`[284] package, and permutation scheme at 1000 to map QTLs in *cis*. Association p value<0.05 after adjustment of number of variants tested in *cis* by beta distribution are marked as eSNPs for eutopic endometrium. Further FDR correction on p value (FDR p) was also performed. The chromosome and position of eSNPs from the eQTL map were lifted down to GRCh37 to match with this study; overlapping of genes showing nominal DE results ($p<0.05$) from case-control comparisons with eSNPs (within 1Mb range) were reported.

5.2.v Differential expression of genes around GWAS loci for endometriosis sub-phenotypes

The 16 genetic variants showing nominal associations with at least one of the endometriosis sub-phenotypes, as described in chapter 3 (Table 3.5), were selected. Genes locating ±500kb around these 16 loci was generated by the UCSC genome browser[285]. A Bonferroni correction for the 16 lookups was applied on top of the *p* and adj.*p* from the DE analysis. The same method was applied to lookup DE status of any gene families of interest.

5.3 Results

5.3.i Cluster analysis of samples by proliferative and secretory phase groups and model determination for DE analysis

5.3.i.a Proliferative phase samples

The MDS plot for the 44 proliferative phase samples (bottom left of Supplementary Figure S 5.1) shows no obvious outlying clusters nor separate clustering between cases and controls, with 16.3% variance explained by the first PC, and 10.9% by PC2; 63.1% of variation is cumulatively explained by the top 10 PCs. There is also no outlier in normalised log-transformed gene counts. Among the 44 samples, 18 are stage I/II cases, 16 are stage III/IV cases, and 10 are clinic controls. The smaller subset of a tighter cluster of 38 samples for sensitivity testing consisted of 17 stage I/II, 14 stage III/IV, and 7 controls. The results from the analysis of 44 samples was presented, and supplemented with those from the smaller subset for sensitivity testing purpose. The one-term design model, `design1 ~ endometriosis case-control status`, was used.

5.3.i.b Secretory phase samples

Differences in expression profiles of three sub-secretory phases were firstly reported by Talbi et al[82]. Results presented in chapter 4 (Figure 4.12) supported the existence of differences, and led to the investigation of variation in DE between sub-secretory phases in this secretory group. The MDS plot for the 91 secretory phase samples (Figure 5.1 & bottom left of Supplementary Figure S 5.2) shows no clustering by sub-secretory phases nor by case-control status, with 25.9% variance explained by the first PC, and 13.7% by PC2; 63.4% of variation is cumulatively explained by the top 10 PCs. There are also no outliers in normalised log-transformed counts as shown in bottom right of Supplementary Figure S 5.2 after data cleaning. The endometriosis case-control and ASRM stage breakdown across sub-secretory phases is shown in Table 5.2. A higher proportion of late-secretory (42%) and lower proportion of mid-secretory (23%) samples are in stage III/IV case group when compared to stage I/II (38% vs. 32%) or to the control group (39% vs. 26%).

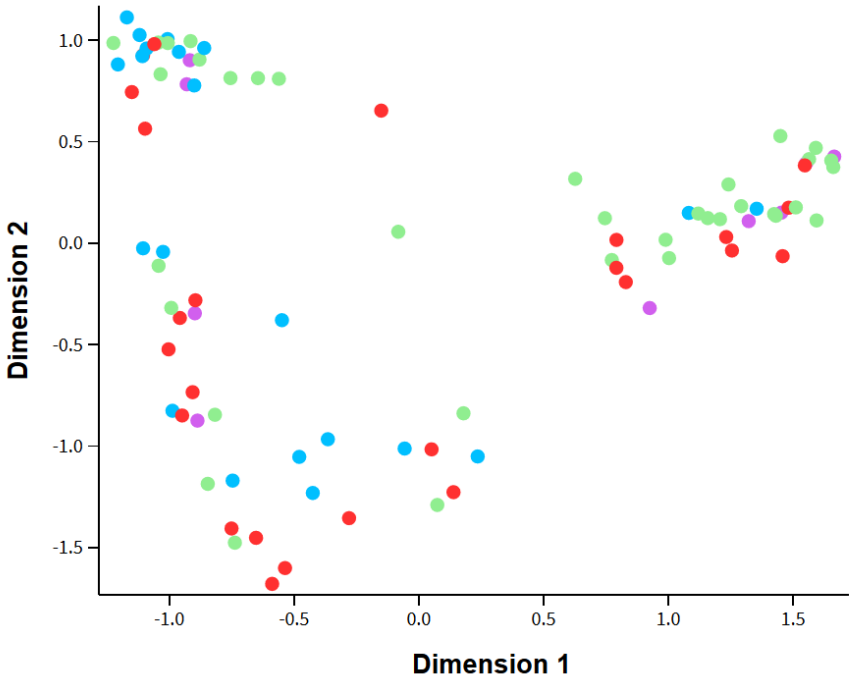


Figure 5.1 MDS plot for secretory samples (n=91) for subsequent differential expression analysis on endometriosis disease status.
Key: blue dots = early-secretory phase, red dots = mid-secretory phase, green dots = late-secretory phase, purple dots = secretory samples without sub-secretory phase information

There is no specific clustering of endometriosis cases-controls within sub-secretory phases in the PCA plot (Supplementary Figure S 5.3). However, when comparing the two-term model `(design2 ~ endometriosis case-control status + sub-secretory phase)` to the simpler model (`design1 ~ endometriosis case-control status)`, result of the likelihood ratio (LRT) test showed that 7.9% of genes (3,375 out of 42,997) are significantly different (adj.p<0.05) between the two model designs, suggesting the inclusion of the term for sub-secretory phase as a covariate can better explain the differences between cases and controls. Thus, results based on the two-term model will be presented, supplemented with the results from the simpler model for comparison only.

Table 5.2 Number of endometriosis cases and clinic controls across three sub-secretory phases in the secretory phase group (n=91).

	Early-secretory		**Mid-secretory**		**Late-secretory**		**No Sub-secretory Classification Available**		**Total**	
	n	%	n	%	n	%	n	%	n	%
Control	9	29	8	26	12	39	2	6.5	31	34
Case	12	20	17	28	24	40	7	12	60	66
Stage III/IV	6	23	6	23	11	42	3	12	26	29
Stage I/II	6	18	11	32	13	38	4	12	34	37
Total	21	23	25	27	36	40	9	10	91	100

5.3.ii Differential gene expression in eutopic endometrium between endometriosis cases and clinic controls

5.3.ii.a Proliferative phase

Out of 52,190 genes, 40,735 were tested for differential expression between endometriosis cases versus controls, and stage III/IV cases versus controls. For stage I/II versus controls, and stage I/II versus III/IV, 29,678 and 24,940 genes were tested, respectively, after independent gene filtering with FDR correction.

In the comparison of expression profiles from all endometriosis cases (n=34) and clinic controls (n=10) in the proliferative phase of the endometrium, three genes (*ZNF26*, *MLXIPL*, *SAA2*) and three pseudogenes were differentially expressed (adj.p<0.05), and 2,386 genes were nominally differentially expressed (p<0.05) (Volcano plot in Supplementary Figure S 5.4). None of the DE genes were featured in the top 50 genes from the comparison of expression profiles markedly different between proliferative and secretory phases described in chapter 4 (Supplementary Table S 4.2). From the case-control comparison, out of the top 50 genes, ranked by p value, three (*GIPR*, *FAT2*, *TTC29*) were located within the 1Mb window of an eQTL SNP (eSNPs, p<0.05) in endometrium. Details of these genes and the overlap with endometrium eSNPs can be found in Supplementary Table S 5.1. The DE results from the reduced set (n=38), and matching to literature reports, are also noted in the table. It is worth noting that the top DE gene *ZNF26* is not featured in the sensitivity analysis of 38 samples (ranked 10,735th by adj.p); further investigation is presented in Section 5.3.iii.a.

The largest number of significantly differentially expressed genes, 51 (adj.p<0.05), was observed in the comparison of stage I/II cases (n=18) with controls (n=10) within the proliferative phase in the endometrium, with 2,700 nominally significant differentially expressed genes (Figure 5.2, Supplementary Table S 5.2). Again, no DE genes overlapped with those reported in the previous differential expression analysis comparing between proliferative and secretory transcriptomic profiles. Almost all DE genes (98%), except for *HBA1*, are under-expressed in stage I/II when compared to controls as shown in Figure 5.2. There are endometrium eSNPs with regulation ranges covering two out of 50 top DE genes, one upstream of *SPATA6L* and one downstream of *ANKRD65*.

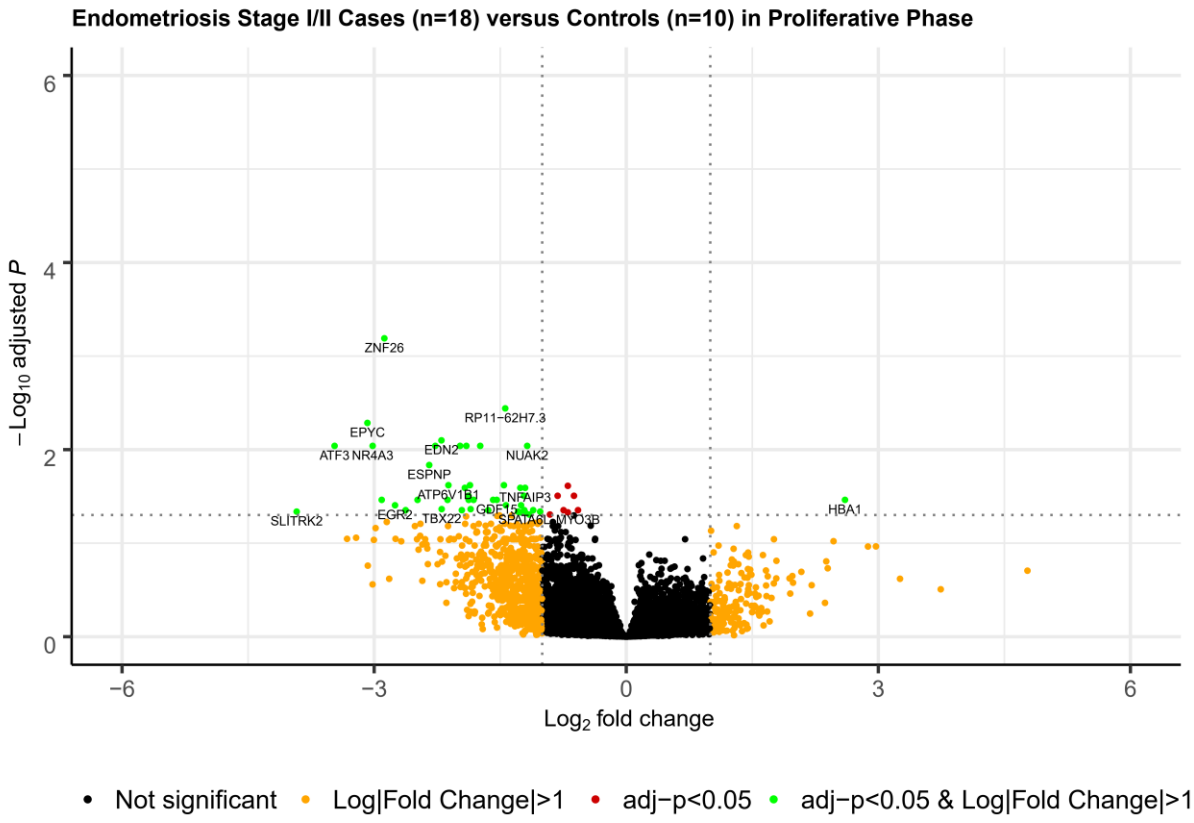


Figure 5.2 Volcano plot for differentially expressed genes between endometriosis stage I/II cases (n=18) and clinic controls (n=10) in the proliferative phase of the endometrium.
Key: green dots = gene with both adj.p<0.05 & $\log_2$ of absolute fold-change (Log_2 |FC|) >1, red dots = gene with adj.p <0.05 only, orange dots = gene with Log_2 |FC|>1 only, and black dots = gene with adj.p>0.05.

In contrast, only one gene (*ZNF26*) showed significant differential expression when comparing stage III/IV cases (n=16) to controls (n=10), with 2,321 genes passing p<0.05 (Supplementary Figure S 5.5, Supplementary Table S 5.3). *ZNF26* is not a significantly differentially expressed gene between proliferative and secretory phases. One of the top 50 genes, *C1orf85* (also named *GLMP*), was within 1Mb of an endometrium eSNP.

The comparison of endometrium from stage I/II (n=18) with stage III/IV cases (n=16) showed 22 significant DE genes, with 2,574 genes in total showing nominal $p<0.05$ (Figure 5.3, Supplementary Table S 5.4). Two DE genes (*C10orf10* ranked 16th also named *DEPP1,* and *CITED2* at 21st) were in the gene list for differentially expressed genes between proliferative and secretory phase, ranked 5th and 9th respectively by adj.*p* in the previous analysis described in Supplementary Table S 4.3. Most DE genes (91%) are over-expressed (with positive fold changes) in stage III/IV when compared to stage I/II cases (Figure 5.3). It is notable that the top three DE pseudogenes, *ANKRD20A11P*, *CYP4F29P* & *SNX18P13*, are in close proximity with each other in the genome at the same cytoband 21q11.2; follow-up of this genomic region is presented in Section 5.3.iii.b. Out of the top 50 genes ranked by *p*-value, three genes (*THBS1*, *KBTBD11* & *FAT3*) and a pseudogene *RP11-347C12.2* were within the 1 Mb window of endometrium eSNPs. Thrombospondin 1 (*THBS1*) is also located within ± 500kb of an endometriosis risk variant, identified via GWAS, situated at 3'UTR of *BMF* (and intronic to *SRP14-AS1*); details in Section 5.3.iv .

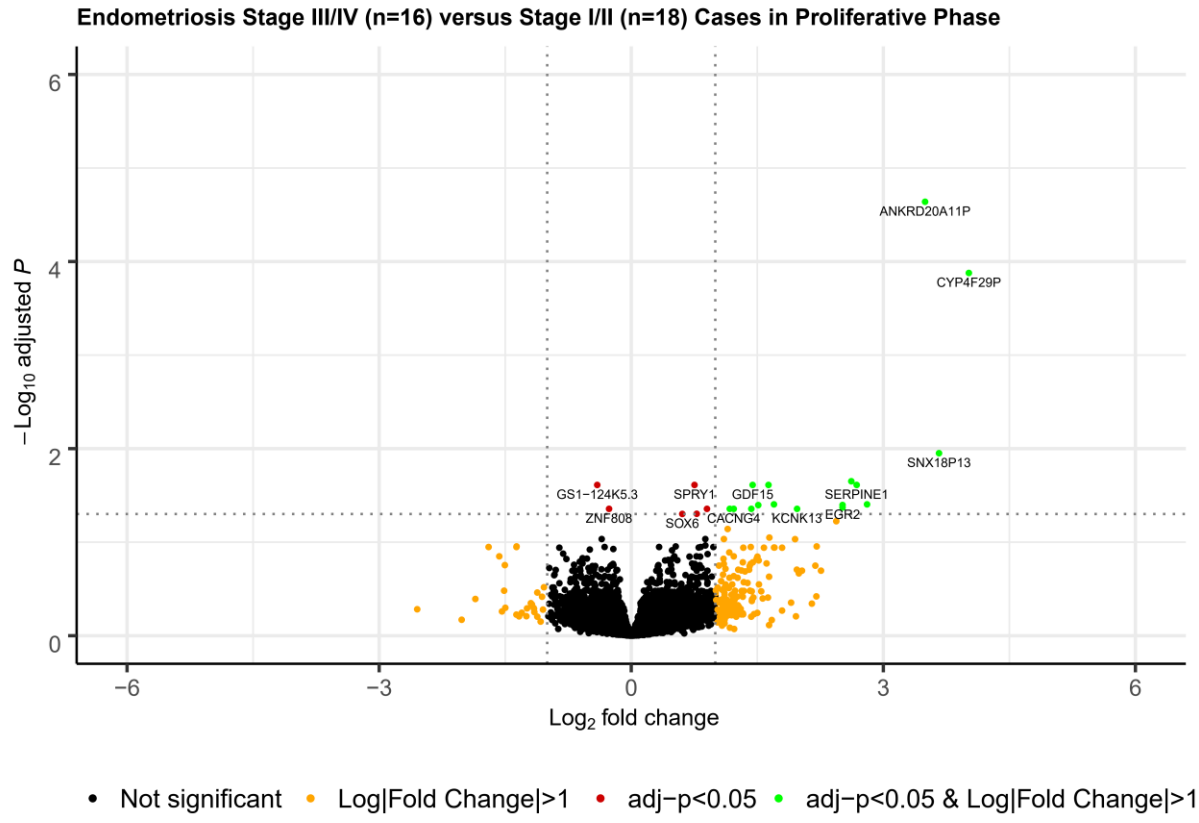


Figure 5.3 Volcano plot for differentially expressed genes between endometriosis stage I/II (n=18) and stage III/IV cases (n=16) in the proliferative phase of the endometrium.
Key: green dots = gene with both adj.p<0.05 & Log_2 |FC|>1, red dots = gene with adj.p <0.05 only, orange dots = gene with Log_2 |FC|>1 only, and black dots = gene with adj.p>0.05.

Similar to the pattern of the above, in the reduced subset of 38 proliferative samples, the largest number of DE genes (n=22) is observed in the comparison between stage I/II cases and controls; where the least number of DE genes (none) is observed in the comparison of stage III/IV to controls. Out of 22 significant DE genes in the stage I/II-control comparison, 17 are also significant in the larger set. In proliferative samples, more DE genes are observed in stage I/II cases against controls than stage III/IV.

5.3.ii.b Secretory phase

Out of 52,190 mapped genes, 42,997 were tested for comparison of endometriosis cases versus controls, and 42,962 for stage I/II cases versus controls. For stage III/IV versus controls, and stage I/II versus III/IV, 27,146 and 35,465 genes were tested respectively after independent gene filtering with FDR correction.

No significant differentially expressed genes (adj.p<0.05) were identified between all endometriosis cases (n=60) vs. clinic controls (n=31) in the secretory phase of the endometrium (Supplementary Figure S 5.6). A total of 1,086 genes are differentially expressed with nominal p<0.05. The gene list for the top 50 ranked by nominal p-value is shown in Supplementary Table S 5.5. There is no report of these 50 genes in DE analysis for endometriosis cases against control (within secretory samples) in previously published papers[144-146]; there is also no overlap with previous differential expression analysis comparing mid and late-secretory phases (Chapter 4) regardless of disease status. The DE results from the simple design model without adjustment of sub-secretory phases were slightly different to those with the adjustment, and can be found in the table. Out of the top 50 genes, one gene *F2RL2* and one pseudogene *AC073109.2* were located within 1Mb window of endometrium eQTL SNPs. It is worth noting that more than half (57%) of the hits with nominal p-values were not protein-coding genes, but predominantly pseudogenes (20%) and long intergenic non-coding RNA (lincRNA) (12%).

Only one gene (*ALAS2*) showed significant differential expression when comparing stage I/II (n=34) versus controls (n=31), with 1,160 genes passing nominal p<0.05 (Supplementary Figure S 5.7, Supplementary Table S 5.6) in the secretory phase. Similar to the above comparison, only 47% of nominal hits mapped to protein-coding genes. *ALAS2* is not a significantly DE gene comparing mid and late-secretory phases. The *F2RL2* as mentioned above, and four other pseudogenes, are among the top 50 hits of the stage I/II-control comparison by p-value, and within 1Mb window of eQTL SNPs in the endometrium.

Conversely, the comparison of stage III/IV (n=26) versus control (n=31) yielded 33 significant DE genes, and 2,175 nominally significant genes ($p<0.05$, Figure 5.4, Supplementary Table S 5.7) in the secretory phase. Ninety-one percent of significant DE genes were protein coding; however, protein-coding gene contributed only 51% of all nominal signals. Four DE genes (*PRELP* ranked 1st, *DES* in 3rd, *CHRDL1* in 14th, and *MYH11* at 16th) were among top 21 genes with significant expression differences (adj.$p<0.05$) between mid and late-secretory phase, according to the previous analysis (Supplementary Table S 4.4). Both the direction and magnitude of fold change of these four genes in the comparison of 'stage III/IV versus control' agreed with those obtained in the 'mid vs. late-secretory' contrast. Most of the DE genes (94%), except for *PLG* and *KCND2*, were over-expressed in stage III/IV when compared to controls as shown in Figure 5.4. There was no overlap of the top 50 DE genes within 1Mb window of eQTL SNPs in the endometrium. However there was one gene, *CRELD1*, which ranked 72th, and one pseudogene, *ZDHHC8P1*, ranked 76th within the 1Mb range of the eQTL SNPs at chromosome:position (hg19) of 3:9,975,386 (FDR $p = 1.93 \times 10^{-4}$) and 22:23,750,312 (FDR $p = 0.09$), respectively.

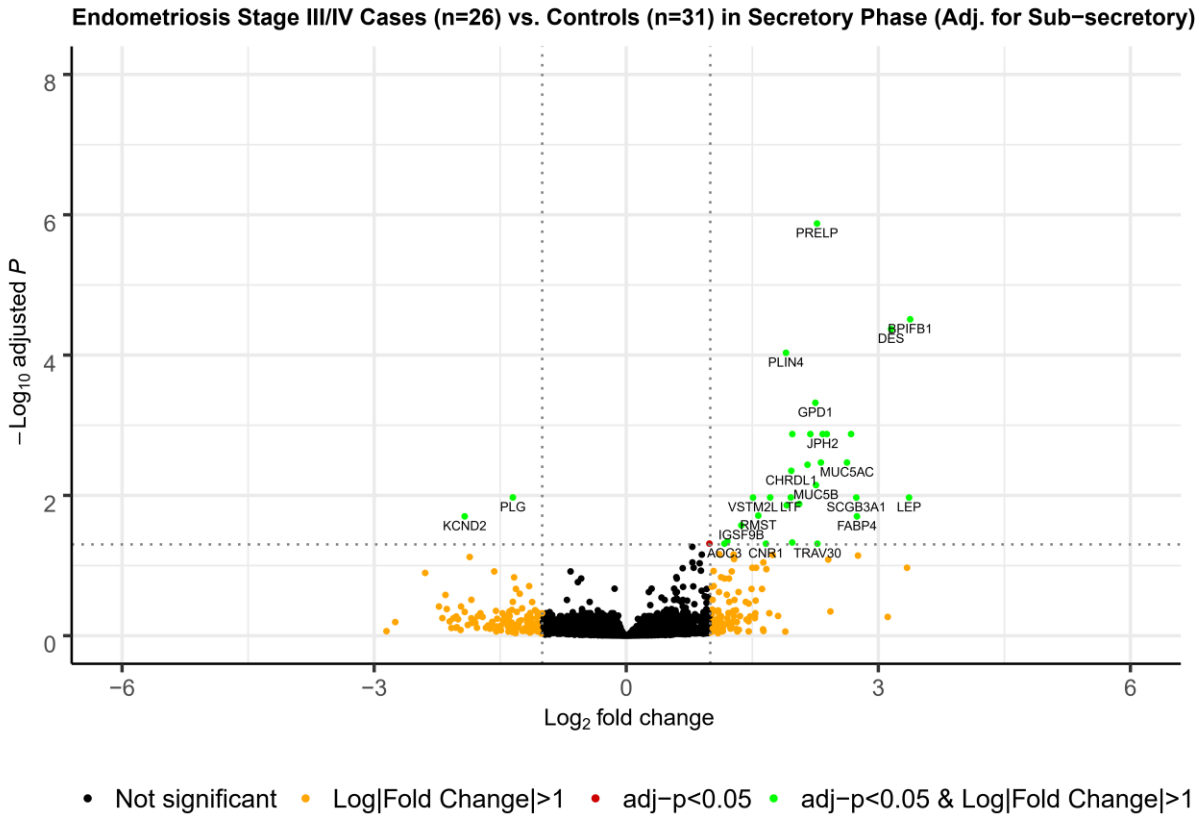


Figure 5.4 Volcano plot for differentially expressed genes between endometriosis stage III/IV cases (n=26) and clinic controls (n=31) in the secretory phase of the endometrium.
Key: green dots = gene with both adj.p<0.05 & Log_2 |FC|>1, red dots = gene with adj.p <0.05 only, orange dots = gene with Log_2 |FC|>1 only, and black dots = gene with adj.p>0.05.

The largest number of DE genes, 134 with adj.p<0.05, was observed in the comparison of stage I/II (n=34) against stage III/IV cases (n=26), with 3,329 genes in total showing nominal p-values (Figure 5.5, Supplementary Table S 5.8). Nevertheless, 12 significant DE genes are also among the top 21 genes with significant expression differences between mid and late-secretory phase; 8 of those genes are denoted in Supplementary Table S 5.8, alongside with *TAGLN*, *SORBS1*, *COL8A1* & *ACTA2* which are ranked >50th in this analysis. Again, the direction and magnitude of fold changes of these 12 genes correlated with those reported in the 'mid vs. late-secretory' contrast (r^2=0.62, Supplementary Figure S 5.8). One out of the top 50 genes, *PTGIS*, was within 1Mb window of an eQTL SNP in the endometrium; however, it ranked fifth in differential expression between mid vs. late-secretory phase of the endometrium. Most of the DE genes (90%) were over-expressed in stage III/IV when compared to stage I/II cases apart from 14 with negative log_2FC, including three pseudogenes (*ANKRD20A11P*, *CYP4F29P* & *SNX18P13*) as shown in Figure 5.5. These are the same pseudogenes located at 21q11.2 cytoband as mentioned in the stage III/IV with I/II comparison in proliferative samples; however, they show opposite direction of fold changes in the secretory phase (the top three most significantly under-expressed in stage III/IV) than in the proliferative phase.

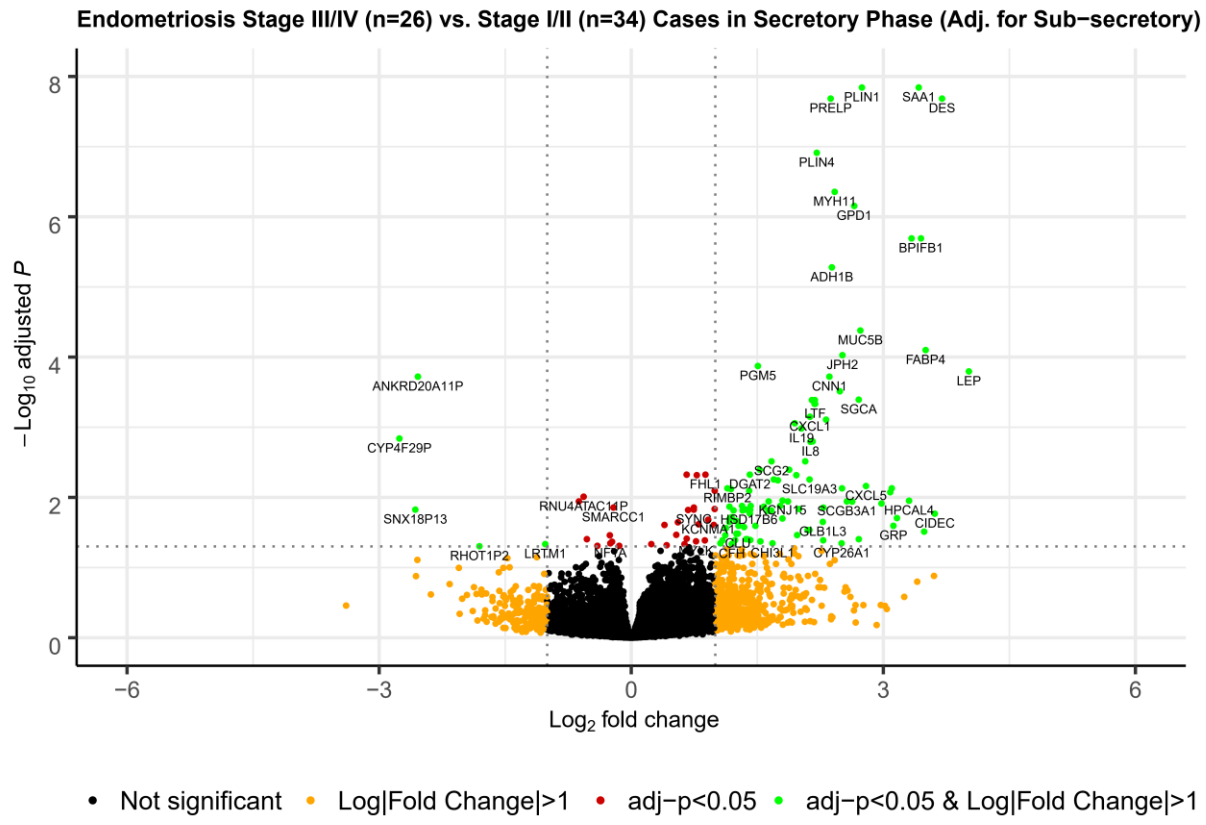


Figure 5.5 Volcano plot for differentially expressed genes between endometriosis stage I/II (n=34) and stage III/IV cases (n=26) in the secretory phase of the endometrium.
Key: green dots = gene with both adj.p<0.05 & Log_2 |FC|>1, red dots = gene with adj.p <0.05 only, orange dots = gene with Log_2 |FC|>1 only, and black dots = gene with adj.p>0.05.

Similar to the pattern of the above results, in the simple design model without adjusting for sub-secretory phase, the largest number of DE genes (n=127) was observed in the comparison between stage I/II and III/IV cases; where the least number of DE genes (none) was observed in the comparison of stage I/II to controls. Out of 127 significant DE genes in the stage I/II-III/IV comparison, 105 are also significant in the two-term model; similarly, 25 out of 30 significant DE genes from the stage III/IV-control comparison by the simple model were also significant in the two-term model. In secretory samples, in contrast to the pattern in proliferative samples, more DE genes were observed in stage III/IV cases against controls than stage I/II.

5.3.iii Further investigations of significantly differentially expressed genes

5.3.iii.a Signal investigation for *ZNF26*

The gene Zinc Finger Protein 26 (*ZNF26*) showed significant differential expression in three comparisons: (1) all endometriosis cases, (2) stage I/II only cases, (3) stage III/IV only cases, against clinic controls in proliferative samples (n=44); however, the signal was not observed in the smaller subset of proliferative samples (n=38). When exploring the distribution of raw counts of reads mapped to the gene from stage I/II samples, stage III/IV, and controls independently, the means (medians) were 8.9 (7), 6.3(6), and 57.2 (9.5) respectively. Box plot (Supplementary Figure S 5.9) showed there are two control samples with extreme raw counts of the gene (at 149 and 342 counts) that drove the differential results. The two outliers were excluded in the smaller subset due to further data cleaning at MSD dimension 1 <-0.5 (bottom left of Supplementary Figure S 5.1).

5.3.iii.b Signal investigation for *ANKRD20A11P*, *CYP4F29P* & *SNX18P13*

The three pseudogenes (*ANKRD20A11P*, *CYP4F29P* & *SNX18P13*) showed significant differential expression in stage I/II versus stage III/IV comparison in both proliferative and secretory groups, but in opposite direction of expression. All three were over-expressed in stage III/IV cases in the proliferative phase (Figure 5.3) when compared to stage I/II, but under-expressed in stage III/IV cases in the secretory phase (Figure 5.5). They were not only located in close proximity (~65kb apart) in the same cytoband 21q11.2, but also showed a clear co-expression pattern. The Spearman's rank correlation of the raw counts of each pseudogene pairs range from ρ=0.74-0.76 (p=1.9x10^{-9}-9.3x10^{-9}) in 44 proliferative samples, to ρ=0.58-0.70 (p=1.8x10^{-9}-1.6x10^{-14}) in 91 secretory samples (Supplementary Figure S 5.10), showing strong positive correlation in their expression patterns.

Box plots for raw counts of the three pseudogenes by stage I/II, stage III/IV and clinic controls for both proliferative and secretory samples are presented in Supplementary Figure S 5.11. In the proliferative group, one control sample showed extreme counts for all three pseudogenes; the sample was also included in the smaller subset for sensitivity testing as it was not clustering separately in MDS plot in the first two dimensions (bottom panel of Supplementary Figure S 5.1), but at the 3rd dimension. In addition, there was one stage III/IV sample that also showed extreme counts in all three pseudogenes; this was excluded in the sensitivity testing set. Despite the exclusion of one of the outlying samples in the sensitivity testing set, *ANKRD20A11P* (adj.p=6.27x10^{-4}) and *CYP4F29P* (adj.p=2.02x10^{-3}) still showed significant differential expression between stage I/II and III/IV, with *SNX18P13* showing a nominally significant result (adj.p=6.51x10^{-2}, p=8.22x10^{-6}). In box plots for the secretory group, one stage I/II sample showed the highest counts, and two controls showed consistently higher counts for all three pseudogenes; these samples were not clustered separately in the top 8 dimensions of the MDS plots. Signals of these three pseudogenes were also captured by the simpler design model (without adjustment of sub-secretory phases); moreover, they did not feature in previous DE analyses (chapter 4) focusing on differences between proliferative and secretory phase, nor in the sub-secretory comparison.

Despite the co-expression pattern observed, the recombination of SNPs in the region is broadly in linkage equilibrium within British European population — the main ancestry of the women recruited — with r^2 ranging from 0 to 0.014 for 67 variants intronic to the three pseudogenes. A heat map of r^2 of variants in the region of chr21:15,200,585-15,352,881 in relation to genomic locations of *ANKRD20A11P* (36.5 kb downstream of *SNX18P13*) and *CYP4F29P* is shown in Supplementary Figure S 5.12; it is generated by the LDmatrix[286] online tool from the NIH national cancer institute. For each of the pseudogenes, its closest protein-coding gene family member was recorded (Table 5.3); there were 190 genes located within ±500kb of these listed genes, of which 114 were mapped and had DE results available in this dataset. However, none of them were significantly differentially expressed in any of the comparisons, after correcting for multiple testing, neither in proliferative nor secretory groups.

Table 5.3 List of protein-coding genes in the same gene family with *ANKRD20A11P*, *CYP4F29P* & *SNX18P13* pseudogenes

Pseudogene (at cytoband 21q11.2)	**Gene Family Name**	**Closest Protein-coding Gene Family Member**	**Cytoband of Gene Family Member**
ANKRD20A11P	Ankyrin Repeat Domain 20	*ANKRD20A2* (also named as *ANKRD20A3*) & *ANKRD20A4*	9p12 & 9q21.11
CYP4F29P	Cytochrome P450 Family 4 Subfamily F	*CYP4F2*, *CYP4F3*, *CYP4F8*, *CYP4F11*, *CYP4F12* & *CYP4F22*	19p13.12
SNX18P13	Sorting Nexin 18	*SNX18*	5q11.2

In addition to the three pseudogenes of interest, *RHOT1P2*, another pseudogene in the same cytoband, showed nominal DE in stage III/IV versus I/II in both proliferative (adj.p=0.06, p= 5.5×10^{-5}, $\log_2$FC=2.44) and secretory phase groups (adj.p=0.05, p=1.9×10^{-4}, $\log_2$FC=-1.81); its direction of fold change agreed with the other three pseudogenes. Three were no eSNPs in the region recorded in the endometrium eQTL map.

5.3.iv Differential expression of genes around GWAS loci associated with endometriosis sub-phenotypes

There were 355 genes located ±500kb around 16 genetic risk variants showing nominal associations with at least one of the endometriosis sub-phenotypes (Chapter 3). Out of the 355 genes, 260 and 273 genes had DE results available in the proliferative and secretory analyses, respectively. Although there were no significantly DE genes among the lookup list after correcting for multiple-testing in both groups, there were two genes with p values (but not adj.p from DE analysis) passing the multiple-testing threshold in the proliferative group.

C2orf50 located 431kb upstream of the GWAS variant rs11674184 (intronic to *GREB1*) showed nominal DE (p=2.43x10^{-4}, adj.p=0.075, $\log_2$FC= -1.28) in the stage I/II versus clinic control comparison. Its DE result in the stage III/IV vs. control analysis was not significant after Bonferroni correction (p=9.91x10^{-3}, adj.p=0.79, $\log_2$FC=-0.91). The expression of *GREB1* itself was not significantly different in any of the analyses performed.

Another gene of interest was *THBS1*, also briefly mentioned in Section 5.3.ii.a in the proliferative group, as it is also within the 1Mb range of an eQTL SNP in endometrium (chr15:38,758,027). The gene is located 462kb upstream of rs12441483 (chr15:40,353,316), a GWAS variant at the 3' untranslated region (UTR) of *BMF* and intronic to *SRP14-AS1*. *THBS1* showed nominal DE (p=1.74x10^{-5}, adj.p=0.039, $\log_2$FC= 1.70) in the stage III/IV versus I/II comparison. Its DE result in the stage I/II versus control comparison was not significant after Bonferroni correction (p=4.42x10^{-4}, adj.p=0.091, $\log_2$FC= -1.59). Similarly, neither the expression of *SRP14*-AS1 nor *BMF* were significantly different in any comparisons performed.

5.4 Discussion

This study is the first in the field to look at transcriptomic differences in the eutopic endometrium of endometriosis cases by ASRM surgical sub-phenotype in two main cycle phases. The results showed a limited number of differentially expressed genes (none for secretory phase, six for proliferative phase) between overall endometriosis case and controls after adjusting for the number of genes tested, which is in line with previous studies by Fung[145] and Mortlock[146]. Power calculation by our team in previous publication[287] has estimated that more than 500 tissue samples are required for robust detection of differential expression of transcript. However, when the ASRM surgical sub-phenotypes were compared with controls (1) more DE genes were observed in stage I/II vs. controls (n=51) in contrast to stage III/IV vs. controls (n=1) in proliferative phase, and (2) more DE genes were observed in stage III/IV vs. controls (n=33) in contrast to stage I/II vs. controls (n=1) in secretory phase.

As seen in Chapter 4, the differential expression profiles of the endometrium between proliferative phase and secretory phase are striking (n=7,417 differentially expressed genes), which could help hypothesizing how the eutopic endometrium behaves throughout the cycle in different surgical sub-phenotypes of endometriosis. One bold hypothesis is that the potentially perturbed biological pathway(s) in stage I/II endometriosis may mainly affect the functioning of (or be the consequence of abnormal) proliferative endometrium, but not as much of the secretory phase endometrium. Likewise, the perturbed biological pathway(s) manifesting in stage III/IV cases may mainly affect functioning of (or be the consequence of abnormal) secretory phase endometrium. It is this different pattern of the transcriptome between cases with certain surgical staging vs. controls that could mask/dilute the different effects contributed by the other staging group when only looking into overall endometriosis cases vs. controls.

Limitation of comparison to previous reports

There are a few limitations when comparing the results of this study to those reported in previous publications: Fung et al.[145] and Mortlock et al.[146] reported no genome-wide significant DE genes (Table 5.1). Given the limited number of significant DE genes, here I have looked at all the nominally differentially expressed genes from these studies and compared with my results. There were no overlap between the nominal DE genes between the two studies[145,146] (Mortlock et al.[146] number of nominal DE genes within mid-secretory phase only= 43; Fung et al.[145] nominal DE genes = 486). When combining the gene lists from all three studies[144-146] for lookup, a long list (1,373 for proliferative & 2,326 for secretory phase) of mainly nominally significant genes was created; this was also due to the minimal overlap of results from these studies. The direction of fold changes was not consistent between studies when they reported the same gene being DE in the same cycle phase. Although both significant and nominally significant genes were reported in my study, the extent of overlap with previous reports was minimal (Supplementary Table S 5.2-8), demonstrating the difficulties in replicating these signals.

It is also important to note that genes reported in this study and mentioned in previous studies are not directly comparable in terms of the experimental design and the type of effect sizes reported (e.g. $\log_2$ fold change vs. expression ratio). The comparison of previous studies was made for overall endometriosis case-control status, while this study involved four different comparisons. In Fung's[145] study, the list of genes was obtained without testing within individual cycle phase, but only adjusting its effect by including the phase as a term in the model on all case-control samples; if the case-control samples they used were unbalanced in terms of cycle phases (e.g. more secretory samples in case group than in control), the adjustment alone may not have been adequate in addressing drastic differences between phases, and would potentially mask the effect of genes with opposite effects due to expression fluctuation in a normal cycle. The magnitude of their estimates of expressed and unexpressed (on/off) gene in endometrium cannot be directly compared to $\log_2$FCs from the expression experiment.

A similar issue arose when comparing with data from Burney et al.[144]; ratios of expression levels between cases and controls were reported instead of fold changes, making it incomparable to our results. In fact, out of the three case-control comparisons within different cycle phases, they reported the highest number of DE genes (>1k) in a set with the smallest sample size (i.e. 6 cases to 3 controls in early-secretory phase). The huge differences in transcriptomic signature they observed was more likely to be due to the inter-sample variations than actual case-control differences. Generally, comparison of results to a report lacking power of detection should not be done; however, no other studies is available in the literature. Despite these limitations, previous reports of genes with their *p* values were referenced in the subsequent discussion where appropriate.

Results from the proliferative group

The differential expression results between overall endometriosis cases vs. controls in the reduced subset (n=38) were comparable to the main analysis (n=44). The smaller subset served as a 'cleaner' dataset with fewer outliers, but compromised by a smaller sample size, especially with a limited number of clinic controls (n=7) and thus reduced power. Only four DE genes were observed in stage I/II versus III/IV comparison; all of them were also reported as significant in the full 44-sample set with 22 DE genes in total.

The only significant gene in the overall endometriosis case vs. control comparison from the reduced subset in proliferative phase is Endothelin 2 (*EDN2*). It is ranked 24th and nominally significant in case-control analysis of the full set, while ranked 4th and significant in the stage I/II-control analysis. It is under-expressed in overall cases (log_2FC=-1.65) and in stage I/II (log_2FC=-2.20). *EDN2* is one of the three members of the endothelin family; the protein family (mainly *EDN1*) modulate growth of endometrium in the proliferative phase, and endometrial blood flow[288] in response to oestrogen and progesterone[289] in rhesus macaque. It also stimulates prostaglandin production *in vitro*[290]. In a mouse study[291], *Edn2* expression in granulosa cells of ovarian follicles was shown to be essential for ovulation through induction of vasoconstriction. Apart from its role in normal physiology, *EDN2* acts as a chemoattractant for macrophage[292] in pathology, especially in breast tumors[293]. One of the hypotheses of pathogenesis of endometriosis, especially for pain sub-phenotypes[294], is that it is associated with inflammation-mediated macrophage action[19,295]. This under-expression of *EDN2* in endometriosis ASRM stage I/II cases, compared to clinic controls, in human proliferative eutopic endometrium is a novel result.

The top signal from the *ZNF26* gene, shown in three comparisons in the proliferative group (n=44) (Section 5.3.iii.a) is likely to be caused by outliers in the control group, and cannot be regarded as a valid signal in this analysis. Its protein localised to the mitochondria[296]; not much is known for the function of this gene, apart from its involvement in the herpes simplex virus 1 infection KEGG pathway[87]. Future analysis in an independent dataset with more clinic controls would be useful to see if the signal was a false positive from this relatively small dataset.

Among the other top three significant genes in the overall endometriosis case vs. control comparison, serum amyloid A2 (*SAA2*) had with the highest absolute log_2FC (under-expressed in cases by 3-fold). It ranked 5th in stage III/IV-control comparison (nominally significant $p=1.42x10^{-5}$, $log_2FC=-3.30$) and 59th in stage I/II-control ($p=1.18Ex10^{-4}$, $log_2FC=-2.85$). SAA2 is involved in acute inflammatory response as chemoattractant for injury and infection[297]. Its protein family is reported to be a biomarker for endometrial cancer in human[298,299], and endometritis in cows[300] and horses[301]. Their expressions are also reported to be down-regulated in menstrual phase endometrium in women suffering from heavy menstrual bleeding (HMB) compared with controls without HMB[302]. Although the gene is mentioned in Burney's[144] analysis of proliferative phase, it showed an expression ratio of diseased over control larger than one (1.84, $p=0.009$), suggesting an opposite direction of effect. The paralog of this gene, *SAA1* (locating in the same cytoband 11p15.1, and also involving in acute-phase response), is not differentially expressed in the proliferative phase analysis; however, *SAA1* is among the top genes that are over-expressed in stage III/IV when compared to controls (adj.$p=3.40x10^{-3}$, $log_2FC=2.32$) in secretory phase, and when compare to stage I/II cases (adj.$p=1.44x10^{-8}$, $log_2FC=3.42$).

One of the nominal genes from the case-control proliferative comparison, FAT Atypical Cadherin 2 (*FAT2*, ranked 37th by $p=1.87x10^{-4}$, $log_2FC=-1.24$), is located within 1Mb of an eQTL SNP in endometrium at chr5:150,964,798 (FDR $p=4.46x10^{-4}$). *FAT2* is expressed primarily in cerebellum and in female reproductive tissues such as uterine, cervix and vagina[303]. The FAT gene family plays a role in neural development and proliferation[304]; mutation in *FAT2* leads to spinocerebellar ataxia[305] — a neurodegenerative disease. The encoded proteins are with cadherin repeats and epidermal growth factor (EGF)-like domain[306] at the end for cell-to-cell interactions. Although the function of this protein family specifically for the female reproductive system is not understood, it is reported that *FAT1* expression is detected in breast cancer, while *FAT2* is expressed in ovarian cancer[307]. Interestingly, one of its member *FAT3* also show nominal DE (ranked 44th by $p=2.10x10^{-4}$, $log_2FC=1.80$) in the stage III/IV-I/II comparison in proliferative samples, and is within regulatory range of another eQTL SNP in endometrium at chr11:91,868,024 (FDR $p=1.46x10^{-9}$).

Nominal DE genes around GWAS loci associated with endometriosis sub-phenotypes

Differential expression analysis between stage I/II vs. control revealed *C2orf50* (chromosome 2 open reading frame 50), (ranked 96th) as nominally under-expressed in stage I/II. This gene is located 431kb upstream of rs11674184 (intronic to *GREB1*) which shows nominal association with stage III/IV and endometrioma sub-phenotypes (Chapter 3, Table 3.5). *C2orf50* encodes for an uncharacterised protein, of which much is not known for its function. Although the assumption of looking up genes nearby GWAS variants is to try to discover any potential genes with expression that may be affected by the SNP nearby, these SNPs may not affect only genes in closest proximity. In this case, the gene is not differentially expressed in the stage III/IV-control comparison, nor was *GREB1,* in any comparison.

Thrombospondin 1 (*THBS1*) is another gene located near to an endometriosis GWAS locus (rs12441483 at 3'UTR of *BMF*) that exhibits differential expression after multiple-testing correction in the stage III/IV vs. stage I/II comparison. The GWAS variant showed nominal association with superficial lesions, endometrioma, infertility, and multiple pain sub-phenotypes as described in chapter 3. *THBS1* ranked 11th by adj.*p* value in the analysis and is within 1MB of an eQTL SNP at chr15:38,758,027 (p=5.32x10^{-3}, FDR p=0.13). It is over-expressed in stage III/IV vs stage I/II in proliferative phase ($\log_2$FC=1.70). The order of *THBS1* expression level is as follows: Stage I/II cases < Clinic controls ≈ Stage III/IV cases. A tentative hypothesis could be that the perturbation of downstream biological pathways by the eQTL SNP (chr15:37,758,027) and GWAS SNP (chr15:40,353,316) may be partially mediated by, or act via, the differential expression of *THBS1* locating midway between both SNPs (chr15: 39,873,280-39,891,667). THBS1 is an adhesive glycoprotein secreted mainly by fibroblasts and endothelial cells with dual anti- angiogenic[308] and pro-angiogenic[309,310] properties depending on which molecules it is interacting with. It mediates cell-to-cell interactions, best known for platelets, and interacts with fibrinogen[311] and integrins[309]. It is shown to have a role in macrophage recruitment in tumors in mouse model and *in vitro* experiments[312]. For its function in relation to female reproductive tract, it is mentioned in studies mainly focusing on embryo receptivity; *THBS1* is significantly expressed in receptive (mid-secretory) than pre-receptive (early-secretory) endometrium in two human endometrial epithelial cell lines[313].

Overlap of genes showing DE in both 'stage III/IV versus I/II' and 'proliferative versus secretory' comparison

Two genes, *DEPP1* (previously known as *DEPP*/*C10orf10*) and *CITED2*, were amongst 22 significant DE genes between stage I/II and III/IV in the proliferative phase. Yet, they also featured within the top 10 DE genes between proliferative and secretory phase transcriptomic profiles as described in Chapter 4. A detail description of *DEPP1* is given in Section 4.4. The DE of these two genes is likely to be driven by the residual cycle phase effects within the proliferative phase, capturing signals possibly contributed by self-reported 'proliferative' samples dated near to the transition day of the cycle, but not reflecting differences due to disease status.

Signals in the 21q11.2 cytoband

Three pseudogenes on cytoband 21q11.2 were differentially expressed between ASRM stage I/II vs stage III/IV in both proliferative and secretory phases, showing opposite effect sizes. There is also another nominally DE pseudogene *RHOT1P2* featured in the same cytoband (Section 5.3.iii.b). A pseudogene is normally shorter in length when compared to a normal protein-coding gene, and there could be a chance for reads to be misaligned to their relatively short reference sequence, and wrongly annotated as pseudogenes. Nonetheless, the chance of misalignment for four different pseudogenes, which happen to be in the same genomic region and showing consistent direction of result, would be very slim. The opposite directions of effects observed between the two cycle phases suggest that the genomic region may also play a role in the normal physiology of menstrual cycle, but with different degree of regulations in stage I/II and III/IV diseases potentially.

Not much is known about the functions of these pseudogenes. According to the latest Ensemble biotype classification[314], only *SNX18P13* is a processed pseudogene (i.e. without introns and thought to be the product of retro-transposition), and the other two are transcribed unprocessed pseudogenes, meaning they contain 'fossil' introns conserved from gene duplication, but with 'original' expression disabled by a premature stop codon[315] when compared to their functioning counterparts. For expression patterns of these three pseudogenes by human tissue types, esophagus shows the highest expression for all three, followed by testis and urinary bladder[316]. Expression is also detected in the female reproductive track, such as the uterus, ovary and vagina[237], but at lower levels. Eutopic endometrium is not part of any databases for expression data; this study shows the co-expression pattern of these three pseudogenes for the first time (Section 5.3.iii.b).

Apart from co-expression, these pseudogenes are also in close proximity in the genome. Therefore, SNPs situated in these pseudogenes were tested for linkage disequilibrium (LD) to investigate if they were inherited together in the same LD block. However, they were not in LD within the British European population, suggesting the co-expression may not be linked to conservational patterns of the genomic region.

The genomic region within 21q11.2 cytoband harbours several GWAS variants for cytokine level in blood circulation, such as rs190160997 in association with serum level of chitinase-3-like protein 1 (CHI3L1 or YKL-40), an inflammatory[317] and fibrosis responder[318,319] reported in a small Scandinavian ancestry dataset[320]. Likewise, a medium size Finnish study[321] reported rs72490194 in nominal association with tumor necrosis factor alpha (TNFα), an inflammatory mediator, and rs56254584 for a growth regulator, interleukin-2 (IL-2) in blood.

Although there were no endometrium eQTL SNPs reported to confer any regulatory elements for expression in this region, it is worth mentioning that there is a gene *NRIP1* (nuclear receptor interacting protein 1, also called *RIP140)* related to the oestrogen pathway and located near to the region of interest (~1Mb downstream of *ANKRD20A11P*). It encodes a protein that modulates transcriptional activity of the oestrogen receptor[322]. The Comparative Toxicogenomics Database (CTD)[323] showed that its gene expression is also mediated by oestradiol directly or through interaction between oestradiol and oestrogen receptor 2 (ESR2) protein. The expression of *NRIP1* is, however, not significantly different in any comparisons performed in this study.

From the above, the 21q11.2 cytoband may harbour regulatory elements involving inflammation and/or oestrogen pathways, which may predispose endometriosis in different ways across the disease spectrum, and thus, the signal captured by different surgical ASRM sub-phenotypes in this analysis. Indeed, genes and their encoded proteins can work together in a biologically functional module[324], regardless of their genomic proximity. To further the investigation of these signals beyond the cytoband in future, weighted gene co-expression network analysis[325] could be done, on larger sample sizes, to identify co-expression modules in association with the condition[326] and their enriched pathways.

Results from the secretory group

The results from the two-term design model for secretory phase `(design2 ~ endometriosis case-control status + sub-secretory phase)` were slightly different with those from the simple model (without inclusion of the sub-secretory term), but comparable in terms of the trend and the number of genes discovered. It is important to note the limitation of the two-term model; despite the inclusion of the sub-secretory term, not all sub-secretory effects can be fully accounted for in the dataset, especially when the proportion of samples from each sub-secretory phase is different across comparison groups (i.e. stage I/II cases, stage III/IV cases, and clinic controls) (Table 5.2). The residual sub-secretory cycle phase effect and their transcriptomic variations were revealed when some of the significant DE genes tested for the disease status were also captured in the mid and late-secretory phase comparison, regardless of disease status. As both DE gene lists were generated using the same RNA-seq dataset, there is better confidence in interpreting the overlap; correlation of direction of effects between the two can help to pinpoint potential false signals contributed by the residual sub-secretory effect. Ideally, if transcriptomic profiles comparing early to mid-secretory phase were available (the comparison was not done due to small sample size), this would have been helpful to add to the comparison. Ultimately, future studies should explore the comparison of disease vs. controls within each sub-secretory phase where the required sample size is available.

In the case-control comparison using the simple model on secretory samples (n=91), hemoglobin subunit gamma 2 (*HBG2*) was the only significantly DE gene; it is also ranked 6th in the two-terms model (p=5.88x10^{-5}, adj.p=0.42, log2FC=-1.24), and was nominally significant in the stage III/IV-control comparison (ranked 61th by p=3.31x10^{-4}, adj.p=0.15, log2FC=-1.33). It was under-expressed in overall cases and in stage III/IV cases compared to clinic controls. *HBG2* was not differentially expressed between the mid and late-secretory comparison (regardless of endometriosis status).

HBG2 is not only well known for formation of fetal haemoglobin; it is a member of the IL-2 pathway[327] and also mentioned in the literature for its response to 17β-oestradiol and progesterone[328]. In the context of endometriosis, it is reported to be up-regulated in ectopic against eutopic endometrium from women suffering endometriosis[329]. It is also involved in normal menstrual breakdown of the cycle; a study shows that *HBG2* is up-regulated 4-fold in day 4 of menstruation when compared to day 3 endometrium[330]. Another study by Girling et al.[302] showed its up-regulation in menstrual endometrium of women with heavy menstrual (HMB) when compared to controls; it is also nominally up-regulated in endometrium from women with uterine fibroids (with or without comorbid HMB).

The following discussion is based on results from the two-term design model: no gene passed the significant threshold in the case-control comparison, while only one gene was significantly different between stage I/II and controls — 5'-aminolevulinate synthase 2 (*ALAS2*). It was the only gene from the top 50 nominal hits that was mentioned in previous studies (Fung et al.'s[145] on/off genes estimation in endometrium).

Traditionally, there is a tendency to report protein-coding genes with good level of annotation in the literature over non-protein-coding biotypes, such as pseudogenes and lncRNA, many of which have a clear function. In fact, half (53-57%) of the nominally significant DE hits from the above two comparisons were non protein-coding in biotypes, making comparison to previous reports challenging. When more significant hits (adj.p<0.05) were found in a comparison, most of them were protein-coding genes (e.g the stage III/IV vs. controls comparison). If considering all nominally significant signals, again, the proportion of protein-coding genes dropped to almost half, which is consistent with the pattern previously mentioned. Extra caution is needed to interpret nominally significant signals for validity, especially those that are not protein-coding in biotype.

The *ALAS2* gene showed significant DE in the stage I/II vs. controls with the two-term model (adj.p=0.03, $\log_2$FC=-2.14); however, it is not significant and ranked 7005th by p-value in the simpler model (without attempting to adjust for any sub-secretory phase effect). The gene is not differentially expressed between mid vs. late-secretory analysis. *ALAS2* is located on the X chromosome and is also haemoglobin-related, similar to *HBG2*; the encoded protein is a mitochondrial enzyme specific to erythrocytes, responsible for the first step of heme synthesis[331]. A study[302] suggested its nominal differential expression in menstrual phase endometrium between fibroid patients (±HMB) and the control group. My analyses suggest possible under-expression of both haemoglobin-related genes in the secretory endometrium based on endometriosis status, a finding which would need replication given the variation in modelling results. A hypothesis could be that the down-regulation of these genes in secretory phase 'prime' the endometrium to have abnormal menstrual phase functioning, which may contribute to endometriosis.

In the analysis of stage III/IV against clinic controls in the secretory phase, four out of 33 significant DE genes are also featured in the previous mid vs. late-secretory transcriptome comparison. As both the direction and magnitude of fold changes from two separate analyses correlate well, it is highly likely that these signals are driven by the higher proportion of late-secretory and lower proportion of mid-secretory samples in the stage III/IV group than in the control group (Table 5.2). One of the four genes, *MYH11*, was also significant (p=0.014) in Burney's[144] study of endometriosis case-control status in early-secretory samples.

The second most significant DE gene between stage III/IV vs. controls, without the concern of sub-secretory effect, was BPI fold containing family B member 1 (*BPIFB1*, previously termed *LPLUNC1*/*C20orf114*). It was also significant and ranked 8th in the stage III/IV vs. I/II comparison. It was 3-fold over-expressed in stage III/IV in both comparisons. The protein is involved in the innate immunity against bacteria in the mucosa of the upper respiratory and digestive tracks[332]. Although its functional annotation does not have a clear link to reproductive biology, the gene is also mentioned in Girling's study[302] as the top down-regulated gene in menstrual endometrium from women with HMB, and in fibroid patients when compared to controls.

The 4th and 5th topmost DE genes were perilipin 4 (*PLIN4*) and glycerol-3-phosphate dehydrogenase 1 (*GPD1*) in the stage III/IV vs. control analysis. Both genes were also significant in the analysis contrasting stage III/IV vs. I/II cases (ranked 5th and 7th by adj.*p* value). They were over-expressed in stage III/IV secretory phase (log2FC=1.90-2.37) when compared to controls and to stage I/II cases. Both genes contribute to lipid storage and metabolism. *PLIN4* encodes a protein that helps the packaging of lipid droplet within cells[333]. One of its family members, *PLIN1,* is also a significant DE gene in both stage III/IV-control and stage III/IV-I/II comparisons (ranked 8th and 2nd respectively by adj.*p* value); its protein is involved in lipid droplet enlargement, transfer, and lipolysis[334]. GPD1 is a catalytic protein for a carbohydrate and lipid reaction (dihydroxyacetone phosphate (DHAP) + reduced nicotine adenine dinucleotide (NADH) ⇌ glycerol-3-phosphate (G3P) + NAD^+)[335]. Proteins involved in lipid metabolism are likely to indirectly affect steroidogenesis[336], which may play a role in modulating hormones production such as for oestrogen and progesterone.

There is also an overlap between the genes that show differential expression in stage III/IV vs. I/II and in mid vs late-secretory phases. More than half of the DE genes (12 out of 21 significant genes) from the sub-secretory comparison also feature in the comparison between ASRM stages. The reason for this may be an imbalance in mid to late-secretory samples presence in each group: 1:1.8 in stage III/IV and 1:1.2 in stage I/II. The inclusion of the sub-secretory term in the model may not adequately address the differences in sub-secretory makeup in each group; thus, the inflation of DE gene discovery. The correlation of effect sizes of expression of these overlapping genes (Supplementary Figure S 5.8) supports the assumption that the signals are from the residual sub-secretory effect, rather than ASRM status itself.

For those significant DE genes that are without concern for confounding effects by cycle phase, *MUC5AC* (Mucin-5AC) and *MUC5B*, ranked 9th and 11th in the stage III/IV vs. I/II comparison, encode for proteins in a close family. They are significantly differentially expressed (ranked 12th and 15th by adj.*p*) in the comparison between stage III/IV vs. controls, suggesting similar over-expression patterns (log_2FC=2.26-3.34) in stage III/IV cases. Similar to the function of BPIFB1, they are glycoproteins[337] situated on upper respiratory and digestive tracks responsible for binding of bacteria as part of the innate immunity barrier. Both genes map near genetic risk loci for asthma[338]. Their gene expression is not only reported in the female reproductive track (mainly in cervix[339] and endometrium[340]), but also proposed to be associated with normal functioning such as embryo implantation[341], and pathology such as endometrial metaplasia[342] and tumors[343]. *MUC5AC* up-regulation is also reported in endometriosis cases in Fung's[145] on/off gene expression analysis.

One of the significant DE genes, *PTGIS* (Prostaglandin I2 Synthase) in the stage III/IV vs. I/II comparison is located within 1Mb of an eQTL SNP in endometrium. Its expression is 2-fold higher in stage III/IV vs. stage I/II in secretory phase. However, the gene showed significant differential expression between mid vs. late-secretory phases. This may suggest that the signal, and its neighbouring eQTL SNP, impact variation and regulation in the normal menstrual cycle, and not a difference by ASRM sub-phenotype. The gene is mentioned in Burney's[144] analysis as down-regulated in early-secretory samples from endometriosis cases vs. controls. Another study[344] showed most variation lied between mid to late-secretory phases among any two consecutive phases, supporting our findings. *PTGIS* encodes a protein for the conversion of prostaglandin H2 to prostaglandin I2 as part of the eicosanoid synthesis network. Eicosanoids are signaling molecules made by oxidation of twenty-carbon essential fatty acids, known to be involved in infection and inflammatory response[345]. The PTGIS protein sequence is similar to the cytochrome P450 family, and with functions involving vasodilation and platelet aggregation linked with cardiovascular disorders[346]. Moreover, a small study[347] characterising transcriptomic differences between endometrioma lesions and paired eutopic endometrium in the proliferative phase showed a very large over-expression (>70-fold) of *PTGIS* in the lesions, making the team to propose the eicosanoid pathway as one of the possible perturbed pathways for endometriosis.

To summarise, the results of differentially expressed genes between endometriosis disease/ASRM surgical status and clinic controls highlighted one genome region and five biological gene clusters: 1) the 21q11.2 cytoband, which harbour three pseudogenes showing DE between stage I/II vs. III/IV cases, that maybe involved in inflammation and/or oestrogen pathways; 2) Innate immune response (*BPIFB1*, *MUC5AC*, *MUC5B*), including macrophage attraction (*EDN2*, *THBS1*); 3) acute inflammation response (*SAA1*, *SAA2*); 4) neural proliferation (*FAT2*, *FAT3*); 5) haemoglobin-related (*HBG2*, *ALAS2*); and 6) possible steroidogenesis via regulation of lipid metabolism (*PLIN1*, *PLIN4*, *GPD1*). Replication with an independent dataset and a larger sample size (minimum of 500 in total[287], ideally 100 per cycle phase) would be required to replicate and further evaluate the signals and proposed biological gene clusters involved in endometriosis. Cell-type specific RNA-seq on endothelial, epithelial, stromal, and immune cell populations of the endometrium[348] can also be conducted to further understand the origin driving the signal observed.

5.5 Conclusion

This work confirmed the limited discoveries from differential expression analysis for overall endometriosis cases vs. controls (in line with the recent literature); this may have been due to heterogeneity of the disease definition. Stratification by ASRM surgical staging information revealed a number of novel potential signals. Transcriptional signatures in eutopic endometrium of stage I/II endometriosis cases in proliferative phase, and stage III/IV cases in secretory phase were established. Cytoband 21q11.2 as a genomic region of interest for the differences between stage I/II and III/IV disease, as well as five gene clusters with similarity in biological functions, i.e. innate immunity, inflammation, neuro-proliferation, haemoglobin, and lipid metabolism, are proposed for potential expression differences between endometriosis cases and controls. Further replication analysis is needed to validate the signals reported in this study.

Chapter 6 Epigenome-wide association study of endometrium in endometriosis

6.1 Introduction

Endometriosis is a complex disease potentially caused by the interplay of both genetic variations (inheritance) and environmental factors (acquired)[7]. After exploring the common genetic variants associated with sub-phenotypes of endometriosis (Chapter 3), and the transcriptomic changes between sub-phenotypes of endometriosis and controls (Chapter 5), the epigenetic component — DNA methylome of the disease state — are investigated in this chapter. To date, the 42 genetic variants associated with endometriosis explain 5.6% of the disease variation; most of these known genetic variants are located in intergenic regions with limited knowledge of their roles. Tissue-specific DNA methylation is one of the ways for regulation of gene expression; DNA methylation markers near these loci may help elucidate potential regulatory role of the genomic region of interest.

Epigenetic processes are multi-layered. DNA methylation, the addition of methyl group to the cytosine residues of CG dinucleotides by a group of DNA methyltransferases, is an epigenetic mechanism that has probably been most frequently explored[349]. The nature of epigenetic marks are dynamic, and their changes may affect downstream gene transcription and interaction. These marks can be influenced by both genetic (e.g. pattern of CpG island) and environmental factors (e.g. the action of DNA methyltransferases); their pattern can be the cause and/or consequence of normal physiology and pathological conditions.

The eutopic endometrium is thought to be the origin of lesions in endometriosis based on transcriptomic, and DNA methylation evidence[164,350]. Similar to gene expression, the methylation profile of eutopic endometrium varies across the menstrual cycle[84,351]; and it is one of the largest sources of variation[159]. Cycle phase will therefore be adjusted for in the analyses presented in this chapter.

Previous studies focused mainly on detecting differential methylation sites around genes for which expression levels appear to be altered in endometriosis. Around 20 promoters, mainly in *GATA*[164], *HOX*[158,352], and *ESR*[353] gene families, have been suggested to show aberrant methylation patterns. After the introduction of DNA methylation bead chip array, genome-wide methylome studies of endometriosis mainly focused on contrasting signatures of ectopic lesions to paired eutopic endometrium[171,354] (Section 1.12); less were focusing on the differences of eutopic endometrium between endometriosis cases and controls. No epigenome-wide association study (EWAS) so far has been conducted that addressed differential patterns according to endometriosis sub-phenotypes.

The aim of this chapter is to explore genome-wide DNA methylation patterns in eutopic endometrium associated with endometriosis sub-phenotypes. An EWAS for endometriosis and its surgical sub-phenotypes in human eutopic endometrium was conducted using DNA methylation array data from Oxford studies, which contributed to the NIH Epigenetic Consortium study (Chapter 2). Quality control (QC) of methylation array data was performed, followed by genome-wide differential methylation analysis between endometriosis and surgical sub-phenotypes (ASRM Stage I/II and Stage III/IV) vs. clinic controls in eutopic endometrium. Based on the genome-wide differential methylation analysis results, nominally-significant loci were investigated for co-localised endometrium *cis*-expression quantitative trait (eQTL) signals, and proximity (±500kb) to the 16 endometriosis genetic risk loci nominally associated with endometriosis sub-phenotypes (from Chapter 3). Lastly, DNA methylation status (hyper/hypomethylation) of genes near (±200kb) to EWAS loci were visualised.

6.2 Materials and methods

6.2.i Criteria for sample selection and arrays

Eutopic endometrium samples from the ENDOX and FENOX studies in Oxford were selected for the NIH Epigenetics Consortium project (Chapter 2). Women with surgically diagnosed endometriosis (cases) and those without surgical evidence of disease (controls) were chosen, limited to those who did not use hormones in three months prior to surgery. We defined patients (n=9 from ENDOX) who were previously diagnosed with endometriosis, but without macroscopic evidence of disease by laparoscopy as at the time when samples were collected as clinic controls for DNA methylation analysis, but as cases for genetic analysis.

In the entire NIH project, there were two sample analysis components of the study: DNA SNP genotyping and DNA methylation profiling. The source of DNA for both was eutopic endometrium (Chapter 2). Two batches of Oxford samples, one for the discovery phase (n=186), and the other for replication phase (n=38), were sent to the Centre for Reproductive Sciences, UCSF for genotyping using the Affymetrix Axiom™ Precision Medicine Genotyping array, and for DNA methylation profiling through hybridizing bisulfite-converted DNA samples using the Illumina® Methylation EPIC Beadchips array (EPIC). All samples contributed to the discovery phase were from the ENDOX study; both ENDOX (n=30) and FENOX studies (n=8) contributed samples to the replication phase. In this chapter, samples from both phases were analysed together.

The genetic ethnicity of women was determined by the genotyping efforts in both NIH project and ENDOX study. For the DNA methylation analysis, only samples identified genetically as being of European ancestry against 1000 Genome Phase 3[355] (genotyped in NIH project, n=156) and HapMap 3[179] (genotyped in ENDOX study, n=7) reference panels were included, as the number of other defined ancestry groups were small.

6.2.ii DNA extraction and histological cycle phase dating of eutopic endometrium

Two sets of frozen sections were obtained from each fresh frozen eutopic endometrial biopsy (n= 224) for histology (6-7μm), and for subsequent DNA extraction (10μm). Methods for slide production for histological cycle phase staging was the same as described in Section 4.2.ii.a, with five phases determined in a double-blinded fashion. If no conclusion could be made based on histology, self-reported cycle phases were used instead (in the same manner as described in Section 4.2.ii.b). DNA extraction was done by QIAamp® Fast DNA Tissue Kit (Qiagen, ID: 51404) with quality checked by NanoDrop and Qubit fluorometric quantitation.

6.2.iii Quality control of DNA methylation data

6.2.iii.a Data pre-processing, preliminary QC, and visualisation for discovery and replication phases data

The raw DNA methylation dataset was limited to samples of European genetic ancestry. I conducted all QC. The raw data was read using `Minfi`[356] R package. Low-level[357] pre-processing, ratio conversion, and mapping to genome annotation by `preprocessNoob` and `ratioConvert/ mapToGenome` functions, respectively, was conducted to facilitate visualisation.

Another auxiliary package, `shinyMethyl`[358], was used for visualisation of both raw and pre-processed DNA methylation signals through density (beta and M-value), and principal component analysis (PCA) plots. Preliminary QC plots for control probes were also generated. Comparison of signals from samples processed in separate batches (i.e. discovery and replication phases) were made by density plots; and the combined dataset (n=163) was used for subsequent QC steps and data analysis.

6.2.iii.b Sample QC

Detailed sample QC was performed with the `meffil`[359] package.

Samples with one or more of the following were removed iteratively until no further outliers were detected:

1) <90% of methylation probe detection (detection p<0.01);
2) >10% of methylation probe with bead number <3;
3) outside of ±5 standard deviations (SDs) from the mean differences between X and Y chromosome signals;
4) outside of ±3 SDs of the regression line between median of methylated and unmethylated signals;
5) deviations from mean values of control probes.

In addition, samples that clustered separately by the first principal component (PC) in PCA, after probe QC and normalisation, were also removed.

6.2.iii.c Removal of array-specific cross-reactive methylation probes and probe QC

A list of cross-reactive probes (n= 108,861) specific to the EPIC array, found to be under-performing by Pidsley et al.[157], were removed from the analysis. In addition, probes with one or more of the following characteristics were also removed in the QC process:

1) detection (p<0.01) in <90% of samples;
2) bead number <3 in >10% of samples.

6.2.iii.d Normalisation and PCA for cluster identification

The functional quantile normalisation of signals was applied via `meffil.normalize.quantiles` and `meffil.normalize.samples` on 751,776 probes (on autosomes and sex chromosomes), and across 145 samples passing QC, to remove global differences in raw probe intensities. The first 15 PCs were used to adjust for technical batch effects on DNA methylation levels; the number of PCs was determined by the least residual overserved in the model fitting process. PCA plots were generated based on normalized beta values from the top 20k most variable probes for cluster identification. PCs were also regressed against technical (e.g. plate number, well position, and amount of input DNA, etc) and biological variables (e.g. cycle phase of the sample) to identify patterns of association (p<0.01 by t-test). Scree plot and separate sets of PCA plots for control probes were also generated for QC purposes.

6.2.iv Genome-wide differential methylation analysis (EWAS)

EWAS for eutopic endometrial DNA methylation according to endometriosis case-control status (cases n=100, clinic controls n=38) was performed using `meffil.ewas`, as well as comparisons of ASRM stage I/II cases (n=68) vs. clinic controls, and stage III/IV cases (n=32) vs. clinic controls.

No prior adjustment for heterogeneity of cellular composition was possible; potential variations were partly addressed by including cycle phase of the sample as one of the covariates for the analysis. In addition, surrogate variables analysis (SVA) was employed as one of the analysis models to account for heterogeneity of data structure[360,361]. The estimation of surrogate variables was based on the top 50k most variable probes, without trimming of DNA methylation outliers by winsorization. The regression method was chosen to assess quantile-quantile (Q-Q) inflation or deflation. Epigenome-wide significant threshold for EWAS by the EPIC array[362] was set at $9x10^{-8}$, and nominal threshold at $1x10^{-5}$.

Four models were implemented in each comparison, and their results were compared: 1) without any covariates, 2) with user-provided covariates only, 3) with surrogate variables and user-provided covariates (SVA), and 4) with independent surrogate variables[363] and user-provided covariates (ISVA).

For the case-control comparison, four user-provided covariates, i.e. plate number, well position, amount of input DNA, and cycle phase, were included. For the stage I/II and stage III/IV comparisons against clinic controls, only two covariates, i.e. well position and cycle phases, were supplied due to the limitation of model fitting in these much smaller subsets.

The manifest version B5 released in March 2020 by Illumina, and the `IlluminaHumanMethylationEPICanno.ilm10b2.hg19` package were used for annotation, with genome build information matched to GRCh37/hg19.

6.2.v EWAS loci in relation to human eutopic endometrium eQTL map

EWAS results from the three comparisons (i.e. case-control, stage I/II-control, and stage III/IV-control) were compared with eQTL data. Using the *cis*-eQTL map of human eutopic endometrium (details in Section 5.2.iv.c), the overlap of expression SNPs (eSNPs) with DNA methylation probes showing epigenome-wide nominal results ($p<1x10^{-5}$) from SVA and the 'user-provided covariates' models were analysed. The regulatory range of eSNPs was assumed to be within 1Mb.

6.2.vi EWAS loci near GWAS variants associated with endometriosis sub-phenotypes

The 16 genetic variants showing nominally significant associations with at least one sub-phenotype of endometriosis (Chapter 3; Table 3.5) were compared with EWAS results. The genomic location ±500kb around these 16 loci were investigated for the presence of DNA methylation probes with epigenome-wide nominal results $p<5x10^{-4}$ from the SVA model in each comparison (case-control, stage I/II-control, and stage III/IV-control).

6.2.vii Visualisation of DNA methylation status of nearby genes of EWAS loci

Nearby genes within ±200kb of EWAS loci showing epigenome-wide nominal results ($p<1x10^{-5}$) were mapped. Using the pathway interactive visualization tool (PiiL)[364], the average DNA methylation status of those genes, based on multiple probes that mapped around them, in endometriosis cases and clinic controls was visualised.

6.3 Results

6.3.i Sample data structure before QC

Of the total 224 Oxford endometrial samples, 163 were of European ancestry (Section 6.2.i) and included in this analysis. Within these, 117 were endometriosis cases (ASRM stage I n=61, II n=20, III n=17, and IV n=19); 45 were clinic controls; and one was of unknown disease status. Table 6.1 shows the breakdown of samples in five cycle phases by endometriosis case-control status. Cycle phase was determined by histological staging for 82.8% of samples (n=135) and 12.5% were from self-reported data (n=20), and the rest (4.91%, n=8) were without any cycle phase data.

Table 6.1 Number of eutopic endometrium samples with and without endometriosis in each of five cycle phases for Oxford European dataset (n=163)

	Menstrual phase	**Proliferative phase**	**Secretory phase**				**Unknown Phase**	**Total**
			Early	**Mid**	**Late**	**No Sub-Secretory**		
Cases	10	25	12	39	21	3	7	117
Controls	3	11	6	17	7	0	1	45
Unknown Disease Status	0	0	0	0	1	0	0	1
Total	13	36	18	56	29	3	8	163

6.3.ii Quality control of DNA methylation data and determination of covariates

6.3.ii.a Visualisation of raw and pre-processed data of samples in discovery and replication phases

The density plot (Figure 6.1 bottom right panel) with average distribution of beta values of type I green probes in eutopic endometrium samples from the discovery and replication batches showed similar features and were comparable to each other after pre-processing. Thus, DNA methylation profiles of samples from both batches were analysed together.

After pre-processing, density plots were also similar for endometriosis cases vs. clinic controls (Supplementary Figure S 6.1), between different ASRM surgical stages, and five cycle phases (Supplementary Figure S 6.2- 3).

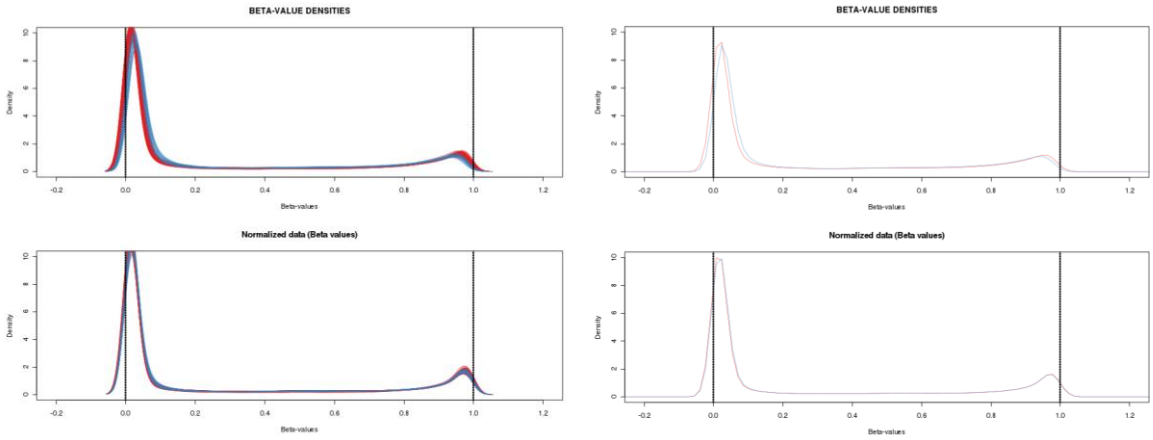


Figure 6.1 Density plots showing (top panel) raw beta value and (bottom panel) pre-processed normalised beta value of type I green probes in eutopic endometrium samples from discovery (red curves) and replication (blue curves) batches.
Left panel shows curves from every samples, while right panel shows the average of all samples.

6.3.ii.b Sample QC

Three rounds of sample QC was performed on the initial dataset (n=163) to remove outliers from various parameters. All samples had correctly predicted sex (all females, Supplementary Figure S 6.4), low proportion of undetected probes, and low proportion of probes with low bead number (Supplementary Figure S 6.5). Three samples were excluded based on the ratio of median methylated to unmethylated signals (Supplementary Figure S 6.6). Two samples were excluded based on dye bias detected in the control probes. Further, 13 samples were excluded based on clustering separate from other samples at PC1 >20 (Figure 6.2). Five of these outlying samples were in menstrual phase (green dots); while one sample was from the early proliferative phase (self-reported), and the two other samples were also in proliferative phase. One sample from each of the three sub-secretory phases, and two samples with unknown phase also clustered at PC1>20.

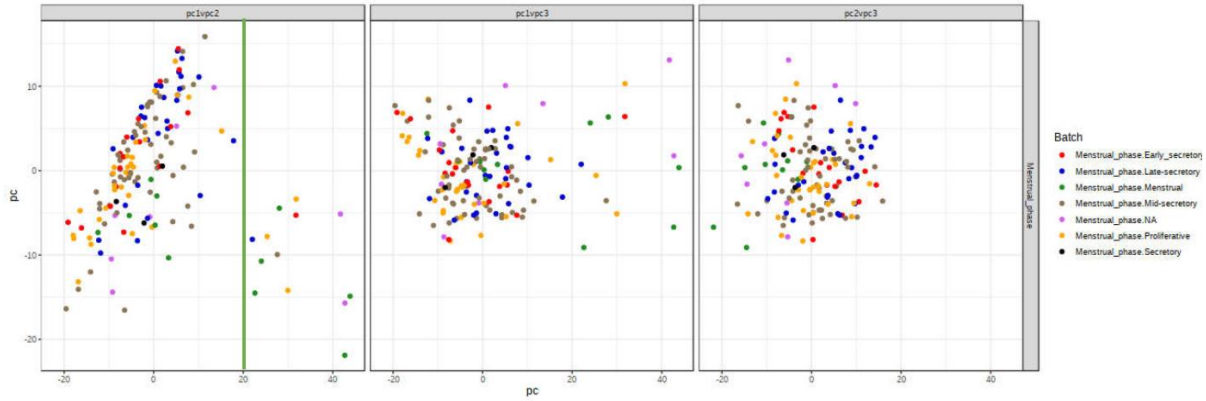


Figure 6.2 PCA plots of samples before 3rd round of QC (n=159) colour-coded by five cycle phases: PC1 vs PC2 (left), PC1 vs PC3 (middle), PC2 vs PC3 (right).
Key: green line = PC1 value >20 for defining outliners; green dots = menstrual phase, orange dots = proliferative phase, red dots = early secretory phase, grey dots = mid-secretory, blue dots = late-secretory, black dots = secretory phase without sub-secretory information, purple dots = unknown phase.

Out of 145 remaining samples, one did not have confirmed endometriosis disease status, and a further six had an unknown cycle phase. These were also excluded, resulting in a total of 100 endometriosis case and 38 clinic control samples available, for the main EWAS analyses.

6.3.ii.c Methylation probe QC and normalisation

After removal of 108,861 probes found to be underperforming in the EPIC array[157], 5,857 additional probes were removed based on poor signal detection and with low bead numbers (236 probes) in high proportion of samples passing QC (Supplementary Figure S 6.7), resulting in 751,835 probes carried over to the main analysis. The scree plot of fitted residuals for determining the number of PCs to adjust for technical batch effects on DNA methylation levels upon normalisation is shown in Supplementary Figure S 6.8. It shows minimal residual variation after the first 15 PCs. The scree plot of control probes also supports this threshold, showing the majority of variance captured by the first 15 PCs. It is a standard practise to remove probes from sex chromosomes to prevent bias of signal due to sex imbalance in datasets, however, this is a female-only study, and with the similarity of normalized data shown in PCA plots with and without probes on sex chromosomes, probes on the X chromosome were retained.

6.3.ii.d PCA plots for DNA methylation data and control probes

The PCA plots after sample and probe QC showed minimal clustering based on cycle phase (Figure 6.3).

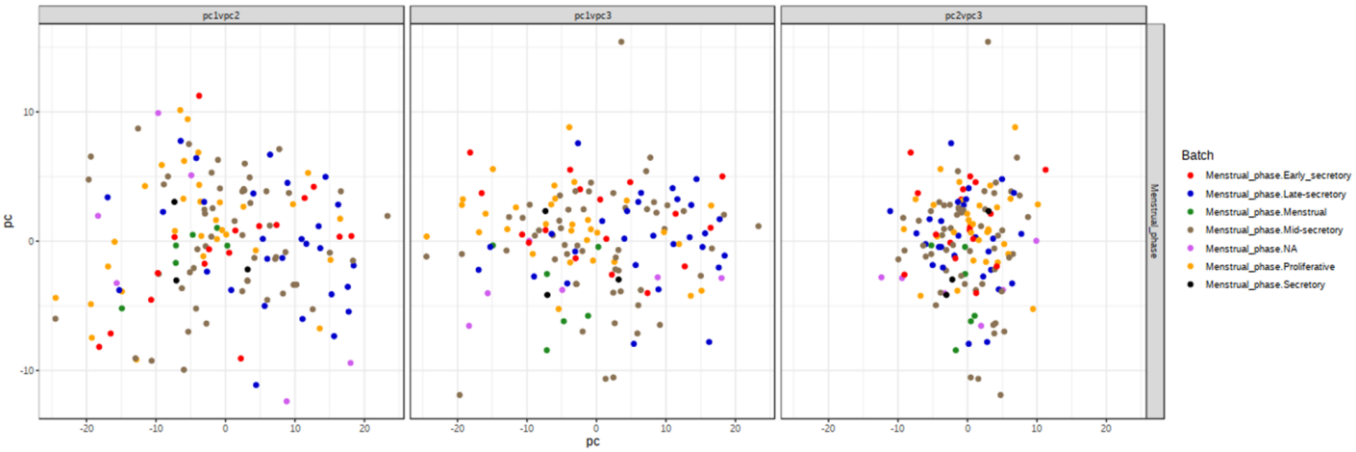


Figure 6.3 PCA plots of samples after QC (n=145) colour-coded by five cycle phases: PC1 vs PC2 (left), PC1 vs PC3 (middle), PC2 vs PC3 (right).
Key: green dots = menstrual phase, orange dots = proliferative phase, red dots = early secretory phase, grey dots = mid-secretory, blue dots = late-secretory, black dots = secretory phase without sub-secretory information, purple dots = unknown phase.

No particular clustering patterns were observed by study batch (i.e. discovery vs. replication phases), endometriosis status, and ASRM surgical stages (Supplementary Figure S 6.9) that raised concern for subsequent analysis. No separate clustering in PCA plots for control probes based on the above parameters was observed.

6.3.ii.e Association of PCs with technical and biological variables

The regression of top 15 PCs against technical/procedural and biological variables showed significant associations mainly with cycle phases, well positions, plate number, and amount of DNA input for profiling. Proliferative and late-secretory endometrium were associated with the first PC (p= 9.67x10^{-3} and 5.50x10^{-4}), where menstrual phase was associated with the thrid PC (p=1.36x10^{-3}). The lowest amount of input DNA (550ng) was associated with the first PC; a summary can be found in Supplementary Table S 6.1. Correlations between each variable associated with the top three PCs were checked; however, there was no specific pattern observed, therefore, they were included as user-provided covariates in subsequent EWAS analyses whenever possible.

6.3.iii EWAS on human eutopic endometrium

6.3.iii.a Endometriosis cases vs. clinic controls

The Q-Q plots from models without consideration of covariates (λ=0.75), and with user-provided covariates only (λ=0.91) were severely deflated (Supplementary Figure S 6.10, top panel); however, the Q-Q plot from the ISVA model showed severe inflation (bottom left, λ=1.17). Only results from the SVA model with a normal Q-Q plot (bottom right, λ=0.88) were considered.

No epigenome-wide significant results (at p<9x10^{-8}) for endometriosis were suggested by the SVA model; two probes at chr11:69,340,400 and chr12:1,068,597 showed nominal associations with p value of 4.47x10^{-6} and 5.13x10^{-6} respectively (Figure 6.4, Table 6.2). Both loci are located at open sea (i.e. not near to a CpG island), one upstream of *CCND1* (Cyclin D1) and one intronic to *RAD52* (DNA repair protein RAD52 homolog).

The distribution of DNA methylation levels of these two loci (Supplementary Figure S 6.11) showed hypomethylation (i.e. lower ratio of methylated to unmethylated signals) at chr11:69,340,400 (left), and hypermethylation at chr12:1,068,597 (right) in eutopic endometrium of endometriosis cases than in clinic controls. The frequencies of five cycle phases of samples included in this analysis can be found in Supplementary Table S 6.2.

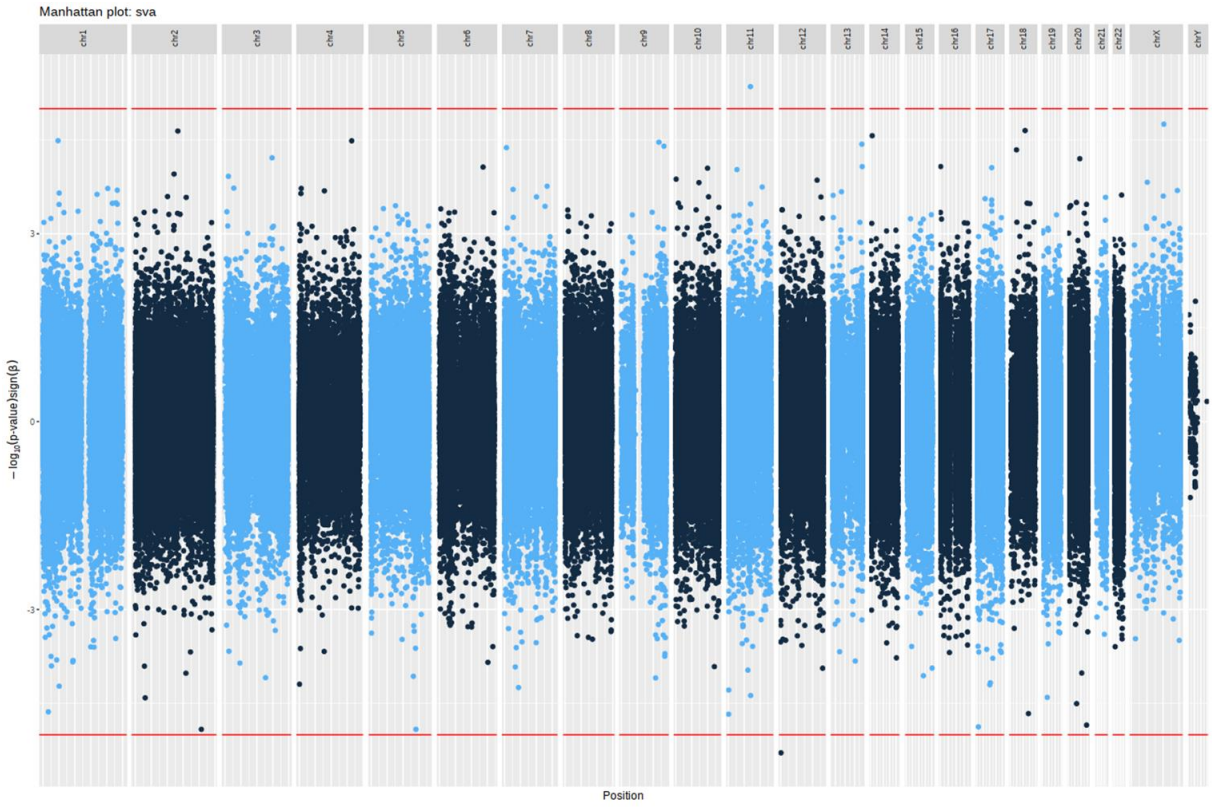


Figure 6.4 Manhattan plot of eutopic endometrium EWAS for endometriosis cases (n=100) vs. clinic controls (n=38) by the SVA model.

Key: dark and light blue dots = DNA methylation probes locate across chromosome number marked at the top; red lines = nominal epigenome-wide threshold at $1x10^{-5}$; blue dots passing red line threshold in the negative panel = negative effect size in test group versus control, and vice versa.

Table 6.2 Table of 14 epigenome-wide nominal EWAS loci for endometriosis, ASRM surgical stage I/II, and stage III/IV cases

Comparison (vs. Clinic Controls)	Chr	Position (hg19)	Adjusted for Surrogate & Supplied Covariates (SVA)		Adjusted for User-provided Covariates		No Adjustment		Nearest Gene	Probe Location Relative to Nearest Genes	Other Nearby Genes	Probe Location Relative to Nearby Genes	Cytoband	Probe Location Relative to CpG Island	CpG Island Position	Illumina EPIC Probe ID	Probe Type	Potential Regulatory Component
			Effect size	P value	Effect size	P value	Effect size	P value										
All Endometriosis Cases	11	69340400	0.051	$4.47x10^{-6}$	0.040	$1.27x10^{-3}$	0.016	$6.95x10^{-2}$	*CCND1*	115kb (5')	*ORAOV1; FGF19*	140kb (3'); 173kb (3')	11q13.3	Open Sea	—	cg02374789	II	—
	12	1068597	-0.129	$5.13x10^{-6}$	-0.101	$1.26x10^{-3}$	-0.059	$3.57x10^{-3}$	*RAD52*	Intronic	*ERC1; WNK1*	32kb (5'); 48kb (5')	12p13.33	Open Sea	—	cg17573903	II	—
ASRM Stage I/II Cases	6	5133188	-0.100	$3.34x10^{-7}$	-0.091	$3.35x10^{-3}$	-0.031	0.24	*LYRM4*	Intronic	*PPP1R3G; FARS2*	46kb (3'); 128kb (5')	6p25.1	5' Shore	chr6:5135172 -5135795	cg01712812	II	Enhancer^
	4	**187070303**	**-0.068**	**$2.09x10^{-6}$**	**-0.047**	**$2.76x10^{-2}$**	**-0.026**	**0.15**	***FAM149A***	**Intronic**	***CYP4V2; TLR3***	**42kb (5'); 64kb (3')**	**4q35.1**	**3' Shelf**	**chr4:187066155 -187066460**	**cg10408467**	**I - Red**	**—**
	12	6813142	0.123	$2.63x10^{-6}$	0.119	$7.52x10^{-6}$	0.024	0.22	*PIANP*	3kb (5')	*ZNF384; COPS7A*	15kb (5'); 20kb (5')	12p13.31	3' Shelf	chr12:6809214 -6809816	cg17286157	II	—
	16	29831109	0.080	$5.80x10^{-6}$	0.083	$8.69x10^{-6}$	0.023	$6.55x10^{-2}$	*PAGR1; BOLA2*	Exonic; Intronic	*MVP*	606bp (5')	16p11.2	3' Shelf	chr16:29826662 -29828302	cg12007977	II	—
	16	85734731	0.033	$6.04x10^{-6}$	0.026	$7.58x10^{-3}$	0.008	0.30	*C16orf74*	Intronic	*GINS2; GSE1*	12kb (5'); 25kb (3')	16q24.1	Open Sea	—	cg27598122	II	Enhancer#
	4	89069031	0.073	$6.74x10^{-6}$	0.069	$6.02x10^{-6}$	0.022	$4.98x10^{-2}$	*ABCG2*	Intronic	*PKD2; SPP1*	70kb (3'); 164kb (3')	4q22.1	Open Sea	—	cg02016771	II	Enhancer^
	4	186990202	-0.039	$6.89x10^{-6}$	-0.024	$1.79x10^{-2}$	-0.005	0.45	*TLR3*	107bp (5')	*FAM149A*	36kb (5')	4q35.1	Open Sea	—	cg17671280	II	Associated to Promoter*, Enhancer^
	11	94883474	0.090	$7.74x10^{-6}$	0.075	$9.36x10^{-4}$	0.017	0.31	*SESN3*	15kb (3')	*ENDOD1*	18kb (3')	11q21	Island	chr11:94883335 -94883565	cg15594471	II	Associated to Promoter*, Enhancer^ [Reprogramming -specific DMR]
	19	12937649	0.090	$7.94x10^{-6}$	0.063	$3.47x10^{-3}$	0.016	0.35	*RTBDN*	Exonic	*MAST1; RNASEH2A*	12kb (5'); 13kb (3')	19p13.2	5' Shore	chr19:12939415 -12939662	cg03204305	II	—
ASRM Stage III/IV Cases	2	119870911	-0.036	$8.95x10^{-6}$	-0.046	$2.65x10^{-3}$	-0.036	$1.25x10^{-2}$	*C1QL2*	43kb (3')	*STEAP3; MARCO*	110kb (5'); 119kb (3')	2q14.2	Open Sea	—	cg12149189	II	Enhancer^
	10	25465430	0.053	$9.41x10^{-4}$	0.091	$6.72x10^{-6}$	0.088	$6.89x10^{-6}$	*GPR158*	Intronic	*ENKUR; TMNSL1*	114kb (5'); 150kb (3')	10p12.1	Island	chr10:25463756 -25465639	cg26249873	I - Grn	[DMR]
	1	154475139	0.060	$1.26x10^{-3}$	0.121	$7.93x10^{-6}$	0.104	$3.79x10^{-5}$	*TDRD10*	Exonic	*SHE; UBE2Q1*	613bp (5'); 46kb (3')	1q21.3	Island	chr1:154474107 -154475699	cg02457680	I - Grn	Enhancer^ [DMR]

Key: EWAS loci ordered by smallest *p* value in SVA model. **Bold** = within regulatory range of an endometrium eQTL SNP (1Mb).

Chr = chromosome; 5'= upstream, 3'= downstream; Shore = 0-2kb from CpG island; Shelf = 2-4kb from CpG island; I – Grn = type I green channel probes.

Potential regulatory component information (right-most column) are suggested by `IlluminaHumanMethylationEPICanno.ilm10b2.hg19` annotation package: ^predicted by Illumina; #by Phantom5; *by DNA Methylation Consortium; [DMR] = Differentially Methylated Region determined by open-source experimental data.

6.3.iii.b Endometriosis ASRM stage I/II cases vs. clinic controls

Similar to the case-control EWAS, Q-Q plots of analysis of stage I/II cases showed deflation/inflation problems in three models except for the SVA (λ=1.12; Supplementary Figure S 6.12). From the SVA model, no epigenome-wide significant associations were detected for stage I/II cases, but nine nominally significant signals were (Figure 6.5, Table 6.2). Three out of nine signals also passed nominal threshold (p<1x10^{-5}) in the model only adjusting for user-provided covariates (λ=1.01), the only three nominal signals detected in that model.

These three probes, in order of smallest p value from SVA model, were located at chr12:6,813,142 (3kb upstream of *PIANP*), chr16:29,831,109 (in exon of *PAGR1* and intronic to *BOLA2*), and chr4:89,069,031 (intronic to *ABCG2*). The first two located to a shelf downstream (i.e. 2-4 kb at 3') from the nearest CpG island, whereas the last one was at open sea, in a predicted enhancer region (Illumina database).

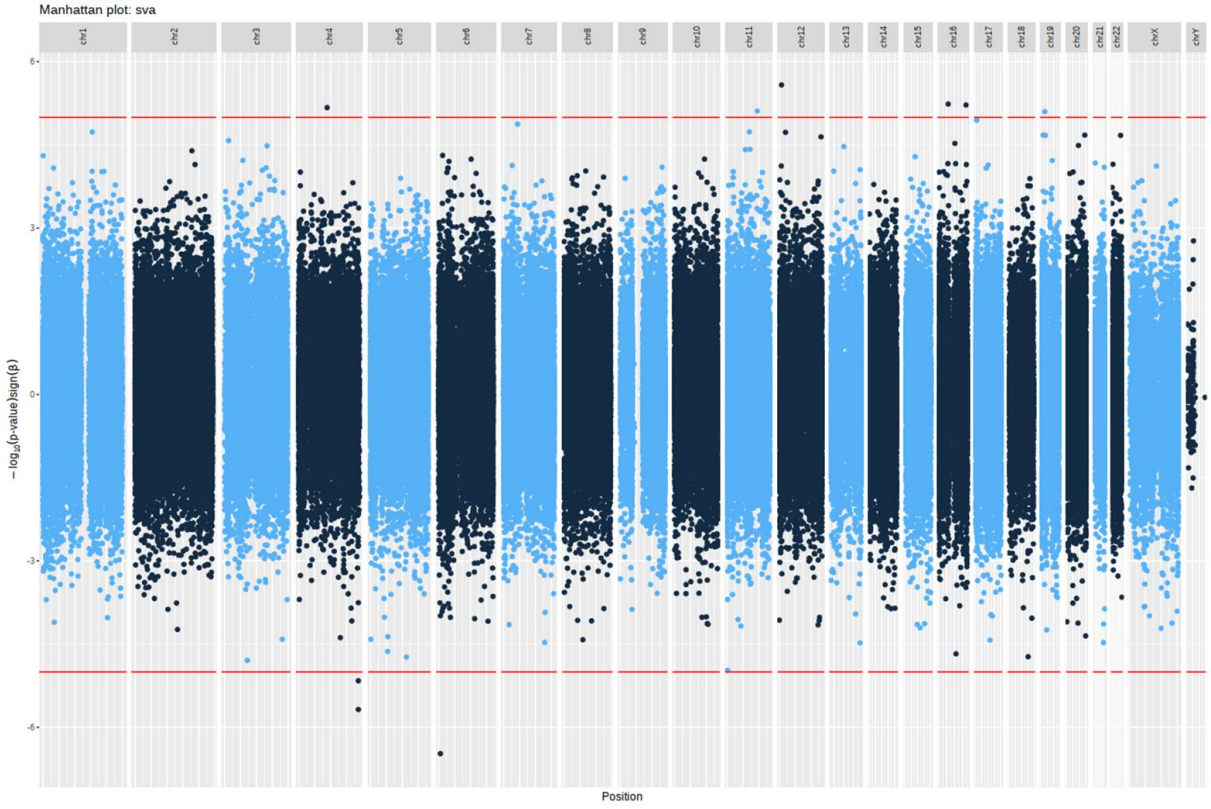


Figure 6.5 Manhattan plot of eutopic endometrium EWAS for endometriosis ASRM stage I/II cases (n=68) vs. clinic controls (n=38) by the SVA model.
Key: dark and light blue dots = methylation probes locate across chromosome number marked at the top; red lines = nominal epigenome-wide threshold at 1x10^{-5}.

One of the probes showing nominal results, chr4:187,070,303 (intronic to *FAM149A*), is located within the 1Mb regulatory range of an endometrium eQTL SNP at chr4:187,120,753 (intronic to a nearby gene *CYP4V2*, FDR p=2.28x10^{-8}).

6.3.iii.c Endometriosis ASRM stage III/IV cases vs. clinic controls

The SVA model (λ=0.87) showed one locus with nominal association to stage III/IV cases at chr2:119,870,911 (43kb downstream of *C1QL2*, p=8.95x10^{-6}, Figure 6.6); it was located at open sea and in the region of a predicted enhancer. Two additional signals were found by the model only adjusting for cycle phase (λ=0.79; Figure 6.7, breakdown of cycle phases in Supplementary Table S 6.2) in the stage III/IV-control comparison (Table 6.2). They were located at chr10:25,465,430 (intronic to *GPR158*, p= 6.72x10^{-6}), and chr1:154,475,139 (in exon of *TDRD10*, p= 7.93x10^{-6}); both probes were on CpG islands. The locus on chromosome 1 was also predicted to be within an enhancer region. Q-Q plots of analyses by the above two models can be found in (Supplementary Figure S 6.13, right panel).

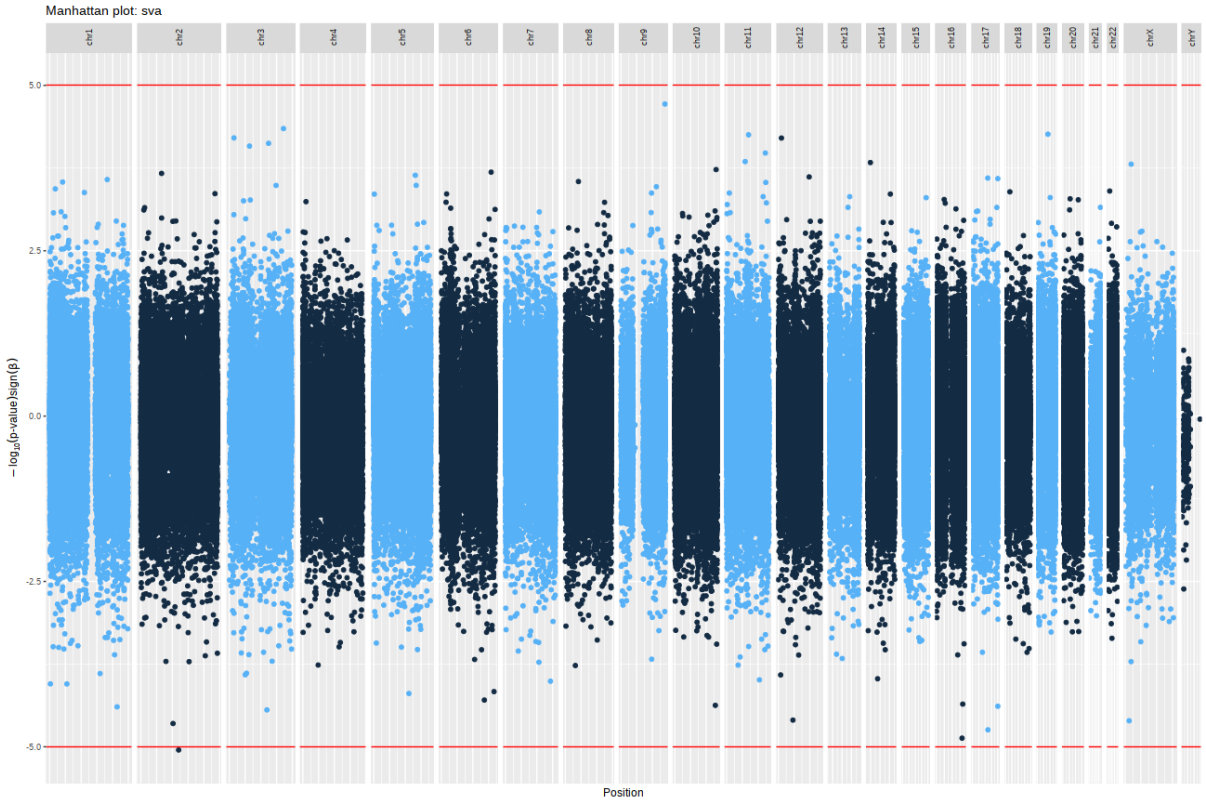


Figure 6.6 Manhattan plot of eutopic endometrium EWAS for endometriosis ASRM stage III/IV cases (n=32) vs. clinic controls (n=38) by the SVA model.
Key: dark and light blue dots = methylation probes locate across chromosome number marked at the top; red lines = nominal epigenome-wide threshold at 1x10^{-5}.

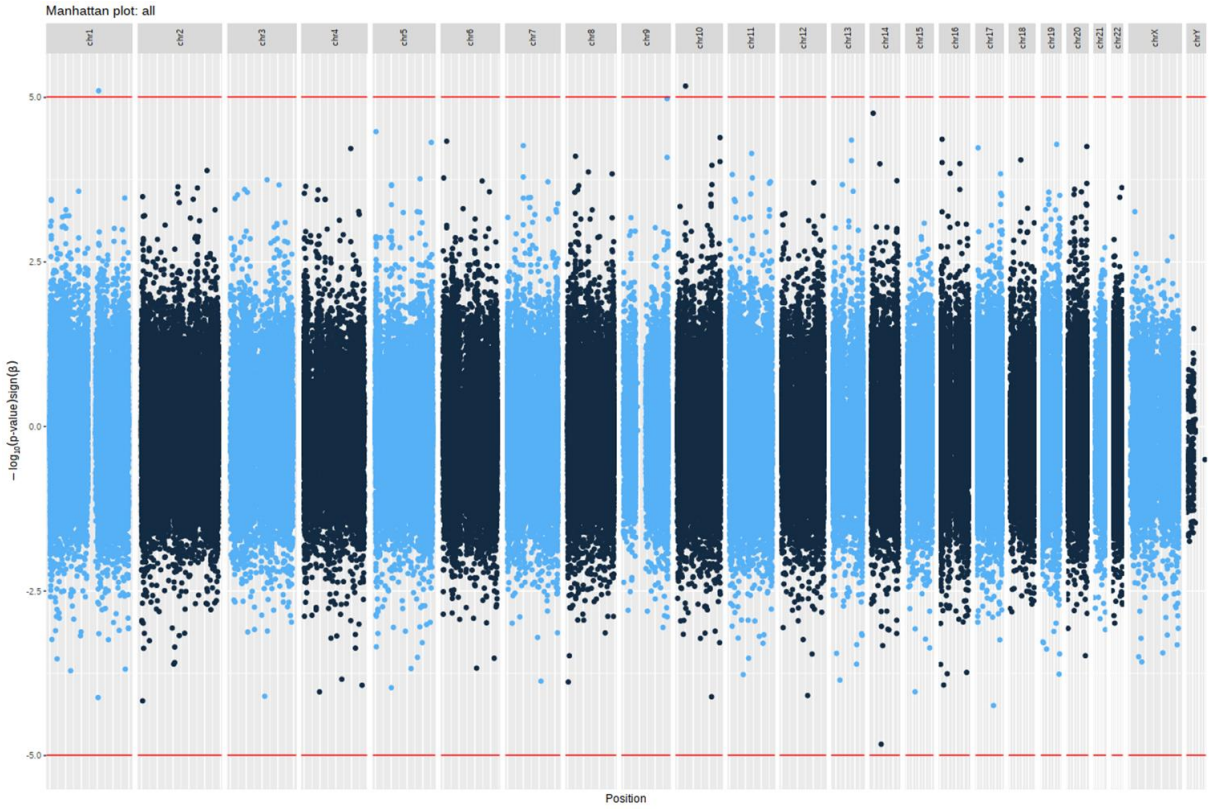


Figure 6.7 Manhattan plot of eutopic endometrium EWAS for endometriosis ASRM stage III/IV cases (n=32) vs. clinic controls (n=38) by the model adjusting only for user-provided covariate (i.e. cycle phase).

Key: dark and light blue dots = methylation probes locate across chromosome number marked at the top; red lines = nominal epigenome-wide threshold at $1x10^{-5}$.

From the above three comparisons of endometriosis cases, stage I/II, stage III/IV to clinic controls, a total of 14 nominal EWAS loci were identified with nominal evidence, which mapped to 15 nearest genes and 23 additional nearby genes within ±200kb. Some of these genes were in the same gene set/family according to the Gene Set Enrichment Analysis[365] (GSEA) online resource (Table 6.3), highlighting their roles in cellular signalling and differentiation regulation.

Table 6.3 Gene sets of nearby genes of 14 nominal EWAS loci for endometriosis and its ASRM surgical sub-phenotypes

Comparison (vs. Control)	Methylation Probe (Chr:Pos in hg19)	Distance between Probe and Nearby Genes	Nearest/Nearby Genes of EWAS Loci	Gene Set/Family
All cases	11:69340400	173kb (3')	*FGF19*	Cytokines and growth factors
ASRM stage I/II cases	4:89069031	164kb (3')	*SPP1*	
		Intronic	*ABCG2*	Cell differentiation markers
	4:186990202	107bp (5')	*TLR3*	
	19:12937649	12kb (5')	*MAST1*	Protein kinases
All cases	12:1068597	48kb (5')	*WNK1*	
		32kb (5')	*ERC1*	Oncogenes & Translocated cancer genes
	11:69340400	115kb (5')	*CCND1*	
ASRM stage I/II cases	12:6813142	15kb (5')	*ZNF384*	

Out of all reported signals, three loci (21.4%) mapped to CpG islands. Other loci were mostly located at open seas (six, 42.9%); two (14.3%) were at CpG island shores (i.e. 0-2kb from island), and three (21.4%) at the shelfs (i.e. 2-4kb from island). The background distribution of probes in EPIC array located in CpG islands is 19%, and for shores is 18%[157]. For annotation of probes relative to location of genes (of reported signals), 64.3% (nine) were in the gene body (intragenic), 21.4% (three) are intergenic, and 14.3% (two) were associated to promoter region. The background distribution of probes in gene body, intergenic, and promoter were 30%, 16%, and 54% respectively[157].

6.3.iii.d Nominally-significant EWAS loci near GWAS variants associated with endometriosis sub-phenotypes

Two of the nominally ($p<5x10^{-4}$) EWAS loci (chr6: 151,667,733 and chr9:136,293,231) between overall endometriosis cases vs. clinic controls in the SVA model were located within ±500kb of two genetic variants associated with at least one sub-phenotype of endometriosis nominally. The first probe was located between two pseudogenes *RN7SKP268* and *snoU13* and near the GWAS SNP rs2967694 (chr6: 151,826,249, intronic to *CCDC170*) with EWAS association $p=1.43x10^{-4}$. The second probe was intronic to *ADAMTS13* and near the GWAS SNP rs507666 (chr9: 136,149,399, intronic to *ABO*) with EWAS association $p=1.79x10^{-4}$.

There was no nominal EWAS signal, near the 16 GWAS variants, within the stage III/IV vs. control analysis. However, there were two nominal loci (chr1:22,486,183 and chr17:64,347,388) in the stage I/II vs. control analysis by the SVA model. The first probe locating between *WNT4* and *MIR4418* is near to rs10917151 (chr1: 22,422,721) with EWAS association $p= 4.43x10^{-4}$; where the second probe is intronic to *PRKCA* and is near the GWAS SNP rs7214750 (chr17: 63,903,119, intronic to *CEP112*) with EWAS association $p=2.73x10^{-4}$.

6.3.iii.e Visualisation of DNA methylation status of nearby genes of nominally-significant EWAS loci

There were a total of 38 genes mapping ±200kb around the 14 nominal EWAS loci. For each of these genes, the average DNA methylation level of a nearby methylation probe (ranging from 3 to 125 Illumina cg probes mapped to a particular gene) was colour-coded according to the endometriosis case-control status. There were no differences observed in the average DNA methylation level of probes near to 38 genes between cases and clinic controls (Figure 6.8); however, the visualization gave a sense of the DNA methylation status of genes (near EWAS nominal hits) in eutopic endometrium regardless of endometriosis status.

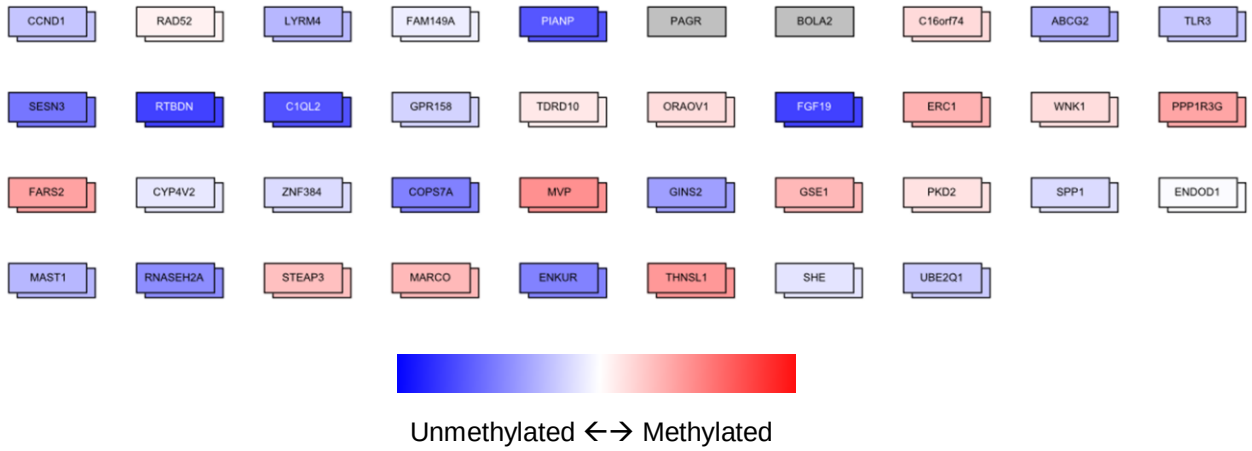


Figure 6.8 Average DNA methylation level of CpG probes around 38 genes mapped to 14 nominal EWAS loci

(Permission for re-distribution (both print and digital) of copyrighted figure is granted on 7 October 2020 by developer of PiiL[364], permission letter in supplementary materials F.iv)

Key: Upper box = DNA methylation status of gene in endometriosis case group; lower box = DNA methylation status in clinic control group. Grey = undetermined average DNA methylation level from limited number of CpG probes available around the gene of interest for calculation; shades of red = methylated; shades of blue = unmethylated.

6.4 Discussion

Oxford European samples (n=163, 136 from discovery and 27 from replication batch) were used for EWAS analyses. To maximise the statistical power, samples from both batches were quality checked to confirm there is no critical structural differences in the data prior to merging (Figure 6.1). Among the technical/procedural variables found to be associated with the first three PCs of the DNA methylation data, the position of well on the array showed the highest 'residual variation' effect. This may be due to residual variation that the first 15 PCs could not account for in the normalisation step (which is designed for dealing with potential technical counfounding effects, Supplementary Figure S 6.8). Among the biological variables, cycle phase showed association with the largest PCs, as expected, as it is a major source variation in the endometrium methylome[159]. Therefore, these variables were included as covariates in the linear regression model for the genome-wide differential methylation analysis.

It has become standard practice to try and adjust for differences in cellular composition between samples in DNA methylation analyses; however, cellular reference sets were not available for our endometrial samples. Given that most variation in the proportion of cell types between samples (e.g. epithelial or stromal cells) could be coming from differences in cycle phase, adjusting for cycle phase effects in the analysis goes some way to countering confounding by cellular heterogeneity. Other variations not captured by cycle phase, nor by any other physical or quantifiable parameter, can be addressed through the testing of so-called surrogate variables in the SVA model. The SVA model was developed originally to address similar heterogeneity problems in expression microarray data[361], and has now been adopted to be used in DNA methylation beadchip array data as well. The ISVA model was specifically developed for large-scale DNA methylation array data[363] like the EPIC array. However, too many independent surrogate variables (>10) were passing the significant threshold of the algorithm, and were included in the model. Confounding effects of these covariates, and over-fitting of the linear regression model resulted in globally inflated test statistics: the entire Q-Q curve sits on the upper triangle of the plot as shown in Supplementary Figure S 6.10. Results suggested by the ISVA model were thus not reported.

In contrast, the model employing only user-provided covariates, and the model without using any covariates showed deflated results for this dataset. It is worth noting that despite the power calculation (specific for the EPIC array[362]) suggesting a 80% power of detecting 2% effect sizes in 81% of probes with a thousand samples, the same deflated pattern is observed in the joint dataset from the consortium (with cases n>1,000) when addressing for cofounders; and no EWAS significant loci are found in the preliminary analysis (results not shown here).

In this Oxford subset of the consortium data, the EWAS loci associated with endometriosis were all nominally significant signals that did not pass the epigenome-wide significant threshold ($9x10^{-8}$), nor did they show clear signal towers in the Manhattan plots; extra caution is needed when interpreting the results. The overlap of these loci with endometrium eQTLs and GWAS variants nominally associated with endometriosis sub-phenotypes are minimal. Two EWAS loci nominally associated with endometriosis ($p<1x10^{-5}$) did not feature in the EWAS analyses for ASRM stage I/II, nor for stage III/IV disease. However, when stratifying by ASRM stage, more loci become nominally associated: nine with stage I/II disease and three with stage III/IV disease; with the difference in sample size for cases (68 vs 32, respectively) likely influencing power of detection. This may suggest a heterogeneity in EWAS results according to ASRM stage, an observation mirrored by previous GWAS analyses[49]. Although there is no overlap between nominal loci discovered in three analyses (EWAS for endometriosis, stage I/II only, and stage III/IV only), the biological functions of genes nearby (in *cis*) are similar for stage I/II and III/IV.

The proportion of reported loci relative to the genome DNA methylation structure (e.g. CpG island, shore, shelf, open sea) is similar to that of the array background distribution (Section 6.3.iii.c). However, the proportion of loci relative to gene structure are biased towards gene bodies (64.3% > 30% background distribution), and away from promoter sequences (14.3% < 54% background distribution). The traditional understanding of role of DNA methylation in the genome mostly focused on its inhibitory role on (and in some case, promoting) gene expression through its action with transcription start site (TSS), promoter, and enhancer regions[366]. However, DNA methylation in the gene body is also demonstrated to be functional in altering gene expression in human cells[367,368], and often exhibit a positive[27,369,370] than an inverse[366] correlation between DNA methylation and expression level. The discussion below focuses on details of nominal EWAS loci associated with overall endometriosis and the ASRM stage sub-phenotypes vs. clinic controls.

Differentially methylated loci for overall endometriosis cases

One of the nominally significant EWAS loci (chr11:69,340,400) is located 115kb upstream of *CCND1*, a gene that belongs to the cyclin family and responsible for the regulation of cell cycle phases[371]. It is also one of the direct targets of the WNT signalling pathway[372], which has been implicated in the pathogenesis of endometriosis[373]. Aberrant methylation patterns[374] and mutation[375] of *CCND1* was reported previously in endometrial cancer; an altered expression pattern, via cyclic AMP signalling, has been described also in endometrial stromal fibroblasts from endometriosis patients[376].

The second locus (chr12:1,068,597) is intronic to *RAD52*, a key gene in DNA break repairment[377], and thus a candidate gene for cancer therapeutics[378]. Its expression was reported to be down-regulated four-fold in late-secretory eutopic endometrium from endometriosis patients compared with controls[379]; however, no differential expression (DE) between endometriosis vs. control secretory endometrium was observed (Chapter 5 analysis).

Differentially methylated loci for ASRM stage I/II cases

The top EWAS locus (chr6: 5,133,188) in ASRM stage I/II cases vs. controls is intronic to *LYRM4* (LYR Motif Containing 4) and located in a predicted enhancer region. The encoded protein is responsible for iron-sulfur biosynthesis and metabolism[380] in nucleus and mitochondria[381]. It was reported to be down-regulated in both colorectal and endometrial cancer[382], but there is little knowledge about its expression specifically in endometrium in general, nor in endometriosis.

The second most significant locus (chr4:187,070,303), intronic to *FAM149A*, is within the regulatory range (1Mb) of an eQTL SNP intronic to the nearby gene *CYP4V2*. There is another nominally significant hit (chr4:186,990,202) in the same cytoband 4q35.1, mapped to the 5' promoter and enhancer region of *TLR3*, and 40kb upstream of *FAM149A*. The function of *FAM149A* is largely unknown. It is included in a study of a candidate 100-gene panel and found to be differentially regulated throughout sub-secretory phases of normal endometrium[260]. Its expression is also significantly (adj.p=3.41x10^{-8}, $\log_2$ fold change=1.19) different in my analysis comparing secretory with proliferative endometrium (Chapter 4), regardless of endometriosis status.

In contrast, the function of *TLR3* is well documented; it is responsible for the innate immune response[383] by sensing the presence of viral RNA[384] with their specific sequence. The protein initiates pro-inflammatory responses. In the context of endometrium, there is a debate if its expression pattern in epithelial cells is menstrual-phase dependent[385] or not[386]. It is also detected in the stromal cells of endometrium[387]. The cytokine induction property[388] of TLR3 is partly regulated by 17β-oestradiol in endometrial epithelial cell lines[389]. Members of the toll-like receptor system are widely studied in endometriosis[390] and in other endometrial disorders. Jiang et al.[391] found that the expression of *TLR3* is significantly higher in eutopic than ectopic endometrium from adenomyosis patients. Whereas Allhorn et al.[386] reported its up-regulation in ectopic endometriosis lesions compared with eutopic endometrium. Its expression is also found to be down-regulated in proliferative eutopic endometrium from endometriosis patients than those from controls[386].

The cytoband 4q35.1 is an interesting region to explore with two signals adjacent to each other. The protein FAM149A is predicted by the integrated interactions databases (IID)[258] to interact with C2CD4A, and its gene is the top DE gene between secretory and proliferative endometrium (Supplementary Table S 4.3). In addition, FAM149A is predicted to be interacting with DEPP1 (also called C10orf10) by the STRING protein network[392]; *DEPP1* is also one of the top five DE genes in the proliferative-secretory comparison (Chapter 4). A previous study proposed that DNA methylation plays a role in hormonal milieu of the cyclic endometrium[351]; these two EWAS loci and the eSNP may therefore play an important role in normal regulation of the cycle cycle.

The following discussion will focus on three loci showing nominally significant association for stage I/II cases in both SVA and only user-provided covariate models, as the support of two models increases confidence in these signals. Of the three loci, chr4:89,069,031 intronic to *ABCG2* is located in a predicted enhancer region, 164kb downstream of *SPP1*. According to the gene set information (Table 6.3) *ABCG2* is a cell differentiation marker, similar to *TLR3*. *SPP1* is one of the top six DE genes which shows a four-fold over-expression in secretory vs. proliferative endometrium (regardless of disease status, Chapter 4). *SPP1* encodes its protein to form the bone mineral matrix; it is an important cytokine of type I immunity and involved in toll-like receptor signalling[87] and vitamin D[393,394] pathways. SPP1 is involved in oestradiol production, promotes aromatase expression[395], and has been found to play a role in decidualisation of the endometrium[396]. No difference in expression between eutopic endometrium from endometriosis vs. controls[397] has been shown, similar to my analysis in Chapter 5.

ABCG2 encodes a transmembrane transporter for a wide range of substances, including urate[398], heme[399], drugs, and steroids such as 17β oestradiol[400]. The protein is a target for dealing with the problem of drug resistance[401] in the treatment of tumours. Its aberrant expression has been reported in multiple myeloma and endometrial adenocarcinoma[402]; and its expression appears altered by level of 17β oestradiol in a endometrial adenocarcinoma cell line experiment[403]. *ABCG2* is one of the markers for human endometrial side population cells[16], both presence in functionalis and basalis layers of endometrium, which may infer stemness (i.e. stem-cell-like properties) in endometrium regeneration during the menstrual cycle[404,405]. The gene is also mentioned in two studies focusing on endometriosis: Matsuzaki et al.[406] showed that the density of *ABCG2*-expressing endothelial cells is different between 1) paired ectopic lesions and eutopic endometrium, 2) ovarian and deep infiltrating lesions, and 3) menstrual phase eutopic endometrium from endometriosis patients and controls. Another study[407] suggested aberrant expression of *ABCG2* in ectopic vs. paired eutopic endometrium, or control endometrium. The findings from my analysis could suggest that the DNA methylation status of the enhancer region near the *ABCG2* in eutopic endometrium for stage I/II endometriosis cases may contribute to differential expression patterns reported in previous literature for endometriosis.

The second locus (chr12:6,813,142) was located at the 5'end of *PIANP*, a gene that encodes a ligand for an immune inhibitory receptor (PILRα) in neural tissues[408], and believed to be involved in the interplay of immune and neural regulations. There is limited knowledge of its role in human endometrium or in endometriosis.

The third locus (chr16:29,831,109) is within an exon of *PAGR1* which encodes for part of a histone methyltransferase complex. It may play a role in epigenetic transcriptional activation[409]. One study[410] has suggested its function as a cofactor for transcription of nuclear hormone receptors. It is involved in oestrogen receptor-induced gene transcriptions (e.g. ERα)[411], and cell cycle transition (G1/S phases) stimulated by oestrogen in particular. There is no study to date that has investigated *PAGR1* in the context of endometrium or endometriosis. Its expression was not tested for in differential expression analysis for endometriosis in Chapter 5 neither as no transcript was mapped to the gene.

Differentially methylated Loci for ASRM stage III/IV cases

The only nominally significant signal suggested by the SVA model to be associated with stage III/IV endometriosis is chr2: 119,870,911 located at a proposed enhancer region of *C1QL2* (3') and 110kb upstream of *STEAP3*. *STEAP3* encodes for a transmembrane iron transporter mainly for iron uptake into red blood cells. It can be regulated by p53[412], and is involved in apoptotic processes as annotated in the Reactome database[413]. Overloading of iron is one of the hypotheses for maintenance endometrioma lesions[414], and STEAP3 as part of iron hemostasis system may play a role in it. The function of *C1QL2* is not well annotated; it may regulate excitatory synapses between hippocampus neurons[415]. One mouse study with surgically-induced endometriosis[416] showed that *C1QL2* is over-represented in gene ontology in ectopic lesions compared with eutopic endometrium.

Two nominally significant loci associated with stage III/IV disease by the 'user-provided covariate' model are both located on CpG islands and in regions that are commonly reported as differentially methylated (DMR) by experimental data (not specific to endometriosis nor tissue-specific) in public database. The locus chr10:25,465,430 is intronic to *GPR158*, which encodes a transmembrane protein for cell signals transduction. Similar to *C1QL2*, *GPR158* is also involved in neural development (e.g. hippocampal formation) as a brain osteocalcin (i.e. bone-derived hormone) receptor in mouse[417]. It is proposed to be a mediator for stress-induced depression in human[418,419], and its gene family member (*GPR88*) is related to depression in endometriosis-induced mice[420]. For its relation with reproductive track disorder, expression of GPR158 is found to be one of the prognostic biomarkers in early-stage ovarian carcinomas[421,422].

The other locus (chr1:154,475,139) is in an exon of *TDRD10*, a gene that encodes an RNA binding protein according to gene ontology[423]. Not much is known about its exact function; however, its DNA methylation status is reported to be one of the three CpG regions that may potentially help with diagnosis and prognosis of oestrogen receptor (ER)-positive breast cancer[424]. The locus is 600bp upstream of another gene, *SHE*, that harbours a GWAS locus for C-reactive protein level in circulation from a large British-ancestry study[425]. Messenger RNA (mRNA) of *TDRD10* and *SHE* are the *cis*-targets of a long non-coding RNA (lncRNA), *XLOC_001035*, which is differentially expressed in endometrioma lesions vs. eutopic endometrium[426]. This microarray study, however, reported a thousand differentially expressed lncRNA after multiple-testing correction based on 25 samples, and the level of evidence is therefore tenuous.

Nominally significant EWAS loci near GWAS variants for endometriosis sub-phenotypes

There is minimal overlap between nominally associated ($p<1x10^{-5}$) EWAS loci with GWAS variants for endometriosis; a lower threshold at $5x10^{-4}$ was applied in this dataset with generally deflated test statistics; however, results are more likely to be chance findings. One EWAS locus (chr9:136,293,231) nominally associated with endometriosis was located near the GWAS variant rs507666 (the *ABO* locus). It is at 5' shore of CpG island and intronic to *ADAMTS13*. ADAMTS13 is responsible for von Willebrand factor cleavage, and thus involves in platelet-thrombus formation[427]. The genetic variant (rs507666) 145kb upstream of the EWAS locus is nominally associated with severe dysmenorrhea sub-phenotype of endometriosis (Table 3.5). The other nominal signal (chr6:151,667,733) is at open sea and 159kb upstream of the GWAS SNP (rs2967694) intronic to *CCDC170* associated with stage III/IV endometriosis.

For association with stage I/II cases, a nominal locus (chr17:64,347,388) around rs7214750 (the *CEP112* locus) is at open sea and intronic to *PRKCA*. It is a kinase involved in broad range of biological processes, from immunity[428], thrombogenesis[429], and apoptosis[430,431], to cell migration[432], angiogenesis[433,434], and cell cycle moderation[435]. Despite its wide range of actions seemingly in line with hypotheses of endometriosis pathology, there is limited literature around the topic. One study reported an inverse relation between PRKCA and ERα protein expressions in breast and ER-negative endometrial cancers[436]. However, the genetic variant (rs7214750) 444kb downstream of this EWAS signal (for stage I/II cases) is nominally associated with stage III/IV disease instead. Similarly, the variant rs10917151 (the *WNT4* locus) 63kb upstream of the other EWAS nominal signal for stage I/II cases (chr1:22,486,183 at open sea) is only nominally associated with stage III/IV disease.

Among the 16 GWAS variants nominally associated with at least one sub-phenotype of endometriosis, six were only associated with stage III/IV disease (Chapter 3). It is interesting to see half of these locating near to a EWAS nominal loci, and constituting three out of four findings of the overlap (±500kb) between EWAS and GWAS loci, which suggests the potential for genetic influence on differential methylation at these sites.

Studying genes nearest to the EWAS loci nominally associated with endometriosis and its ASRM surgical sub-phenotypes may give insight into how aberrant methylation signals may work locally (in *cis*). The gene sets (Table 6.3) to which the nearby genes of the 14 nominal EWAS loci belong only provides a cursory overview of possible biological function of the locus, and assume that the differential methylation loci all impact on these genes based on proximity alone. Although *cis*-acting DNA methylation loci are described in the literature more frequently[437], methylation markers may work beyond immediate proximity to a gene (*trans*-acting)[438]. The investigation of these gene sets is limited by not shedding light on regulation of more distal genes (>±200kb). However, such trans effects are typically much more modest than effects in *cis*, requiring much larger sample sizes[439]. Some highlighted gene sets such as cellular signalling and differentiation are relevant to hypotheses for endometriosis pathogenesis, but they provide no novel, key insights into endometriosis biology per se.

In visualising DNA methylation status of genes near to nominally significant EWAS loci (Figure 6.8), there appears to be no differences in their average methylation status between endometriosis cases and controls. Even where a EWAS signal was mapped to a particular gene, the methylation status of other probes mapping to the same gene did not show association to the condition; thus, with the limited EWAS loci discovery in this dataset, the visualisation can only serve as an indicator for gene DNA methylation status in human endometrium in general and any patterns related to disease status would need to be checked in larger datasets. The larger NIH Epigenetic Consortium analysis including 790 cases and 504 control will improve power of detection, but given results from much larger EWAS studies[440] may still be underpowered.

In summary, the EWAS discovery for endometriosis in this analysis is minimal. However, the six biological gene clusters suggested to be differentially methylated are: 1) DNA repairment, cell cycle and apoptosis (*CCND1*, *RAD52*, *PAGR1*, *STEAP3*); 2) iron transport and metabolism (*LYRM4*, *BOLA2*, *STEAP3*); 3) immune response and inflammation (*PIANP*, *SPP1*, *TLR3*, *SHE*); 4) neural regulation and transduction (*PIANP*, *C1QL2*, *GPR158*); 5) oestrogen pathway (*PAGR1*, *ABCG2*, *SPP1*, *TLR3*, *TDRD10*); and lastly 6) histone methylation (*PAGR1*).

Future Work

The results of this Oxford dataset will contribute to an EWAS meta-analysis in the NIH Epigenome project with much greater power of discovery; gene set enrichment analysis for DNA methylation[441] can then be carried out to gain insight of enriched pathways these methylation marks may have their effect on. In addition, an endometrium methylation QTL (mQTL) map investigating genetic drivers of DNA methylation in this tissue is also being generated by the consortium. There is a lack of robust, large-scale DNA methylation data for human eutopic endometrium; thus, the generation of endometrium mQTL map is an essential first step for studying endometrial disease such as endometriosis in the field.

Future work in the Oxford dataset could include assessment of correlation between differential methylation and differential gene expression, as previous studies looking into endometrium[351,369] have shown. Moreover, future analyses need to address the heterogeneity of cell composition in endometrium samples, ideally by conducting cell-type-specific (i.e. stromal and epithelial cells) DNA methylation experiments that can be used to deconvolute whole tissue data. Statistical models and pipelines[442,443] to conduct these analyses have developed rapidly since 2019. The deconvoluted resolution of DNA methylation patterns associated with disease status by cell type would allow detangling of biological pathways affected by DNA methylation marks into vascular (endothelial), epithelial, stromal, and immunological milieu in eutopic endometrium priming for, or as a consequence of, endometriosis occurrence.

6.5 Conclusion

Despite the minimal EWAS loci discovery for endometriosis in my dataset, different nominal signals associated with endometriosis and ASRM surgical sub-phenotypes were highlighted that require replication and further exploration. The analyses show the importance of addressing the disease by sub-phenotypes that may have distinctive features not only on genomic and transcriptomic level, but also on the DNA methylome. Biological functions of gene clusters — based on nearby genes mapped to nominal EWAS loci — were similar among ASRM surgical stage I/II and III/IV cases; they relate to cell cycle, immunity, neural regulation, iron and oestrogen pathways. Moreover, the analysis revealed differentially methylated loci in cytoband 4q35.1 that may be involved in the normal regulation of menstrual cycle. This analysis will contribute to the meta-analysis in the consortium effort for further discovery of endometriosis EWAS loci, and generation of endometrium mQTL map.

Chapter 7 Conclusion and future work

In this thesis, I explored associations of established genetic variants for endometriosis with surgical and symptomatic sub-phenotypes of the condition. In addition, transcriptomic and DNA methylation signatures in eutopic endometrium were investigated in relation to surgical sub-phenotypes with adjustment of menstrual cycle confounding effects.

The sub-phenotype genetic association analyses (Chapter 3) revealed the SNP rs71575922 (6q25.1/*SYNE1*) with a nominal association with ASRM stage I/II, dysmenorrhea, and severe recent dyspareunia. The SNP locates in a genomic region which centres on *ESR1*, harbouring multiple independent signals in the overall endometriosis GWAS, and suggests a potential perturbed biological pathway involving oestrogen receptivity specifically for these sub-phenotypes. Other nominal associations were rs11674184 (2p25.1/*GREB1*) for ASRM stage III/IV and endometrioma, rs12441483 (15q15.1/*BMF*) for multiple pain sub-phenotypes, and rs1903068 (4q12/*KDR*) for endometrioma and multiple pain sub-phenotypes. As most sub-phenotypes investigated, especially those of a symptomatic nature, were not independent variables and overlapped considerably, formal statistical testing of difference of effects between these sub-phenotypes was not possible. Future analyses of mutually exclusive sub-phenotypic groups in much larger sample sizes could resolve this; however, as overlap of symptoms is inherent to endometriosis, a more appropriate future analysis strategy could be statistical clustering analyses, and — where large numbers of different symptoms recorded in different datasets are available together with genomic profiling — machine learning methodology.

In Chapter 4, I attempted to assign menstrual cycle phases in eutopic endometrium by transcriptomic signatures. This proved to be challenging, both within a specific dataset, and in establishing a multivariate model for phase prediction. The choice of genes for the model building and testing is crucial; however, there was limited agreement among published studies of genes significantly differentiating between different phases, further complicated by the use of different methodological platforms for transcript detection (i.e. whole tissue-based microarray vs. RNA-seq). Future availability of independent transcriptomic datasets for endometrium would allow external validation of proposed models. Although accurate menstrual cycle phase assignment based on transcriptomic profiles could not be achieved, the differential expression analysis between cycle phases confirmed previous literature reports of a significant and considerable cycle phase effect on gene expression patterns. It will mask any potential transcriptional differences between endometrium samples from endometriosis cases and controls if analyses are unadjusted for cycle phase. The analysis also highlighted the complement system as one of the most robustly enriched and up-regulated pathways in the secretory phase of the endometrium compared to the proliferative phase. In the differential expression analysis (Chapter 5) eutopic endometrium between endometriosis cases and controls, there were limited discoveries when overall endometriosis was considered. However, potential signals were found between ASRM surgical sub-phenotypes on chromosome 21q11.2 — a genomic region associated with cytokine levels in circulation. Five gene clusters with similarity in biological functions were highlighted comparing ASRM surgical status and clinic controls: 1) innate immune response (*BPIFB1*, *MUC5AC*, *MUC5B*), including macrophage attraction (*EDN2*, *THBS1*); 2) acute inflammation response (*SAA1*, *SAA2*); 3) neural proliferation (*FAT2*, *FAT3*); 4) haemoglobin-related (*HBG2*, *ALAS2*); and 5) possible steroidogenesis via regulation of lipid metabolism (*PLIN1*, *PLIN4*, *GPD1*). Comparison of the above results to the literature was difficult because of non-robust (due to small sample size) and minimal reports of differentially expressed genes in overall endometriosis vs. controls in publications. Another limitation in our analysis was that some samples had only self-reported cycle phase available for the analysis, which is less reliable than histology confirmation in menstrual cycle assignment. The signals outlined in this thesis warrant further validation in an independent dataset, ideally with >500 samples with histological-confirmation on menstrual cycle phases for robust signal detection. Cell-type specific RNA-seq on epithelial and stromal cell populations of the endometrium should also be conducted to further understand the origin driving the signal observed.

Similarly, considering epigenome-wide differential DNA methylation analyses (Chapter 6), there were few discoveries when analysing the unstratified endometriosis case group vs. controls; however, nominally significant signals were observed between ASRM surgical sub-phenotypes and clinic controls, revealing six gene clusters, by similarity in function: 1) DNA repairment, cell cycle and apoptosis (*CCND1*, *RAD52*, *PAGR1*, *STEAP3*); 2) iron transport and metabolism (*LYRM4*, *BOLA2*, *STEAP3*); 3) immune response and inflammation (*PIANP*, *SPP1*, *TLR3*, *SHE*); 4) neural regulation and transduction (*PIANP*, *C1QL2*, *GPR158*); 5) oestrogen pathway (*PAGR1*, *ABCG2*, *SPP1*, *TLR3*, *TDRD10*); and 6) histone methylation (*PAGR1*). There was minimal overlap of differential DNA methylation signals with location of endometrium eQTL SNPs. Ideally, future studies should adjust for cell composition, age, and smoking status. Meta-analysis of these results with other DNA methylation datasets for eutopic endometrium involving women of European ancestry is underway. Correlation of differential methylation with differential gene expression should be explored in samples with both transcriptomic and DNA methylation profiling available.

This thesis addresses genomic, transcriptomic, and epigenomic molecular signatures in sub-phenotypes of endometriosis, with a focus on ASRM surgical sub-phenotypes. Transcriptomic and DNA methylation analyses were not feasible on symptomatic sub-phenotypes due to restricted sample size. Oestrogen receptivity, lipid metabolism, immune response, cell cycle, haemoglobin/iron metabolism, and neural factors are among biological pathways that may explain differences between eutopic endometrium underlying ASRM stage I/II, stage III/IV disease and clinic controls. The methods outlined for transcriptomic and epigenetic analyses can be applied to future larger WERF-EPHect-compliant datasets to study a wider range of symptomatic sub-phenotypes of disease that this thesis could cover.

Supplementary materials

A. List of publications and presentations

Chapter 2 Publications with contribution

T. T. Tapmeier, H. M. Nazri, K. S. Subramaniam, S. Manek, K. Garbutt, E. J. Flint, **C. Cheuk**, C.Hubbard, K. Barrett, E. Shepherd, K. T. Zondervan, and C. M. Becker. (2020). 'Protocol for a longitudinal, prospective cohort study investigating the biology of uterine fibroids and endometriosis, and patients' quality of life: the FENOX study.' BMJ Open, 10: e032220.

- Designing framework of electronic phenotypic data collection, storage, and traceability;
- Discussing workflow from patient recruitment to sample management

N. Rahmioglu, K. Banasik, P. Christofidou, ... **C. Cheuk**, ….S. A. Missmer, G. W. Montgomery, A. P. Morris and K. T. Zondervan, et al. (2018). 'Large-scale genome-wide association meta-analysis of endometriosis reveals 13 novel loci and genetically-associated comorbidity with other pain conditions.' bioRxiv, 406967.

- Responsible for laboratory processing of Oxford samples and parallel analysis of genotyping QC

Chapter 3 Conference paper and presentation

Besides contribution of ENDOX Phase I & II data structure and analysis to the latest IGEC sub-phenotypic endometriosis paper (under submission), preliminary results were also shared in several scientific meetings. A conference paper titled 'Understanding Endometriosis Heterogeneity - Sub-phenotype Analysis of Genetic Associations' was accepted and granted oral presentation in 'The 34th Annual Meeting of European Society of Human Reproduction and Embryology (ESHRE)' conference in Barcelona, Spain on 1-4 July 2018. In addition, preliminary data for deep endometriosis in particular were shared to Prof Krina Zondervan for her presentation on the ESHRE campus course titled: 'Deep Endometriosis – from Pathophysiology to Clinic' in Münster, Germany on 16-17 September 2019.

SELECTED ORAL COMMUNICATIONS

SESSION 03: ENDOMETRIOSIS AND ENDOMETRIUM: DISCOVERY SCIENCE

Monday 2 July 2018 Room 211 + 212 10:00–11:30

O-010 Understanding endometriosis heterogeneity—sub-phenotype analysis of genetic associations

C.S.K. Cheuk[1], N. Rahmioglu[2], C. Becker[1], K. Zondervan[1]

[1]*University of Oxford, Nuffield Department of Women's and Reproductive Health, Oxford, United Kingdom*

[2]*University of Oxford, Wellcome Centre for Human Genetics, Oxford, United Kingdom*

Study question: The aim of this study is to investigate whether sub-phenotype analysis of established genetic variants associated with endometriosis reveals differential mechanisms underlying surgical and symptomatic subtypes of the disease.

Summary answer: Preliminary results from a reduced dataset showed suggestive differential effects for genome-wide association study (GWAS) loci across surgical sub-phenotypes.

What is known already: Most of the known 14 genetic loci have been reported based on association with overall endometriosis, with sub-phenotypic investigations limited to ASRM stages I/II vs. III/IV. Studies using the World Endometriosis Research Foundation-Endometriosis Phenome and Biobanking Harmonisation Project (EPHect) standardized data collection protocols now afford the investigation of these loci for detailed surgical and symptomatic phenotypes.

Study design, size, duration: Cross-sectional study in which genotype information from 358 women in the Oxford ENDOX study (294 cases, 64 controls) undergoing laparoscopic surgery was analysed for association with EPHect-compliant surgical and symptomatic phenotype data.

Participants/materials, setting, methods: Preliminary analysis of 108 endometriosis cases included 51% with an endometrioma (not exclusive to other surgical subtypes), 68% with peritoneal disease and 31% with deep endometriosis; 40%, 14%, 19% and 27% of women were diagnosed with ASRM Stage I to IV respectively.

All 14 known GWAS loci for endometriosis were tested for association with sub-phenotypes using Pearson's chi-squared and linear-by-linear association test.

Main results and the role of chance: The GWAS locus on 2q35 (rs1250241, located in *FN1*), with genotype T/A, showed a borderline association ($p = 0.035$) with peritoneal endometriosis (odds ratio (OR) of TT vs. TA = 0.08, 95% confidence interval (CI): 0.01-0.78; OR of TT vs. AA = 0.13, 95% CI: 0.01-1.25; OR of TA vs. AA = 1.68, 95% CI:0.69-4.07), but not with endometrioma ($p = 0.856$) or deep endometriosis ($p = 0.307$). Analysis on additional genotype data ($n = 250$) with more sub-phenotypes including infertility and pain status are underway and results will be presented at the meeting.

Limitations, reasons for caution: Sample size was small in the preliminary analysis, with limited power to discover differential association pattern.

Wider implications of the findings: Insight into aetiological heterogeneity of endometriosis elucidated by genetics is an important avenue to ultimately develop targeted non-invasive markers of disease and novel treatment strategies.

Trial registration number: Not applicable.

C. S. K. Cheuk, N. Rahmioglu, C. Becker, K. Zondervan (2018). "Understanding endometriosis heterogeneity - sub-phenotype analysis of genetic associations." Human Reproduction, 33, 5.

Chapter 4 Abstract and poster

For the preliminary work on 'Differential Gene Expression between Menstrual Phases in Endometrium', the following abstract and poster presentation were accepted in 'The Genomics of Common Diseases 2017' conference jointly organised by the Nature Genetics and Wellcome Genome Campus Conference Centre on 6-8 September 2017. I was awarded a bursary to attend.

From histology to molecular profiling—Menstrual phase prediction by RNA sequencing signatures in human endometrium

Cecilia SK Cheuk[1], Nilufer Rahmioglu[2], Andrew P. Morris[3], Sanjiv Manek[4], Karin Hellner[1], Christian Becker[1], Krina T Zondervan[1,2]

[1] Nuffield Department of Obstetrics and Gynaecology, University of Oxford, UK. [2] Wellcome Trust Centre for Human Genetics, University of Oxford, UK. [3] Institute of Translational Medicine, University of Liverpool, UK. [4] Department of Cellular Pathology, Oxford University Hospitals NHS Trust, UK.

Endometriosis is defined as the presence of tissue resembling endometrium outside the uterus. It is associated with severe pelvic pain and reduced fertility, affecting an estimated 5-10% of women of reproductive age. Genome-wide association studies have revealed genetic variants associated with endometriosis at 14 loci [141], but for most the causal mechanisms remain to be elucidated in the target tissue: endometrium. Endometrium is a highly dynamic tissue, regulated by various hormones throughout menstrual cycle. Both the gene expression [82] and DNA methylome landscapes of endometrium are known to show significant cyclic changes throughout different menstrual phases; therefore, menstrual phase is crucial to account for in any molecular profiling of endometrium. The menstrual phases exhibit specific structural changes, typically assessed through histological examination of biopsies. However, histological confirmation is not always available, especially for gene expression data extracted from open source databases such as GTEx.
The aim of this study is to investigate whether menstrual phase can be accurately predicted from RNA sequencing data of human endometrium.
Endometrium samples (n=192) were collected by WERF-EPHect standardized protocols from women undergoing laparoscopic surgery in the Oxford ENDOX study (129 endometriosis cases; 63 controls). Tissues were profiled using next-generation RNA sequencing (Illumina HiSeq4000). Histological confirmation of menstrual phase of samples was obtained based on H&E staining of frozen tissues slides. Principal components analysis confirmed the strong menstrual effect on RNA seq gene expression profiles of endometrium, with first three components explaining 62% of the variation. Preliminary analyses revealed 6,510 differentially expressed genes between proliferative and secretory phases, 4,077 between secretory and menstrual phases, and 4,875 between proliferative and menstrual phases (FDR-p<0.05). Some top signals include *MMP7*, *MMP10* and *INHBA.* Building of a predictive algorithm for menstrual phases with analysis of discovery and validation datasets is ongoing. Development of a gene expression signature based classification of menstrual phase will aid studies into functional mechanism involved in any endometrium-related condition, including endometriosis, infertility and endometrial cancer.

From Histology to Molecular Profiling: Menstrual Phase Prediction by RNA Sequencing Signatures in Human Endometrium

Cecilia SK Cheuk[1], Nilufer Rahmioglu[2], Andrew P Morris[3], Sanjiv Manek[4], Karin Hellner[1], Christian M Becker[1], Krina T Zondervan[1,2]

[1]Nuffield Department of Obstetrics and Gynaecology, University of Oxford. [2]Wellcome Trust Centre for Human Genetics, University of Oxford.
[3]Institute of Translational Medicine, University of Liverpool. [4]Department of Cellular Pathology, Oxford University Hospitals NHS Trust, United Kingdom.

Introduction

- **Endometrium** is a dynamic tissue, regulated by hormones during menstrual cycle. Both gene expression(1) and DNA methylome(2) landscapes of endometrium show significant cyclic changes throughout different menstrual phases (i.e. menstrual, proliferative/follicular & secretory/luteal phase)
- **Menstrual phase** is crucial to be accounted for in any molecular profiling of endometrium for understanding functional mechanism involved in any endometrium-related condition, e.g. endometriosis, infertility and endometrial cancer
- **Endometriosis** is defined as the presence of tissue resembling endometrium outside the uterus. Genome-wide association studies have revealed genetic variants associated with endometriosis at 14 loci(3), but for most the causal mechanisms remain to be elucidated in the target tissue: endometrium
- **Aim – to investigate whether menstrual phase can be accurately predicted from RNA sequencing data of human endometrium**

Materials and Methods

1. Endometrium samples (n=190) were collected by WERF-EPHect standardized protocols from women undergoing laparoscopic surgery in the Oxford ENDOX study (128 endometriosis cases; 62 controls)
2. Tissues were profiled using next-generation RNA sequencing (Illumina HiSeq4000)
3. Histological confirmation of menstrual phase of samples was obtained based on H&E staining of frozen tissues slides
4. Histology-confirmed **proliferative** (pro) and **secretory** (sec) phase samples (n=70) (Fig.1b & c) were analyzed with their differential expression pattern by DESeq and edgeR packages in R
5. Gene filter of count-per-million (CPM)>1 in 80% of the samples resulted in analysis of ~18,000 genes

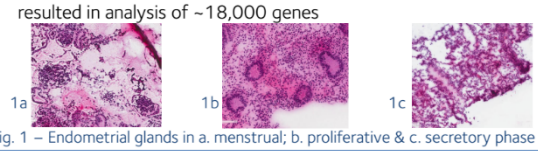


Fig. 1 – Endometrial glands in a. menstrual; b. proliferative & c. secretory phase (200X)

Results

Proliferative (n=24) vs Secretory Phase (n=46)

1. Principal components analysis (PCA) with first three components explaining 75.3% of the variation (Fig. 2)

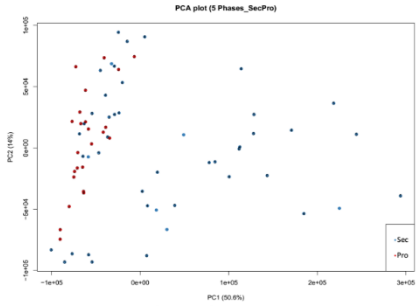


Fig. 2 – PCA plot for Proliferative (red) vs Secretory phase (blue) samples, with PC 1 (50.6%) and PC 2 (14%)

2. Analysis revealed 7945 differentially expressed (DE) genes at FDR-p<0.05 (Fig. 3) out of 18,116 expressed genes (CPM>1 in 80% of 70 samples)
3. Among top 20 differentially expressed genes, two of them, i.e. *CXCL14* and *GPX3* have been reported in previous gene expression studies(1) (Table 1)

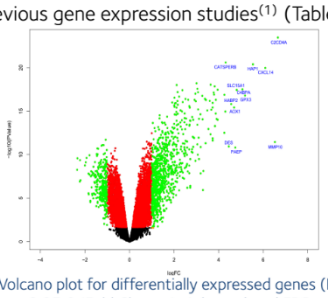


Fig. 3 – Volcano plot for differentially expressed genes (Pro vs Sec), with FDR-p<0.05 & |Fold Change|>1 (green) and FDR-p<0.05 (red)

Table 1 – Top 20 DE Genes for Proliferative vs Secretory

DE Genes	FDR p-value	Log (Fold Change)
C2CD4A	6.25E-20	6.66
CATSPERB	2.31E-17	4.31
HAP1	2.51E-17	5.54
CXCL14	4.87E-17	6.09
SDCBP2	2.27E-15	3.01
MFSD4A	2.58E-15	3.64
ACKR2	2.58E-15	2.12
C2CD4B	3.20E-15	3.82
C4BPA	5.45E-15	5.07
SLC15A1	5.45E-15	4.81
C10orf10	5.45E-15	3.96
APOD	5.45E-15	3.05
RLIMP1	5.45E-15	2.37
RIMKLB	8.11E-15	3.08
FOXO1	8.11E-15	2.16
AC073218.2	8.24E-15	3.78
PHYHIPL	1.05E-14	2.96
SPP1	1.25E-14	4.26
NNMT	1.25E-14	3.80
GPX3	1.45E-14	5.19

Mid-secretory (n=19) vs Late-secretory Phase (n=18)

1. Principal components analysis (PCA) with first three components explaining 77.8% of the variation
2. No global clustering for mid-secretory versus late-secretory phase (Fig.4)
3. Analysis revealed 12 differentially expressed (DE) genes at FDR-p<0.05 (Table 2) out of 18,029 expressed genes (CPM>1 in 80% of 37 samples)
4. Among12 differentially expressed genes, *C7* was reported in previous gene expression studies(1)



Fig. 4 – Cluster plot for Mid-secretory vs Late-secretory phase samples, with Mid-sec denoted with '4' and Late-sec denoted with '5'

Table 2 – DE Genes for Mid-secretory vs Late-secretory

DE Genes	FDR p-value	Log (Fold Change)
DES	0.00097	3.96
HIST2H2AA3	0.002804	-2.61
MYH11	0.002804	2.47
COL8A1	0.002804	2.02
ACTG2	0.003028	3.35
PDLIM3	0.003028	3.26
PTGIS	0.003065	2.60
CHRDL1	0.003323	2.92
KIAA1755	0.022362	1.25
AL592490.1	0.029542	1.63
C7	0.040375	2.53
PRELP	0.040375	2.19

Conclusion & Future Work

- Highly differentially expressed genes between proliferative and secretory phase provide strong basis for menstrual phase prediction
- Predictive algorithm for menstrual phases on RNA-Seq data is being developed based on differentially expressed genes between three menstrual phases, with the possibility to predict sub-secretory phases
- Menstrual phases of samples can be predicted on open-source RNA-Seq data on endometrial-related conditions with the predictive algorithm, assisting future discovery of disease mechanisms with adjustments based on menstrual phase effect

Reference: 1. Ponnampalam AP, et al. Mol Hum Reprod. 2004;10(12):879-93. 2. Saare M, et al. Clin Epigenetics. 2016;8:2. 3. Sapkota Y, et al. Nature Communications. 2017;8:15539.

B List of supplementary figures

Supplementary Figure S 2.1 Flow chart of sample and variant QCs steps for genotype data (n=772) from discovery phase (year 1) of NIH epigenetic of endometriosis consortium. ... 226
Supplementary Figure S 2.2 Q-Q plot of GWAS for endometriosis status of the NIH discovery dataset ... 226
Supplementary Figure S 2.3 Manhattan plot of GWAS for endometriosis status of the NIH discovery dataset ... 227
Supplementary Figure S 3.1 Correlation heat map of 16 major sub-phenotypes of endometriosis based on patients reporting age of onset for dysmenorrhea (n=162), coloured coded by Pearson's correlation coefficient. ... 228
Supplementary Figure S 3.2 forest plot for effect estimates of the C allele of rs3858429 on 11p14.1, upstream of *ARL14EP*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. ... 229
Supplementary Figure S 3.3 Forest plot for effect estimates of the A allele of rs10917151 on 1p36.12, downstream of *CDC42 & WNT4*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. ... 230
Supplementary Figure S 3.4 Forest plot for effect estimates of the G allele of rs6456259 on 6p22.3, upstream of *ID4*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. ... 231
Supplementary Figure S 3.5 Forest plot for effect estimates of the C allele of rs7740302 on 6q25.1, in the intron of *CCDC170*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. ... 232
Supplementary Figure S 3.6 Forest plot for effect estimates of the G allele of rs1451383 on 7p15.2, upstream of *NFE2L3*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. ... 233
Supplementary Figure S 3.7 Forest plot for effect estimates of the C allele of rs7214750 on 17q24.1, in the intron of *CEP112*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. ... 234
Supplementary Figure S 3.8 Forest plot for effect estimates of the T Allele of rs7907732 on 10q23.31, in the intron of *RNLS*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. ... 235
Supplementary Figure S 3.9 Forest plot for effect estimates of the A allele of rs507666 on 9q34.2, in the intron of *ABO*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. ... 236
Supplementary Figure S 3.10 Forest plot for effect estimates of the A allele of rs2967684 on 19p13.2, downstream of *ACTL9*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. ... 237
Supplementary Figure S 3.11 Forest plot for effect estimates of the A allele of rs2510770 on 4q22.3, in the intron of *PDLIM5*, for major surgical, infertility, and pain sub-phenotypes of endometriosis. ... 238
Supplementary Figure S 3.12 Correlation coefficient (Pearson's r^2) of effect estimates of tested genetic variants (n=46) on sub-phenotypic pairs for endometriosis ... 239
Supplementary Figure S 3.13 Correlation plots of effect estimates of 46 genetic variants for a) deep lesion (Pearson's r^2= 0.64) and b) endometrioma (r^2= 0.68) versus superficial lesion sub-phenotype of endometriosis. ... 240

Supplementary Figure S 3.14 Correlation plots of effect estimates of 46 genetic variants for a) severe dysmenorrhea (Pearson's r^2= 0.74), b) acyclical pelvic pain (r^2= 0.82) versus dyspareunia, and c) bowel movement/IBS pain versus bladder pain sub-phenotype (r^2= 0.82) of endometriosis. 241
Supplementary Figure S 3.15 Forest plot for effect estimates of the A allele of rs507666 on 9q34.2, in the intron of *ABO*, for sensitivity testing of endometriosis with dysmenorrhea sub-phenotype. 242
Supplementary Figure S 3.16 Forest plot for effect estimates of the A allele of rs2967684 on 19p13.2, downstream of *ACTL9*, for sensitivity testing of dyspareunia and acyclical pelvic pain sub-phenotypes of endometriosis. 243
Supplementary Figure S 3.17 Forest plot for effect estimates of the A allele of rs2510770 on 4q22.3, in the intron of *PDLIM5*, for sensitivity testing of endometriosis with dyspareunia sub-phenotype. 244
Supplementary Figure S 3.18 WERF-EPHect clinical questionnaire (standard) (EPQ-S v1.0) Question C17 preceding capturing of the dyspareunia pain NRS for recent (last) time. 245
Supplementary Figure S 4.1 Characteristics of endometrial glands in a. menstrual, b. proliferative, c. early-secretory, d. mid-secretory, e. late-secretory phase & f. mucus only (200X; scale bar=100μm). 246
Supplementary Figure S 4.2 PCA plot for samples (n=80) with histological-confirmed five cycle phases, with PC 1 (22.7%) and PC 2 (12.4%). 247
Supplementary Figure S 4.3 PCA plot for samples (n=80) with histological-confirmed five cycle phases, with PC 3 (6.9%) and PC 4 (4.9%). 248
Supplementary Figure S 4.4 PCA plot for secretory samples only (n=53), with PC 3 (7.6%) and PC 4 (4.4%). 249
Supplementary Figure S 4.5 PCA plot for showing interaction between two terms, the histology cycle phases and the endometriosis case-control status, in the design matrix (`design ~ histology phase + case-control status` by `DESeq2`). 250
Supplementary Figure S 4.6 Scatter plot of log-scale adjusted *p*-value of differentially expressed genes between proliferative and secretory phases common to `edgeR` and `DESeq2` analyses. 250
Supplementary Figure S 4.7 Scatter plot of ranking by adjusted *p*-value of differentially expressed genes common to in-house `edgeR` (proliferative vs. all secretory) and Fung et al. (proliferative vs. early-secretory) analyses. 251
Supplementary Figure S 4.8 Scatter plot of log-scale adjusted *p*-value of differentially expressed genes common to in-house `edgeR` (proliferative vs. all secretory) and Fung et al. (proliferative vs. early-secretory) analyses. 252
Supplementary Figure S 4.9 Heat map of correlation of expression patterns between 26 selected genes for building the multivariate prediction model to distinguish proliferative and secretory phases. 253
Supplementary Figure S 4.10 Residuals-versus-fitted plot (Left) and Quantile-quantile (Q-Q) plot (Right) of the 18-gene model for proliferative and secretory phase prediction. 254
Supplementary Figure S 4.11 Accuracy variant signature curves from prediction for proliferative and secretory phases based on the 18-gene model. 254
Supplementary Figure S 5.1 Workflow of further sample QCs (MDS plots) for proliferative samples, with removal of outliers in three steps (from n=52 to n=44 final dataset & n=38 subset for sensitivity testing). 255

Supplementary Figure S 5.2 Workflow of further sample QCs (MDS & box plots for normalised and log-transformed counts) for secretory samples, with removal of outliers in five steps (from n=104 to n=91 final dataset)..... 256
Supplementary Figure S 5.3 PCA plot on variance stabilising transformed (VST) counts of 91 secretory samples for observation of possible interaction between the endometriosis stages-control and the sub-secretory phase. 257
Supplementary Figure S 5.4 Volcano plot for differentially expressed genes between endometriosis cases (n=34) and clinic controls (n=10) in proliferative phase of the endometrium. 258
Supplementary Figure S 5.5 Volcano plot for differentially expressed genes between endometriosis stage III/IV cases (n=16) and clinic controls (n=10) in proliferative phase of the endometrium. 259
Supplementary Figure S 5.6 Volcano plot for differentially expressed genes between endometriosis cases (n=60) and clinic controls (n=31) in secretory phase of the endometrium. 260
Supplementary Figure S 5.7 Volcano plot for differentially expressed genes between endometriosis stage I/II cases (n=34) and clinic controls (n=31) in secretory phase of the endometrium. 261
Supplementary Figure S 5.8 Correlation plot of Log_2 fold changes of 12 differentially expressed genes in both comparisons of (1) Endometriosis stage III/IV vs. I/II in secretory endometrium, and (2) Late-secretory vs. mid-secretory endometrium regardless of disease status. 262
Supplementary Figure S 5.9 Box plot of raw count for *ZNF26* gene in 44 proliferative samples, grouped by endometriosis disease status. 263
Supplementary Figure S 5.10 Correlation plots of raw counts for *ANKRD20A11P*, *CYP4F29P* & *SNX18P13* per sample in proliferative and secretory phase. 264
Supplementary Figure S 5.11 Box plot of raw counts for *ANKRD20A11P*, *CYP4F29P* & *SNX18P13* in proliferative and secretory samples, grouped by endometriosis disease status. 265
Supplementary Figure S 5.12 Heat map matrix of pairwise linkage disequilibrium statistics for variants in the region of chromosome 21: position 15,200,585-15,352,881 (GRCh37) in relation to genomic locations of *ANKRD20A11P* (36.5 kb downstream of *SNX18P13*) and *CYP4F29P* pseudogenes. 266
Supplementary Figure S 6.1 Density plots showing (top panel) raw beta value, and (bottom panel) pre-processed normalised beta value of type I green probes in eutopic endometrium samples for endometriosis cases (red line) and clinic controls (blue line). Left panel shows lines from every samples, while right panel shows the average density of all samples. 267
Supplementary Figure S 6.2 Density plots showing (top panel) raw beta value and (bottom panel) pre-processed normalised beta value of type I green probes in eutopic endometrium samples according to ASRM surgical staging. Left panel shows lines from every samples, while right panel shows the average density of all samples. 267
Supplementary Figure S 6.3 Density plots showing (top panel) raw beta value and (bottom panel) pre-processed normalised beta value of type I green probes in eutopic endometrium samples according to five cycle phases. Left panel shows lines from every samples, while right panel shows the average density of all samples. 268

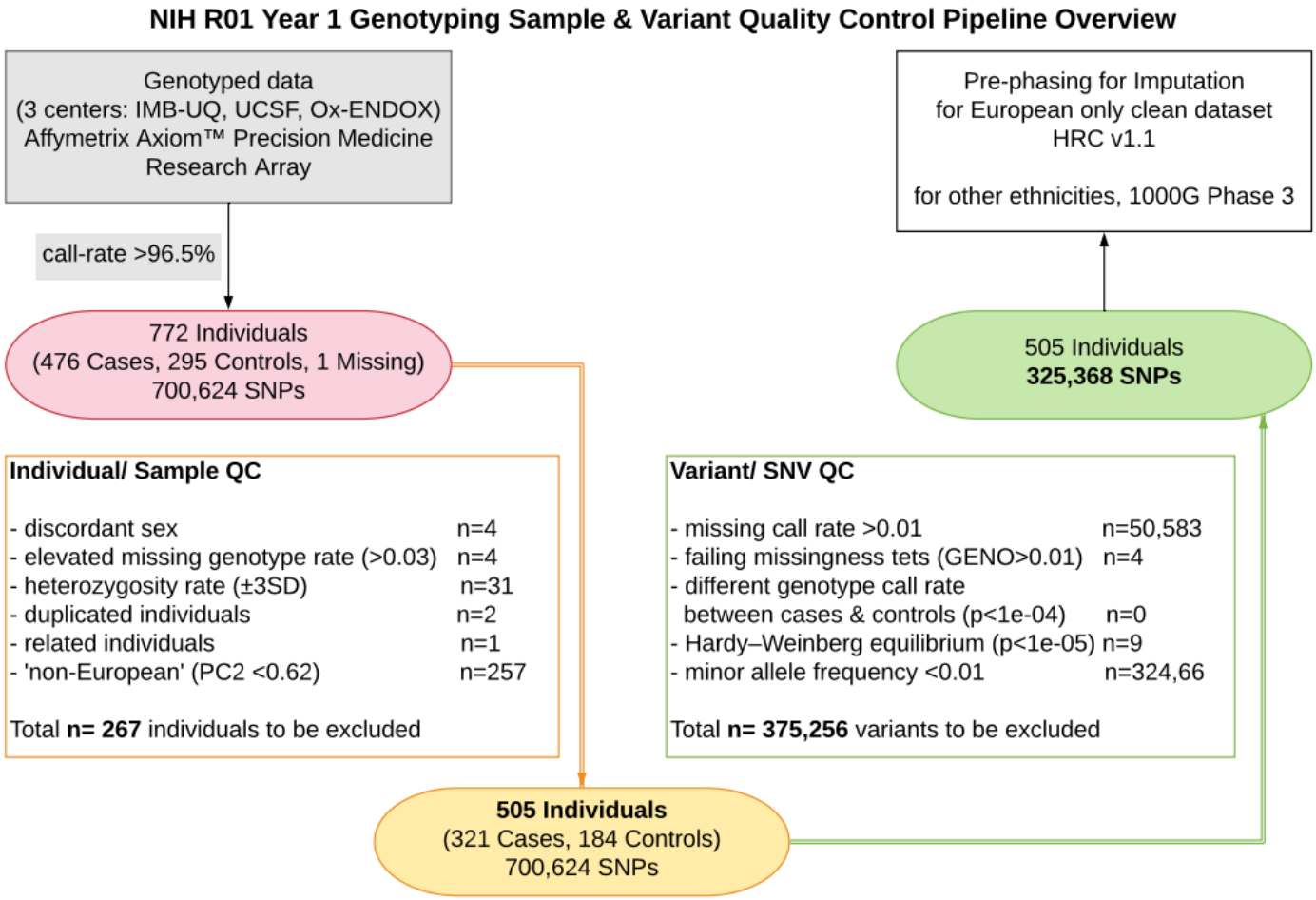
Supplementary Figure S 6.4 Plot of differences between total mean intensity of probes from X and Y chromosomes. 269
Supplementary Figure S 6.5 Samples (n=163) passing DNA methylation detection *p* values (left) and bead number for probes (right). 270
Supplementary Figure S 6.6 Plot of median methylation (M) versus unmethylation (UM) intensity of samples. 270
Supplementary Figure S 6.7 Methylation probes showing proportion of samples with detection *p* values (left) and bead number (right). 271
Supplementary Figure S 6.8 Scree plots for residual fitting of serial number of principal components (PCs) for adjustment of technical effects on DNA methylation signals (left), and variance captured by each PCs for control probes (right). 271
Supplementary Figure S 6.9 PCA plots of samples after QC (n=145) colour-coded by discovery and replication phase (top panel), by endometriosis case-control status (middle panel), and by ASRM surgical stages (bottom panel). 272
Supplementary Figure S 6.10 Quantile-quantile plots of EWAS analysis of endometriosis cases (n=100) and clinic controls (n=38) in four models: without any covariates (top left), with four user-provided covariates (top right), independent surrogate and user-provided variables — ISVA (bottom left), and surrogate and user-provided variables — SVA (bottom right). 273
Supplementary Figure S 6.11 Box plots of DNA methylation level of two probes at chr11:69340400 (Illumina probe ID cg02374789, left) and chr12:1068597 (cg17573903, right) in EWAS of endometriosis case-control comparison by SVA model. 274
Supplementary Figure S 6.12 Quantile-quantile plots of EWAS analysis of endometriosis ASRM stage I/II cases (n=68) and clinic controls (n=38) in four models: without any covariates (top left), with two user-provided covariates (top right), independent surrogate and user-provided variables — ISVA (bottom left), and surrogate and user-provided variables — SVA (bottom right). 275
Supplementary Figure S 6.13 Quantile-quantile plots of EWAS analysis of endometriosis ASRM stage III/IV cases (n=32) and clinic controls (n=38) in four models: without any covariates (top left), with two user-provided covariates (top right), independent surrogate and user-provided variables — ISVA (bottom left), and surrogate and user-provided variables — SVA (bottom right). 276

C List of supplementary tables

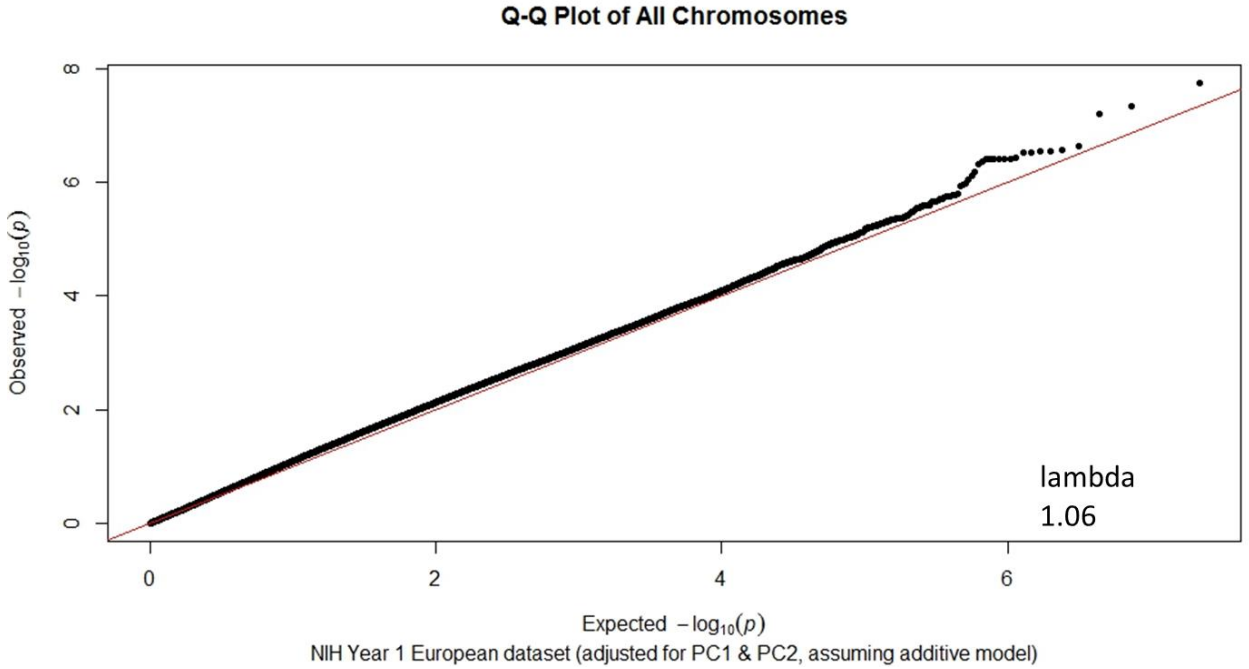
Supplementary Table S 3.1 Preliminary association testing of selected 38 SNPs for overall endometriosis against three tiers of controls in ENDOX Phase I. 277
Supplementary Table S 3.2 Meta-analysed association results of 46 loci for overall endometriosis, and ASRM surgical classification of endometriosis by stages. 278
Supplementary Table S 3.3 Meta-analysed association results of 46 loci for three surgical sub-phenotypes of endometriosis, and endometriosis with bowel movement/IBS pain. 279
Supplementary Table S 3.4 Meta-analysed association results of 46 loci for endometriosis with dysmenorrhea by severity. 280
Supplementary Table S 3.5 Meta-analysed association results of 46 loci for endometriosis with dysmenorrhea by age of onset, and endometriosis with infertility. 281
Supplementary Table S 3.6 Meta-analysed association results of 46 loci for endometriosis with dyspareunia by severity and temporality (part A). 282
Supplementary Table S 3.7 Meta-analysed association results of 46 loci for endometriosis with acyclical pelvic pain by severity and temporality (part A). 283
Supplementary Table S 3.8 Meta-analysed association results of 46 loci for endometriosis with mild to moderate dyspareunia, and acyclical pelvic pain by temporality (part B). 284
Supplementary Table S 3.9 Meta-analysed association results of 46 loci for endometriosis with bladder pain, and endometriosis with ≥ 2, ≥ 3, and ≥ 4 reported pain types. 285
Supplementary Table S 3.10 Chi-square statistics for patterns in missing report of dyspareunia in pain Numeric Rating Scale (NRS, recent) for a) dysmenorrhea severity, b) acyclical pelvic pain severity, c) number of reported pain types, and d) age groups. 286
Supplementary Table S 4.1 Breakdown of number of samples in histological-confirmed and self-reported cycle phases. 288
Supplementary Table S 4.2 Top 50 annotated differentially expressed (DE) genes between histological-confirmed proliferative (n=23) versus all secretory (n=53) samples by `DESeq2` (2nd) analysis, with comparison to results from `edgeR` (1st) pipeline. 289
Supplementary Table S 4.3 Top 50 differentially expressed (DE) genes between histological-confirmed proliferative (n=23) versus all secretory (n=53) samples by `edgeR` pipeline, with comparison to Fung et al. and Talbi et al. reports. 290
Supplementary Table S 4.4 Top 21 differentially expressed (DE) genes between histological-confirmed mid-secretory (n=23) versus late-secretory (n=18) samples by `edgeR` pipeline. 291
Supplementary Table S 4.5 List of 26 genes for multivariate model building for proliferative and secretory phase prediction (with single-gene logistic regression statistics). 292
Supplementary Table S 4.6 List of significantly enriched Ingenuity canonical pathways (Benjamini-Hochberg $p \leq 0.05$, i.e. -log $p \geq 1.3$, n=168) for differentially expressed genes between proliferative and secretory phase endometrium by IPA. 293
Supplementary Table S 4.7 List of significantly enriched diseases and bio-function annotations with predicted activation state (z-score >+2, n=89) for differentially expressed genes between proliferative and secretory phase endometrium by IPA. 296
Supplementary Table S 4.8 List of top 50 significantly enriched diseases and functional annotations, by significance, for differentially expressed genes between proliferative and secretory phase endometrium by IPA. 300

Supplementary Table S 4.9 List of enriched Reactome canonical pathways ($p \leq 0.05$, n=57) for differentially expressed genes between proliferative and secretory phase endometrium by PANTHER. 302
Supplementary Table S 5.1 Top 50 nominally differentially expressed (DE) genes between endometriosis cases (n=34) versus clinic controls (n=10) within proliferative samples, with comparison to results from sensitivity analysis (n=38, subset of proliferative samples), previous reports from the literature, and lookup of endometrium *cis*-eQTL SNPs. 304
Supplementary Table S 5.2 Top 51 differentially expressed (DE) genes between ASRM stage I/II cases (n=18) versus clinic controls (n=10) within proliferative samples, with comparison to results from sensitivity analysis (n=24, subset of proliferative samples), previous reports from the literature, and lookup of endometrium *cis*-eQTL SNPs. 305
Supplementary Table S 5.3 Top 50 nominally differentially expressed (DE) genes between ASRM stage III/IV cases (n=16) versus clinic controls (n=10) within proliferative samples, with comparison to results from sensitivity analysis (n=21, subset of proliferative samples), previous reports from the literature and lookup of endometrium *cis*-eQTL SNPs. 306
Supplementary Table S 5.4 Top 50 nominally differentially expressed (DE) genes between ASRM stage I/II cases (n=18) versus stage III/IV (n=16) within proliferative samples, with comparison to results from sensitivity analysis (n=31, subset of proliferative samples), previous reports from the literature, and lookup of endometrium *cis*-eQTL SNPs. 307
Supplementary Table S 5.5 Top 50 nominally differentially expressed (DE) genes between endometriosis cases (n=60) versus clinic controls (n=31) within secretory samples, with comparison to results from model without adjustment of sub-secretory phase as covariate, results from DE analysis of mid-secretory versus late-secretory samples, and lookup of endometrium *cis*-eQTL SNPs (No genes were reported in 3 literatures). 308
Supplementary Table S 5.6 Top 50 nominally differentially expressed (DE) genes between ASRM stage I/II (n=34) versus clinic controls (n=31) within secretory samples, with comparison to results from DE analysis of mid-secretory versus late-secretory samples, previous reports from the literature, and lookup of endometrium *cis*-eQTL SNPs. 309
Supplementary Table S 5.7 Top 50 nominally differentially expressed (DE) genes between ASRM stage III/IV (n=26) versus clinic controls (n=31) within secretory samples, with comparison to results from DE analysis of mid-secretory versus late-secretory samples, and previous reports from the literature (No genes were overlapped with endometrium *cis*-eQTL SNPs). 310
Supplementary Table S 5.8 Top 50 differentially expressed (DE) genes between ASRM stage III/IV (n=26) versus stage I/II (n=34) within secretory samples, with comparison to results from DE analysis of mid-secretory versus late-secretory samples, previous reports from the literature, and lookup of endometrium *cis*-eQTL SNPs. 311
Supplementary Table S 6.1 Table of technical and biological variables in significant association with the first three principal components (PCs) of DNA methylation data after quality control. 312
Supplementary Table S 6.2 Number of eutopic endometrium samples in each of five cycle phases in EWAS comparisons of endometriosis, ASRM surgical stage I/II, and stage III/IV cases to clinic controls. 312

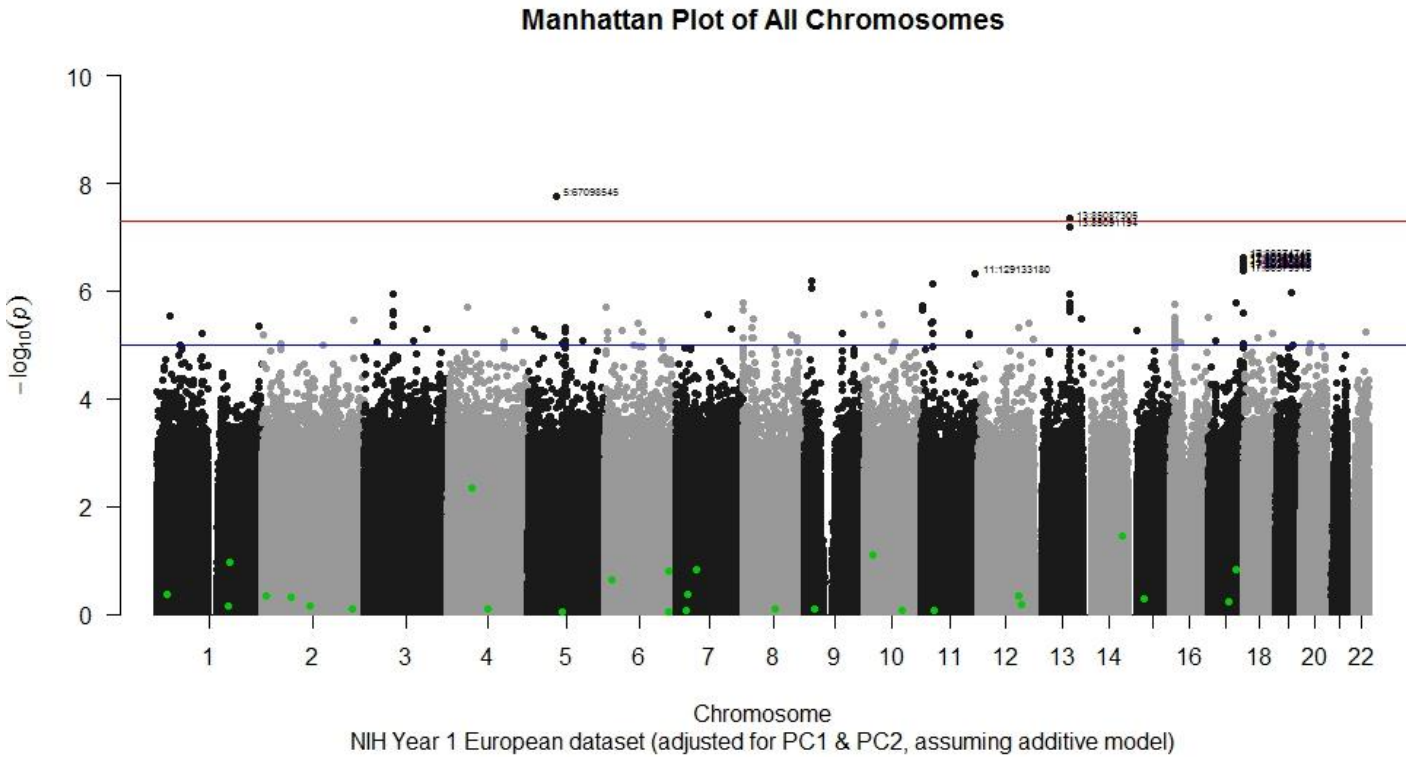
D. Supplementary figures



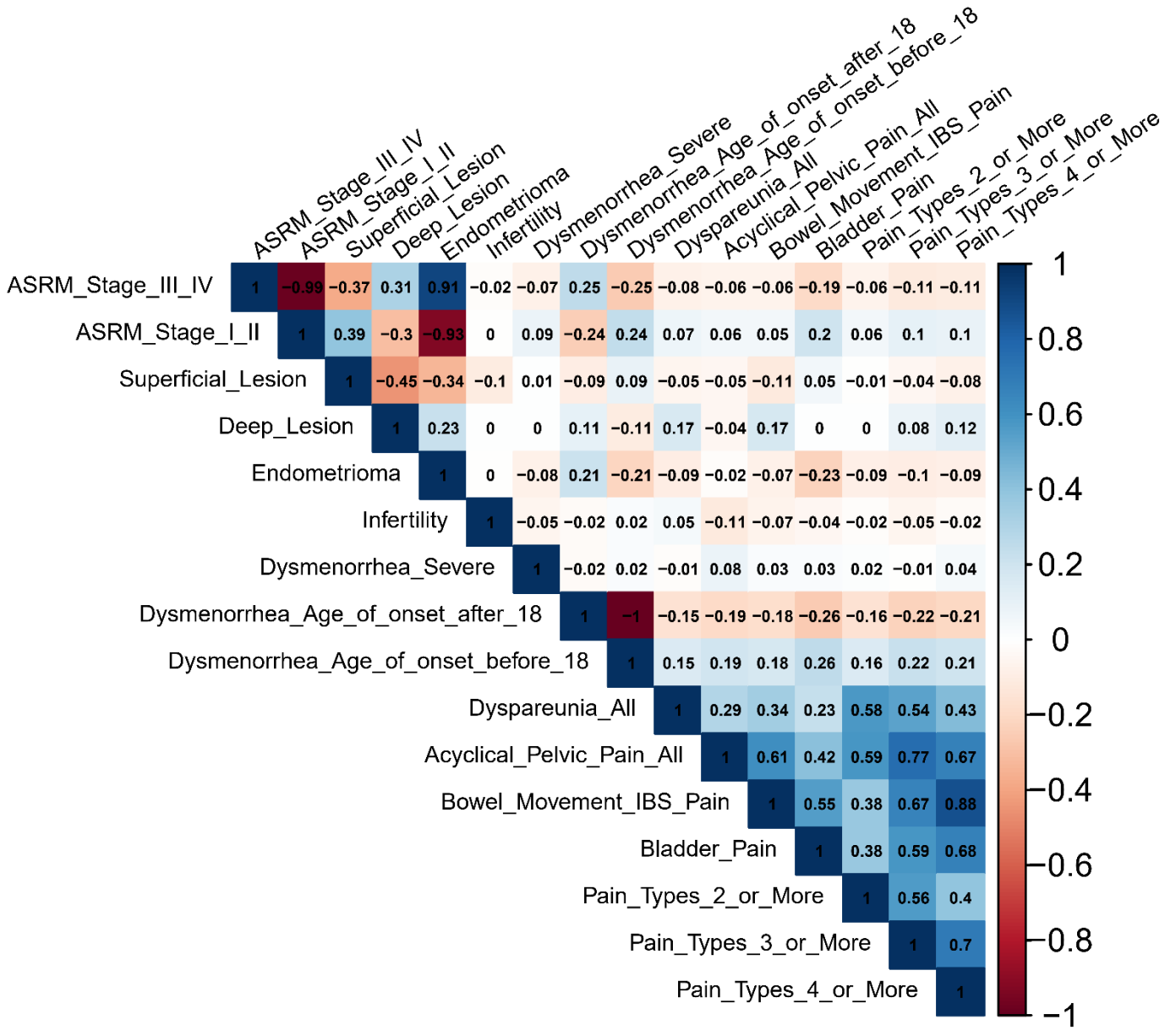
Supplementary Figure S 2.1 Flow chart of sample and variant QCs steps for genotype data (n=772) from discovery phase (year 1) of NIH epigenetic of endometriosis consortium.



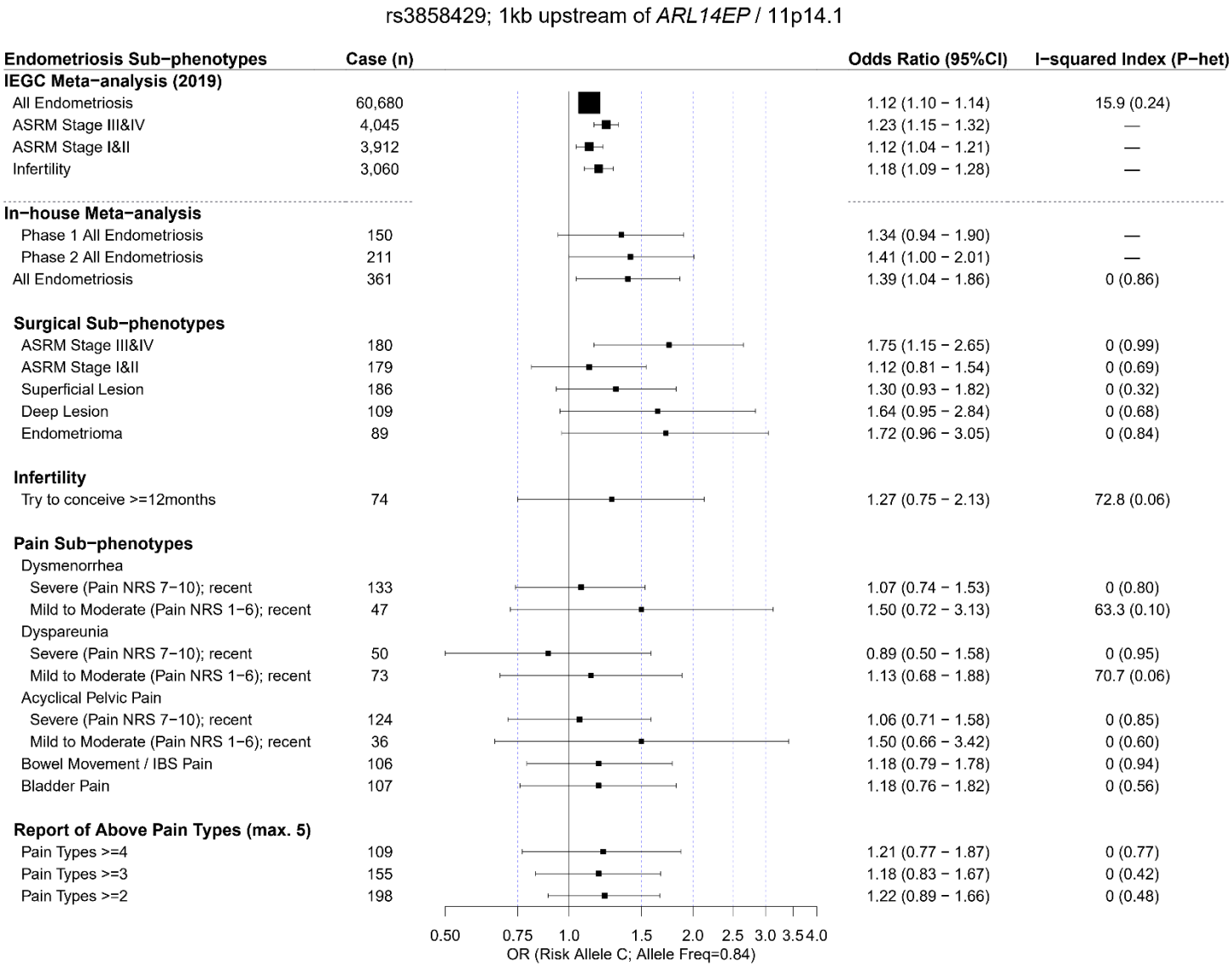
Supplementary Figure S 2.2 Q-Q plot of GWAS for endometriosis status of the NIH discovery dataset.



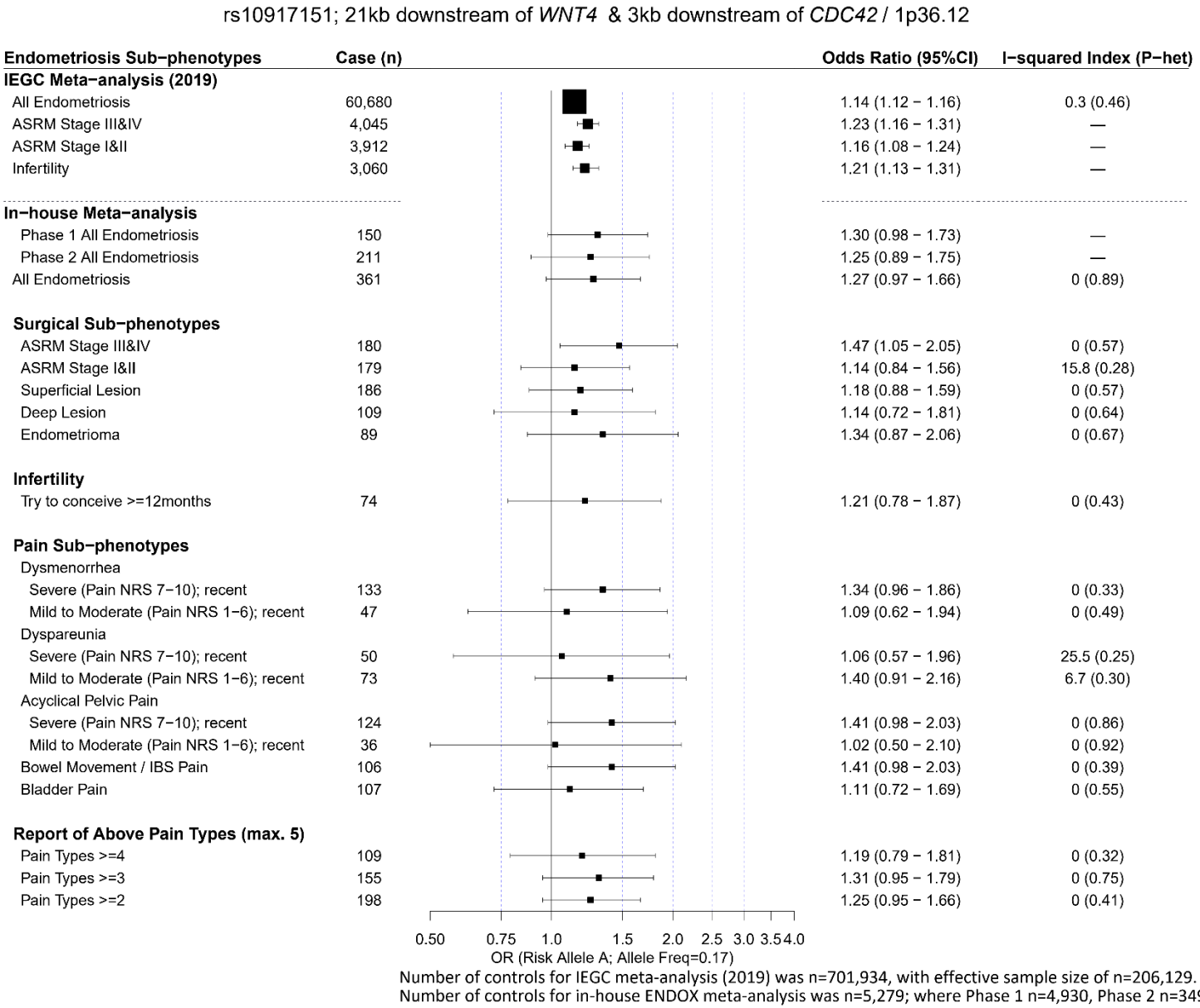
Supplementary Figure S 2.3 Manhattan plot of GWAS for endometriosis status of the NIH discovery dataset
Key: Green dots = previously reported 27 GWAS loci for endometriosis as at 2018[135]; red line = genome-wide significant threshold at $5x10^{-8}$, blue line = suggestive/nominal threshold at $1x10^{-5}$. Chromosome: position labelling = for loci with *p* value < $5x10^{-7}$.



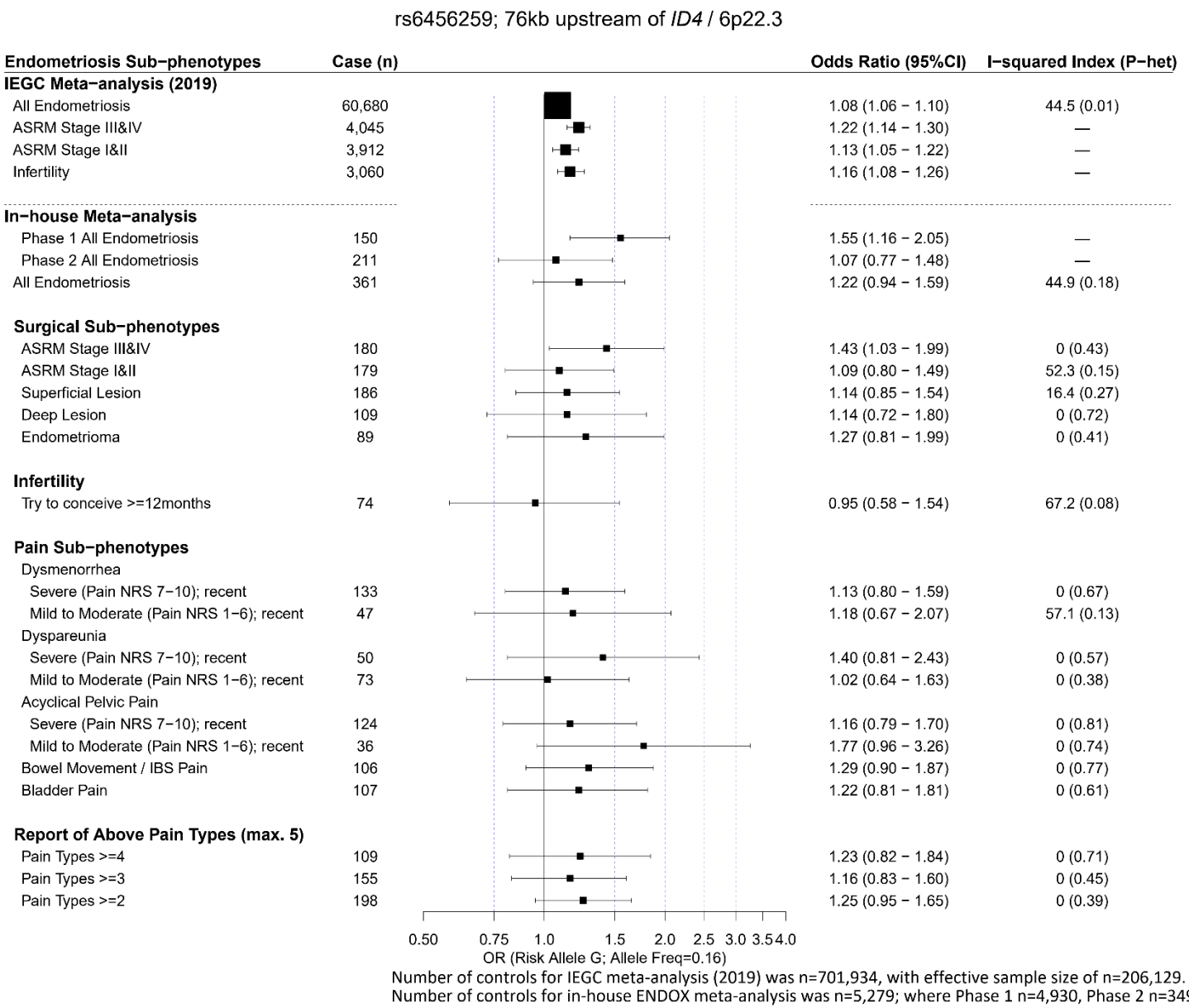
Supplementary Figure S 3.1 Correlation heat map of 16 major sub-phenotypes of endometriosis based on patients reporting age of onset for dysmenorrhea (n=162), coloured coded by Pearson's correlation coefficient. Key: shades of blue = positive correlations; shades of red = negative correlations.



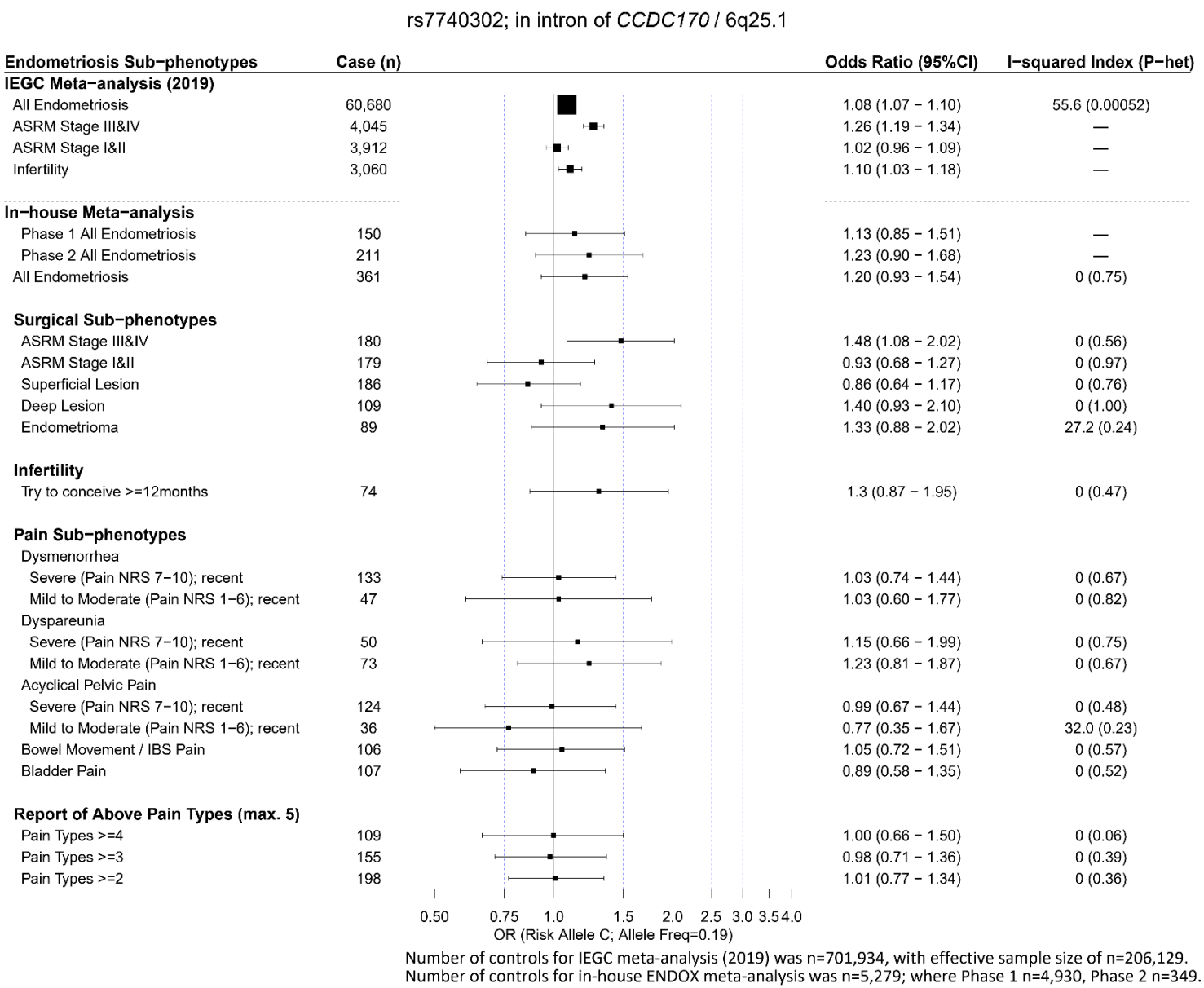
Supplementary Figure S 3.2 forest plot for effect estimates of the C allele of rs3858429 on 11p14.1, upstream of *ARL14EP*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.



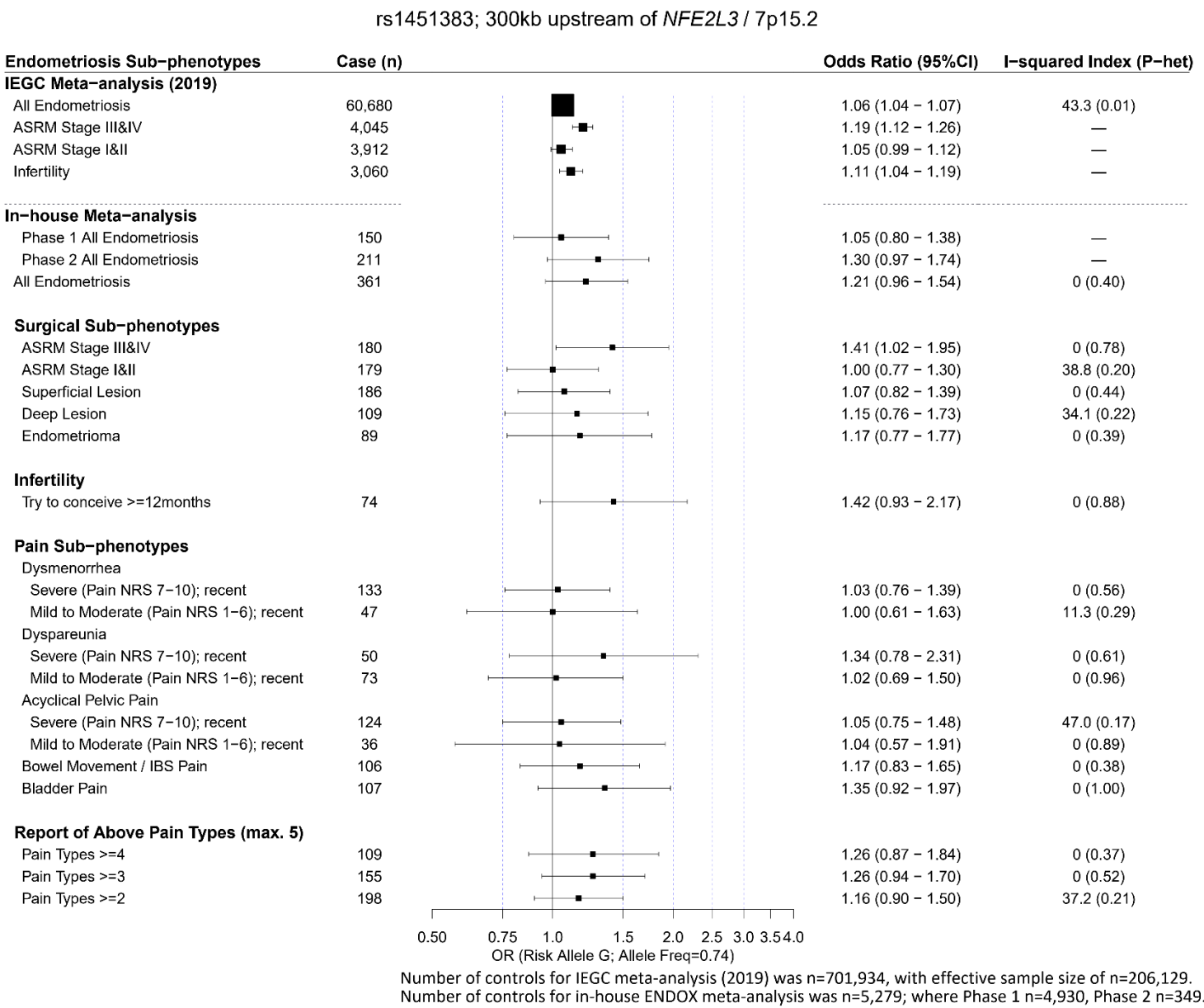
Supplementary Figure S 3.3 Forest plot for effect estimates of the A allele of rs10917151 on 1p36.12, downstream of *CDC42 & WNT4*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.



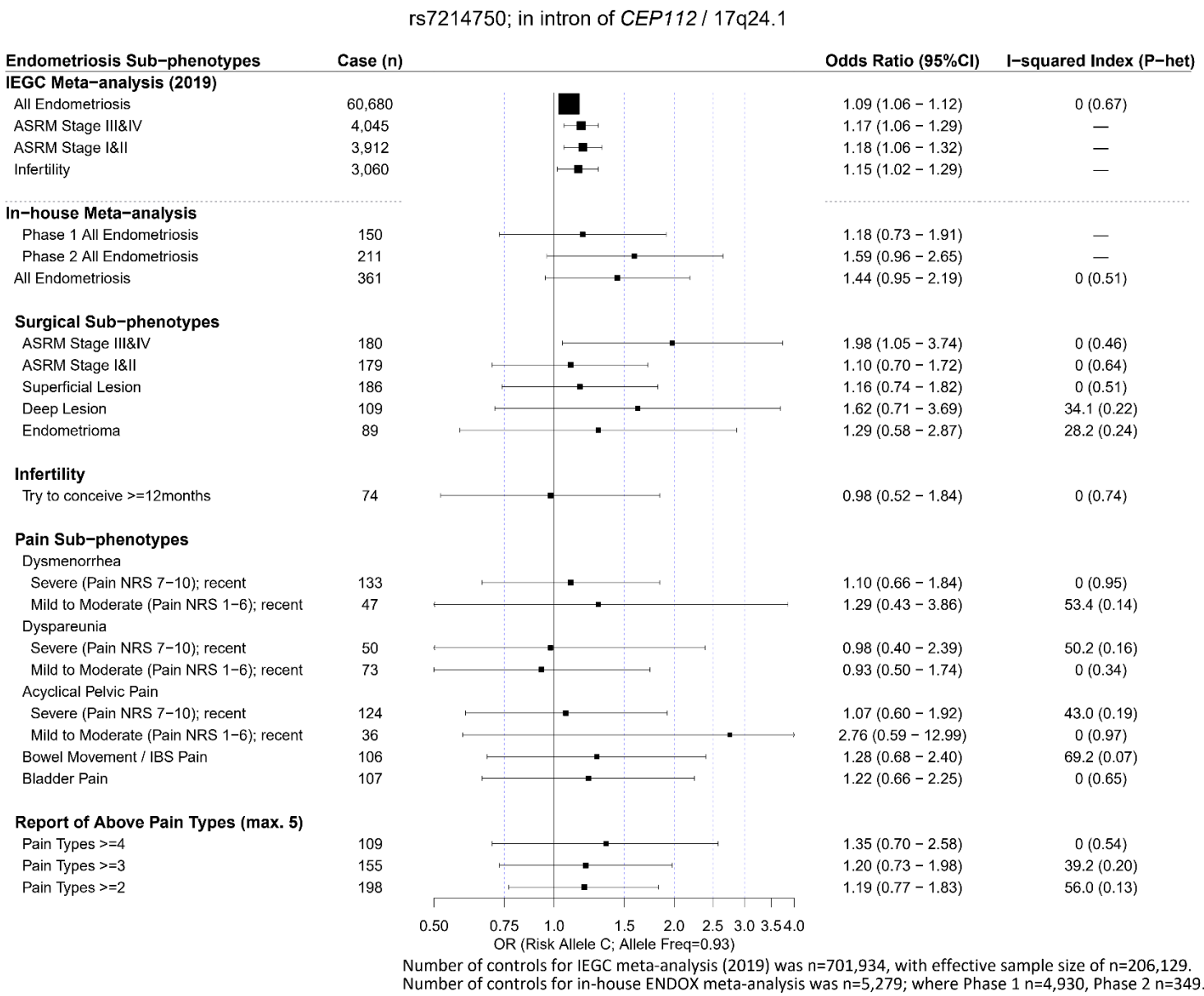
Supplementary Figure S 3.4 Forest plot for effect estimates of the G allele of rs6456259 on 6p22.3, upstream of *ID4*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.



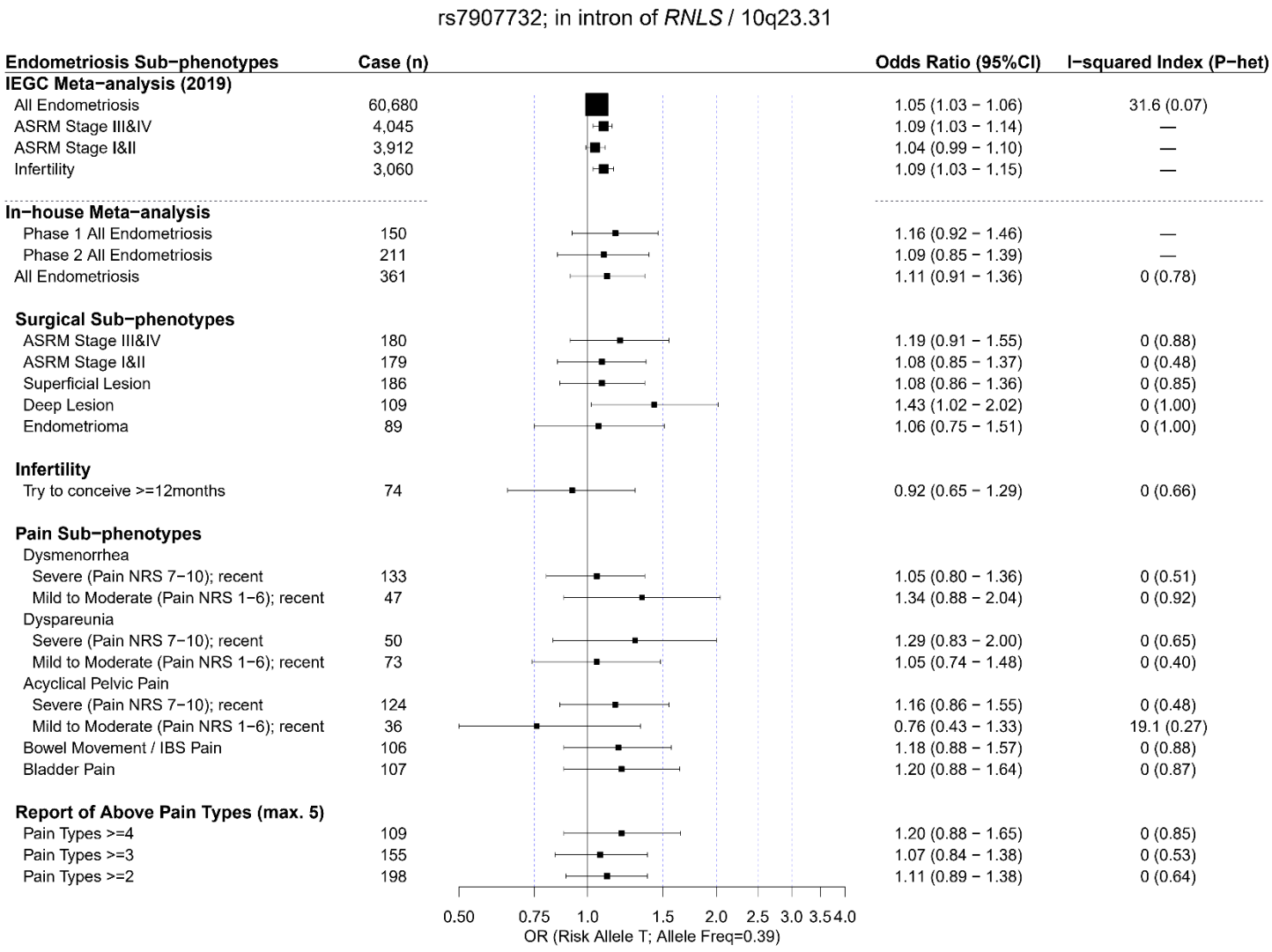
Supplementary Figure S 3.5 Forest plot for effect estimates of the C allele of rs7740302 on 6q25.1, in the intron of *CCDC170*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.



Supplementary Figure S 3.6 Forest plot for effect estimates of the G allele of rs1451383 on 7p15.2, upstream of *NFE2L3*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.

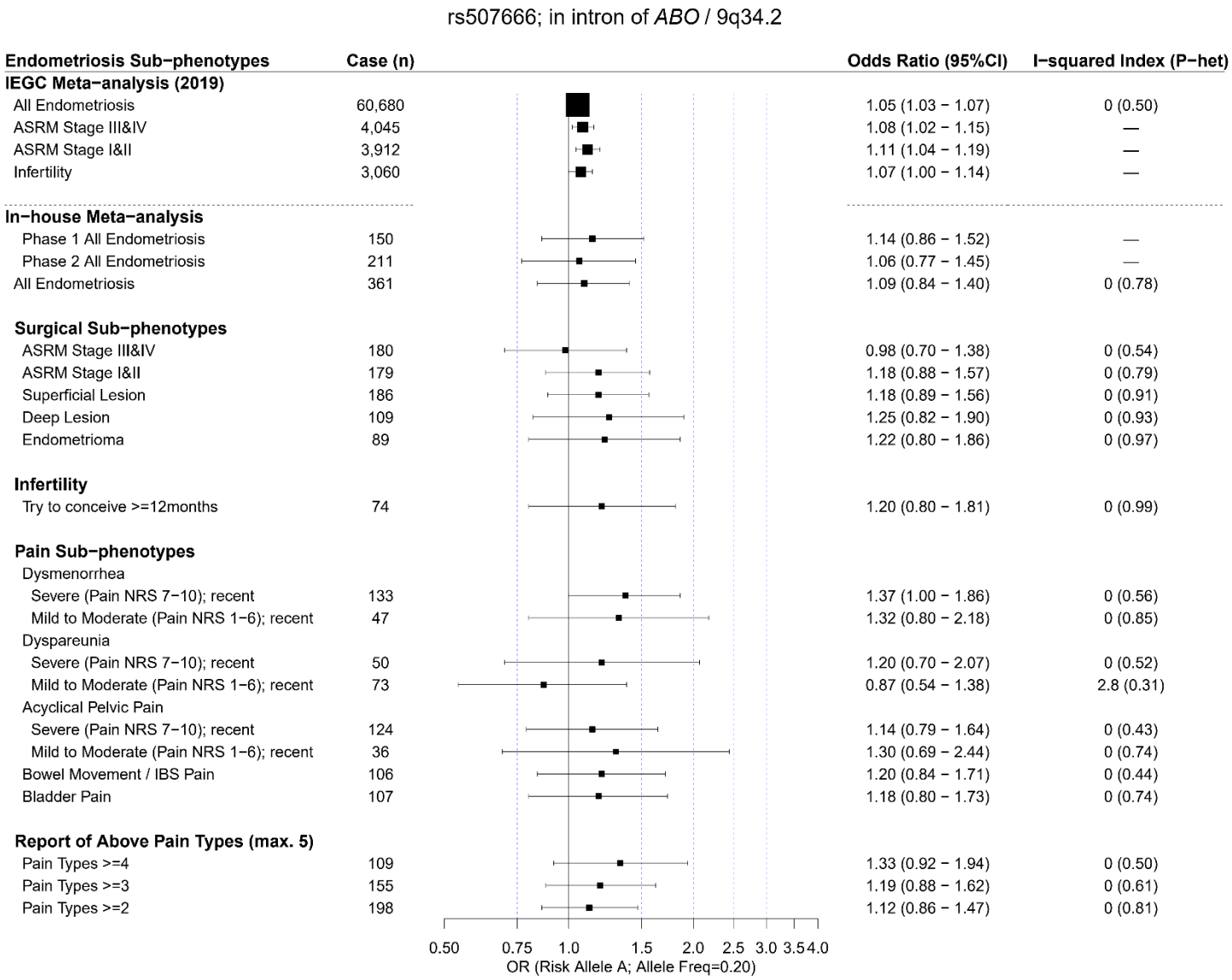


Supplementary Figure S 3.7 Forest plot for effect estimates of the C allele of rs7214750 on 17q24.1, in the intron of *CEP112*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.

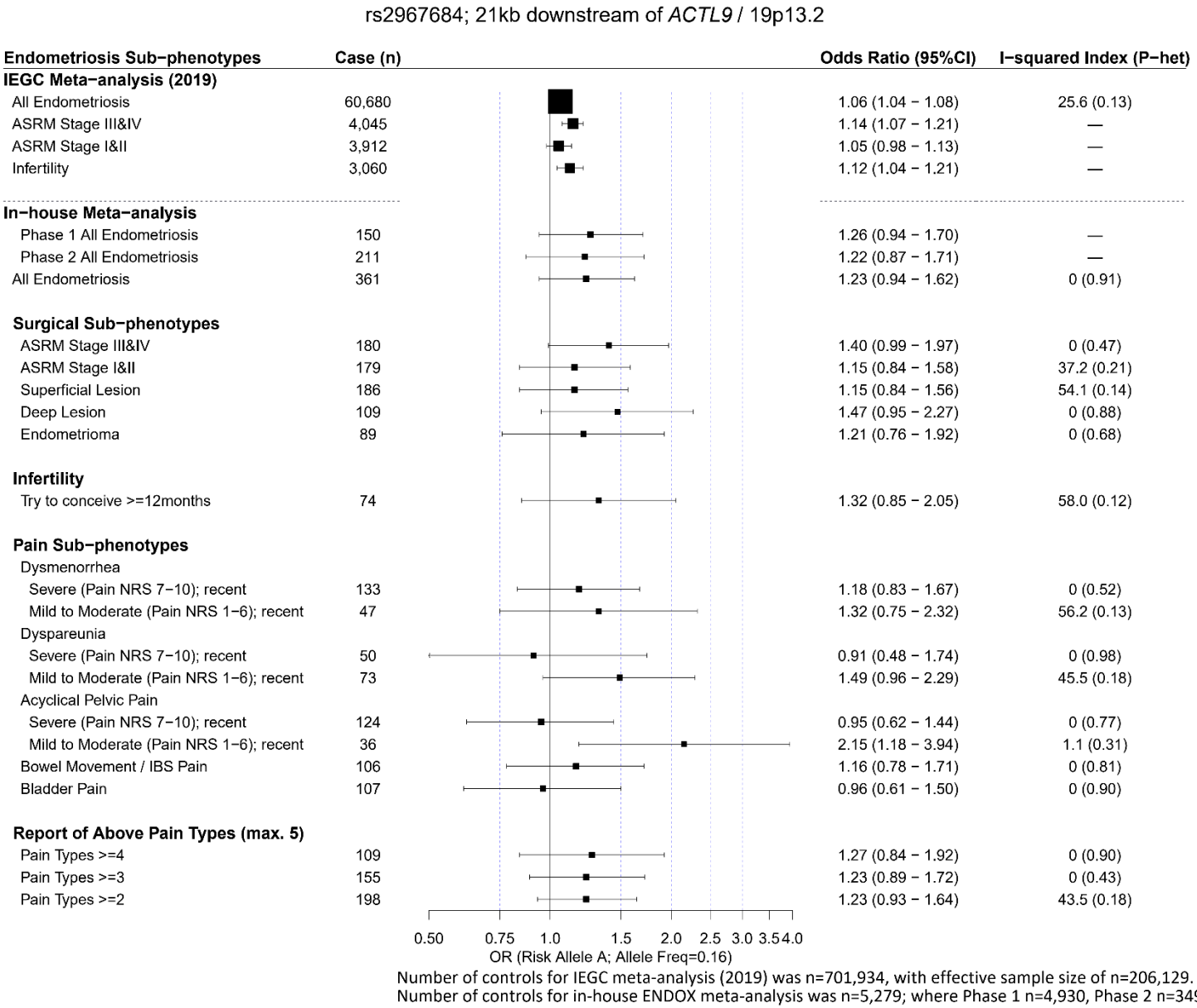


Number of controls for IEGC meta-analysis (2019) was n=701,934, with effective sample size of n=206,129.
Number of controls for in-house ENDOX meta-analysis was n=5,279; where Phase 1 n=4,930, Phase 2 n=349.

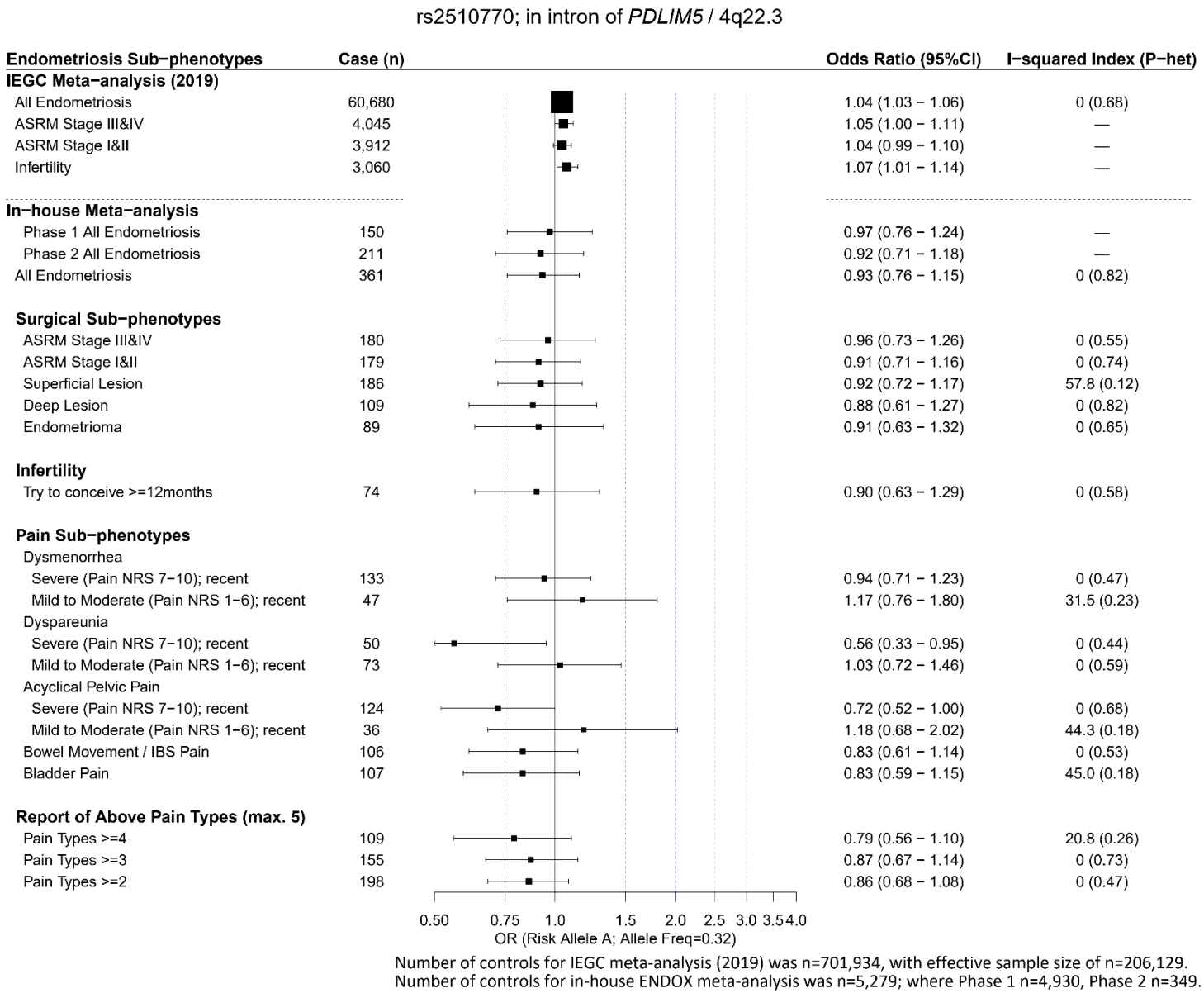
Supplementary Figure S 3.8 Forest plot for effect estimates of the T Allele of rs7907732 on 10q23.31, in the intron of *RNLS*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.



Supplementary Figure S 3.9 Forest plot for effect estimates of the A allele of rs507666 on 9q34.2, in the intron of *ABO*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.



Supplementary Figure S 3.10 Forest plot for effect estimates of the A allele of rs2967684 on 19p13.2, downstream of *ACTL9*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.



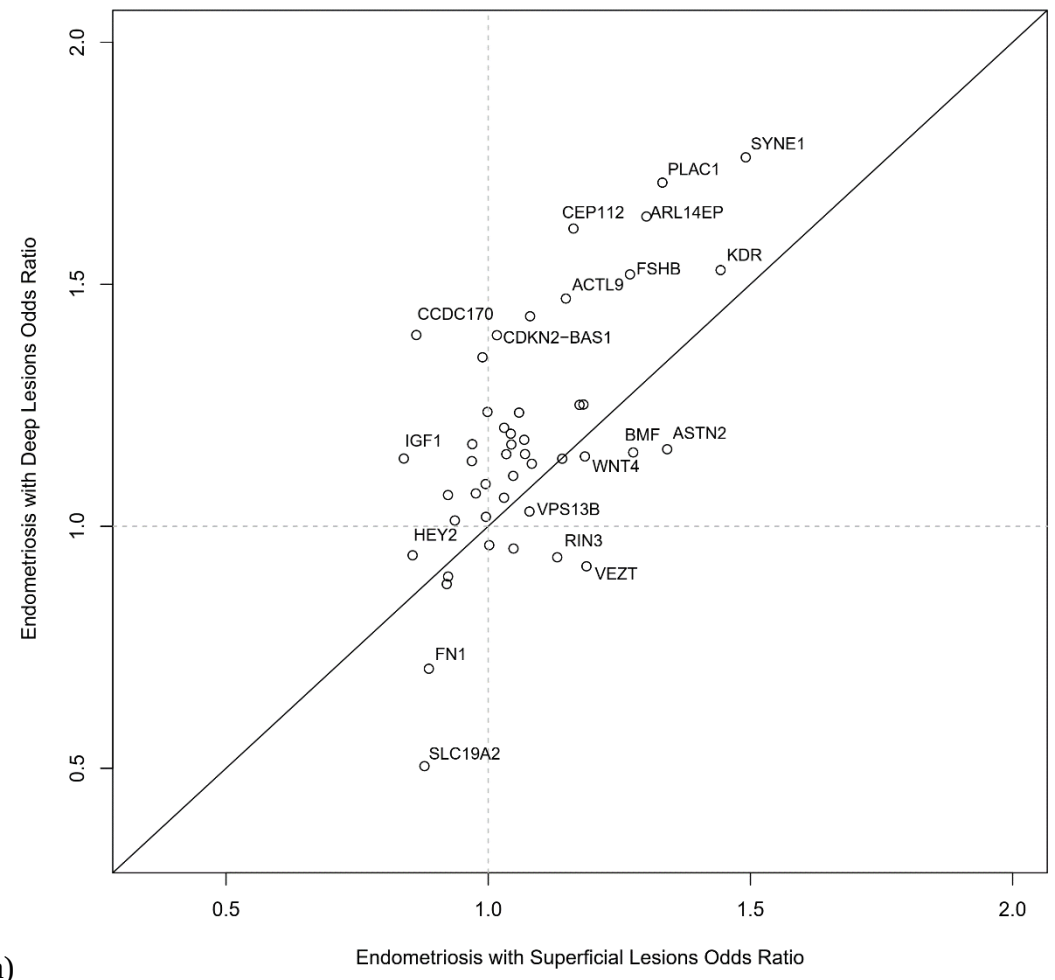
Supplementary Figure S 3.11 Forest plot for effect estimates of the A allele of rs2510770 on 4q22.3, in the intron of *PDLIM5*, for major surgical, infertility, and pain sub-phenotypes of endometriosis.

Number for Sub-phenotypes of Endometriosis		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	All Endometriosis	1.0																														
2	ASRM Stage III/IV	0.8	1.0																													
3	Stage I/II	0.7	0.3	1.0																												
4	Superficial Lesion	0.8	0.5	0.8	1.0																											
5	Deep Lesion	0.8	0.7	0.6	0.6	1.0																										
6	Endometrioma	0.7	0.7	0.3	0.7	0.6	1.0																									
7	Infertility	0.6	0.3	0.7	0.7	0.4	0.4	1.0																								
8	Subfertility	0.6	0.4	0.4	0.6	0.3	0.3	0.5	1.0																							
9	Dysmenorrhea (All)	0.8	0.6	0.7	0.8	0.7	0.8	0.7	0.6	1.0																						
10	Dysmenorrhea (Moderate to Severe; recent)	0.7	0.4	0.7	0.8	0.6	0.6	0.6	0.5	0.9	1.0																					
11	Dysmenorrhea (Severe; recent)	0.7	0.4	0.7	0.7	0.5	0.6	0.6	0.4	0.8	0.9	1.0																				
12	Dysmenorrhea (Mild to Moderate; recent)	0.7	0.4	0.6	0.8	0.6	0.5	0.5	0.5	0.7	0.7	0.5	1.0																			
13	Dysmenorrhea (Age of onset age ≥18)	0.6	0.3	0.6	0.7	0.4	0.6	0.6	0.5	0.8	0.7	0.7	0.6	1.0																		
14	Dysmenorrhea (Age of onset age <18)	0.6	0.6	0.3	0.5	0.4	0.5	0.3	0.5	0.7	0.6	0.5	0.5	0.3	1.0																	
15	Dyspareunia (All)	0.8	0.5	0.7	0.7	0.6	0.6	0.6	0.4	0.8	0.8	0.7	0.6	0.7	0.4	1.0																
16	Dyspareunia (Moderate to Severe; recent)	0.5	0.3	0.5	0.5	0.4	0.3	0.4	0.3	0.5	0.6	0.6	0.4	0.4	0.3	0.8	1.0															
17	Dyspareunia (Severe; recent)	0.5	0.3	0.6	0.5	0.5	0.2	0.5	0.2	0.5	0.5	0.4	0.4	0.4	0.3	0.6	0.8	1.0														
18	Dyspareunia (Mild to Moderate; recent)	0.4	0.3	0.3	0.3	0.3	0.5	0.2	0.1	0.4	0.6	0.6	0.3	0.4	0.3	0.6	0.4	-0.1	1.0													
19	Dyspareunia (Severe; lifetime)	0.4	0.1	0.5	0.5	0.3	0.2	0.5	0.5	0.4	0.3	0.3	0.2	0.4	0.1	0.4	0.4	0.5	0.0	1.0												
20	Dyspareunia (Mild to Moderate; lifetime)	0.7	0.6	0.4	0.5	0.5	0.5	0.3	0.3	0.7	0.6	0.5	0.5	0.5	0.6	0.6	0.2	0.2	0.4	0.0	1.0											
21	Acyclical Pelvic Pain (All)	0.8	0.6	0.7	0.8	0.6	0.6	0.5	0.4	0.8	0.9	0.8	0.6	0.6	0.6	0.8	0.6	0.6	0.4	0.4	0.7	1.0										
22	Acyclical Pelvic Pain (Moderate to Severe; recent)	0.7	0.5	0.7	0.6	0.5	0.5	0.4	0.4	0.8	0.8	0.8	0.5	0.5	0.5	0.8	0.6	0.6	0.4	0.4	0.7	1.0	1.0									
23	Acyclical Pelvic Pain (Severe; recent)	0.5	0.2	0.6	0.5	0.4	0.4	0.4	0.1	0.5	0.6	0.7	0.2	0.4	0.2	0.6	0.6	0.6	0.3	0.3	0.3	0.8	0.8	1.0								
24	Acyclical Pelvic Pain (Mild to Moderate; recent)	0.5	0.6	0.2	0.5	0.3	0.5	0.2	0.5	0.6	0.6	0.4	0.5	0.4	0.6	0.5	0.1	0.1	0.4	0.0	0.7	0.6	0.5	0.0	1.0							
25	Acyclical Pelvic Pain (Severe; lifetime)	0.6	0.4	0.5	0.6	0.4	0.5	0.3	0.4	0.7	0.7	0.7	0.4	0.5	0.6	0.7	0.5	0.5	0.4	0.4	0.6	0.8	0.9	0.7	0.5	1.0						
26	Acyclical Pelvic Pain (Mild to Moderate; lifetime)	0.3	0.3	0.2	0.5	0.3	0.5	0.4	0.1	0.4	0.4	0.4	0.4	0.4	0.3	0.3	0.1	0.2	0.3	0.2	0.3	0.3	0.2	0.1	0.4	0.1	1.0					
27	Bowel Movement/IBS Pain	0.6	0.5	0.5	0.5	0.5	0.5	0.3	0.2	0.6	0.6	0.6	0.4	0.5	0.4	0.7	0.6	0.6	0.3	0.2	0.6	0.8	0.8	0.7	0.3	0.8	0.2	1.0				
28	Bladder Pain	0.6	0.5	0.4	0.6	0.5	0.6	0.3	0.2	0.6	0.5	0.6	0.4	0.6	0.4	0.7	0.4	0.5	0.2	0.2	0.6	0.7	0.7	0.6	0.3	0.7	0.3	0.8	1.0			
29	Pain Types ≥ 2	0.8	0.6	0.7	0.8	0.6	0.6	0.7	0.5	0.9	0.9	0.8	0.7	0.7	0.6	0.9	0.6	0.6	0.5	0.4	0.7	0.9	0.9	0.7	0.6	0.7	0.4	0.7	0.7	1.0		
30	Pain Types ≥ 3	0.8	0.5	0.7	0.8	0.6	0.6	0.6	0.4	0.8	0.8	0.8	0.6	0.7	0.5	0.9	0.7	0.6	0.4	0.4	0.7	0.9	0.9	0.8	0.5	0.8	0.3	0.8	0.8	0.9	1.0	
31	Pain Types ≥ 4	0.6	0.5	0.5	0.6	0.5	0.5	0.3	0.3	0.7	0.6	0.6	0.5	0.6	0.5	0.8	0.5	0.5	0.3	0.3	0.7	0.8	0.9	0.7	0.5	0.8	0.3	0.9	0.9	0.8	0.8	1.0

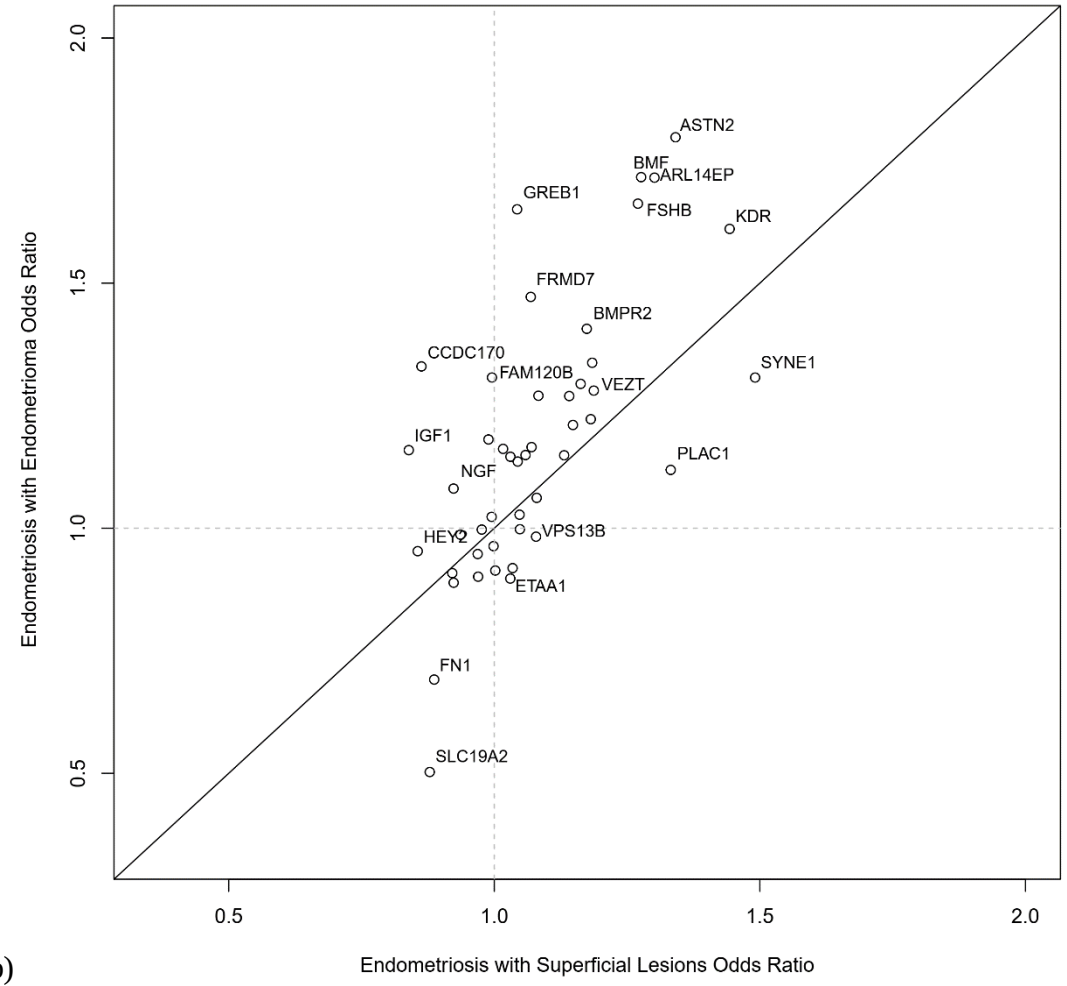
Key:

Pearson's r^2	≥0.9
	0.7-0.8
	0.5-0.6
	<0.5
	0.0

Supplementary Figure S 3.12 Correlation coefficient (Pearson's r^2) of effect estimates of tested genetic variants (n=46) on sub-phenotypic pairs for endometriosis

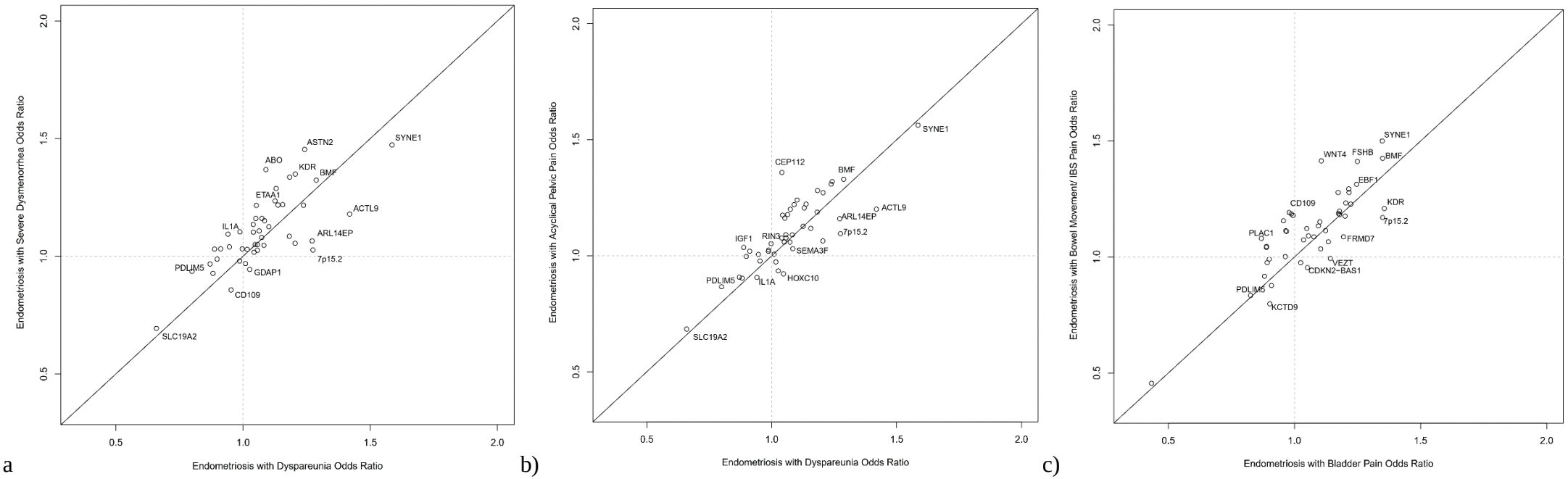


a)



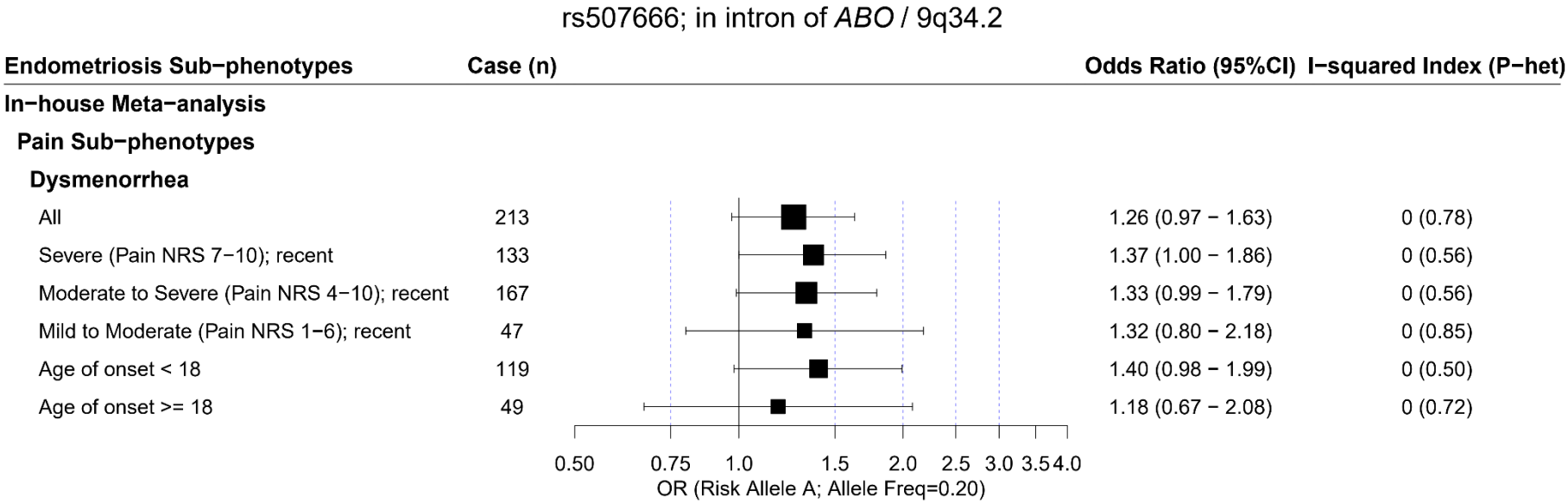
b)

Supplementary Figure S 3.13 Correlation plots of effect estimates of 46 genetic variants for a) deep lesion (Pearson's r^2= 0.64) and b) endometrioma (r^2= 0.68) versus superficial lesion sub-phenotype of endometriosis.

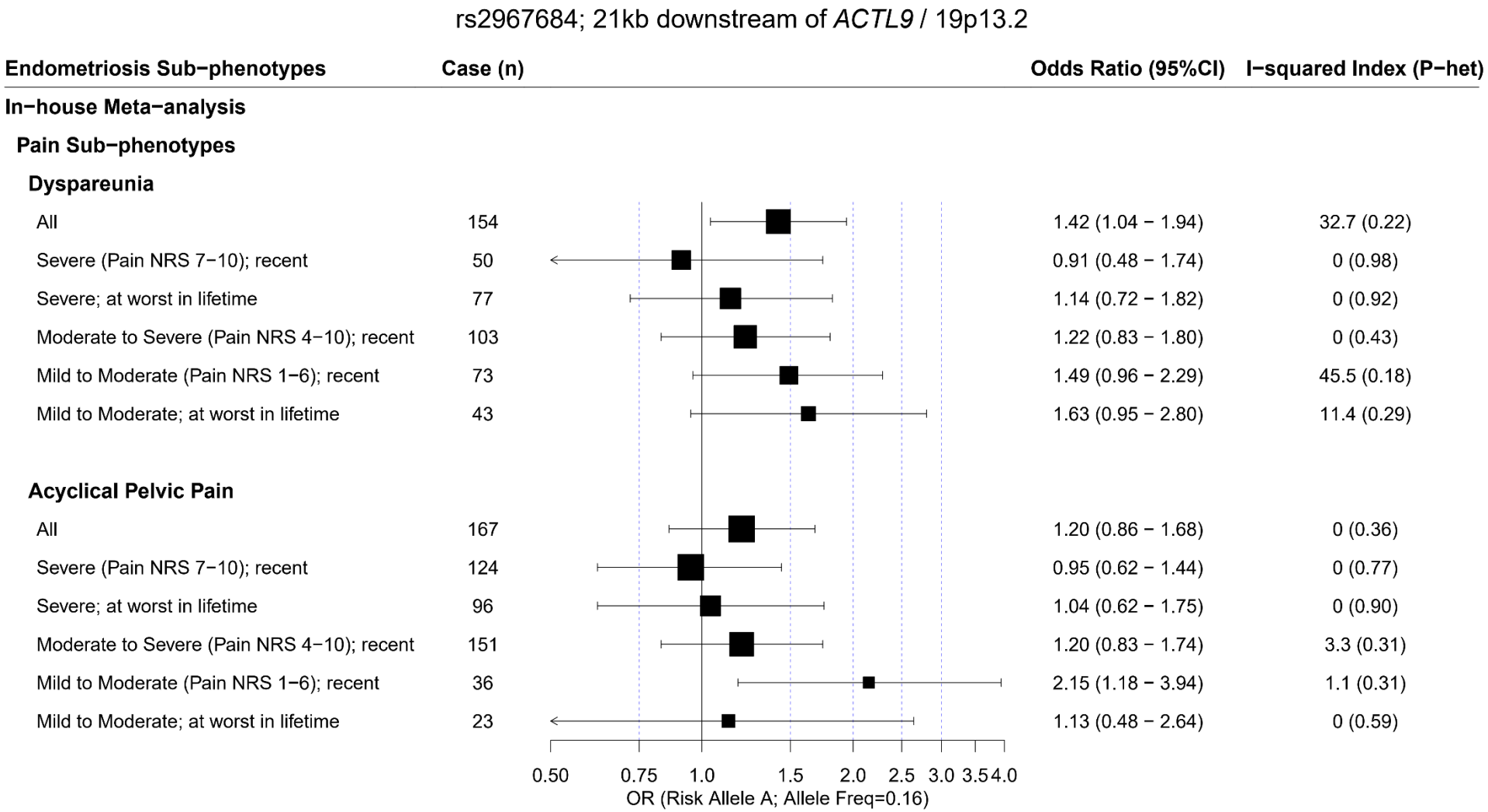


a b) c)

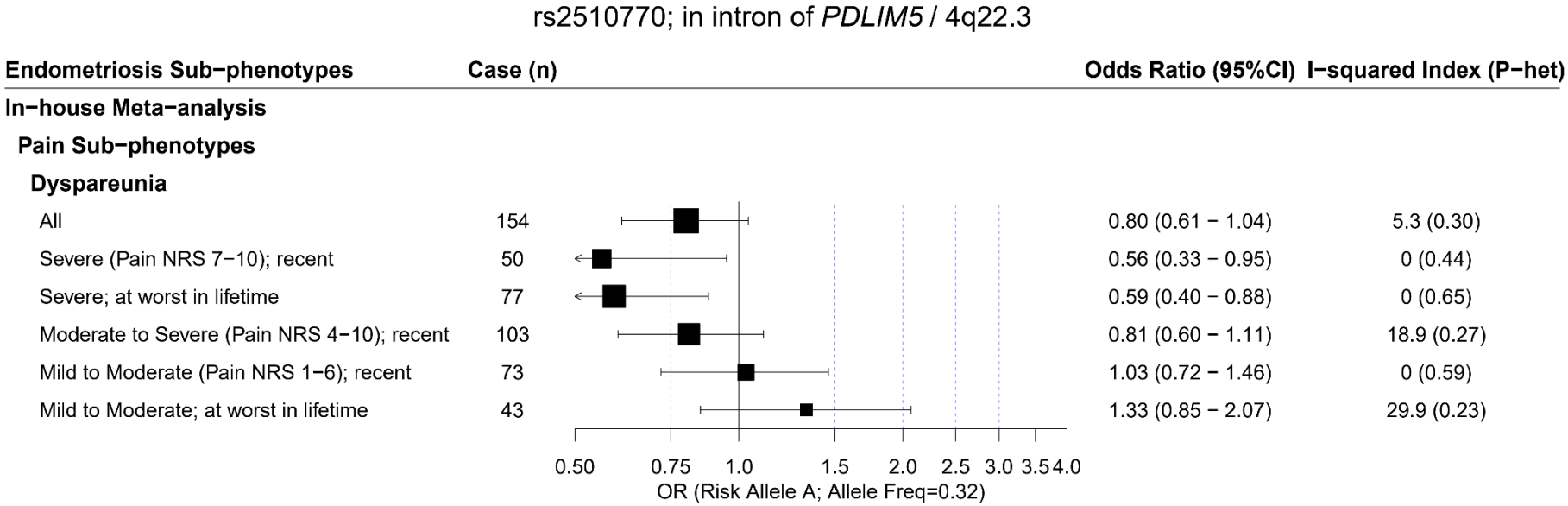
Supplementary Figure S 3.14 Correlation plots of effect estimates of 46 genetic variants for a) severe dysmenorrhea (Pearson's r^2= 0.74), b) acyclical pelvic pain (r^2= 0.82) versus dyspareunia, and c) bowel movement/IBS pain versus bladder pain sub-phenotype (r^2= 0.82) of endometriosis.



Supplementary Figure S 3.15 Forest plot for effect estimates of the A allele of rs507666 on 9q34.2, in the intron of *ABO*, for sensitivity testing of endometriosis with dysmenorrhea sub-phenotype.

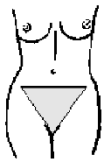


Supplementary Figure S 3.16 Forest plot for effect estimates of the A allele of rs2967684 on 19p13.2, downstream of *ACTL9*, for sensitivity testing of dyspareunia and acyclical pelvic pain sub-phenotypes of endometriosis.



Supplementary Figure S 3.17 Forest plot for effect estimates of the A allele of rs2510770 on 4q22.3, in the intron of *PDLIM5*, for sensitivity testing of endometriosis with dyspareunia sub-phenotype.

The following questions are about pelvic pain during or after vaginal intercourse or penetration.



By 'pelvic pain' we mean any type of pain (cramping, shooting, stabbing, etc.) in the lower part of your belly, as shown by the shaded area in this picture:

We remind you that any information you give will be treated in complete confidence.
*If however you **do not wish to answer these questions**, please check here ___ and skip to question **C28***

*If you have **never had intercourse**, please check here ___ and skip to question **C28***

C16. Have you ever had pelvic pain during intercourse or in the 24 hours following vaginal sexual intercourse/penetration**?**

☐ No → **Please skip to question C29**

☐ Yes → If yes: C16.1. At what age did this pain start? _____

C17. When did you last have vaginal intercourse?

☐ In the last month

☐ 1-3 months ago

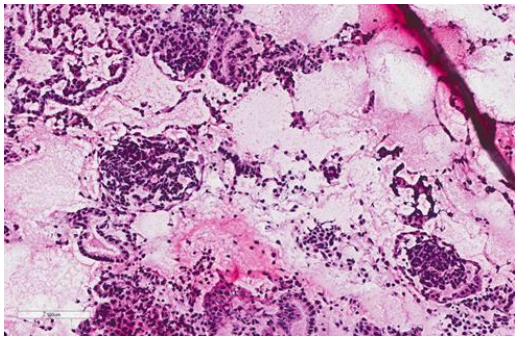
☐ 4-12 months ago

☐ More than 12 months ago → If so, did you avoid intercourse because of pelvic pain? ☐ No ☐ Yes

If you had vaginal intercourse more than 12 months ago, please skip to question C26

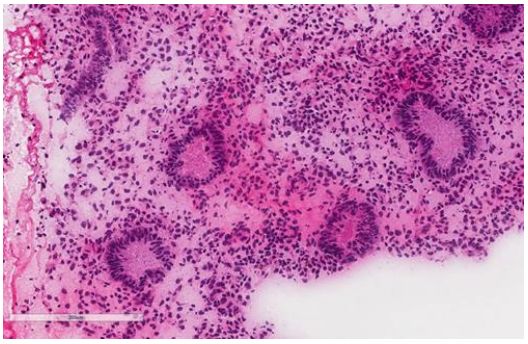
Supplementary Figure S 3.18 WERF-EPHect clinical questionnaire (standard) (EPQ-S v1.0) Question C17 preceding capturing of the dyspareunia pain NRS for recent (last) time.

From http://endometriosisfoundation.org/WERF-EPHect-Clinical-Questionnaire-Standard-2014v1.pdf, last accessed on 07 May 2020.



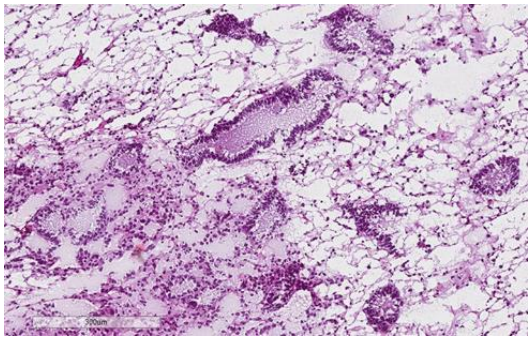
a. Menstrual Phase

blue balls



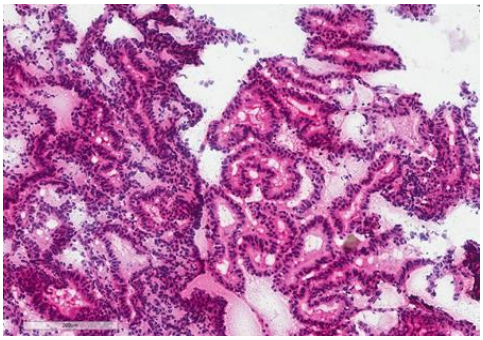
b. Proliferative Phase

blue donuts with mitotic bodies



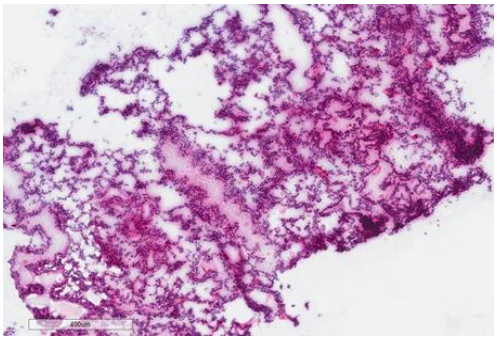
c. Early-secretory Phase

sub-nuclear vacuoles



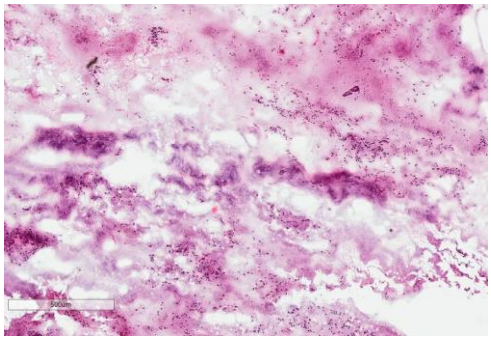
d. Mid-secretory Phase

lip-shaped & coiling-up glands



e. Late-secretory Phase

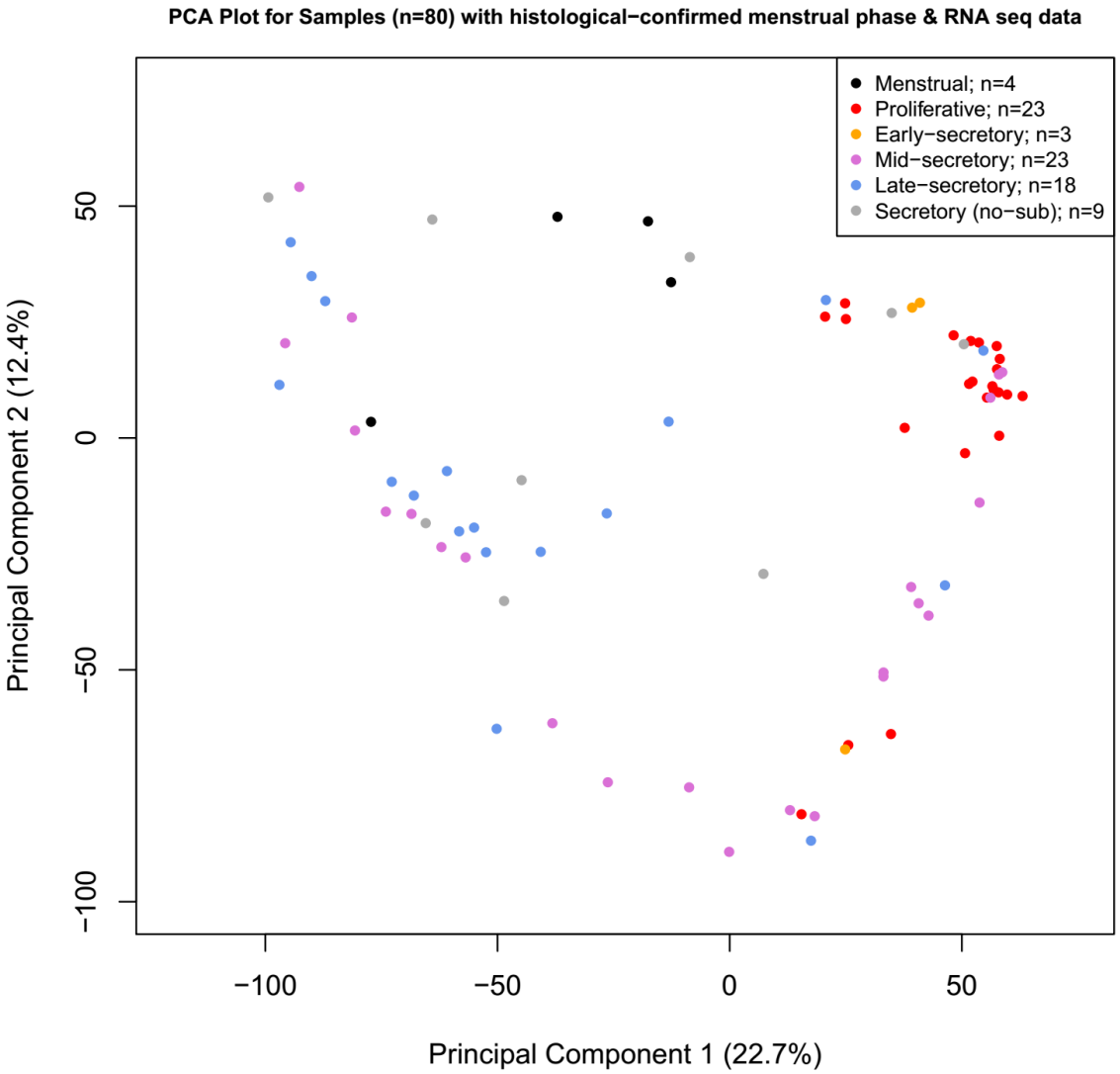
'tree-like' glands



f. Mucus only

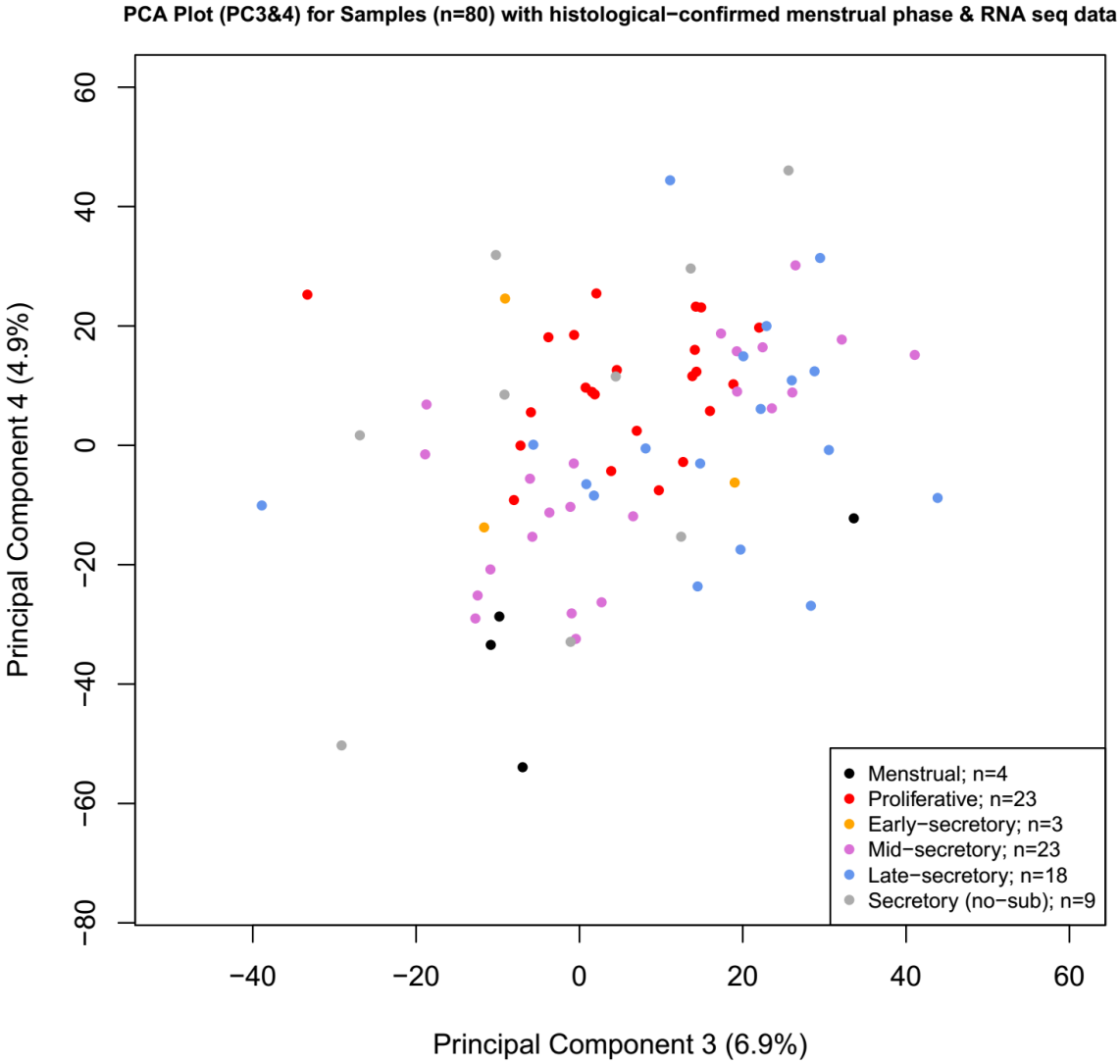
no gland is shown

Supplementary Figure S 4.1 Characteristics of endometrial glands in a. menstrual, b. proliferative, c. early-secretory, d. mid-secretory, e. late-secretory phase & f. mucus only (200X; scale bar=100μm).



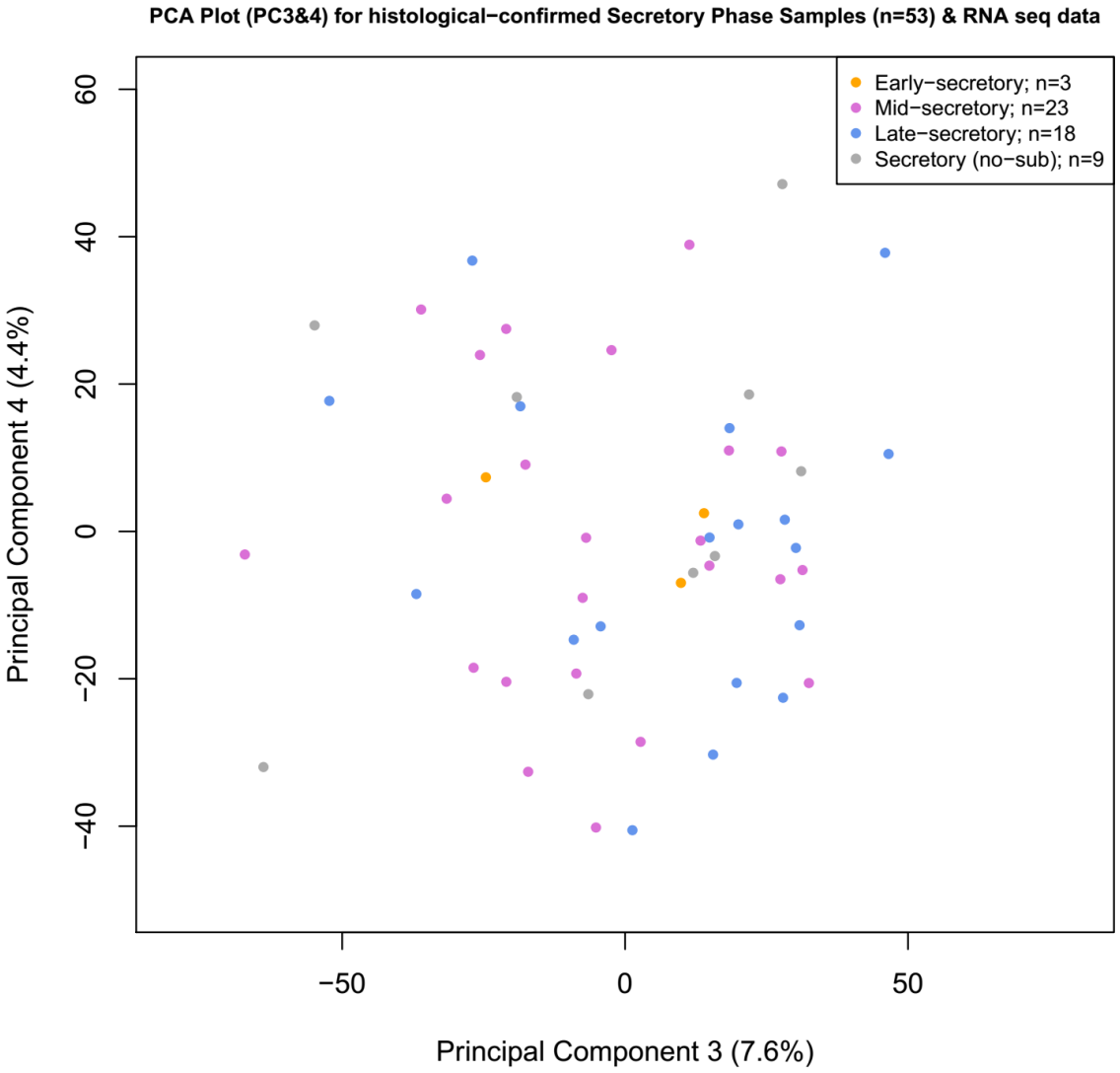
Supplementary Figure S 4.2 PCA plot for samples (n=80) with histological-confirmed five cycle phases, with PC 1 (22.7%) and PC 2 (12.4%).

Key: black dots = menstrual, red dots = proliferative, orange dots = early-secretory, purple dots = mid-secretory, blue dots = late-secretory, and grey dots = no sub-secretory classification.



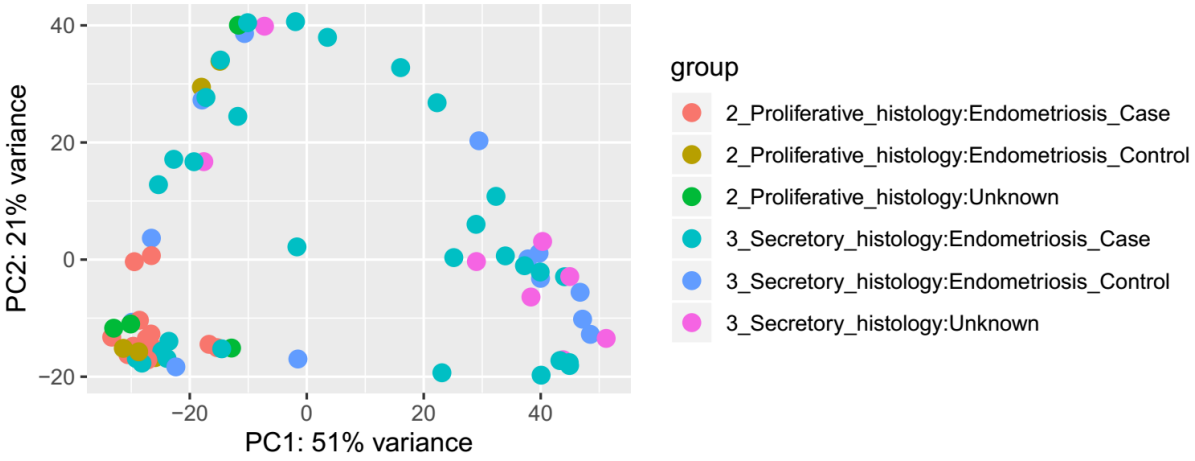
Supplementary Figure S 4.3 PCA plot for samples (n=80) with histological-confirmed five cycle phases, with PC 3 (6.9%) and PC 4 (4.9%).

Key: black dots = menstrual, red dots = proliferative, orange dots = early-secretory, purple dots = mid-secretory, blue dots = late-secretory, and grey dots = no sub-secretory classification.

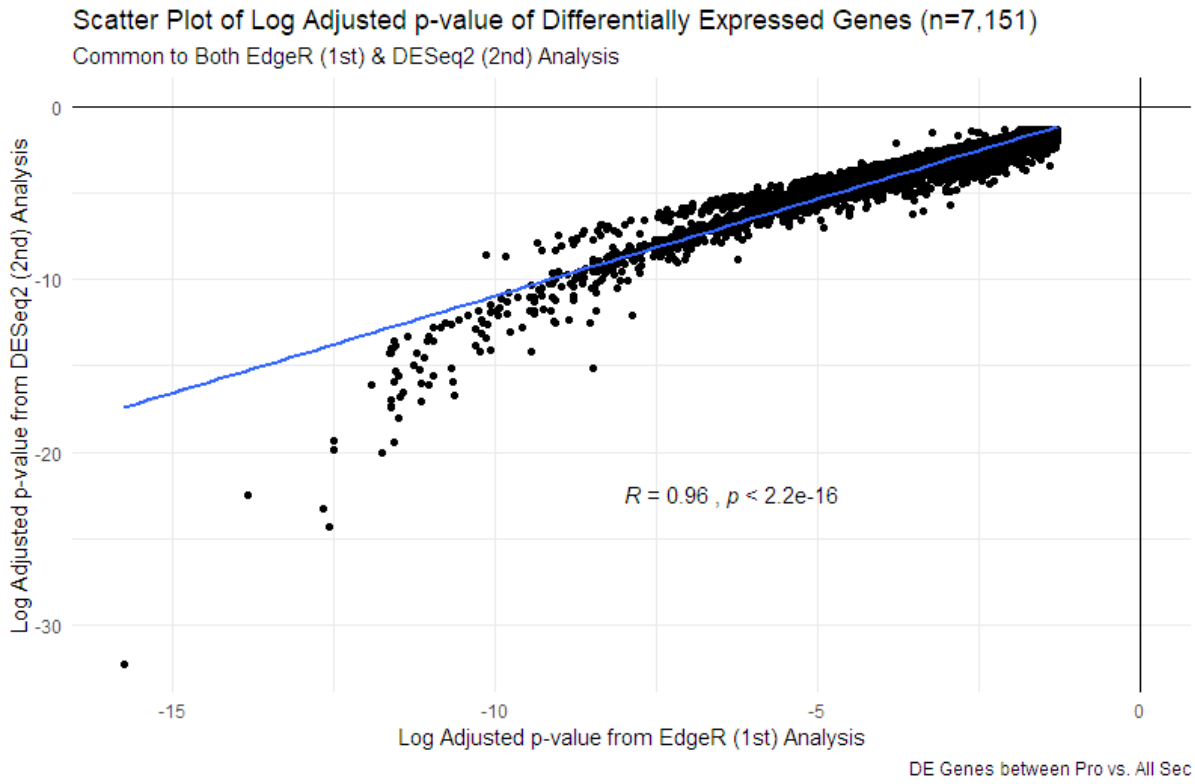


Supplementary Figure S 4.4 PCA plot for secretory samples only (n=53), with PC 3 (7.6%) and PC 4 (4.4%).

Key: orange dots = early-secretory, purple dots = mid-secretory, blue dots = late-secretory, and grey dots = no sub-secretory classification.

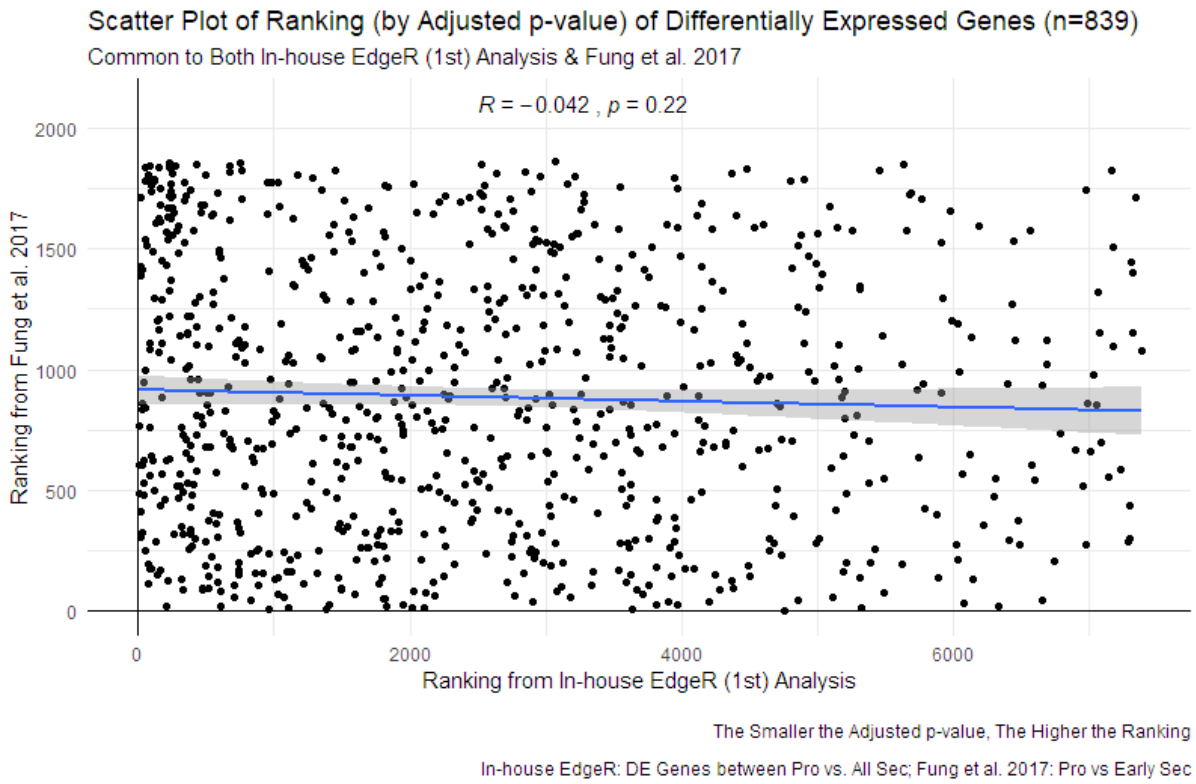


Supplementary Figure S 4.5 PCA plot for showing interaction between two terms, the histology cycle phases and the endometriosis case-control status, in the design matrix (`design ~ histology phase + case-control status` by `DESeq2`).



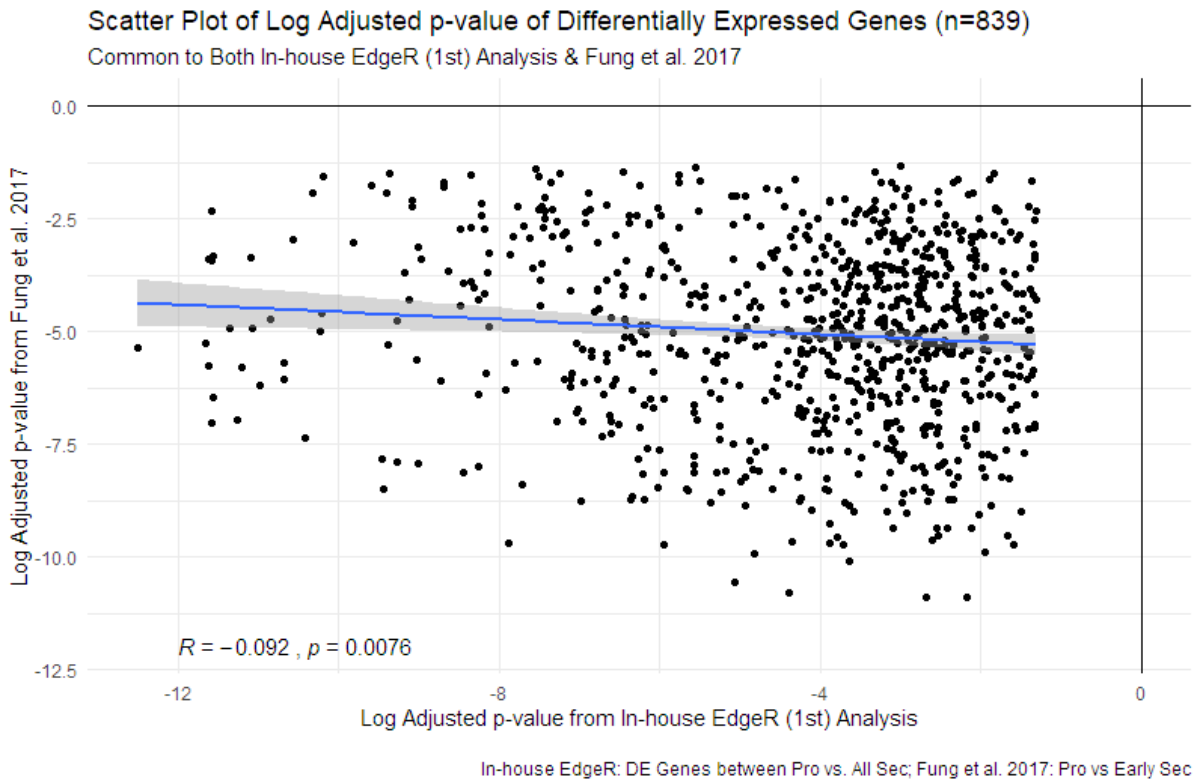
Supplementary Figure S 4.6 Scatter plot of log-scale adjusted *p*-value of differentially expressed genes between proliferative and secretory phases common to `edgeR` and `DESeq2` analyses.

Key: blue line = regression line by linear model.



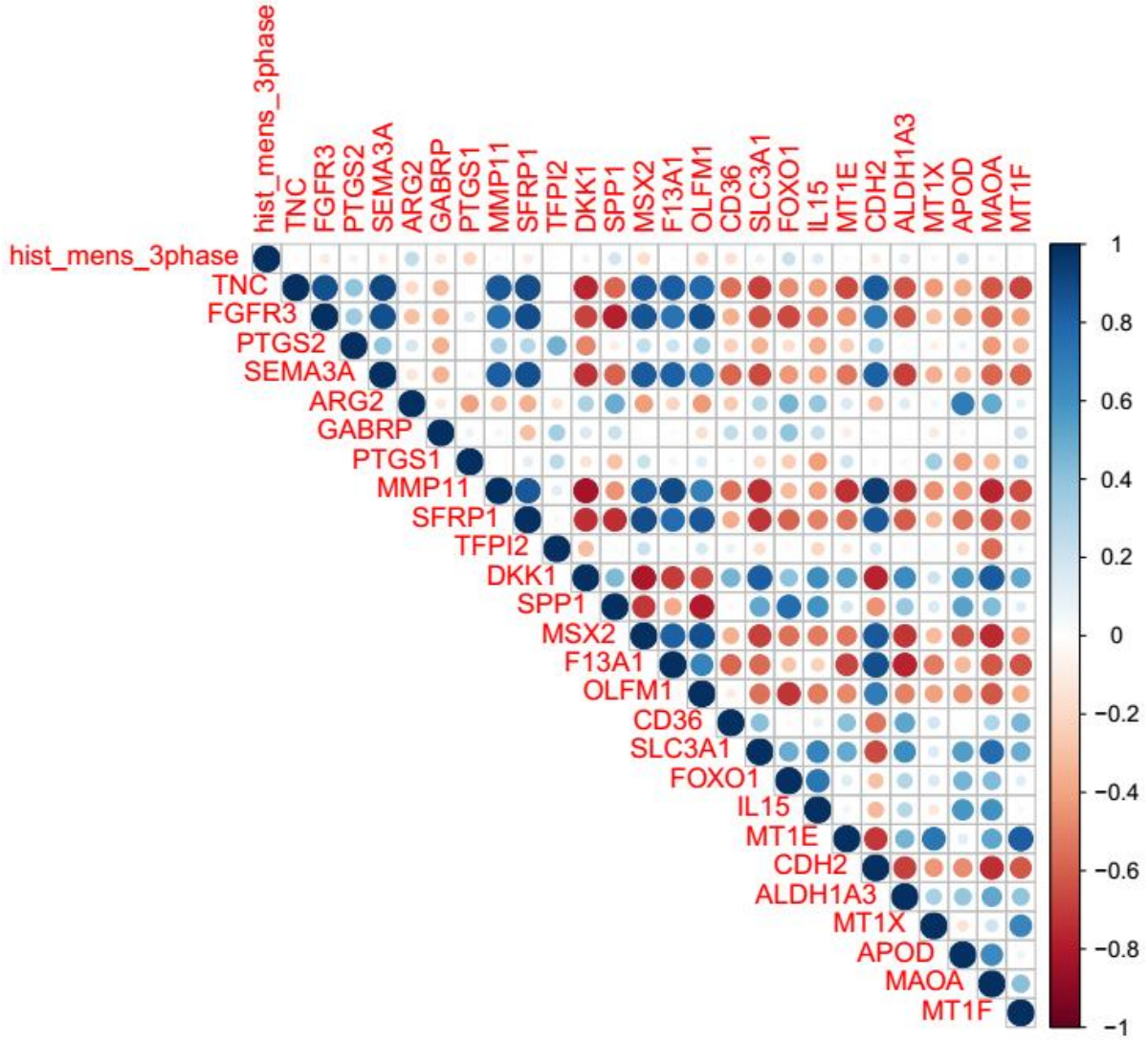
Supplementary Figure S 4.7 Scatter plot of ranking by adjusted *p*-value of differentially expressed genes common to in-house `edgeR` (proliferative vs. all secretory) and Fung et al. (proliferative vs. early-secretory) analyses.

Key: blue line = regression line by linear model, blue shading = 95% CI of regression line; the smaller the adj.*p*, the higher the ranking.



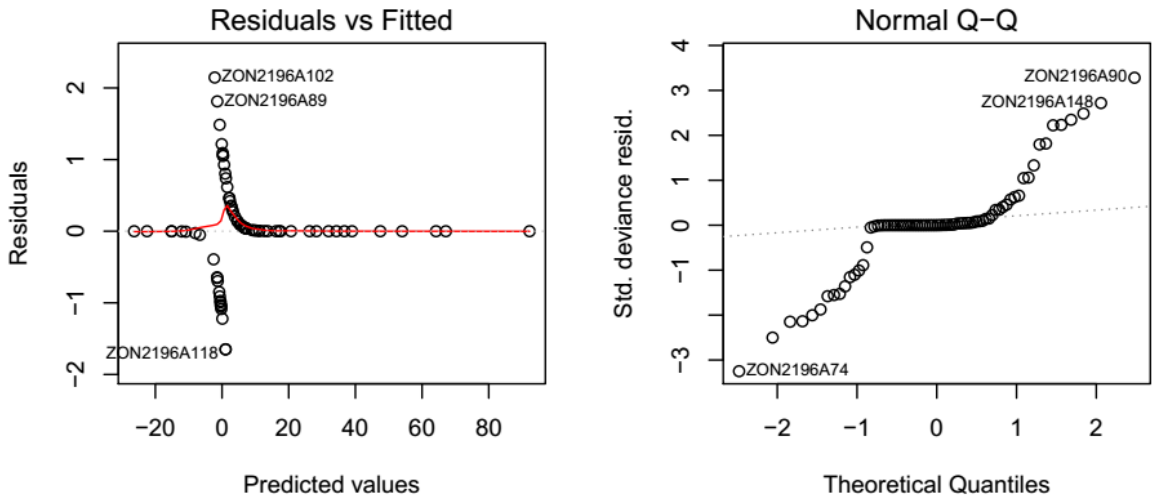
Supplementary Figure S 4.8 Scatter plot of log-scale adjusted *p*-value of differentially expressed genes common to in-house `edgeR` (proliferative vs. all secretory) and Fung et al. (proliferative vs. early-secretory) analyses.

Key: blue line = regression line by linear model, blue shading = 95% CI of regression line.

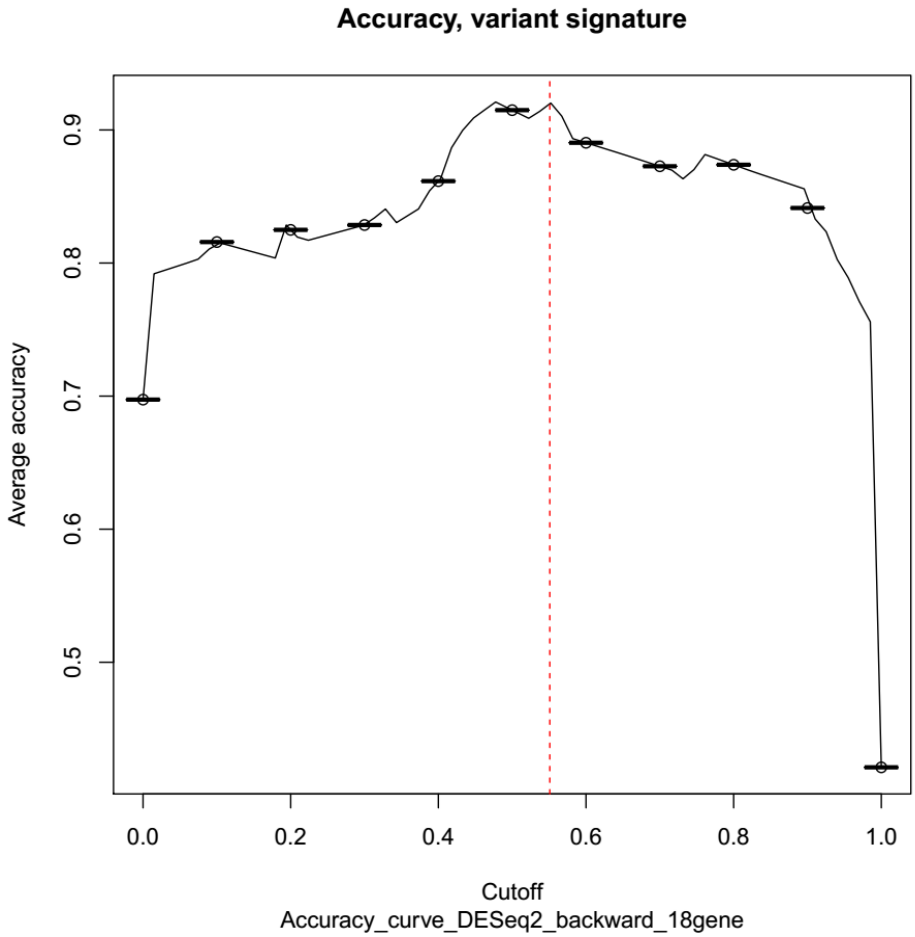


Supplementary Figure S 4.9 Heat map of correlation of expression patterns between 26 selected genes for building the multivariate prediction model to distinguish proliferative and secretory phases.

Key: diagonal line in dark blue = correlation of same gene to same gene equal to +1; intensity and size of blue dot = more positively correlated; intensity and size of red dot = more negatively correlated; row labelled as hist_men_3phase = correlation of a given gene expression to the condition (i.e. proliferative or secretory phase)

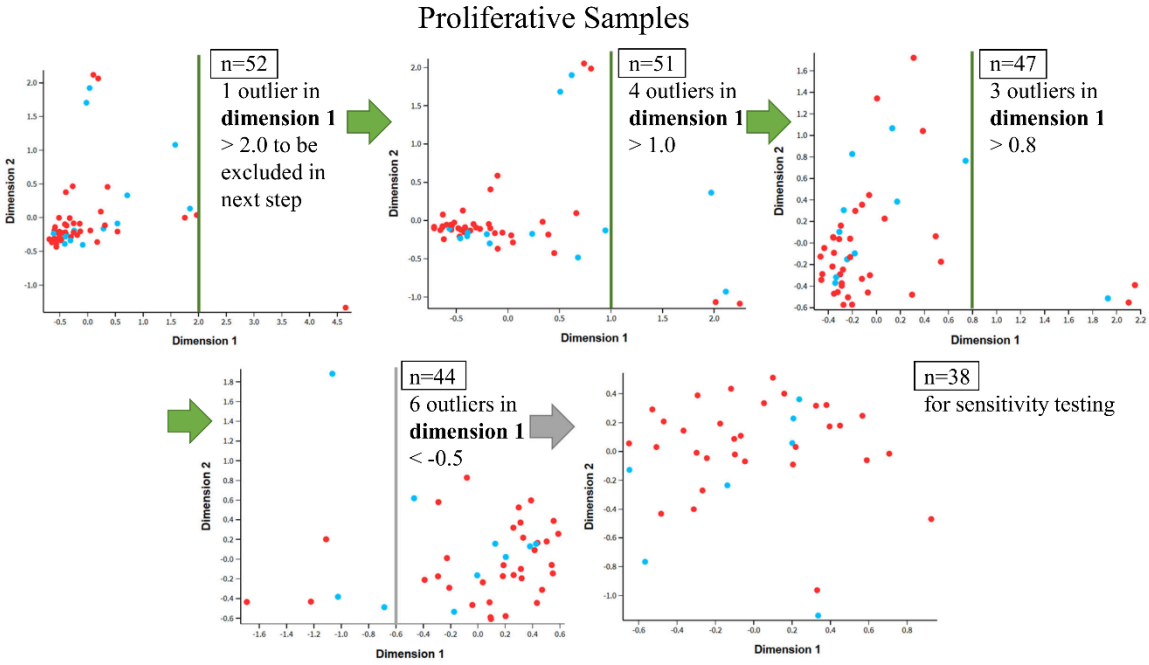


Supplementary Figure S 4.10 Residuals-versus-fitted plot (Left) and Quantile-quantile (Q-Q) plot (Right) of the 18-gene model for proliferative and secretory phase prediction.



Supplementary Figure S 4.11 Accuracy variant signature curves from prediction for proliferative and secretory phases based on the 18-gene model.

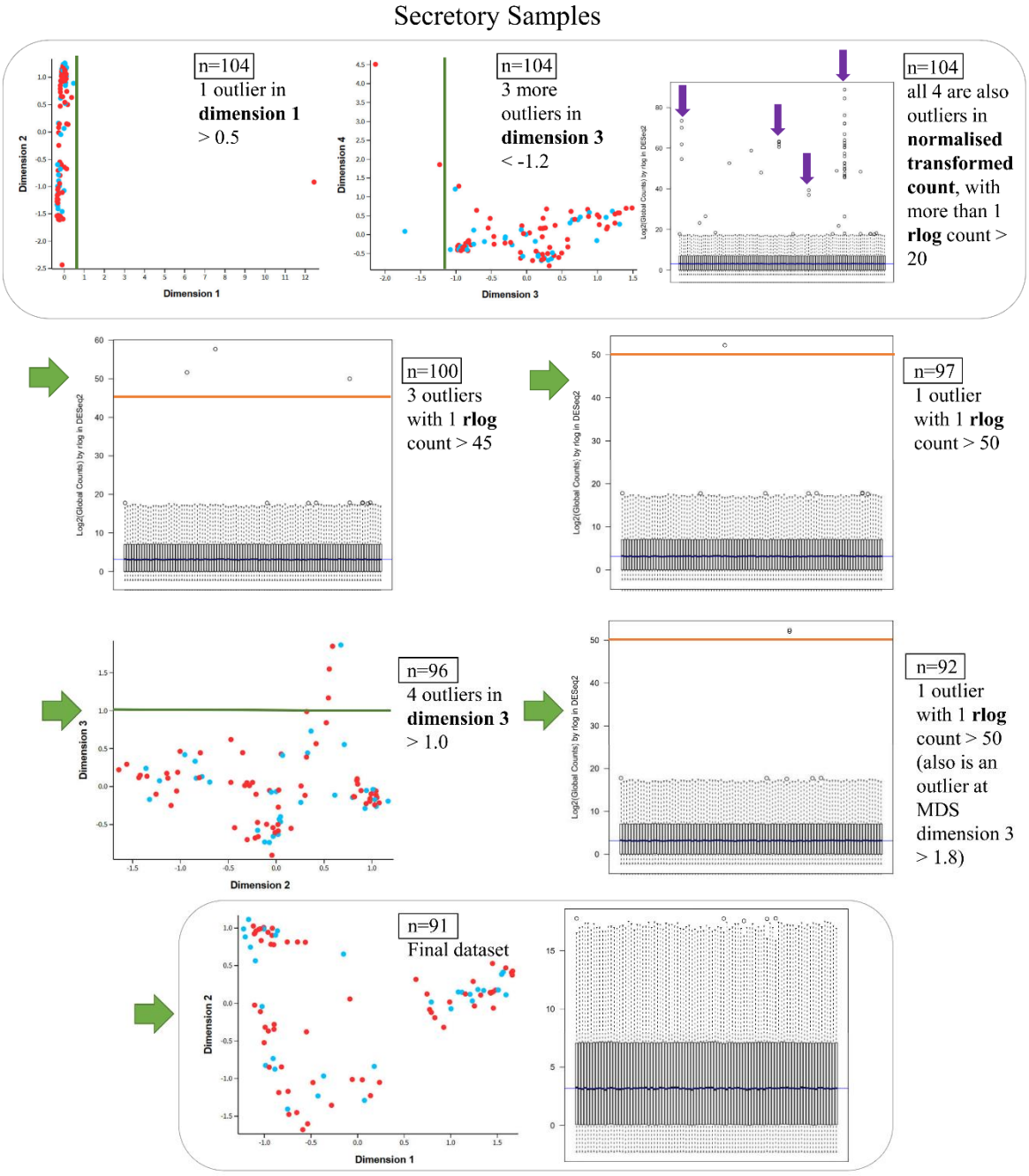
Key: red dotted line = predicted probability cut-off (threshold value of the classifier to give a result of 'proliferative' versus 'secretory') at 0.55, where the average accuracy of the model is at its highest at 92%.



Supplementary Figure S 5.1 Workflow of further sample QCs (MDS plots) for proliferative samples, with removal of outliers in three steps (from n=52 to n=44 final dataset & n=38 subset for sensitivity testing).

Flow: starting from top left corner to the top right, then to bottom left and bottom right.

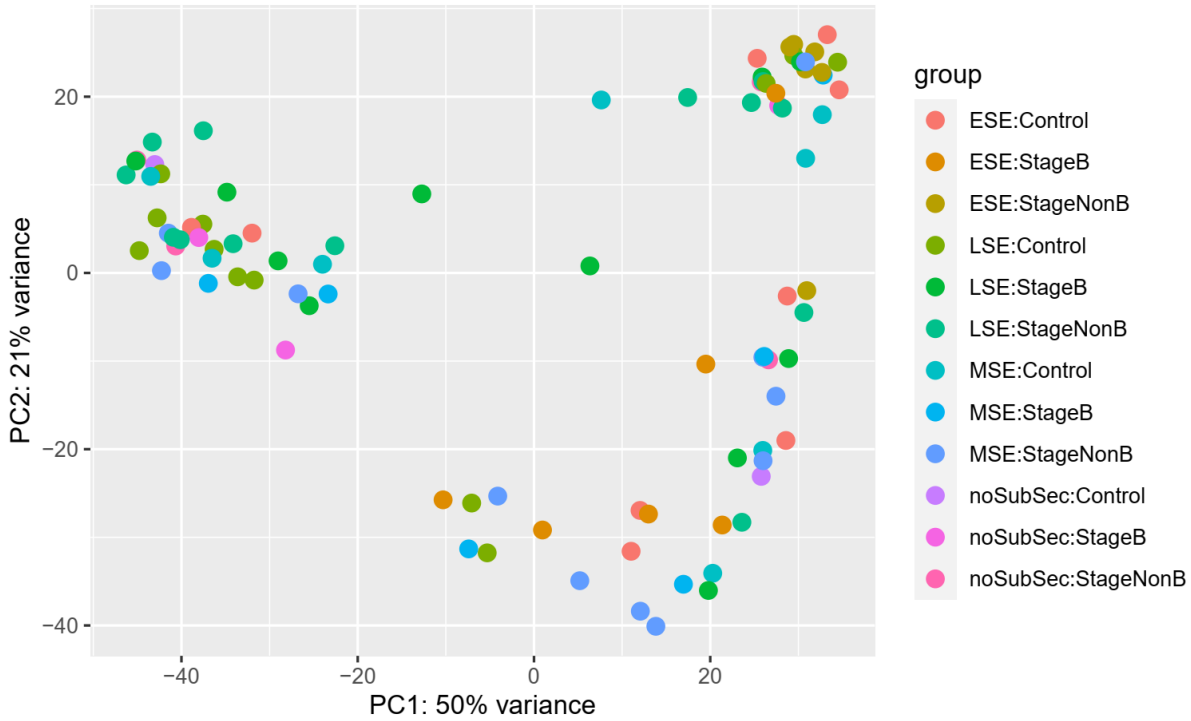
Key: red dots = endometriosis cases, blue dots = clinic controls, green line = dimension value to help defining outliners in each round, green arrow = the flow of data cleaning leading to the final dataset with 44 samples, grey line = dimension value to help defining additional outliners for a smaller dataset for sensitivity testing, grey arrow = the removal of additional outliners for sensitivity testing.



Supplementary Figure S 5.2 Workflow of further sample QCs (MDS & box plots for normalised and log-transformed counts) for secretory samples, with removal of outliers in five steps (from n=104 to n=91 final dataset).

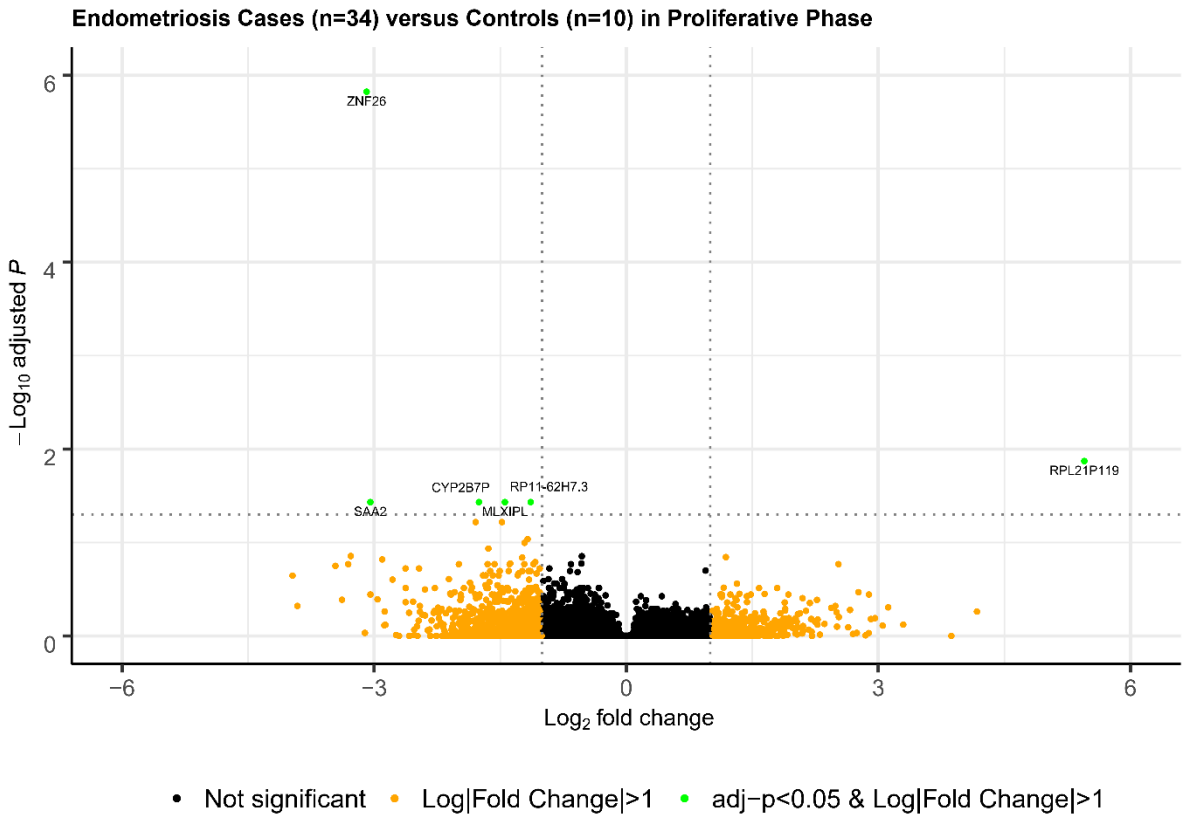
Flow: starting from top left panel to the right, then the rows below from the left to the right, and to the final bottom panel.

Key: red dots = endometriosis cases, blue dots = clinic controls, green line = dimension value to help defining outliners in MDS plots, purple arrow = outliers in box plots, orange line = rlog value to help defining outliners in box plots, green arrow = the flow of data cleaning leading to the final dataset with 91 samples.



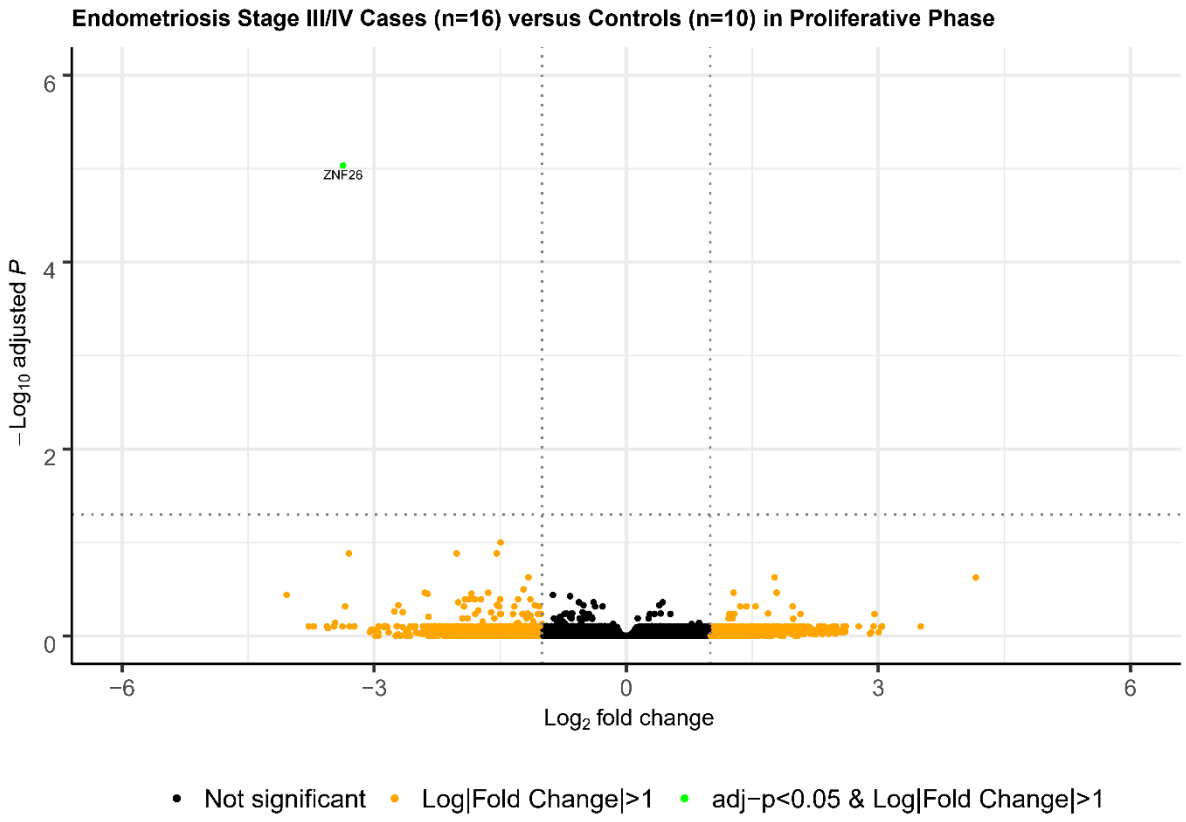
Supplementary Figure S 5.3 PCA plot on variance stabilising transformed (VST) counts of 91 secretory samples for observation of possible interaction between the endometriosis stages-control and the sub-secretory phase.

Key: PC = Principal Component, ESE = early-secretory phase, MSE = mid-secretory phase, LSE = late-secretory phase, Control = clinic controls for endometriosis, StageB = ASRM Stage III/IV cases, StageNonB = ASRM Stage I/II endometriosis cases.



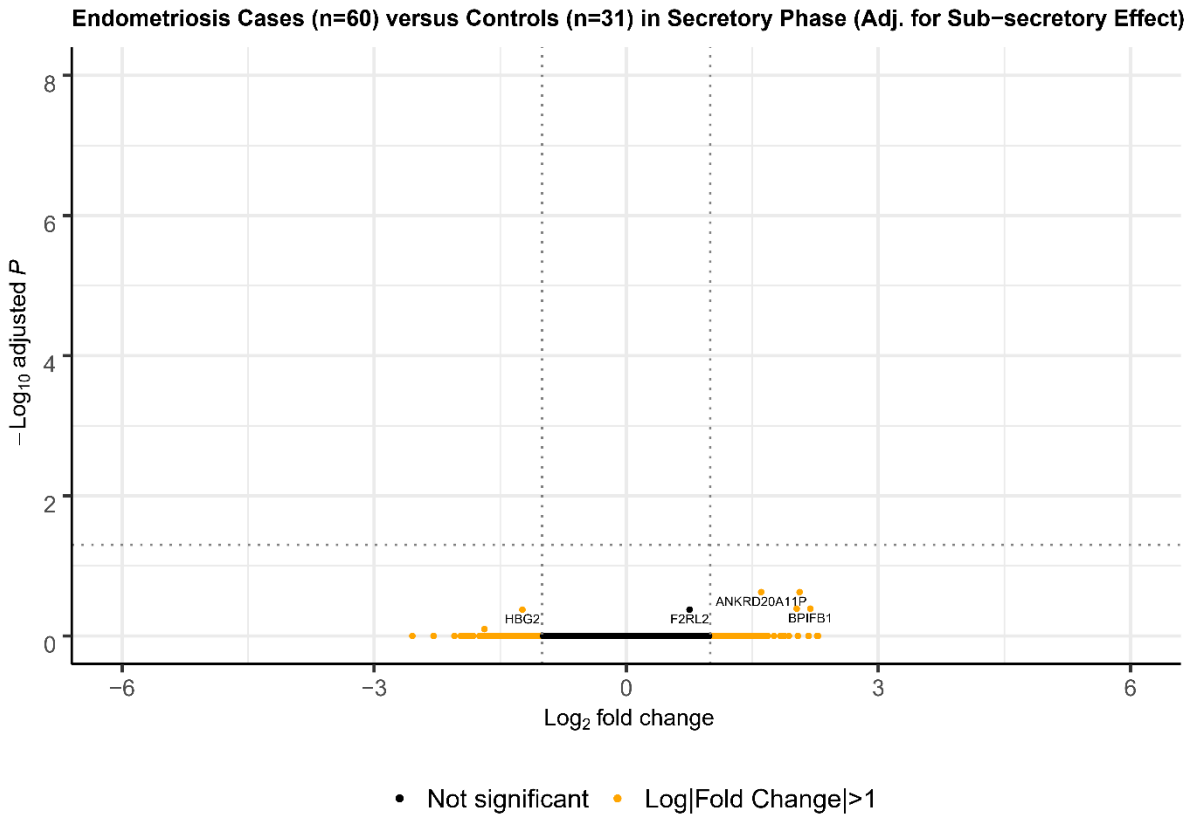
Supplementary Figure S 5.4 Volcano plot for differentially expressed genes between endometriosis cases (n=34) and clinic controls (n=10) in proliferative phase of the endometrium.

Key: green dots = gene with both adj.p<0.05 & Log_2 |FC >1, orange dots = gene with Log_2 |FC|>1 only, and black dots = gene with adj.p>0.05.

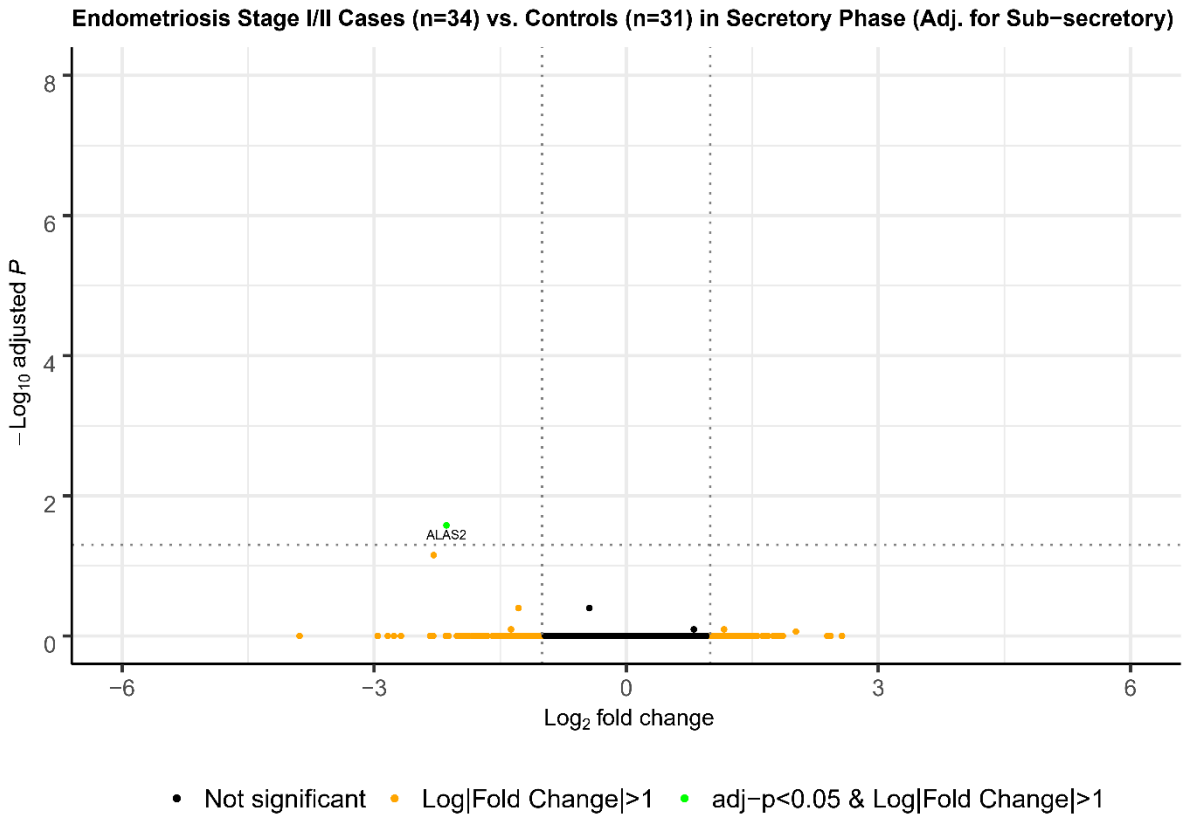


Supplementary Figure S 5.5 Volcano plot for differentially expressed genes between endometriosis stage III/IV cases (n=16) and clinic controls (n=10) in proliferative phase of the endometrium.

Key: green dots = gene with both adj.p<0.05 & Log_2 |FC >1, orange dots = gene with Log_2 |FC|>1 only, and black dots = gene with adj.p>0.05.

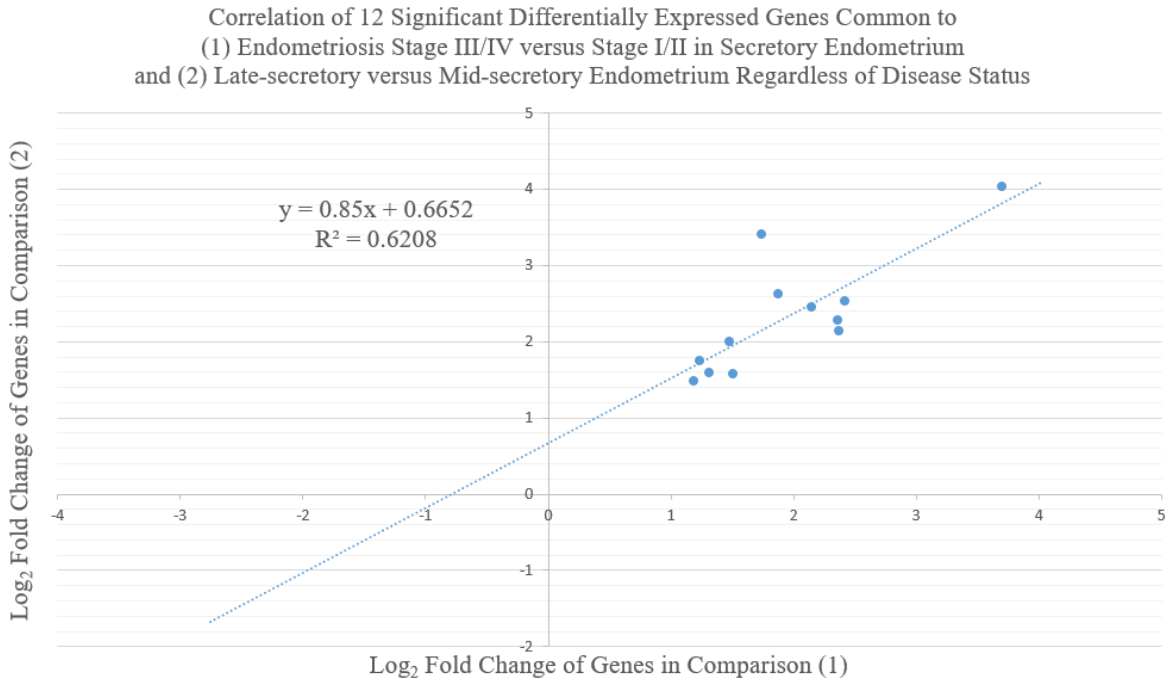


Supplementary Figure S 5.6 Volcano plot for differentially expressed genes between endometriosis cases (n=60) and clinic controls (n=31) in secretory phase of the endometrium. Key: orange dots = gene with Log_2 |FC|>1 only, and black dots = gene with adj.*p*>0.05.

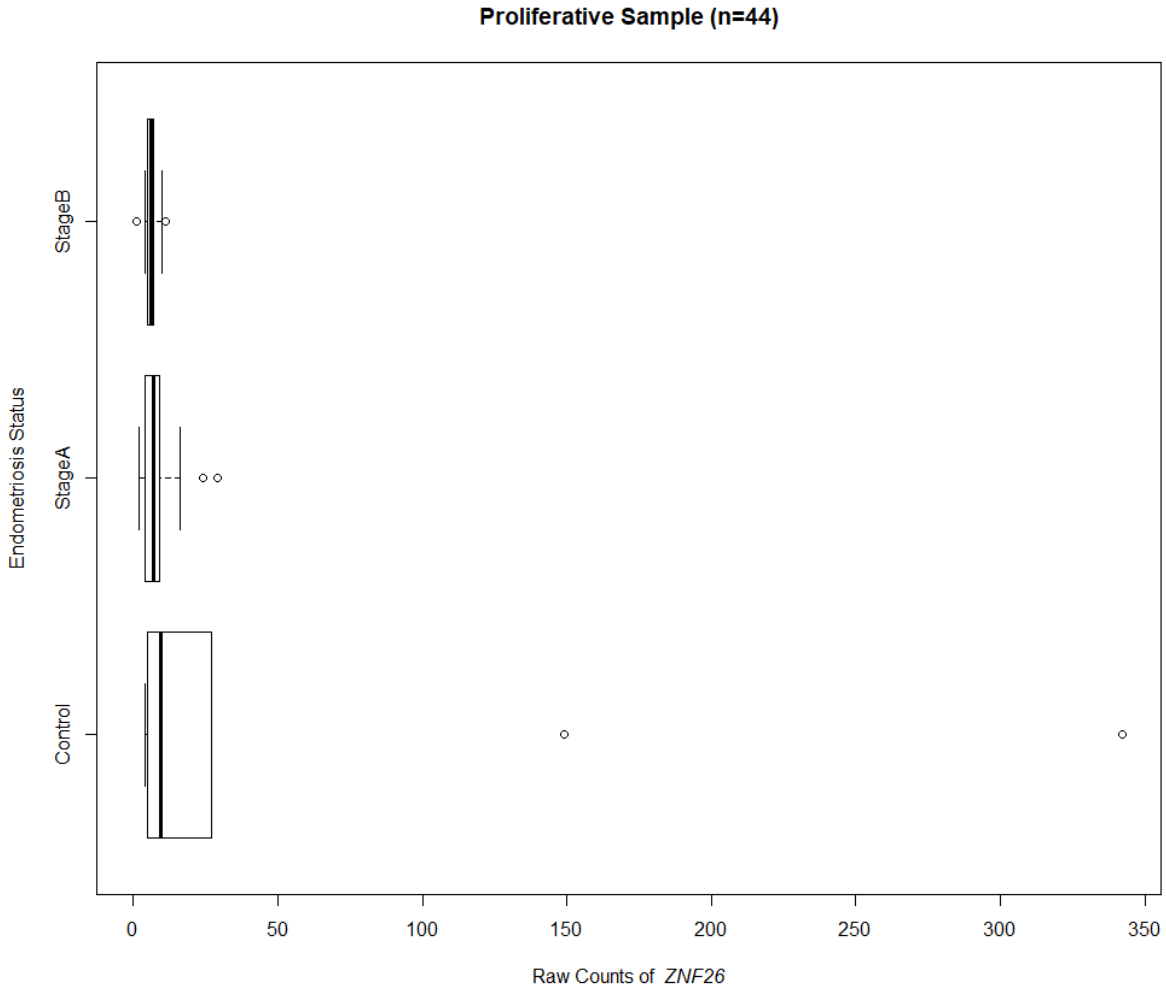


Supplementary Figure S 5.7 Volcano plot for differentially expressed genes between endometriosis stage I/II cases (n=34) and clinic controls (n=31) in secretory phase of the endometrium.

Key: green dots = gene with both adj.p<0.05 & Log_2 |FC >1, orange dots = gene with Log_2 |FC|>1 only, and black dots = gene with adj.p>0.05.



Supplementary Figure S 5.8 Correlation plot of Log_2 fold changes of 12 differentially expressed genes in both comparisons of (1) Endometriosis stage III/IV vs. I/II in secretory endometrium, and (2) Late-secretory vs. mid-secretory endometrium regardless of disease status.



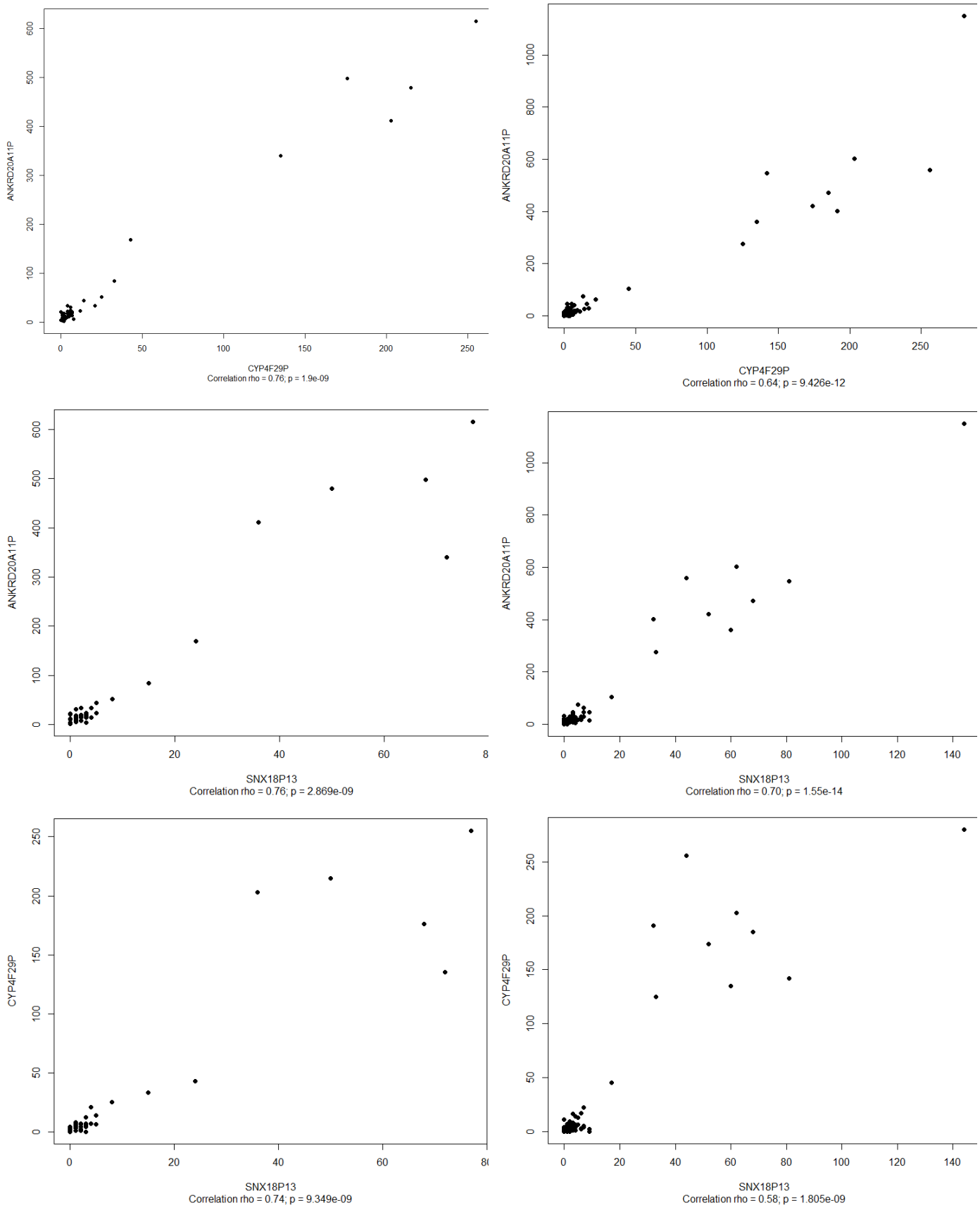
Supplementary Figure S 5.9 Box plot of raw count for *ZNF26* gene in 44 proliferative samples, grouped by endometriosis disease status.

Key: StageA = ASRM stage I/II endometriosis cases, StageB = stage III/IV, control = clinic controls.

Spearman's Rank Correlation of Genes Raw Count in

Proliferative Group (n=44)

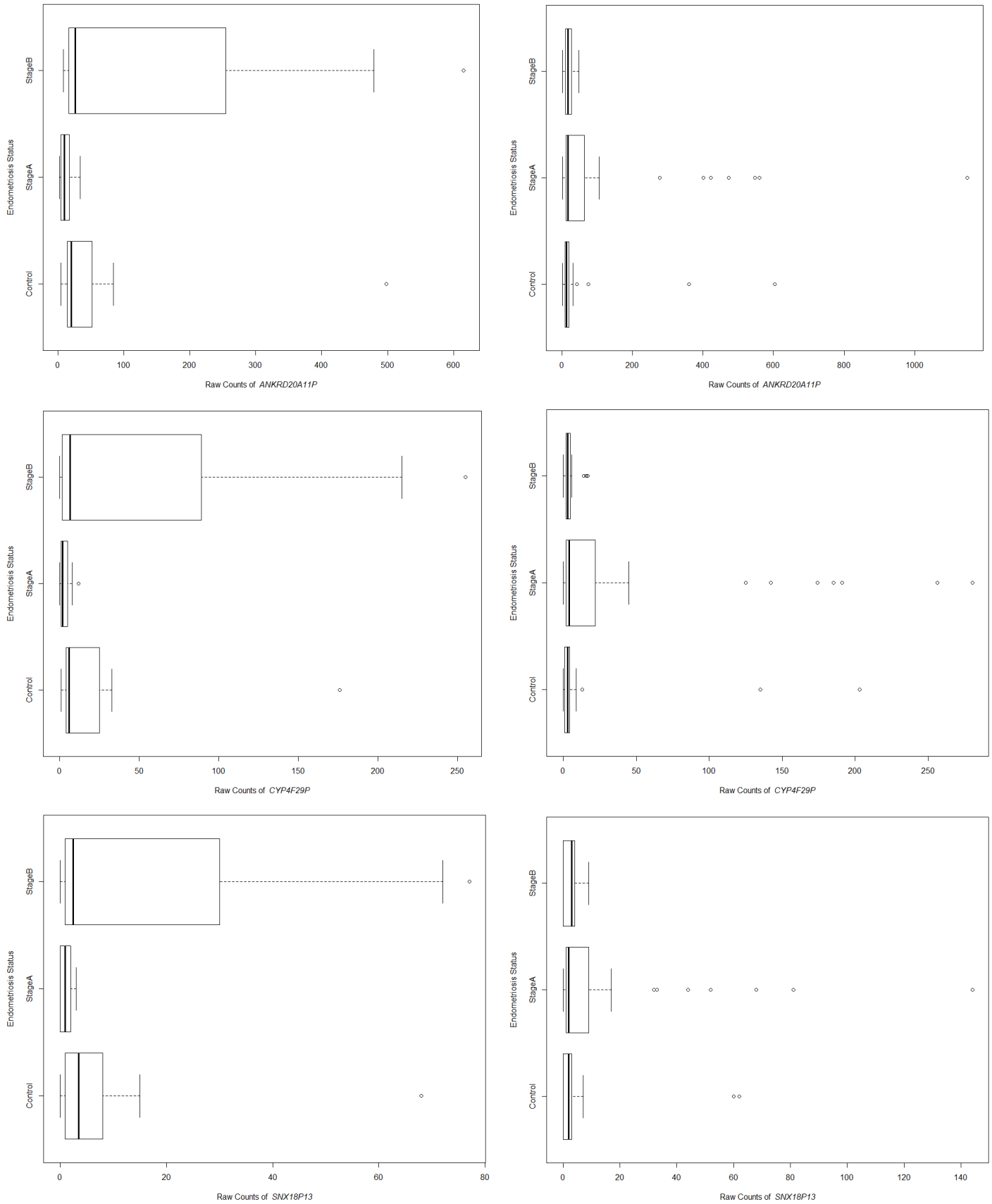
Secretory Group (n=91)



Supplementary Figure S 5.10 Correlation plots of raw counts for *ANKRD20A11P*, *CYP4F29P* & *SNX18P13* per sample in proliferative and secretory phase.

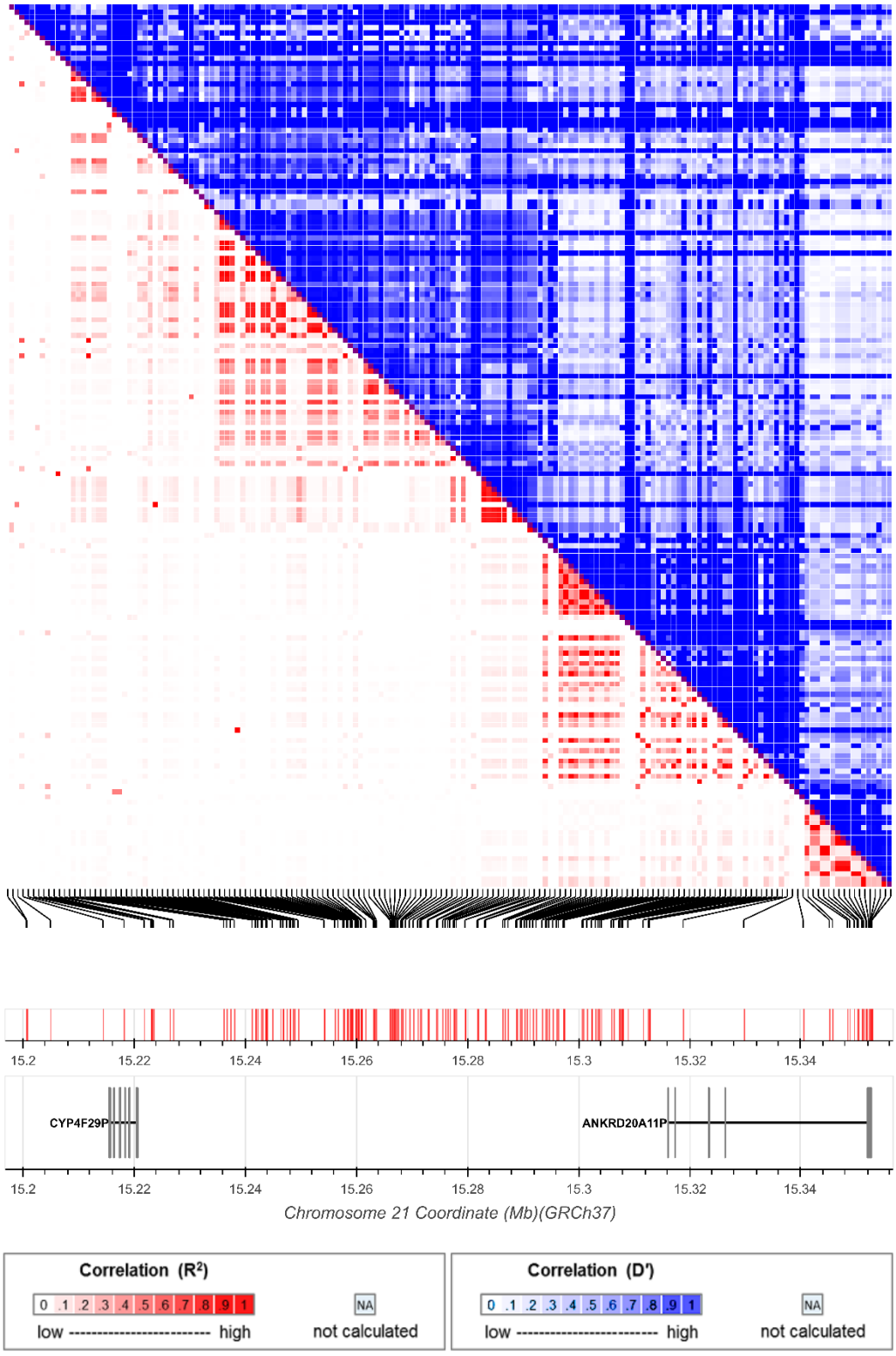
Proliferative Group (n=44)

Secretory Group (n=91)



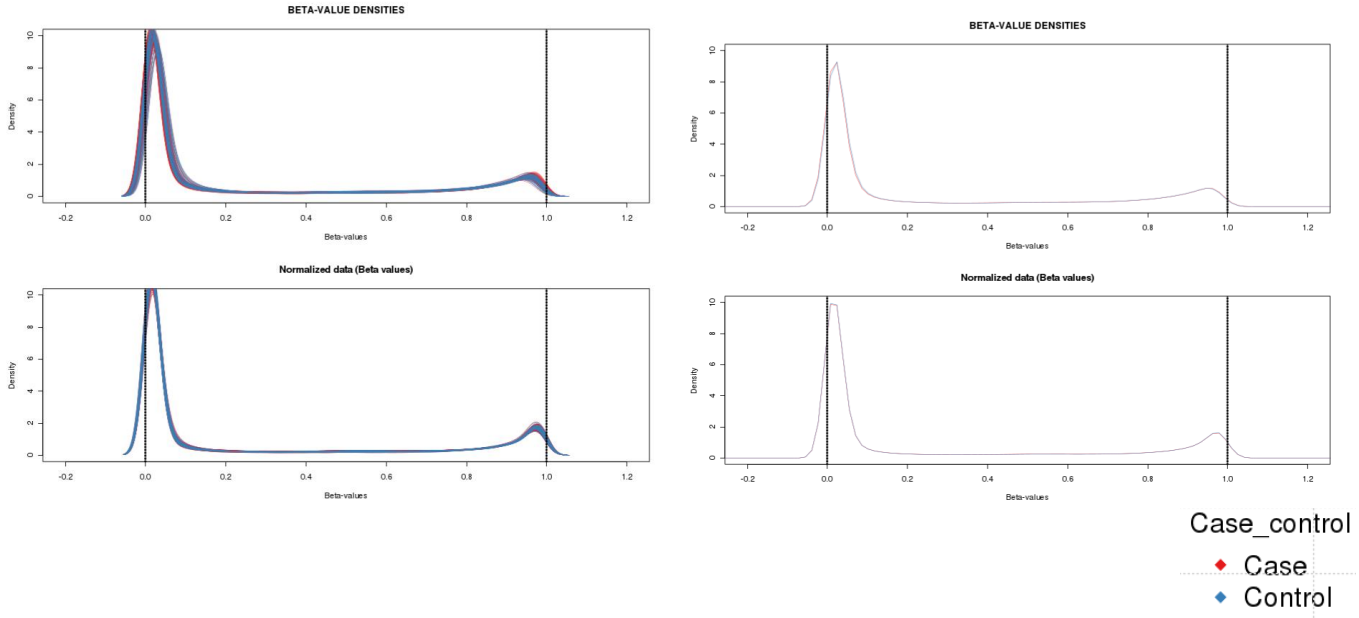
Supplementary Figure S 5.11 Box plot of raw counts for *ANKRD20A11P*, *CYP4F29P* & *SNX18P13* in proliferative and secretory samples, grouped by endometriosis disease status.

Key: StageA = ASRM stage I/II endometriosis cases, StageB = stage III/IV, control = clinic controls.

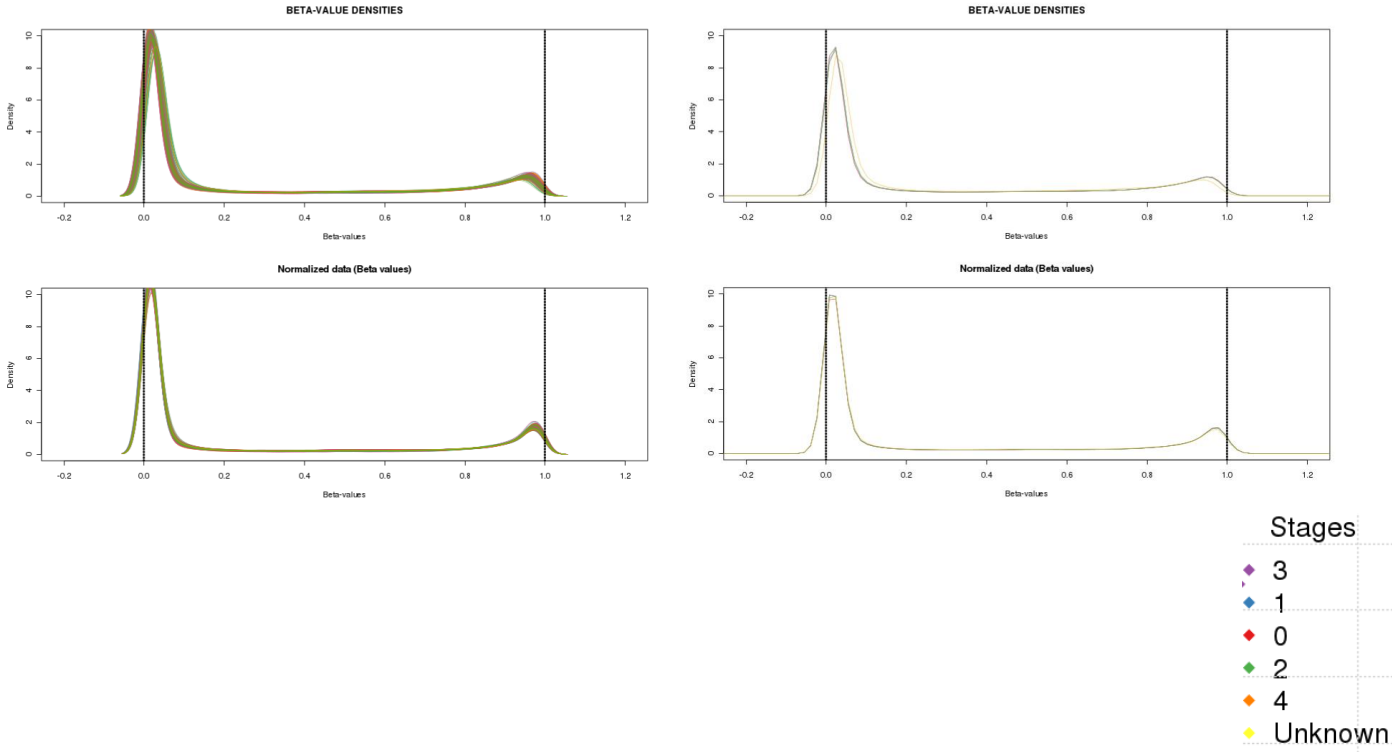


Supplementary Figure S 5.12 Heat map matrix of pairwise linkage disequilibrium statistics for variants in the region of chromosome 21: position 15,200,585-15,352,881 (GRCh37) in relation to genomic locations of *ANKRD20A11P* (36.5 kb downstream of *SNX18P13*) and *CYP4F29P* pseudogenes.

(Permission for re-distribution (both print and digital) of figure generated by LDmatrix online tool is granted on 02 Oct 2020 by NIH/NCI, permission email in supplementary materials F.iii)

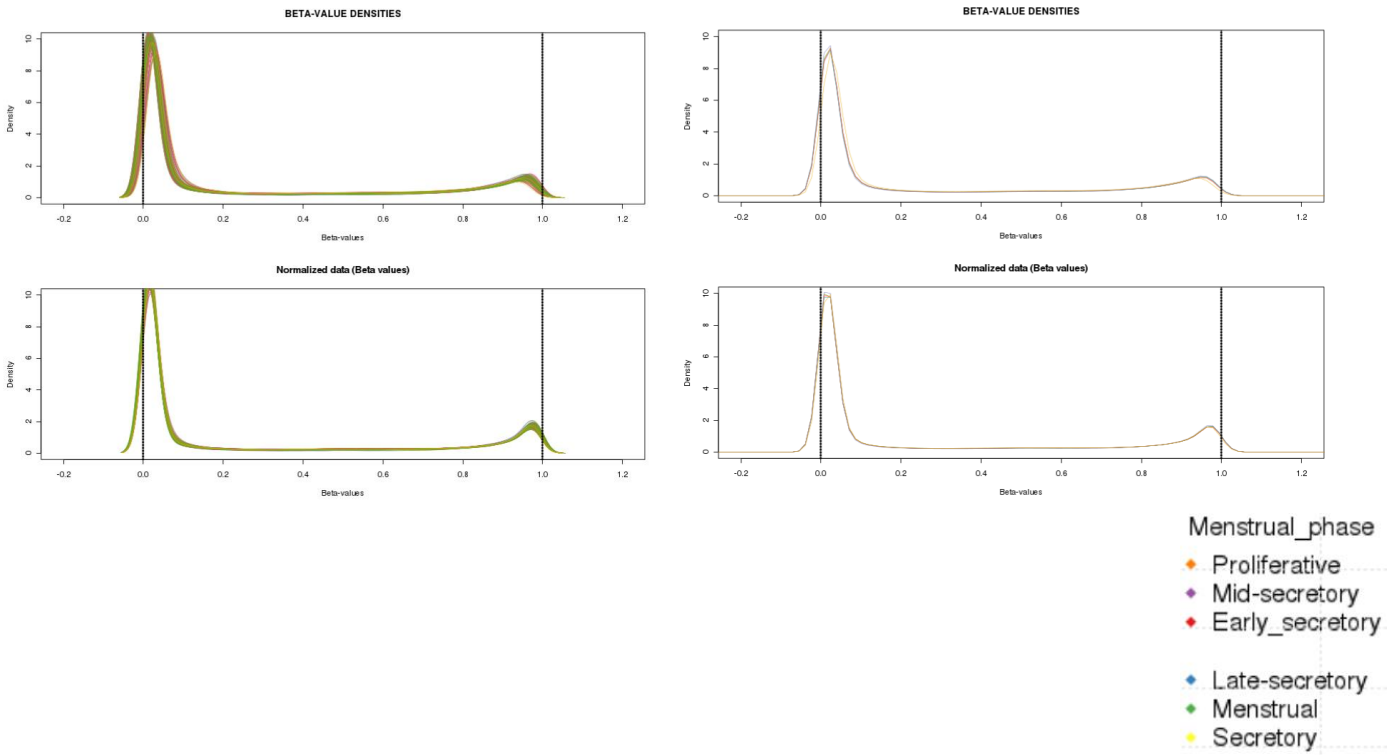


Supplementary Figure S 6.1 Density plots showing (top panel) raw beta value, and (bottom panel) pre-processed normalised beta value of type I green probes in eutopic endometrium samples for endometriosis cases (red line) and clinic controls (blue line). Left panel shows lines from every samples, while right panel shows the average density of all samples.



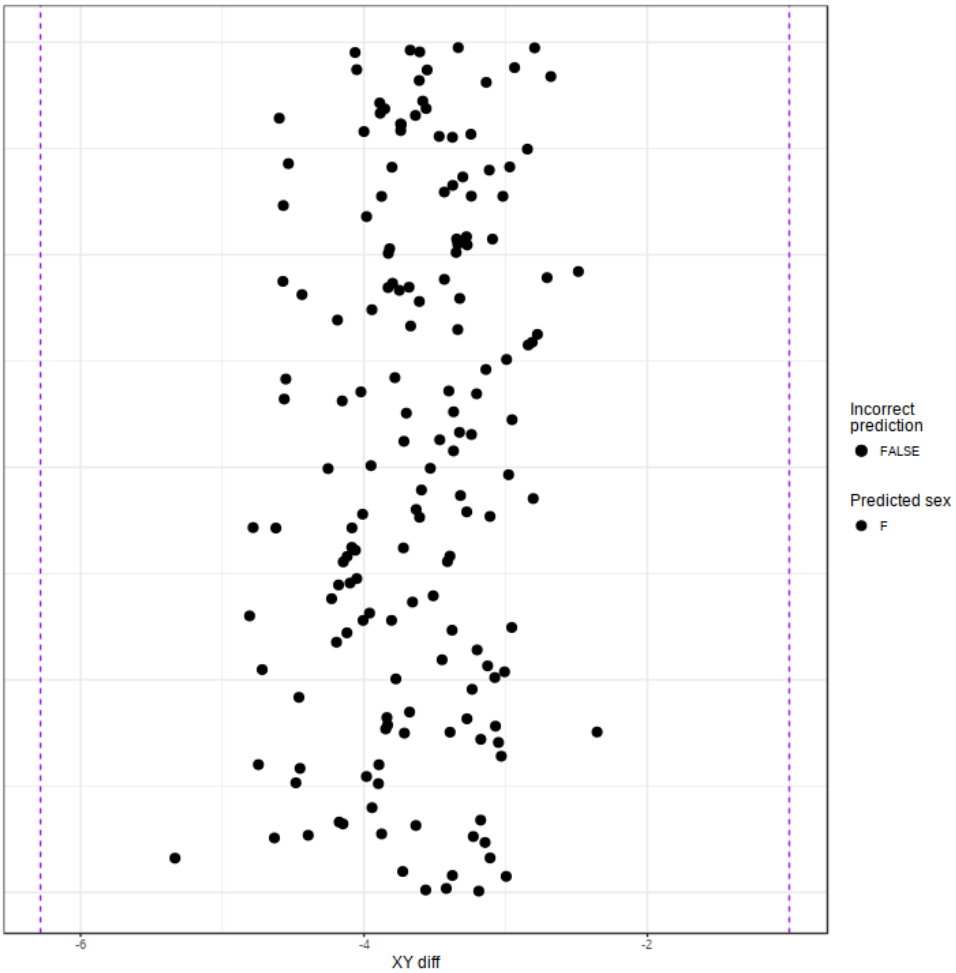
Supplementary Figure S 6.2 Density plots showing (top panel) raw beta value and (bottom panel) pre-processed normalised beta value of type I green probes in eutopic endometrium samples according to ASRM surgical staging. Left panel shows lines from every samples, while right panel shows the average density of all samples.

Key: blue line = ASRM stage I (1), green line = stage II (2), purple line = stage III (3), orange line = stage IV (4), and yellow line = unknown endometriosis status.



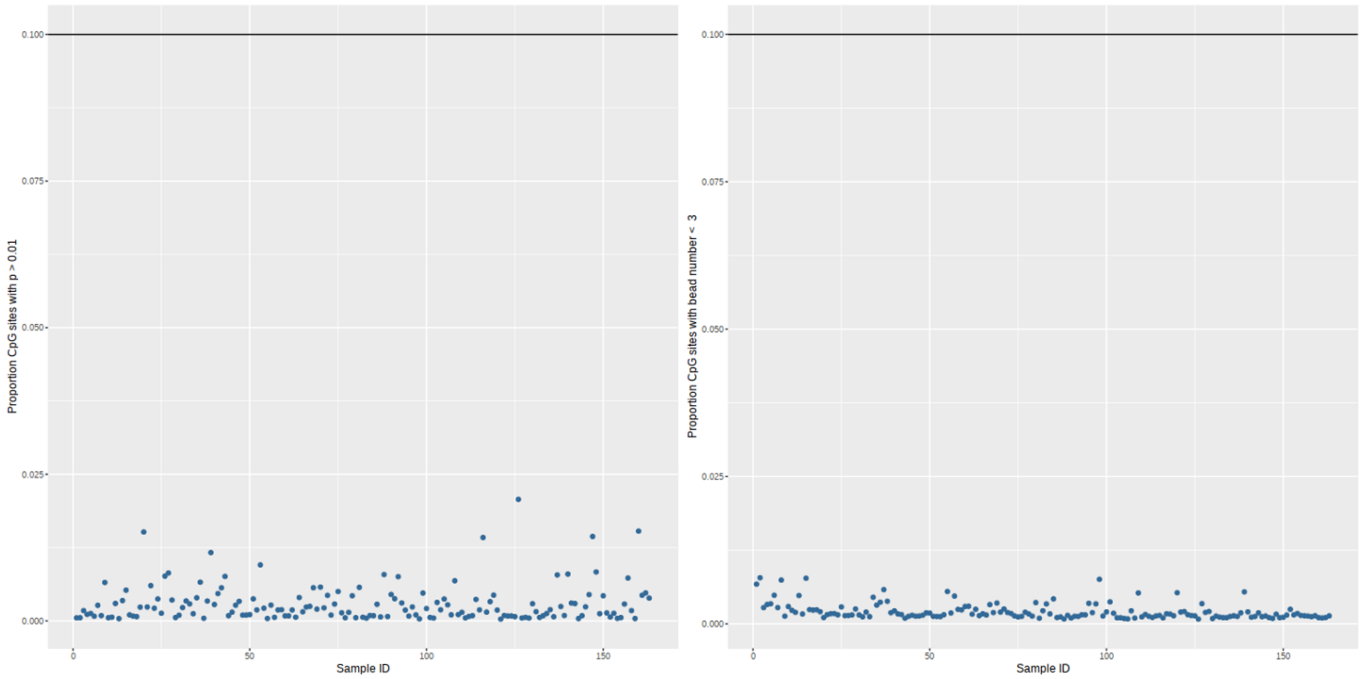
Supplementary Figure S 6.3 Density plots showing (top panel) raw beta value and (bottom panel) pre-processed normalised beta value of type I green probes in eutopic endometrium samples according to five cycle phases. Left panel shows lines from every samples, while right panel shows the average density of all samples.

Key: green line = menstrual phase, orange line = proliferative phase, red line = early-secretory phase, purple line = mid-secretory phase, blue line = late-secretory phase, yellow line = secretory phase without sub-secretory information.



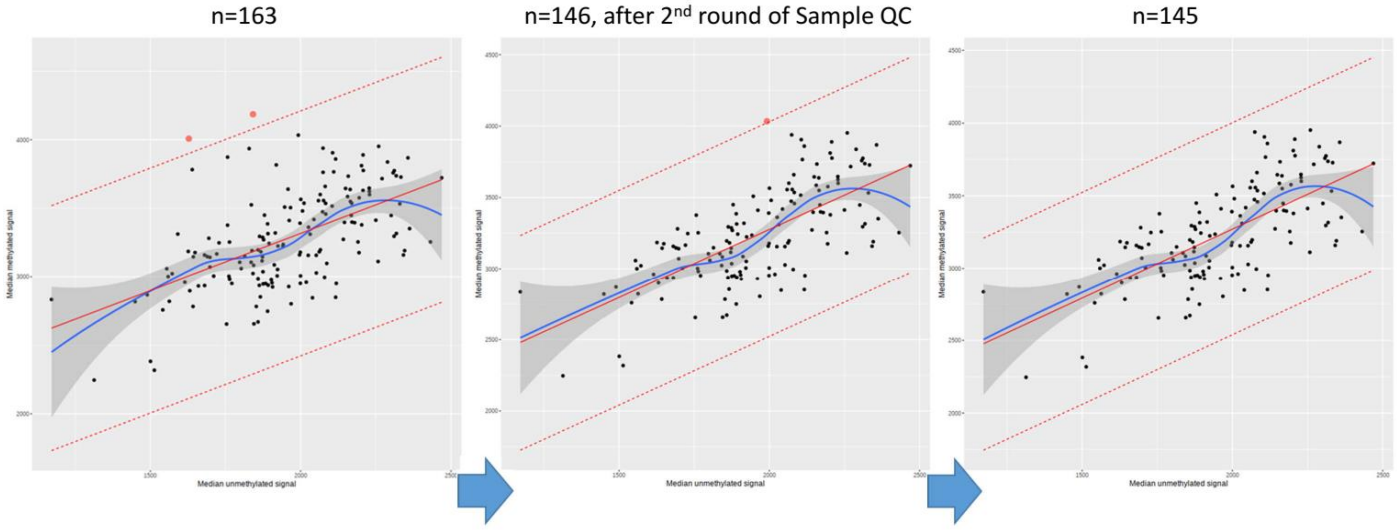
Supplementary Figure S 6.4 Plot of differences between total mean intensity of probes from X and Y chromosomes.

Key: XY diff at -2 = cut-off of sex detection. Black dots = samples with predicted sex as female (n=163). Purple dotted line = ±5SD boundary from the mean XY differences of the whole dataset.



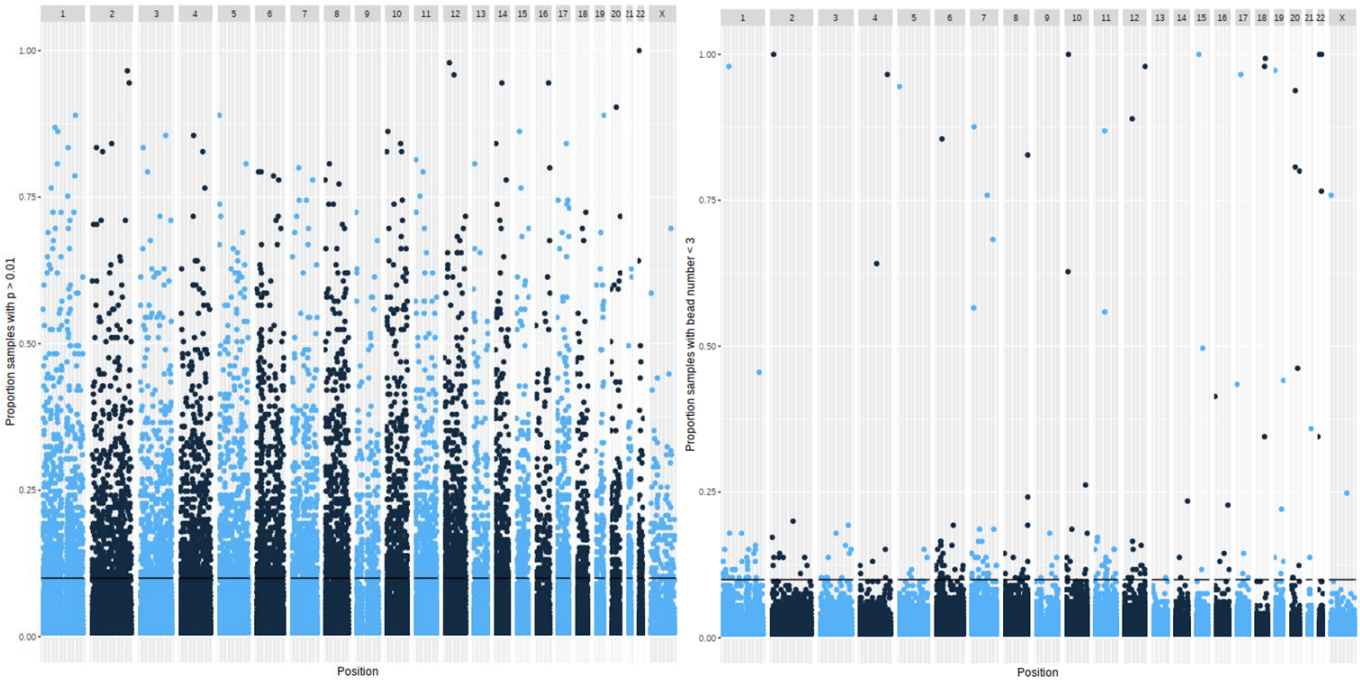
Supplementary Figure S 6.5 Samples (n=163) passing DNA methylation detection *p* values (left) and bead number for probes (right).

Key: blue dots = each samples with proportion of undetected sites (p>0.01, left) and with bead number <3 per CpG (right). Black line = threshold set at 10% of methylation probes.



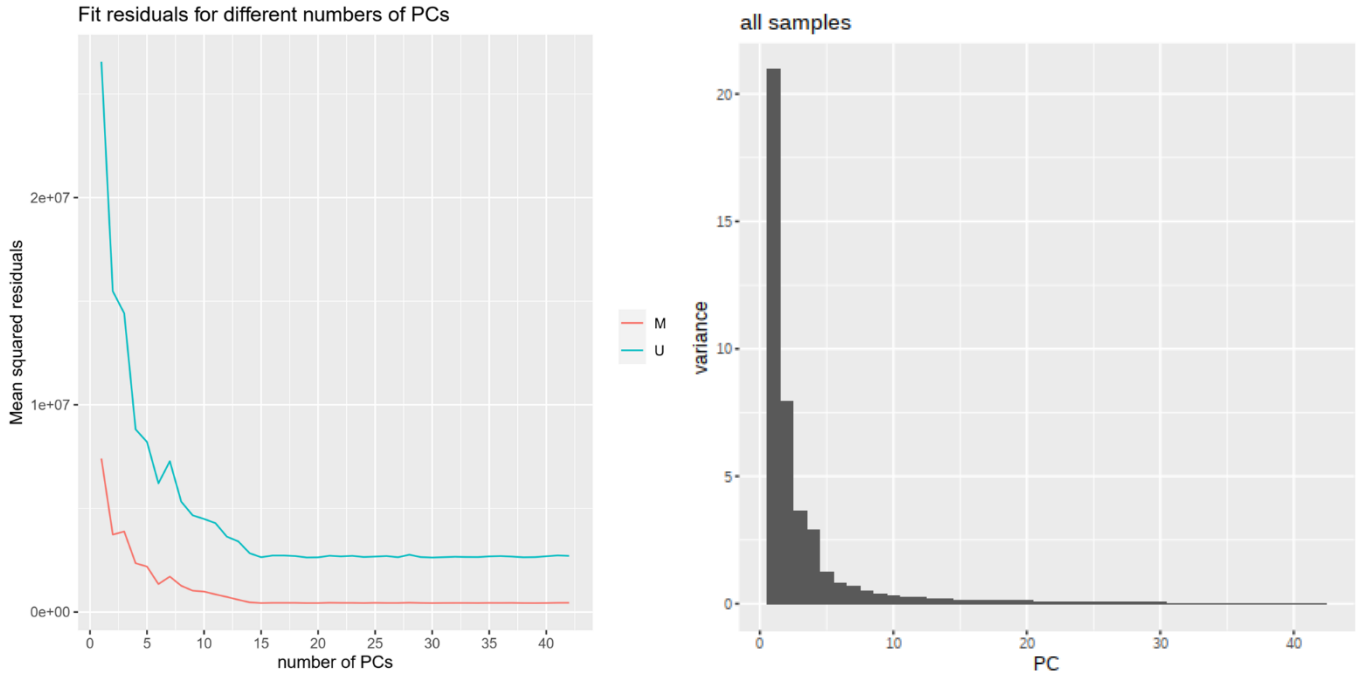
Supplementary Figure S 6.6 Plot of median methylation (M) versus unmethylation (UM) intensity of samples.

Key: red line = regression line of expected ratio of M vs UM, red dotted lines = ±3SD boundaries of the regression line. Black dots = samples passing QC with the M vs UM ratio within ±3SD of the regression line, red dots = outlying samples (n=3) sit outside of ±3SD of the regression line. Blue line = fitted line for dataset.



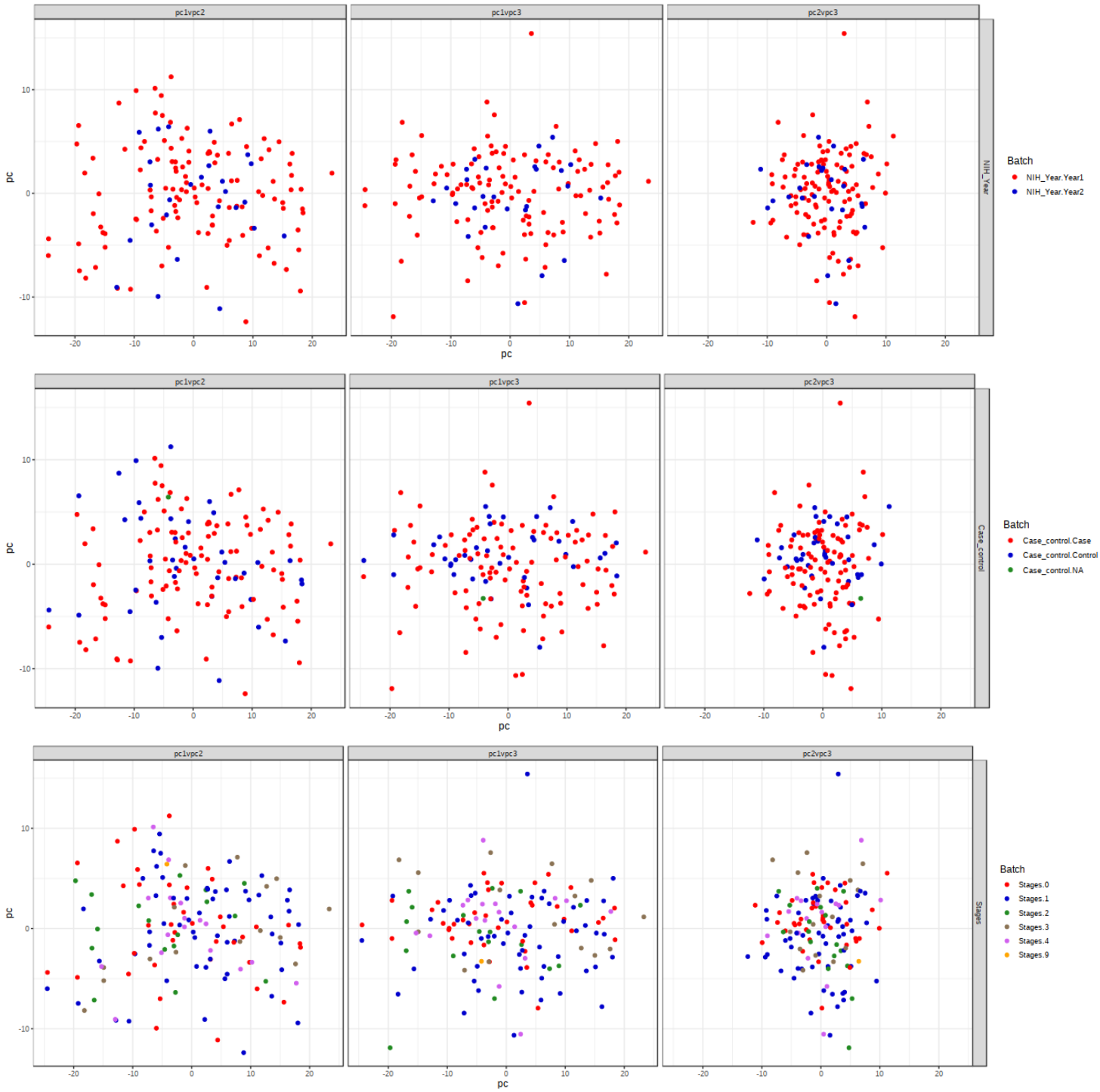
Supplementary Figure S 6.7 Methylation probes showing proportion of samples with detection *p* values (left) and bead number (right).

Key: dark and light blue dots = each probes with proportion of undetected sites (p>0.01, left) and with bead number <3 per CpG (right). Black line = threshold set at 10% of samples.



Supplementary Figure S 6.8 Scree plots for residual fitting of serial number of principal components (PCs) for adjustment of technical effects on DNA methylation signals (left), and variance captured by each PCs for control probes (right).

Key: plot on the left: M (red line) = mean squared residuals of DNA methylation signals; U (blue line) = mean squared residuals of unmethylation signals.



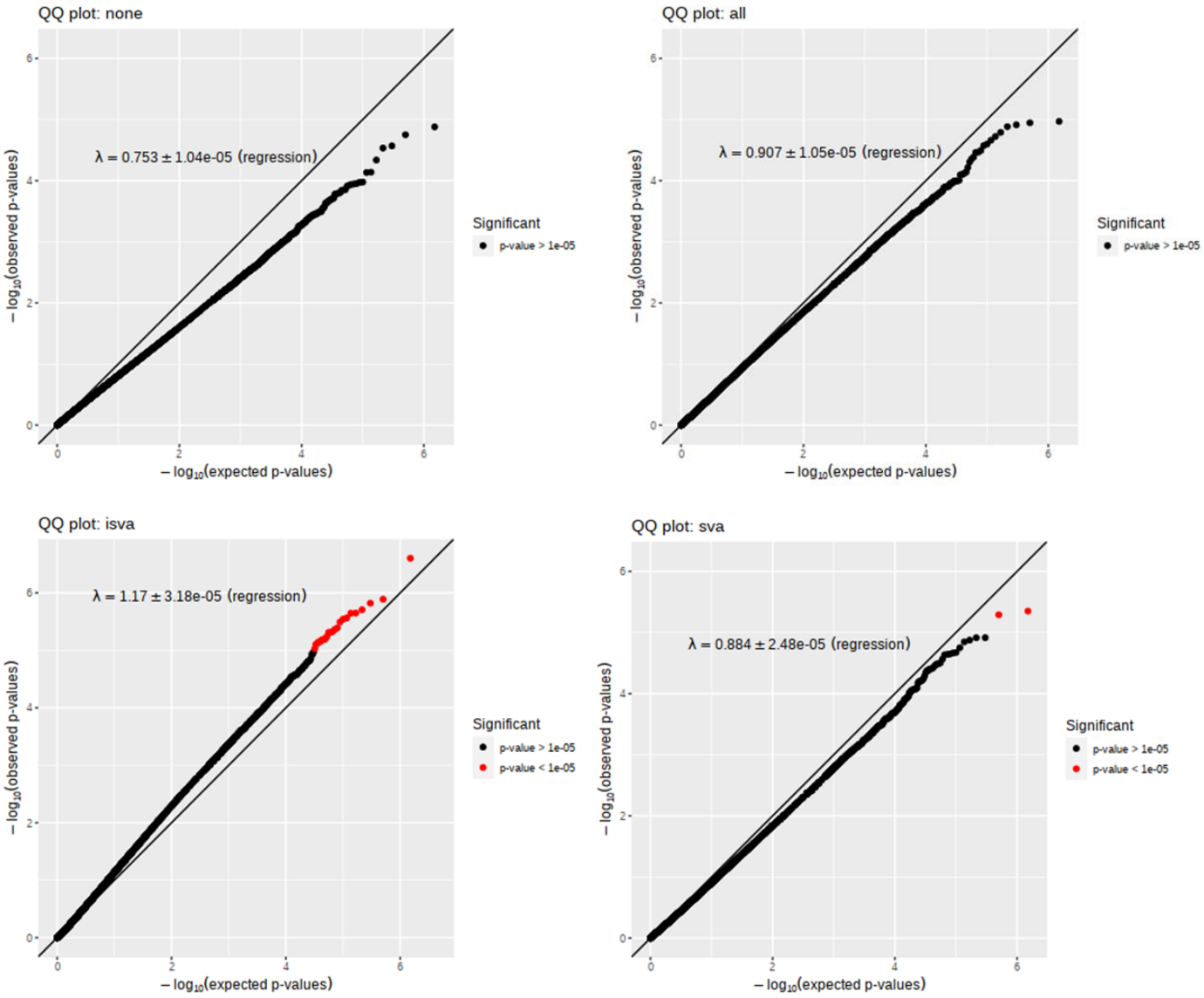
Supplementary Figure S 6.9 PCA plots of samples after QC (n=145) colour-coded by discovery and replication phase (top panel), by endometriosis case-control status (middle panel), and by ASRM surgical stages (bottom panel).

Key: left column is for PC1 vs PC2, central column for PC1 vs PC3, and right column for PC2 vs PC3.

Top panel: red dots =NIH discovery phase (Year 1), blue dots = replication phase (Year 2);

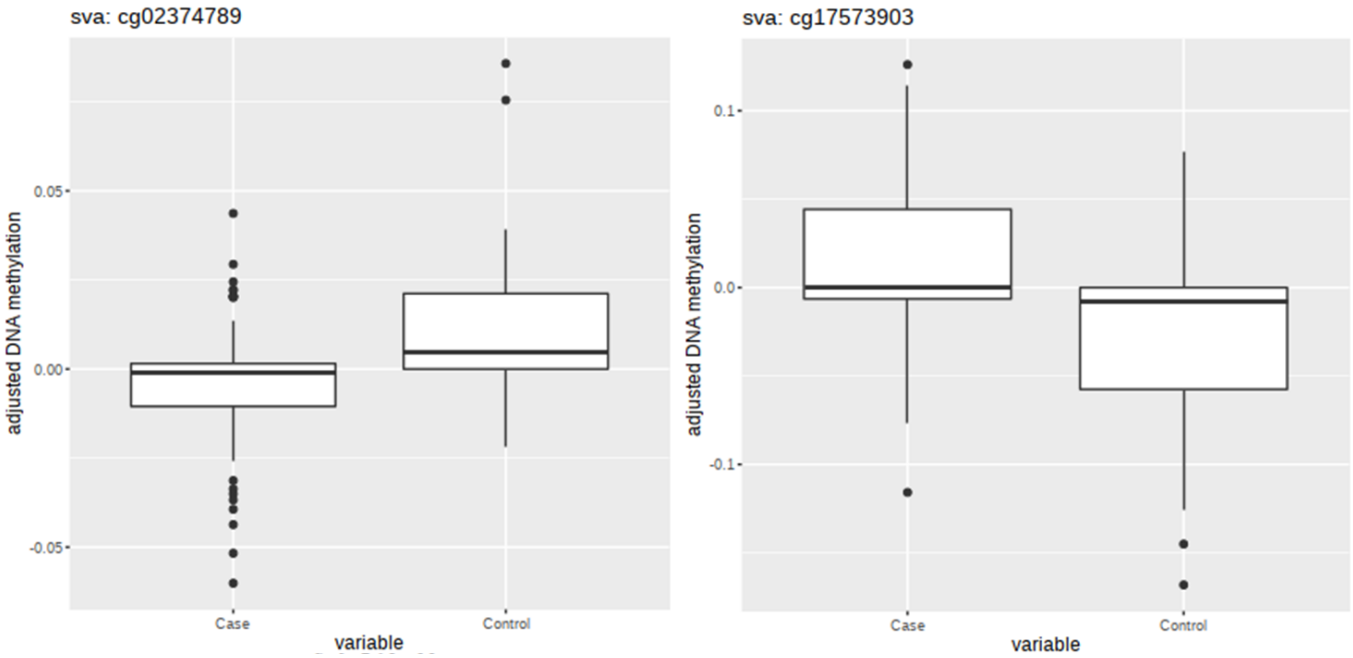
middle panel: red dots = endometriosis cases, blue dots = clinic controls, green dot = unknown disease status;

bottom panel: red dots = clinic controls, blue dots = endometriosis cases with ASRM stage I, green dots = stage II, grey dots = stage III, purple dots = stage IV, orange dots = unknown disease status.

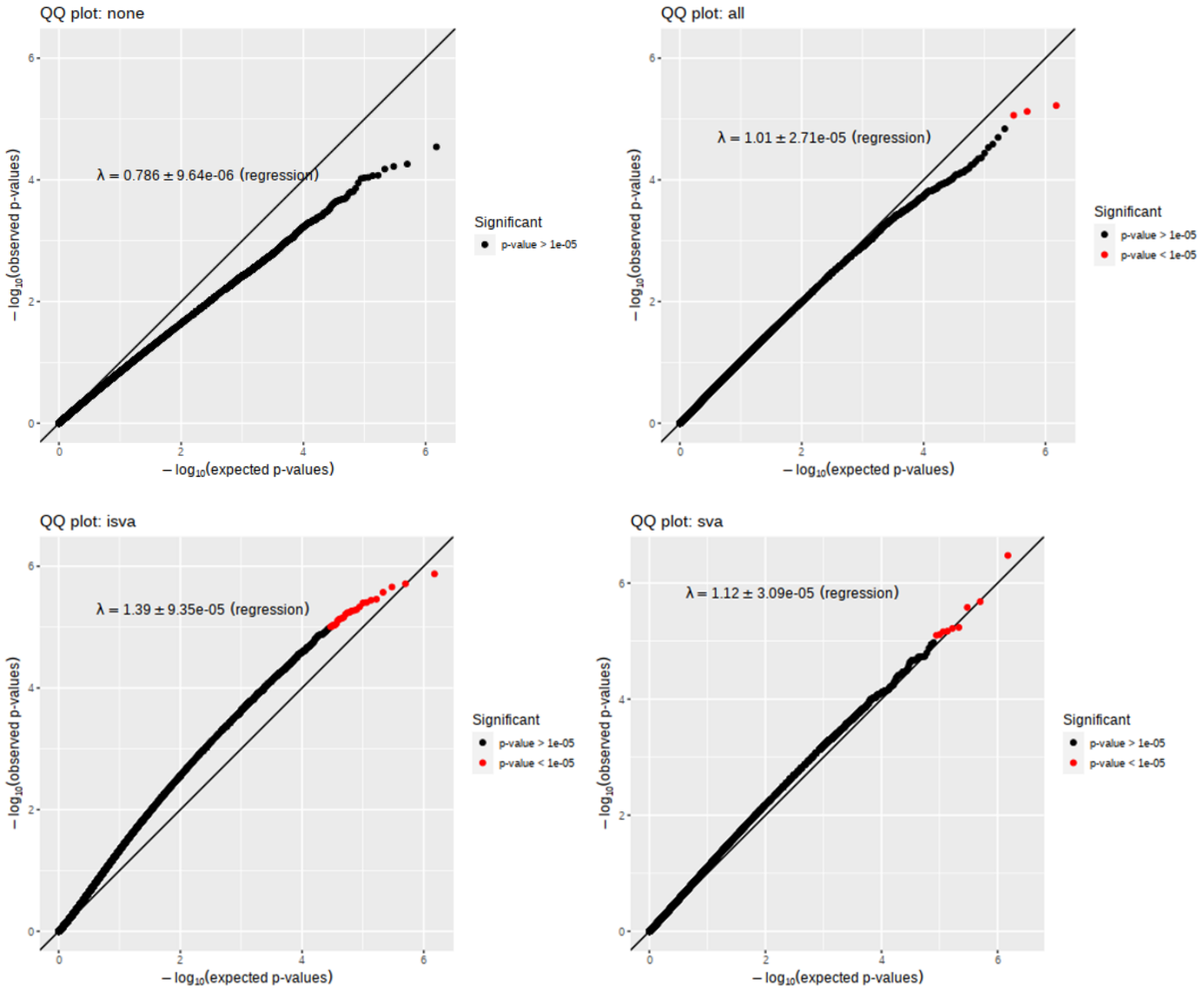


Supplementary Figure S 6.10 Quantile-quantile plots of EWAS analysis of endometriosis cases (n=100) and clinic controls (n=38) in four models: without any covariates (top left), with four user-provided covariates (top right), independent surrogate and user-provided variables — ISVA (bottom left), and surrogate and user-provided variables — SVA (bottom right).

Key: black dots = probes with $p>1x10^{-5}$, red dots = probes passing nominal epigenome-wide threshold with $p<1x10^{-5}$.

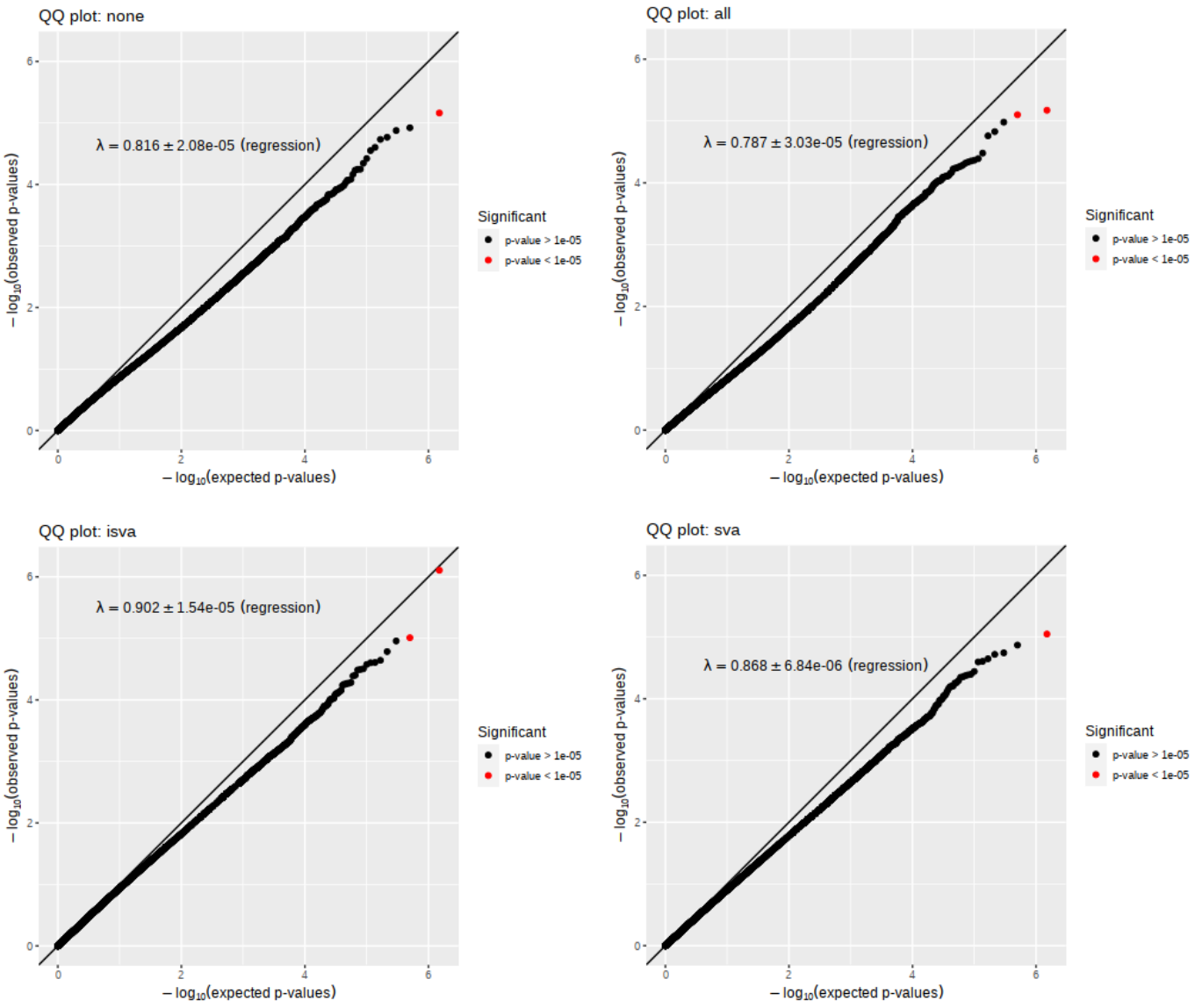


Supplementary Figure S 6.11 Box plots of DNA methylation level of two probes at chr11:69,340,400 (Illumina probe ID cg02374789, left) and chr12:1,068,597 (cg17573903, right) in EWAS of endometriosis case-control comparison by SVA model.



Supplementary Figure S 6.12 Quantile-quantile plots of EWAS analysis of endometriosis ASRM stage I/II cases (n=68) and clinic controls (n=38) in four models: without any covariates (top left), with two user-provided covariates (top right), independent surrogate and user-provided variables — ISVA (bottom left), and surrogate and user-provided variables — SVA (bottom right).

Key: black dots = probes with $p>1x10^{-5}$, red dots = probes passing nominal epigenome-wide threshold with $p<1x10^{-5}$.



Supplementary Figure S 6.13 Quantile-quantile plots of EWAS analysis of endometriosis ASRM stage III/IV cases (n=32) and clinic controls (n=38) in four models: without any covariates (top left), with two user-provided covariates (top right), independent surrogate and user-provided variables — ISVA (bottom left), and surrogate and user-provided variables — SVA (bottom right).

Key: black dots = probes with $p>1x10^{-5}$, red dots = probes passing nominal epigenome-wide threshold with $p<1x10^{-5}$.

E. Supplementary tables

Supplementary Table S 3.1 Preliminary association testing of selected 38 SNPs for overall endometriosis against three tiers of controls in ENDOX Phase I.

Chr	SNP	Position (hg19)	RA (RAF)	**Overall** Endometriosis Cases **(n=150)** in ENDOX Phase I **vs.**						Nearest Gene	Cytoband
				All Controls (n=4,930)		**Clinical Controls Only [in-house + biobank] (n=2,497)**		**In-house Clinical Controls Only (n=77)**			
				OR (95% CI)	**P value**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**		
1	rs12030576	115817221	G (0.65)	1.02 (0.80-1.30)	0.89	1.03 (0.81-1.32)	0.80	1.50 (1.01-2.24)	9.15×10^{-2}	*NGF*	1p13.2
1	rs2421985	172099136	C (0.49)	1.08 (0.86-1.35)	0.54	1.05 (0.83-1.33)	0.71	0.94 (0.63-1.38)	0.83	*DNM3*	1q24.3
2	rs11674184	11721535	T (0.61)	1.31 (1.02-1.67)	3.41×10^{-2}	1.24 (0.97-1.60)	9.48×10^{-2}	1.28 (0.85-1.93)	0.23	*GREB1*	2p25.1
2	rs1430787	67868498	A (0.31)	1.16 (0.91-1.47)	0.23	1.18 (0.92-1.51)	0.18	1.00 (0.66-1.51)	0.99	*ETAA1*	2p14
2	rs10167914	113529183	C (0.32)	1.10 (0.86-1.40)	0.45	1.12 (0.87-1.44)	0.36	1.26 (0.82-1.95)	0.23	*IL1A*	2q13
2	rs6435157	203441224	T (0.77)	1.09 (0.83-1.44)	0.48	1.10 (0.83-1.46)	0.49	0.96 (0.60-1.55)	0.80	*BMPR2*	2q33.1
2	rs1250247	216299629	C (0.28)	0.85 (0.65-1.11)	0.21	0.83 (0.64-1.09)	0.17	0.89 (0.57-1.38)	0.59	*FN1*	2q35
3	rs1352889	49652148	T (0.17)	0.79 (0.57-1.10)	0.13	0.78 (0.56-1.08)	9.17×10^{-2}	0.66 (0.40-1.10)	0.10	*BSN*	3p21.31
3	rs12631337	50198537	A (0.44)	0.89 (0.70-1.12)	0.31	0.89 (0.70-1.13)	0.31	0.87 (0.59-1.29)	0.43	*SEMA3F*	3p21.31
4	rs1903068	56008477	A (0.68)	1.37 (1.06-1.78)	1.38×10^{-2}	1.37 (1.05-1.79)	1.36×10^{-2}	1.31 (0.86-2.01)	0.29	*KDR*	4q12
4	rs2510770	95479372	A (0.32)	0.97 (0.76-1.24)	0.78	0.99 (0.77-1.27)	0.94	0.92 (0.61-1.39)	0.54	*PDLIM5*	4q22.3
6	rs6456259	19761718	G (0.16)	1.55 (1.16-2.05)	3.99×10^{-3}	1.49 (1.12-1.99)	7.73×10^{-3}	2.24 (1.25-4.01)	2.74×10^{-3}	*ID4*	6p22.3
6	rs4540228	74613060	C (0.66)	0.96 (0.75-1.22)	0.82	0.95 (0.75-1.22)	0.77	0.75 (0.49-1.14)	0.19	*CD109*	6q13
6	rs2226158	125995467	G (0.46)	0.88 (0.70-1.11)	0.28	0.89 (0.70-1.12)	0.34	0.99 (0.67-1.46)	0.98	*HEY2*	6q22.31
6	rs7740302	151826249	C (0.19)	1.13 (0.85-1.51)	0.39	1.16 (0.86-1.54)	0.33	1.39 (0.83-2.34)	0.17	*CCDC170*	6q25.1
6	rs71575922	152554014	G (0.15)	1.51 (1.15-2.00)	4.89×10^{-3}	1.47 (1.11-1.96)	9.25×10^{-3}	2.29 (1.29-4.07)	3.51×10^{-3}	*SYNE1*	6q25.1
6	rs11756073	170363580	A (0.17)	1.14 (0.84-1.54)	0.38	1.13 (0.84-1.54)	0.35	0.92 (0.56-1.53)	0.79	*FAM120B*	6q27
7	rs1451383	25891809	G (0.74)	1.05 (0.80-1.38)	0.72	1.06 (0.80-1.39)	0.71	1.19 (0.77-1.86)	0.53	—	7p15.2
7	rs6970537	27217359	G (0.57)	1.12 (0.89-1.42)	0.33	1.10 (0.87-1.40)	0.38	1.36 (0.92-2.02)	0.10	*HOXA10*	7p15.2
7	rs55909142	46673774	C (0.59)	1.12 (0.88-1.42)	0.33	1.14 (0.90-1.45)	0.22	0.87 (0.58-1.31)	0.47	—	7p12.3
8	rs17053711	25311269	G (0.73)	1.04 (0.80-1.36)	0.77	1.04 (0.80-1.36)	0.80	0.79 (0.49-1.26)	0.37	*KCTD9*	8p21.2
8	rs12549438	100150371	T (0.21)	1.12 (0.85-1.47)	0.49	1.12 (0.85-1.48)	0.48	0.89 (0.57-1.41)	0.47	*VPS13B*	8q22.2
9	rs10122243	22158924	T (0.41)	1.23 (0.98-1.55)	7.91×10^{-2}	1.24 (0.98-1.57)	6.49×10^{-2}	1.05 (0.71-1.55)	0.87	*CDKN2-BAS1*	9p21.3
9	rs507666	136149399	A (0.20)	1.14 (0.86-1.52)	0.40	1.13 (0.85-1.50)	0.41	0.81 (0.51-1.29)	0.34	*ABO*	9q34.2
10	rs10828249	21824727	A (0.35)	1.16 (0.91-1.47)	0.22	1.12 (0.88-1.42)	0.37	1.28 (0.85-1.93)	0.23	*MLLT10*	10p12.31
10	rs7907732	90182764	T (0.39)	1.16 (0.92-1.46)	0.22	1.13 (0.89-1.43)	0.31	0.98 (0.66-1.46)	0.96	*RNLS*	10q23.31
11	rs7924571	32350027	C (0.76)	1.44 (1.05-1.98)	1.60×10^{-2}	1.44 (1.05-1.98)	1.95×10^{-2}	0.99 (0.58-1.69)	0.88	*FSHB*	11p14.1
12	rs56090796	15406228	A (0.33)	1.15 (0.90-1.46)	0.26	1.08 (0.85-1.38)	0.49	1.02 (0.68-1.53)	0.67	*PTPRO*	12p12.3
12	rs3803042	54387947	A (0.42)	0.97 (0.77-1.23)	0.82	0.97 (0.76-1.22)	0.76	0.97 (0.66-1.44)	0.95	*HOXC10*	12p13.13
12	rs12320196	95645385	G (0.47)	1.30 (1.03-1.63)	2.25×10^{-2}	1.29 (1.02-1.63)	2.91×10^{-2}	1.64 (1.11-2.44)	1.12×10^{-2}	*VEZT*	12q22
13	rs7334326	50967732	C (0.16)	0.94 (0.67-1.31)	0.71	0.94 (0.67-1.32)	0.70	0.83 (0.49-1.43)	0.47	*DLEU1*	13q14.2
14	rs57281976	93112974	G (0.77)	1.16 (0.87-1.53)	0.30	1.13 (0.85-1.51)	0.37	1.09 (0.68-1.75)	0.61	*RIN3*	14q32.12
15	rs12441483	40353416	C (0.47)	1.25 (0.99-1.57)	5.30×10^{-2}	1.26 (1.00-1.59)	5.91×10^{-2}	1.33 (0.90-1.97)	0.22	*BMF*	15q15.1
17	rs66683298	46277748	C (0.60)	1.19 (0.93-1.51)	0.14	1.19 (0.93-1.53)	0.14	1.64 (1.10-2.43)	1.89×10^{-2}	*SKAP1*	17q21.32
17	rs7214750	63903119	C (0.93)	1.18 (0.73-1.91)	0.50	1.13 (0.69-1.85)	0.61	1.12 (0.51-2.49)	0.59	*CEP112*	17q24.1
X	rs13441059	70108889	A (0.36)	1.09 (0.86-1.38)	0.46	1.06 (0.84-1.35)	0.57	1.02 (0.68-1.52)	0.93	*TEX11*	Xq13.1
X	rs5933091	131278107	T (0.69)	1.15 (0.90-1.49)	0.28	1.21 (0.94-1.57)	0.12	1.18 (0.78-1.79)	0.32	*FRMD7*	Xq26.2
X	rs73241342	133691738	A (0.04)	0.86 (0.47-1.56)	0.58	0.86 (0.47-1.58)	0.60	1.48 (0.47-4.70)	0.33	*PLAC1*	Xq26.3

Supplementary Table S 3.2 Meta-analysed association results of 46 loci for overall endometriosis, and ASRM surgical classification of endometriosis by stages.

Chr	SNP	Position (hg19)	RA (RAF)	**Overall** Endometriosis **(n=361)**		**ASRM** Classification of Endometriosis						Nearest Gene	Cytoband
						Stages III&IV (n=180)		**Stages I&II (n=179)**		**Stages III&IV vs. I&II**			
				OR (95% CI)	**P value**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**		
1	rs10917151	22422721	A (0.17)	1.27 (0.97-1.66)	$7.71x10^{-2}$	1.47 (1.05-2.05)	$2.45x10^{-2}$	1.14 (0.84-1.56)	0.40	1.23 (0.77-1.96)	0.38	*WNT4*	1p36.12
1	rs12030576	115817221	G (0.65)	0.99 (0.80-1.22)	0.92	1.04 (0.78-1.37)	0.81	0.96 (0.75-1.22)	0.74	1.10 (0.75-1.61)	0.63	*NGF*	1p13.2
1	rs2040445	169216412	G (0.97)	0.80 (0.37-1.73)	0.58	0.60 (0.22-1.61)	0.31	0.80 (0.35-1.79)	0.58	1.08 (0.25-4.75)	0.91	*SLC19A2*	1q24.2
1	rs2421985	172099136	C (0.49)	1.06 (0.87-1.30)	0.55	1.13 (0.87-1.47)	0.34	0.96 (0.76-1.21)	0.74	1.20 (0.84-1.72)	0.32	*DNM3*	1q24.3
2	rs11674184	11721535	T (0.61)	1.14 (0.92-1.41)	0.22	1.48 (1.11-1.98)	$7.71x10^{-3}$	0.93 (0.74-1.19)	0.57	1.55 (1.06-2.29)	$2.51x10^{-2}$	*GREB1*	2p25.1
2	rs1430787	67868498	A (0.31)	1.05 (0.85-1.30)	0.65	0.97 (0.73-1.30)	0.86	1.19 (0.93-1.53)	0.16	0.78 (0.53-1.14)	0.20	*ETAA1*	2p14
2	rs10167914	113529183	C (0.32)	0.98 (0.79-1.22)	0.88	0.96 (0.73-1.27)	0.78	1.03 (0.81-1.32)	0.80	0.95 (0.65-1.40)	0.81	*IL1A*	2q13
2	rs6435157	203441224	T (0.77)	1.20 (0.94-1.52)	0.14	1.37 (0.99-1.90)	$5.79x10^{-2}$	1.04 (0.80-1.37)	0.75	1.28 (0.83-1.99)	0.27	*BMPR2*	2q33.1
2	rs1250247	216299629	C (0.28)	0.87 (0.70-1.09)	0.23	0.85 (0.64-1.15)	0.29	0.92 (0.71-1.19)	0.51	0.89 (0.59-1.34)	0.59	*FN1*	2q35
3	rs1352889	49652148	T (0.17)	0.99 (0.76-1.29)	0.96	0.93 (0.65-1.31)	0.66	1.00 (0.73-1.35)	0.98	0.90 (0.55-1.46)	0.67	*BSN*	3p21.31
3	rs12631337	50198537	A (0.44)	1.08 (0.88-1.31)	0.47	1.03 (0.79-1.34)	0.82	1.08 (0.86-1.36)	0.51	0.92 (0.64-1.33)	0.67	*SEMA3F*	3p21.31
4	rs1903068	56008477	A (0.68)	1.26 (1.01-1.57)	$3.65x10^{-2}$	1.40 (1.04-1.87)	$2.50x10^{-2}$	1.17 (0.91-1.50)	0.22	1.21 (0.81-1.81)	0.34	*KDR*	4q12
4	rs2510770	95479372	A (0.32)	0.93 (0.76-1.15)	0.51	0.96 (0.73-1.26)	0.74	0.91 (0.71-1.16)	0.45	1.09 (0.74-1.59)	0.68	*PDLIM5*	4q22.3
5	rs2946160	157891878	A (0.74)	1.03 (0.82-1.29)	0.79	1.00 (0.75-1.34)	0.99	1.04 (0.80-1.35)	0.78	0.98 (0.65-1.47)	0.91	*EBF1*	5q33.3
6	rs6456259	19761718	G (0.16)	1.22 (0.94-1.59)	0.13	1.43 (1.03-1.99)	$3.12x10^{-2}$	1.09 (0.80-1.49)	0.58	1.35 (0.85-2.15)	0.20	*ID4*	6p22.3
6	rs4540228	74613060	C (0.66)	1.03 (0.83-1.27)	0.80	1.18 (0.89-1.57)	0.24	0.90 (0.71-1.14)	0.38	1.28 (0.87-1.87)	0.21	*CD109*	6q13
6	rs2226158	125995467	G (0.46)	0.85 (0.69-1.03)	$9.93x10^{-2}$	0.90 (0.69-1.17)	0.43	0.81 (0.64-1.02)	$7.36x10^{-2}$	1.08 (0.75-1.55)	0.67	*HEY2*	6q22.31
6	rs7740302	151826249	C (0.19)	1.20 (0.93-1.54)	0.16	1.48 (1.08-2.02)	$1.36x10^{-2}$	0.93 (0.68-1.27)	0.64	1.54 (0.98-2.42)	$6.05x10^{-2}$	*CCDC170*	6q25.1
6	rs71575922	152554014	G (0.15)	1.61 (1.25-2.07)	$2.40x10^{-4}$	1.53 (1.09-2.13)	$1.27x10^{-2}$	1.76 (1.33-2.33)	$7.26x10^{-5}$	0.81 (0.52-1.25)	0.33	*SYNE1*	6q25.1
6	rs11756073	170363580	A (0.17)	1.03 (0.78-1.35)	0.84	1.20 (0.85-1.69)	0.31	0.86 (0.61-1.20)	0.37	1.44 (0.88-2.37)	0.15	*FAM120B*	6q27
7	rs1451383	25891809	G (0.74)	1.21 (0.96-1.54)	0.11	1.41 (1.02-1.95)	$4.01x10^{-2}$	1.00 (0.77-1.30)	1.00	1.40 (0.91-2.16)	0.13	—	7p15.2
7	rs6970537	27217359	G (0.57)	1.00 (0.81-1.22)	0.96	1.04 (0.80-1.36)	0.75	0.96 (0.76-1.22)	0.75	1.11 (0.77-1.59)	0.58	*HOXA10*	7p15.2
7	rs55909142	46673774	C (0.59)	1.04 (0.85-1.28)	0.67	1.16 (0.89-1.51)	0.29	0.97 (0.77-1.23)	0.83	1.15 (0.80-1.66)	0.45	—	7p12.3
8	rs17053711	25311269	G (0.73)	0.97 (0.77-1.22)	0.80	0.98 (0.73-1.31)	0.88	0.99 (0.76-1.29)	0.95	0.97 (0.65-1.46)	0.90	*KCTD9*	8p21.2
8	rs10090060	75257608	A (0.57)	1.02 (0.83-1.25)	0.87	1.06 (0.81-1.38)	0.69	1.01 (0.80-1.27)	0.96	1.10 (0.76-1.58)	0.62	*GDAP1*	8q21.11
8	rs12549438	100150371	T (0.21)	1.04 (0.81-1.33)	0.78	0.92 (0.66-1.29)	0.62	1.20 (0.91-1.58)	0.20	0.77 (0.49-1.20)	0.25	*VPS13B*	8q22.2
9	rs10122243	22158924	T (0.41)	1.11 (0.91-1.36)	0.29	1.25 (0.96-1.62)	$9.47x10^{-2}$	1.04 (0.82-1.31)	0.77	1.20 (0.83-1.72)	0.33	*CDKN2-BAS1*	9p21.3
9	rs10983311	119470373	T (0.77)	1.23 (0.96-1.58)	0.11	1.26 (0.91-1.76)	0.17	1.17 (0.88-1.56)	0.29	1.12 (0.71-1.78)	0.62	*ASTN2*	9q33.1
9	rs507666	136149399	A (0.20)	1.09 (0.84-1.40)	0.52	0.98 (0.70-1.38)	0.92	1.18 (0.88-1.57)	0.27	0.87 (0.55-1.37)	0.55	*ABO*	9q34.2
10	rs10828249	21824727	A (0.35)	1.14 (0.92-1.40)	0.23	1.12 (0.85-1.47)	0.42	1.16 (0.91-1.47)	0.23	0.97 (0.67-1.41)	0.89	*MLLT10*	10p12.31
10	rs7907732	90182764	T (0.39)	1.11 (0.91-1.36)	0.32	1.19 (0.91-1.55)	0.20	1.08 (0.85-1.37)	0.53	1.08 (0.75-1.56)	0.67	*RNLS*	10q23.31
11	rs3858429	30343757	C (0.84)	1.39 (1.04-1.86)	$2.71x10^{-2}$	1.75 (1.15-2.65)	$9.09x10^{-3}$	1.12 (0.81-1.54)	0.50	1.57 (0.90-2.73)	0.11	*ARL14EP*	11p14.1
11	rs7924571	32350027	C (0.76)	1.25 (0.97-1.62)	$8.14x10^{-2}$	1.12 (0.80-1.56)	0.50	1.28 (0.95-1.72)	0.11	1.03 (0.64-1.66)	0.91	*FSHB*	11p14.1
12	rs56090796	15406228	A (0.33)	1.17 (0.95-1.45)	0.14	1.18 (0.90-1.56)	0.23	1.18 (0.92-1.50)	0.19	0.96 (0.66-1.40)	0.82	*PTPRO*	12p12.3
12	rs3803042	54387947	A (0.42)	1.09 (0.89-1.34)	0.39	1.14 (0.88-1.49)	0.33	1.04 (0.83-1.32)	0.72	1.03 (0.72-1.49)	0.86	*HOXC10*	12p13.13
12	rs12320196	95645385	G (0.47)	1.12 (0.92-1.37)	0.26	1.20 (0.92-1.56)	0.17	1.07 (0.85-1.35)	0.56	1.16 (0.81-1.66)	0.42	*VEZT*	12q22
12	rs10860864	102822862	C (0.82)	0.94 (0.72-1.22)	0.65	1.06 (0.74-1.53)	0.74	0.80 (0.60-1.06)	0.12	1.45 (0.90-2.34)	0.13	*IGF1*	12q23.2
13	rs7334326	50967732	C (0.16)	0.98 (0.74-1.30)	0.89	0.88 (0.60-1.30)	0.51	1.08 (0.78-1.49)	0.66	0.83 (0.49-1.39)	0.48	*DLEU1*	13q14.2
14	rs57281976	93112974	G (0.77)	1.06 (0.84-1.34)	0.64	1.04 (0.76-1.41)	0.82	1.10 (0.83-1.44)	0.51	0.97 (0.63-1.49)	0.89	*RIN3*	14q32.12
15	rs12441483	40353416	C (0.47)	1.18 (0.97-1.44)	0.10	1.12 (0.86-1.45)	0.40	1.21 (0.96-1.53)	0.10	0.98 (0.68-1.40)	0.89	*BMF*	15q15.1
17	rs66683298	46277748	C (0.60)	1.12 (0.92-1.38)	0.26	1.17 (0.89-1.53)	0.25	1.06 (0.84-1.34)	0.62	1.14 (0.79-1.65)	0.49	*SKAP1*	17q21.32
17	rs7214750	63903119	C (0.93)	1.44 (0.95-2.19)	$8.48x10^{-2}$	1.98 (1.05-3.74)	$3.56x10^{-2}$	1.10 (0.70-1.72)	0.69	1.69 (0.75-3.77)	0.20	*CEP112*	17q24.1
19	rs2967684	8786624	A (0.16)	1.23 (0.94-1.62)	0.13	1.40 (0.99-1.97)	$5.57x10^{-2}$	1.15 (0.84-1.58)	0.38	1.15 (0.71-1.86)	0.57	*ACTL9*	19p13.2
X	rs13441059	70108889	A (0.36)	0.99 (0.81-1.22)	0.93	1.03 (0.78-1.34)	0.85	0.97 (0.76-1.24)	0.82	1.08 (0.74-1.57)	0.69	*TEX11*	Xq13.1
X	rs5933091	131278107	T (0.69)	1.10 (0.89-1.37)	0.38	1.31 (0.97-1.76)	$7.54x10^{-2}$	0.96 (0.75-1.24)	0.77	1.32 (0.89-1.97)	0.17	*FRMD7*	Xq26.2
X	rs73241342	133691738	A (0.04)	1.27 (0.79-2.03)	0.32	1.04 (0.55-1.96)	0.90	1.59 (0.89-2.84)	0.12	0.81 (0.30-2.19)	0.68	*PLAC1*	Xq26.3

Supplementary Table S 3.3 Meta-analysed association results of 46 loci for three surgical sub-phenotypes of endometriosis, and endometriosis with bowel movement/IBS pain.

Chr	SNP	Position (hg19)	RA (RAF)	Endometriosis with								Nearest Gene	Cytoband
				Superficial Lesion (n=186)		Deep Lesion (n=109)		Endometrioma (n=89)		Bowel Movement/ IBS Pain (n=106)			
				OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value		
1	rs10917151	22422721	A (0.17)	1.18 (0.88-1.59)	0.27	1.14 (0.72-1.81)	0.56	1.34 (0.87-2.06)	0.19	1.41 (0.98-2.03)	$6.10x10^{-2}$	*WNT4*	1p36.12
1	rs12030576	115817221	G (0.65)	0.92 (0.73-1.17)	0.51	1.06 (0.74-1.53)	0.74	1.08 (0.74-1.57)	0.68	1.07 (0.78-1.45)	0.68	*NGF*	1p13.2
1	rs2040445	169216412	G (0.97)	0.88 (0.39-2.00)	0.76	0.50 (0.17-1.49)	0.22	0.50 (0.06-3.90)	0.51	0.46 (0.20-1.02)	$5.54x10^{-2}$	*SLC19A2*	1q24.2
1	rs2421985	172099136	C (0.49)	0.99 (0.79-1.24)	0.92	1.35 (0.96-1.90)	$8.76x10^{-2}$	1.18 (0.84-1.67)	0.35	1.11 (0.84-1.48)	0.46	*DNM3*	1q24.3
2	rs11674184	11721535	T (0.61)	1.04 (0.82-1.32)	0.73	1.19 (0.83-1.72)	0.35	1.65 (1.11-2.45)	$1.26x10^{-2}$	1.11 (0.82-1.50)	0.50	*GREB1*	2p25.1
2	rs1430787	67868498	A (0.31)	1.03 (0.81-1.31)	0.81	1.06 (0.73-1.53)	0.76	0.90 (0.61-1.32)	0.58	1.28 (0.95-1.72)	0.11	*ETAA1*	2p14
2	rs10167914	113529183	C (0.32)	1.00 (0.79-1.28)	0.99	0.96 (0.66-1.39)	0.83	0.91 (0.62-1.34)	0.64	0.88 (0.64-1.20)	0.42	*IL1A*	2q13
2	rs6435157	203441224	T (0.77)	1.17 (0.89-1.54)	0.25	1.25 (0.82-1.91)	0.30	1.41 (0.90-2.20)	0.13	1.03 (0.74-1.45)	0.84	*BMPR2*	2q33.1
2	rs1250247	216299629	C (0.28)	0.89 (0.69-1.14)	0.35	0.71 (0.47-1.06)	$9.35x10^{-2}$	0.69 (0.45-1.05)	$8.43x10^{-2}$	1.04 (0.76-1.42)	0.80	*FN1*	2q35
3	rs1352889	49652148	T (0.17)	0.92 (0.68-1.25)	0.61	0.90 (0.56-1.44)	0.65	0.89 (0.55-1.44)	0.63	1.13 (0.79-1.64)	0.50	*BSN*	3p21.31
3	rs12631337	50198537	A (0.44)	1.05 (0.84-1.31)	0.68	0.95 (0.68-1.35)	0.79	1.00 (0.70-1.41)	0.99	1.09 (0.82-1.46)	0.56	*SEMA3F*	3p21.31
4	rs1903068	56008477	A (0.68)	1.44 (1.11-1.87)	$5.34x10^{-3}$	1.53 (1.03-2.26)	$3.38x10^{-2}$	1.61 (1.07-2.42)	$2.13x10^{-2}$	1.21 (0.89-1.65)	0.23	*KDR*	4q12
4	rs2510770	95479372	A (0.32)	0.92 (0.72-1.17)	0.50	0.88 (0.61-1.27)	0.50	0.91 (0.63-1.32)	0.61	0.83 (0.61-1.14)	0.25	*PDLIM5*	4q22.3
5	rs2946160	157891878	A (0.74)	1.04 (0.81-1.35)	0.74	1.17 (0.79-1.74)	0.44	1.14 (0.76-1.70)	0.54	1.31 (0.93-1.85)	0.12	*EBF1*	5q33.3
6	rs6456259	19761718	G (0.16)	1.14 (0.85-1.54)	0.39	1.14 (0.72-1.80)	0.57	1.27 (0.81-1.99)	0.30	1.29 (0.90-1.87)	0.17	*ID4*	6p22.3
6	rs4540228	74613060	C (0.66)	0.94 (0.74-1.18)	0.58	1.01 (0.71-1.45)	0.95	0.99 (0.69-1.42)	0.94	1.19 (0.87-1.63)	0.27	*CD109*	6q13
6	rs2226158	125995467	G (0.46)	0.86 (0.68-1.07)	0.18	0.94 (0.67-1.33)	0.72	0.95 (0.67-1.35)	0.79	0.92 (0.69-1.22)	0.55	*HEY2*	6q22.31
6	rs7740302	151826249	C (0.19)	0.86 (0.64-1.17)	0.35	1.40 (0.93-2.10)	0.11	1.33 (0.88-2.02)	0.18	1.05 (0.72-1.51)	0.81	*CCDC170*	6q25.1
6	rs71575922	152554014	G (0.15)	1.49 (1.12-1.98)	$5.50x10^{-3}$	1.76 (1.17-2.65)	$6.63x10^{-3}$	1.31 (0.84-2.05)	0.24	1.50 (1.05-2.15)	$2.64x10^{-2}$	*SYNE1*	6q25.1
6	rs11756073	170363580	A (0.17)	1.00 (0.73-1.36)	0.98	1.02 (0.64-1.64)	0.94	1.31 (0.84-2.03)	0.23	1.23 (0.85-1.78)	0.28	*FAM120B*	6q27
7	rs1451383	25891809	G (0.74)	1.07 (0.82-1.39)	0.61	1.15 (0.76-1.73)	0.51	1.17 (0.77-1.77)	0.47	1.17 (0.83-1.65)	0.37	—	7p15.2
7	rs6970537	27217359	G (0.57)	1.00 (0.80-1.25)	0.99	1.24 (0.87-1.76)	0.24	0.96 (0.68-1.36)	0.83	1.19 (0.89-1.59)	0.25	*HOXA10*	7p15.2
7	rs55909142	46673774	C (0.59)	1.03 (0.82-1.29)	0.80	1.20 (0.85-1.71)	0.30	1.15 (0.80-1.64)	0.45	1.07 (0.80-1.44)	0.63	—	7p12.3
8	rs17053711	25311269	G (0.73)	0.97 (0.75-1.25)	0.81	1.13 (0.76-1.69)	0.54	0.95 (0.64-1.40)	0.79	0.80 (0.58-1.09)	0.16	*KCTD9*	8p21.2
8	rs10090060	75257608	A (0.57)	1.06 (0.84-1.33)	0.63	1.23 (0.87-1.76)	0.24	1.15 (0.80-1.64)	0.45	0.98 (0.73-1.30)	0.87	*GDAP1*	8q21.11
8	rs12549438	100150371	T (0.21)	1.08 (0.82-1.42)	0.59	1.03 (0.67-1.58)	0.89	0.98 (0.64-1.52)	0.94	1.23 (0.88-1.73)	0.23	*VPS13B*	8q22.2
9	rs10122243	22158924	T (0.41)	1.02 (0.81-1.28)	0.89	1.39 (0.99-1.96)	$5.61x10^{-2}$	1.16 (0.82-1.64)	0.40	0.95 (0.71-1.28)	0.75	*CDKN2-BAS1*	9p21.3
9	rs10983311	119470373	T (0.77)	1.34 (1.00-1.80)	$4.89x10^{-2}$	1.16 (0.76-1.78)	0.50	1.80 (1.09-2.97)	$2.25x10^{-2}$	1.15 (0.81-1.64)	0.44	*ASTN2*	9q33.1
9	rs507666	136149399	A (0.20)	1.18 (0.89-1.56)	0.24	1.25 (0.82-1.90)	0.29	1.22 (0.80-1.86)	0.35	1.20 (0.84-1.71)	0.32	*ABO*	9q34.2
10	rs10828249	21824727	A (0.35)	1.05 (0.83-1.33)	0.70	1.10 (0.77-1.57)	0.59	1.03 (0.71-1.48)	0.88	1.11 (0.83-1.50)	0.48	*MLLT10*	10p12.31
10	rs7907732	90182764	T (0.39)	1.08 (0.86-1.36)	0.51	1.43 (1.02-2.02)	$3.89x10^{-2}$	1.06 (0.75-1.51)	0.74	1.18 (0.88-1.57)	0.27	*RNLS*	10q23.31
11	rs3858429	30343757	C (0.84)	1.30 (0.93-1.82)	0.12	1.64 (0.95-2.84)	$7.67x10^{-2}$	1.72 (0.96-3.05)	$6.65x10^{-2}$	1.18 (0.79-1.78)	0.42	*ARL14EP*	11p14.1
11	rs7924571	32350027	C (0.76)	1.27 (0.95-1.70)	0.11	1.52 (0.94-2.45)	$8.50x10^{-2}$	1.66 (1.01-2.74)	$4.64x10^{-2}$	1.41 (0.96-2.07)	$7.83x10^{-2}$	*FSHB*	11p14.1
12	rs56090796	15406228	A (0.33)	0.98 (0.77-1.25)	0.85	1.07 (0.74-1.54)	0.72	1.00 (0.69-1.45)	0.99	1.00 (0.73-1.37)	0.99	*PTPRO*	12p12.3
12	rs3803042	54387947	A (0.42)	0.97 (0.77-1.22)	0.79	1.17 (0.83-1.65)	0.37	0.90 (0.63-1.28)	0.57	0.99 (0.74-1.33)	0.95	*HOXC10*	12p13.13
12	rs12320196	95645385	G (0.47)	1.19 (0.95-1.49)	0.14	0.92 (0.65-1.29)	0.62	1.28 (0.91-1.81)	0.16	0.99 (0.75-1.33)	0.97	*VEZT*	12q22
12	rs10860864	102822862	C (0.82)	0.84 (0.63-1.12)	0.23	1.14 (0.71-1.84)	0.59	1.16 (0.72-1.88)	0.55	1.12 (0.75-1.67)	0.57	*IGF1*	12q23.2
13	rs7334326	50967732	C (0.16)	1.03 (0.75-1.42)	0.84	1.15 (0.72-1.83)	0.56	0.92 (0.55-1.53)	0.74	1.18 (0.80-1.74)	0.41	*DLEU1*	13q14.2
14	rs57281976	93112974	G (0.77)	1.13 (0.86-1.49)	0.37	0.94 (0.63-1.39)	0.74	1.15 (0.76-1.75)	0.52	0.98 (0.70-1.36)	0.88	*RIN3*	14q32.12
15	rs12441483	40353416	C (0.47)	1.28 (1.02-1.60)	$3.31x10^{-2}$	1.15 (0.82-1.62)	0.41	1.72 (1.21-2.44)	$2.55x10^{-3}$	1.43 (1.07-1.90)	$1.54x10^{-2}$	*BMF*	15q15.1
17	rs66683298	46277748	C (0.60)	1.08 (0.86-1.36)	0.50	1.13 (0.79-1.61)	0.51	1.27 (0.88-1.83)	0.20	1.09 (0.81-1.46)	0.58	*SKAP1*	17q21.32
17	rs7214750	63903119	C (0.93)	1.16 (0.74-1.82)	0.51	1.62 (0.71-3.69)	0.26	1.29 (0.58-2.87)	0.53	1.28 (0.68-2.40)	0.44	*CEP112*	17q24.1
19	rs2967684	8786624	A (0.16)	1.15 (0.84-1.56)	0.38	1.47 (0.95-2.27)	$8.16x10^{-2}$	1.21 (0.76-1.92)	0.42	1.16 (0.78-1.71)	0.47	*ACTL9*	19p13.2
X	rs13441059	70108889	A (0.36)	1.00 (0.79-1.26)	0.97	1.09 (0.77-1.54)	0.64	1.02 (0.72-1.46)	0.90	1.19 (0.89-1.59)	0.25	*TEX11*	Xq13.1
X	rs5933091	131278107	T (0.69)	1.07 (0.84-1.36)	0.60	1.18 (0.81-1.72)	0.40	1.47 (0.98-2.21)	$6.23x10^{-2}$	1.09 (0.79-1.49)	0.60	*FRMD7*	Xq26.2
X	rs73241342	133691738	A (0.04)	1.33 (0.79-2.25)	0.28	1.71 (0.86-3.39)	0.12	1.12 (0.48-2.62)	0.80	1.08 (0.54-2.14)	0.83	*PLAC1*	Xq26.3

Supplementary Table S 3.4 Meta-analysed association results of 46 loci for endometriosis with dysmenorrhea by severity.

Chr	SNP	Position (hg19)	RA (RAF)	Endometriosis with **Dysmenorrhea**								Nearest Gene	Cytoband
				All (n=213)		**Severe (NRS 7-10) Recent (n=133)**		**Moderate to Severe (4-10) Recent (n=167)**		**Mild to Moderate (1-6) Recent (n=47)**			
				OR (95% CI)	**P value**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**		
1	rs10917151	22422721	A (0.17)	1.15 (0.86-1.53)	0.35	1.34 (0.96-1.86)	$8.54x10^{-2}$	1.28 (0.93-1.76)	0.13	1.09 (0.62-1.94)	0.76	*WNT4*	1p36.12
1	rs12030576	115817221	G (0.65)	0.95 (0.76-1.19)	0.65	0.93 (0.71-1.22)	0.59	0.94 (0.72-1.21)	0.62	1.02 (0.65-1.60)	0.92	*NGF*	1p13.2
1	rs2040445	169216412	G (0.97)	0.84 (0.38-1.88)	0.68	0.69 (0.28-1.73)	0.43	0.78 (0.33-1.87)	0.58	0.79 (0.18-3.38)	0.75	*SLC19A2*	1q24.2
1	rs2421985	172099136	C (0.49)	1.02 (0.83-1.26)	0.85	1.08 (0.83-1.40)	0.56	1.06 (0.83-1.36)	0.64	0.79 (0.51-1.20)	0.27	*DNM3*	1q24.3
2	rs11674184	11721535	T (0.61)	1.08 (0.86-1.35)	0.52	1.16 (0.88-1.53)	0.29	1.09 (0.84-1.42)	0.51	0.82 (0.53-1.26)	0.36	*GREB1*	2p25.1
2	rs1430787	67868498	A (0.31)	1.05 (0.84-1.32)	0.67	1.22 (0.93-1.60)	0.16	1.14 (0.88-1.48)	0.32	0.87 (0.54-1.40)	0.55	*ETAA1*	2p14
2	rs10167914	113529183	C (0.32)	0.97 (0.77-1.22)	0.81	1.09 (0.83-1.44)	0.52	1.03 (0.79-1.34)	0.83	0.81 (0.51-1.30)	0.39	*IL1A*	2q13
2	rs6435157	203441224	T (0.77)	1.27 (0.98-1.65)	$7.31x10^{-2}$	1.22 (0.89-1.68)	0.22	1.20 (0.89-1.63)	0.23	1.25 (0.75-2.11)	0.39	*BMPR2*	2q33.1
2	rs1250247	216299629	C (0.28)	0.85 (0.67-1.08)	0.19	1.04 (0.78-1.38)	0.79	0.93 (0.71-1.23)	0.62	0.65 (0.39-1.09)	0.10	*FN1*	2q35
3	rs1352889	49652148	T (0.17)	0.98 (0.74-1.30)	0.87	1.03 (0.73-1.44)	0.89	1.06 (0.77-1.46)	0.73	1.02 (0.59-1.76)	0.95	*BSN*	3p21.31
3	rs12631337	50198537	A (0.44)	1.01 (0.82-1.25)	0.91	1.15 (0.89-1.49)	0.29	1.09 (0.85-1.40)	0.48	0.86 (0.56-1.32)	0.50	*SEMA3F*	3p21.31
4	rs1903068	56008477	A (0.68)	1.32 (1.04-1.67)	$2.07x10^{-2}$	1.35 (1.01-1.80)	$4.33x10^{-2}$	1.35 (1.02-1.78)	$3.42x10^{-2}$	1.50 (0.93-2.43)	$9.66x10^{-2}$	*KDR*	4q12
4	rs2510770	95479372	A (0.32)	0.89 (0.71-1.12)	0.32	0.94 (0.71-1.23)	0.64	0.98 (0.75-1.27)	0.85	1.17 (0.76-1.80)	0.49	*PDLIM5*	4q22.3
5	rs2946160	157891878	A (0.74)	1.09 (0.85-1.39)	0.49	1.08 (0.81-1.46)	0.59	1.10 (0.83-1.46)	0.52	0.99 (0.62-1.59)	0.98	*EBF1*	5q33.3
6	rs6456259	19761718	G (0.16)	1.19 (0.90-1.58)	0.21	1.13 (0.80-1.59)	0.51	1.10 (0.79-1.54)	0.56	1.18 (0.67-2.07)	0.57	*ID4*	6p22.3
6	rs4540228	74613060	C (0.66)	0.90 (0.72-1.13)	0.36	0.86 (0.65-1.12)	0.26	0.89 (0.68-1.15)	0.36	0.90 (0.58-1.39)	0.63	*CD109*	6q13
6	rs2226158	125995467	G (0.46)	0.92 (0.74-1.14)	0.45	0.99 (0.76-1.28)	0.93	0.95 (0.74-1.21)	0.67	0.79 (0.52-1.22)	0.29	*HEY2*	6q22.31
6	rs7740302	151826249	C (0.19)	1.16 (0.89-1.52)	0.28	1.03 (0.74-1.44)	0.87	1.05 (0.76-1.44)	0.78	1.03 (0.60-1.77)	0.91	*CCDC170*	6q25.1
6	rs71575922	152554014	G (0.15)	1.58 (1.21-2.06)	$7.46x10^{-4}$	1.47 (1.07-2.04)	$1.92x10^{-2}$	1.57 (1.16-2.13)	$3.76x10^{-3}$	2.01 (1.24-3.26)	$4.72x10^{-3}$	*SYNE1*	6q25.1
6	rs11756073	170363580	A (0.17)	1.02 (0.77-1.37)	0.87	1.03 (0.72-1.47)	0.86	0.95 (0.67-1.35)	0.78	0.76 (0.40-1.44)	0.40	*FAM120B*	6q27
7	rs1451383	25891809	G (0.74)	1.11 (0.87-1.43)	0.40	1.03 (0.76-1.39)	0.86	1.03 (0.78-1.38)	0.82	1.00 (0.61-1.63)	1.00	—	7p15.2
7	rs6970537	27217359	G (0.57)	1.02 (0.82-1.26)	0.87	0.97 (0.75-1.26)	0.81	1.02 (0.79-1.31)	0.88	1.28 (0.83-1.97)	0.27	*HOXA10*	7p15.2
7	rs55909142	46673774	C (0.59)	1.06 (0.85-1.32)	0.61	1.02 (0.78-1.32)	0.90	1.11 (0.86-1.44)	0.41	1.43 (0.92-2.24)	0.11	—	7p12.3
8	rs17053711	25311269	G (0.73)	0.95 (0.75-1.21)	0.67	0.97 (0.72-1.30)	0.82	0.92 (0.70-1.21)	0.53	0.95 (0.59-1.53)	0.84	*KCTD9*	8p21.2
8	rs10090060	75257608	A (0.57)	0.99 (0.80-1.24)	0.96	0.94 (0.73-1.23)	0.67	1.02 (0.80-1.32)	0.85	1.06 (0.69-1.63)	0.79	*GDAP1*	8q21.11
8	rs12549438	100150371	T (0.21)	1.01 (0.77-1.32)	0.94	1.06 (0.77-1.45)	0.74	0.96 (0.70-1.31)	0.80	0.85 (0.49-1.48)	0.58	*VPS13B*	8q22.2
9	rs10122243	22158924	T (0.41)	1.08 (0.87-1.33)	0.50	1.10 (0.85-1.43)	0.46	1.02 (0.80-1.31)	0.87	0.87 (0.56-1.33)	0.51	*CDKN2-BAS1*	9p21.3
9	rs10983311	119470373	T (0.77)	1.41 (1.07-1.86)	$1.61x10^{-2}$	1.45 (1.03-2.06)	$3.44x10^{-2}$	1.51 (1.09-2.11)	$1.46x10^{-2}$	1.50 (0.85-2.64)	0.16	*ASTN2*	9q33.1
9	rs507666	136149399	A (0.20)	1.26 (0.97-1.63)	$8.44x10^{-2}$	1.37 (1.00-1.86)	$4.69x10^{-2}$	1.33 (0.99-1.79)	$6.09x10^{-2}$	1.32 (0.80-2.18)	0.29	*ABO*	9q34.2
10	rs10828249	21824727	A (0.35)	1.11 (0.89-1.38)	0.37	1.22 (0.93-1.59)	0.15	1.24 (0.96-1.59)	$9.66x10^{-2}$	1.07 (0.69-1.65)	0.78	*MLLT10*	10p12.31
10	rs7907732	90182764	T (0.39)	1.08 (0.87-1.34)	0.50	1.05 (0.80-1.36)	0.74	1.13 (0.88-1.45)	0.34	1.34 (0.88-2.04)	0.18	*RNLS*	10q23.31
11	rs3858429	30343757	C (0.84)	1.26 (0.92-1.72)	0.15	1.07 (0.74-1.53)	0.73	1.19 (0.83-1.70)	0.34	1.50 (0.72-3.13)	0.28	*ARL14EP*	11p14.1
11	rs7924571	32350027	C (0.76)	1.27 (0.96-1.67)	$9.04x10^{-2}$	1.29 (0.92-1.81)	0.14	1.28 (0.93-1.77)	0.13	1.44 (0.81-2.55)	0.22	*FSHB*	11p14.1
12	rs56090796	15406228	A (0.33)	1.11 (0.89-1.39)	0.36	1.24 (0.94-1.62)	0.13	1.22 (0.94-1.57)	0.14	0.96 (0.61-1.51)	0.85	*PTPRO*	12p12.3
12	rs3803042	54387947	A (0.42)	0.98 (0.79-1.21)	0.83	1.05 (0.81-1.36)	0.72	1.01 (0.79-1.30)	0.92	0.86 (0.55-1.32)	0.48	*HOXC10*	12p13.13
12	rs12320196	95645385	G (0.47)	1.08 (0.87-1.33)	0.50	1.05 (0.81-1.36)	0.71	1.11 (0.87-1.42)	0.39	1.41 (0.92-2.15)	0.11	*VEZT*	12q22
12	rs10860864	102822862	C (0.82)	0.95 (0.72-1.27)	0.75	1.03 (0.73-1.46)	0.87	0.98 (0.70-1.35)	0.88	0.74 (0.44-1.24)	0.25	*IGF1*	12q23.2
13	rs7334326	50967732	C (0.16)	1.10 (0.82-1.47)	0.55	1.16 (0.82-1.65)	0.41	1.25 (0.90-1.74)	0.18	1.18 (0.67-2.08)	0.56	*DLEU1*	13q14.2
14	rs57281976	93112974	G (0.77)	1.12 (0.87-1.45)	0.38	1.03 (0.76-1.40)	0.84	1.12 (0.83-1.50)	0.45	1.40 (0.81-2.42)	0.23	*RIN3*	14q32.12
15	rs12441483	40353416	C (0.47)	1.37 (1.11-1.69)	$3.96x10^{-3}$	1.32 (1.02-1.71)	$3.40x10^{-2}$	1.27 (0.99-1.63)	$5.51x10^{-2}$	1.12 (0.74-1.70)	0.59	*BMF*	15q15.1
17	rs66683298	46277748	C (0.60)	1.09 (0.87-1.35)	0.45	1.14 (0.87-1.48)	0.35	1.10 (0.85-1.42)	0.46	0.99 (0.65-1.53)	0.98	*SKAP1*	17q21.32
17	rs7214750	63903119	C (0.93)	1.24 (0.80-1.92)	0.33	1.10 (0.66-1.84)	0.71	1.14 (0.70-1.87)	0.60	1.29 (0.43-3.86)	0.66	*CEP112*	17q24.1
19	rs2967684	8786624	A (0.16)	1.21 (0.91-1.61)	0.20	1.18 (0.83-1.67)	0.36	1.25 (0.90-1.74)	0.18	1.32 (0.75-2.32)	0.34	*ACTL9*	19p13.2
X	rs13441059	70108889	A (0.36)	1.01 (0.81-1.27)	0.90	0.98 (0.75-1.28)	0.88	1.04 (0.81-1.35)	0.74	1.23 (0.80-1.88)	0.34	*TEX11*	Xq13.1
X	rs5933091	131278107	T (0.69)	1.14 (0.90-1.45)	0.26	1.22 (0.91-1.63)	0.19	1.14 (0.87-1.50)	0.34	0.84 (0.54-1.31)	0.44	*FRMD7*	Xq26.2
X	rs73241342	133691738	A (0.04)	1.03 (0.61-1.74)	0.91	1.11 (0.60-2.04)	0.74	1.09 (0.60-1.96)	0.78	1.08 (0.38-3.09)	0.89	*PLAC1*	Xq26.3

Supplementary Table S 3.5 Meta-analysed association results of 46 loci for endometriosis with dysmenorrhea by age of onset, and endometriosis with infertility.

Chr	SNP	Position (hg19)	RA (RAF)	Endometriosis with **Dysmenorrhea**				Endometriosis with **Infertility**				Nearest Gene	Cytoband
				Age of Onset < 18 (n=119)		**Age of Onset ≥ 18 (n=49)**		**Try to conceive ≥ 12 months (n=74)**		**Try to conceive ≥ 6 months (n=140)**			
				OR (95% CI)	**P value**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**		
1	rs10917151	22422721	A (0.17)	0.97 (0.63-1.49)	0.90	1.60 (0.92-2.80)	$9.88x10^{-2}$	1.21 (0.78-1.87)	0.40	1.44 (1.01-2.05)	$4.53x10^{-2}$	*WNT4*	1p36.12
1	rs12030576	115817221	G (0.65)	1.00 (0.73-1.37)	0.99	1.06 (0.65-1.73)	0.81	0.93 (0.66-1.32)	0.68	1.07 (0.79-1.45)	0.68	*NGF*	1p13.2
1	rs2040445	169216412	G (0.97)	0.87 (0.27-2.80)	0.82	0.65 (0.15-2.88)	0.57	1.24 (0.30-5.18)	0.77	1.68 (0.42-6.66)	0.46	*SLC19A2*	1q24.2
1	rs2421985	172099136	C (0.49)	1.07 (0.79-1.43)	0.67	0.82 (0.52-1.29)	0.39	1.07 (0.77-1.49)	0.70	1.02 (0.77-1.35)	0.90	*DNM3*	1q24.3
2	rs11674184	11721535	T (0.61)	0.92 (0.68-1.25)	0.59	1.46 (0.88-2.42)	0.15	0.88 (0.63-1.24)	0.48	1.12 (0.83-1.51)	0.45	*GREB1*	2p25.1
2	rs1430787	67868498	A (0.31)	1.20 (0.88-1.64)	0.25	0.73 (0.43-1.25)	0.25	1.12 (0.78-1.60)	0.55	1.10 (0.81-1.49)	0.55	*ETAA1*	2p14
2	rs10167914	113529183	C (0.32)	0.90 (0.65-1.24)	0.51	1.15 (0.71-1.86)	0.57	0.97 (0.67-1.39)	0.86	1.10 (0.81-1.48)	0.56	*IL1A*	2q13
2	rs6435157	203441224	T (0.77)	1.33 (0.92-1.93)	0.13	1.14 (0.66-1.97)	0.64	1.55 (0.99-2.40)	$5.31x10^{-2}$	1.16 (0.82-1.65)	0.39	*BMPR2*	2q33.1
2	rs1250247	216299629	C (0.28)	0.93 (0.67-1.29)	0.65	0.60 (0.33-1.09)	$9.54x10^{-2}$	0.90 (0.62-1.32)	0.60	0.91 (0.66-1.25)	0.54	*FN1*	2q35
3	rs1352889	49652148	T (0.17)	1.12 (0.77-1.63)	0.56	0.72 (0.37-1.40)	0.34	0.76 (0.48-1.23)	0.27	0.99 (0.67-1.45)	0.94	*BSN*	3p21.31
3	rs12631337	50198537	A (0.44)	1.14 (0.85-1.53)	0.40	0.71 (0.44-1.13)	0.15	1.10 (0.79-1.54)	0.56	1.04 (0.78-1.38)	0.78	*SEMA3F*	3p21.31
4	rs1903068	56008477	A (0.68)	1.29 (0.93-1.78)	0.13	1.22 (0.74-2.02)	0.44	1.45 (0.99-2.13)	$5.51x10^{-2}$	1.29 (0.94-1.77)	0.12	*KDR*	4q12
4	rs2510770	95479372	A (0.32)	0.87 (0.63-1.19)	0.38	0.81 (0.49-1.32)	0.39	0.90 (0.63-1.29)	0.58	0.86 (0.64-1.17)	0.35	*PDLIM5*	4q22.3
5	rs2946160	157891878	A (0.74)	1.04 (0.75-1.46)	0.80	1.11 (0.66-1.86)	0.70	0.88 (0.61-1.26)	0.48	0.99 (0.72-1.37)	0.97	*EBF1*	5q33.3
6	rs6456259	19761718	G (0.16)	1.13 (0.76-1.67)	0.54	1.45 (0.83-2.56)	0.19	0.95 (0.58-1.54)	0.84	1.21 (0.82-1.77)	0.34	*ID4*	6p22.3
6	rs4540228	74613060	C (0.66)	0.88 (0.65-1.19)	0.41	1.01 (0.63-1.64)	0.96	0.91 (0.64-1.28)	0.58	0.92 (0.68-1.24)	0.58	*CD109*	6q13
6	rs2226158	125995467	G (0.46)	0.85 (0.63-1.15)	0.29	1.08 (0.69-1.70)	0.74	0.90 (0.65-1.26)	0.56	0.78 (0.58-1.04)	$8.78x10^{-2}$	*HEY2*	6q22.31
6	rs7740302	151826249	C (0.19)	1.14 (0.79-1.66)	0.49	1.28 (0.73-2.23)	0.38	1.30 (0.87-1.95)	0.20	1.18 (0.83-1.69)	0.35	*CCDC170*	6q25.1
6	rs71575922	152554014	G (0.15)	1.60 (1.11-2.31)	$1.19x10^{-2}$	1.76 (1.02-3.03)	$4.30x10^{-2}$	1.92 (1.30-2.83)	$1.12x10^{-3}$	1.54 (1.08-2.19)	$1.59x10^{-2}$	*SYNE1*	6q25.1
6	rs11756073	170363580	A (0.17)	1.02 (0.68-1.53)	0.92	1.18 (0.64-2.17)	0.60	0.97 (0.61-1.54)	0.89	0.98 (0.66-1.46)	0.92	*FAM120B*	6q27
7	rs1451383	25891809	G (0.74)	1.12 (0.79-1.59)	0.52	1.27 (0.73-2.22)	0.40	1.42 (0.93-2.17)	0.10	1.14 (0.81-1.60)	0.46	—	7p15.2
7	rs6970537	27217359	G (0.57)	0.91 (0.67-1.22)	0.52	1.18 (0.74-1.87)	0.50	1.03 (0.73-1.45)	0.86	0.97 (0.73-1.30)	0.86	*HOXA10*	7p15.2
7	rs55909142	46673774	C (0.59)	0.99 (0.73-1.33)	0.94	1.01 (0.64-1.61)	0.96	0.93 (0.66-1.30)	0.66	1.14 (0.85-1.52)	0.39	—	7p12.3
8	rs17053711	25311269	G (0.73)	0.92 (0.65-1.28)	0.61	0.81 (0.48-1.36)	0.43	1.03 (0.70-1.52)	0.87	0.90 (0.65-1.23)	0.51	*KCTD9*	8p21.2
8	rs10090060	75257608	A (0.57)	1.01 (0.75-1.36)	0.97	0.88 (0.55-1.38)	0.57	0.80 (0.57-1.11)	0.18	0.98 (0.74-1.31)	0.92	*GDAP1*	8q21.11
8	rs12549438	100150371	T (0.21)	1.08 (0.75-1.56)	0.68	0.92 (0.51-1.66)	0.78	1.07 (0.71-1.62)	0.75	1.16 (0.82-1.64)	0.40	*VPS13B*	8q22.2
9	rs10122243	22158924	T (0.41)	1.06 (0.79-1.42)	0.71	1.16 (0.73-1.84)	0.52	1.03 (0.74-1.44)	0.86	1.13 (0.85-1.50)	0.42	*CDKN2-BAS1*	9p21.3
9	rs10983311	119470373	T (0.77)	1.36 (0.93-1.99)	0.12	2.53 (1.20-5.31)	$1.44x10^{-2}$	1.30 (0.84-2.02)	0.24	1.39 (0.95-2.03)	$9.23x10^{-2}$	*ASTN2*	9q33.1
9	rs507666	136149399	A (0.20)	1.40 (0.98-1.99)	$6.45x10^{-2}$	1.18 (0.67-2.08)	0.56	1.20 (0.80-1.81)	0.38	1.40 (1.00-1.97)	$5.16x10^{-2}$	*ABO*	9q34.2
10	rs10828249	21824727	A (0.35)	1.07 (0.79-1.46)	0.66	0.95 (0.59-1.54)	0.84	1.18 (0.84-1.66)	0.35	1.16 (0.86-1.56)	0.32	*MLLT10*	10p12.31
10	rs7907732	90182764	T (0.39)	1.02 (0.76-1.38)	0.88	1.42 (0.90-2.23)	0.13	0.92 (0.65-1.29)	0.62	1.22 (0.91-1.62)	0.18	*RNLS*	10q23.31
11	rs3858429	30343757	C (0.84)	1.35 (0.87-2.09)	0.18	1.17 (0.61-2.22)	0.64	1.27 (0.75-2.13)	0.37	1.41 (0.92-2.18)	0.12	*ARL14EP*	11p14.1
11	rs7924571	32350027	C (0.76)	1.36 (0.92-2.00)	0.12	1.14 (0.64-2.03)	0.65	1.20 (0.79-1.84)	0.40	1.29 (0.89-1.87)	0.18	*FSHB*	11p14.1
12	rs56090796	15406228	A (0.33)	1.05 (0.76-1.44)	0.78	1.07 (0.66-1.75)	0.77	1.34 (0.95-1.90)	$9.17x10^{-2}$	1.12 (0.83-1.51)	0.47	*PTPRO*	12p12.3
12	rs3803042	54387947	A (0.42)	0.89 (0.66-1.21)	0.45	1.02 (0.65-1.63)	0.92	0.96 (0.68-1.35)	0.80	0.97 (0.73-1.30)	0.84	*HOXC10*	12p13.13
12	rs12320196	95645385	G (0.47)	1.08 (0.80-1.45)	0.62	0.98 (0.61-1.55)	0.92	1.17 (0.84-1.63)	0.36	1.05 (0.79-1.40)	0.73	*VEZT*	12q22
12	rs10860864	102822862	C (0.82)	0.87 (0.60-1.28)	0.49	0.91 (0.50-1.66)	0.77	1.10 (0.68-1.77)	0.69	0.80 (0.56-1.14)	0.22	*IGF1*	12q23.2
13	rs7334326	50967732	C (0.16)	0.84 (0.54-1.31)	0.45	1.68 (0.97-2.93)	$6.47x10^{-2}$	1.18 (0.75-1.85)	0.47	1.13 (0.76-1.67)	0.55	*DLEU1*	13q14.2
14	rs57281976	93112974	G (0.77)	0.87 (0.63-1.22)	0.43	1.13 (0.63-2.04)	0.67	1.22 (0.81-1.83)	0.35	1.13 (0.80-1.59)	0.49	*RIN3*	14q32.12
15	rs12441483	40353416	C (0.47)	1.53 (1.14-2.06)	$4.60x10^{-3}$	1.14 (0.71-1.81)	0.59	1.54 (1.10-2.16)	$1.15x10^{-2}$	1.35 (1.02-1.80)	$3.67x10^{-2}$	*BMF*	15q15.1
17	rs66683298	46277748	C (0.60)	0.94 (0.70-1.26)	0.68	1.03 (0.65-1.64)	0.90	1.01 (0.72-1.41)	0.97	1.01 (0.76-1.35)	0.93	*SKAP1*	17q21.32
17	rs7214750	63903119	C (0.93)	0.85 (0.51-1.44)	0.55	1.87 (0.35-9.89)	0.46	0.98 (0.52-1.84)	0.95	1.51 (0.79-2.90)	0.21	*CEP112*	17q24.1
19	rs2967684	8786624	A (0.16)	1.20 (0.80-1.79)	0.38	1.19 (0.64-2.21)	0.58	1.32 (0.85-2.05)	0.22	1.33 (0.92-1.93)	0.13	*ACTL9*	19p13.2
X	rs13441059	70108889	A (0.36)	0.99 (0.73-1.34)	0.94	1.28 (0.81-2.02)	0.30	0.95 (0.67-1.34)	0.77	1.03 (0.77-1.38)	0.84	*TEX11*	Xq13.1
X	rs5933091	131278107	T (0.69)	0.99 (0.72-1.36)	0.93	1.07 (0.64-1.79)	0.80	1.14 (0.79-1.65)	0.48	0.95 (0.70-1.28)	0.73	*FRMD7*	Xq26.2
X	rs73241342	133691738	A (0.04)	0.89 (0.41-1.92)	0.77	0.77 (0.21-2.76)	0.68	1.45 (0.72-2.92)	0.30	0.95 (0.45-2.01)	0.90	*PLAC1*	Xq26.3

Supplementary Table S 3.6 Meta-analysed association results of 46 loci for endometriosis with dyspareunia by severity and temporality (part A).

Chr	SNP	Position (hg19)	RA (RAF)	Endometriosis with **Dyspareunia**								Nearest Gene	Cytoband
				All (n=154)		Severe (NRS 7-10)				Moderate to Severe (4-10)			
						Recent (n=50)		At worse in lifetime (n=77)		Recent (n=103)			
				OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value		
1	rs10917151	22422721	A (0.17)	1.18 (0.86-1.63)	0.31	1.06 (0.57-1.96)	0.86	1.26 (0.79-2.00)	0.34	1.28 (0.87-1.86)	0.21	*WNT4*	1p36.12
1	rs12030576	115817221	G (0.65)	0.88 (0.69-1.13)	0.33	0.87 (0.55-1.37)	0.54	1.05 (0.73-1.50)	0.80	0.86 (0.64-1.16)	0.33	*NGF*	1p13.2
1	rs2040445	169216412	G (0.97)	0.66 (0.29-1.51)	0.33	0.56 (0.15-2.11)	0.39	1.43 (0.19-10.81)	0.73	0.79 (0.29-2.16)	0.65	*SLC19A2*	1q24.2
1	rs2421985	172099136	C (0.49)	1.07 (0.84-1.36)	0.57	1.31 (0.84-2.04)	0.24	1.07 (0.76-1.50)	0.70	1.24 (0.93-1.65)	0.14	*DNM3*	1q24.3
2	rs11674184	11721535	T (0.61)	1.05 (0.82-1.35)	0.71	1.15 (0.71-1.86)	0.58	1.03 (0.72-1.48)	0.86	1.19 (0.87-1.63)	0.27	*GREB1*	2p25.1
2	rs1430787	67868498	A (0.31)	1.05 (0.81-1.36)	0.70	0.75 (0.45-1.26)	0.28	0.76 (0.51-1.12)	0.17	0.99 (0.72-1.35)	0.93	*ETAA1*	2p14
2	rs10167914	113529183	C (0.32)	0.94 (0.72-1.22)	0.65	0.94 (0.58-1.51)	0.79	1.09 (0.76-1.56)	0.63	0.97 (0.71-1.33)	0.86	*IL1A*	2q13
2	rs6435157	203441224	T (0.77)	1.16 (0.86-1.55)	0.33	1.24 (0.72-2.13)	0.44	0.97 (0.66-1.44)	0.90	1.14 (0.81-1.62)	0.45	*BMPR2*	2q33.1
2	rs1250247	216299629	C (0.28)	0.95 (0.72-1.24)	0.69	1.07 (0.67-1.72)	0.78	0.99 (0.69-1.44)	0.98	1.01 (0.74-1.39)	0.94	*FN1*	2q35
3	rs1352889	49652148	T (0.17)	1.06 (0.77-1.44)	0.73	0.70 (0.37-1.33)	0.28	1.15 (0.75-1.76)	0.51	1.08 (0.75-1.55)	0.69	*BSN*	3p21.31
3	rs12631337	50198537	A (0.44)	1.08 (0.85-1.38)	0.51	0.97 (0.62-1.50)	0.88	1.11 (0.79-1.56)	0.54	1.06 (0.79-1.41)	0.71	*SEMA3F*	3p21.31
4	rs1903068	56008477	A (0.68)	1.21 (0.93-1.57)	0.16	1.51 (0.90-2.52)	0.12	1.38 (0.95-2.02)	$9.41x10^{-2}$	1.16 (0.85-1.58)	0.35	*KDR*	4q12
4	rs2510770	95479372	A (0.32)	0.80 (0.61-1.04)	$9.33x10^{-2}$	0.56 (0.33-0.95)	$3.03x10^{-2}$	0.59 (0.40-0.88)	$9.04x10^{-3}$	0.81 (0.60-1.11)	0.20	*PDLIM5*	4q22.3
5	rs2946160	157891878	A (0.74)	1.18 (0.89-1.56)	0.24	1.33 (0.78-2.24)	0.29	1.23 (0.82-1.84)	0.31	1.21 (0.86-1.69)	0.27	*EBF1*	5q33.3
6	rs6456259	19761718	G (0.16)	1.10 (0.80-1.52)	0.56	1.40 (0.81-2.43)	0.23	0.93 (0.58-1.50)	0.77	1.11 (0.75-1.63)	0.61	*ID4*	6p22.3
6	rs4540228	74613060	C (0.66)	0.95 (0.74-1.23)	0.71	1.11 (0.69-1.78)	0.66	0.96 (0.67-1.36)	0.81	0.97 (0.72-1.32)	0.86	*CD109*	6q13
6	rs2226158	125995467	G (0.46)	0.90 (0.71-1.14)	0.38	0.84 (0.54-1.32)	0.45	0.85 (0.60-1.19)	0.349	0.83 (0.62-1.11)	0.21	*HEY2*	6q22.31
6	rs7740302	151826249	C (0.19)	1.02 (0.74-1.39)	0.92	1.15 (0.66-1.99)	0.62	1.13 (0.73-1.74)	0.59	1.15 (0.80-1.66)	0.44	*CCDC170*	6q25.1
6	rs71575922	152554014	G (0.15)	1.59 (1.18-2.14)	$2.48x10^{-3}$	2.62 (1.60-4.28)	$1.24x10^{-4}$	1.50 (0.98-2.29)	$6.27x10^{-2}$	1.64 (1.15-2.33)	$5.93x10^{-3}$	*SYNE1*	6q25.1
6	rs11756073	170363580	A (0.17)	0.91 (0.65-1.29)	0.60	1.18 (0.66-2.11)	0.58	0.94 (0.58-1.53)	0.81	1.05 (0.71-1.56)	0.81	*FAM120B*	6q27
7	rs1451383	25891809	G (0.74)	1.27 (0.95-1.71)	0.11	1.34 (0.78-2.31)	0.29	1.40 (0.92-2.15)	0.12	1.08 (0.77-1.51)	0.66	—	7p15.2
7	rs6970537	27217359	G (0.57)	1.01 (0.79-1.29)	0.94	1.20 (0.75-1.89)	0.45	0.94 (0.67-1.32)	0.72	1.11 (0.83-1.49)	0.49	*HOXA10*	7p15.2
7	rs55909142	46673774	C (0.59)	1.04 (0.82-1.33)	0.74	1.43 (0.89-2.30)	0.14	1.00 (0.71-1.43)	0.98	1.12 (0.83-1.51)	0.45	—	7p12.3
8	rs17053711	25311269	G (0.73)	0.87 (0.67-1.14)	0.31	0.83 (0.51-1.34)	0.44	0.83 (0.57-1.20)	0.32	0.84 (0.61-1.15)	0.27	*KCTD9*	8p21.2
8	rs10090060	75257608	A (0.57)	1.03 (0.80-1.31)	0.84	0.63 (0.41-0.98)	$4.01x10^{-2}$	0.99 (0.70-1.40)	0.96	0.95 (0.71-1.27)	0.72	*GDAP1*	8q21.11
8	rs12549438	100150371	T (0.21)	1.20 (0.90-1.61)	0.21	1.70 (1.04-2.75)	$3.28x10^{-2}$	1.31 (0.88-1.96)	0.18	1.37 (0.98-1.91)	$6.84x10^{-2}$	*VPS13B*	8q22.2
9	rs10122243	22158924	T (0.41)	0.99 (0.77-1.26)	0.92	0.99 (0.64-1.55)	0.98	1.03 (0.73-1.45)	0.86	0.93 (0.69-1.25)	0.63	*CDKN2-BAS1*	9p21.3
9	rs10983311	119470373	T (0.77)	1.24 (0.91-1.69)	0.17	0.91 (0.54-1.53)	0.73	1.14 (0.75-1.75)	0.54	1.14 (0.79-1.62)	0.49	*ASTN2*	9q33.1
9	rs507666	136149399	A (0.20)	1.09 (0.80-1.48)	0.58	1.20 (0.70-2.07)	0.50	1.37 (0.92-2.06)	0.13	1.01 (0.69-1.46)	0.97	*ABO*	9q34.2
10	rs10828249	21824727	A (0.35)	1.14 (0.89-1.46)	0.31	1.40 (0.89-2.18)	0.14	1.26 (0.89-1.78)	0.20	1.19 (0.89-1.60)	0.25	*MLLT10*	10p12.31
10	rs7907732	90182764	T (0.39)	1.08 (0.85-1.38)	0.53	1.29 (0.83-2.00)	0.26	1.05 (0.75-1.49)	0.77	1.29 (0.97-1.73)	$8.28x10^{-2}$	*RNLS*	10q23.31
11	rs3858429	30343757	C (0.84)	1.27 (0.89-1.82)	0.19	0.89 (0.50-1.58)	0.70	0.93 (0.59-1.45)	0.74	1.06 (0.71-1.58)	0.78	*ARL14EP*	11p14.1
11	rs7924571	32350027	C (0.76)	1.13 (0.84-1.53)	0.43	0.90 (0.53-1.53)	0.70	1.02 (0.68-1.55)	0.91	1.11 (0.77-1.59)	0.58	*FSHB*	11p14.1
12	rs56090796	15406228	A (0.33)	1.13 (0.87-1.45)	0.36	0.89 (0.55-1.44)	0.63	0.94 (0.65-1.36)	0.75	1.09 (0.80-1.48)	0.60	*PTPRO*	12p12.3
12	rs3803042	54387947	A (0.42)	1.05 (0.82-1.34)	0.71	1.11 (0.71-1.74)	0.64	1.07 (0.76-1.51)	0.69	1.29 (0.96-1.72)	$8.91x10^{-2}$	*HOXC10*	12p13.13
12	rs12320196	95645385	G (0.47)	1.06 (0.83-1.34)	0.65	0.92 (0.59-1.43)	0.70	0.94 (0.67-1.32)	0.72	0.97 (0.73-1.30)	0.85	*VEZT*	12q22
12	rs10860864	102822862	C (0.82)	0.89 (0.65-1.21)	0.46	0.83 (0.47-1.44)	0.50	0.89 (0.57-1.38)	0.60	0.91 (0.63-1.33)	0.64	*IGF1*	12q23.2
13	rs7334326	50967732	C (0.16)	1.07 (0.77-1.50)	0.68	1.40 (0.79-2.49)	0.25	0.76 (0.45-1.29)	0.31	1.18 (0.80-1.74)	0.42	*DLEU1*	13q14.2
14	rs57281976	93112974	G (0.77)	1.00 (0.75-1.32)	0.98	1.05 (0.63-1.76)	0.85	0.98 (0.66-1.45)	0.92	1.15 (0.81-1.62)	0.44	*RIN3*	14q32.12
15	rs12441483	40353416	C (0.47)	1.29 (1.01-1.64)	$3.90x10^{-2}$	1.31 (0.85-2.03)	0.23	1.29 (0.92-1.80)	0.15	1.13 (0.85-1.50)	0.42	*BMF*	15q15.1
17	rs66683298	46277748	C (0.60)	1.04 (0.81-1.33)	0.75	1.24 (0.78-1.95)	0.36	0.83 (0.59-1.16)	0.27	1.05 (0.78-1.41)	0.75	*SKAP1*	17q21.32
17	rs7214750	63903119	C (0.93)	1.04 (0.65-1.66)	0.87	0.98 (0.40-2.39)	0.96	1.11 (0.56-2.20)	0.76	0.83 (0.49-1.39)	0.47	*CEP112*	17q24.1
19	rs2967684	8786624	A (0.16)	1.42 (1.04-1.94)	$2.75x10^{-2}$	0.91 (0.48-1.74)	0.78	1.14 (0.72-1.82)	0.58	1.22 (0.83-1.80)	0.31	*ACTL9*	19p13.2
X	rs13441059	70108889	A (0.36)	0.99 (0.77-1.27)	0.92	1.28 (0.82-1.99)	0.28	0.94 (0.66-1.34)	0.74	1.07 (0.80-1.44)	0.65	*TEX11*	Xq13.1
X	rs5933091	131278107	T (0.69)	1.24 (0.94-1.62)	0.12	1.03 (0.63-1.66)	0.92	1.15 (0.79-1.67)	0.47	1.27 (0.92-1.76)	0.15	*FRMD7*	Xq26.2
X	rs73241342	133691738	A (0.04)	1.06 (0.60-1.90)	0.84	1.40 (0.52-3.75)	0.50	1.56 (0.77-3.16)	0.21	1.06 (0.53-2.14)	0.86	*PLAC1*	Xq26.3

Supplementary Table S 3.7 Meta-analysed association results of 46 loci for endometriosis with acyclical pelvic pain by severity and temporality (part A).

Chr	SNP	Position (hg19)	RA (RAF)	Endometriosis with **Acyclical Pelvic Pain**								Nearest Gene	Cytoband
				All (n=167)		Severe (NRS 7-10)				Moderate to Severe (4-10)			
						Recent (n=124)		At worse in lifetime (n=96)		Recent (n=151)			
				OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value		
1	rs10917151	22422721	A (0.17)	1.28 (0.93-1.77)	0.13	1.41 (0.98-2.03)	$6.29x10^{-2}$	1.31 (0.82-2.11)	0.26	1.37 (0.97-1.94)	$7.60x10^{-2}$	*WNT4*	1p36.12
1	rs12030576	115817221	G (0.65)	0.90 (0.70-1.17)	0.45	0.83 (0.61-1.12)	0.22	0.92 (0.63-1.36)	0.69	0.89 (0.67-1.19)	0.43	*NGF*	1p13.2
1	rs2040445	169216412	G (0.97)	0.68 (0.29-1.60)	0.38	0.70 (0.26-1.89)	0.48	0.68 (0.18-2.50)	0.56	0.57 (0.24-1.39)	0.22	*SLC19A2*	1q24.2
1	rs2421985	172099136	C (0.49)	1.06 (0.83-1.36)	0.65	1.14 (0.85-1.52)	0.39	1.13 (0.78-1.63)	0.52	1.13 (0.86-1.48)	0.39	*DNM3*	1q24.3
2	rs11674184	11721535	T (0.61)	1.06 (0.81-1.38)	0.67	1.08 (0.80-1.47)	0.61	1.08 (0.74-1.60)	0.68	1.03 (0.77-1.37)	0.84	*GREB1*	2p25.1
2	rs1430787	67868498	A (0.31)	1.16 (0.89-1.51)	0.27	1.22 (0.90-1.65)	0.21	1.15 (0.78-1.70)	0.47	1.18 (0.89-1.58)	0.25	*ETAA1*	2p14
2	rs10167914	113529183	C (0.32)	0.91 (0.69-1.19)	0.48	0.94 (0.69-1.30)	0.73	1.03 (0.69-1.52)	0.90	0.97 (0.72-1.31)	0.85	*IL1A*	2q13
2	rs6435157	203441224	T (0.77)	1.12 (0.83-1.51)	0.46	0.96 (0.68-1.34)	0.80	0.98 (0.64-1.49)	0.91	1.03 (0.75-1.42)	0.84	*BMPR2*	2q33.1
2	rs1250247	216299629	C (0.28)	1.01 (0.76-1.32)	0.97	1.03 (0.75-1.42)	0.85	1.07 (0.72-1.58)	0.76	0.96 (0.71-1.30)	0.80	*FN1*	2q35
3	rs1352889	49652148	T (0.17)	1.09 (0.79-1.50)	0.60	1.12 (0.77-1.62)	0.56	1.34 (0.85-2.09)	0.21	1.13 (0.80-1.60)	0.48	*BSN*	3p21.31
3	rs12631337	50198537	A (0.44)	1.03 (0.80-1.32)	0.82	1.16 (0.87-1.56)	0.31	1.06 (0.73-1.54)	0.76	1.13 (0.86-1.49)	0.37	*SEMA3F*	3p21.31
4	rs1903068	56008477	A (0.68)	1.27 (0.97-1.67)	$8.71x10^{-2}$	1.15 (0.84-1.58)	0.37	1.14 (0.77-1.70)	0.51	1.17 (0.87-1.57)	0.30	*KDR*	4q12
4	rs2510770	95479372	A (0.32)	0.87 (0.66-1.13)	0.30	0.72 (0.52-1.00)	$5.09x10^{-2}$	0.72 (0.48-1.09)	0.12	0.77 (0.57-1.04)	$9.38x10^{-2}$	*PDLIM5*	4q22.3
5	rs2946160	157891878	A (0.74)	1.19 (0.89-1.59)	0.25	1.29 (0.92-1.83)	0.15	1.37 (0.88-2.13)	0.16	1.35 (0.98-1.88)	$6.85x10^{-2}$	*EBF1*	5q33.3
6	rs6456259	19761718	G (0.16)	1.24 (0.90-1.72)	0.20	1.16 (0.79-1.70)	0.46	1.35 (0.84-2.17)	0.22	1.28 (0.90-1.82)	0.17	*ID4*	6p22.3
6	rs4540228	74613060	C (0.66)	0.98 (0.75-1.27)	0.87	1.03 (0.75-1.40)	0.87	1.07 (0.72-1.59)	0.75	0.94 (0.70-1.25)	0.66	*CD109*	6q13
6	rs2226158	125995467	G (0.46)	1.00 (0.78-1.28)	0.98	1.01 (0.76-1.35)	0.94	0.91 (0.63-1.32)	0.62	0.99 (0.76-1.30)	0.96	*HEY2*	6q22.31
6	rs7740302	151826249	C (0.19)	0.97 (0.70-1.35)	0.87	0.99 (0.67-1.44)	0.95	1.09 (0.68-1.75)	0.73	0.99 (0.69-1.42)	0.96	*CCDC170*	6q25.1
6	rs71575922	152554014	G (0.15)	1.56 (1.15-2.13)	$4.63x10^{-3}$	1.55 (1.09-2.22)	$1.58x10^{-2}$	1.60 (1.02-2.51)	$4.20x10^{-2}$	1.60 (1.14-2.24)	$6.28x10^{-3}$	*SYNE1*	6q25.1
6	rs11756073	170363580	A (0.17)	1.02 (0.72-1.44)	0.91	1.14 (0.77-1.67)	0.52	1.14 (0.70-1.86)	0.60	1.04 (0.72-1.51)	0.83	*FAM120B*	6q27
7	rs1451383	25891809	G (0.74)	1.10 (0.82-1.47)	0.54	1.05 (0.75-1.48)	0.77	1.16 (0.75-1.79)	0.51	1.14 (0.82-1.57)	0.43	—	7p15.2
7	rs6970537	27217359	G (0.57)	1.01 (0.78-1.30)	0.96	1.09 (0.81-1.47)	0.56	0.94 (0.65-1.36)	0.73	0.98 (0.75-1.30)	0.91	*HOXA10*	7p15.2
7	rs55909142	46673774	C (0.59)	1.18 (0.91-1.52)	0.22	1.10 (0.82-1.48)	0.53	1.05 (0.72-1.53)	0.79	1.13 (0.85-1.49)	0.40	—	7p12.3
8	rs17053711	25311269	G (0.73)	0.91 (0.69-1.20)	0.50	0.97 (0.70-1.34)	0.84	0.83 (0.56-1.25)	0.38	0.90 (0.66-1.22)	0.50	*KCTD9*	8p21.2
8	rs10090060	75257608	A (0.57)	0.93 (0.73-1.20)	0.60	0.95 (0.71-1.27)	0.72	0.98 (0.68-1.43)	0.94	0.91 (0.69-1.20)	0.50	*GDAP1*	8q21.11
8	rs12549438	100150371	T (0.21)	1.06 (0.78-1.45)	0.69	1.09 (0.77-1.56)	0.62	1.25 (0.81-1.94)	0.31	1.13 (0.81-1.58)	0.47	*VPS13B*	8q22.2
9	rs10122243	22158924	T (0.41)	1.02 (0.80-1.32)	0.85	1.04 (0.78-1.40)	0.79	1.01 (0.70-1.47)	0.94	1.00 (0.76-1.32)	0.97	*CDKN2-BAS1*	9p21.3
9	rs10983311	119470373	T (0.77)	1.32 (0.96-1.82)	$9.19x10^{-2}$	1.08 (0.75-1.54)	0.69	1.61 (0.97-2.67)	$6.65x10^{-2}$	1.29 (0.91-1.84)	0.15	*ASTN2*	9q33.1
9	rs507666	136149399	A (0.20)	1.22 (0.90-1.66)	0.21	1.14 (0.79-1.64)	0.48	1.36 (0.88-2.12)	0.17	1.21 (0.87-1.70)	0.26	*ABO*	9q34.2
10	rs10828249	21824727	A (0.35)	1.22 (0.95-1.58)	0.13	1.32 (0.98-1.77)	$6.83x10^{-2}$	1.11 (0.76-1.62)	0.61	1.23 (0.93-1.63)	0.15	*MLLT10*	10p12.31
10	rs7907732	90182764	T (0.39)	1.09 (0.85-1.40)	0.51	1.16 (0.86-1.55)	0.33	1.09 (0.75-1.58)	0.66	1.09 (0.82-1.43)	0.55	*RNLS*	10q23.31
11	rs3858429	30343757	C (0.84)	1.16 (0.81-1.65)	0.41	1.06 (0.71-1.58)	0.77	1.11 (0.66-1.84)	0.70	1.13 (0.77-1.66)	0.53	*ARL14EP*	11p14.1
11	rs7924571	32350027	C (0.76)	1.21 (0.88-1.66)	0.25	1.28 (0.87-1.87)	0.21	1.20 (0.75-1.92)	0.46	1.14 (0.81-1.60)	0.46	*FSHB*	11p14.1
12	rs56090796	15406228	A (0.33)	1.13 (0.87-1.47)	0.38	1.12 (0.83-1.52)	0.46	1.00 (0.67-1.48)	1.00	1.12 (0.84-1.50)	0.43	*PTPRO*	12p12.3
12	rs3803042	54387947	A (0.42)	0.92 (0.71-1.19)	0.53	0.91 (0.67-1.22)	0.53	0.97 (0.67-1.41)	0.87	0.98 (0.74-1.29)	0.86	*HOXC10*	12p13.13
12	rs12320196	95645385	G (0.47)	1.08 (0.84-1.38)	0.56	1.05 (0.79-1.41)	0.73	1.10 (0.76-1.59)	0.62	1.02 (0.78-1.34)	0.90	*VEZT*	12q22
12	rs10860864	102822862	C (0.82)	1.04 (0.74-1.46)	0.84	1.29 (0.84-1.96)	0.24	1.09 (0.65-1.80)	0.75	1.04 (0.72-1.52)	0.82	*IGF1*	12q23.2
13	rs7334326	50967732	C (0.16)	1.20 (0.86-1.68)	0.29	1.22 (0.83-1.81)	0.31	1.36 (0.85-2.19)	0.20	1.22 (0.85-1.76)	0.28	*DLEU1*	13q14.2
14	rs57281976	93112974	G (0.77)	1.05 (0.78-1.41)	0.74	1.07 (0.76-1.51)	0.69	0.93 (0.61-1.41)	0.73	1.02 (0.74-1.40)	0.91	*RIN3*	14q32.12
15	rs12441483	40353416	C (0.47)	1.33 (1.04-1.71)	$2.51x10^{-2}$	1.38 (1.03-1.85)	$2.90x10^{-2}$	1.50 (1.04-2.17)	$3.08x10^{-2}$	1.33 (1.02-1.75)	$3.81x10^{-2}$	*BMF*	15q15.1
17	rs66683298	46277748	C (0.60)	1.08 (0.83-1.39)	0.57	1.14 (0.85-1.54)	0.39	0.99 (0.68-1.44)	0.96	1.12 (0.85-1.48)	0.42	*SKAP1*	17q21.32
17	rs7214750	63903119	C (0.93)	1.36 (0.79-2.33)	0.27	1.07 (0.60-1.92)	0.81	1.56 (0.63-3.86)	0.34	1.35 (0.75-2.41)	0.32	*CEP112*	17q24.1
19	rs2967684	8786624	A (0.16)	1.20 (0.86-1.68)	0.29	0.95 (0.62-1.44)	0.80	1.04 (0.62-1.75)	0.88	1.20 (0.83-1.74)	0.33	*ACTL9*	19p13.2
X	rs13441059	70108889	A (0.36)	1.02 (0.79-1.32)	0.89	1.04 (0.77-1.41)	0.79	1.06 (0.73-1.54)	0.77	1.00 (0.75-1.32)	0.98	*TEX11*	Xq13.1
X	rs5933091	131278107	T (0.69)	1.31 (0.98-1.74)	$6.44x10^{-2}$	1.43 (1.02-2.01)	$3.75x10^{-2}$	1.25 (0.83-1.90)	0.28	1.28 (0.94-1.74)	0.12	*FRMD7*	Xq26.2
X	rs73241342	133691738	A (0.04)	1.18 (0.66-2.10)	0.58	1.31 (0.69-2.48)	0.41	1.04 (0.42-2.57)	0.93	1.15 (0.61-2.17)	0.66	*PLAC1*	Xq26.3

Supplementary Table S 3.8 Meta-analysed association results of 46 loci for endometriosis with mild to moderate dyspareunia, and acyclical pelvic pain by temporality (part B).

Chr	SNP	Position (hg19)	RA (RAF)	Endometriosis with **Dyspareunia**				Endometriosis with **Acyclical Pelvic Pain**				Nearest Gene	Cytoband
				Mild to Moderate (NRS 1-6)				**Mild to Moderate (NRS 1-6)**					
				Recent (n=73)		**At worse in lifetime (n=43)**		**Recent (n=36)**		**At worse in lifetime (n=23)**			
				OR (95% CI)	**P value**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**		
1	rs10917151	22422721	A (0.17)	1.40 (0.91-2.16)	0.13	1.05 (0.57-1.91)	0.88	1.02 (0.50-2.10)	0.95	1.28 (0.57-2.89)	0.55	*WNT4*	1p36.12
1	rs12030576	115817221	G (0.65)	0.75 (0.53-1.07)	0.11	0.91 (0.57-1.47)	0.70	1.00 (0.57-1.75)	1.00	1.51 (0.73-3.09)	0.26	*NGF*	1p13.2
1	rs2040445	169216412	G (0.97)	0.63 (0.21-1.93)	0.42	0.21 (0.06-0.75)	$1.65x10^{-2}$	0.60 (0.12-2.97)	0.53	0.48 (0.06-3.84)	0.49	*SLC19A2*	1q24.2
1	rs2421985	172099136	C (0.49)	1.02 (0.73-1.44)	0.90	0.84 (0.54-1.31)	0.45	0.93 (0.55-1.56)	0.77	0.69 (0.36-1.29)	0.24	*DNM3*	1q24.3
2	rs11674184	11721535	T (0.61)	1.16 (0.81-1.66)	0.43	0.90 (0.58-1.41)	0.66	1.03 (0.60-1.78)	0.91	1.36 (0.69-2.66)	0.37	*GREB1*	2p25.1
2	rs1430787	67868498	A (0.31)	1.31 (0.92-1.86)	0.14	1.42 (0.90-2.25)	0.13	0.98 (0.55-1.73)	0.94	1.07 (0.56-2.06)	0.83	*ETAA1*	2p14
2	rs10167914	113529183	C (0.32)	0.95 (0.66-1.37)	0.78	0.71 (0.43-1.17)	0.18	0.85 (0.47-1.52)	0.58	0.66 (0.32-1.33)	0.24	*IL1A*	2q13
2	rs6435157	203441224	T (0.77)	1.23 (0.81-1.87)	0.33	1.32 (0.76-2.31)	0.32	1.91 (0.91-4.03)	$8.96x10^{-2}$	3.29 (1.15-9.41)	$2.59x10^{-2}$	*BMPR2*	2q33.1
2	rs1250247	216299629	C (0.28)	0.93 (0.64-1.36)	0.71	0.66 (0.39-1.11)	0.12	0.95 (0.53-1.69)	0.85	0.75 (0.37-1.52)	0.43	*FN1*	2q35
3	rs1352889	49652148	T (0.17)	1.39 (0.93-2.09)	0.11	1.22 (0.71-2.09)	0.47	1.04 (0.53-2.06)	0.90	0.82 (0.35-1.89)	0.64	*BSN*	3p21.31
3	rs12631337	50198537	A (0.44)	1.17 (0.83-1.63)	0.38	0.87 (0.56-1.35)	0.55	0.71 (0.41-1.23)	0.22	0.69 (0.37-1.31)	0.26	*SEMA3F*	3p21.31
4	rs1903068	56008477	A (0.68)	1.16 (0.80-1.67)	0.44	1.12 (0.70-1.78)	0.64	1.45 (0.79-2.66)	0.24	1.75 (0.84-3.64)	0.13	*KDR*	4q12
4	rs2510770	95479372	A (0.32)	1.03 (0.72-1.46)	0.88	1.33 (0.85-2.07)	0.21	1.18 (0.68-2.02)	0.56	1.18 (0.63-2.2)	0.61	*PDLIM5*	4q22.3
5	rs2946160	157891878	A (0.74)	1.06 (0.73-1.56)	0.75	1.39 (0.81-2.38)	0.23	1.10 (0.60-1.99)	0.77	1.5 (0.7-3.19)	0.30	*EBF1*	5q33.3
6	rs6456259	19761718	G (0.16)	1.02 (0.64-1.63)	0.93	1.28 (0.73-2.23)	0.39	1.77 (0.96-3.26)	$6.83x10^{-2}$	0.86 (0.3-2.44)	0.78	*ID4*	6p22.3
6	rs4540228	74613060	C (0.66)	0.77 (0.54-1.08)	0.13	1.19 (0.74-1.92)	0.47	0.82 (0.48-1.41)	0.48	0.98 (0.52-1.86)	0.96	*CD109*	6q13
6	rs2226158	125995467	G (0.46)	1.04 (0.74-1.46)	0.81	0.85 (0.53-1.36)	0.49	1.01 (0.60-1.70)	0.975	0.86 (0.47-1.58)	0.62	*HEY2*	6q22.31
6	rs7740302	151826249	C (0.19)	1.23 (0.81-1.87)	0.33	0.77 (0.41-1.44)	0.41	0.77 (0.35-1.67)	0.51	0.88 (0.38-2.04)	0.77	*CCDC170*	6q25.1
6	rs71575922	152554014	G (0.15)	1.14 (0.72-1.82)	0.57	1.72 (1.02-2.89)	$4.27x10^{-2}$	1.52 (0.80-2.89)	0.2	0.93 (0.38-2.27)	0.87	*SYNE1*	6q25.1
6	rs11756073	170363580	A (0.17)	0.84 (0.51-1.39)	0.50	1.07 (0.60-1.93)	0.81	0.68 (0.30-1.56)	0.36	0.92 (0.36-2.35)	0.86	*FAM120B*	6q27
7	rs1451383	25891809	G (0.74)	1.02 (0.69-1.50)	0.93	1.46 (0.84-2.54)	0.18	1.04 (0.57-1.91)	0.90	1.79 (0.75-4.25)	0.19	—	7p15.2
7	rs6970537	27217359	G (0.57)	0.98 (0.69-1.38)	0.91	0.98 (0.63-1.53)	0.94	0.87 (0.52-1.47)	0.60	1.05 (0.57-1.94)	0.88	*HOXA10*	7p15.2
7	rs55909142	46673774	C (0.59)	1.04 (0.74-1.46)	0.83	0.83 (0.53-1.30)	0.42	1.71 (0.97-3.02)	$6.59x10^{-2}$	1.81 (0.9-3.61)	$9.39x10^{-2}$	—	7p12.3
8	rs17053711	25311269	G (0.73)	0.86 (0.59-1.25)	0.43	0.83 (0.52-1.34)	0.45	0.64 (0.37-1.10)	0.11	0.85 (0.43-1.66)	0.63	*KCTD9*	8p21.2
8	rs10090060	75257608	A (0.57)	1.31 (0.92-1.88)	0.14	0.95 (0.61-1.48)	0.83	1.09 (0.64-1.87)	0.74	0.65 (0.36-1.19)	0.16	*GDAP1*	8q21.11
8	rs12549438	100150371	T (0.21)	0.98 (0.64-1.50)	0.93	0.92 (0.52-1.63)	0.78	1.00 (0.52-1.95)	0.99	0.54 (0.2-1.43)	0.22	*VPS13B*	8q22.2
9	rs10122243	22158924	T (0.41)	0.92 (0.65-1.30)	0.65	1.06 (0.68-1.64)	0.80	0.84 (0.49-1.43)	0.52	0.93 (0.5-1.72)	0.81	*CDKN2-BAS1*	9p21.3
9	rs10983311	119470373	T (0.77)	1.61 (1.01-2.59)	$4.70x10^{-2}$	1.68 (0.87-3.22)	0.12	2.55 (1.08-6.04)	$3.36x10^{-2}$	2.06 (0.82-5.13)	0.12	*ASTN2*	9q33.1
9	rs507666	136149399	A (0.20)	0.87 (0.54-1.38)	0.55	0.99 (0.56-1.75)	0.97	1.30 (0.69-2.44)	0.41	1.3 (0.62-2.72)	0.49	*ABO*	9q34.2
10	rs10828249	21824727	A (0.35)	0.81 (0.56-1.18)	0.28	0.92 (0.57-1.46)	0.71	0.89 (0.50-1.60)	0.71	1.09 (0.57-2.09)	0.78	*MLLT10*	10p12.31
10	rs7907732	90182764	T (0.39)	1.05 (0.74-1.48)	0.80	0.95 (0.60-1.49)	0.81	0.76 (0.43-1.33)	0.33	1.31 (0.71-2.41)	0.39	*RNLS*	10q23.31
11	rs3858429	30343757	C (0.84)	1.13 (0.68-1.88)	0.63	1.48 (0.75-2.90)	0.25	1.50 (0.66-3.42)	0.34	1.56 (0.58-4.17)	0.38	*ARL14EP*	11p14.1
11	rs7924571	32350027	C (0.76)	1.25 (0.81-1.94)	0.32	1.25 (0.71-2.19)	0.43	1.03 (0.54-1.95)	0.94	1.48 (0.62-3.51)	0.38	*FSHB*	11p14.1
12	rs56090796	15406228	A (0.33)	1.26 (0.88-1.79)	0.20	1.22 (0.77-1.92)	0.39	1.13 (0.64-2.00)	0.66	0.99 (0.51-1.92)	0.97	*PTPRO*	12p12.3
12	rs3803042	54387947	A (0.42)	1.24 (0.88-1.75)	0.22	0.82 (0.52-1.29)	0.39	0.89 (0.52-1.54)	0.69	0.55 (0.28-1.11)	$9.46x10^{-2}$	*HOXC10*	12p13.13
12	rs12320196	95645385	G (0.47)	1.13 (0.81-1.58)	0.48	1.21 (0.79-1.87)	0.38	1.22 (0.72-2.05)	0.46	1.06 (0.58-1.95)	0.84	*VEZT*	12q22
12	rs10860864	102822862	C (0.82)	0.96 (0.61-1.51)	0.87	0.69 (0.40-1.17)	0.17	0.68 (0.36-1.28)	0.24	0.5 (0.25-0.99)	$4.63x10^{-2}$	*IGF1*	12q23.2
13	rs7334326	50967732	C (0.16)	1.07 (0.67-1.72)	0.77	1.22 (0.68-2.20)	0.51	1.36 (0.70-2.67)	0.37	1.12 (0.46-2.7)	0.80	*DLEU1*	13q14.2
14	rs57281976	93112974	G (0.77)	1.00 (0.68-1.49)	0.99	0.75 (0.47-1.20)	0.23	0.89 (0.49-1.60)	0.69	0.83 (0.41-1.67)	0.60	*RIN3*	14q32.12
15	rs12441483	40353416	C (0.47)	1.11 (0.79-1.55)	0.56	1.52 (0.98-2.36)	$6.07x10^{-2}$	1.20 (0.71-2.02)	0.50	1.09 (0.6-1.99)	0.78	*BMF*	15q15.1
17	rs66683298	46277748	C (0.60)	1.03 (0.73-1.46)	0.86	1.13 (0.72-1.76)	0.60	1.02 (0.60-1.73)	0.95	0.8 (0.44-1.46)	0.47	*SKAP1*	17q21.32
17	rs7214750	63903119	C (0.93)	0.93 (0.50-1.74)	0.82	2.24 (0.71-7.10)	0.17	2.76 (0.59-12.99)	0.20	0.73 (0.25-2.09)	0.55	*CEP112*	17q24.1
19	rs2967684	8786624	A (0.16)	1.49 (0.96-2.29)	$7.23x10^{-2}$	1.63 (0.95-2.80)	$7.76x10^{-2}$	2.15 (1.18-3.94)	$1.29x10^{-2}$	1.13 (0.48-2.64)	0.79	*ACTL9*	19p13.2
X	rs13441059	70108889	A (0.36)	0.83 (0.58-1.20)	0.32	0.96 (0.61-1.51)	0.86	0.94 (0.54-1.62)	0.82	0.93 (0.49-1.76)	0.83	*TEX11*	Xq13.1
X	rs5933091	131278107	T (0.69)	1.42 (0.96-2.11)	$7.96x10^{-2}$	1.26 (0.77-2.06)	0.36	1.10 (0.62-1.95)	0.76	1.25 (0.63-2.47)	0.53	*FRMD7*	Xq26.2
X	rs73241342	133691738	A (0.04)	0.97 (0.41-2.29)	0.94	0.82 (0.25-2.66)	0.74	0.49 (0.06-4.11)	0.51	2.1 (0.61-7.26)	0.24	*PLAC1*	Xq26.3

Supplementary Table S 3.9 Meta-analysed association results of 46 loci for endometriosis with bladder pain, and endometriosis with ≥ 2, ≥ 3, and ≥ 4 reported pain types.

Chr	SNP	Position (hg19)	RA (RAF)	Endometriosis with								Nearest Gene	Cytoband
				Bladder Pain (n=107)		Pain Types ≥ 4 (n=109)		Pain Types ≥ 3 (n=155)		Pain Types ≥ 2 (n=198)			
				OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value		
1	rs10917151	22422721	A (0.17)	1.11 (0.72-1.69)	0.64	1.19 (0.79-1.81)	0.40	1.31 (0.95-1.79)	0.10	1.25 (0.95-1.66)	0.11	*WNT4*	1p36.12
1	rs12030576	115817221	G (0.65)	1.13 (0.81-1.58)	0.46	1.07 (0.77-1.49)	0.70	0.90 (0.70-1.17)	0.43	0.96 (0.77-1.20)	0.73	*NGF*	1p13.2
1	rs2040445	169216412	G (0.97)	0.43 (0.17-1.09)	$7.48x10^{-2}$	0.40 (0.17-0.96)	$4.05x10^{-2}$	0.61 (0.27-1.37)	0.23	0.77 (0.36-1.66)	0.51	*SLC19A2*	1q24.2
1	rs2421985	172099136	C (0.49)	0.97 (0.71-1.31)	0.82	0.98 (0.72-1.33)	0.90	1.07 (0.84-1.37)	0.58	1.06 (0.86-1.31)	0.59	*DNM3*	1q24.3
2	rs11674184	11721535	T (0.61)	0.97 (0.70-1.33)	0.84	1.05 (0.76-1.45)	0.78	1.06 (0.82-1.37)	0.68	1.03 (0.83-1.30)	0.77	*GREB1*	2p25.1
2	rs1430787	67868498	A (0.31)	1.17 (0.85-1.63)	0.34	1.21 (0.88-1.68)	0.25	1.10 (0.84-1.42)	0.50	1.14 (0.91-1.43)	0.26	*ETAA1*	2p14
2	rs10167914	113529183	C (0.32)	0.91 (0.65-1.27)	0.58	0.87 (0.62-1.22)	0.43	0.93 (0.71-1.21)	0.59	0.96 (0.76-1.22)	0.76	*IL1A*	2q13
2	rs6435157	203441224	T (0.77)	1.10 (0.77-1.59)	0.60	1.07 (0.74-1.54)	0.71	1.14 (0.85-1.53)	0.39	1.19 (0.91-1.54)	0.20	*BMPR2*	2q33.1
2	rs1250247	216299629	C (0.28)	0.89 (0.63-1.25)	0.50	0.97 (0.69-1.37)	0.88	0.96 (0.73-1.26)	0.75	0.94 (0.74-1.19)	0.60	*FN1*	2q35
3	rs1352889	49652148	T (0.17)	1.09 (0.74-1.62)	0.65	1.19 (0.81-1.76)	0.38	1.10 (0.80-1.50)	0.56	1.06 (0.80-1.40)	0.68	*BSN*	3p21.31
3	rs12631337	50198537	A (0.44)	1.06 (0.78-1.44)	0.73	1.12 (0.82-1.54)	0.46	1.10 (0.86-1.40)	0.46	1.03 (0.83-1.27)	0.82	*SEMA3F*	3p21.31
4	rs1903068	56008477	A (0.68)	1.36 (0.97-1.91)	$7.90x10^{-2}$	1.12 (0.80-1.56)	0.51	1.37 (1.04-1.80)	$2.42x10^{-2}$	1.29 (1.02-1.63)	$3.70x10^{-2}$	*KDR*	4q12
4	rs2510770	95479372	A (0.32)	0.83 (0.59-1.15)	0.26	0.79 (0.56-1.10)	0.17	0.87 (0.67-1.14)	0.31	0.86 (0.68-1.08)	0.20	*PDLIM5*	4q22.3
5	rs2946160	157891878	A (0.74)	1.25 (0.87-1.79)	0.23	1.37 (0.95-1.98)	$9.64x10^{-2}$	1.21 (0.91-1.61)	0.19	1.08 (0.85-1.38)	0.54	*EBF1*	5q33.3
6	rs6456259	19761718	G (0.16)	1.22 (0.81-1.81)	0.34	1.23 (0.82-1.84)	0.32	1.16 (0.83-1.60)	0.39	1.25 (0.95-1.65)	0.12	*ID4*	6p22.3
6	rs4540228	74613060	C (0.66)	0.98 (0.71-1.35)	0.90	1.02 (0.74-1.42)	0.88	0.98 (0.76-1.27)	0.89	0.97 (0.77-1.21)	0.78	*CD109*	6q13
6	rs2226158	125995467	G (0.46)	0.88 (0.65-1.20)	0.42	0.89 (0.65-1.21)	0.45	0.95 (0.74-1.21)	0.68	0.93 (0.75-1.16)	0.54	*HEY2*	6q22.31
6	rs7740302	151826249	C (0.19)	0.89 (0.58-1.35)	0.58	1.00 (0.66-1.50)	0.99	0.98 (0.71-1.36)	0.93	1.01 (0.77-1.34)	0.93	*CCDC170*	6q25.1
6	rs71575922	152554014	G (0.15)	1.35 (0.90-2.01)	0.14	1.52 (1.03-2.25)	$3.48x10^{-2}$	1.63 (1.20-2.20)	$1.65x10^{-3}$	1.50 (1.15-1.97)	$3.12x10^{-3}$	*SYNE1*	6q25.1
6	rs11756073	170363580	A (0.17)	1.22 (0.82-1.83)	0.33	1.22 (0.81-1.83)	0.34	1.03 (0.74-1.45)	0.85	0.98 (0.73-1.32)	0.88	*FAM120B*	6q27
7	rs1451383	25891809	G (0.74)	1.35 (0.92-1.97)	0.12	1.26 (0.87-1.84)	0.23	1.26 (0.94-1.70)	0.13	1.16 (0.90-1.50)	0.25	—	7p15.2
7	rs6970537	27217359	G (0.57)	1.17 (0.86-1.61)	0.32	1.07 (0.78-1.47)	0.66	0.99 (0.78-1.27)	0.96	1.06 (0.85-1.31)	0.63	*HOXA10*	7p15.2
7	rs55909142	46673774	C (0.59)	1.04 (0.76-1.41)	0.82	1.11 (0.81-1.53)	0.50	1.12 (0.87-1.44)	0.38	1.06 (0.85-1.32)	0.59	—	7p12.3
8	rs17053711	25311269	G (0.73)	0.90 (0.64-1.27)	0.56	0.80 (0.57-1.12)	0.19	0.88 (0.67-1.16)	0.36	0.95 (0.75-1.21)	0.68	*KCTD9*	8p21.2
8	rs10090060	75257608	A (0.57)	0.89 (0.65-1.22)	0.47	0.91 (0.67-1.24)	0.55	0.92 (0.72-1.18)	0.53	1.00 (0.80-1.24)	1.00	*GDAP1*	8q21.11
8	rs12549438	100150371	T (0.21)	1.20 (0.83-1.74)	0.33	1.25 (0.86-1.81)	0.24	1.12 (0.83-1.51)	0.46	1.08 (0.83-1.40)	0.59	*VPS13B*	8q22.2
9	rs10122243	22158924	T (0.41)	1.05 (0.77-1.43)	0.75	1.02 (0.74-1.39)	0.92	0.99 (0.78-1.28)	0.97	1.01 (0.81-1.25)	0.95	*CDKN2-BAS1*	9p21.3
9	rs10983311	119470373	T (0.77)	1.10 (0.75-1.61)	0.62	1.25 (0.84-1.85)	0.27	1.21 (0.89-1.65)	0.22	1.29 (0.98-1.69)	$7.32x10^{-2}$	*ASTN2*	9q33.1
9	rs507666	136149399	A (0.20)	1.18 (0.80-1.73)	0.40	1.33 (0.92-1.94)	0.13	1.19 (0.88-1.62)	0.26	1.12 (0.86-1.47)	0.41	*ABO*	9q34.2
10	rs10828249	21824727	A (0.35)	1.12 (0.82-1.54)	0.47	1.01 (0.73-1.40)	0.95	1.23 (0.96-1.59)	0.11	1.20 (0.96-1.49)	0.11	*MLLT10*	10p12.31
10	rs7907732	90182764	T (0.39)	1.20 (0.88-1.64)	0.25	1.20 (0.88-1.65)	0.24	1.07 (0.84-1.38)	0.57	1.11 (0.89-1.38)	0.35	*RNLS*	10q23.31
11	rs3858429	30343757	C (0.84)	1.18 (0.76-1.82)	0.46	1.21 (0.77-1.87)	0.41	1.18 (0.83-1.67)	0.36	1.22 (0.89-1.66)	0.22	*ARL14EP*	11p14.1
11	rs7924571	32350027	C (0.76)	1.25 (0.84-1.85)	0.27	1.21 (0.81-1.79)	0.36	1.33 (0.96-1.83)	$8.46x10^{-2}$	1.18 (0.90-1.55)	0.22	*FSHB*	11p14.1
12	rs56090796	15406228	A (0.33)	0.96 (0.69-1.35)	0.83	0.91 (0.65-1.29)	0.60	1.09 (0.84-1.42)	0.50	1.18 (0.95-1.48)	0.14	*PTPRO*	12p12.3
12	rs3803042	54387947	A (0.42)	0.90 (0.66-1.24)	0.52	0.87 (0.63-1.20)	0.40	0.94 (0.73-1.21)	0.63	0.97 (0.78-1.21)	0.80	*HOXC10*	12p13.13
12	rs12320196	95645385	G (0.47)	1.14 (0.84-1.55)	0.40	0.96 (0.70-1.31)	0.79	1.16 (0.91-1.48)	0.23	1.10 (0.89-1.37)	0.37	*VEZT*	12q22
12	rs10860864	102822862	C (0.82)	1.05 (0.69-1.59)	0.82	1.01 (0.66-1.53)	0.97	1.01 (0.73-1.41)	0.94	0.95 (0.71-1.26)	0.71	*IGF1*	12q23.2
13	rs7334326	50967732	C (0.16)	0.99 (0.64-1.54)	0.98	1.16 (0.76-1.77)	0.49	1.13 (0.81-1.58)	0.48	1.13 (0.84-1.52)	0.40	*DLEU1*	13q14.2
14	rs57281976	93112974	G (0.77)	1.02 (0.72-1.46)	0.89	1.00 (0.70-1.44)	0.98	1.02 (0.77-1.36)	0.89	1.07 (0.83-1.38)	0.59	*RIN3*	14q32.12
15	rs12441483	40353416	C (0.47)	1.35 (0.99-1.83)	$5.55x10^{-2}$	1.33 (0.98-1.81)	$7.13x10^{-2}$	1.36 (1.06-1.74)	$1.46x10^{-2}$	1.34 (1.08-1.66)	$7.29x10^{-3}$	*BMF*	15q15.1
17	rs66683298	46277748	C (0.60)	1.08 (0.79-1.47)	0.64	1.14 (0.83-1.57)	0.41	1.09 (0.85-1.41)	0.49	1.03 (0.83-1.28)	0.79	*SKAP1*	17q21.32
17	rs7214750	63903119	C (0.93)	1.22 (0.66-2.25)	0.53	1.35 (0.70-2.58)	0.37	1.20 (0.73-1.98)	0.47	1.19 (0.77-1.83)	0.44	*CEP112*	17q24.1
19	rs2967684	8786624	A (0.16)	0.96 (0.61-1.50)	0.85	1.27 (0.84-1.92)	0.26	1.23 (0.89-1.72)	0.21	1.23 (0.93-1.64)	0.15	*ACTL9*	19p13.2
X	rs13441059	70108889	A (0.36)	0.99 (0.72-1.36)	0.94	1.06 (0.77-1.46)	0.71	1.04 (0.81-1.34)	0.76	1.02 (0.81-1.27)	0.89	*TEX11*	Xq13.1
X	rs5933091	131278107	T (0.69)	1.19 (0.85-1.68)	0.31	1.24 (0.87-1.75)	0.23	1.24 (0.94-1.63)	0.13	1.20 (0.94-1.53)	0.14	*FRMD7*	Xq26.2
X	rs73241342	133691738	A (0.04)	0.87 (0.39-1.96)	0.73	1.05 (0.49-2.22)	0.91	1.05 (0.58-1.90)	0.87	1.04 (0.62-1.76)	0.87	*PLAC1*	Xq26.3

Supplementary Table S 3.10 Chi-square statistics for patterns in missing report of dyspareunia in pain Numeric Rating Scale (NRS, recent) for a) dysmenorrhea severity, b) acyclical pelvic pain severity, c) number of reported pain types, and d) age groups.

Key: X^2_{Yates} = Total chi-square statistic with Yates's correction for non-continuous variable; X^2 = Chi-square statistic without correction; ☒ = Without report of; ☑ = With Report of; O = Observed count; P = Predicted count. The phenotypic breakdown shown in table a) & b) are the same, thus the statistics.

a) Dysmenorrhea NRS (Recent)

Women Reported of Dyspareunia Experience	Severe			Non-severe / No / No report of			Row Totals
	Observed Count	Predicted Count	X^2 Statistic	Observed Count	Predicted Count	X^2 Statistic	
☒ Dyspareunia NRS, Recent	11	15.74	1.43	15	10.26	2.19	26
☑ Dyspareunia NRS, Recent	78	73.26	0.31	43	47.74	0.47	121
						X^2_{Yates} = **3.5 (p=0.061)**	
Column Totals	89			58			147

b) Acyclical Pelvic Pain NRS (Recent)

Women Reported of Dyspareunia Experience	Severe			Non-severe / No / No report of			Row Totals
	Observed Count	Predicted Count	X^2 Statistic	Observed Count	Predicted Count	X^2 Statistic	
☒ Dyspareunia NRS, Recent	11	15.74	1.43	15	10.26	2.19	26
☑ Dyspareunia NRS, Recent	78	73.26	0.31	43	47.74	0.47	121
						X^2_{Yates} = **3.5 (p=0.061)**	
Column Totals	89			58			147

c) Number of Pain Types Reported

Women Reported of Dyspareunia Experience	≥ 3			< 3 / No / No report of			Row Totals
	Observed Count	Predicted Count	X^2 Statistic	Observed Count	Predicted Count	X^2 Statistic	
☒ Dyspareunia NRS, Recent	20	22.64	0.31	6	3.36	2.07	26
☑ Dyspareunia NRS, Recent	108	105.36	0.07	13	15.64	0.45	121
						X^2_{Yates} =**1.9 (p=0.17)**	

d) Age Group

Women Reported of Dyspareunia Experience	18-26			27-30			31-33			34-38			39-48			Row Totals
	O	P	X^2	O	P	X^2	O	P	X^2	O	P	X^2	O	P	X^2	
☒ Dyspareunia NRS, Recent	2	4.95	1.76	4	6.01	0.67	8	6.01	0.66	8	4.24	3.32	4	4.78	0.13	26
☑ Dyspareunia NRS, Recent	26	23.05	0.38	30	27.99	0.14	26	27.99	0.14	16	19.76	0.71	23	22.22	0.03	121
															X^2= **7.9 (p=0.094)**	
Column Totals	28			34			34			24			27			147

Age Range of All Women Reported of Dyspareunia Experience in ENDOX Phase I & II	Frequency
18-26	30
27-30	34
31-33	34
34-38	28
39-48	28

Supplementary Table S 4.1 Breakdown of number of samples in histological-confirmed and self-reported cycle phases.

Key: Ca = Case, Co = Control, Irregular = self-reported irregular cycle length; blue boxes = matched phases, grey boxes = numbers within three sub-secretory phases.

		Self-reported Cycle phase (n)								Histology Total:
		Menstrual	Proliferative	All Secretory	Early-Secretory	Mid-Secretory	Late-Secretory	No Sub-Secretory	Missing cycle length	
Histology based Cycle phase (n)	Menstrual	3 Ca3; Co0	0	1 Ca0; Co1	0	0	1 Ca0; Co1	—	0	4 Ca3; Co1
	Proliferative	1 Ca1; Co0	10 Ca8; Co2	9 Ca7; Co2	8 (5 on transition day) Ca6; Co2	1 Ca1; Co0	0	—	3 Ca2; Co1 Irregular1	23 Ca18; Co5
	All Secretory	3 Ca3; Co0 Irregular1	8 Ca5; Co3	32 Ca25; Co7	—	—	—	—	10 Ca5; Co5 Irregular3	53 Ca38; Co15
	Early-Secretory	0	0	—	0	2 Ca2; Co0	0	—	1 Ca0; Co1	3 Ca2; Co1
	Mid-Secretory	0	5 Ca3; Co2	—	4 Ca2; Co2	3 Ca3; Co0	6 Ca5; Co1	—	5 Ca2; Co3 Irregular1	23 Ca15; Co8
	Late-Secretory	1 Ca1; Co0	1 Ca1; Co0	—	3 Ca3; Co0	3 Ca1; Co2	8 Ca7; Co1	—	2 Ca2; Co0 Irregular1	18 Ca15; Co3
	No Sub-Secretory	2 Ca2; Co0 Irregular1	2 Ca1; Co1	—	1 Ca1; Co0	0	2 Ca1; Co1	—	2 Ca1;Co1 Irregular1	9 Ca6; Co3
	Histology not available	10 Ca6; Co3 Irregular1	21 Ca14; Co7 Irregular2	34 Ca21; Co13	10 Ca6; Co4	8 Ca6; Co2	12 Ca7; Co5	4 Ca2; Co2	45 Ca27; Co18 Irregular15 (Co6)	110 Ca69; Co41
Self-report Total:		**17 Ca14; Co3**	**39 Ca27; Co12**	**76 Ca53; Co23**	**26 Ca18; Co8**	**17 Ca13; Co4**	**29 Ca20; Co9**	**4 Ca2; Co2**	**58 Ca34; Co24 Irregular19**	

Supplementary Table S 4.2 Top 50 annotated differentially expressed (DE) genes between histological-confirmed proliferative (n=23) versus all secretory (n=53) samples by DESeq2 (2nd) analysis, with comparison to results from edgeR (1st) pipeline.

Key: Rank = rank of gene by adjusted *p* value, Y = presence in edgeR DE list, lncRNA = long non-coding RNA, *marker ranked 28th in DESeq2 was without annotation; positive log_2 fold change indicates higher expression in secretory compared to proliferative phase samples

Top 50 Annotated DE Gene in DESeq2				Proliferative vs. All Secretory						
				DESeq2 (2nd) Pipeline			edgeR (1st) Pipeline			
Generic Gene Name (GRCh38)	Alternative Name (GRCh37)	Entrezgene ID	Gene Biotype	Adj.*p*	Log_2 FC	Rank	Presence	Adj.*p*	Log_2 FC	Rank
C2CD4A	—	145741	protein coding	5.23E-33	6.57	1	Y	2E-16	6.55	1
LINC02532	*RP1-60O19.1*	—	lncRNA	4.37E-30	7.08	2	—	—	—	—
CXCL14	—	9547	protein coding	4.72E-25	5.89	3	Y	3E-13	5.86	4
LINC01502	*RP11-98L5.2*	—	lncRNA	4.72E-25	7.68	4	—	—	—	—
IGFBP1	—	3484	protein coding	6.65E-25	8.67	5	—	—	—	—
HAP1	—	9001	protein coding	5.20E-24	5.37	6	Y	2E-13	5.34	3
CATSPERB	—	79820	protein coding	3.21E-23	4.23	7	Y	1E-14	4.23	2
ANKRD55	—	79722	protein coding	1.73E-21	6.74	8	—	—	—	—
TMEM252	—	169693	protein coding	3.21E-21	5.63	9	—	—	—	—
ADGRF1	*GPR110*	266977	protein coding	4.63E-21	6.12	10	—	—	—	—
C4BPA	—	722	protein coding	9.59E-21	4.96	11	Y	2E-12	4.93	8
SPP1	—	6696	protein coding	1.19E-20	4.15	12	Y	3E-13	4.27	6
AL121839.1	*RP11-747H7.1*	—	processed pseudogene	1.32E-20	4.52	13	—	—	—	—
COMP	—	1311	protein coding	3.17E-20	5.99	14	—	—	—	—
GPX3	—	2878	protein coding	3.73E-20	5.02	15	Y	3E-12	5.03	17
DEPP1	*C10orf10*	11067	protein coding	5.14E-20	3.90	16	Y	3E-13	3.94	5
SLC15A1	—	6564	protein coding	1.07E-18	4.59	17	Y	3E-12	4.61	24
SLC6A12	—	6539	protein coding	1.12E-18	3.96	18	—	—	—	—
LINC02432	*RP11-362F19.1*	—	lncRNA	1.48E-18	5.00	19	—	—	—	—
C1GALT1P2	*RP11-365H8.2*	—	processed pseudogene	1.82E-18	7.99	20	—	—	—	—
IRX3	—	79191	protein coding	3.69E-18	4.65	21	—	—	—	—
C2CD4B	—	388125	protein coding	3.76E-18	3.70	22	Y	2E-12	3.69	12
AC073218.1	*RP11-529E15.1*	—	lncRNA	5.28E-18	3.71	23	Y	2E-12	3.70	13
HABP2	—	3026	protein coding	8.35E-18	4.48	24	Y	7E-12	4.48	31
NNMT	—	4837	protein coding	1.05E-17	3.59	25	Y	2E-12	3.67	11
MFSD4A	*MFSD4*	148808	protein coding	1.05E-17	3.47	26	Y	2E-12	3.46	16
SAMMSON	*RP11-460N16.1*	—	lncRNA	1.05E-17	5.03	27	—	—	—	—
LAMB3	—	3914	protein coding	1.67E-17	3.74	29*	Y	3E-12	3.85	25
AOX1	—	316	protein coding	1.94E-17	4.55	30	Y	2E-11	4.52	46
SLCO4A1	—	28231	protein coding	2.93E-17	3.72	31	Y	4E-12	3.85	26
LEFTY2	—	7044	protein coding	3.50E-17	6.16	32	—	—	—	—
RPSAP51	—	—	processed pseudogene	5.77E-17	3.60	33	—	—	—	—
GALNT15	—	117248	protein coding	7.24E-17	3.84	34	Y	9E-12	3.82	37
SDCBP2	—	27111	protein coding	7.65E-17	2.89	35	Y	1E-12	2.93	7
HAL	—	3034	protein coding	9.15E-17	3.50	36	Y	7E-12	3.48	32
DPP4	—	1803	protein coding	1.23E-16	4.20	37	Y	2E-11	4.18	45
CD55	—	1604	protein coding	1.35E-16	3.15	38	Y	3E-12	3.19	20
FGB	—	2244	protein coding	1.62E-16	7.18	39	—	—	—	—
LYPLAL1-AS1	*RP11-392O17.1*	—	lncRNA	1.77E-16	3.58	40	—	—	—	—
TCN1	—	6947	protein coding	2.07E-16	4.23	41	—	—	—	—
SORCS1	—	114815	protein coding	2.59E-16	3.62	42	Y	1E-11	3.64	40
PHYHIPL	—	84457	protein coding	2.59E-16	2.86	43	Y	3E-12	2.83	23
NANOGP7	—	—	processed pseudogene	3.97E-16	4.64	44	—	—	—	—
SOD2	—	6648	protein coding	5.52E-16	2.77	45	Y	3E-12	2.87	22
PTPRR	—	5801	protein coding	6.71E-16	3.75	46	—	—	—	—
ATG9B	—	285973	protein coding	6.95E-16	3.27	47	Y	7E-12	3.30	30
MAOA	—	4128	protein coding	7.42E-16	3.23	48	Y	2E-11	3.21	43
MMP10	—	4319	protein coding	7.86E-16	6.27	49	Y	3E-09	6.48	134
RIMKLB	—	57494	protein coding	9.92E-16	2.93	50	Y	5E-12	2.95	28

Supplementary Table S 4.3 Top 50 differentially expressed (DE) genes between histological-confirmed proliferative (n=23) versus all secretory (n=53) samples by `edgeR` pipeline, with comparison to Fung et al. and Talbi et al. reports.

Key: List ordered by adj.*p*; Y = reported in publications; positive log_2 fold change indicates higher expression in secretory compared to proliferative phase samples

Top 50	In-house (Proliferative vs. All Secretory)		Reported in previous publications (Proliferative vs. Early-secretory)	
DE Gene	**Adj.*p***	**Log_2 Fold Change**	**Fung et al. 2017**	**Talbi et al. 2006**
C2CD4A	1.75E-16	6.55	—	—
CATSPERB	1.46E-14	4.23	—	—
HAP1	2.19E-13	5.34	—	—
CXCL14	2.62E-13	5.86	—	—
DEPP1	3.19E-13	3.94	—	—
SPP1	3.19E-13	4.27	Y	Y
SDCBP2	1.21E-12	2.93	Y	—
C4BPA	1.79E-12	4.93	—	—
CITED2	2.24E-12	2.16	Y	—
LINC-PINT	2.29E-12	2.14	—	—
NNMT	2.36E-12	3.67	—	—
C2CD4B	2.36E-12	3.69	Y	—
AC073218.2	2.36E-12	3.70	—	—
RLIMP1	2.40E-12	2.27	—	—
ADAMTS1	2.40E-12	2.07	Y	—
MFSD4A	2.40E-12	3.46	—	—
GPX3	2.62E-12	5.03	—	—
BCL6	2.62E-12	2.12	Y	—
NFIL3	2.62E-12	2.00	Y	—
CD55	2.62E-12	3.19	Y	—
FOXO1	2.78E-12	2.05	Y	Y
SOD2	2.78E-12	2.87	Y	—
PHYHIPL	3.24E-12	2.83	—	—
SLC15A1	3.24E-12	4.61	—	—
LAMB3	3.33E-12	3.85	—	—
SLCO4A1	3.72E-12	3.85	—	—
KLF6	4.32E-12	1.94	Y	—
RIMKLB	5.39E-12	2.95	Y	—
CEBPD	6.16E-12	2.80	Y	—
ATG9B	6.81E-12	3.30	—	—
HABP2	6.92E-12	4.48	—	—
HAL	6.92E-12	3.48	Y	—
APOD	7.95E-12	2.91	Y	Y
MAP3K5	8.58E-12	2.50	Y	—
AL121839.2	9.32E-12	2.42	—	—
NABP1	9.32E-12	2.30	—	—
GALNT15	9.32E-12	3.82	—	—
GABARAPL1	1.06E-11	2.10	Y	—
AL024508.1	1.08E-11	2.58	—	—
SORCS1	1.08E-11	3.64	—	—
NDRG1	1.43E-11	2.17	Y	—
ACKR2	1.67E-11	1.95	—	—
MAOA	2.12E-11	3.21	Y	Y
IL15	2.13E-11	2.20	Y	Y
DPP4	2.22E-11	4.18	—	—
AOX1	2.29E-11	4.52	—	—
GBP2	2.68E-11	2.03	Y	—
TC2N	3.75E-11	1.90	Y	—
LINC01320	4.78E-11	2.97	—	—
SERPING1	4.80E-11	2.58	Y	—

Supplementary Table S 4.4 Top 21 differentially expressed (DE) genes between histological-confirmed mid-secretory (n=23) versus late-secretory (n=18) samples by `edgeR` pipeline.

Key: List ordered by adj.*p*; positive log_2 fold change indicates higher expression in late-secretory compared to mid-secretory phase samples

Top 21 DE Gene		**In-house (Mid-secretory vs. Late-secretory)**	
Generic Gene Name (GRCh38)	**Alternative Name (GRCh37)**	**Adj.*p***	**Log_2 Fold Change**
DES	—	1.77E-05	4.03
MYH11	—	1.10E-04	2.54
PDLIM3	—	1.24E-04	3.41
COL8A1	*C3orf7*	1.50E-04	1.99
PTGIS	—	1.82E-03	2.45
KIAA1755	—	1.82E-03	1.31
CHRDL1	*MGC1*	5.24E-03	2.63
HIST2H2AA3	—	8.75E-03	-2.39
ACTA2-AS1	—	1.02E-02	1.07
ACTA2	—	1.02E-02	1.75
PRELP	—	1.02E-02	2.14
CCNO	*CCNU*	1.08E-02	-1.07
TAGLN	—	3.05E-02	1.58
PGM5	—	3.29E-02	1.58
CNN1	—	3.68E-02	2.28
SNHG9	—	3.89E-02	-0.78
SLC25A1	*SLC20A3*	3.89E-02	-0.91
CCDC80	*LINC01279*	3.89E-02	1.33
C7	—	3.89E-02	2.27
SORBS1	*SH3D5*	3.89E-02	1.48
SPINK5	—	4.43E-02	-1.26

Supplementary Table S 4.5 List of 26 genes for multivariate model building for proliferative and secretory phase prediction (with single-gene logistic regression statistics).

Key: FC = Fold Change; Beta = beta coefficient in logistic regression statistics; SE = standard error. * Rank of gene by adjusted *p* value out of n=7,417 significant DE list from in-house analysis, ^ Rank out of n=1,864 reported DE list from Fung et al. 2017. $^{+}$ Rank out of n=570 reported 'on' gene list from Fung et al., $^{\#}$ Rank out of n=570 reported 'off' gene list from Fung et al.

	In-house (by edgeR) (Proliferative vs. All Secretory)			**Fung et al. 2017 (Proliferative vs. Early-secretory)**			**Talbi et al. 2006**	**In-house Single-gene Logistic Regression**		
Genes	**Adj.*p***	**Log_2FC**	**Rank***	**Adj.*p***	**Log_2FC**	**Rank^**	**Direction of FC**	**Beta**	**SE**	***p***
SPP1	3.19E-13	4.27	6	4.22E-06	2.46	485	positive	0.34	0.21	1.05E-01
FOXO1	2.78E-12	2.05	21	3.59E-07	1.55	413	positive	0.84	0.44	5.85E-02
APOD	7.95E-12	2.91	33	4.50E-04	1.13	1418	positive	0.32	0.27	2.27E-01
MAOA	2.12E-11	3.21	43	2.02E-06	1.89	623	positive	0.12	0.29	6.74E-01
IL15	2.13E-11	2.20	44	9.01E-07	1.35	531	positive	0.39	0.38	3.04E-01
ARG2	3.39E-10	2.02	74	1.53E-08	2.46	190	positive	0.84	0.42	**4.40E-02**
SLC3A1	5.50E-10	2.61	86	—	—	291^{+}	positive	0.13	0.24	5.82E-01
DKK1	3.56E-09	3.55	137	7.17E-09	3.20	147	positive	0.07	0.11	5.47E-01
FGFR3	3.72E-08	-1.58	247	3.22E-03	-1.08	1673	negative	-0.09	0.12	4.47E-01
MSX2	3.47E-07	-1.67	429	6.55E-05	-1.50	1106	negative	-0.18	0.14	1.98E-01
MT1F	2.63E-06	1.93	739	1.07E-08	1.99	169	positive	-0.12	0.22	5.76E-01
GABRP	3.58E-06	1.10	781	3.52E-05	1.24	1030	positive	-0.40	0.37	2.79E-01
SFRP1	5.77E-06	-1.70	886	2.65E-09	-2.16	102	negative	-0.11	0.14	4.06E-01
TNC	8.87E-06	-1.35	975	1.99E-05	-1.45	960	negative	-0.04	0.16	8.12E-01
MT1E	1.99E-05	1.97	1163	2.91E-07	2.12	393	positive	-0.11	0.21	5.96E-01
CDH2	2.50E-05	-1.07	1218	3.23E-09	-1.60	112	negative	-0.17	0.18	3.42E-01
MMP11	4.40E-05	-1.62	1408	2.15E-10	-3.48	21	negative	-0.05	0.12	6.97E-01
F13A1	6.18E-05	-1.22	1526	1.54E-06	-1.93	595	negative	-0.02	0.17	9.07E-01
ALDH1A3	9.10E-05	1.83	1684	1.70E-07	2.94	260	positive	0.33	0.26	2.12E-01
MT1X	1.41E-04	1.50	1876	3.47E-07	1.82	411	positive	0.08	0.21	7.14E-01
SEMA3A	2.20E-04	-1.45	2086	—	—	197$^{\#}$	negative	-0.08	0.09	4.10E-01
OLFM1	2.71E-04	-1.30	2196	1.51E-06	-2.05	560	negative	-0.23	0.14	9.00E-02
PTGS1	2.13E-03	1.17	3661	1.84E-04	1.87	1265	positive	-0.73	0.41	7.28E-02
CD36	2.66E-03	1.40	3859	3.23E-06	2.46	681	positive	-0.31	0.23	1.82E-01
PTGS2	1.46E-02	1.35	5631	3.30E-02	1.21	1849	positive	0.07	0.14	6.04E-01
TFPI2	2.99E-02	1.04	6578	1.73E-06	1.57	605	positive	0.05	0.21	8.03E-01

Supplementary Table S 4.6 List of significantly enriched Ingenuity canonical pathways (Benjamini-Hochberg $p \leq 0.05$, i.e. -log $p \geq 1.3$, n=168) for differentially expressed genes between proliferative and secretory phase endometrium by IPA.

Key: Pathway order by significance (-log *p* value); z-score = IPA activation score, >+2 the pathway is activated in secretory phase, <-2 the pathway is inhibited in secretory phase relative to proliferative phase, without a z-score = no prediction of direction available from the database

Ingenuity Canonical Pathways	-log *p*	z-score	Representative Genes
Hepatic Fibrosis / Hepatic Stellate Cell Activation	9.54	—	CCR7, FGF1, MMP2, EDNRB, COL27A1
Leukocyte Extravasation Signaling	8.09	3.00	CLDN10, ITGAM, RHOH, ITGAL, MMP2
Complement System	7.40	1.00	ITGAM, MASP1, C4A/C4B, C2, C3
Th1 and Th2 Activation Pathway	7.09	—	CCR1, SOCS3, JAK3, IL2RB, IL2RG
Axonal Guidance Signaling	6.69	—	MMP2, ADAMTS16, BMP2, ROBO3, ENPEP
Osteoarthritis Pathway	6.51	1.63	PTCH1, ITGA3, ITGA2, GLI1, MTF1
Natural Killer Cell Signaling	6.22	—	FGFR3, LAIR1, PIK3R3, IRS2, LILRB1
Phagosome Formation	5.72	—	FGFR3, ITGAM, RHOH, PIK3R3, IRS2
Th1 Pathway	5.56	2.14	FGFR3, CD3E, STAT4, PIK3R3, IRS2
Th2 Pathway	5.34	2.07	FGFR3, CD3E, STAT4, PIK3R3, IRS2
Granulocyte Adhesion and Diapedesis	5.12	—	CLDN10, ITGAM, ITGAL, MMP11, MMP2
GP6 Signaling Pathway	4.98	1.89	FGFR3, GRAP2, LAMC3, PIK3R3, IRS2
Atherosclerosis Signaling	4.76	—	IL1B, APOD, PLA2G16, MSR1, APOC1
iCOS-iCOSL Signaling in T Helper Cells	4.43	1.73	FGFR3, GRAP2, CD3E, PIK3R3, IRS2
LXR/RXR Activation	4.39	0.50	C3, IL1B, CD14, APOD, ABCA1
VDR/RXR Activation	4.33	0.82	TRPV6, CD14, TGFB2, KLF4, THBD
Dendritic Cell Maturation	4.29	3.27	CCR7, FGFR3, IL15, STAT4, PIK3R3
HMGB1 Signaling	4.29	2.50	FGFR3, RHOH, IL15, FASLG, PIK3R3
Tumoricidal Function of Hepatic Natural Killer Cells	4.12	1.34	ITGAL, FASLG, SRGN, SERPINB9, PRF1
Cardiac Hypertrophy Signaling (Enhanced)	4.09	5.10	MYOCD, IL15, FGF1, NPR1, TNFSF15
Caveolar-mediated Endocytosis Signaling	4.04	—	ITGAM, ITGAL, ITGA1, ITGB6, ITGA3
Reelin Signaling in Neurons	3.99	—	FGFR3, ARHGEF15, DAB1, ITGAL, PIK3R3
Bladder Cancer Signaling	3.89	0.82	FGFR3, MMP11, FGF1, MMP2, DAPK1
Paxillin Signaling	3.84	1.13	FGFR3, ITGAM, ITGAL, PIK3R3, IRS2
Coagulation System	3.81	-0.707	PROS1, TFPI, THBD, F3, PLAT
STAT3 Pathway	3.79	1.27	FGFR3, KDR, IL1B, SOCS3, IL2RA
Tec Kinase Signaling	3.72	3.00	FGFR3, RHOH, STAT4, FASLG, PIK3R3
Crosstalk between Dendritic Cells and Natural Killer Cells	3.66	—	CCR7, ITGAL, IL15, FASLG, PRF1
Acute Phase Response Signaling	3.66	2.89	MAP3K5, PIK3R3, C3, IL1B, RBP7
T Helper Cell Differentiation	3.52	—	TNFRSF11B, BCL6, STAT4, IL2RA, FCER1G
PI3K/AKT Signaling	3.48	1.39	SFN, MAP3K5, PIK3R3, GDF15, ITGA3
T Cell Exhaustion Signaling Pathway	3.46	1.61	FGFR3, STAT4, PIK3R3, IRS2, KDR
Endothelin-1 Signaling	3.43	2.07	FGFR3, PIK3R3, IRS2, PLA2G16, GUCY1B1
T Cell Receptor Signaling	3.41	—	FGFR3, GRAP2, CD3E, PIK3R3, IRS2
IL-23 Signaling Pathway	3.41	1.00	FGFR3, STAT4, IRS2, PIK3R3, IL1B
Sperm Motility	3.38	3.16	FGFR3, KDR, NPR1, PLA2G16, GUCY1B1
Prostanoid Biosynthesis	3.33	2.00	PTGS2, PTGS1, TBXAS1, PTGIS
Adrenomedullin signaling pathway	3.26	2.07	FGFR3, PIK3R3, IRS2, MMP2, GAD1
Extrinsic Prothrombin Activation Pathway	3.23	0.00	PROS1, TFPI, THBD, F3, F13A1
CTLA4 Signaling in Cytotoxic T Lymphocytes	3.16	—	FGFR3, GRAP2, CD3E, PIK3R3, IRS2
Angiopoietin Signaling	3.15	0.82	FGFR3, ANGPTL1, PIK3R3, RASD1, IRS2
Glioma Invasiveness Signaling	3.15	1.27	FGFR3, RHOH, PIK3R3, RASD1, IRS2
PTEN Signaling	3.15	-0.832	FGFR3, FASLG, PIK3R3, KDR, ITGA3
LPS/IL-1 Mediated Inhibition of RXR Function	3.12	-0.333	CHST2, HS3ST3B1, IL1B, CD14, ABCA1
HIF1α Signaling	3.11	—	FGFR3, MMP11, PIK3R3, IRS2, MMP2
SPINK1 General Cancer Pathway	3.10	-1.265	FGFR3, MT1E, MT2A, MT1X, PIK3R3
IL-8 Signaling	3.10	3.44	FGFR3, HBEGF, ITGAM, RHOH, PIK3R3
CD28 Signaling in T Helper Cells	3.04	2.31	FGFR3, GRAP2, CD3E, PIK3R3, IRS2
Role of JAK1 and JAK3 in γc Cytokine Signaling	3.01	—	FGFR3, IL15, PIK3R3, RASD1, IRS2
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	2.97	2.52	FGFR3, RHOH, MAP3K5, PIK3R3, IRS2
FGF Signaling	2.97	1.16	FGFR3, RPS6KA5, PIK3R3, FGF1, MAP3K5
Role of Tissue Factor in Cancer	2.97	—	FGFR3, HBEGF, CYR61, PIK3R3, IRS2
IL-15 Production	2.96	—	FGFR3, IL15, KDR, CSF2RA, JAK3
Fcγ Receptor-mediated Phagocytosis in Macrophages and Monocytes	2.93	2.89	LCP2, WAS, PIK3R3, INPP5D, FGR
Agranulocyte Adhesion and Diapedesis	2.87	—	CLDN10, MMP11, MMP2, IL1B, ITGA3
Integrin Signaling	2.87	3.13	FGFR3, ITGAM, RHOH, ITGAL, IRS2

Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	2.80	—	IL15, DKK3, SOCS3, SFRP1, DKK1
Inhibition of Matrix Metalloproteases	2.80	0.38	MMP11, MMP2, MMP19, THBS2, TIMP3
HGF Signaling	2.79	1.27	FGFR3, MAP3K5, PIK3R3, IRS2, ITGA3
PKCθ Signaling in T Lymphocytes	2.78	3.00	FGFR3, GRAP2, CD3E, MAP3K5, PIK3R3
ERK5 Signaling	2.77	2.12	RPS6KA5, SFN, RASD1, CREB3L4, SGK1
Germ Cell-Sertoli Cell Junction Signaling	2.71	—	FGFR3, CDH2, RHOH, LAMC3, MAP3K5
Actin Cytoskeleton Signaling	2.67	2.31	FGFR3, PIK3R3, FGF1, IRS2, CD14
IL-15 Signaling	2.66	—	FGFR3, IL15, PIK3R3, RASD1, FASLG
Antiproliferative Role of Somatostatin Receptor 2	2.66	-0.447	FGFR3, PIK3R3, RASD1, IRS2, GAD1
Th17 Activation Pathway	2.62	2.53	STAT4, IL1B, SOCS3, DEFB1, JAK3
Phospholipase C Signaling	2.61	1.60	CD3E, RHOH, ARHGEF15, GRAP2, ITGA3
Cytotoxic T Lymphocyte-mediated Apoptosis of Target Cells	2.58	2.45	CD3E, FASLG, PRF1, FCER1G, CD247
Retinoate Biosynthesis I	2.58	1.63	DHRS3, RBP7, BMP2, ALDH8A1, ALDH1A3
Superpathway of Inositol Phosphate Compounds	2.56	3.58	FGFR3, IRS2, PIK3R3, PPFIA2, SOCS3
Tryptophan Degradation to 2-amino-3-carboxymuconate Semialdehyde	2.52	—	IDO2, IDO1, KMO
IL-12 Signaling and Production in Macrophages	2.50	—	FGFR3, STAT4, PIK3R3, IRS2, APOD
Eicosanoid Signaling	2.49	—	PLA2G16, PTGS2, RARRES3, PTGDR, PTGS1
Oncostatin M Signaling	2.48	2.24	MT2A, RASD1, JAK3, TIMP3, EPAS1
Macropinocytosis Signaling	2.47	0.82	FGFR3, PIK3R3, RASD1, IRS2, ITGB6
Ephrin Receptor Signaling	2.43	2.31	ARHGEF15, FGF1, ITGA3, ITGA2, WIPF1
Neuroinflammation Signaling Pathway	2.42	1.71	CX3CR1, GAD1, JAK3, GABRP, TREM2
Type I Diabetes Mellitus Signaling	2.41	2.33	TNFRSF11B, CD3E, FASLG, MAP3K5, IL1B
Virus Entry via Endocytic Pathways	2.41	—	FGFR3, ITGAL, PIK3R3, IRS2, ITGA3
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	2.27	—	FGFR3, ADAMTS4, MAP3K5, PIK3R3, IRS2
IL-9 Signaling	2.26	0.38	FGFR3, IRS2, PIK3R3, SOCS3, JAK3
HER-2 Signaling in Breast Cancer	2.22	—	FGFR3, PIK3R3, RASD1, MAP3K5, IRS2
NF-κB Signaling	2.21	2.50	FGFR3, PIK3R3, IRS2, KDR, IL1B
Leptin Signaling in Obesity	2.17	-0.816	ADCY1, FGFR3, PIK3R3, IRS2, SOCS3
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	2.16	1.41	FGFR3, IL15, FASLG, PIK3R3, IRS2
Xenobiotic Metabolism Signaling	2.16	—	FGFR3, CHST2, HS3ST3B1, GCLC, MAP3K5
IL-7 Signaling Pathway	2.14	1.41	FGFR3, BCL6, PIK3R3, IRS2, JAK3
Relaxin Signaling	2.13	1.27	RXFP1, FGFR3, PIK3R3, IRS2, GAD1
Acute Myeloid Leukemia Signaling	2.13	-1.134	FGFR3, CCNA1, PIK3R3, RASD1, IRS2
IL-2 Signaling	2.12	1.41	FGFR3, PIK3R3, RASD1, IRS2, JAK3
Sphingosine-1-phosphate Signaling	2.12	1.73	FGFR3, RHOH, PIK3R3, IRS2, SPHK1
Telomerase Signaling	2.05	1.90	FGFR3, PIK3R3, RASD1, IRS2, IL2RA
Colorectal Cancer Metastasis Signaling	2.02	2.18	FGFR3, RHOH, IRS2, MMP11, PIK3R3
Antioxidant Action of Vitamin C	2.01	-2.646	SLC2A3, PLA2G16, CSF2RA, SLC2A8, RARRES3
GADD45 Signaling	2.00	—	CDKN1A, GADD45A, GADD45G, CCNB1
ILK Signaling	1.96	2.32	FERMT2, FGFR3, RHOH, PIK3R3, IRS2
3-phosphoinositide Biosynthesis	1.96	3.00	FGFR3, PIK3R3, IRS2, SOCS3, PPFIA2
Pancreatic Adenocarcinoma Signaling	1.94	1.27	FGFR3, HBEGF, PIK3R3, IRS2, PTGS2
Melatonin Degradation II	1.92	—	MAOB, MAOA
NRF2-mediated Oxidative Stress Response	1.90	1.63	FGFR3, DNAJC6, GCLC, MAP3K5, PIK3R3
RAR Activation	1.88	—	MAP3K5, PIK3R3, RBP7, ZBTB16, TGFB2
Rac Signaling	1.86	1.51	FGFR3, IQGAP3, PIK3R3, RASD1, IRS2
Inhibition of Angiogenesis by TSP1	1.83	2.00	CD36, KDR, GUCY1B1, NOS3, THBS1
TREM1 Signaling	1.82	2.83	TLR4, IL1B, NOD2, NLRC5, TYROBP
NF-κB Activation by Viruses	1.82	1.67	FGFR3, ITGAL, ITGA1, PIK3R3, RASD1
Graft-versus-Host Disease Signaling	1.82	—	FASLG, IL1B, PRF1, FCER1G, CD86
Actin Nucleation by ARP-WASP Complex	1.82	2.24	RHOH, WAS, RASD1, ITGA3, RHOB
p53 Signaling	1.81	1.41	FGFR3, SFN, PIK3R3, IRS2, BIRC5
Semaphorin Signaling in Neurons	1.81	—	RHOH, RHOB, SEMA7A, SEMA3A, MET
Signaling by Rho Family GTPases	1.81	1.70	FGFR3, CDH2, RHOH, ARHGEF15, IRS2
Clathrin-mediated Endocytosis Signaling	1.80	—	FGFR3, PIK3R3, FGF1, IRS2, APOD
Role of NFAT in Regulation of the Immune Response	1.80	2.67	FGFR3, CD3E, PIK3R3, IRS2, FCER1G
B Cell Receptor Signaling	1.78	0.83	FGFR3, MAP3K5, PIK3R3, IRS2, RAC2
Circadian Rhythm Signaling	1.77	—	BHLHE40, CREB3L4, PER2, PER1, ADCYAP1R1
Apelin Cardiac Fibroblast Signaling Pathway	1.77	—1	TGFB2, SPHK1, NOS3, SERPINE1
IGF-1 Signaling	1.74	-0.447	FGFR3, SFN, CYR61, PIK3R3, RASD1
Serine Biosynthesis	1.71	—	PSAT1, PHGDH
Glutamate Degradation III (via 4-aminobutyrate)	1.71	—	GAD1, ABAT
NAD biosynthesis II (from tryptophan)	1.70	—	IDO2, IDO1, KMO

Neurotrophin/TRK Signaling	1.69	0.00	FGFR3, PIK3R3, RASD1, MAP3K5, IRS2
MSP-RON Signaling Pathway	1.69	—	TLR4, FGFR3, ITGAM, IRS2, PIK3R3
PI3K Signaling in B Lymphocytes	1.67	2.53	TLR4, RASD1, INPP5D, IRS2, C3
Phospholipases	1.66	1.89	PLA1A, PLA2G16, RARRES3, PLCH2, PLD1
Cardiac Hypertrophy Signaling	1.65	2.84	FGFR3, RHOH, MAP3K5, IRS2, PIK3R3
Tryptophan Degradation III (Eukaryotic)	1.64	0.00	IDO2, HACD1, IDO1, KMO
Ephrin A Signaling	1.63	—	FGFR3, PIK3R3, IRS2, EPHA5, EFNA1
Agrin Interactions at Neuromuscular Junction	1.62	2.65	ITGAL, ITGA1, ARHGEF6, RASD1, ITGA3
FLT3 Signaling in Hematopoietic Progenitor Cells	1.61	0.00	FGFR3, RPS6KA5, PIK3R3, RASD1, INPP5D
Synaptogenesis Signaling Pathway	1.58	1.28	FGFR3, DAB1, CDH2, PIK3R3, IRS2
Molecular Mechanisms of Cancer	1.57	—	RHOH, ARHGEF15, PTCH1, ITGA3, ITGA2
Corticotropin Releasing Hormone Signaling	1.56	1.00	ADCY1, FASLG, CREB3L4, GAD1, NPR1
Superpathway of Citrulline Metabolism	1.53	—	ARG2, NOS3, LOC102724788/PRODH
JAK/Stat Signaling	1.53	0.00	FGFR3, PIK3R3, RASD1, STAT4, IRS2
14-3-3-mediated Signaling	1.52	1.67	FGFR3, SFN, MAP3K5, PIK3R3, RASD1
Gap Junction Signaling	1.51	—	FGFR3, GJA4, PIK3R3, IRS2, GAD1
ERK/MAPK Signaling	1.51	1.29	FGFR3, KSR1, PIK3R3, IRS2, ITGA3
CD40 Signaling	1.51	0.82	FGFR3, PIK3R3, IRS2, PTGS2, JAK3
Endocannabinoid Cancer Inhibition Pathway	1.50	0.28	FGFR3, PIK3R3, IRS2, MMP2, TRIB3
p70S6K Signaling	1.50	1.90	FGFR3, SFN, PIK3R3, RASD1, IRS2
P2Y Purigenic Receptor Signaling Pathway	1.50	1.00	ADCY1, FGFR3, PIK3R3, RASD1, IRS2
VEGF Signaling	1.49	1.41	FGFR3, SFN, PIK3R3, RASD1, IRS2
Protein Kinase A Signaling	1.47	0.00	PTCH1, PTPN22, PDE4B, PTCH2, ADCY1
Glioblastoma Multiforme Signaling	1.45	1.16	FGFR3, RHOH, PIK3R3, IRS2, RAC2
RhoGDI Signaling	1.45	-2.111	CDH2, RHOH, ARHGEF15, ITGA3, ITGA2
Hepatic Cholestasis	1.43	—	IL15, FASLG, IL1B, TNFSF15, CD14
Docosahexaenoic Acid (DHA) Signaling	1.43	—	FGFR3, IRS2, PIK3R3, IL1B, FOXO1
Altered T Cell and B Cell Signaling in Rheumatoid Arthritis	1.42	—	TLR4, IL15, FASLG, IL1B, CSF1
SAPK/JNK Signaling	1.42	1.00	FGFR3, MAP3K5, PIK3R3, RASD1, IRS2
Superpathway of Serine and Glycine Biosynthesis I	1.42	—	PSAT1, PHGDH
Fc Epsilon RI Signaling	1.41	1.00	FGFR3, LCP2, GRAP2, PIK3R3, RASD1
IL-10 Signaling	1.40	—	IL1B, CD14, CCR1, SOCS3, ARG2
GM-CSF Signaling	1.40	0.82	FGFR3, PIK3R3, RASD1, IRS2, CSF2RA
Iron homeostasis signaling pathway	1.39	—	SLC39A14, GDF15, CP, STEAP3, JAK3
Granzyme A Signaling	1.39	—	HIST1H1A, GZMA, PRF1
Sertoli Cell-Sertoli Cell Junction Signaling	1.38	—	CLDN10, MAP3K5, ITGA3, GUCY1B1, ITGA2
TR/RXR Activation	1.38	—	FGFR3, PPARGC1A, PIK3R3, IRS2, SLC2A1
Neuropathic Pain Signaling In Dorsal Horn Neurons	1.37	1.90	FGFR3, GRIA3, PIK3R3, IRS2, PLCH2
CNTF Signaling	1.37	0.82	FGFR3, RPS6KA5, PIK3R3, RASD1, IRS2
Sonic Hedgehog Signaling	1.37	—	PTCH1, GLI1, PTCH2, CCNB1
Autoimmune Thyroid Disease Signaling	1.37	—	FASLG, PRF1, FCER1G, CD86, GZMB
Hematopoiesis from Pluripotent Stem Cells	1.37	—	CD3E, CSF1, LIF, FCER1G, CD247
Regulation of the Epithelial-Mesenchymal Transition Pathway	1.36	—	FGFR3, CDH2, PIK3R3, FGF1, IRS2
PDGF Signaling	1.36	1.41	FGFR3, PIK3R3, RASD1, INPP5D, IRS2
Synaptic Long Term Depression	1.35	2.89	GAD1, NPR1, PLA2G16, GUCY1B1, PPP2R2C
Epithelial Adherens Junction Signaling	1.34	—	CDH2, WAS, RASD1, FGF1, DLL1
Thrombin Signaling	1.34	1.16	FGFR3, RHOH, ARHGEF15, PIK3R3, IRS2
Xanthine and Xanthosine Salvage	1.34	—	PNP
Superoxide Radicals Degradation	1.30	—	SOD2, SOD3

Supplementary Table S 4.7 List of significantly enriched diseases and bio-function annotations with predicted activation state (z-score >+2, n=89) for differentially expressed genes between proliferative and secretory phase endometrium by IPA.

Key: Annotation order by |z-score|; z-score = IPA activation score, >+2 the pathway is activated in secretory relative to proliferative phase; No. of genes = total number of genes involve in one functional annotation

Categories	Diseases or Functions Annotation	-log *p*	z-score	Representative Genes	No. of Genes
Cellular Movement	Cell movement	37.02	5.80	ABCA1, ABI3, ACKR2, ADAMTS1, ADAMTS5	239
Cellular Movement	Migration of cells	32.08	5.48	ABI3, ACKR2, ADAMTS1, ADGRE2, ADGRG3	212
Cell-To-Cell Signaling and Interaction	Binding of blood cells	22.09	5.21	ADRA2A, APOE, C3, C4A/C4B, CADM1	60
Cell-To-Cell Signaling and Interaction,Hematological System Development and Function	Binding of leukocytes	18.58	5.05	APOE, CADM1, CCR1, CD14, CD2	50
Cellular Movement	Cell movement of tumor cell lines	16.91	4.70	ABCA1, ADAMTS5, ARHGEF6, ATF3, BMP2	136
Cell-To-Cell Signaling and Interaction	Adhesion of blood cells	19.44	4.59	APOE, C3, C4A/C4B, CADM1, CD14	52
Cell-To-Cell Signaling and Interaction	Binding of tumor cell lines	18.25	4.54	APOE, CD14, CD36, CD44, CD55	74
Cell-To-Cell Signaling and Interaction	Interaction of tumor cell lines	18.09	4.44	APOE, CD14, CD36, CD44, CD55	75
Cell-To-Cell Signaling and Interaction,Hematological System Development and Function	Interaction of mononuclear leukocytes	13.73	4.34	CCR1, CD14, CD2, CD44, CD58	32
Cell-To-Cell Signaling and Interaction,Hematological System Development and Function,Immune Cell Trafficking	Adhesion of immune cells	16.32	4.26	APOE, CADM1, CD14, CD2, CD44	44
Molecular Transport	Transport of molecule	8.73	4.26	ABCA1, ABCA4, ABCC3, ABCC8, ABCC9	108
Cell-To-Cell Signaling and Interaction,Hematological System Development and Function	Binding of mononuclear leukocytes	13.39	4.24	CCR1, CD2, CD44, CD58, CD86	31
Infectious Diseases	Viral Infection	7.71	4.23	ABCD2, ABLIM3, ACPP, ADAMTS5, ADRA2A	150
Free Radical Scavenging	Metabolism of reactive oxygen species	8.00	4.14	ABCA1, ALOX5, APOE, CD36, CD44	41
Cell Death and Survival	Cell viability	12.99	4.11	ABCA1, ABCC3, APOD, APOE, AQP3	136
Cell Death and Survival	Cell survival	14.56	4.08	ABCA1, ABCC3, APOD, APOE, AQP3	144
Free Radical Scavenging	Synthesis of reactive oxygen species	8.08	4.02	ABCA1, ALOX5, APOE, CD36, CD44	40
Cellular Movement	Homing of cells	19.31	4.01	ACKR2, ADGRE2, ANOS1, BIN2, C3	71
Cellular Movement	Cell movement of blood cells	20.53	3.94	ACKR2, ADGRE2, ADGRG3, C3, C4A/C4B	85
Cell-To-Cell Signaling and Interaction,Hematological System Development and Function	Binding of lymphocytes	12.12	3.92	CCR1, CD2, CD44, CD58, CD86	26
Cellular Movement	Chemotaxis	19.19	3.90	ACKR2, ADGRE2, ANOS1, BIN2, C3	70
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking	Cell movement of leukocytes	20.09	3.88	ACKR2, ADGRE2, C3, CCR1, CCR7	68
Cellular Movement	Migration of blood cells	19.15	3.79	ACKR2, ADGRE2, ADGRG3, C3, C4A/C4B	82
Cellular Movement	Chemotaxis of myeloid cells	6.94	3.75	C3, CCR1, CCR7, CORO1A, CSF1	27
Cellular Movement,Immune Cell Trafficking	Leukocyte migration	18.63	3.71	ACKR2, ADGRE2, ADGRG3, C3, C4A/C4B	81
Free Radical Scavenging	Production of reactive oxygen species	7.14	3.57	ABCA1, ALOX5, CD36, CD44, CD52	31

Cell-To-Cell Signaling and Interaction,Hematological System Development and Function	Binding of T lymphocytes	10.04	3.52	CCR1, CD2, CD44, CD58, CD86	21
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking	Cell movement of mononuclear leukocytes	13.90	3.50	ACKR2, CCR1, CCR7, CD2, CD44	47
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking,Inflammatory Response	Cell movement of neutrophils	7.83	3.50	CCR1, CCR7, CORO1A, CSF1, CSF2RA	23
Cell-To-Cell Signaling and Interaction,Hematological System Development and Function,Immune Cell Trafficking	Adhesion of mononuclear leukocytes	12.01	3.49	CD2, CD44, CD58, CXCR3, CXCR4	25
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking	Lymphocyte migration	13.86	3.49	ACKR2, CCR1, CCR7, CD2, CD44	36
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking	Cell movement of lymphocytes	13.61	3.47	ACKR2, CCR1, CCR7, CD2, CD44	38
Cellular Movement	Cell movement of myeloid cells	10.32	3.47	C3, CCR1, CCR7, CORO1A, CSF1	40
Cell Morphology,Cellular Movement	Cell spreading	7.74	3.45	C3, CD36, CORO1A, CRYAB, CYR61	29
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking,Inflammatory Response	Cell movement of phagocytes	9.60	3.45	C3, CCR1, CCR7, CD38, CD44	39
Cellular Function and Maintenance	Cellular homeostasis	14.74	3.43	ABCC3, ADM, ADRA2A, APOL1, AQP3	114
Cellular Movement,Immune Cell Trafficking	Migration of lymphatic system cells	14.74	3.43	ACKR2, ADGRG3, CCR1, CCR7, CD2	39
Cell-To-Cell Signaling and Interaction	Binding of lymphatic system cells	12.78	3.43	ADGRG3, CCR1, CD2, CD44, CD58	28
Cell-To-Cell Signaling and Interaction,Hematological System Development and Function,Inflammatory Response	Binding of professional phagocytic cells	7.84	3.43	APOE, CADM1, CCR1, CD14, CD44	22
Cellular Movement,Immune Cell Trafficking	Cell movement of lymphatic system cells	14.50	3.42	ACKR2, ADGRG3, CCR1, CCR7, CD2	41
Cellular Movement	Invasion of cells	22.00	3.41	ADAM8, ADAMTS15, ADM, ANGPTL4, ARHGAP24	133
Cell Death and Survival	Cell death of blood cells	12.42	3.37	BCL6, C3, CCR7, CD247, CD300A	51
Cellular Movement	Migration of tumor cell lines	11.82	3.37	ARHGEF6, ATF3, BMP2, CCR7, CD36	109
Cell-To-Cell Signaling and Interaction	Adhesion of tumor cell lines	12.85	3.35	APOE, CD44, CD55, CDH2, CDH3	55
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking	Migration of mononuclear leukocytes	15.54	3.33	ACKR2, CCR1, CCR7, CD2, CD44	42
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking,Inflammatory Response	Chemotaxis of phagocytes	6.89	3.30	C3, CCR1, CCR7, CORO1A, CSF1	27
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking	Cell movement of granulocytes	8.52	3.25	CCR1, CCR7, CORO1A, CSF1, CSF2RA	28
Gene Expression	Expression of RNA	6.65	3.18	AFF2, ANK2, ANXA4, APOE, ARID5A	156
Inflammatory Response	Inflammatory response	9.03	3.17	ACKR2, ADAM8, APOE, C3, CCR1	51
Cell-To-Cell Signaling and Interaction,Hematological System Development and Function,Immune Cell Trafficking	Adhesion of lymphocytes	10.73	3.16	CD2, CD44, CD58, CXCR3, CXCR4	21
Cell Death and Survival	Cell viability of tumor cell lines	7.44	3.16	ABCA1, ABCC3, APOE, AQP3, ATF3	103

Cell Death and Survival	Cell death of immune cells	12.01	3.15	BCL6, C3, CCR7, CD247, CD300A	49
Cell-mediated Immune Response,Cell-To-Cell Signaling and Interaction,Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking	Adhesion of T lymphocytes	8.75	3.08	CD2, CD44, CD58, CXCR3, CXCR4	17
Cellular Movement	Invasion of tumor cell lines	15.88	3.05	ADAM8, ADAMTS15, ADM, ANGPTL4, ARHGAP24	112
Cellular Movement	Migration of tumor cells	10.38	3.01	CCR7, CD44, CSF1, CSF2RA, CSF3R	21
Cell-To-Cell Signaling and Interaction	Adhesion of lymphatic system cells	11.04	2.96	ADGRG3, CD2, CD44, CD58, CXCR3	22
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking	Homing of leukocytes	8.22	2.89	ACKR2, C3, CCR1, CCR7, CORO1A	35
Cell-To-Cell Signaling and Interaction	Binding of myeloid cells	9.60	2.87	APOE, CADM1, CCR1, CD14, CD44	26
Cell-To-Cell Signaling and Interaction	Binding of leukemia cell lines	12.05	2.79	CD14, CD36, CD44, CSF1, CXCR4	31
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking,Inflammatory Response	Chemotaxis of leukocytes	7.77	2.78	ACKR2, C3, CCR1, CCR7, CORO1A	34
Cell-To-Cell Signaling and Interaction,Cellular Function and Maintenance,Inflammatory Response	Phagocytosis of cells	7.23	2.76	ABCA1, C3, CD14, CD36, CLEC7A	24
Cell Death and Survival,Cellular Compromise	Cytotoxicity of natural killer cells	9.51	2.75	CD2, CD247, CD300A, CD38, CD58	17
Cellular Movement	Cell movement of tumor cells	8.95	2.70	CCR7, CD44, CSF1, CSF2RA, CSF3R	19
Cell-mediated Immune Response,Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking	T cell migration	11.80	2.69	ACKR2, CCR1, CCR7, CD44, CD86	28
Cellular Function and Maintenance	Ion homeostasis of cells	6.71	2.60	ABCC3, ADM, APOL1, ATP2C2, ATP6V1A	41
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking	Cell movement of antigen presenting cells	6.11	2.58	CCR1, CCR7, CD38, CD44, CSF1	18
Cellular Function and Maintenance,Inflammatory Response	Phagocytosis	7.02	2.57	ABCA1, C3, CD14, CD36, CLEC7A	26
Cardiovascular System Development and Function,Cellular Movement	Cell movement of endothelial cells	16.12	2.54	ADAMTS1, ADGRG3, ADM, ANGPTL1, APOE	59
Cell-To-Cell Signaling and Interaction,Hematological System Development and Function,Immune Cell Trafficking,Inflammatory Response	Activation of mononuclear leukocytes	10.76	2.51	CD14, CD2, CD247, CD3E, CD58	34
Cardiovascular System Development and Function,Cellular Movement	Migration of endothelial cells	13.61	2.51	ADAMTS1, ADGRG3, ADM, ANGPTL1, APOE	52
Cell-mediated Immune Response,Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking	Cell movement of T lymphocytes	11.50	2.50	ACKR2, CCR1, CCR7, CD86, CXCR3	26
Cellular Movement	Cell movement of cancer cells	7.64	2.50	CCR7, CD44, CSF1, CSF2RA, CSF3R	16
Cell-To-Cell Signaling and Interaction,Reproductive System Development and Function	Binding of gonadal cell lines	6.45	2.36	APOE, CD44, CDH2, CHST1, ITGA1	12
Cancer,Organismal Injury and Abnormalities	Abdominal neoplasm	17.01	2.35	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	850

Cell-To-Cell Signaling and Interaction	Activation of cells	17.14	2.34	APOE, BMP2, BMPER, C3, CADM1	75
Cell-To-Cell Signaling and Interaction,Hematological System Development and Function,Immune Cell Trafficking,Inflammatory Response	Activation of leukocytes	14.86	2.33	APOE, C3, CADM1, CD14, CD2	50
Cell-To-Cell Signaling and Interaction	Adhesion of leukemia cell lines	7.64	2.31	CD44, CYTIP, DYSF, FOXP1, FYB1	21
Lipid Metabolism,Molecular Transport,Small Molecule Biochemistry	Concentration of lipid	6.38	2.28	ABCA1, ABCG1, ADM, ADRA2A, ALDH8A1	33
Cell-mediated Immune Response,Cellular Function and Maintenance,Hematological System Development and Function	T cell homeostasis	6.46	2.24	BCL6, CD2, CD69, IDO1, IL12RB1	19
Cancer,Organismal Injury and Abnormalities,Tumor Morphology	Invasion of tumor	7.31	2.22	APOE, CD14, CD53, CLU, CTSL	15
Cell-To-Cell Signaling and Interaction,Hematological System Development and Function,Hematopoiesis	Binding of red blood cells	6.26	2.21	C4A/C4B, CD2, CD36, CD55, DMBT1	10
Cardiovascular System Development and Function,Cell-To-Cell Signaling and Interaction	Adhesion of endothelial cells	7.01	2.14	ADGRG3, CD36, CD44, CDH2, CXCR3	23
Cancer,Organismal Injury and Abnormalities	Abdominal cancer	16.75	2.14	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	841
Molecular Transport,Nucleic Acid Metabolism,Small Molecule Biochemistry	Quantity of cyclic nucleotides	6.41	2.10	ADM, ANXA4, APOE, BDKRB2, CXCR4	20
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking,Inflammatory Response	Migration of phagocytes	6.73	2.09	CCR1, CCR7, CD38, CD44, CSF1	22
Molecular Transport	Secretion of molecule	6.66	2.07	ABCA1, ABCG1, ADM, ADRA2A, APOE	32
Inflammatory Response	Immune response of cells	7.02	2.06	ABCA1, APOE, C3, CD14, CD36	40
Cell-To-Cell Signaling and Interaction,Hematological System Development and Function,Immune Cell Trafficking,Inflammatory Response	Activation of lymphocytes	9.45	2.03	CD2, CD247, CD3E, CD58, CD7	30
Cellular Development,Cellular Growth and Proliferation,Hematological System Development and Function,Hematopoiesis,Lymphoid Tissue Structure and Development,Tissue Development	Leukopoiesis	11.81	2.03	ADM, B4GALNT1, BCL6, CD14, CD2	45

Supplementary Table S 4.8 List of top 50 significantly enriched diseases and functional annotations, by significance, for differentially expressed genes between proliferative and secretory phase endometrium by IPA.

Key: Annotations order by significance (-log *p* value); z-score = IPA activation score, >+2 the pathway is activated in secretory relative to proliferative phase, without a z-score = no prediction of direction available from the database; No. of genes = total number of genes involve in one functional annotation

Categories	Diseases or Functions Annotation	-log *p*	z-score	Representative Genes	No. of Genes
Cellular Movement	Cell movement	37.02	5.80	ABCA1, ABI3, ACKR2, ADAMTS1, ADAMTS5	239
Cancer,Dermatological Diseases and Conditions,Organismal Injury and Abnormalities	Skin cancer	33.16	—	ABAT, ABCA4, ABCA6, ABCA8, ABCA9	626
Cellular Movement	Migration of cells	32.08	5.48	ABI3, ACKR2, ADAMTS1, ADGRE2, ADGRG3	212
Cancer,Dermatological Diseases and Conditions,Organismal Injury and Abnormalities	Cutaneous melanoma	30.04	—	ABAT, ABCA4, ABCA6, ABCA8, ABCA9	596
Cancer,Organismal Injury and Abnormalities,Reproductive System Disease	Female genital tract serous carcinoma	29.21	—	ABCA8, APOD, APOE, ARID5B, BUB1B	95
Cancer,Organismal Injury and Abnormalities	Malignant genitourinary solid tumor	28.29	0.85	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	672
Cancer,Organismal Injury and Abnormalities	Genitourinary tumor	27.74	0.31	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	678
Cancer,Organismal Injury and Abnormalities,Reproductive System Disease	Genital tumor	27.40	—	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	593
Cancer,Organismal Injury and Abnormalities,Reproductive System Disease	Breast or gynecological cancer	27.21	—	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	527
Cancer,Organismal Injury and Abnormalities,Reproductive System Disease	Genital tract cancer	26.93	—	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	585
Organismal Injury and Abnormalities,Reproductive System Disease	**Endometriosis**	26.76	—	ABCG1, ALDH1A3, APOE, BCL6, BIRC5	89
Cell Death and Survival	Necrosis	26.31	1.76	ABCC3, ADAMTS16, ADM, ALOX5, AMIGO2	248
Cancer,Organismal Injury and Abnormalities,Reproductive System Disease	Breast or ovarian cancer	25.91	—	ABAT, ABCA1, ABCA4, ABCA8, ABCA9	376
Cell Death and Survival	Apoptosis	25.65	0.49	ADM, ALOX5, AMIGO2, ANGPTL4, ANXA4	235
Cancer,Organismal Injury and Abnormalities	Breast or pancreatic cancer	25.55	—	ABAT, ABCA1, ABCA4, ABCA8, ABCA9	476
Organismal Injury and Abnormalities,Reproductive System Disease	Benign pelvic disease	25.47	—	ABCA8, ABCG1, ADRA2A, ALDH1A3, APOE	112
Cancer,Organismal Injury and Abnormalities	Melanoma	24.72	—	ABAT, ABCA4, ABCA6, ABCA8, ABCA9	651
Cancer,Organismal Injury and Abnormalities	Pelvic cancer	24.61	—	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	607
Cancer,Organismal Injury and Abnormalities	Pelvic tumor	24.46	—	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	613
Cancer,Organismal Injury and Abnormalities	Genitourinary adenocarcinoma	23.94	—	ABCA1, ABCA4, ABCA6, ABCA8, ABCA9	541
Cancer,Organismal Injury and Abnormalities	Genitourinary carcinoma	23.48	—	ABCA1, ABCA4, ABCA6, ABCA8, ABCA9	591
Cancer,Organismal Injury and Abnormalities	Lymphatic system tumor	22.67	—	ABAT, ABCC9, ABCG1, ABI3BP, ADAMTS15	225
Cancer,Organismal Injury and Abnormalities	Incidence of tumor	22.61	0.81	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	661
Cancer,Organismal Injury and Abnormalities	Lymphoid cancer	22.47	—	ABAT, ABCC9, ABCG1, ABI3BP, ADAMTS15	222
Cell-To-Cell Signaling and Interaction	Binding of blood cells	22.09	5.21	ADRA2A, APOE, C3, C4A/C4B, CADM1	60
Cellular Movement	Invasion of cells	22.00	3.41	ADAM8, ADAMTS15, ADM, ANGPTL4, ARHGAP24	133
Hematological Disease,Immunological Disease	Lymphoproliferative disorder	21.77	—	ABAT, ABCC9, ABCG1, ABI3BP, ADAMTS15	219

Cancer,Organismal Injury and Abnormalities,Reproductive System Disease	Female genital neoplasm	21.76	—	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	439
Connective Tissue Disorders,Inflammatory Disease,Organismal Injury and Abnormalities,Skeletal and Muscular Disorders	Rheumatic Disease	21.74	—	ADAM8, ADAMTS1, ADAMTS15, ADAMTS16, ADAMTS19	173
Cancer,Organismal Injury and Abnormalities,Reproductive System Disease	Tumorigenesis of reproductive tract	21.64	—	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	438
Cancer,Hematological Disease,Organismal Injury and Abnormalities	Lymphocytic neoplasm	21.47	—	ABAT, ABCC9, ABCG1, ABI3BP, ADAMTS15	218
Cancer,Organismal Injury and Abnormalities	Advanced malignant tumor	21.45	0.43	ABAT, ABCA1, ABCA8, ABCC3, ADAMTS1	126
Cancer,Hematological Disease,Immunological Disease,Organismal Injury and Abnormalities	Neoplasia of leukocytes	21.45	—	ABAT, ABCC9, ABCG1, ABI3BP, ADAMTS15	218
Dermatological Diseases and Conditions,Organismal Injury and Abnormalities	Psoriasis	21.39	—	ABCA8, ACPP, ADM, ALDH1A3, APOC1	102
Neurological Disease,Organismal Injury and Abnormalities	Epileptic seizure	21.10	—	ADM, ANGPTL4, ANOS1, ARID5B, ATF3	51
Cancer,Organismal Injury and Abnormalities	Neoplasia of cells	21.09	-0.04	ABAT, ABCA1, ABCA4, ABCA6, ABCC3	425
Cancer,Organismal Injury and Abnormalities,Reproductive System Disease	Uterine serous papillary cancer	21.09	—	ABCA8, APOD, BUB1B, C7, CCNB1	56
Cancer,Organismal Injury and Abnormalities,Reproductive System Disease	Breast cancer	21.01	—	ABAT, ABCA1, ABCA8, ABCA9, ABCC9	266
Cancer,Organismal Injury and Abnormalities	Adenocarcinoma	20.99	—	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	808
Cancer,Organismal Injury and Abnormalities	Cancer of cells	20.81	0.51	ABAT, ABCA1, ABCA4, ABCA6, ABCC3	418
Cancer,Organismal Injury and Abnormalities	Advanced malignant solid tumor	20.80	—	ABAT, ABCA1, ABCA8, ABCC3, ADAMTS1	115
Cellular Movement	Cell movement of blood cells	20.53	3.94	ACKR2, ADGRE2, ADGRG3, C3, C4A/C4B	85
Cancer,Organismal Injury and Abnormalities	Abdominal adenocarcinoma	20.35	—	ABCA1, ABCA4, ABCA6, ABCA8, ABCA9	796
Cancer,Organismal Injury and Abnormalities,Reproductive System Disease	Female genital tract cancer	20.29	—	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	424
Cell Death and Survival	Apoptosis of tumor cell lines	20.21	0.39	ADM, ALOX5, ARHGEF6, ARSB, ATF3	170
Cancer,Organismal Injury and Abnormalities	Advanced extracranial solid tumor	20.15	—	ABCA1, ABCA8, ABCC3, ADAMTS1, C1R	98
Cancer,Hematological Disease,Organismal Injury and Abnormalities	Mature lymphocytic neoplasm	20.09	—	ABCC9, ABCG1, ABI3BP, ADAMTS15, ADAMTS16	192
Cellular Movement,Hematological System Development and Function,Immune Cell Trafficking	Cell movement of leukocytes	20.09	3.88	ACKR2, ADGRE2, C3, CCR1, CCR7	68
Cancer,Hematological Disease,Immunological Disease,Organismal Injury and Abnormalities	Lymphocytic cancer	20.07	—	ABAT, ABCC9, ABCG1, ABI3BP, ADAMTS15	208
Cancer,Organismal Injury and Abnormalities	Breast or colorectal cancer	20.02	0.28	ABAT, ABCA1, ABCA4, ABCA6, ABCA8	547

Supplementary Table S 4.9 List of enriched Reactome canonical pathways ($p \leq 0.05$, n=57) for differentially expressed genes between proliferative and secretory phase endometrium by PANTHER.

Key: Direction up/down = up/down-regulated in secretory than in proliferative phase; No. of genes = total number of genes from the DE list involved in a particular pathway

Reactome Pathway Code	Canonical Pathways	*p*	Direction	No. of Genes
R-HSA-69620	Cell Cycle Checkpoints	4.78E-04	down	13
R-HSA-69278	Cell Cycle, Mitotic	7.00E-04	down	18
R-HSA-73894	DNA Repair	9.98E-04	down	7
R-HSA-1640170	Cell Cycle	1.57E-03	down	22
R-HSA-5693532	DNA Double-Strand Break Repair	1.70E-03	down	6
R-HSA-5693567	HDR through Homologous Recombination or Single Strand Annealing	4.71E-03	down	5
R-HSA-5693538	Homology Directed Repair	4.71E-03	down	5
R-HSA-176409	APC/C:Cdc20 mediated degradation of mitotic proteins	5.98E-03	down	5
R-HSA-174143	APC/C-mediated degradation of cell cycle proteins	5.98E-03	down	5
R-HSA-176814	Activation of APC/C and APC/C:Cdc20 mediated degradation of mitotic proteins	5.98E-03	down	5
R-HSA-453276	Regulation of mitotic cell cycle	5.98E-03	down	5
R-HSA-68886	M Phase	8.07E-03	down	11
R-HSA-69618	Mitotic Spindle Checkpoint	8.69E-03	down	5
R-HSA-141444	Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal	8.69E-03	down	5
R-HSA-141424	Amplification of signal from the kinetochores	8.69E-03	down	5
R-HSA-166658	Complement cascade	8.98E-03	up	14
R-HSA-5693607	Processing of DNA double-strand break ends	9.83E-03	down	4
R-HSA-5689880	Ub-specific processing proteases	1.03E-02	down	5
R-HSA-76002	Platelet activation, signaling and aggregation	1.06E-02	up	32
R-HSA-179419	APC:Cdc20 mediated degradation of cell cycle proteins prior to satisfation of the cell cycle checkpoint	1.07E-02	down	4
R-HSA-114608	Platelet degranulation	1.09E-02	up	19
R-HSA-76005	Response to elevated platelet cytosolic Ca2+	1.09E-02	up	19
R-HSA-1280215	Cytokine Signaling in Immune system	1.09E-02	up	59
R-HSA-5688426	Deubiquitination	1.16E-02	down	6
R-HSA-176408	Regulation of APC/C activators between G1/S and early anaphase	1.33E-02	down	4
R-HSA-9018519	Estrogen-dependent gene expression	1.61E-02	down	5
R-HSA-68877	Mitotic Prometaphase	1.66E-02	down	8
R-HSA-75153	Apoptotic execution phase	1.96E-02	down	3
R-HSA-877300	Interferon gamma signaling	1.96E-02	up	7
R-HSA-913531	Interferon Signaling	1.96E-02	up	7
R-HSA-2559586	DNA Damage/Telomere Stress Induced Senescence	2.09E-02	down	4
R-HSA-432040	Vasopressin regulates renal water homeostasis via Aquaporins	2.13E-02	up	3
R-HSA-445717	Aquaporin-mediated transport	2.13E-02	up	3
R-HSA-1538133	G0 and Early G1	2.29E-02	down	3
R-HSA-174184	Cdc20:Phospho-APC/C mediated degradation of Cyclin A	2.39E-02	down	3
R-HSA-391251	Protein folding	2.52E-02	up	4
R-HSA-390466	Chaperonin-mediated protein folding	2.52E-02	up	4
R-HSA-982772	Growth hormone receptor signaling	2.60E-02	up	3
R-HSA-8878166	Transcriptional regulation by RUNX2	2.61E-02	down	11
R-HSA-8953897	Cellular responses to external stimuli	2.70E-02	up	24
R-HSA-5619102	SLC transporter disorders	2.71E-02	up	8
R-HSA-8941326	RUNX2 regulates bone development	2.73E-02	down	2
R-HSA-69306	DNA Replication	2.99E-02	down	3
R-HSA-69239	Synthesis of DNA	2.99E-02	down	3
R-HSA-69242	S Phase	3.44E-02	down	4
R-HSA-453279	Mitotic G1-G1/S phases	3.46E-02	down	7
R-HSA-2500257	Resolution of Sister Chromatid Cohesion	3.49E-02	down	7
R-HSA-977606	Regulation of Complement cascade	3.57E-02	up	11
R-HSA-69206	G1/S Transition	3.75E-02	down	4
R-HSA-179409	APC-Cdc20 mediated degradation of Nek2A	3.76E-02	down	3

R-HSA-5654687	Downstream signaling of activated FGFR1	3.78E-02	down	3
R-HSA-168256	Immune System	4.00E-02	up	167
R-HSA-196854	Metabolism of vitamins and cofactors	4.02E-02	up	16
R-HSA-2586552	Signaling by Leptin	4.16E-02	up	2
R-HSA-449147	Signaling by Interleukins	4.41E-02	up	51
R-HSA-1632852	Macroautophagy	4.61E-02	up	2
R-HSA-6785807	Interleukin-4 and Interleukin-13 signaling	4.62E-02	up	26

Supplementary Table S 5.1 Top 50 nominally differentially expressed (DE) genes between endometriosis cases (n=34) versus clinic controls (n=10) within proliferative samples, with comparison to results from sensitivity analysis (n=38, subset of proliferative samples), previous reports from the literature, and lookup of endometrium *cis*-eQTL SNPs.

Key: DE = differentially expressed based on *p* value, adj p = adjusted *p* value in DE analysis, log_2 FC = log_2 fold change (positive value indicates higher expression in endometriosis cases compared to controls), Chr:pos = chromosome: position from GRCh37/hg19, FDR p = false discovery rate *p* value in eQTL analysis (eQTLs shown are all with *p* value <0.05); Gene marks with * are non-protein coding in gene biotype (i.e. pseudogene, lincRNA, antisense, etc); On/Off = mentioned in on/off gene prediction analysis.

Endometriosis Cases vs. Control in Proliferative Top 50				**Sensitivity Testing (n=38) DE genes ranked by adj p** (Top 150 DE genes with details)				**Reported in Previous Publications**	**Endometrium *cis*-eQTL SNP (1Mb)**	
DE Gene	adj p	p	Log_2 FC	Rank	adj p	p	Log_2 FC		Chr:pos in hg19	FDR p
ZNF26	1.50E-06	3.67E-11	-3.09	10735	—	—	—	—	—	—
*RPL21P119**	1.34E-02	6.59E-07	5.45	4	0.14	1.38E-05	5.76	—	—	—
MLXIPL	3.69E-02	4.13E-06	-1.44	128	0.56	1.79E-03	-0.73	—	—	—
SAA2	3.69E-02	5.44E-06	-3.05	140	0.59	2.07E-03	-1.46	Burney et al. 2007	—	—
*RP11-62H7.3**	3.69E-02	4.34E-06	-1.14	20	0.26	1.25E-04	-1.04	—	—	—
*CYP2B7P**	3.69E-02	4.96E-06	-1.75	7604	—	—	—	—	—	—
PRKCG	6.05E-02	1.11E-05	-1.79	16	0.26	1.15E-04	-1.72	—	—	—
GRAMD2	6.05E-02	1.19E-05	-1.48	13929	—	—	—	—	—	—
THSD4	9.21E-02	2.03E-05	-1.17	247	—	—	—	—	—	—
PRB4	0.10	2.47E-05	-1.21	49	0.38	4.70E-04	-1.21	—	—	—
*AC068138.1**	0.12	3.13E-05	-1.64	11783	—	—	—	—	—	—
AREG	0.14	4.11E-05	-3.28	22	0.28	1.67E-04	-3.00	—	—	—
MYO3B	0.14	4.47E-05	-0.53	42	0.35	3.71E-04	-0.52	—	—	—
ENPP3	0.14	4.92E-05	1.19	7	0.20	3.46E-05	1.39	—	—	—
CYP2S1	0.14	5.33E-05	-1.24	1183	—	—	—	—	—	—
*PSMA2P3**	0.15	5.96E-05	-2.90	50	0.38	4.73E-04	-2.63	—	—	—
CDH6	0.16	6.76E-05	-1.08	1420	—	—	—	—	—	—
*RP11-317B3.2**	0.17	7.40E-05	-1.38	84	0.47	9.84E-04	-1.20	—	—	—
HAPLN3	0.17	7.83E-05	-0.53	18	0.26	1.31E-04	-0.51	—	—	—
TRIM31	0.17	8.66E-05	-1.21	353	—	—	—	—	—	—
*GBP1P1**	0.17	8.79E-05	-1.10	16475	—	—	—	—	—	—
NXNL2	0.17	9.65E-05	-0.66	43	0.38	4.17E-04	-0.55	—	—	—
ALAS2	0.17	9.71E-05	2.53	15112	—	—	—	Fung et al. 2018 (On/Off)	—	—
EDN2	0.17	1.07E-04	-1.65	1	2.37E-02	5.91E-07	-2.13	Fung et al. 2018 (On/Off)	—	—
VSTM1	0.17	1.08E-04	-1.99	11842	—	—	—	—	—	—
SLITRK2	0.17	1.09E-04	-3.31	18018	—	—	—	—	—	—
GIPR	0.17	1.13E-04	-1.39	1439	—	—	—	—	19:46255858	1.16E-02
GRID2	0.17	1.17E-04	-1.57	13049	—	—	—	—	—	—
C1orf141	0.18	1.27E-04	-3.46	11067	—	—	—	—	—	—
NUAK2	0.19	1.41E-04	-0.91	59	0.38	5.62E-04	-0.86	—	—	—
EREG	0.19	1.51E-04	-2.63	187	—	—	—	—	—	—
*RP1-35C21.2**	0.19	1.51E-04	-2.47	13033	—	—	—	—	—	—
*ZNF663P**	0.19	1.54E-04	-1.03	474	—	—	—	—	—	—
C1QL3	0.20	1.67E-04	0.94	129	0.56	1.80E-03	0.93	—	—	—
*RP11-466P24.6**	0.20	1.96E-04	-1.14	19	0.26	1.22E-04	-1.26	—	—	—
TCTEX1D1	0.20	1.80E-04	-1.51	15243	—	—	—	—	—	—
FAT2	0.20	1.87E-04	-1.24	1284	—	—	—	—	5:150964798	4.46E-04
CDC42BPG	0.20	1.88E-04	-0.67	320	—	—	—	—	—	—
RUFY4	0.20	1.94E-04	-1.40	329	—	—	—	—	—	—
ALPP	0.20	1.99E-04	-1.62	89	0.48	1.10E-03	-1.70	—	—	—
VCAM1	0.20	2.03E-04	-1.16	1425	—	—	—	—	—	—
CD1D	0.21	2.14E-04	-0.58	254	—	—	—	Fung et al. 2018 (DE)	—	—
RND1	0.21	2.28E-04	-1.28	6	0.20	3.11E-05	-1.65	—	—	—
PNLDC1	0.21	2.32E-04	-1.07	72	0.44	8.07E-04	-1.16	—	—	—
TTC29	0.22	2.48E-04	-1.67	13394	—	—	—	—	4:147560065	0.41
CTLA4	0.22	2.61E-04	-1.29	375	—	—	—	—	—	—
TBX22	0.22	2.64E-04	-1.70	301	—	—	—	—	—	—
MUC4	0.22	2.65E-04	-1.59	664	—	—	—	Burney et al. 2007	—	—
TMEM190	0.23	2.74E-04	-1.52	20021	—	—	—	—	—	—
*ARGFXP1**	0.23	2.78E-04	-3.97	14841	—	—	—	—	—	—

Supplementary Table S 5.2 Top 51 differentially expressed (DE) genes between ASRM stage I/II cases (n=18) versus clinic controls (n=10) within proliferative samples, with comparison to results from sensitivity analysis (n=24, subset of proliferative samples), previous reports from the literature, and lookup of endometrium *cis*-eQTL SNPs.

Key: DE = differentially expressed based on adj.*p* value, adj p = adjusted *p* value in DE analysis, log_2 FC = log_2 fold change (positive value indicates higher expression in ASRM stage I/II cases compared to controls), Chr:pos = chromosome: position from GRCh37/hg19, FDR p = false discovery rate *p* value in eQTL analysis (eQTLs shown are all with *p* value <0.05); Gene marks with * are non-protein coding in gene biotype (i.e. pseudogene, lincRNA, antisense, etc); On/Off = mentioned in on/off gene prediction analysis.

Endometriosis Stage I/II vs. Control in Proliferative Top 51				**Sensitivity Testing (n=24) DE genes ranked by adj p** (Top 150 DE genes with details)				**Reported in Previous Publications**	**Endometrium *cis*-eQTL SNP (1Mb)**	
DE Gene	adj p	p	Log_2 FC	Rank	adj p	p	Log_2 FC		Chr:pos in hg19	FDR p
ZNF26	6.46E-04	2.18E-08	-2.88	5722	—	—	—	—	—	—
*RP11-62H7.3**	3.62E-03	2.44E-07	-1.44	20	4.54E-02	3.20E-05	-1.26	—	—	—
EPYC	5.18E-03	5.24E-07	-3.08	11723	—	—	—	—	—	—
EDN2	7.96E-03	1.07E-06	-2.20	1	2.00E-03	7.05E-08	-2.45	Fung et al. 2018 (On/Off)	—	—
GRAMD2	9.12E-03	2.01E-06	-1.74	1380	—	—	—	—	—	—
*CYP2B7P**	9.12E-03	2.86E-06	-1.98	17	3.91E-02	2.34E-05	-2.02	—	—	—
KRT17	9.12E-03	1.85E-06	-2.27	4	6.45E-03	9.07E-07	-2.10	Burney et al. 2007	—	—
DUSP5	9.12E-03	2.95E-06	-1.90	3	6.45E-03	7.27E-07	-2.25	—	—	—
NR4A3	9.12E-03	3.15E-06	-3.02	7	1.80E-02	4.42E-06	-3.46	—	—	—
NUAK2	9.12E-03	3.19E-06	-1.18	18	4.40E-02	2.94E-05	-1.08	—	—	—
ATF3	9.12E-03	3.38E-06	-3.47	15	3.87E-02	2.04E-05	-3.78	—	—	—
*ESPNP**	1.46E-02	5.91E-06	-2.34	1970	—	—	—	—	—	—
*RP11-6J24.3**	2.40E-02	1.15E-05	-1.86	284	—	—	—	—	—	—
S100A2	2.40E-02	1.21E-05	-1.46	340	—	—	—	—	—	—
ATP6V1B1	2.40E-02	1.10E-05	-2.12	7302	—	—	—	—	—	—
EFNA1	2.44E-02	1.31E-05	-0.70	22	4.98E-02	3.85E-05	-0.68	—	—	—
TNFAIP3	2.56E-02	1.53E-05	-1.20	6	8.55E-03	1.81E-06	-1.46	—	—	—
ELF3	2.56E-02	1.61E-05	-1.26	11	2.96E-02	1.16E-05	-1.33	—	—	—
GRID2	2.56E-02	1.64E-05	-1.92	1036	—	—	—	—	—	—
*LINC00284**	3.12E-02	2.41E-05	-1.88	1832	—	—	—	—	—	—
CDKN1A	3.12E-02	2.24E-05	-1.22	149	0.26	1.36E-03	-0.82	—	—	—
HAPLN3	3.12E-02	2.29E-05	-0.62	34	7.24E-02	8.66E-05	-0.57	—	—	—
SPRY1	3.12E-02	2.36E-05	-0.82	10	2.60E-02	9.15E-06	-0.90	—	—	—
GDF15	3.45E-02	2.80E-05	-1.54	21	4.85E-02	3.58E-05	-1.52	—	—	—
LYPD1	3.45E-02	2.98E-05	-2.91	12933	—	—	—	—	—	—
NKD2	3.45E-02	3.33E-05	-2.48	12819	—	—	—	—	—	—
CDKL2	3.45E-02	3.51E-05	-1.82	2755	—	—	—	—	—	—
TNFSF9	3.45E-02	3.60E-05	-2.13	47	8.94E-02	1.46E-04	-1.82	—	—	—
RND1	3.45E-02	3.50E-05	-1.58	13	2.96E-02	1.35E-05	-1.87	—	—	—
HBA1	3.45E-02	3.51E-05	2.60	170	—	—	—	Burney et al. 2007	—	—
RP11-386G21.1	3.45E-02	3.55E-05	-1.88	38	7.96E-02	1.12E-04	-1.75	—	—	—
BHLHE40	3.94E-02	4.40E-05	-1.43	2	6.45E-03	5.16E-07	-1.70	—	—	—
EGR2	3.94E-02	4.51E-05	-2.75	16	3.91E-02	2.28E-05	-3.28	—	—	—
GBP1	3.94E-02	4.51E-05	-1.25	50	9.48E-02	1.67E-04	-1.37	Fung et al. 2018 (DE)	—	—
TBX22	4.33E-02	5.19E-05	-2.20	28432	—	—	—	—	—	—
PRKCG	4.33E-02	5.25E-05	-1.85	77	0.14	3.89E-04	-1.74	—	—	—
*ANKRD20A11P**	4.44E-02	6.05E-05	-2.63	27	6.39E-02	6.12E-05	-2.91	—	—	—
C2orf54	4.44E-02	6.19E-05	-1.96	1684	—	—	—	—	—	—
*RP11-403D15.2**	4.44E-02	6.34E-05	-1.11	37	7.96E-02	1.11E-04	-1.20	—	—	—
*RP3-416H24.1**	4.44E-02	6.40E-05	-1.64	56	0.11	2.29E-04	-1.51	—	—	—
MYO3B	4.44E-02	6.08E-05	-0.57	86	0.16	4.86E-04	-0.55	—	—	—
SYTL3	4.44E-02	6.42E-05	-0.75	52	0.11	1.96E-04	-0.64	—	—	—
SPATA6L	4.44E-02	6.43E-05	-1.21	2076	—	—	—	—	9:4775634	0.27
ANKRD65	4.60E-02	6.88E-05	-1.29	4133	—	—	—	—	1:1350601	0.28
SLITRK2	4.60E-02	7.13E-05	-3.92	33606	—	—	—	—	—	—
ANXA1	4.60E-02	7.11E-05	-1.02	164	—	—	—	—	—	—
KCNJ2	4.68E-02	7.41E-05	-1.26	2248	—	—	—	—	—	—
THSD4	4.68E-02	7.72E-05	-1.20	464	—	—	—	—	—	—
*RNF157-AS1**	4.68E-02	7.73E-05	-0.70	43	8.94E-02	1.51E-04	-0.63	—	—	—
VWA7	4.94E-02	8.37E-05	-1.15	726	—	—	—	—	—	—
*RP11-356I2.4**	4.94E-02	8.49E-05	-0.91	9	2.60E-02	9.07E-06	-1.09	—	—	—

Supplementary Table S 5.3 Top 50 nominally differentially expressed (DE) genes between ASRM stage III/IV cases (n=16) versus clinic controls (n=10) within proliferative samples, with comparison to results from sensitivity analysis (n=21, subset of proliferative samples), previous reports from the literature and lookup of endometrium *cis*-eQTL SNPs.

Key: DE = differentially expressed based on *p* value, adj p = adjusted *p* value in DE analysis, log_2 FC = log_2 fold change (positive value indicates higher expression in ASRM stage III/IV cases compared to controls), Chr:pos = chromosome: position from GRCh37/hg19, FDR p = false discovery rate *p* value in eQTL analysis (eQTLs shown are all with *p* value <0.05); Gene marks with * are non-protein coding in gene biotype (i.e. pseudogene, lincRNA, snRNA, antisense, etc).

Endometriosis Stage III/IV vs. Control in Proliferative Top 50				**Sensitivity Testing (n=21) DE genes ranked by adj p** (Top 150 DE genes with details)				**Reported in Previous Publications**	**Endometrium *cis*-eQTL SNP (1Mb)**	
DE Gene	adj p	p	Log_2 FC	Rank	adj p	p	Log_2 FC		Chr:pos in hg19	FDR p
ZNF26	9.21E-06	2.26E-10	-3.37	757	—	—	—	—	—	—
PRB4	0.10	4.91E-06	-1.50	11	0.41	1.63E-04	-1.48	—	—	—
MLXIPL	0.13	1.61E-05	-1.54	47	0.56	5.34E-04	-0.91	—	—	—
*AC068138.1**	0.13	1.04E-05	-2.02	900	—	—	—	—	—	—
SAA2	0.13	1.42E-05	-3.30	1088	—	—	—	Burney et al. 2007	—	—
*RP11-728K20.2**	0.24	3.94E-05	-1.16	25	0.54	3.38E-04	-1.09	—	—	—
XKR9	0.24	4.05E-05	1.77	9	0.41	1.51E-04	1.91	—	—	—
*RPL21P119**	0.24	4.66E-05	4.16	1	0.12	3.10E-06	6.56	—	—	—
*ZNF663P**	0.32	7.01E-05	-1.22	225	—	—	—	—	—	—
HSD17B2	0.34	8.87E-05	1.79	31	0.55	4.23E-04	1.80	—	—	—
SPAG6	0.34	1.04E-04	-2.40	7025	—	—	—	—	—	—
ENPP3	0.34	1.01E-04	1.28	3	0.41	5.83E-05	1.50	—	—	—
*RP11-396C23.2**	0.34	1.10E-04	-1.64	21	0.42	2.21E-04	-1.81	—	—	—
EPYC	0.35	1.30E-04	-2.36	20401	—	—	—	—	—	—
*ECEL1P2**	0.35	1.28E-04	-1.84	14	0.41	1.18E-04	-1.96	—	—	—
*RP11-92K15.1**	0.36	1.49E-04	-0.87	18	0.42	2.14E-04	-0.94	—	—	—
C1orf141	0.36	1.52E-04	-4.04	1103	—	—	—	—	—	—
CD1D	0.38	1.66E-04	-0.67	235	—	—	—	Fung et al. 2018 (DE)	—	—
*RP11-317B3.2**	0.41	2.45E-04	-1.50	119	0.62	1.76E-03	-1.33	—	—	—
CADM3	0.41	2.45E-04	-1.92	5590	—	—	—	—	—	—
THSD4	0.41	2.47E-04	-1.14	460	—	—	—	—	—	—
ALPP	0.41	2.49E-04	-1.87	78	0.59	1.19E-03	-1.88	—	—	—
PRKCG	0.41	2.11E-04	-1.73	60	0.56	8.27E-04	-1.69	—	—	—
ATP6V1B1	0.41	2.24E-04	-1.80	21833	—	—	—	—	—	—
GRIK5	0.41	2.49E-04	-1.29	49	0.56	6.08E-04	-1.31	—	—	—
GAS8	0.44	2.81E-04	-0.39	52	0.56	6.53E-04	-0.35	—	—	—
*RPS14P3**	0.44	2.92E-04	-0.56	32	0.55	4.44E-04	-0.63	—	—	—
GATA5	0.44	3.10E-04	-2.00	75	0.59	1.11E-03	-2.12	—	—	—
C12orf57	0.44	3.10E-04	0.43	134	0.64	2.24E-03	0.42	Fung et al. 2018 (DE)	—	—
*CYP2B7P**	0.47	3.44E-04	-1.54	179	—	—	—	—	—	—
*RP11-22L13.1**	0.47	3.65E-04	-2.71	16	0.41	1.25E-04	-3.35	—	—	—
*RP11-1396O13.1**	0.47	3.79E-04	-1.09	354	—	—	—	—	—	—
C1orf85	0.47	3.91E-04	0.39	91	0.60	1.37E-03	0.39	—	1:156154860	0.30
DNAH1	0.47	3.85E-04	-0.51	451	—	—	—	—	—	—
AREG	0.48	4.63E-04	-3.35	96	0.62	1.54E-03	-3.00	—	—	—
FGL1	0.48	4.93E-04	1.36	63	0.56	8.88E-04	1.25	—	—	—
HELB	0.48	5.38E-04	-0.37	8	0.41	1.49E-04	-0.46	—	—	—
*AC068134.5**	0.48	4.19E-04	-1.93	57	0.56	7.51E-04	-2.04	—	—	—
KCNK13	0.48	4.51E-04	1.98	35	0.55	4.77E-04	2.32	—	—	—
OR1N1	0.48	4.54E-04	1.54	112	0.62	1.93E-03	1.49	—	—	—
*RPL3P5**	0.48	4.76E-04	-1.04	19	0.42	2.18E-04	-1.22	—	—	—
*RNU6ATAC27P**	0.48	4.93E-04	-1.04	860	—	—	—	—	—	—
RUFY4	0.48	5.02E-04	-1.51	450	—	—	—	—	—	—
*CTD-3010D24.3**	0.48	5.28E-04	-1.33	1274	—	—	—	—	—	—
LMLN	0.48	5.39E-04	-0.28	109	0.62	1.86E-03	-0.27	—	—	—
OR2A4	0.48	5.45E-04	1.43	141	0.64	2.36E-03	1.52	—	—	—
ALPPL2	0.53	6.14E-04	-1.76	61	0.56	8.48E-04	-1.80	—	—	—
*CTC-550B14.1**	0.55	6.44E-04	-2.76	17	0.42	1.92E-04	-3.28	—	—	—
*RP1-35C21.2**	0.56	6.69E-04	-2.66	6787	—	—	—	—	—	—
*RP11-68I3.2**	0.56	6.89E-04	-1.61	1009	—	—	—	—	—	—

Supplementary Table S 5.4 Top 50 nominally differentially expressed (DE) genes between ASRM stage I/II cases (n=18) versus stage III/IV (n=16) within proliferative samples, with comparison to results from sensitivity analysis (n=31, subset of proliferative samples), previous reports from the literature, and lookup of endometrium *cis*-eQTL SNPs.

Key: **Bold gene name =** located within ±500kb of GWAS SNP at 3'UTR of ***BMF*** for Endometriosis Sub-phenotype; DE = differentially expressed based on *p* value, adj p = adjusted *p* value in DE analysis, log_2 FC = log_2 fold change (positive value indicates higher expression in ASRM stage III/IV cases compared to stage I/II), Chr:pos = chromosome: position from GRCh37/hg19, FDR p = false discovery rate *p* value in eQTL analysis (eQTLs shown are all with *p* value <0.05); Gene with * are non-protein coding in gene biotype.

Endometriosis Stage III/IV vs. I/II in Proliferative Top 50				**Sensitivity Testing (n=31) DE genes ranked by adj p** (Top 150 DE genes with details)				**Reported in Previous Publications**	**Endometrium *cis*-eQTL SNP (1Mb)**	
DE Gene	adj p	p	Log_2 FC	Rank	adj p	p	Log_2 FC		Chr:pos in hg19	FDR p
*ANKRD20A11P**	2.30E-05	9.20E-10	3.50	2	6.27E-04	3.13E-08	3.24	—	—	—
*CYP4F29P**	1.32E-04	1.06E-08	4.02	3	2.02E-03	1.51E-07	3.71	—	—	—
*SNX18P13**	1.12E-02	1.35E-06	3.66	7	6.51E-02	8.22E-06	3.41	—	—	—
NR4A3	2.23E-02	3.58E-06	2.62	5	6.51E-02	1.13E-05	2.67	—	—	—
GDF15	2.45E-02	7.42E-06	1.44	67	0.31	5.31E-04	1.04	—	—	—
SPRY1	2.45E-02	7.75E-06	0.75	36	0.24	2.15E-04	0.60	—	—	—
IER3	2.45E-02	8.34E-06	1.63	39	0.25	2.51E-04	1.34	—	—	—
SERPINE1	2.45E-02	8.82E-06	2.68	1	3.89E-04	9.70E-09	3.24	—	—	—
*GS1-124K5.3**	2.45E-02	6.60E-06	-0.40	4	3.89E-02	3.88E-06	-0.42	—	—	—
ATF3	3.94E-02	1.63E-05	2.81	44	0.25	2.99E-04	2.58	—	—	—
THBS1	3.94E-02	1.74E-05	1.70	340	—	—	—	—	15:38758027	0.13
EGR2	4.01E-02	1.93E-05	2.51	30	0.21	1.64E-04	2.36	—	—	—
DUSP5	4.02E-02	2.09E-05	1.51	68	0.31	5.35E-04	1.27	—	—	—
CXCL3	4.26E-02	2.39E-05	2.51	42	0.25	2.72E-04	2.49	—	—	—
RRAD	4.40E-02	3.17E-05	1.17	24	0.21	1.13E-04	1.14	—	—	—
C10orf10	4.40E-02	3.26E-05	1.43	1082	—	—	—	—	—	—
ARL4C	4.40E-02	3.39E-05	0.90	185	—	—	—	—	—	—
ZNF808	4.40E-02	3.49E-05	-0.26	13	0.16	5.32E-05	-0.27	—	—	—
KCNK13	4.40E-02	2.84E-05	1.97	10	0.14	3.51E-05	2.05	—	—	—
CACNG4	4.40E-02	3.53E-05	1.22	23	0.21	8.01E-05	1.06	—	—	—
CITED2	4.98E-02	4.19E-05	0.78	29	0.21	1.63E-04	0.67	—	—	—
SOX6	4.99E-02	4.40E-05	0.61	182	—	—	—	Fung et al. 2018 (On/Off)	—	—
*RHOT1P2**	5.97E-02	5.51E-05	2.44	52	0.26	3.42E-04	2.14	—	—	—
*ADH1C**	7.23E-02	6.96E-05	1.15	244	—	—	—	—	—	—
KRT17	8.94E-02	8.96E-05	1.64	645	—	—	—	Burney et al. 2007	—	—
SUSD4	9.26E-02	1.01E-04	0.88	56	0.26	3.70E-04	0.84	—	—	—
MAP10	9.26E-02	9.76E-05	-0.35	77	0.33	6.32E-04	-0.32	—	—	—
C8orf4	9.26E-02	1.04E-04	1.10	149	0.45	1.67E-03	0.87	—	—	—
CDH16	9.29E-02	1.08E-04	1.95	458	—	—	—	—	—	—
RND3	0.11	1.30E-04	0.88	83	0.36	7.57E-04	0.79	—	—	—
TDRD12	0.11	1.38E-04	-1.36	169	—	—	—	—	—	—
EFNA1	0.11	1.43E-04	0.53	226	—	—	—	—	—	—
SPOCK3	0.11	1.47E-04	2.21	4691	—	—	—	—	—	—
CAPN8	0.11	1.54E-04	1.42	388	—	—	—	—	—	—
KBTBD11	0.11	1.65E-04	0.83	161	—	—	—	—	8:1920411	2.28E-01
CD55	0.11	1.74E-04	0.98	360	—	—	—	—	—	—
*HSPA8P18**	0.11	1.77E-04	-1.37	69	0.31	5.26E-04	-1.29	—	—	—
TIGD1	0.11	1.81E-04	-0.32	131	0.42	1.36E-03	-0.26	—	—	—
CDIPT	0.11	1.70E-04	0.33	101	0.36	9.20E-04	0.29	—	—	—
IGFN1	0.11	1.78E-04	-1.70	188	—	—	—	—	—	—
INHBA	0.11	1.94E-04	1.70	830	—	—	—	—	—	—
PID1	0.11	1.97E-04	0.80	87	0.36	7.83E-04	0.79	—	—	—
PLA2G4F	0.11	1.96E-04	1.70	28	0.21	1.60E-04	1.79	—	—	—
FAT3	0.11	2.10E-04	1.80	43	0.25	2.88E-04	1.77	—	11:91868024	1.46E-09
*RP3-416H24.1**	0.11	2.14E-04	1.33	592	—	—	—	—	—	—
*RP11-347C12.2**	0.11	2.15E-04	-0.86	12	0.16	5.15E-05	-0.91	—	16:30323810	5.97E-22
B3GNT3	0.11	2.16E-04	1.58	86	0.36	7.69E-04	1.22	—	—	—
TFCP2	0.12	2.28E-04	-0.21	70	0.33	5.72E-04	-0.19	—	—	—
*PABPC1P1**	0.12	2.36E-04	-0.49	75	0.33	6.17E-04	-0.47	—	—	—
TINAGL1	0.12	2.45E-04	0.49	88	0.36	8.12E-04	0.41	—	—	—

Supplementary Table S 5.5 Top 50 nominally differentially expressed (DE) genes between endometriosis cases (n=60) versus clinic controls (n=31) within secretory samples, with comparison to results from model without adjustment of sub-secretory phase as covariate, results from DE analysis of mid-secretory versus late-secretory samples, and lookup of endometrium *cis*-eQTL SNPs (No genes were reported in 3 literatures).

Key: DE = differentially expressed based on *p* value, adj p = adjusted *p* value in DE analysis, $\log_2$ FC = $\log_2$ fold change (positive value indicates higher expression in endometriosis cases compared to controls), Chr:pos = chromosome: position from GRCh37/hg19, FDR p = false discovery rate *p* value in eQTL analysis (eQTLs shown are all with *p* value <0.05); Gene marks with ^ are protein coding in gene biotype.

Endometriosis Case vs. Control in Secretory (With Sub-secretory Phases Adjustment) Top 50				**Look-up from Analysis without Sub-secretory Phases Adjustment DE genes ranked by adj p** (Top 150 DE genes with details)				**Rank of Gene in Mid vs. Late-secretory DE regardless of disease status by adj p**	**Endometrium *cis*-eQTL SNP (1Mb)**	
DE Gene	adj p	p	$\log_2$ FC	Rank	adj p	p	$\log_2$ FC		Chr:pos in hg19	FDR p
*SLC5A8^*	0.24	6.96E-06	2.06	2	0.15	7.05E-06	2.08	—	—	—
ANKRD20A11P	0.24	1.10E-05	1.61	9	1.00	3.12E-04	1.53	—	—	—
CYP4F29P	0.41	3.80E-05	2.03	617	—	—	—	—	—	—
*BPIFB1^*	0.41	3.36E-05	2.19	11	1.00	5.56E-04	1.74	—	—	—
*F2RL2^*	0.42	5.45E-05	0.75	7	1.00	1.66E-04	0.71	13425	5:75919633	4.60E-04
*HBG2^*	0.42	5.88E-05	-1.24	1	0.02	5.59E-07	-1.57	—	—	—
*VNN3^*	0.80	1.30E-04	-1.69	283	—	—	—	—	—	—
KB-1205A7.2	1.00	4.37E-04	-0.34	13	1.00	6.86E-04	-0.32	9649	—	—
RP11-599B13.7	1.00	5.78E-04	1.51	39871	—	—	—	—	—	—
TCEB1P2	1.00	7.21E-04	0.94	19	1.00	1.06E-03	0.90	—	—	—
AC007278.3	1.00	8.12E-04	-1.53	10	1.00	3.86E-04	-1.59	—	—	—
RN7SL418P	1.00	1.06E-03	-0.45	45	1.00	2.59E-03	-0.41	11403	—	—
RP3-466I7.1	1.00	1.15E-03	0.71	25	1.00	1.40E-03	0.69	—	—	—
IFNA11P	1.00	1.48E-03	-1.60	20	1.00	1.08E-03	-1.59	—	—	—
RP11-95P2.3	1.00	1.62E-03	-0.67	67	1.00	4.32E-03	-0.60	—	—	—
*LILRA6^*	1.00	1.67E-03	-0.85	50	1.00	3.30E-03	-0.81	—	—	—
RPL19P14	1.00	1.85E-03	-0.37	38	1.00	2.25E-03	-0.37	—	—	—
AC117947.1	1.00	1.90E-03	-1.61	46	1.00	2.77E-03	-1.48	—	—	—
RP11-755B10.3	1.00	2.10E-03	1.19	41	1.00	2.34E-03	1.16	—	—	—
RP11-6D1.3	1.00	2.16E-03	1.44	36	1.00	2.07E-03	1.43	—	—	—
RMST	1.00	2.18E-03	0.99	57	1.00	3.76E-03	0.94	—	—	—
RP11-369F10.2	1.00	2.26E-03	0.50	75	1.00	5.79E-03	0.45	—	—	—
MIR544B	1.00	2.34E-03	0.58	39872	—	—	—	—	—	—
RP11-20I23.8	1.00	2.34E-03	-0.64	39873	—	—	—	—	—	—
RP11-64D22.5	1.00	2.39E-03	1.02	60	1.00	3.89E-03	0.97	—	—	—
*NPIPB11^*	1.00	2.44E-03	0.42	176	—	—	—	15279	—	—
RP6-65G23.1	1.00	2.47E-03	1.19	34	1.00	2.01E-03	1.19	—	—	—
RN7SKP97	1.00	2.62E-03	-0.39	49	1.00	2.95E-03	-0.38	—	—	—
RP11-350D17.3	1.00	2.67E-03	1.21	39875	—	—	—	—	—	—
RP11-329A14.1	1.00	2.80E-03	-0.35	56	1.00	3.74E-03	-0.33	14358	—	—
TRAV30	1.00	2.88E-03	1.51	26	1.00	1.44E-03	1.65	—	—	—
RP11-417E7.1	1.00	2.94E-03	-1.05	159	—	—	—	—	—	—
RP4-539M6.21	1.00	3.09E-03	0.33	39874	—	—	—	3069	—	—
RP11-95H11.1	1.00	3.18E-03	-0.44	43	1.00	2.55E-03	-0.44	—	—	—
XXbac-B476C20.11	1.00	3.42E-03	0.34	47	1.00	2.81E-03	0.34	—	—	—
AC008268.2	1.00	3.54E-03	-0.88	69	1.00	4.72E-03	-0.83	—	—	—
RP4-760C5.3	1.00	3.55E-03	0.50	274	—	—	—	—	—	—
RP11-135A1.2	1.00	3.65E-03	1.30	37	1.00	2.11E-03	1.37	—	—	—
RP11-692N5.2	1.00	3.68E-03	-0.55	39878	—	—	—	—	—	—
MIR5687	1.00	3.71E-03	-0.40	66	1.00	4.22E-03	-0.39	11544	—	—
RN7SKP150	1.00	3.75E-03	0.49	171	—	—	—	—	—	—
RPS11P6	1.00	4.17E-03	-0.77	42	1.00	2.47E-03	-0.80	—	—	—
RP11-251G23.5	1.00	4.47E-03	0.23	39876	—	—	—	15972	—	—
RP11-111A22.1	1.00	4.53E-03	0.37	81	1.00	6.19E-03	0.36	—	—	—
SUMO2P17	1.00	4.77E-03	0.73	85	1.00	6.47E-03	0.69	—	—	—
*RP11-701P16.2^*	1.00	4.86E-03	-0.57	48	1.00	2.90E-03	-0.60	—	—	—
RP11-299H22.6	1.00	5.17E-03	-1.27	134	1.00	9.57E-03	-1.19	—	—	—
AC073109.2	1.00	5.20E-03	-0.69	203	—	—	—	4314	7:11564331	1.01E-07
RP11-20I23.13	1.00	5.21E-03	-0.68	40	1.00	2.28E-03	-0.74	—	—	—
RP11-100L22.3	1.00	5.33E-03	1.55	109	1.00	8.33E-03	1.46	—	—	—

Supplementary Table S 5.6 Top 50 nominally differentially expressed (DE) genes between ASRM stage I/II (n=34) versus clinic controls (n=31) within secretory samples, with comparison to results from DE analysis of mid-secretory versus late-secretory samples, previous reports from the literature, and lookup of endometrium *cis*-eQTL SNPs.

Key: DE = differentially expressed based on *p* value, adj p = adjusted *p* value in DE analysis, log_2 FC = log_2 fold change (positive value indicates higher expression in ASRM stage I/II cases compared to controls), Chr:pos = chromosome: position from GRCh37/hg19, FDR p = false discovery rate *p* value in eQTL analysis (eQTLs shown are all with *p* value <0.05); Gene marks with ^ are protein coding in gene biotype.

Endometriosis Stage I/II vs. Control in Secretory (With Sub-secretory Phases Adjustment) Top 50				**Mid vs. Late-secretory DE regardless of disease status DE genes ranked by adj p-value** (Top 150 DE genes with details)				**Reported in Previous Publications**	**Endometrium *cis*-eQTL SNP (1Mb)**	
DE Gene	adj p	p	Log_2 FC	Rank	adj p	p	Log_2 FC		Chr:pos in hg19	FDR p
*ALAS2^*	0.03	6.15E-07	-2.14	—	—	—	—	Fung et al. 2018 (On/Off)	—	—
*VNN3^*	0.07	3.27E-06	-2.29	—	—	—	—	—	—	—
KB-1205A7.2	0.40	3.72E-05	-0.44	9649	—	—	—	—	—	—
*MGAM^*	0.40	3.04E-05	-1.28	—	—	—	—	—	—	—
UBE2Q2P3	0.81	1.12E-04	1.16	—	—	—	—	—	—	—
*COL8A1^*	0.81	1.17E-04	-1.37	4	1.50E-04	3.3E-08	1.99	—	—	—
*F2RL2^*	0.81	1.31E-04	0.80	13425	—	—	—	—	5:75919633	4.60E-04
*HIST2H3A^*	0.87	1.61E-04	2.02	—	—	—	—	—	—	—
RP11-299H22.6	1.00	4.61E-04	-1.82	—	—	—	—	—	—	—
RP11-554D15.3	1.00	1.80E-03	-1.98	—	—	—	—	—	—	—
RP11-45P15.4	1.00	2.06E-03	0.50	—	—	—	—	—	14:103812320	0.26
RP11-255H23.3	1.00	2.07E-03	0.47	11923	—	—	—	—	—	—
RP11-734E19.1	1.00	2.09E-03	1.25	—	—	—	—	—	—	—
RP4-539M6.21	1.00	2.34E-03	0.38	3069	—	—	—	—	—	—
RP1-261D10.2	1.00	2.35E-03	0.42	—	—	—	—	—	—	—
RP11-486O13.4	1.00	2.61E-03	-0.57	—	—	—	—	—	—	—
CTD-3247F14.2	1.00	2.99E-03	1.78	—	—	—	—	—	—	—
RP11-678G14.3	1.00	3.24E-03	-0.89	—	—	—	—	—	—	—
RP11-599B13.7	1.00	3.56E-03	1.40	—	—	—	—	—	—	—
MIR544B	1.00	4.55E-03	0.60	—	—	—	—	—	—	—
RP3-405J10.3	1.00	4.95E-03	-0.44	9616	—	—	—	—	—	—
RP11-20I23.8	1.00	5.06E-03	-0.67	—	—	—	—	—	—	—
RP11-95P2.3	1.00	5.32E-03	-0.68	—	—	—	—	—	—	—
RP11-781P6.1	1.00	5.46E-03	1.42	—	—	—	—	—	—	—
*SPIB^*	1.00	6.16E-03	-0.87	—	—	—	—	—	—	—
RP11-692N5.2	1.00	6.32E-03	-0.59	—	—	—	—	—	—	—
RP11-408A13.4	1.00	6.36E-03	0.27	11474	—	—	—	—	9:14578127	0.48
KB-1836B5.4	1.00	6.52E-03	1.04	—	—	—	—	—	—	—
MIR3922	1.00	6.88E-03	-1.17	—	—	—	—	—	—	—
KRT8P34	1.00	7.13E-03	0.70	—	—	—	—	—	—	—
AC003102.3	1.00	7.37E-03	0.58	—	—	—	—	—	—	—
RP1-97G4.1	1.00	7.54E-03	-1.02	—	—	—	—	—	—	—
*GRIN2B^*	1.00	7.91E-03	-0.97	—	—	—	—	—	—	—
RN7SL602P	1.00	8.00E-03	-1.18	—	—	—	—	—	—	—
AC005592.3	1.00	8.21E-03	1.31	—	—	—	—	—	—	—
RP11-615I2.3	1.00	8.28E-03	-1.23	—	—	—	—	—	—	—
ZNF818P	1.00	8.34E-03	-0.29	9501	—	—	—	—	19:53698273	7.10E-02
RP11-95H11.1	1.00	8.37E-03	-0.45	—	—	—	—	—	—	—
KB-1440D3.13	1.00	8.49E-03	0.70	—	—	—	—	—	—	—
RP11-63G10.3	1.00	8.64E-03	-0.57	—	—	—	—	—	—	—
TPST2P1	1.00	9.49E-03	0.45	—	—	—	—	—	—	—
*RP11-297M9.1^*	1.00	9.55E-03	1.47	—	—	—	—	—	—	—
RP11-632K20.2	1.00	9.77E-03	-0.89	—	—	—	—	—	15:32810421	1.57E-08
RP11-111A22.1	1.00	1.03E-02	0.37	—	—	—	—	—	—	—
RP11-465I4.3	1.00	1.06E-02	-0.42	—	—	—	—	—	—	—
RP11-697E2.10	1.00	1.09E-02	0.79	—	—	—	—	—	—	—
LINC00519	1.00	1.20E-02	-1.20	—	—	—	—	—	—	—
CTD-3025N20.3	1.00	1.20E-02	0.86	—	—	—	—	—	—	—
RP3-395M20.12	1.00	1.24E-02	-0.63	—	—	—	—	—	—	—
RP11-20I23.13	1.00	1.25E-02	-0.68	—	—	—	—	—	—	—

Supplementary Table S 5.7 Top 50 nominally differentially expressed (DE) genes between ASRM stage III/IV (n=26) versus clinic controls (n=31) within secretory samples, with comparison to results from DE analysis of mid-secretory versus late-secretory samples, and previous reports from the literature (No genes were overlapped with endometrium *cis*-eQTL SNPs).

Key: DE = differentially expressed based on *p* value, adj p = adjusted *p* value in DE analysis, log_2 FC = log_2 fold change (positive value indicates higher expression in ASRM stage III/IV cases compared to controls); Gene marks with * are non-protein coding in gene biotype (i.e. pseudogene, lincRNA, processed transcript, etc); On/Off = mentioned in on/off gene prediction analysis, ESE = early-secretory, MSE = mid-secretory.

Endometriosis Stage III/IV vs. Control in Secretory (With Sub-secretory Phases Adjustment) Top 50				**Mid vs. Late-secretory DE regardless of disease status DE genes ranked by adj p-value** (Top 150 DE genes with details)				**Reported in Previous Publications**
DE Gene	adj p	p	Log_2 FC	Rank	adj p	p	Log_2 FC	
PRELP	1.33E-06	4.91E-11	2.27	11	1.02E-02	6.22E-06	2.14	—
BPIFB1	3.09E-05	2.28E-09	3.38	—	—	—	—	—
DES	4.25E-05	4.70E-09	3.15	1	1.77E-05	9.79E-10	4.03	—
PLIN4	9.28E-05	1.37E-08	1.90	463	—	—	—	—
GPD1	4.79E-04	8.81E-08	2.25	—	—	—	—	—
SLC5A8	1.34E-03	3.22E-07	2.68	—	—	—	—	—
TFF3	1.34E-03	3.98E-07	2.19	—	—	—	—	—
PLIN1	1.34E-03	3.98E-07	1.98	—	—	—	—	—
SLC24A2	1.34E-03	4.72E-07	2.39	—	—	—	—	Fung et al. 2018 (On/Off)
JPH2	1.34E-03	4.92E-07	2.34	—	—	—	—	—
SAA1	3.40E-03	1.39E-06	2.32	—	—	—	—	—
MUC5AC	3.40E-03	1.50E-06	2.63	—	—	—	—	Fung et al. 2018 (On/Off)
*ANAPC1P1**	3.65E-03	1.75E-06	2.16	—	—	—	—	—
CHRDL1	4.46E-03	2.30E-06	1.96	7	5.24E-03	2.03E-06	2.63	—
MUC5B	7.13E-03	3.94E-06	2.26	10615	—	—	—	—
MYH11	1.07E-02	6.36E-06	1.71	2	1.10E-04	1.22E-08	2.54	Burney et al. 2007 (within ESE)
PLG	1.07E-02	7.04E-06	-1.35	—	—	—	—	—
LTF	1.07E-02	7.09E-06	1.96	16381	—	—	—	—
SCGB3A1	1.07E-02	7.63E-06	2.74	—	—	—	—	—
LEP	1.07E-02	8.17E-06	3.37	—	—	—	—	—
VSTM2L	1.07E-02	8.30E-06	1.51	—	—	—	—	—
SIX1	1.33E-02	1.07E-05	2.06	—	—	—	—	—
UNC5D	1.38E-02	1.17E-05	1.91	—	—	—	—	—
*RMST**	1.94E-02	1.72E-05	1.57	—	—	—	—	—
FABP4	1.99E-02	1.88E-05	2.75	—	—	—	—	—
KCND2	1.99E-02	1.90E-05	-1.92	6537	—	—	—	—
IGSF9B	2.68E-02	2.66E-05	1.37	—	—	—	—	—
DMD	4.55E-02	4.69E-05	1.20	1197	—	—	—	Burney et al. 2007 (within ESE)
TCF23	4.69E-02	5.01E-05	1.98	—	—	—	—	—
THRSP	4.87E-02	5.38E-05	0.99	—	—	—	—	—
*TRAV30**	4.87E-02	5.79E-05	2.27	—	—	—	—	—
CNR1	4.87E-02	5.80E-05	1.66	—	—	—	—	—
AOC3	4.87E-02	5.93E-05	1.17	118	0.20	1.32E-03	1.11	Burney et al. 2007 (within MSE)
FHL1	5.41E-02	6.77E-05	0.79	498	—	—	—	—
LPL	6.88E-02	8.87E-05	1.11	17365	—	—	—	Fung et al. 2018 (DE)
*KB-1269D1.8**	6.97E-02	9.24E-05	1.28	—	—	—	—	—
SYNM	6.98E-02	9.51E-05	0.90	22	5.22E-02	6.37E-05	1.23	Mortlock et al. 2020 (within MSE)
*IGHA1**	7.05E-02	9.87E-05	1.74	—	—	—	—	—
TUSC5	7.21E-02	1.04E-04	2.76	—	—	—	—	—
RAB3C	7.55E-02	1.11E-04	-1.86	—	—	—	—	—
ATP6V1B1	8.08E-02	1.22E-04	1.28	—	—	—	—	—
*LINC00906**	8.23E-02	1.27E-04	2.40	—	—	—	—	—
CXCL1	9.02E-02	1.45E-04	1.63	—	—	—	—	—
SMOC2	9.02E-02	1.46E-04	0.79	8943	—	—	—	—
KCNMA1	9.29E-02	1.54E-04	0.87	1284	—	—	—	Burney et al. 2007 (within ESE)
REEP1	0.11	1.86E-04	0.80	990	—	—	—	—
CIDEC	0.11	1.90E-04	3.34	—	—	—	—	—
ADH1B	0.11	1.90E-04	1.50	3763	—	—	—	—
IL19	0.11	1.94E-04	1.55	—	—	—	—	Fung et al. 2018 (On/Off)
NEXN	0.11	2.01E-04	0.67	64	0.18	6.78E-04	0.76	—

Supplementary Table S 5.8 Top 50 differentially expressed (DE) genes between ASRM stage III/IV (n=26) versus stage I/II (n=34) within secretory samples, with comparison to results from DE analysis of mid-secretory versus late-secretory samples, previous reports from the literature, and lookup of endometrium *cis*-eQTL SNPs.

Key: DE = differentially expressed based on adj.*p* value, log_2 FC = log_2 fold change (positive value indicates higher expression in ASRM stage III/IV cases compared to I/II), FDR p = false discovery rate *p* value in eQTL analysis (eQTLs shown are all with *p* value <0.05); Gene marks with * are non-protein coding in gene biotype; On/Off = on/off gene prediction, ESE = early-secretory, MSE = mid-secretory.

Endometriosis Stage III/IV vs. I/II in Secretory (With Sub-secretory Phases Adjustment) Top 50				**Mid vs. Late-secretory DE regardless of disease status DE genes ranked by adj p-value** (Top 150 DE genes with details)				**Reported in Previous Publications**	**Endometrium *cis*-eQTL SNP (1Mb)**	
DE Gene	adj p	p	Log_2 FC	Rank	adj p	p	Log_2 FC		Chr:pos in hg19	FDR p
SAA1	1.44E-08	7.06E-13	3.42	—	—	—	—	—	—	—
PLIN1	1.44E-08	8.10E-13	2.75	—	—	—	—	—	—	—
PRELP	2.07E-08	2.32E-12	2.37	11	1.02E-02	6.22E-06	2.14	—	—	—
DES	2.07E-08	2.33E-12	3.70	1	1.77E-05	9.79E-10	4.03	—	—	—
PLIN4	1.22E-07	1.73E-11	2.21	463	—	—	—	—	—	—
MYH11	4.41E-07	7.46E-11	2.42	2	1.10E-04	1.22E-08	2.54	Burney et al. 2007 (within ESE)	—	—
GPD1	7.01E-07	1.38E-10	2.65	—	—	—	—	—	—	—
BPIFB1	2.03E-06	4.81E-10	3.45	—	—	—	—	—	—	—
MUC5AC	2.03E-06	5.16E-10	3.34	—	—	—	—	Fung et al. 2018 (On/Off)	—	—
ADH1B	5.26E-06	1.48E-09	2.39	3763	—	—	—	—	—	—
MUC5B	4.18E-05	1.30E-08	2.73	10615	—	—	—	—	—	—
FABP4	7.95E-05	2.69E-08	3.50	—	—	—	—	—	—	—
JPH2	9.38E-05	3.44E-08	2.51	—	—	—	—	—	—	—
PGM5	1.34E-04	5.29E-08	1.51	14	3.29E-02	2.55E-05	1.58	—	—	—
LEP	1.60E-04	6.77E-08	4.02	—	—	—	—	—	—	—
CNN1	1.90E-04	9.07E-08	2.36	15	3.68E-02	3.06E-05	2.28	Fung et al. 2018 (On/Off)	—	—
*ANKRD20A11P**	1.90E-04	9.11E-08	-2.54	—	—	—	—	—	—	—
CXCL3	3.07E-04	1.56E-07	2.48	—	—	—	—	—	—	—
SGCA	4.05E-04	2.17E-07	2.71	—	—	—	—	—	—	—
PTGIS	4.10E-04	2.33E-07	2.15	5	1.82E-03	5.31E-07	2.45	Burney et al. 2007 (within ESE)	20:48660144	0.43
TFF3	4.10E-04	2.43E-07	2.18	—	—	—	—	—	—	—
SAA2	4.65E-04	2.94E-07	2.18	—	—	—	—	Burney et al. 2007 (within MSE)	—	—
LTF	4.65E-04	3.01E-07	2.19	16381	—	—	—	—	—	—
CXCL1	7.09E-04	4.80E-07	2.13	—	—	—	—	—	—	—
SLC24A2	7.80E-04	5.49E-07	2.32	—	—	—	—	Fung et al. 2018 (On/Off)	—	—
*IGHG1**	8.77E-04	6.43E-07	1.95	—	—	—	—	—	—	—
IL19	1.04E-03	7.92E-07	2.03	—	—	—	—	Fung et al. 2018 (On/Off)	—	—
*CYP4F29P**	1.45E-03	1.14E-06	-2.76	—	—	—	—	—	—	—
*PGM5-AS1**	1.59E-03	1.30E-06	2.16	—	—	—	—	—	—	—
IL8	1.61E-03	1.36E-06	2.13	—	—	—	—	—	—	—
*ANAPC1P1**	3.06E-03	2.74E-06	2.07	—	—	—	—	—	—	—
SCG2	3.06E-03	2.76E-06	1.67	—	—	—	—	—	—	—
VSTM2L	4.04E-03	3.78E-06	1.53	—	—	—	—	—	—	—
CHRDL1	4.04E-03	3.87E-06	1.88	7	5.24E-03	2.03E-06	2.63	—	—	—
DGAT2	4.74E-03	4.80E-06	1.41	—	—	—	—	—	—	—
*CTD-2007H13.3**	4.74E-03	4.89E-06	0.66	—	—	—	—	—	—	—
FHL1	4.74E-03	4.95E-06	0.88	498	—	—	—	—	—	—
PI16	4.82E-03	5.28E-06	1.96	—	—	—	—	—	—	—
GJC2	4.82E-03	5.30E-06	0.78	11182	—	—	—	—	—	—
*HLA-U**	5.55E-03	6.30E-06	1.70	—	—	—	—	—	—	—
SLC19A3	5.55E-03	6.42E-06	2.12	—	—	—	—	—	—	—
PDLIM3	5.71E-03	6.76E-06	1.75	3	1.24E-04	2.06E-08	3.41	Fung et al. 2018 (On/Off)	—	—
CXCL5	6.89E-03	8.35E-06	2.79	—	—	—	—	Fung et al. 2018 (On/Off)	—	—
TUSC5	7.42E-03	9.20E-06	3.10	—	—	—	—	—	—	—
RIMBP2	7.43E-03	9.55E-06	1.14	9168	—	—	—	—	—	—
CD300LG	7.43E-03	9.63E-06	2.51	—	—	—	—	—	—	—
FCER1A	7.60E-03	1.01E-05	1.19	—	—	—	—	—	—	—
SYNM	8.05E-03	1.09E-05	0.99	22	5.22E-02	6.37E-05	1.23	Mortlock et al. 2020 (within MSE)	—	—
IGSF9B	8.05E-03	1.11E-05	1.40	—	—	—	—	—	—	—
*IGLV3-21**	8.44E-03	1.19E-05	3.08	—	—	—	—	—	—	—

Supplementary Table S 6.1 Table of technical and biological variables in significant association with the first three principal components (PCs) of DNA methylation data after quality control.

Technical/Biological Variable	Variable Group	PC	t-test *p*	Estimate	Lower 95% CI	Upper 95% CI
Cycle Phases	Proliferative	PC1	9.67E-03	-5.34	-9.89	-0.79
	Late-secretory		5.50E-04	7.33	2.69	11.97
	Menstrual	PC3	1.36E-03	-4.43	-7.48	-1.37
Well Positions	C01	PC1	8.18E-03	15.86	2.51	29.22
	D01	PC3	1.42E-05	15.29	7.62	22.97
	F10		5.48E-04	-12.02	-19.70	-4.34
	E06		2.07E-03	-10.67	-18.35	-2.99
Amount of DNA (ng)	550	PC1	4.59E-03	-14.75	-26.31	-3.19
	1026	PC3	1.88E-03	-10.77	-18.45	-3.09
Plate Number	1477	PC2	1.25E-03	3.71	1.18	6.24
	1549		2.74E-03	-4.83	-8.40	-1.26

Supplementary Table S 6.2 Number of eutopic endometrium samples in each of five cycle phases in EWAS comparisons of endometriosis, ASRM surgical stage I/II, and stage III/IV cases to clinic controls.

	Menstrual	Proliferative	Secretory		
			Early	Mid	Late
Endometriosis cases (n=100) vs. clinic controls (n=38)					
Cases	6	22	11	37	21
Controls	0	9	6	17	6
ASRM Stage I/II cases (n=68) vs. clinic controls (n=38)					
Cases	4	17	8	25	14
Controls	0	9	6	17	6
ASRM Stage III/IV cases (n=32) vs. clinic controls (n=38)					
Cases	2	5	3	12	7
Controls	0	9	6	17	6

F. Permission of use for copy-righted figures

F.i Chapter 1

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Sep 13, 2020

This Agreement between Miss. Sze Kan Cecilia Cheuk ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4906910884794
License date	Sep 13, 2020
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Immunology
Licensed Content Title	The role of sex hormones in immune protection of the female reproductive tract
Licensed Content Author	Charles R. Wira et al
Licensed Content Date	Mar 6, 2015
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	2
High-res required	no

Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Elucidating the Heterogeneity of Endometriosis using Multi-omics
Institution name	University of Oxford
Expected presentation date	Oct 2020
Portions	Figure 2
Requestor Location	Miss. Sze Kan Cecilia Cheuk Linacre College Oxford, OX1 3JA United Kingdom Attn: Miss. Sze Kan Cecilia Cheuk
Total	0.00 GBP

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to

license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. ***BOOKS ONLY:*** Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.2

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

No text material on this page.

This page is the last page of the Springer Nature License.

Cecilia Cheuk

From: noreply@salesforce.com on behalf of customercare@copyright.com
Sent: 16 September 2020 01:36
To: Cecilia Cheuk
Subject: Case #01170965 - Request for permission of re-use of figure for a PhD Thesis [ref:_00D30oeGz._5004Q2235GD:ref]

Follow Up Flag: Follow up
Flag Status: Flagged

Dear Cecilia,

Thank you for reaching out to Copyright Clearance Center's RightsLink service, where we act on behalf of publishers in providing permissions to our customers. The availability of permissions can vary depending on the publisher and its type. My name is Lee and it will be my pleasure to assist you.

Based on your request, I would like to note that the intended use selected on license number **4906910884794** is correct. As this license is proof that permission has been granted to reuse this figure, you may attach a copy of this license to your thesis if you wish. I would like to recommend consulting the terms and conditions included with this license as this will provide you with guidelines on how to cite the figure.

I hope the above is in order and wish you all the best with your upcoming thesis.

Kind Regards,
Lee

Lee Ferns
Customer Account Specialist
Copyright Clearance Center
222 Rosewood Drive
Danvers, MA 01923
www.copyright.com
Toll Free US +1.855.239.3415
International +1.978-646-2600
Facebook - Twitter - LinkedIn

ref:_00D30oeGz._5004Q2235GD:ref

---------- Original Message ----------
From: Cecilia Cheuk [cecilia.cheuk@linacre.ox.ac.uk]
Sent: 9/13/2020 2:38 AM
To: customercare@copyright.com
Subject: Request for permission of re-use of figure for a PhD Thesis

Dear team,

I would like to apply for a permission to use a figure https://www.nature.com/articles/nri3819/figures/2 in a PhD thesis.

I have gone through the online order/application steps, please see the attached pdf. (License Number: 4906910884794). However, I am wondering if that is the correct class of permission that fits for the purpose.

Would there be a permission letter that I can use for my PhD thesis submission, stating I have been granted permission by the publisher to use copyrighted figures (both print and digital) from this paper for my PhD thesis, and state that the nature of this license (e.g. CC-BY 4.0) for purposes of PhD thesis?

Thank you very much in looking into this.

Best wishes,
Cecilia

This message (including attachments) is confidential, unless marked otherwise. It is intended for the addressee(s) only. If you are not an intended recipient, please delete it without further distribution and reply to the sender that you have received the message in error.

ref:_00D30oeGz._5004Q2235GD:ref

F.ii Chapter 4

QIAGEN Bioinformatics Technical Support May 19, 2020

To whom this may concern:

Cecilia Cheuk, Wellcome Centre for Human Genetics, University of Oxford has been granted permission by QIAGEN Silicon Valley to use copyrighted figures (both print and digital) generated from QIAGEN Ingenuity Pathway Analysis (QIAGEN IPA) for her PhD thesis.

Figures produced from IPA are available under an open-access CC-BY 4.0 license for purposes of publication.

If you have any questions, please contact QIAGEN Bioinformatics Technical Support at ts-bioinformatics@qiagen.com

Best regards,

(on behalf of QIAGEN Silicon Valley)

Gaurav Kadu

Gaurav Kadu

Senior Scientific Bioinformatics Application Scientist, Biox Scientific Support India

QIAGEN - Sample to Insight

Phone American: +1-866-464-3684 | Danish: +45-80 82 0167
ts-bioinformatics@qiagen.com
www.digitalinsights.qiagen.com

Making improvements in life possible

F.iii Chapter 5

Cecilia Cheuk

From: Machiela, Mitchell (NIH/NCI) [E] <mitchell.machiela@nih.gov>
Sent: 02 October 2020 21:44
To: Cecilia Cheuk; NCILDlinkWebAdmin
Subject: Re: Seek Permission to Use Output from LDmatrix Tool for PhD Thesis

Hi Cecilia,
Here is a link to the LDlink license: https://dceg.cancer.gov/tools/analysis/ldlink/license
Mitch

From: Cecilia Cheuk <cecilia.cheuk@linacre.ox.ac.uk>
Date: Friday, October 2, 2020 at 5:33 AM
To: "Machiela, Mitchell (NIH/NCI) [E]" <mitchell.machiela@nih.gov>, NCILDlinkWebAdmin <NCILDlinkWebAdmin@mail.nih.gov>
Subject: RE: Seek Permission to Use Output from LDmatrix Tool for PhD Thesis

Dear Mitch and the team,

Thank you for the permission.

I should like to include this work within the electronic version of my thesis, which I am required to deposit in the Oxford University Research Archive (ORA) (http://ora.ox.ac.uk/). ORA is a repository for the research publications and other research output produced by members of the University of Oxford. It is non-commercial and openly available to all.

I am wondering which kind of copyright this online-generated LD matrix image is, and the type of license I may need to include for both the digital (as above) and 2 physical copies of the thesis to be available to the university archive?

Thanks again.

With best wishes,
Cecilia

From: Machiela, Mitchell (NIH/NCI) [E] <mitchell.machiela@nih.gov>
Sent: 02 October 2020 01:40
To: Cecilia Cheuk <cecilia.cheuk@linacre.ox.ac.uk>; NCILDlinkWebAdmin <NCILDlinkWebAdmin@mail.nih.gov>
Subject: Re: Seek Permission to Use Output from LDmatrix Tool for PhD Thesis

Hi Cecilia,
You are more than welcome to use LDlink output plots for your PhD thesis.
Glad you have found this tool useful.
Best wishes as you defend your thesis.
Mitch

From: Cecilia Cheuk <cecilia.cheuk@linacre.ox.ac.uk>
Date: Thursday, October 1, 2020 at 12:46 PM
To: NCILDlinkWebAdmin <NCILDlinkWebAdmin@mail.nih.gov>
Subject: Seek Permission to Use Output from LDmatrix Tool for PhD Thesis

Dear admin team,

1

I am seeking permission to use an output figure (attached) from your LD matrix online tool for my PhD thesis. If any other information is needed, please do not hesitate to get in touch.

Thank you.

With best wishes,
Cecilia

--
Cecilia Cheuk
D.Phil Candidate in Obstetrics and Gynaecology

Endometriosis CaRe Centre Oxford
Nuffield Department of Women's and Reproductive Health

Wellcome Centre for Human Genetics
University of Oxford

John Radcliffe Hospital - Women's Centre
Oxford OX3 9DU
United Kingdom

https://dceg.cancer.gov/tools/analysis/ldlink/license; accessed on 02 October 2020

NIH NATIONAL CANCER INSTITUTE
Division of Cancer Epidemiology & Genetics

LDlink License Agreement

LDlink is available under the MIT license, an Open Source Initiative approved license.

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

This permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

open source initiative
Approved License®

F.iv Chapter 6

Cecilia Cheuk

From: Manfred Grabherr <manfred.grabherr@imbim.uu.se>
Sent: 07 October 2020 15:32
To: Cecilia Cheuk
Subject: Re: Seeking Permission to Use Output from PiiL Tool for PhD Thesis

Hi Cecilia,

No worries, you can use the images in any way you want. Since it is your data, the copyright is yours. Thanks for using PiiL! :)

Cheers,

- Manfred

On 2020-10-07 07:49, Cecilia Cheuk wrote:

> Dear PiiL team,
>
> I am seeking permission to use an output figure (attached) from your PiiL tool for my PhD thesis.
> I really appreciate the function of your software, and I hope the graphic from it can be used to help displaying my results for the thesis.
>
> I should like to include this work within the electronic version of my thesis, which I am required to deposit in the Oxford University Research Archive (ORA) (http://ora.ox.ac.uk/). ORA is a repository for the research publications and other research output produced by members of the University of Oxford. It is non-commercial and openly available to all.
>
> I am wondering which kind of copyright this PiiL-generated image is, and the type of license I may need to include for both the digital (as above) and 2 physical copies of the thesis to be available to the university archive?
>
> Thank you very much
>
> With best wishes,
> Cecilia
>
> --
> Cecilia Cheuk
> D.Phil Candidate in Obstetrics and Gynaecology
>
> Endometriosis CaRe Centre Oxford
> Nuffield Department of Women's and Reproductive Health
>
> Wellcome Centre for Human Genetics
> University of Oxford
>
> John Radcliffe Hospital - Women's Centre
> Oxford OX3 9DU
> United Kingdom

1

G. Abbreviations

Abbreviations

5-hmC: Cytosine Hydroxymethylation 26
5-mC: Cytosine Methylation 26
95% CI: 95% Confidence Interval 53
A: Adenine 26
adj.*p:* Adjusted *p*-value 106, 149
AFS: American Fertility Society 7
AIC: Akaike Information Criterion 107
ASRM: American Society for Reproductive Medicine 5, 38
AUC: Area Under Curve 123
BBID: Biological Biochemical Image Database 108
BCE: Boston Centre for Endometriosis 35
BDNF: Brain-derived Neurotrophic Factor 98
C: Cytosine 25
CA: Cancer Antigen 3
cAMP: Cyclic Adenosine Monophosphate 24
CCKR: Cholecystokinin Receptor 28
CDCV: Common Disease-Common Variant Hypothesis 19
CHI3L1: Chitinase-3-like Protein 1 173
CI: Confidence Interval 121
CIN: Cervical Intra-epithelial Neoplasia 47
CNS: Central Nervous System 10
CpG: 5'—C—phosphate—G—3' 25
CPM: Counts Per Million 105, 150
CTD: Comparative Toxicogenomics Database 174
DE: Differential Expression or Differentially Expressed 101, 145, 203
DHEAS: Dehydroepiandrosterone Sulphate 98
DMRs: Differentially Methylated Regions 16
EFI: Endometriosis Fertility Index 9
EGF: Epidermal Growth Factor 170
EGFR: Epidermal Growth Factor Receptor 24
ENCODE: Encyclopedia of DNA Elements 26
ENZIAN-Score: Enzian Classification for Deep Endometriosis 9
EPHect: Endometriosis Phenome and Biobanking Harmonisation Project 33
EPIC: Illumina® Methylation EPIC Beadchips Array 183
eQTL SNPs: expression Quantitative Trait Locus Single Nucleotide Polymorphism 134
eQTL: Expression Quantitative Trait Loci 24, 146, 182
ER: Oestrogen Receptor 87, 208
ERα: Oestrogen Receptor Alpha 206
eSNPs: Expression Quantitative Trait Locus Single Nucleotide Polymorphism 98, 146, 187
ESR2: Oestrogen Receptor 2 174
EWAS: Epigenome-wide Association Study 29, 35, 182
EXPPECT: Excellence in Pelvic Pain and Endometriosis Care and Treatment 44

FAC1: Fetal Alz-50 Clone 1 98
FANTOM5: Functional Annotation of the Mammalian Genome 26
FDR: False Discovery Rate 106, 148
FSH: Follicle-stimulating Hormone 4, 92
G: Guanine 25
GLM: Generalised log-Linear Model 105, 148
GnRH: Gonadotropin-releasing Hormone 1
GO: Gene Ontology 15
GSEA: Gene Set Enrichment Analysis 197
GTEx: Genotype-Tissue Expression Project 98
GWAS: Genome-wide Association Study 20, 34, 39
HMB: Heavy Menstrual Bleeding 93, 170
HPG: Hypothalamic-pituitary-gonadal 4
HRC: Haplotype Reference Consortium 36, 45
HSD17B2: 17β-hydroxysteroid Dehydrogenase Type 2 17
HWE: Hardy-Weinberg Equilibrium 45
IBS: Irritable Bowel Syndrome 5, 56
ICD-10-CM: International Classification of Diseases, 10th revision, Clinical Modification . 45
IEGC: International Endometriosis Genomics Consortium 20, 32, 39
IID: Integrated Interactions Database 135, 205
IL-2: Interleukin-2 173
IPA: Ingenuity Pathway Analysis 16
iPSC: Induced Pluripotent Stem Cell 92
ISVA: Independent Surrogate Variables Analysis 186
KEGG: Kyoto Encyclopedia of Genes and Genomes 16, 108
LD: Linkage Disequilibrium 90, 173
LH: Luteinizing Hormone 4, 93
lincRNA: Long Intergenic Non-coding RNA 158
LKB1: Liver Kinase B1 94
LMP: Last Menstrual Period 103
lncRNA: Long Non-coding RNA 22, 97, 133, 208
Log_2 |FC|: log2 of Absolute Fold-change 155
log_2FC: log_2 Fold-Change 106
LRT: Likelihood Ratio Test 114, 149
MAF: Minor Allele Frequency 19, 40
Mb: Megabase 150
MDS: Multi-dimensional Scaling 148
MHC-II: Major Histocompatibility Class II Molecule 93
miRNA: micro RNA 22
mQTL: methylation QTL 210
MRC: Medical Research Council, UK 102
MRI: Magnetic Resonance Imaging 1
mRNA: Messenger RNA 22, 208
MSD: Medical Science Division 35
mTOR: Mechanistic Target Of Rapamycin Kinase 94
MYO1E: Actin-based Motor Protein Myosin 1E 93

NGF: Nerve Growth Factor 98
NICE: National Institute for Health and Care Excellence, the United Kingdom 3
NIH: National Institute of Health 32, 182
NRS: Numeric Rating Scale 47
OGC: Oxford Genomics Centre 102
OR: Odds Ratio 41
PC: Principal Component 111, 148, 185
PCA: Principal Component Analysis 15, 105, 149, 184
PCOS: Polycystic Ovarian Syndrome 93
PCR: Polymerase Chain Reaction 25
PI3K-Akt: Phosphatidylinositol 3-kinase-Protein kinase B 16
PRA: Progesterone Receptors A 18
PRB: Progesterone Receptors B 18
QC: Quality Control 16, 36, 147, 182
Q-Q: Quantile-Quantile 107
r^2: Pearson's correlation coefficient 53
RA: Risk Allele 68
rAFS/ASRM/rASRM: Revised American Fertility Society/American Society for Reproductive Medicine Endometriosis Staging 7
REDCap: Research Electronic Data Capture 33
RIN: RNA Integrity Number 102
RNA-seq: RNA-sequencing 22, 101, 145
ROS: Reactive Oxygen Species 137
rRNA: ribosomal RNA 22
SDs: Standard Deviations 185
SF-1: Steroidogenic Factor 1 17
SNPs: Single Nucleotide Polymorphisms 19, 39
SOPs: Standard Operating Protocols 7, 33
SQL: Structured Query Language 33
SSF: Standard Surgical Form v3.0, Dec 2014 46
SVA: Surrogate Variables Analysis 186
T: Thymines 26
T2D: Type 2 Diabetes Mellitus 39
TMM: Trimmed Mean of M-values 105
TNF: Tumour Necrosis Factor 10
TNFα: Tumor Necrosis Factor Alpha 173
TPM: Transcripts Per Million 104
TSS: Transcription Start Site 150, 203
U: Uracils 26
UCSF: University of California, San Francisco 35
UKBB: UK BioBank 45
UTR: Untranslated Region 88, 166
VCAN: Extracellular Matrix Proteoglycan Versican 24
VEGF: Vascular Endothelial Growth Factor 89
VST: Variance Stabilising Transformed 149
WCHG: The Wellcome Centre for Human Genetics 102

WERF: World Endometriosis Research Foundation 33
WERF–EPHect: World Endometriosis Research Foundation – Endometriosis Phenome and Biobanking Harmonisation Project 7

Gene Name Abbreviations

ABCG2: ATP Binding Cassette Subfamily G Member 2 195
ABO: ABO Blood Group, Alpha 1-3-N-Acetylgalactosaminyltransferase And Alpha 1-3-Galactosyltransferase 199
ABO: Alpha 1-3-N-Acetylgalactosaminyltransferase and Alpha 1-3-Galactosyltransferase .. 74
ACTL9: Actin Like 9 75
ADAMTS13: ADAM Metallopeptidase With Thrombospondin Type 1 Motif 13 199
AHRR: Aryl-Hydrocarbon Receptor Repressor 29
AKR1B1: Aldo-Keto Reductase Family 1 Member B 29
ALAS2: 5'-Aminolevulinate Synthase 2 158
ANKRD65: Ankyrin Repeat Domain 65 154
APEH: Acylaminoacyl-Peptide Hydrolase 29
ARG2: Arginase 2 123
ARL14EP: ADP Ribosylation Factor Like GTPase 14 Effector Protein 61
ASTN2: Astrotactin 2 60
BDNF: Brain Derived Neurotrophic Factor 29
BMF: B-cell Lymphoma 2 (BCL2) Modifying Factor 20, 60
BMP4: Bone Morphogenetic Protein 4 23
BOLA2: BolA Family Member 2 195
BPIFB1: BPI Fold Containing Family B Member 1 177
C1QL2: Complement C1q Like 2 196
C2CD4A: C2 Calcium Dependent Domain Containing 4A 135
C2orf50: Chromosome 2 Open Reading Frame 50 166
CATSPERB: Cation Channel Sperm Associated Auxiliary Subunit Beta 136
CCDC170: Coiled-Coil Domain Containing 170 ..... 68, 199
CCND1: Cyclin D1 192
CEP112: Centrosomal Protein 112 ..... 68, 199
CITED2: Cbp/P300 Interacting Transactivator With Glu/Asp Rich Carboxy-Terminal Domain 2 156
CLCF1: Cardiotrophin Like Cytokine Factor 1 29
CLOCK: Circadian Locomotor Output Cycles Kaput 23
COX-2: Prostaglandin-endoperoxide Synthase 2 27
CRELD1: Cysteine Rich With EGF Like Domains 1 159
CST11: Cystatin 11 29
CXCL14: C-X-C Motif Chemokine Ligand 14 137
CYP2C19: Cytochrome P450 Family 2 Subfamily C Member 19 19
CYP4V2: Cytochrome P450 Family 4 Subfamily V Member 2 196
DAF: Decay Accelerating Factor 142
DAPK1: Death Associated Protein Kinase 1 29
DEPP1: Decidual Protein Induced by Progesterone 1 Autophagy Regulator 137, 156
DES: Desmin 138
EDN2: Endothelin 2 169
ELOVL4: ELOVL Fatty Acid Elongase 4 29
ESR: Oestrogen Receptor 182
ESR1: Oestrogen Receptor 1 ..... 18, 94

ESR2: Oestrogen Receptor 2 18
F2RL2: Coagulation Factor II Thrombin Receptor Like 2 158
FAM149A: Family With Sequence Similarity 149 Member A 196
FAM181A: Family With Sequence Similarity 181 Member A 28
FAT2: FAT Atypical Cadherin 2 154
FAT3: FAT Atypical Cadherin 3 156
FSHB: Follicle Stimulating Hormone Beta Subunit 69
GATA6: GATA Binding Protein 6 27
GIPR: Gastric Inhibitory Polypeptide Receptor 154
GLMP: Glycosylated Lysosomal Membrane Protein 155
GPD1: Glycerol-3-Phosphate Dehydrogenase 1 178
GPR158: G Protein-Coupled Receptor 158 196
GREB1: Growth Regulating Oestrogen Receptor Binding 1 41
GREM1: Gremlin 1, DAN Family BMP Antagonist 23
HAND2: Heart And Neural Crest Derivatives Expressed 2 29
HAP1: Huntingtin Associated Protein 1 136
HBA1: Hemoglobin Subunit Alpha 1 154
HBE1: Hemoglobin Subunit Epsilon 1 29
HBG2: Hemoglobin Subunit Gamma 2 175
HOX: Homeobox 182
HOXA10: Homeobox A10 39
ID4: Inhibitor of DNA Binding 4 68
IGF1: Insulin Like Growth Factor 1 20
IGFBP1: Insulin Like Growth Factor Binding Protein 1 134
IL16: Interleukin 16 20
ITGAM: Integrin Subunit Alpha M 142
ITGAX: Integrin Subunit Alpha X 142
KBTBD11: Kelch Repeat And BTB Domain Containing 11 156
KCND2: Potassium Voltage-Gated Channel Subfamily D Member 2 159
KDR: Kinase Insert Domain Receptor 60
LEFTY2: Left-Right Determination Factor 2 23
LYRM4: LYR Motif Containing 4 204
MLXIPL: MLX Interacting Protein Like 154
MUC5AC: Mucin-5AC 179
MUC5B: Mucin-5B 179
MYC: MYC Proto-Oncogene 23
MYH11: Myosin Heavy Chain 11 ..... 138, 159
NEGR1: Neuronal Growth Regulator 1 29
NFE2L3: Nuclear Factor, Erythroid 2 Like 3 68
NGF: Nerve Growth Factor 39
NR5A1: Nuclear Receptor Subfamily 5 Group A Member 1 29
NRIP1: Nuclear Receptor Interacting Protein 1 174
OSR2: Odd-Skipped Related Transciption Factor 2 29
PAGR1: PAXIP1 Associated Glutamate Rich Protein 1 195
PDLIM3: PDZ and LIM Domain 3 138
PDLIM5: PDZ And LIM Domain 5 ..... 20, 75

PER1: Period Circadian Regulator 1 28
PGR: Progesterone Receptor 18
PIANP: PILR Alpha Associated Neural Protein 195
PLG: Plasminogen 159
PLIN1: Perilipin 1 178
PLIN4: Perilipin 4 178
PRKCA: Protein Kinase C Alpha 199
PTGIS: Prostaglandin I2 Synthase 161
PTGS2:Prostaglandin-endoperoxide Synthase 2 27
RAD52: DNA repair protein RAD52 homolog 192
RNLS: Renalase – FAD Dependent Amine Oxidase 69
ROR1: Receptor Tyrosine Kinase Like Orphan Receptor 1 29
RPF2: Ribosome Production Factor 2 Homolog 28
SAA1: Serum Amyloid A1 170
SAA2: Serum Amyloid A2 154
SF-1: Steroidogenic Factor 1 27
SHE: Src Homology 2 Domain Containing E 208
snoU13: Small nucleolar RNA U13 199
SPATA6L: Spermatogenesis Associated 6 Like 154
SPP1: Secreted Phosphoprotein 1 205
SRP14-AS1: Signal Recognition Particle 14 Anti-sense RNA1 97
STEAP3: STEAP3 Metalloreductase 207
SYNE1: Spectrin Repeat Containing Nuclear Envelope Protein 1 60
TBX3: T-Box Transcription Factor 3 29
TDRD10: Tudor Domain Containing 10 196
THBS1: Thrombospondin 1 156
TLR3: Toll Like Receptor 3 204
TTC29: Tetratricopeptide Repeat Domain 29 154
VEGF: Vascular Endothelial Growth Factor 20
WNT4: Wingless-Type MMTV Integration Site Family, Member 4 ..... 68, 199
WNT5A: Wnt Family Member 5A 29
WT1: Wilms' Tumor 1 92
ZNF26: Zinc Finger Protein 26 154

H. Software packages (versions)

Software & Packages

biomaRt: Interface to BioMart databases (Ensembl) (v2.30.0) 106, 148
DAVID: Database for Annotation, Visualization and Integrated Discovery (v6.8); https://david.ncifcrf.gov 107
DESeq: Differential gene expression analysis based on the negative binomial distribution (v1.26.0) 105
DESeq2: Differential gene expression analysis based on the negative binomial distribution 2 (v1.22.2) 105, 148
edgeR: Empirical Analysis of Digital Gene Expression Data in R (v3.16.5) 105
EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling (v1.6.0) 148
forestplot: Advanced Forest Plot Using 'grid' Graphics in R (v1.9) 52
Hisat2: Hierarchical Indexing for Spliced Alignment of Transcripts2 (v2.0.4) 102
ic: Post-imputation data checking programme (http://www.well.ox.ac.uk/~wrayner/tools/Post-Imputation.html) 37
IlluminaHumanMethylationEPICanno.ilm10b2.hg19: Annotation for Illumina's EPIC methylation arrays (v0.6.0) 187
IPA: Ingenuity Pathway analysis; https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis 107
meffil: Efficient algorithms for DNA methylation (v1.1.1) 185
Minfi: Analyze Illumina Infinium DNA methylation arrays (v1.24.0) 184
PANTHER: Protein ANalysis THrough Evolutionary Relationships (v14.1); http://www.pantherdb.org/ 107
PiiL: Pathway Interactive Visualization Tool (v0.14) 187
PLINK: Whole genome association analysis toolset (v1.07) 37
QTLtools: Tool set for molecular QTL discovery and analysis 150
SHAPEIT: method for estimation of haplotypes (aka phasing) from genotype or sequencing data (v2) 37
shinyMethyl: Interactive visualization for Illumina methylation arrays (v1.14.0) 184
snptest: Analysis of Single SNP Association in Genome-wide Studies (v2.5.4-beta3) ... 37, 52
StringTie: Transcript assembly and quantification for RNA-Seq 150
subread: Subread Sequence Alignment and Counting for R (v1.5.0) 102

I. Reference

1. Eskenazi B, Warner ML. Epidemiology of endometriosis. Obstet Gynecol Clin North Am 1997;24:235-58.
2. Adamson GD, Kennedy S, Hummelshoj L. Creating solutions in endometriosis: global collaboration through the World Endometriosis Research Foundation. Journal of Endometriosis and Pelvic Pain Disorders 2010;2:3-6.
3. Dunselman GA, Vermeulen N, Becker C, et al. ESHRE guideline: management of women with endometriosis. Hum Reprod 2014.
4. Kennedy S, Bergqvist A, Chapron C, et al. ESHRE guideline for the diagnosis and treatment of endometriosis. Hum Reprod 2005;20:2698-704.
5. Nnoaham KE, Hummelshoj L, Webster P, et al. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. Fertil Steril 2011;96:366-73.e8.
6. Simoens S, Dunselman G, Dirksen C, et al. The burden of endometriosis: costs and quality of life of women with endometriosis and treated in referral centres. Human Reproduction 2012;27:1292-9.
7. Bulun SE. Endometriosis. N Engl J Med 2009;360:268-79.
8. von Rokitansky C. Uterine gland proliferation in uterine and ovarian sarcomas. Zeitschrift Gesellschaft fur Aerzte zu Wien 1860;37:577.
9. Sampson J. Peritoneal endometriosis due to menstrual dissemination of endometrial tissue into the peritoneal cavity. Am J Obstet Gynecol 1927;14:422-69.
10. Halme J, Hammond MG, Hulka JF, Raj SG, Talbert LM. Retrograde menstruation in healthy women and in patients with endometriosis. Obstet Gynecol 1984;64:151-4.
11. Iv D, Va P, Bm A, et al. Pulmonary endometriosis: Clinico-pathological approach to diagnostics and treatment. Lungs and Breathing 2018;2.
12. Fukuda S, Hirata T, Neriishi K, et al. Thoracic endometriosis syndrome: Comparison between catamenial pneumothorax or endometriosis-related pneumothorax and catamenial hemoptysis. Eur J Obstet Gynecol Reprod Biol 2018;225:118-23.
13. Ferguson BR, Bennington JL, Haber SL. Histochemistry of mucosubstances and histology of mixed müllerian pelvic lymph node glandular inclusions. Evidence for histogenesis by müllerian metaplasia of coelomic epithelium. Obstet Gynecol 1969;33:617-25.
14. Du H, Taylor HS. Contribution of bone marrow-derived stem cells to endometrium and endometriosis. Stem Cells 2007;25:2082-6.
15. Gargett CE, Masuda H. Adult stem cells in the endometrium. Mol Hum Reprod 2010;16:818-34.
16. Tsuji S, Yoshimoto M, Takahashi K, Noda Y, Nakahata T, Heike T. Side population cells contribute to the genesis of human endometrium. Fertil Steril 2008;90:1528-37.
17. Sampson JA. Metastatic or Embolic Endometriosis, due to the Menstrual Dissemination of Endometrial Tissue into the Venous Circulation. Am J Pathol 1927;3:93-110.43.
18. Mechsner S, Weichbrodt M, Riedlinger WF, et al. Estrogen and progestogen receptor positive endometriotic lesions and disseminated cells in pelvic sentinel lymph nodes of patients with deep infiltrating rectovaginal endometriosis: a pilot study. Hum Reprod 2008;23:2202-9.
19. Symons LK, Miller JE, Kay VR, et al. The Immunopathophysiology of Endometriosis. Trends Mol Med 2018;24:748-62.

20. Burney RO, Giudice LC. Pathogenesis and pathophysiology of endometriosis. Fertil Steril 2012;98:511-9.
21. Zondervan KT, Becker CM, Missmer SA. Endometriosis. N Engl J Med 2020;382:1244-56.
22. Walter AJ, Hentz JG, Magtibay PM, Cornella JL, Magrina JF. Endometriosis: correlation between histologic and visual findings at laparoscopy. Am J Obstet Gynecol 2001;184:1407-11; discussion 11-3.
23. Sumathi VP, McCluggage WG. CD10 is useful in demonstrating endometrial stroma at ectopic sites and in confirming a diagnosis of endometriosis. Journal of clinical pathology 2002;55:391-2.
24. Bedaiwy MA, Gupta S, Noriega J, Brainard J, Agarwal A, Falcone T. Diagnostic value of hemosiderin laden macrophages in histologically proven endometriosis. Fertility and Sterility 2008;90:S438.
25. Noventa M, Saccardi C, Litta P, et al. Ultrasound techniques in the diagnosis of deep pelvic endometriosis: algorithm based on a systematic review and meta-analysis. Fertil Steril 2015;104:366-83 e2.
26. Bazot M, Bharwani N, Huchon C, et al. European society of urogenital radiology (ESUR) guidelines: MR imaging of pelvic endometriosis. Eur Radiol 2017;27:2765-75.
27. Jacobs IJ, Menon U. Progress and challenges in screening for early detection of ovarian cancer. Mol Cell Proteomics 2004;3:355-66.
28. Kenemans P, Yedema CA, Bon GG, von Mensdorff-Pouilly S. CA 125 in gynecological pathology--a review. Eur J Obstet Gynecol Reprod Biol 1993;49:115-24.
29. Muyldermans M, Cornillie FJ, Koninckx PR. CA125 and endometriosis. Human Reproduction Update 1995;1:173-87.
30. National Institute for Health and Care Excellence. Endometriosis: diagnosis and management. NICE guideline 2017;NG73; available from: https://www.nice.org.uk/guidance/ng73 [Accessed 09 Sep 2020].
31. Fassbender A, Burney RO, O DF, D'Hooghe T, Giudice L. Update on Biomarkers for the Detection of Endometriosis. Biomed Res Int 2015;2015:130854.
32. Ghai V, Jan H, Shakir F, Haines P, Kent A. Diagnostic delay for superficial and deep endometriosis in the United Kingdom. J Obstet Gynaecol 2020;40:83-9.
33. Ballard K, Lowton K, Wright J. What's the delay? A qualitative study of women's experiences of reaching a diagnosis of endometriosis. Fertil Steril 2006;86:1296-301.
34. Zondervan KT, Becker CM, Koga K, Missmer SA, Taylor RN, Vigano P. Endometriosis. Nat Rev Dis Primers 2018;4:9.
35. Vercellini P, Fedele L, Aimi G, De Giorgi O, Consonni D, Crosignani PG. Reproductive performance, pain recurrence and disease relapse after conservative surgical treatment for endometriosis: the predictive value of the current classification system. Hum Reprod 2006;21:2679-85.
36. Vignali M, Bianchi S, Candiani M, Spadaccini G, Oggioni G, Busacca M. Surgical treatment of deep endometriosis and risk of recurrence. Journal of Minimally Invasive Gynecology 2005;12:508-13.
37. Farquhar C, Prentice A, Singla AA, Selak V. Danazol for pelvic pain associated with endometriosis. Cochrane Database of Systematic Reviews 2007.
38. Bedaiwy MA, Alfaraj S, Yong P, Casper R. New developments in the medical treatment of endometriosis. Fertil Steril 2017;107:555-65.
39. Ulukus M, Cakmak H, Arid A. The Role of Endometrium in Endometriosis. The Journal of the Society for Gynecologic Investigation: JSGI 2006;13:467-76.

40. Slopien R, Meczekalski B. Aromatase inhibitors in the treatment of endometriosis. Prz Menopauzalny 2016;15:43-7.
41. Montgomery GW, Giudice LC. New Lessons about Endometriosis - Somatic Mutations and Disease Heterogeneity. N Engl J Med 2017;376:1881-2.
42. Fauconnier A, Chapron C. Endometriosis and pelvic pain: epidemiological evidence of the relationship and implications. Hum Reprod Update 2005;11:595-606.
43. Prescott J, Farland LV, Tobias DK, et al. A prospective cohort study of endometriosis and subsequent risk of infertility. Hum Reprod 2016;31:1475-82.
44. Sinaii N, Plumb K, Cotton L, et al. Differences in characteristics among 1,000 women with endometriosis based on extent of disease. Fertil Steril 2008;89:538-45.
45. Rogers PA, Adamson GD, Al-Jefout M, et al. Research Priorities for Endometriosis. Reprod Sci 2017;24:202-26.
46. Surrey ES, Soliman AM, Johnson SJ, Davis M, Castelli-Haley J, Snabes MC. Risk of Developing Comorbidities Among Women with Endometriosis: A Retrospective Matched Cohort Study. Journal of Women's Health 2018;27:1114-23.
47. DiVasta AD, Zimmerman LA, Vitonis AF, Fadayomi AB, Missmer SA. Overlap Between Irritable Bowel Syndrome Diagnosis and Endometriosis in Adolescents. Clinical Gastroenterology and Hepatology 2020.
48. Schomacker ML, Hansen KE, Ramlau-Hansen CH, Forman A. Is endometriosis associated with irritable bowel syndrome? A cross-sectional study. Eur J Obstet Gynecol Reprod Biol 2018;231:65-9.
49. Painter JN, Anderson CA, Nyholt DR, et al. Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. Nat Genet 2011;43:51-4.
50. Zondervan KT, Rahmioglu N, Morris AP, et al. Beyond Endometriosis Genome-Wide Association Study: From Genomics to Phenomics to the Patient. Semin Reprod Med 2016;34:242-54.
51. Poli-Neto OB, Meola J, Rosa ESJC, Tiezzi D. Transcriptome meta-analysis reveals differences of immune profile between eutopic endometrium from stage I-II and III-IV endometriosis independently of hormonal milieu. Sci Rep 2020;10:313.
52. Vercellini P, Fedele L, Aimi G, Pietropaolo G, Consonni D, Crosignani PG. Association between endometriosis stage, lesion type, patient characteristics and severity of pelvic pain symptoms: a multivariate analysis of over 1000 patients. Hum Reprod 2007;22:266-71.
53. Urteaga I, McKillop M, Elhadad N. Learning endometriosis phenotypes from patient-generated data. npj Digital Medicine 2020;3.
54. Becker CM, Laufer MR, Stratton P, et al. World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: I. Surgical phenotype data collection in endometriosis research. Fertil Steril 2014.
55. Vitonis AF, Vincent K, Rahmioglu N, et al. World Endometriosis Research Foundation Endometriosis Phenome and biobanking harmonization project: II. Clinical and covariate phenotype data collection in endometriosis research. Fertil Steril 2014.
56. Rahmioglu N, Fassbender A, Vitonis AF, et al. World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonization Project: III. Fluid biospecimen collection, processing, and storage in endometriosis research. Fertil Steril 2014.
57. Fassbender A, Rahmioglu N, Vitonis AF, et al. World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: IV. Tissue collection, processing, and storage in endometriosis research. Fertil Steril 2014.
58. Koninckx PR, Ussia A, Adamyan L, Wattiez A, Donnez J. Deep endometriosis: definition, diagnosis, and treatment. Fertil Steril 2012;98:564-71.

59. Koninckx PR, Martin DC. Deep endometriosis: a consequence of infiltration or retraction or possibly adenomyosis externa? Fertility and Sterility 1992;58:924-8.
60. Johnson NP, Hummelshoj L, Adamson GD, et al. World Endometriosis Society consensus on the classification of endometriosis. Hum Reprod 2017;32:315-24.
61. American Society for Reproductive Medicine. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. Fertil Steril 1997;67:817-21.
62. Andres MP, Borrelli GM, Abrao MS. Endometriosis classification according to pain symptoms: can the ASRM classification be improved? Best Pract Res Clin Obstet Gynaecol 2018;51:111-8.
63. Adamson GD, Pasta DJ. Endometriosis fertility index: the new, validated endometriosis staging system. Fertil Steril 2010;94:1609-15.
64. Haas D, Shebl O, Shamiyeh A, Oppelt P. The rASRM score and the Enzian classification for endometriosis: their strengths and weaknesses. Acta Obstet Gynecol Scand 2013;92:3-7.
65. Tuttlies F, Keckstein J, Ulrich U, et al. ENZIAN-Score, eine Klassifikation der tief infiltrierenden Endometriose. Zentralbl Gynakol 2005;127:275-81.
66. Bulletti C, Coccia ME, Battistoni S, Borini A. Endometriosis and infertility. J Assist Reprod Genet 2010;27:441-7.
67. Yong PJ, Williams C, Bedaiwy MA, Allaire C. A Proposed Platform for Phenotyping Endometriosis-Associated Pain: Unifying Peripheral and Central Pain Mechanisms. Current Obstetrics and Gynecology Reports 2020;9:89-97.
68. Arnold J, Vercellino GF, Chiantera V, Schneider A, Mechsner S, Barcena de Arellano ML. Neuroimmunomodulatory alterations in non-lesional peritoneum close to peritoneal endometriosis. Neuroimmunomodulation 2013;20:9-18.
69. Pongratz G, Straub RH. The sympathetic nervous response in inflammation. Arthritis Research & Therapy 2014;16:504.
70. Coxon L, Horne AW, Vincent K. Pathophysiology of endometriosis-associated pain: A review of pelvic and central nervous system mechanisms. Best Pract Res Clin Obstet Gynaecol 2018;51:53-67.
71. Tamburro S, Canis M, Albuisson E, Dechelotte P, Darcha C, Mage G. Expression of transforming growth factor β1 in nerve fibers is related to dysmenorrhea and laparoscopic appearance of endometriotic implants. Fertility and Sterility 2003;80:1131-6.
72. Kajitani T, Maruyama T, Asada H, et al. Possible involvement of nerve growth factor in dysmenorrhea and dyspareunia associated with endometriosis. Endocr J 2013;60:1155-64.
73. Barcena de Arellano ML, Arnold J, Vercellino GF, et al. Influence of nerve growth factor in endometriosis-associated symptoms. Reprod Sci 2011;18:1202-10.
74. As-Sanie S, Harris RE, Napadow V, et al. Changes in regional gray matter volume in women with chronic pelvic pain: a voxel-based morphometry study. Pain 2012;153:1006-14.
75. As-Sanie S, Kim J, Schmidt-Wilcke T, et al. Functional Connectivity Is Associated With Altered Brain Chemistry in Women With Endometriosis-Associated Chronic Pelvic Pain. The Journal of Pain 2016;17:1-13.
76. Bull JR, Rowland SP, Scherwitzl EB, Scherwitzl R, Danielsson KG, Harper J. Real-world menstrual cycle characteristics of more than 600,000 menstrual cycles. NPJ Digit Med 2019;2:83.
77. Noyes RW, Hertig AT, Rock J. Dating the Endometrial Biopsy. Fertility and Sterility 1950;1:3-25.
78. Vollman RF. The menstrual cycle. Major Probl Obstet Gynecol 1977;7:1-193.
79. Wilcox AJ, Baird DD, Weinberg CR. Time of Implantation of the Conceptus and Loss of Pregnancy. New England Journal of Medicine 1999;340:1796-9.

80. Martín J, Domínguez F, Avila S, et al. Human endometrial receptivity: gene regulation. J Reprod Immunol 2002;55:131-9.
81. Wilcox AJ, Dunson D, Baird DD. The timing of the "fertile window" in the menstrual cycle: day specific estimates from a prospective study. Bmj 2000;321:1259-62.
82. Talbi S, Hamilton AE, Vo KC, et al. Molecular phenotyping of human endometrium distinguishes menstrual cycle phases and underlying biological processes in normo-ovulatory women. Endocrinology 2006;147:1097-121.
83. Wira CR, Rodriguez-Garcia M, Patel MV. The role of sex hormones in immune protection of the female reproductive tract. Nat Rev Immunol 2015;15:217-30.
84. Saare M, Modhukur V, Suhorutshenko M, et al. The influence of menstrual cycle and endometriosis on endometrial methylome. Clin Epigenetics 2016;8:2.
85. Ponnampalam AP, Weston GC, Trajstman AC, Susil B, Rogers PA. Molecular classification of human endometrial cycle stages by transcriptional profiling. Mol Hum Reprod 2004;10:879-93.
86. Fung JN, Girling JE, Lukowski SW, et al. The genetic regulation of transcription in human endometrial tissue. Hum Reprod 2017;32:893-904.
87. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 2000;28:27-30.
88. Grimstad FW, Decherney A. A Review of the Epigenetic Contributions to Endometriosis. Clinical Obstetrics and Gynecology 2017;60.
89. Guo SW. Epigenetics of endometriosis. Mol Hum Reprod 2009;15:587-607.
90. Pellegrini C, Gori I, Achtari C, et al. The expression of estrogen receptors as well as GREB1, c-MYC, and cyclin D1, estrogen-regulated genes implicated in proliferation, is increased in peritoneal endometriosis. Fertil Steril 2012;98:1200-8.
91. Vercellini P, Vigano P, Somigliana E, Fedele L. Endometriosis: pathogenesis and treatment. Nat Rev Endocrinol 2014;10:261-75.
92. Kobayashi H, Uekuri C, Shigetomi H. Towards an understanding of the molecular mechanism of endometriosis: unbalancing epithelial-stromal genetic conflict. Gynecol Endocrinol 2014;30:7-15.
93. Marquardt RM, Kim TH, Shin JH, Jeong JW. Progesterone and Estrogen Signaling in the Endometrium: What Goes Wrong in Endometriosis? Int J Mol Sci 2019;20.
94. Noble LS, Takayama K, Zeitoun KM, et al. Prostaglandin E2 stimulates aromatase expression in endometriosis-derived stromal cells. J Clin Endocrinol Metab 1997;82:600-6.
95. Steinach E, Kun H. TRANSFORMATION OF MALE SEX HORMONES INTO A SUBSTANCE WITH THE ACTION OF A FEMALE HORMONE. The Lancet 1937;230:845.
96. Santen RJ, Brodie H, Simpson ER, Siiteri PK, Brodie A. History of aromatase: saga of an important biological mediator and therapeutic target. Endocr Rev 2009;30:343-75.
97. Zeitoun KM, Bulun SE. Aromatase: a key molecule in the pathophysiology of endometriosis and a therapeutic target. Fertility and Sterility 1999;72:961-9.
98. Bulun SE, Cheng YH, Pavone ME, et al. 17Beta-hydroxysteroid dehydrogenase-2 deficiency and progesterone resistance in endometriosis. Semin Reprod Med 2010;28:44-50.
99. Han SJ, Jung SY, Wu SP, et al. Estrogen Receptor beta Modulates Apoptosis Complexes and the Inflammasome to Drive the Pathogenesis of Endometriosis. Cell 2015;163:960-74.
100. Fujimoto J, Hirose R, Sakaguchi H, Tamaya T. Expression of oestrogen receptor-α and -β in ovarian endometriomata. Molecular Human Reproduction 1999;5:742-7.
101. Xue Q, Lin Z, Cheng YH, et al. Promoter methylation regulates estrogen receptor 2 in human endometrium and endometriosis. Biol Reprod 2007;77:681-7.

102. Han SJ, O'Malley BW. The dynamics of nuclear receptors and nuclear receptor coregulators in the pathogenesis of endometriosis. Hum Reprod Update 2014;20:467-84.
103. Chae U, Min JY, Kim SH, et al. Decreased Progesterone Receptor B/A Ratio in Endometrial Cells by Tumor Necrosis Factor-Alpha and Peritoneal Fluid from Patients with Endometriosis. Yonsei Med J 2016;57:1468-74.
104. Gentilini D, Vigano P, Vignali M, et al. Endometrial stromal progesterone receptor-A/progesterone receptor-B ratio: no difference between women with and without endometriosis. Fertil Steril 2010;94:1538-40.
105. Igarashi TM, Bruner-Tran KL, Yeaman GR, et al. Reduced expression of progesterone receptor-B in the endometrium of women with endometriosis and in cocultures of endometrial cells exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. Fertil Steril 2005;84:67-74.
106. Bedaiwy MA, Dahoud W, Skomorovska-Prokvolit Y, et al. Abundance and Localization of Progesterone Receptor Isoforms in Endometrium in Women With and Without Endometriosis and in Peritoneal and Ovarian Endometriotic Implants. Reprod Sci 2015;22:1153-61.
107. Wu Y, Strawn E, Basir Z, Halverson G, Guo SW. Promoter hypermethylation of progesterone receptor isoform B (PR-B) in endometriosis. Epigenetics 2006;1:106-11.
108. Treloar SA, O'Connor DT, O'Connor VM, Martin NG. Genetic influences on endometriosis in an Australian twin sample. . Fertil Steril 1999;71:701-10.
109. Saha R, Pettersson HJ, Svedberg P, et al. Heritability of endometriosis. Fertility and Sterility 2015;104:947-52.
110. Kennedy S, Mardon H, Barlow D. Familial endometriosis. Journal of Assisted Reproduction and Genetics 1995;12:32-4.
111. Nouri K, Ott J, Krupitz B, Huber JC, Wenzl R. Family incidence of endometriosis in first-, second-, and third-degree relatives: case-control study. Reprod Biol Endocrinol 2010;8:85.
112. Coxhead D, Thomas EJ. Familial inheritance of endometriosis in a British population. A case control study. Journal of Obstetrics and Gynaecology 1993;13:42-4.
113. Moen MH, Magnus P. The familial risk of endometriosis. Acta Obstet Gynecol Scand 1993;72:560-4.
114. Lander ES. The new genomics: Global views of biology. Science 1996;274:536-9.
115. Risch N, Merikangas K. The Future of Genetic Studies of Complex Human Diseases. Science 1996;273:1516.
116. Collins FS, Guyer MS, Chakravarti A. Variations on a Theme: Cataloging Human DNA Sequence Variation. Science 1997;278:1580.
117. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 2003;33:177-82.
118. Lee SH, Harold D, Nyholt DR, et al. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. Hum Mol Genet 2013;22:832-41.
119. Zondervan KT, Treloar SA, Lin J, et al. Significant evidence of one or more susceptibility loci for endometriosis with near-Mendelian inheritance on chromosome 7p13-15. Hum Reprod 2007;22:717-28.
120. Treloar SA, Wicks J, Nyholt DR, et al. Genomewide linkage study in 1,176 affected sister pair families identifies a significant susceptibility locus for endometriosis on chromosome 10q26. Am J Hum Genet 2005;77:365-76.

121. Painter JN, Nyholt DR, Morris A, et al. High-density fine-mapping of a chromosome 10q26 linkage peak suggests association between endometriosis and variants close to CYP2C19. Fertil Steril 2011;95:2236-40.
122. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. Endocrine reviews 2011;32:81-151.
123. Azimzadeh P, Khorram Khorshid HR, Akhondi MM, Shirazi A. Association of interleukin-16 polymorphisms with disease progression and susceptibility in endometriosis. Int J Immunogenet 2016;43:297-302.
124. Matalliotakis M, Zervou MI, Eliopoulos E, et al. The role of IL16 gene polymorphisms in endometriosis. Int J Mol Med 2018;41:1469-76.
125. Fang F, Gong L, Wang X, Zhang L. The association between vascular endothelial growth factor (VEGF) +405G>C genetic polymorphism and endometriosis. Exp Biol Med (Maywood) 2015;240:1177-82.
126. Liang S, Huang Y, Fan Y. Vascular endothelial growth factor gene polymorphisms and endometriosis risk: a meta-analysis. Arch Gynecol Obstet 2012;286:139-46.
127. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nature Reviews Genetics 2002;3:391-7.
128. Rahmioglu N, Montgomery GW, Zondervan KT. Genetics of Endometriosis. Women's Health 2015;11:577-86.
129. Tam V, Patel N, Turcotte M, Bosse Y, Pare G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet 2019;20:467-84.
130. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 2008;9:356-69.
131. Pearson TA, Manolio TA. How to Interpret a Genome-wide Association Study. JAMA 2008;299:1335-44.
132. Hunter DJMBBSS, Altshuler DMDP, Rader DJMD. From Darwin's Finches to Canaries in the Coal Mine -- Mining the Genome for New Biology. The New England Journal of Medicine 2008;358:2760-3.
133. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 2005;6:109-18.
134. Hemminki K, Forsti A, Bermejo JL. The 'common disease-common variant' hypothesis and familial risks. PLoS One 2008;3:e2504.
135. Rahmioglu N, Banasik K, Christofidou P, et al. Large-scale genome-wide association meta-analysis of endometriosis reveals 13 novel loci and genetically-associated comorbidity with other pain conditions. bioRxiv 2018.
136. Nyholt DR, Low SK, Anderson CA, et al. Genome-wide association meta-analysis identifies new endometriosis risk loci. Nat Genet 2012;44:1355-9.
137. Albertsen HM, Chettier R, Farrington P, Ward K. Genome-wide association study link novel loci to endometriosis. PLoS One 2013;8:e58257.
138. Rahmioglu N, Nyholt DR, Morris AP, Missmer SA, Montgomery GW, Zondervan KT. Genetic variants underlying risk of endometriosis: insights from meta-analysis of eight genome-wide association and replication datasets. Hum Reprod Update 2014;20:702-16.
139. Sapkota Y, Low SK, Attia J, et al. Association between endometriosis and the interleukin 1A (IL1A) locus. Hum Reprod 2015;30:239-48.
140. Steinthorsdottir V, Thorleifsson G, Aradottir K, et al. Common variants upstream of KDR encoding VEGFR2 and in TTC39B associate with endometriosis. Nat Commun 2016;7:12350.

141. Sapkota Y, Steinthorsdottir V, Morris AP, et al. Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. Nature Communications 2017;8:15539.
142. Witz CA, Monotoya-Rodriguez IA, Schenken RS. Whole explants of peritoneum and endometrium: a novel model of the early endometriosis lesion. Fertility and Sterility 1999;71:56-60.
143. Chung H-W, Wen Y, Chun S-H, Nezhat C, Woo B-H, Lake Polan M. Matrix metalloproteinase-9 and tissue inhibitor of metalloproteinase-3 mRNA expression in ectopic and eutopic endometrium in women with endometriosis: a rationale for endometriotic invasiveness. Fertility and Sterility 2001;75:152-9.
144. Burney RO, Talbi S, Hamilton AE, et al. Gene expression analysis of endometrium reveals progesterone resistance and candidate susceptibility genes in women with endometriosis. Endocrinology 2007;148:3814-26.
145. Fung JN, Mortlock S, Girling JE, et al. Genetic regulation of disease risk and endometrial gene expression highlights potential target genes for endometriosis and polycystic ovarian syndrome. Sci Rep 2018;8:11424.
146. Mortlock S, Kendarsari RI, Fung JN, et al. Tissue specific regulation of transcription in endometrium and association with disease. Hum Reprod 2020;35:377-93.
147. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One 2014;9:e78644.
148. Herbert ZT, Kershner JP, Butty VL, et al. Cross-site comparison of ribosomal depletion kits for Illumina RNAseq library construction. BMC Genomics 2018;19:199.
149. 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 2012;10:84.
150. Suryawanshi S, Huang X, Elishaev E, et al. Complement pathway is frequently altered in endometriosis and endometriosis-associated ovarian cancer. Clin Cancer Res 2014;20:6163-74.
151. Crispi S, Piccolo MT, D'Avino A, et al. Transcriptional profiling of endometriosis tissues identifies genes related to organogenesis defects. Journal of Cellular Physiology 2013;228:1927-34.
152. Fassbender A, Verbeeck N, Bornigen D, et al. Combined mRNA microarray and proteomic analysis of eutopic endometrium of women with and without endometriosis. Hum Reprod 2012;27:2020-9.
153. Aghajanova L, Giudice LC. Molecular evidence for differences in endometrium in severe versus mild endometriosis. Reprod Sci 2011;18:229-51.
154. Edgar R, Tan PPC, Portales-Casamar E, Pavlidis P. Meta-analysis of human methylomes reveals stably methylated sequences surrounding CpG islands associated with high gene expression. Epigenetics & Chromatin 2014;7:28.
155. Lovkvist C, Dodd IB, Sneppen K, Haerter JO. DNA methylation in human epigenomes depends on local topology of CpG sites. Nucleic Acids Res 2016;44:5123-32.
156. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res 2017;45:e22.
157. Pidsley R, Zotenko E, Peters TJ, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 2016;17:208.

158. Wu Y, Halverson G, Basir Z, Strawn E, Yan P, Guo SW. Aberrant methylation at HOXA10 may be responsible for its aberrant expression in the endometrium of patients with endometriosis. Am J Obstet Gynecol 2005;193:371-80.
159. Rahmioglu N, Drong AW, Lockstone H, et al. Variability of genome-wide DNA methylation and mRNA expression profiles in reproductive and endocrine disease related tissues. Epigenetics 2017;12:897-908.
160. Xue Q, Lin Z, Yin P, et al. Transcriptional activation of steroidogenic factor-1 by hypomethylation of the 5' CpG island in endometriosis. J Clin Endocrinol Metab 2007;92:3261-7.
161. Attar E, Tokunaga H, Imir G, et al. Prostaglandin E2 via steroidogenic factor-1 coordinately regulates transcription of steroidogenic genes necessary for estrogen synthesis in endometriosis. J Clin Endocrinol Metab 2009;94:623-31.
162. Zidan HE, Rezk NA, Alnemr AA, Abd El Ghany AM. COX-2 gene promoter DNA methylation status in eutopic and ectopic endometrium of Egyptian women with endometriosis. J Reprod Immunol 2015;112:63-7.
163. Ito T, Kato T, Iwama Y, et al. Prostaglandin H2 as an endothelium-derived contracting factor and its interaction with endothelium-derived nitric oxide. J Hypertens 1991;9:729-36.
164. Dyson MT, Roqueiro D, Monsivais D, et al. Genome-wide DNA methylation analysis predicts an epigenetic switch for GATA factor expression in endometriosis. PLoS Genet 2014;10:e1004158.
165. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc 2013;8:1551-66.
166. Naqvi H, Ilagan Y, Krikun G, Taylor HS. Altered genome-wide methylation in endometriosis. Reprod Sci 2014;21:1237-43.
167. 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;95:93.
168. Houshdaran S, Oke AB, Fung JC, Vo KC, Nezhat C, Giudice LC. Steroid hormones regulate genome-wide epigenetic programming and gene transcription in human endometrial cells with marked aberrancies in endometriosis. PLoS Genet 2020;16:e1008601.
169. Yotova I, Hsu E, Do C, et al. Epigenetic Alterations Affecting Transcription Factors and Signaling Pathways in Stromal Cells of Endometriosis. PLoS One 2017;12:e0170859.
170. Xiong Z, Li M, Yang F, et al. EWAS Data Hub: a resource of DNA methylation array data and metadata. Nucleic Acids Res 2020;48:D890-D5.
171. Wang L, Zhao J, Li Y, Wang Z, Kang S. Genome-wide analysis of DNA methylation in endometriosis using Illumina Human Methylation 450 K BeadChips. Mol Reprod Dev 2019;86:491-501.
172. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019;95:103208.
173. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377-81.
174. Tapmeier TT, Nazri HM, Subramaniam KS, et al. Protocol for a longitudinal, prospective cohort study investigating the biology of uterine fibroids and endometriosis, and patients' quality of life: the FENOX study. BMJ Open 2020;10:e032220.
175. Montgomery GW, Nyholt DR, Zhao ZZ, et al. The search for genes contributing to endometriosis risk. Hum Reprod Update 2008;14:447-57.
176. Giudice LC, Kao LC. Endometriosis. Lancet 2004;364:1789-99.

177. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445:881-5.
178. Du H, Taylor HS. The Role of Hox Genes in Female Reproductive Tract Development, Adult Function, and Fertility. Cold Spring Harb Perspect Med 2015;6:a023002.
179. International HapMap C, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. Nature 2010;467:52-8.
180. Bourdel N, Alves J, Pickering G, Ramilo I, Roman H, Canis M. Systematic review of endometriosis pain assessment: how to choose a scale? Hum Reprod Update 2015;21:136-52.
181. Drossman DA. Functional Gastrointestinal Disorders: History, Pathophysiology, Clinical Features and Rome IV. Gastroenterology 2016.
182. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 2007;39:906-13.
183. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-1.
184. American Society for Reproductive Medicine. Endometriosis and infertility: a committee opinion. Fertil Steril 2012;98:591-8.
185. Herrick S, Evers DM, Lee JY, Udagawa N, Pak DT. Postsynaptic PDLIM5/Enigma Homolog binds SPAR and causes dendritic spine shrinkage. Mol Cell Neurosci 2010;43:188-200.
186. Gallagher CS, Makinen N, Harris HR, et al. Genome-wide association and epidemiological analyses reveal common genetic origins between uterine leiomyomata and endometriosis. Nat Commun 2019;10:4857.
187. Bogart LM, Berry SH, Clemens JQ. Symptoms of interstitial cystitis, painful bladder syndrome and similar diseases in women: a systematic review. J Urol 2007;177:450-6.
188. Haines CN, Braunreiter KM, Mo XM, Burd CJ. GREB1 isoforms regulate proliferation independent of ERalpha co-regulator activities in breast cancer. Endocr Relat Cancer 2018;25:735-46.
189. Rae JM, Johnson MD, Scheys JO, Cordero KE, Larios JM, Lippman ME. GREB 1 is a critical regulator of hormone dependent breast cancer growth. Breast Cancer Res Treat 2005;92:141-9.
190. Mohammed H, D'Santos C, Serandour AA, et al. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. Cell Rep 2013;3:342-9.
191. Shang Y, Hu X, DiRenzo J, Lazar MA, Brown M. Cofactor Dynamics and Sufficiency in Estrogen Receptor–Regulated Transcription. Cell 2000;103:843-52.
192. Rafnar T, Gunnarsson B, Stefansson OA, et al. Variants associating with uterine leiomyoma highlight genetic background shared by various cancers and hormone-related traits. Nat Commun 2018;9:3636.
193. Klattig J, Sierig R, Kruspe D, Makki MS, Englert C. WT1-mediated gene regulation in early urogenital ridge development. Sex Dev 2007;1:238-54.
194. Edwards TL, Giri A, Hellwege JN, et al. A Trans-Ethnic Genome-Wide Association Study of Uterine Fibroids. Front Genet 2019;10:511.
195. O'Mara TA, Glubb DM, Amant F, et al. Identification of nine new susceptibility loci for endometrial cancer. Nat Commun 2018;9:3166.
196. Paul P, van den Hoorn T, Jongsma ML, et al. A Genome-wide multidimensional RNAi screen reveals pathways controlling MHC class II antigen presentation. Cell 2011;145:268-83.

197. Stolk L, Perry JR, Chasman DI, et al. Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. Nat Genet 2012;44:260-8.
198. Ruth KS, Campbell PJ, Chew S, et al. Genome-wide association study with 1000 genomes imputation identifies signals for nine sex hormone-related phenotypes. Eur J Hum Genet 2016;24:284-90.
199. Hayes MG, Urbanek M, Ehrmann DA, et al. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. Nat Commun 2015;6:7502.
200. Day F, Karaderi T, Jones MR, et al. Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. PLoS Genet 2018;14:e1007813.
201. Matalliotakis M, Zervou MI, Matalliotaki C, et al. The role of gene polymorphisms in endometriosis. Mol Med Rep 2017;16:5881-6.
202. Nicholson RI, Gee JM. Oestrogen and growth factor cross-talk and endocrine insensitivity and acquired resistance in breast cancer. Br J Cancer 2000;82:501-13.
203. Straub RH. The complex role of estrogens in inflammation. Endocr Rev 2007;28:521-74.
204. Vincent K, Tracey I. Hormones and their Interaction with the Pain Experience. Rev Pain 2008;2:20-4.
205. Porter W, Saville B, Hoivik D, Safe S. Functional Synergy between the Transcription Factor Sp1 and the Estrogen Receptor. Molecular Endocrinology 1997;11:1569-80.
206. Shaw RJ, Bardeesy N, Manning BD, et al. The LKB1 tumor suppressor negatively regulates mTOR signaling. Cancer Cell 2004;6:91-9.
207. Nath-Sain S, Marignani PA. LKB1 catalytic activity contributes to estrogen receptor alpha signaling. Mol Biol Cell 2009;20:2785-95.
208. Shaw RJ. LKB1 and AMP-activated protein kinase control of mTOR signalling and growth. Acta Physiol (Oxf) 2009;196:65-80.
209. Gros-Louis F, Dupre N, Dion P, et al. Mutations in SYNE1 lead to a newly discovered form of autosomal recessive cerebellar ataxia. Nat Genet 2007;39:80-5.
210. Synofzik M, Smets K, Mallaret M, et al. SYNE1 ataxia is a common recessive ataxia with major non-cerebellar features: a large multi-centre study. Brain 2016;139:1378-93.
211. Jiang P, Li Y, Poleshko A, et al. The Protein Encoded by the CCDC170 Breast Cancer Gene Functions to Organize the Golgi-Microtubule Network. EBioMedicine 2017;22:28-43.
212. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 2019;47:D1005-D12.
213. Albuquerque RJ, Hayashi T, Cho WG, et al. Alternatively spliced vascular endothelial growth factor receptor-2 is an essential endogenous inhibitor of lymphatic vessel growth. Nat Med 2009;15:1023-30.
214. Jakeman LB, Armanini M, Phillips HS, Ferrara N. Developmental expression of binding sites and messenger ribonucleic acid for vascular endothelial growth factor suggests a role for this protein in vasculogenesis and angiogenesis. Endocrinology 1993;133:848-59.
215. Ferrara N, Davis-Smyth T. The Biology of Vascular Endothelial Growth Factor. Endocrine Reviews 1997;18:4-25.
216. Leung DW, Cachianes G, Kuang WJ, Goeddel DV, Ferrara N. Vascular endothelial growth factor is a secreted angiogenic mitogen. Science 1989;246:1306.

217. Ferrara N. Vascular endothelial growth factor: basic science and clinical progress. Endocr Rev 2004;25:581-611.
218. Reinders MEJ, Sho M, Izawa A, et al. Proinflammatory functions of vascular endothelial growth factor in alloimmunity. Journal of Clinical Investigation 2003;112:1655-65.
219. Morales-Ruiz M, Fulton D, Sowa G, et al. Vascular Endothelial Growth Factor–Stimulated Actin Reorganization and Migration of Endothelial Cells Is Regulated via the Serine/Threonine Kinase Akt. Circulation Research 2000;86:892-6.
220. Hamilton JL, Nagao M, Levine BR, Chen D, Olsen BR, Im HJ. Targeting VEGF and Its Receptors for the Treatment of Osteoarthritis and Associated Pain. J Bone Miner Res 2016;31:911-24.
221. Matsuoka A, Maeda O, Mizutani T, et al. Bevacizumab Exacerbates Paclitaxel-Induced Neuropathy: A Retrospective Cohort Study. PLOS ONE 2016;11:e0168707.
222. Llorian-Salvador M, Gonzalez-Rodriguez S. Painful Understanding of VEGF. Front Pharmacol 2018;9:1267.
223. Sugino N, Kashida S, Karube-Harada A, Takiguchi S, Kato H. Expression of vascular endothelial growth factor (VEGF) and its receptors in human endometrium throughout the menstrual cycle and in early pregnancy. Reproduction 2002;123:379-87.
224. Herve MA, Meduri G, Petit FG, et al. Regulation of the vascular endothelial growth factor (VEGF) receptor Flk-1/KDR by estradiol through VEGF in uterus. J Endocrinol 2006;188:91-9.
225. Bourlev V, Volkov N, Pavlovitch S, Lets N, Larsson A, Olovsson M. 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 2006;132:501-9.
226. Wilson PM, Fryer RH, Fang Y, Hatten ME. Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration. J Neurosci 2010;30:8529-40.
227. Freilinger T, Anttila V, de Vries B, et al. Genome-wide association analysis identifies susceptibility loci for migraine without aura. Nat Genet 2012;44:777-82.
228. Gormley P, Anttila V, Winsvold BS, et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. Nat Genet 2016;48:856-66.
229. Yang MH, Wang PH, Wang SJ, Sun WZ, Oyang YJ, Fuh JL. Women with endometriosis are more likely to suffer from migraines: a population-based study. PLoS One 2012;7:e33941.
230. Adewuyi EO, Sapkota Y, International Endogene Consortium I, et al. Shared Molecular Genetic Mechanisms Underlie Endometriosis and Migraine Comorbidity. Genes (Basel) 2020;11.
231. Puthalakath H, Villunger A, Reilly LA, et al. Bmf: A Proapoptotic BH3-Only Protein Regulated by Interaction with the Myosin V Actin Motor Complex, Activated by Anoikis. Science 2001;293:1829.
232. Hausmann M, Leucht K, Ploner C, et al. BCL-2 modifying factor (BMF) is a central regulator of anoikis in human intestinal epithelial cells. J Biol Chem 2011;286:26533-40.
233. Jordan-Sciutto KL, Dragich JM, Bowser R. DNA Binding Activity of the Fetal Alz-50 Clone 1 (FAC1) Protein Is Enhanced by Phosphorylation. Biochemical and Biophysical Research Communications 1999;260:785-9.
234. Schoonover S, Davies P, Bowser R. Immunolocalization and Redistribution of the FAC1 Protein in Alzheimer's Disease. Journal of Neuropathology & Experimental Neurology 1996;55:444-54.

235. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012;40:D930-4.
236. Kheradpour P, Kellis M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. Nucleic Acids Res 2014;42:2976-87.
237. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science (New York, NY) 2015;348:648-60.
238. Zhai G, Teumer A, Stolk L, et al. Eight common genetic variants associated with serum DHEAS levels suggest a key role in ageing mechanisms. PLoS Genet 2011;7:e1002025.
239. Pott J, Bae YJ, Horn K, et al. Genetic Association Study of Eight Steroid Hormones and Implications for Sexual Dimorphism of Coronary Artery Disease. J Clin Endocrinol Metab 2019;104:5008-23.
240. Rahmani A, Shoae-Hassani A, Keyhanvar P, Kheradmand D, Darbandi-Azar A. Dehydroepiandrosterone stimulates nerve growth factor and brain derived neurotrophic factor in cortical neurons. Adv Pharmacol Sci 2013;2013:506191.
241. Labrie F. Intracrinology. Molecular and Cellular Endocrinology 1991;78:C113-C8.
242. Kramer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics 2014;30:523-30.
243. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44-57.
244. Mi H, Muruganujan A, Huang X, et al. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat Protoc 2019;14:703-21.
245. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res 2019;47:D419-D26.
246. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. Nucleic Acids Res 2019;47:D590-D5.
247. Nishimura D. BioCarta. Biotech Software & Internet Report 2001;2:117-20.
248. Becker KG, White SL, Muller J, Engel J. BBID: the biological biochemical image database. Bioinformatics 2000;16:745-6.
249. Seo M, Kim K, Yoon J, et al. RNA-seq analysis for detecting quantitative trait-associated genes. Sci Rep 2016;6:24375.
250. Bolte AC, Dutta AB, Hurt ME, et al. Meningeal lymphatic dysfunction exacerbates traumatic brain injury pathogenesis. Nat Commun 2020;11:4524.
251. Nativio R, Lan Y, Donahue G, et al. An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. Nat Genet 2020.
252. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
253. Crow M, Lim N, Ballouz S, Pavlidis P, Gillis J. Predictability of human differential gene expression. Proc Natl Acad Sci U S A 2019;116:6491-500.
254. Giudice LC. Multifaceted roles for IGFBP-1 in human endometrium during implantation and pregnancy. Ann N Y Acad Sci 1997;828:146-56.
255. Warton K, Foster NC, Gold WA, Stanley KK. A novel gene family induced by acute inflammation in endothelial cells. Gene 2004;342:85-95.
256. Mathew D, Drury JA, Valentijn AJ, Vasieva O, Hapangama DK. In silico, in vitro and in vivo analysis identifies a potential role for steroid hormone regulation of FOXD3 in endometriosis-associated genes. Hum Reprod 2016;31:345-54.

257. Chan C, Virtanen C, Winegarden NA, Colgan TJ, Brown TJ, Greenblatt EM. Discovery of biomarkers of endometrial receptivity through a minimally invasive approach: a validation study with implications for assisted reproduction. Fertil Steril 2013;100:810-7.
258. Kotlyar M, Pastrello C, Malik Z, Jurisica I. IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. Nucleic Acids Res 2019;47:D581-D9.
259. Liu J, Xia J, Cho KH, Clapham DE, Ren D. CatSperbeta, a novel transmembrane protein in the CatSper channel complex. J Biol Chem 2007;282:18945-52.
260. Tseng LH, Chen I, Chen MY, Yan H, Wang CN, Lee CL. Genome-based expression profiling as a single standardized microarray platform for the diagnosis of endometrial disorder: an array of 126-gene model. Fertil Steril 2010;94:114-9.
261. Wu LL-y, Zhou X-F. Huntingtin associated protein 1 and its functions. Cell Adh Migr 2009;3:71-6.
262. Hu S, Yao G, Wang Y, et al. Transcriptomic changes during the pre-receptive to receptive transition in human endometrium detected by RNA-Seq. J Clin Endocrinol Metab 2014;99:E2744-53.
263. D'Aurora M, Romani F, Franchi S, et al. MRAP2 regulates endometrial receptivity and function. Gene 2019;703:7-12.
264. Hromas R, Broxmeyer HE, Kim C, et al. Cloning of BRAK, a Novel Divergent CXC Chemokine Preferentially Expressed in Normal versus Malignant Cells. Biochemical and Biophysical Research Communications 1999;255:703-6.
265. Frick IM, Nordin SL, Baumgarten M, et al. Constitutive and inflammation-dependent antimicrobial peptides produced by epithelium are differentially processed and inactivated by the commensal Finegoldia magna and the pathogen Streptococcus pyogenes. J Immunol 2011;187:4300-9.
266. Mokhtar NM, Cheng CW, Cook E, Bielby H, Smith SK, Charnock-Jones DS. Progestin regulates chemokine (C-X-C motif) ligand 14 transcript level in human endometrium. Mol Hum Reprod 2010;16:170-7.
267. Diaz-Gimeno P, Horcajadas JA, Martinez-Conejero JA, et al. A genomic diagnostic tool for human endometrial receptivity based on the transcriptomic signature. Fertil Steril 2011;95:50-60, e1-15.
268. Watanabe H, Nonoguchi K, Sakurai T, Masuda T, Itoh K, Fujita J. A novel protein Depp, which is induced by progesterone in human endometrial stromal cells activates Elk-1 transcription factor. Mol Hum Reprod 2005;11:471-6.
269. Salcher S, Hermann M, Kiechl-Kohlendorfer U, Ausserlechner MJ, Obexer P. C10ORF10/DEPP-mediated ROS accumulation is a critical modulator of FOXO3-induced autophagy. Mol Cancer 2017;16:95.
270. Choi J, Jo M, Lee E, Oh YK, Choi D. The role of autophagy in human endometrium. Biol Reprod 2012;86:70.
271. Hnia K, Tronchere H, Tomczak KK, et al. Myotubularin controls desmin intermediate filament architecture and mitochondrial dynamics in human and mouse skeletal muscle. J Clin Invest 2011;121:70-85.
272. Muñoz-Mármol AM, Strasser G, Isamat M, et al. A dysfunctional desmin mutation in a patient with severe generalized myopathy. Proceedings of the National Academy of Sciences 1998;95:11312.
273. Kohnen G, Campbell S, Jeffers MD, Cameron IT. Spatially regulated differentiation of endometrial vascular smooth muscle cells. Human Reproduction 2000;15:284-92.

274. Nucci MR, O'Connell JT, Huettner PC, Cviko A, Sun D, Quade BJ. h-Caldesmon Expression Effectively Distinguishes Endometrial Stromal Tumors From Uterine Smooth Muscle Tumors. The American Journal of Surgical Pathology 2001;25.
275. Chu PG, Arber DA, Weiss LM, Chang KL. Utility of CD10 in Distinguishing between Endometrial Stromal Sarcoma and Uterine Smooth Muscle Tumors: An Immunohistochemical Comparison of 34 Cases. Modern Pathology 2001;14:465-71.
276. Xia H, Winokur ST, Kuo W-L, Altherr MR, Bredt DS. Actinin-associated LIM Protein: Identification of a Domain Interaction between PDZ and Spectrin-like Repeat Motifs. Journal of Cell Biology 1997;139:507-15.
277. Morris K, Ihnatovych I, Ionetz E, Reed J, Braundmeier A, Strakova Z. Cofilin and slingshot localization in the epithelium of uterine endometrium changes during the menstrual cycle and in endometriosis. Reprod Sci 2011;18:1014-24.
278. Slater M, Cooper M, Murphy CR. The cytoskeletal proteins alpha-actinin, Ezrin, and talin are De-expressed in endometriosis and endometrioid carcinoma compared with normal uterine epithelium. Appl Immunohistochem Mol Morphol 2007;15:170-4.
279. Messaoudi S, El Kasmi I, Bourdiec A, et al. 15 years of transcriptomic analysis on endometrial receptivity: what have we learnt? Fertil Res Pract 2019;5:9.
280. García-Campos MA, Espinal-Enríquez J, Hernández-Lemus E. Pathway Analysis: State of the Art. Front Physiol 2015;6:383-.
281. Cousins FL, Kirkwood PM, Saunders PT, Gibson DA. Evidence for a dynamic role for mononuclear phagocytes during endometrial repair and remodelling. Sci Rep 2016;6:36748.
282. Regal JF, Gilbert JS, Burwick RM. The complement system and adverse pregnancy outcomes. Mol Immunol 2015;67:56-70.
283. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 2016;11:1650-67.
284. Delaneau O, Ongen H, Brown AA, Fort A, Panousis NI, Dermitzakis ET. A complete tool set for molecular QTL discovery and analysis. Nat Commun 2017;8:15452.
285. Karolchik D, Hinrichs AS, Furey TS, et al. The UCSC Table Browser data retrieval tool. Nucleic Acids Research 2004;32:D493-D6.
286. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 2015;31:3555-7.
287. Rahmioglu N, Drong A, Lockstone H, Lindgren CM, Becker CM, Zondervan KT. Variability of Genome-Wide Gene Expression and DNA Methylation Profiles Across Tissue Samples From Women With and Without Endometriosis. REPRODUCTIVE SCIENCES; 2015: SAGE PUBLICATIONS INC 2455 TELLER RD, THOUSAND OAKS, CA 91320 USA. p. 209A-A.
288. Economos K, MacDonald PC, Casey ML. Endothelin-1 gene expression and protein biosynthesis in human endometrium: potential modulator of endometrial blood flow. The Journal of Clinical Endocrinology & Metabolism 1992;74:14-9.
289. Keator CS, Mah K, Ohm L, Slayden OD. Estrogen and progesterone regulate expression of the endothelins in the rhesus macaque endometrium. Hum Reprod 2011;26:1715-28.
290. Ahmed A, Cameron IT, Ferriani RA, Smith SK. Activation of phospholipase A2 and phospholipase C by endothelin-1 in human endometrium. Journal of Endocrinology 1992;135:383-90.

291. Cacioppo JA, Lin PP, Hannon PR, McDougle DR, Gal A, Ko C. Granulosa cell endothelin-2 expression is fundamental for ovulatory follicle rupture. Sci Rep 2017;7:817.
292. Grimshaw MJ, Wilson JL, Balkwill FR. Endothelin-2 is a macrophage chemoattractant: implications for macrophage distribution in tumors. European Journal of Immunology 2002;32:2393-400.
293. Grimshaw MJ, Hagemann T, Ayhan A, Gillett CE, Binder C, Balkwill FR. A Role for Endothelin-2 and Its Receptors in Breast Tumor Cell Invasion. Cancer Research 2004;64:2461.
294. Wu J, Xie H, Yao S, Liang Y. Macrophage and nerve interaction in endometriosis. J Neuroinflammation 2017;14:53.
295. Capobianco A, Rovere-Querini P. Endometriosis, a disease of the macrophage. Front Immunol 2013;4:9.
296. Thul PJ, Akesson L, Wiking M, et al. A subcellular map of the human proteome. Science 2017;356.
297. Baba S, Takahashi T, Kasama T, Shirasawa H. Identification of two novel amyloid A protein subsets coexisting in an individual patient of AA-amyloidosis. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1992;1180:195-200.
298. Cocco E, Bellone S, El-Sahwi K, et al. Serum amyloid A: a novel biomarker for endometrial cancer. Cancer 2010;116:843-51.
299. Cocco E, Bellone S, El-Sahwi K, et al. Serum amyloid A (SAA): a novel biomarker for uterine serous papillary cancer. Br J Cancer 2009;101:335-41.
300. Zhang SD, Dong SW, Wang DS, et al. Differential proteomic profiling of endometrium and plasma indicate the importance of hydrolysis in bovine endometritis. J Dairy Sci 2017;100:9324-37.
301. Sikora M, Krol J, Nowak M, Stefaniak T, Aubertsson G, Kozdrowski R. The usefulness of uterine lavage and acute phase protein levels as a diagnostic tool for subclinical endometritis in Icelandic mares. Acta Vet Scand 2016;58:50.
302. Girling JE, Lockhart MG, Olshansky M, et al. Differential Gene Expression in Menstrual Endometrium From Women With Self-Reported Heavy Menstrual Bleeding. Reprod Sci 2016.
303. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. Science 2015;347:1260419.
304. Ahmed AF, de Bock CE, Lincz LF, et al. FAT1 cadherin acts upstream of Hippo signalling through TAZ to regulate neuronal differentiation. Cell Mol Life Sci 2015;72:4653-69.
305. Nibbeling EAR, Duarri A, Verschuuren-Bemelmans CC, et al. Exome sequencing and network analysis identifies shared mechanisms underlying spinocerebellar ataxia. Brain 2017;140:2860-78.
306. Sigrist CJ, de Castro E, Cerutti L, et al. New and continuing developments at PROSITE. Nucleic Acids Res 2013;41:D344-7.
307. Katoh Y, Katoh M. Comparative integromics on FAT1, FAT2, FAT3 and FAT4. Int J Mol Med 2006;18:523-8.
308. Simantov R, Febbraio M, Crombie R, Asch AS, Nachman RL, Silverstein RL. Histidine-rich glycoprotein inhibits the antiangiogenic effect of thrombospondin-1. J Clin Invest 2001;107:45-52.
309. Staniszewska I, Zaveri S, Del Valle L, et al. Interaction of alpha9beta1 integrin with thrombospondin-1 promotes angiogenesis. Circ Res 2007;100:1308-16.

310. Nunes SS, Outeiro-Bernstein MA, Juliano L, et al. Syndecan-4 contributes to endothelial tubulogenesis through interactions with two motifs inside the pro-angiogenic N-terminal domain of thrombospondin-1. J Cell Physiol 2008;214:828-37.
311. Mumby SM, Raugi GJ, Bornstein P. Interactions of thrombospondin with extracellular matrix proteins: selective binding to type V collagen. Journal of Cell Biology 1984;98:646-52.
312. Martin-Manso G, Galli S, Ridnour LA, Tsokos M, Wink DA, Roberts DD. Thrombospondin 1 promotes tumor macrophage recruitment and enhances tumor cell cytotoxicity of differentiated U937 cells. Cancer Res 2008;68:7090-9.
313. Bhagwat SR, Chandrashekar DS, Kakar R, et al. Endometrial receptivity: a revisit to functional genomics studies on human endometrium and creation of HGEx-ERdb. PLoS One 2013;8:e58419.
314. Yates AD, Achuthan P, Akanni W, et al. Ensembl 2020. Nucleic Acids Res 2020;48:D682-D8.
315. Tutar Y. Pseudogenes. Comp Funct Genomics 2012;2012:424526.
316. Fagerberg L, Hallstrom BM, Oksvold P, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics 2014;13:397-406.
317. Recklies AD, Ling H, White C, Bernier SM. Inflammatory cytokines induce production of CHI3L1 by articular chondrocytes. J Biol Chem 2005;280:41213-21.
318. Johansen JS, Moller S, Price PA, et al. Plasma YKL-40: a new potential marker of fibrosis in patients with alcoholic cirrhosis? Scand J Gastroenterol 1997;32:582-90.
319. Korthagen NM, van Moorsel CH, Barlo NP, et al. Serum and BALF YKL-40 levels are predictors of survival in idiopathic pulmonary fibrosis. Respir Med 2011;105:106-13.
320. Zhang R, Song J, Isgren A, et al. Genome-wide study of immune biomarkers in cerebrospinal fluid and serum from patients with bipolar disorder and controls. Transl Psychiatry 2020;10:58.
321. Ahola-Olli AV, Wurtz P, Havulinna AS, et al. Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. Am J Hum Genet 2017;100:40-50.
322. Cavaillès V, Dauvois S, L'Horset F, et al. Nuclear factor RIP140 modulates transcriptional activation by the estrogen receptor. The EMBO Journal 1995;14:3741-51.
323. Davis AP, Grondin CJ, Johnson RJ, et al. The Comparative Toxicogenomics Database: update 2019. Nucleic Acids Res 2019;47:D948-D54.
324. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol 2007;1:54.
325. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559.
326. Song L, Langfelder P, Horvath S. Comparison of co-expression measures: mutual information, correlation, and model based indices. BMC Bioinformatics 2012;13:328.
327. Belinky F, Nativ N, Stelzer G, et al. PathCards: multi-source consolidation of human biological pathways. Database (Oxford) 2015;2015.
328. Dassen H, Punyadeera C, Kamps R, et al. Progesterone regulation of implantation-related genes: new insights into the role of oestrogen. Cell Mol Life Sci 2007;64:1009-32.
329. Meola J, Rosa e Silva JC, Dentillo DB, et al. Differentially expressed genes in eutopic and ectopic endometrium of women with endometriosis. Fertil Steril 2010;93:1750-73.
330. Paiva P, Lockhart MG, Girling JE, et al. Identification of genes differentially expressed in menstrual breakdown and repair. Mol Hum Reprod 2016;22:898-912.

331. Cox TC, Sadlon TJ, Schwarz QP, et al. The major splice variant of human 5-aminolevulinate synthase-2 contributes significantly to erythroid heme biosynthesis. The International Journal of Biochemistry & Cell Biology 2004;36:281-95.
332. Shin OS, Uddin T, Citorik R, et al. LPLUNC1 modulates innate immune responses to Vibrio cholerae. J Infect Dis 2011;204:1349-57.
333. Wolins NE, Skinner JR, Schoenfish MJ, Tzekov A, Bensch KG, Bickel PE. Adipocyte protein S3-12 coats nascent lipid droplets. J Biol Chem 2003;278:37713-21.
334. Grahn TH, Zhang Y, Lee MJ, et al. FSP27 and PLIN1 interaction promotes the formation of large lipid droplets in human adipocytes. Biochem Biophys Res Commun 2013;432:296-301.
335. Menaya J, González-Manchón C, Parrilla R, Ayuso MS. Molecular cloning, sequencing and expression of a cDNA encoding a human liver NAD-dependent α-glycerol-3-phosphate dehydrogenase. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1995;1262:91-4.
336. Kraemer FB, Khor VK, Shen WJ, Azhar S. Cholesterol ester droplets and steroidogenesis. Mol Cell Endocrinol 2013;371:15-9.
337. Perez-Vilar J, Randell SH, Boucher RC. C-Mannosylation of MUC5AC and MUC5B Cys subdomains. Glycobiology 2004;14:325-37.
338. Han Y, Jia Q, Jahani PS, et al. Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. Nat Commun 2020;11:1776.
339. Gipson IK, Moccia R, Spurr-Michaud S, et al. The Amount of MUC5B Mucin in Cervical Mucus Peaks at Midcycle1. The Journal of Clinical Endocrinology & Metabolism 2001;86:594-600.
340. Alameda F, Mejias-Luque R, Garrido M, de Bolos C. Mucin genes (MUC2, MUC4, MUC5AC, and MUC6) detection in normal and pathological endometrial tissues. Int J Gynecol Pathol 2007;26:61-5.
341. Horne AW, White JO, Lalani E-N. Adhesion molecules and the normal endometrium. BJOG: An International Journal of Obstetrics & Gynaecology 2002;109:610-7.
342. Moritani S, Kushima R, Ichihara S, et al. Eosinophilic cell change of the endometrium: a possible relationship to mucinous differentiation. Mod Pathol 2005;18:1243-8.
343. Hebbar V, Damera G, Sachdev GP. Differential expression of MUC genes in endometrial and cervical tissues and tumors. BMC Cancer 2005;5:124.
344. Catalano RD, Wilson MR, Boddy SC, Jabbour HN. Comprehensive expression analysis of prostanoid enzymes and receptors in the human endometrium across the menstrual cycle. Mol Hum Reprod 2011;17:182-92.
345. Dennis EA, Norris PC. Eicosanoid storm in infection and inflammation. Nat Rev Immunol 2015;15:511-23.
346. Nakayama T, Soma M, Watanabe Y, et al. Splicing mutation of the prostacyclin synthase gene in a family associated with hypertension. Biochemical and Biophysical Research Communications 2002;297:1135-9.
347. Monsivais D, Bray JD, Su E, et al. Activated glucocorticoid and eicosanoid pathways in endometriosis. Fertil Steril 2012;98:117-25.
348. Krjutskov K, Katayama S, Saare M, et al. Single-cell transcriptome analysis of endometrial tissue. Hum Reprod 2016;31:844-53.
349. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol 2019;20:590-607.

350. Borghese B, Barbaux S, Mondon F, et al. Research resource: genome-wide profiling of methylated promoters in endometriosis reveals a subtelomeric location of hypermethylation. Mol Endocrinol 2010;24:1872-85.
351. 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;28:1118-35.
352. Fambrini M, Sorbi F, Bussani C, Cioni R, Sisti G, Andersson KL. Hypermethylation of HOXA10 gene in mid-luteal endometrium from women with ovarian endometriomas. Acta Obstet Gynecol Scand 2013;92:1331-4.
353. Bulun SE, Monsivais D, Kakinuma T, et al. Molecular Biology of Endometriosis: From Aromatase to Genomic Abnormalities. Semin Reprod Med 2015;33:220-4.
354. Barjaste N, Shahhoseini M, Afsharian P, Sharifi-Zarchi A, Masoudi-Nejad A. Genome-wide DNA methylation profiling in ectopic and eutopic of endometrial tissues. J Assist Reprod Genet 2019;36:1743-52.
355. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature 2015;526:68-74.
356. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30:1363-9.
357. Triche TJ, Jr., Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res 2013;41:e90.
358. Fortin JP, Fertig E, Hansen K. shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R. F1000Res 2014;3:175.
359. Min J, Hemani G, Davey Smith G, Relton CL, Suderman M. Meffil: efficient normalisation and analysis of very large DNA methylation samples. 2017.
360. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012;28:882-3.
361. Leek JT, Storey JD. Capturing Heterogeneity in Gene Expression Studies by "Surrogate Variable Analysis". PLoS Genetics 2005;preprint.
362. Mansell G, Gorrie-Stone TJ, Bao Y, et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. BMC Genomics 2019;20:366.
363. Teschendorff AE, Zhuang J, Widschwendter M. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. Bioinformatics 2011;27:1496-505.
364. Moghadam BT, Zamani N, Komorowski J, Grabherr M. PiiL: visualization of DNA methylation and gene expression data in gene pathways. BMC Genomics 2017;18:571.
365. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.
366. Anastasiadi D, Esteve-Codina A, Piferrer F. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. Epigenetics Chromatin 2018;11:37.
367. Ball MP, Li JB, Gao Y, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol 2009;27:361-8.
368. Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell 2014;26:577-90.

369. Kukushkina V, Modhukur V, Suhorutsenko M, et al. DNA methylation changes in endometrium and correlation with gene expression during the transition from pre-receptive to receptive phase. Sci Rep 2017;7:3916.
370. Lister R, Pelizzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009;462:315-22.
371. Santra MK, Wajapeyee N, Green MR. F-box protein FBXO31 mediates cyclin D1 degradation to induce G1 arrest after DNA damage. Nature 2009;459:722-5.
372. Alvarez-Medina R, Le Dreau G, Ros M, Marti E. Hedgehog activation is required upstream of Wnt signalling to control neural progenitor proliferation. Development 2009;136:3301-9.
373. Matsuzaki S, Botchorishvili R, Pouly JL, Canis M. Targeting the Wnt/β-catenin pathway in endometriosis: a potentially effective approach for treatment and prevention. Molecular and Cellular Therapies 2014;2:36.
374. Balch C, Matei DE, Huang THM, Nephew KP. Role of epigenomics in ovarian and endometrial cancers. Epigenomics 2010;2:419-47.
375. Moreno-Bueno G, Rodriguez-Perales S, Sanchez-Estevez C, et al. Cyclin D1 gene (CCND1) mutations in endometrial cancer. Oncogene 2003;22:6115-8.
376. Velarde MC, Aghajanova L, Nezhat CR, Giudice LC. Increased mitogen-activated protein kinase kinase/extracellularly regulated kinase activity in human endometrial stromal fibroblasts of women with endometriosis reduces 3',5'-cyclic adenosine 5'-monophosphate inhibition of cyclin D1. Endocrinology 2009;150:4701-12.
377. de Souza-Pinto NC, Maynard S, Hashiguchi K, Hu J, Muftuoglu M, Bohr VA. The recombination protein RAD52 cooperates with the excision repair protein OGG1 for the repair of oxidative lesions in mammalian cells. Mol Cell Biol 2009;29:4441-54.
378. Ma J, Setton J, Lee NY, Riaz N, Powell SN. The therapeutic significance of mutational signatures from DNA repair deficiency in cancer. Nat Commun 2018;9:3292.
379. Matsuzaki S, Canis M, Vaurs-Barriere C, Boespflug-Tanguy O, Dastugue B, Mage G. DNA microarray analysis of gene expression in eutopic endometrium from patients with deep endometriosis using laser capture microdissection. Fertil Steril 2005;84 Suppl 2:1180-90.
380. Shi Y, Ghosh MC, Tong WH, Rouault TA. Human ISD11 is essential for both iron-sulfur cluster assembly and maintenance of normal cellular iron homeostasis. Hum Mol Genet 2009;18:3014-25.
381. Lim SC, Friemel M, Marum JE, et al. Mutations in LYRM4, encoding iron-sulfur cluster biogenesis factor ISD11, cause deficiency of multiple respiratory chain complexes. Hum Mol Genet 2013;22:4460-73.
382. Rahman F, Mahmud P, Karim R, Hossain T, Islam F. Determination of novel biomarkers and pathways shared by colorectal cancer and endometrial cancer via comprehensive bioinformatics analysis. Informatics in Medicine Unlocked 2020;20.
383. Olson JK, Miller SD. Microglia initiate central nervous system innate and adaptive immune responses through multiple TLRs. J Immunol 2004;173:3916-24.
384. Brightbill HD, Libraty DH, Krutzik SR, et al. Host Defense Mechanisms Triggered by Microbial Lipoproteins Through Toll-Like Receptors. Science 1999;285:732.
385. Jorgenson RL, Young SL, Lesmeister MJ, Lyddon TD, Misfeldt ML. Human endometrial epithelial cells cyclically express Toll-like receptor 3 (TLR3) and exhibit TLR3-dependent responses to dsRNA. Hum Immunol 2005;66:469-82.
386. Allhorn S, Boing C, Koch AA, Kimmig R, Gashaw I. TLR3 and TLR4 expression in healthy and diseased human endometrium. Reprod Biol Endocrinol 2008;6:40.
387. Aflatoonian R, Tuckerman E, Elliott SL, et al. Menstrual cycle-dependent changes of Toll-like receptors in endometrium. Hum Reprod 2007;22:586-93.

388. Eriksson M, Meadows SK, Basu S, Mselle TF, Wira CR, Sentman CL. TLRs mediate IFN-gamma production by human uterine NK cells in endometrium. J Immunol 2006;176:6219-24.
389. Lesmeister MJ, Jorgenson RL, Young SL, Misfeldt ML. 17Beta-estradiol suppresses TLR3-induced cytokine and chemokine production in endometrial epithelial cells. Reprod Biol Endocrinol 2005;3:74.
390. Khan KN, Kitajima M, Fujishita A, Nakashima M, Masuzaki H. Toll-like receptor system and endometriosis. J Obstet Gynaecol Res 2013;39:1281-92.
391. Jiang C, Liu C, Guo J, et al. The Expression of Toll-like receptors in eutopic and ectopic endometrium and its implication in the inflammatory pathogenesis of adenomyosis. Sci Rep 2017;7:7365.
392. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019;47:D607-D13.
393. Vigano P, Lattuada D, Mangioni S, et al. Cycling and early pregnant endometrium as a site of regulated expression of the vitamin D system. J Mol Endocrinol 2006;36:415-24.
394. Christakos S, Dhawan P, Liu Y, Peng X, Porta A. New insights into the mechanisms of vitamin D action. Journal of Cellular Biochemistry 2003;88:695-705.
395. Leitner L, Jurets A, Itariu BK, et al. Osteopontin promotes aromatase expression and estradiol production in human adipocytes. Breast Cancer Res Treat 2015;154:63-9.
396. Johnson GA, Burghardt RC, Bazer FW, Spencer TE. Osteopontin: Roles in Implantation and Placentation1. Biology of Reproduction 2003;69:1458-71.
397. Broi MGD, Rocha CVJ, Meola J, et al. Expression of PGR, HBEGF, ITGAV, ITGB3 and SPP1 genes in eutopic endometrium of infertile women with endometriosis during the implantation window: a pilot study. JBRA Assist Reprod 2017;21:196-202.
398. Woodward OM, Köttgen A, Coresh J, Boerwinkle E, Guggino WB, Köttgen M. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. Proceedings of the National Academy of Sciences 2009;106:10338.
399. Desuzinges-Mandon E, Arnaud O, Martinez L, Huche F, Di Pietro A, Falson P. ABCG2 transports and transfers heme to albumin through its large extracellular loop. J Biol Chem 2010;285:33123-33.
400. Suzuki M, Suzuki H, Sugimoto Y, Sugiyama Y. ABCG2 transports sulfated conjugates of steroids and xenobiotics. J Biol Chem 2003;278:22644-9.
401. Taylor NMI, Manolaridis I, Jackson SM, Kowal J, Stahlberg H, Locher KP. Structure of the human multidrug transporter ABCG2. Nature 2017;546:504-9.
402. Turner JG, Gump JL, Zhang C, et al. ABCG2 expression, function, and promoter methylation in human multiple myeloma. Blood 2006;108:3881-9.
403. Yaguchi T, Onishi T. Estrogen induces cell proliferation by promoting ABCG2-mediated efflux in endometrial cancer cells. Biochem Biophys Rep 2018;16:74-8.
404. Masuda H, Maruyama T, Gargett CE, et al. Endometrial side population cells: potential adult stem/progenitor cells in endometrium. Biol Reprod 2015;93:84.
405. Masuda H, Matsuzaki Y, Hiratsu E, et al. Stem cell-like properties of the endometrial side population: implication in endometrial regeneration. PLoS One 2010;5:e10387.
406. Matsuzaki S, Darcha C. Adenosine triphosphate-binding cassette transporter G2 expression in endometriosis and in endometrium from patients with and without endometriosis. Fertil Steril 2012;98:1512-20 e3.
407. Liu S, Zhou J, Wen J. Expression and significance of CD133 and ABCG2 in endometriosis. Clinical and experimental obstetrics & gynecology 2015;42:771-5.

408. Kogure A, Shiratori I, Wang J, Lanier LL, Arase H. PANP is a novel O-glycosylated PILRalpha ligand expressed in neural tissues. Biochem Biophys Res Commun 2011;405:428-33.
409. Gong Z, Cho YW, Kim JE, Ge K, Chen J. Accumulation of Pax2 transactivation domain interaction protein (PTIP) at sites of DNA breaks via RNF8-dependent pathway is required for cell survival after DNA damage. J Biol Chem 2009;284:7284-93.
410. Zhang Z, Sun Y, Cho YW, Chow CC, Simons SS, Jr. PA1 protein, a new competitive decelerator acting at more than one step to impede glucocorticoid receptor-mediated transactivation. J Biol Chem 2013;288:42-58.
411. Liang J, Zhang H, Zhang Y, Zhang Y, Shang Y. GAS, a new glutamate-rich protein, interacts differentially with SRCs and is involved in oestrogen receptor function. EMBO Rep 2009;10:51-7.
412. Amzallag N, Passer BJ, Allanic D, et al. TSAP6 facilitates the secretion of translationally controlled tumor protein/histamine-releasing factor via a nonclassical pathway. J Biol Chem 2004;279:46104-12.
413. Joshi-Tope G, Gillespie M, Vastrik I, et al. Reactome: a knowledgebase of biological pathways. Nucleic Acids Res 2005;33:D428-32.
414. Mori M, Ito F, Shi L, et al. Ovarian endometriosis-associated stromal cells reveal persistently high affinity for iron. Redox Biol 2015;6:578-86.
415. UniProt C. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 2019;47:D506-D15.
416. Pelch KE, Schroder AL, Kimball PA, Sharpe-Timms KL, Davis JW, Nagel SC. Aberrant gene expression profile in a mouse model of endometriosis mirrors that observed in women. Fertil Steril 2010;93:1615-27 e18.
417. Kosmidis S, Polyzos A, Harvey L, et al. RbAp48 Protein Is a Critical Component of GPR158/OCN Signaling and Ameliorates Age-Related Memory Loss. Cell Rep 2018;25:959-73 e6.
418. Watkins LR, Orlandi C. Orphan G Protein Coupled Receptors in Affective Disorders. Genes (Basel) 2020;11.
419. Sutton LP, Orlandi C, Song C, et al. Orphan receptor GPR158 controls stress-induced depression. Elife 2018;7.
420. Li T, Mamillapalli R, Ding S, et al. Endometriosis alters brain electrophysiology, gene expression and increases pain sensitization, anxiety, and depression in female mice. Biol Reprod 2018;99:349-59.
421. Engqvist H, Parris TZ, Kovacs A, et al. Validation of Novel Prognostic Biomarkers for Early-Stage Clear-Cell, Endometrioid and Mucinous Ovarian Carcinomas Using Immunohistochemistry. Front Oncol 2020;10:162.
422. Engqvist H, Parris TZ, Kovacs A, et al. Immunohistochemical validation of COL3A1, GPR158 and PITHD1 as prognostic biomarkers in early-stage ovarian carcinomas. BMC Cancer 2019;19:928.
423. Gaudet P, Livstone MS, Lewis SE, Thomas PD. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. Brief Bioinform 2011;12:449-62.
424. de Almeida BP, Apolonio JD, Binnie A, Castelo-Branco P. Roadmap of DNA methylation in breast cancer identifies novel prognostic biomarkers. BMC Cancer 2019;19:219.
425. Han X, Ong JS, An J, Hewitt AW, Gharahkhani P, MacGregor S. Using Mendelian randomization to evaluate the causal relationship between serum C-reactive protein levels and age-related macular degeneration. Eur J Epidemiol 2020;35:139-46.

426. Sun PR, Jia SZ, Lin H, Leng JH, Lang JH. Genome-wide profiling of long noncoding ribonucleic acid expression patterns in ovarian endometriosis by microarray. Fertil Steril 2014;101:1038-46 e7.
427. Akiyama M, Takeda S, Kokame K, Takagi J, Miyata T. Crystal structures of the noncatalytic domains of ADAMTS13 reveal multiple discontinuous exosites for von Willebrand factor. Proceedings of the National Academy of Sciences 2009;106:19274.
428. Luo SF, Chang CC, Lee IT, et al. Activation of ROS/NF-kappaB and Ca2+/CaM kinase II are necessary for VCAM-1 induction in IL-1beta-treated human tracheal smooth muscle cells. Toxicol Appl Pharmacol 2009;237:8-21.
429. Tabuchi A, Yoshioka A, Higashi T, et al. Direct demonstration of involvement of protein kinase Calpha in the Ca2+-induced platelet aggregation. J Biol Chem 2003;278:26374-9.
430. Gschwend JE, Fair WR, Powell CT. Bryostatin 1 Induces Prolonged Activation of Extracellular Regulated Protein Kinases in and Apoptosis of LNCaP Human Prostate Cancer Cells Overexpressing Protein Kinase Cα. Molecular Pharmacology 2000;57:1224.
431. Shen L, Dean NM, Glazer RI. Induction of p53-Dependent, Insulin-Like Growth Factor-Binding Protein-3-Mediated Apoptosis in Glioblastoma Multiforme Cells by a Protein Kinase Cα Antisense Oligonucleotide. Molecular Pharmacology 1999;55:396.
432. Evensen NA, Kuscu C, Nguyen HL, et al. Unraveling the role of KIAA1199, a novel endoplasmic reticulum protein, in cancer cell migration. J Natl Cancer Inst 2013;105:1402-16.
433. Wang A, Nomura M, Patan S, Ware JA. Inhibition of protein kinase Calpha prevents endothelial cell migration and vascular tube formation in vitro and myocardial neovascularization in vivo. Circ Res 2002;90:609-16.
434. Maffucci T, Raimondi C, Abu-Hayyeh S, et al. A phosphoinositide 3-kinase/phospholipase Cgamma1 pathway regulates fibroblast growth factor-induced capillary tube formation. PLoS One 2009;4:e8285.
435. Besson A, Yong VW. Involvement of p21^{Waf1/Cip1} in Protein Kinase C Alpha-Induced Cell Cycle Progression. Molecular and Cellular Biology 2000;20:4580.
436. Fournier DB, Chisamore M, Lurain JR, Rademaker AW, Jordan VC, Tonetti DA. Protein kinase C alpha expression is inversely related to ER status in endometrial carcinoma: possible role in AP-1-mediated proliferation of ER-negative endometrial cancer. Gynecol Oncol 2001;81:366-72.
437. van Eijk KR, de Jong S, Boks MPM, et al. Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. BMC genomics 2012;13:636-.
438. Bell JT, Pai AA, Pickrell JK, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol 2011;12:R10.
439. McRae AF, Marioni RE, Shah S, et al. Identification of 55,000 Replicated DNA Methylation QTL. Sci Rep 2018;8:17605.
440. Cardona A, Day FR, Perry JRB, et al. Epigenome-Wide Association Study of Incident Type 2 Diabetes in a British Population: EPIC-Norfolk Study. Diabetes 2019;68:2315-26.
441. Dong D, Tian Y, Zheng SC, Teschendorff AE. ebGSEA: an improved Gene Set Enrichment Analysis method for Epigenome-Wide-Association Studies. Bioinformatics 2019;35:3514-6.
442. Luo X, Yang C, Wei Y. Detection of cell-type-specific risk-CpG sites in epigenome-wide association studies. Nat Commun 2019;10:3113.

443. Rahmani E, Schweiger R, Rhead B, et al. Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology. Nat Commun 2019;10:3417.