# Global alterations of the epigenetic landscape define endometrial receptivity and provide a hallmark for endometriosis


Natasha Poppy Morgan

The Babraham Institute

Jesus College

University of Cambridge


This dissertation is submitted for the degree of Doctor of Philosophy

September 2019

# Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee

Natasha Poppy Morgan

# Global alterations of the epigenetic landscape define endometrial receptivity and provide a hallmark for endometriosis

Natasha Poppy Morgan

The endometrium plays an essential role in the female reproductive cycle. It undergoes extensive morphological and transcriptional changes during the menstrual cycle, in response to steroid hormones. For implantation to occur successfully, the endometrium must attain a receptive state which is synchronised with the presence of a competent blastocyst. Failure in the establishment or timing of receptivity is a primary cause of infertility. The endometrium is also the tissue chiefly involved in the aetiology of endometriosis, a condition in which endometrial tissue is found outside the womb where it forms lesions. Endometriosis affects ~10% of women of reproductive age, can cause severe pain and poses an increased risk of infertility. Overall, the molecular regulation required to fine-tune the cyclical changes of the endometrium and that may also play a role in the development of endometriosis remains poorly understood.

In this thesis, I set out to investigate how the epigenome may contribute to achieving gene expression changes that underpin the attainment of the window of receptivity as well as the disease state of endometriosis. To this end, in conjunction with histological analyses, I generated a comprehensive set of genome-wide profiles, including transcriptomes by RNA-seq, histone modification profiles by ChIP-seq for H3K4me3, H3K9me3 and H3K27me3, as well as DNA methylation (5mC) and hydroxymethylation (5hmC) profiles by meDIP-seq and hmeDIP-seq, respectively, of endometrial biopsies from control and endometriosis patients.

My data reveal considerable changes to the epigenome that occur during the transition from the non-receptive to the receptive state. I identify a cohort of regions marked by H3K4me3 and H3K27me3, known as bivalent domains, that resolve to H3K4me3-only in the receptive state and globally correlate with increased expression of the associated genes. Notably, I show that this subset of bivalent domains that will become activated can be predicted already in the non-receptive state based on marking with 5hmC. These regions are also significantly enriched for Kruppel-like factor 9 (KLF9) and Forkhead box O1A (FOXO1) binding motifs, which are progesterone-regulated transcription factors with essential roles in reproduction.

Perhaps the most dramatic epigenetic change is observed in endometrial epithelial cells that acquire high levels of 5hmC during the establishment of the receptive state. This increase of 5hmC correlates with increased gene expression, in particular at genes that are crucial for the attainment of receptivity. These data are also in line with my finding that the majority of gene expression changes observed in the whole tissue can be attributed to epithelial cells. Thus, 5hmC appears to play a key role in the molecular functionality of epithelial cells, which have a high secretory output and are critical for implantation to occur.

Finally, I report on alterations to the epigenetic profile of endometrium from patients with endometriosis, a noteworthy discovery. Differential 5mC and 5hmC enrichment is particularly prevalent at genes involved in the Hippo signalling pathway, which regulates core aspects of cell and tissue biology including cell polarity, adhesion, proliferation and apoptosis. These epigenetic changes may potentially predispose endometrial cells of endometriosis patients to exhibit increased levels of adherence and proliferation at ectopic sites, thus contributing to the formation of lesions. In addition, and perhaps more importantly, I identified and validated a set of genomic regions based on their differential DNA (hydroxyl)methylation state that separates patient samples according to their receptivity and disease status. Although these epigenetic probes require independent validation on many more independent patient samples, identification of such a cohort of loci holds great promise for the establishment of an epigenetic biomarker capable of determining receptivity state and as a diagnostic tool for endometriosis. If proven successful, this would be a great advance in the field of reproductive biology and in particular female infertility. My data prompt further investigations into the contribution of the epigenome to the transformation of the endometrium during the normal menstrual cycle and to the aetiology of endometriosis.

# Acknowledgements

This PhD would not have been possible without all of the people mentioned below. I would like to offer my sincere thanks to you all.

Firstly, thank you to Dr. Myriam Hemberger for initially giving me this fantastic opportunity to undertake my PhD within her lab. You have been a wonderful mentor and provided great scientific and personal support throughout the whole PhD and more recently with reading my thesis. You have always been on hand to show me techniques and when I would pop in for a quick chat you were there to offer your opinion or advice and for that I am very grateful and thankful for.

The Hemberger Lab has provided an exceptional space to develop as a scientist. All the members have provided experimental support and scientific guidance and I will forever be in their debt. In particular, I would like to mention Claire and Sarah for all your scientific input, the words of wisdom, advice and insight that you provided throughout the project. A special thank you to Laura, Claire, Sarah, Georgia and Elena for all your time and support not only professionally but personally as well. You were always there to have a chat, a laugh and importantly a cuppa with. You splashed bright colours and sparkles into the whole experience, so our lab was always full of fun and smiles.

Thank you to Cambridge Epigenetixs, in particular Jo, Jens, Kishen and Rita, who kindly taught me important techniques, provided excellent guidance and generally were a great support throughout my time at CEGX and beyond.

To Jesus College: I want to thank Amanda and Joe for their recommendation; it really was the right college for me. I will always be thankful for meeting Chris and Rachel there, and popping over for a cup of tea, some wonderful cake and a catch up. I miss you guys.

To my friends, my wonderful, wonderful friends. In alphabetical order Adam, Amanda, Chris, Jayeta and Laura. You have been there for all the ups and downs, and provided motivational pep talks to drag me back up when required. I will never forget the endless conversations, the board game nights (got to love all those sheep), as well as your generous friendship. I know that wherever in the world we are, we will always be the very best of friends.

To my magnificent family that stretches the world, I am so grateful to have you all as my family. You have provided encouragement and support from the very start, shaping the person I am today. Mum and Wayne, thank you for instilling in me the belief that I could accomplish this feat and the support and guidance that you have provided throughout my life, as well as my career.

To Tom, you truly are amazing. What would I have done without you. Thank you for being my rock through thick and thin; you have always supported me with your love, affection and cheekiness. The last four years have been a fantastic journey together with so many landmarks on the way. I am so thankful I could share it with you. You have looked after my heart and soul throughout the whole experience. You have been there through it all - the lows and the highs. You are the tether to my kite, keeping me grounded and the wind to lift me up. I can't wait for our next adventures together…

# Assistance to Acknowledge

**Initial training in techniques, laboratory practice and subsequent mentoring:**

Myriam Hemberger (Supervisor), Claire Senner (Mentor), Joanne Mason (Mentor), Jens Fullgrabe (Mentor), Laura Woods (Colleague), Kishen Chahwala (Colleague), Anne Corcoran (Assessor), Andrew Sharkey (Assessor)

**Data obtained from a technical service provider:**

Next-generation sequencing was carried out by the Babraham Next Generation Sequencing Facility department (The Babraham Institute) or at Cambridge Epigenetix, Cambridge

Next-generation sequencing data was processed and aligned by the Babraham Bioinformatics department (The Babraham Institute) or at Cambridge Epigenetix, Cambridge

Advice on statistical analysis was provided by Dr. Anne Segonds-Pichon in the Babraham Bioinformatics department (The Babraham Institute)

CellProfiler training and protocols were provided by the Babraham Imaging department (The Babraham Institute)

**Data produced in collaboration:**

Liquid chromatography- mass spectrometry was conducted in collaboration with Prof. Petra. Hajkova's group at Imperial Collage London

Epithelial and Stromal cell organoids were provided by collaborator Prof. Jan Brosens at the University of Warwick, Coventry.

**Data provided or analysed by someone else:**

ChromstaR analysis of Histone-ChIP data was conducted by Dr. Sarah Burge

TMOT analysis was conducted by Dr. Sarah Burge

I would also like to acknowledge and thank the women that donated tissue which enabled this research to be conducted.

# Summary

The endometrium plays an essential role in the female reproductive cycle. It undergoes extensive morphological and transcriptional changes during the menstrual cycle, in response to steroid hormones. For implantation to occur successfully, the endometrium must attain a receptive state which is synchronised with the presence of a competent blastocyst. Failure in the establishment or timing of receptivity is a primary cause of infertility. The endometrium is also the tissue chiefly involved in the aetiology of endometriosis, a condition in which endometrial tissue is found outside the womb where it forms lesions. Endometriosis affects ~10% of women of reproductive age, can cause severe pain and poses an increased risk of infertility. Overall, the molecular regulation required to fine-tune the cyclical changes of the endometrium and that may also play a role in the development of endometriosis remains poorly understood.

In this thesis, I set out to investigate how the epigenome may contribute to achieving gene expression changes that underpin the attainment of the window of receptivity as well as the disease state of endometriosis. To this end, in conjunction with histological analyses, I generated a comprehensive set of genome-wide profiles, including transcriptomes by RNA-seq, histone modification profiles by ChIP-seq for H3K4me3, H3K9me3 and H3K27me3, as well as DNA methylation (5mC) and hydroxymethylation (5hmC) profiles by meDIP-seq and hmeDIP-seq, respectively, of endometrial biopsies from control and endometriosis patients.

My data reveal considerable changes to the epigenome that occur during the transition from the non-receptive to the receptive state. I identify a cohort of regions marked by H3K4me3 and H3K27me3, known as bivalent domains, that resolve to H3K4me3-only in the receptive state and globally correlate with increased expression of the associated genes. Notably, I show that this subset of bivalent domains that will become activated can be predicted already in the non-receptive state based on marking with 5hmC. These regions are also significantly enriched for Kruppel-like factor 9 (KLF9) and Forkhead box O1A (FOXO1) binding motifs, which are progesterone-regulated transcription factors with essential roles in reproduction.

Perhaps the most dramatic epigenetic change is observed in endometrial epithelial cells that acquire high levels of 5hmC during the establishment of the receptive state. This

increase of 5hmC correlates with increased gene expression, in particular at genes that are crucial for the attainment of receptivity. These data are also in line with my finding that the majority of gene expression changes observed in the whole tissue can be attributed to epithelial cells. Thus, 5hmC appears to play a key role in the molecular functionality of epithelial cells, which have a high secretory output and are critical for implantation to occur.

Finally, I report on alterations to the epigenetic profile of endometrium from patients with endometriosis, a noteworthy discovery. Differential 5mC and 5hmC enrichment is particularly prevalent at genes involved in the Hippo signalling pathway, which regulates core aspects of cell and tissue biology including cell polarity, adhesion, proliferation and apoptosis. These epigenetic changes may potentially predispose endometrial cells of endometriosis patients to exhibit increased levels of adherence and proliferation at ectopic sites, thus contributing to the formation of lesions. In addition, and perhaps more importantly, I identified and validated a set of genomic regions based on their differential DNA (hydroxyl)methylation state that separates patient samples according to their receptivity and disease status. Although these epigenetic probes require independent validation on many more independent patient samples, identification of such a cohort of loci holds great promise for the establishment of an epigenetic biomarker capable of determining receptivity state and as a diagnostic tool for endometriosis. If proven successful, this would be a great advance in the field of reproductive biology and in particular female infertility. My data prompt further investigations into the contribution of the epigenome to the transformation of the endometrium during the normal menstrual cycle and to the aetiology of endometriosis.

# Abbreviations

| | |
|---|---|
| 5hmC | 5-hydroxymethlycytosine |
| 5mC | 5-methylcytosine |
| bp | base pair |
| CGI | CpG Island |
| ChIP-seq | Chromatin Immunoprecipitation sequencing |
| DAPI | 4',6-diamidino-2-phenylindole |
| DhMR | Differentially hydroxymethylated region |
| DMR | Differentially methylated region |
| DNMT | DNA methyltransferase |
| E | Embryonic Day |
| ER | Estrogen Receptor |
| ERA | Endometrial Receptivity Array |
| ERT | Endometrial Receptivity Transcriptome |
| GO | Gene Ontology |
| GWAS | Genome Wide Association Studies |
| H3K4me3 | Trimethylation of Lysine 4 Histone 3 |
| H3K9me3 | Trimethylation of Lysine 9 Histone 3 |
| H3K27me3 | Trimethylation of Lysine 27 Histone 3 |
| HMCP | Hydroxymethylation pull-down sequencing |
| hmeDIP-seq | Hydroxymethylated DNA immunoprecipitation sequencing |
| IF | Immunofluorescence |
| IHC | Immunohistochemistry |
| IVF | In vitro Fertilisation |
| LCM | Laser Capture Microdissection |
| LH | Luteinizing Hormone |
| meDIP-seq | Methylated DNA immunoprecipitation sequencing |
| NR | Non-Receptive |

| | |
|---|---|
| NRC | Non-Receptive Control |
| NRE | Non-Receptive Endometriosis |
| OE | Over-Expressing |
| PCA | Principle component analysis |
| PGR | Progesterone receptor |
| R | Receptive |
| RC | Receptive Control |
| RE | Receptive Endometriosis |
| RNA-seq | High-throughput mRNA sequencing |
| RT-qPCR | Quantitative Reverse Transcription PCR |
| TET | Ten Eleven Translocation |
| TMOT | TrueMethyl® On Target |
| TSS | Transcription Start Site |
| uNK | uterine Natural Killer cells |
| WOI | Window of Implantation |

# Contents

**Acknowledgements** ..... **i**

**Assistance to Acknowledge** ..... **iii**

**Summary** ..... **v**

**Abbreviations** ..... **vii**

**Chapter I** ..... **1**

**Introduction** ..... **1**

**1.1 The Endometrium** ..... **3**

1.1.1 Menstrual cycle: The natural cycle of endometrial tissue ..... 3

1.1.2 Morphological changes of the endometrium throughout the menstrual cycle ..... 5

1.1.3 Transcriptional changes associated with the menstrual cycle ..... 9

1.1.4 Model systems to study the endometrium ..... 11

**1.2 Endometriosis: The disease, diagnosis and treatments** ..... **13**

1.2.1 What is endometriosis? ..... 13

1.2.2 How is endometriosis diagnosed and classified? ..... 14

1.2.3 Treatments for endometriosis ..... 15

1.2.4 Theories of the pathogeneses of endometriosis ..... 16

**1.3 Epigenetic Modifications** ..... **19**

1.3.1 Histone modifications ..... 20

1.3.2 DNA Modifications ..... 23

1.3.3 Epigenetic modifications associated with the menstrual cycle ..... 27

1.3.4 Epigenetic modifications associated with endometriosis ..... 31

**1.4 Project Aims** ..... **36**

**Chapter II** ..... **39**

**Materials and Methods** ..... **39**

**2.1 Human Endometrial Biopsies and Organoids** ..... **41**

**2.2 DNA and RNA extraction** ..... **44**

**2.3 RNA-sequencing** ..... **44**

**2.4 Native histone chromatin immunoprecipitation and sequencing (ChIP-seq)** ..... **45**

**2.5 High-throughput sequencing of Methylated or Hydroxymethylated Immunoprecipitated DNA (meDIP-seq and hmeDIP-seq)** ..... **47**

**2.6 Cell culture and transfection** ..... **48**

**2.7 Immunofluorescence (IF) staining** ..... **49**

2.7.1 IF staining of paraffin-embedded human endometrial biopsies ..... 49

2.7.2 IF staining of cell lines ..... 50

2.7.3 IF imaging and image analysis ..... 51

**2.8 Laser Capture Microdissection (LCM)** ..... **51**

**2.9 Hydroxymethylation Capture Pulldown (HMCP)** ..... **52**

**2.10 TrueMethyl® OnTarget (TMOT)** ..... **53**

**2.11 RT-qPCR expression analysis ..... 55**

**2.12 Liquid chromatography- mass spectrometry ..... 55**

**2.13 Bioinformatic analysis ..... 56**

2.13.1 Expression profile analysis ..... 56
2.13.2 Cell-type specific gene expression signatures ..... 56
2.13.3 Gene Ontology analysis (GO) ..... 57
2.13.4 Histone ChIP-sequencing analysis ..... 57
2.13.5 Repeat element analysis ..... 58
2.13.6 Motif analysis ..... 58
2.13.7 Laser capture microdissection -hydroxymethylation pulldown analysis (LCM- HMCP) ..... 59
2.13.8 meDIP and hmeDIP analysis ..... 59
2.13.9 True Methyl® On Target (TMOT) analysis ..... 60

**Chapter III ..... 63**

**Characterising the precise receptivity state of the endometrium ..... 63**

**3.1 Introduction ..... 65**

**3.2 Genome-wide expression profiles reveal the precise receptivity stage, splitting the samples into distinct groups ..... 68**

**3.3 Virtual deconvolution of cell type composition of endometrial tissue biopsies ..... 73**

3.3.1 Determination of overall proportion of the two main cell types ..... 73
3.3.2 Epithelial cell signatures correlate with receptivity state ..... 74

**3.4 Characterisation of histone modification dynamics ..... 76**

3.4.1 Global histone patterns ..... 76
3.4.2 Identification of a loss of bivalency in receptive tissue ..... 80

**3.5 Discussion ..... 84**

**Chapter IV ..... 95**

**Hydroxymethylation is a major hallmark of receptivity in endometrial tissue ..... 95**

**4.1 Introduction ..... 97**

**4.2 Identification of cell-type specific 5hmC levels between the non-receptive and receptive state ..... 100**

**4.3 The role of 5hmC in Epithelial cells of the Endometrium ..... 104**

**4.4 Investigating mechanisms underlying the 5hmC dynamics ..... 109**

**4.5 NNMT as a potential candidate regulating 5hmC levels ..... 110**

**4.6 Discussion ..... 118**

**Chapter V ..... 127**

**Epigenetic changes associated with endometriosis ..... 127**

**5.1 Introduction ..... 129**

**5.2 Transcriptional changes associated with endometriosis ..... 132**

**5.3 Characterisation of histone modifications with respect to endometriosis ..... 134**

**5.4 DNA hydroxymethylation dynamics are altered in endometriosis ..... 141**
5.4.1 Global distribution of DNA methylation and hydroxymethylation ..... 141
5.4.2 Locus-specific changes in DNA methylation and hydroxymethylation reveal epigenetic de-regulation of the Hippo signalling pathway ..... 147
5.4.3 Validation of an epigenetic signature for endometriosis ..... 152

**5.5 Discussion ..... 159**

**Chapter VI ..... 167**

**Discussion ..... 167**

**6.1 Receptivity and its foremost importance ..... 169**
6.1.1 Expression profiles that define the immediate window of receptivity ..... 169
6.1.2 Overview of epigenetics and the endometrium ..... 170
6.1.3 The identification of bivalently marked genes in the endometrium ..... 172
6.1.4 The most dramatic changes occur in glandular and luminal epithelial cells ..... 176
6.1.5 Epithelial cells undergo dramatic changes in hydroxymethylation levels ..... 177

**6.2 The dynamic alterations of the epigenetic profile in endometriosis ..... 181**
6.2.1 Transcriptional differences associated with endometriosis ..... 181
6.2.2 Endometriosis is associated with altered epigenetic profiles in the endometrium ..... 183
6.2.3 Changes in covalent DNA modifications often affect Hippo pathway components ..... 185

**6.3 5mC and 5hmC biomarker cohorts for endometriosis and receptivity ..... 187**

**6.4 Conclusions and implications ..... 190**

**6.5 Future directions ..... 192**

**References ..... 197**

**Appendix ..... 227**

**Supplementary Figures ..... 229**

**Supplementary Tables ..... 234**

# Chapter I

# Introduction

## 1.1 The Endometrium

The endometrium is the mucous membrane that lines the uterus. It thickens during the menstrual cycle in preparation for a potential embryo to implant. If no implantation occurs then the functional endometrial layer is shed during the menstrual phase. In humans the endometrium prepares for a potential pregnancy every month, regardless if fertilisation has occurred, through the process of the menstrual cycle.

### 1.1.1 Menstrual cycle: The natural cycle of endometrial tissue

The menstrual cycle consists of three phases, the menstrual phase, the proliferative phase and the secretory phase, which are repeated every ~28 days (Figure 1.1). These phases involve dramatic morphological and physiological changes to the endometrium and are controlled by steroid hormones.

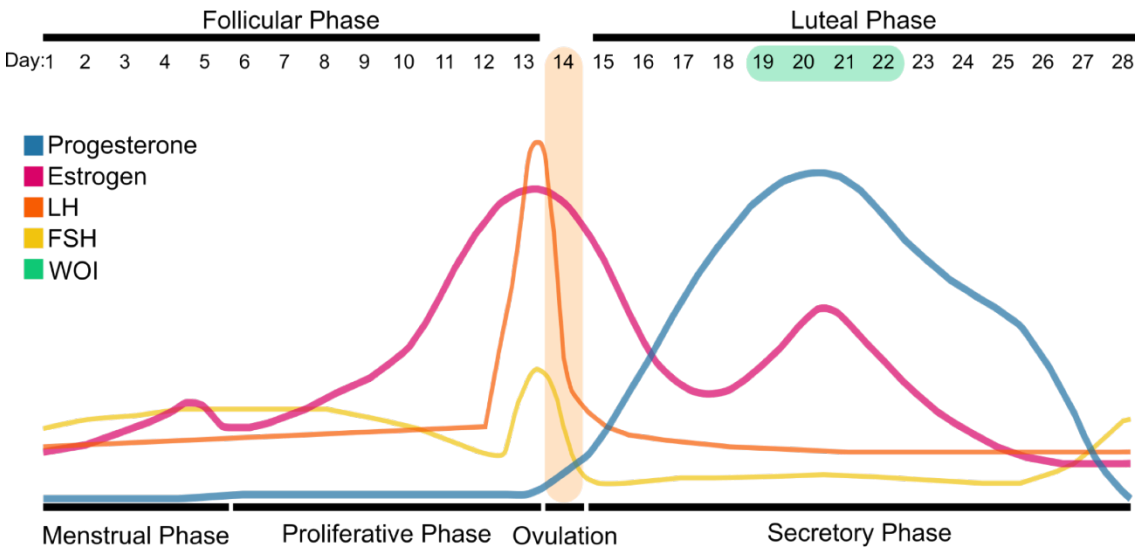



**Figure 1.1 The menstrual cycle and hormones.** A schematic of the hormone levels throughout the menstrual cycle. LH = Luteinising Hormone, FSH = Follicle Stimulating Hormone, WOI = Window of Implantation.

The menstrual and proliferative phase together comprise the follicular phase, which starts on the first day of menses until ovulation. At the end of the previous menstrual cycle the levels of progesterone and estrogen drop. This is sensed by the hypothalamus that in response produces gonadotropin-releasing hormone (GnRH). GnRH in turn stimulates the pituitary gland which synthesises and releases follicle stimulating hormone (FSH) and to a lesser extent luteinising hormone (LH) into the bloodstream. As the name suggests, FSH stimulates the growth of a cluster of ovarian follicles, these consist of the oocyst which is surrounded by granulosa and theca cells (Macklon and Fauser 2001). The production of estrogen is reliant on the collaboration of theca and granulosa cells (Raju et al. 2013). The theca cells in response to LH produce androgens which diffuse into the granulosa cells. FSH activates aromatase production in granulosa cells, an enzyme that in turn converts androgens into estrogen, resulting in a rise in estrogen levels. As the follicles grow they produce increasing amounts of estrogen and inhibin B. This results in a negative feedback loop reducing GnRH and FSH levels. As the cycle progresses one follicle out of the cluster becomes dominant and will give rise to a mature oocyte during ovulation. The levels of estrogen continue to rise and once the concentration and the duration reach a critical level, they stimulate a positive feedback loop and a surge in FSH and LH at ~day 13. The LH surge stimulates luteinisation and the release of a mature oocyte from the dominant follicle, a process known as ovulation. Ovulation occurs ~10-12 hours after the LH peak (Hoff, Quigley, and Yen 1983). The remnants of the follicle form the corpus luteum.

Following ovulation and the formation of the corpus luteum, the menstrual cycle now enters the luteal phase which is also known as the secretory phase. LH stimulates the corpus luteum to now produce high levels of progesterone which prepares the estrogen-primed endometrium for a potential implantation if fertilisation occurs. Estrogen levels rise again in the mid-secretory phase, now produced by the corpus luteum. The corpus luteum declines during the late-secretory phase and this results in a reduction of progesterone and estrogen levels, stimulating the start of the next menstrual cycle and menstruation. If pregnancy does occur the hormone - human chorionic gonadotropin (hCG) is produced by the blastocyst and this is thought to support the corpus luteum to continue with progesterone production for the first trimester when the placenta takes over the production of this pregnancy-defining hormone.

### 1.1.2 Morphological changes of the endometrium throughout the menstrual cycle

The endometrium undergoes dramatic morphological changes in response to the differing hormone levels to ensure that it is prepared for a potential implantation (Figure 1.2). The endometrium has a bilayer structure which consists of an upper functionalis layer and a basalis layer that is positioned on top of the myometrium.

The two main cell types that make up the endometrium are epithelial and stromal cells. The epithelial cells form a monolayer on the surface of the endometrium; these cells are called luminal epithelium. Where the luminal epithelium invaginates and loops into the endometrium they become glandular epithelial cells. Luminal and glandular epithelial cells have differing roles and correspondingly different expression profiles (G. E. Evans et al. 2014). Other cell types present in the endometrium are part of the immune system. The immune cells include uterine natural killer cells (uNK), dendritic cells, macrophages and T cells (J. Y. Lee, Lee, and Lee 2011). The function and characteristics of these cells develop and adjust in response to hormones as the menstrual cycle progresses. In addition, the endometrium also contains blood vessels.

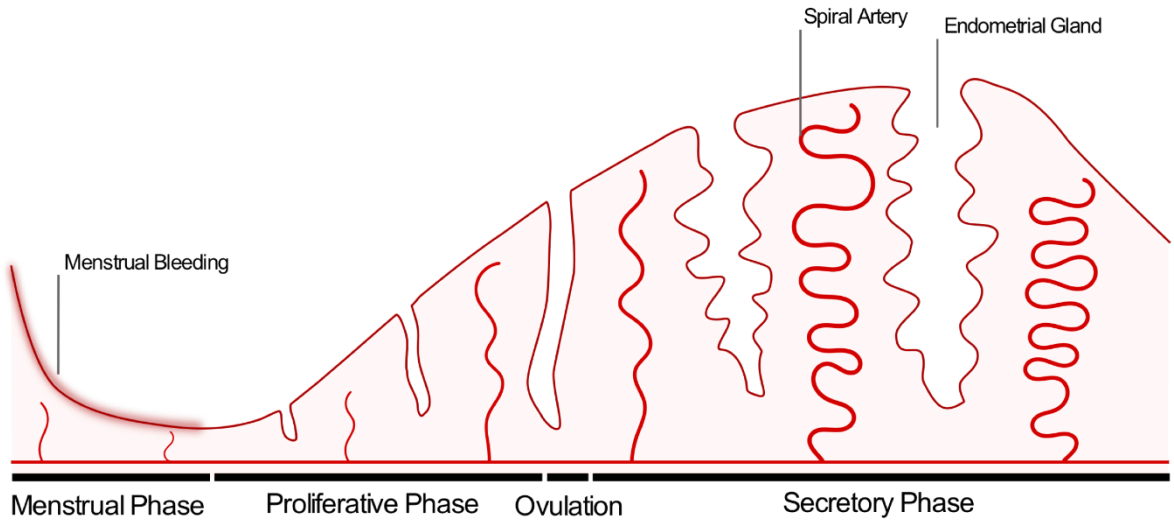



**Figure 1.2 Overview of the endometrium during the menstrual cycle.** Abstract representation of the endometrium dynamics throughout the menstrual cycle.

The menstrual phase is initiated by a reduction in progesterone and estrogen levels, which leads a sloughing of the upper functionalis layer of the endometrium. This occurs when fertilisation and subsequent implantation fails to transpire and is initiated as part of the resetting of the uterine environment. During this phase, there is a rise in

inflammatory factors and an infiltration of multiple leukocytes cells including, neutrophils, macrophages, and uNK cells just prior to menses (J. Evans and Salamonsen 2012, 2014). The shedding of the endometrium includes the rapid degradation of the extracellular matrix, reducing tissue stability; lysosomal enzymes are released and constriction of the blood vessels occurs (Cornillie et al. 1991; J. Evans and Salamonsen 2012). The loss and initial regeneration of the endometrium occurs simultaneously at different sites within the uterus (Garry et al. 2009). As one section is being shed another section is already undergoing re-epithelisation. This healing process by re-epithelisation covers the underlying basalis layer from which these cells are derived (Garry et al. 2009). This requires that the microenvironment induces changes from a pro- to an anti-inflammatory response along with the production of growth factors to support the tissue and its integrity at re-epithelising sections (J. Evans and Salamonsen 2012). The basalis layer itself remains intact during the menstrual cycle.

As the endometrium enters the proliferative phase, the basalis layer responds to increasing estrogen levels by proliferation of epithelial and stromal cells to re-form the shed functionalis layer. Several endometrial progenitor cells have been identified that may contribute to the proliferation process, including epithelial progenitor cells, mesenchymal stem cells and side-population cells (Gargett, Schwab, and Deane 2015). This results in an increase in the thickness of the endometrium from ~6 mm immediately after menstruation to ~11 mm at ovulation (Bakos, Lundkvist, and Bergh 1993; Raine-Fenning et al. 2004).

Following the proliferative phase, the growth of the endometrium plateaus as it enters the secretory phase and the endometrium prepares for a potential implantation. Humans are one of a few species that undergo preparation for implantation regardless of fertilisation having occurred (Emera, Romero, and Wagner 2012). The secretory phase undergoes substantial morphological and transcriptional changes and consequently, this phase is often further subdivided into the early, mid and late secretory phase (Noyes, Hertig, and Rock 1950; Talbi et al. 2006). In particular during the mid-secretory phase, the window of implantation (WOI) is significant as it is the period in which the endometrium is receptive and a viable blastocyst can implant. Therefore, these two events, i.e. the establishment of the WOI and the presence of a viable blastocyst, must be highly synchronised for implantation to be successful (Figure 1.1). There are several reports on the timing of the WOI with the majority of the field

agreeing that the endometrium is in its most receptive state 6-10 days post luteal hormone (LH) surge ("LH+6 to LH+10") (Acosta et al. 2000; Díaz-Gimeno et al. 2011).

*Epithelial cells*

Multiple cells contribute to the morphological changes that occur during the secretory phase. In particular glandular epithelial cells transform into highly secretory cells that produce and release a number of cytokines into the uterus that form part of the uterine fluid. The secretion of these cytokines may contribute to a dialogue with the blastocyst (Filant and Spencer 2014; Rashid et al. 2011). Mouse knockout models of a number of genes, including *Ctnnb1*, *Foxa2* and *Wnt4*, result in the adult uterus lacking endometrial glands. As a consequence, these mice show defects in blastocyst attachment as well as in stromal cell decidualisation, revealing the importance of endometrial glands for a successful pregnancy (Filant and Spencer 2014). The uterine fluid that the glandular secretions contribute to contains a number of nutrients, cytokines, and growth factors (Burton et al. 2002; Hannan et al. 2010, 2011; Scotchie et al. 2009). The role and impact that many of these cytokines have on the blastocyst remains to be established in humans. The supportive function of endometrial glands in the implantation process is, however, exemplified by the handful of molecules that have been investigated; including vascular endothelial growth factor (VEGF), leukemia inhibitory factor (LIF) as well as a maternal microRNA hsa-miR-30d (Dimitriadis et al. 2010; Hannan et al. 2011; Rashid et al. 2011; Vilella et al. 2015). VEGF is secreted into the uterus and has been shown *in vitro* to increase epithelial adhesiveness to blastocysts as well as increasing blastocyst outgrowth (Hannan et al. 2011). Hsa-miR-30d is produced in the glands and secreted via exosomes into the uterus lumen in humans. Using a mouse model, it was established that hsa-miR-30d was taken up by the embryo and resulted in an up-regulation of genes required for embryo adhesion (Vilella et al. 2015). LIF has been identified to be present in epithelial cells as well as in uterine flushings. Indeed, women that are infertile have lower levels of LIF (Dimitriadis et al. 2006, 2010). Mouse models that are LIF-deficient are infertile due to a failure to support embryo implantation, and a human *in vitro* 3D culture model re-capitulates this when LIF is neutralised (Rashid et al. 2011). LIF is suspected to have a direct interaction with human blastocysts as the mRNA for LIF-receptors has been detected in the blastocysts (Charnock-Jones et al. 1994). These examples demonstrate the vital role that epithelial glands and their secretions play in a

successful implantation, and this support continues throughout the first trimester of pregnancy by providing histiotrophic nutrition (Burton et al. 2002).

Epithelial cells not only provide histotrophic support but the luminal epithelium also forms a barrier towards the uterine lumen. During the secretory phase the luminal epithelial cells become less adherent to each other, allowing interactions with the blastocyst to occur and creating a permissive environment for the blastocyst to penetrate through the luminal epithelium and implant (Whitby, Salamonsen, and Evans 2018). Direct interaction of the blastocyst occurs with luminal epithelial cells that have pinopodes, which are only present when the endometrium is receptive (Bentin-Ley et al. 1999; Noyes, Hertig, and Rock 1950). The exact role and mechanism of pinopodes is currently unknown but this specialised interaction highlights the crucial role that luminal epithelial cells play in the preparation and facilitation of implantation.

Additionally, the number of uNK cells increases in the secretory phase; they produce cytokines while exhibiting a limited cytotoxic capacity, unlike peripheral NK cells. It has been suggested that uNK cells are involved in the remodelling of spiral arties where they are abundant and produce metalloproteinases (MMPs). In conjunction with this, uNK cells moderate the invasive capacity of extravillous trophoblast cells, an embryo-derived cell type involved in the further remodelling of uterine spiral arties after implantation (Gaynor and Colucci 2017).

*Stromal cells*

Stromal cells are highly prevalent in the endometrium and are one of the main cell types along with epithelial cells. Stromal cells undergo decidualisation during the mid to late secretory phase in response to progesterone and local cyclic adenosine monophosphate (cAMP). This is initiated in stromal cells near spiral arteries and then results in a ‘wave’ of decidualisation outwards (J. Evans et al. 2016). The appearance of stromal cells changes in the secretory phase with cells undergoing oedema (swelling) as well as the accumulation of glycogen and lipid droplets (Noyes, Hertig, and Rock 1950; Okada, Tsuzuki, and Murata 2018). They also influence the immune response through interactions with uNK cells and the remodelling of spiral arties (Okada, Tsuzuki, and

Murata 2018). Stromal cells are important in facilitating implantation through modulating trophoblast invasion.

Humans are one of the few species which undergo spontaneous decidualisation, it is thought to have evolved as part of the maternal-fetal conflict (Emera, Romero, and Wagner 2012). This concept describes the hypothesis that foetal (and notably paternally expressed) genes have evolved to extract as much nutrition as possible to ensure foetal growth and survival, whereas maternally expressed genes have evolved to ensure the success of the current foetus as well as future offspring, resulting in a tug-of-war effect. The monoallelic expression pattern can be achieved by imprinting, where one parental allele is repressed, often by DNA methylation. A number of imprinted genes are expressed in the placenta and some have been implicated in decidualisation in humans. Within the maternal-fetal concept, decidualisation may be one mechanism that has evolved to protect the mother. The decidualised stromal cells are large tightly joined cells that influence the extracellular matrix and produce a number of proteins including IGF binding proteins. This helps to inhibit the fetuses invasive capacity if implantation occurs (Emera, Romero, and Wagner 2012; Okada, Tsuzuki, and Murata 2018). When invasion remains unchecked it can result in placenta accrete, where the placenta invades into the uterus wall and can result in life-threatening haemorrhaging when the placenta detaches after child-birth. This conflict is further emphasised as stromal cells have been suggested to be involved in the rejection of developmentally impaired blastocysts, reducing the investment of the mother at an early stage (Teklenburg et al. 2010). In any case, spontaneous decidualisation leads to the requirement for menstruation.

Overall, the endometrium undergoes dramatic morphological changes throughout the menstrual cycle in response to hormones. This remodelling is associated with extensive transcriptional changes to facilitate it.

### 1.1.3 Transcriptional changes associated with the menstrual cycle

Morphological changes in the endometrium have been well studied since Noyes et al. in the 1950's, and to this day this fine-histological characterisation is used to determine the menstrual stage of biopsies (Noyes, Hertig, and Rock 1950). However, several reports

have acknowledged that determining the correct menstrual phase of biopsies by histology is not always accurate due to the lack of consistency between different, and sometimes even the same, evaluators (Gibson et al. 1991; Murray et al. 2004). There are additional ways to determine the menstrual phase of endometrial samples; self-reporting, LH-surge timing, and more recently with new technological advances using gene expression profiles (Altmäe et al. 2017; Enciso et al. 2018; Talbi et al. 2006).

Gene expression analysis of multiple biopsies across the menstrual cycle clearly segregates the samples into the menstrual cycle phases, allowing undetermined samples to be staged by their clustering position (Díaz-Gimeno et al. 2011; Ponnampalam et al. 2004; Talbi et al. 2006). Although multiple studies have shown that the transcriptional profile differs between the menstrual cycle phases, there is limited overlap between the studies in terms of genes that identify each stage. Altmäe et al. conducted a meta-analysis of transcriptome studies between the proliferative/early-secretory and mid-secretory phase and found that only 39 genes were consistent between the studies, perhaps identifying crucial and reliable genes such as nicotinamide N-methyltransferase (NNMT) and progestagen-associated endometrial protein (PAEP) (Altmäe et al. 2017). The overall inconsistency between studies is most likely due to the differences in samples, for instance, the timing of the samples, analysis approach as well as the population assessed in terms of their age, genetic background and diet. One of the studies included in this evaluation was the Endometrial Receptivity Array (ERA), which was designed as an endometrial dating tool (Díaz-Gimeno et al. 2011). It uses the expression levels of a subsets of genes to determine if the sample is in the non-receptive (outside the WOI) or receptive state (equivalent to the WOI) (Díaz-Gimeno et al. 2011). The ERA is being used commercially by Igenomix as well as in multiple research studies, particularly for personalised embryo transfer (Bassil et al. 2018; Hashimoto et al. 2017; Tan et al. 2018). Within the research setting, it is crucial to determine the exact timing of samples as this will affect the analysis and comparisons being carried out.

Generally, the proliferative phase sees an upregulation of genes involved in proliferation, including cell-cycle regulation, cell division, mitosis, and DNA replication. In addition, genes categorised in the gene ontology terms involved in extracellular matrix, ion channels and steroid binding are upregulated during the proliferative phase (Burney et al. 2007; Ruiz-Alonso, Blesa, and Simon 2012; Talbi et al. 2006). These processes are all involved in the cellular proliferation and remodelling that occurs during this stage.

Gene ontology terms characteristic of transcripts upregulated in the secretory phase are more variable, as this phase undergoes dramatic changes within itself. One of the terms that is shared across the various sub-stages within the secretory phase is the inhibition of cell proliferation. That aside, the early secretory phase is characterised by an upregulation of factors involved in response to oxygen species as well as metabolism and transport (Ruiz-Alonso, Blesa, and Simon 2012; Talbi et al. 2006). By contrast, the mid-secretory phase shows terms involved in the response to increased inflammation as well as glandular secretions such as immune response, wound healing, chemotaxis, cell-cell adhesion junctions as well as high cellular metabolism and synthesis of amino acids, steroids and lipids (Altmäe et al. 2017; Ruiz-Alonso, Blesa, and Simon 2012; Talbi et al. 2006). The late secretory phase is characterised by terms involved in cell cycle arrest, apoptosis, interactions with the extracelluar matrix and immune response regulation (Ruiz-Alonso, Blesa, and Simon 2012; Talbi et al. 2006).

There have been reports that have used laser capture microdissection to isolate key cell types to establish their expression profiles during the WOI. This revealed that luminal and glandular epithelium as well as stroma express some shared but also some distinct sets of genes, confirming that each cell type plays differing roles within the endometrium (G. E. Evans et al. 2012, 2014).

### 1.1.4 Model systems to study the endometrium

The endometrium is difficult to study due to a lack of appropriate model systems. Humans are one of the few species that have a menstrual cycle. For instance, in mice the uterine environment adapts as a result of fertilisation and implantation rather than in preparation for a potential pregnancy, and therefore does not undergo spontaneous decidualisation followed by menstruation. There are a number of higher primates as well as four species of bats and elephant shrews that have evolved menstruation (Emera, Romero, and Wagner 2012). *Papio anubis*, a primate species, is one such primate that does menstruate and is a good model for human endometrium (Afshar et al. 2013; Nothnick et al. 2016; Sherwin et al. 2010), but difficult to use experimentally for ethical and practical reasons.

Cultured cell lines can be useful models to investigate the endometrium, although the culture system lacks the complexity and communication of a multi-cell type tissue. There are several immortalized epithelial cell lines that originate from endometrial cancers, such as HEC-1A, Ishikawa, RL95-2 and EEC-1 (Bhagwat et al. 2013; Buck et al. 2015; Colón-Caraballo, Monteiro, and Flores 2015; Lessey, Vendrov, and Yuan 2003; Mo et al. 2006; Tamm et al. 2009). These cell lines have been widely used in research. Of note also is that these different cell lines can be used to reflect different stages of the menstrual cycle. For example, HEC-1A cells are more representative of the pre-receptive state, whereas RL95-2 cells better reflect the receptive state (Bhagwat et al. 2013; Thie and Denker 2002). However, it is worth noting that not all cell lines retain their responsiveness to hormones (Lessey, Vendrov, and Yuan 2003). The longevity of these cells means that they can be manipulated and functional studies can be carried out which cannot be conducted in vivo. Cultured primary cells are an option where available, however, these often can only be cultured for a low number of passages, restricting manipulation (Grimaldi et al. 2016; Tamura et al. 2014). More recently there has been the development of cultured endometrial gland organoids from primary cells (Turco et al. 2017). These can be passaged seemingly indefinitely and are responsive to hormones, opening up new avenues for research.

To study the human endometrium, perhaps one of the best options is still the primary tissue itself, and accordingly biopsies are often used for analysis (Díaz-Gimeno et al. 2011; G. E. Evans et al. 2018; Talbi et al. 2006). They allow scientists to study the exact tissue within the context of its complex environment including exposure and response to hormones and cell to cell communications. As long as biopsies are processed quickly after retrieval, there is little tissue degradation or metabolic changes that occur to the tissue. Although the primary tissue has many advantages, the disadvantage is that it does contain several cell types. Therefore, cell type composition and tissue complexity may obscure more subtle changes that occur in a small cell population only. Another potential problem is the difficulty to obtain true control samples, as the women who kindly provide the biopsies often attend a clinic for fertility-related issues or tests.

The system to be used to study the endometrium depends on the questions asked. Different systems have their own pros and cons and are suitable for different experiments. These all need to be considered when designing the experiments to pursue the answer to the scientific question.

## 1.2 Endometriosis: The disease, diagnosis and treatments

### 1.2.1 What is endometriosis?

Endometriosis is a condition where lesions composed of cells originating from the endometrium are found outside (i.e. ectopically) of the uterus. These lesions can be composed of both endometrial epithelial and stromal cells. They are predominantly found in the peritoneum, ovaries and rectovaginal septum (Vercellini et al. 2014). They also undergo similar cyclical processes to eutopic endometrium during the menstrual cycle by responding to hormones. Endometriosis is an estrogen-dependant condition and it has been shown that lesions contain estrogen receptors (ER), although the ER expression levels differ between patients. The estrogen responsiveness of the lesions is demonstrated in a mouse model where the number and weight of lesions as  as the proliferation marker Ki67 were significantly increased when estradiol was administered (Burns et al. 2012). Moreover, there are consistently low levels of estrogen in peritoneal fluid and this in turn stimulates continuous growth of tissue, mediated through ER activation (Khan et al. 2008; Vercellini et al. 2014). Endometriosis is also associated with progesterone resistance, although the exact mechanisms remain unknown (Burney et al. 2007). This creates a scenario of sustained pro-proliferative signals through the action of estrogen, while the anti-proliferative effects of progesterone are suppressed.

The prevalence rates of endometriosis are difficult to determine but it is estimated to affect approximately 10% of women of reproductive age (Nnoaham et al. 2011). The symptoms of endometriosis depend greatly on the individual woman; they include infertility, chronic pelvic pain, dyspareunia, dysmenorrhea, and abnormal or heavy menstrual flow. Yet endometriosis can also remain asymptomatic (Guo 2009a; Holt and Weiss 2000). A study carried out on 221 women with infertility for 1 year or more, with regular cycles and partners with normal semen according to WHO, showed that the prevalence of endometriosis was 47% (Meuleman et al. 2009). This demonstrates that there is a high proportion of women that are infertile and who may also have undiagnosed endometriosis. Endometriosis has a substantial detrimental effect on the women's physical quality of life (Nnoaham et al. 2011). It also incurs a high financial burden to society with one report estimating that the direct (medical) and indirect (loss of productivity and leisure time due to unpaid caregiving) cost of endometriosis is $1.8 billion annually to the Canadian society alone (Levy et al. 2011). Another report

suggested that the cost of endometriosis to society solely due to loss of work productivity ranges from $4 (USD) in Nigeria to $456 (USD) in Italy per woman per week (Nnoaham et al. 2011).

The progression of endometriosis in humans has not been extensively studied. This is mostly due to the fact that diagnosis is usually not established until several years after initial symptoms. It would also require multiple surgeries to be performed without any intervention undertaken to be able to track the condition (Nnoaham et al. 2011; Sourial, Tempest, and Hapangama 2014).

### 1.2.2 How is endometriosis diagnosed and classified?

Endometriosis is very difficult to diagnose due to the wide and varying symptoms. Many women have repeated visits to their primary health care professional, with on average 7 visits, until they are referred. Following on from this it takes on average 6.7 years to get a conclusive diagnosis of endometriosis (Nnoaham et al. 2011).

The gold standard for a conclusive diagnosis is a laparoscopic medical examination followed by a histological examination of the biopsies obtained (Dunselman et al. 2014). This has to be carried out by a surgical team and resection of endometriotic lesions is usually carried out simultaneously, if required. The stage of endometriosis is scored in accordance with the revised American Society for Reproductive Medicine classification of endometriosis: 1996 (rASRM) (Canis et al. 1997). This scoring system is based on location, size and depth of the lesions, which results in a classification of Stages I to IV; minimal, mild, moderate and severe, respectively (Canis et al. 1997). There are three well described types of lesions; peritoneal lesions, endometriomas (a chocolate fluid filled cyst on the ovary) and deep infiltrating endometriosis. The rASFM stages classify the overall endometriosis level, with minimal and mild usually associated with superficial peritoneal lesions, moderate with an endometrioma >3cm and severe with bilateral endometriomas or ablation of Douglas pouch (the extension of the peritoneal cavity between the uterus and the rectum). There is still some dispute in the field whether endometriosis is a single condition or a condition made up of several different pathology types (Gordts, Koninckx, and Brosens 2017; Hufnagel et al. 2015; Nisolle and Donnez 1997).

Diagnostic imaging is being encouraged pre-operatively by the Research Priorities for Endometriosis (Rogers et al. 2017). However, this approach is limited by the availability of equipment as well as the experience of the operator. It is also unable to diagnose mild or minimal endometriosis through imaging alone. It can, however, help surgeons plan for what to expect as well as the extent of the surgery needed to be conducted.

Biomarkers for endometriosis have been investigated but many demonstrate opposing findings in different studies. This could be due to the heterogeneity in endometriosis itself as well as in controls and the lack of validation experiments. Currently biomarkers are not recommended for the diagnosis of endometriosis (Dunselman et al. 2014).

### 1.2.3 Treatments for endometriosis

Currently there is no cure for endometriosis and only a few treatment options available, these can be split into hormonal or surgical. Usually the initial response is hormone therapy to aid with pain management. Several hormones can be prescribed these include the combined oral contraceptive (estrogen and progesterone), which reduces the menstrual flow and can reduce associated pain, it is usually used in milder forms of endometriosis. Progestogens (also known as progestagens), i.e. synthetic hormones that bind to and activate the progesterone receptor, can also be used as they behave like natural progesterone and are thought to reduce symptoms by repressing the growth of endometriotic tissue and the associated inflammation. An alternative route is to use GnRH analogues, which simulate natural GnRH. The use of GnRH causes the suppression of the production of estrogen, resulting in temporary menopause, this in turn reduces the size of endometriotic lesions and the associated detrimental effects.

Hormone therapy is often prescribed without endometriosis being diagnosed laparoscopically. These therapy types replicate either temporary menopause or pseudo-pregnancy, with the latter also able to act as a contraceptive. Evidence for their effectiveness is sparse and sporadic and the associated side-effects with each hormonal treatment need to be considered (Dunselman et al. 2014; Guidice 2010). Symptoms often recur at discontinuation of the drugs, making this an unsuitable strategy for women who wish to conceive.

Another avenue of treatment for endometriosis is to surgically remove the lesions. The effectiveness of this treatment depends on the level and location of the disease. Surgery to remove the lesions can be extensive and complicated, and often it is impossible to target all lesions. This could be due to their location, size or the difficulty in locating the lesions due to their variability in presentation. Surgery can elevate pain symptoms as well as increase the chances of spontaneous conception potentially through a reduction in inflammation (Vercellini et al. 2009). Even if a suspected complete removal of endometriosis lesions is achieved, there is still a 40-50% chance of recurrence after five years (Guo 2009b; Shen et al. 2009). Surgery removing the uterus, the ovaries and lesions (complete hysterectomy) can be performed in women where other treatment options have not been successful and pregnancy is not desired. However, even then this does not necessarily lead to a cure of endometriosis or its symptoms with persistent or reoccurring pain in 10% of patients (Dunselman et al. 2014; Guidice 2010).

If the desired outcome of treatment is to overcome infertility, then IVF is usually recommended. The success rate of IVF is lower in women with endometriosis compared to women with other causes of infertility and this drops further with increasing severity of endometriosis (Vercellini et al. 2014). Treatment to suppress endometriosis prior to IVF such as with GnRH analogues, oral contraception or surgery can increase IVF success, perhaps via the reduction of an inflammatory environment although evidence is limited (Guidice 2010; Vercellini et al. 2014).

Treatments for endometriosis are limited and very much depend on the individual patient, but must take into consideration the symptoms, the extent of the condition and if there are fertility concerns.

### 1.2.4 Theories of the pathogeneses of endometriosis

As to the aetiology of endometriosis, there are several theories as to how this condition may develop; the most recognised hypothesis is Sampson's theory (Sampson 1927). It suggests that endometriosis is the result of retrograde menstruation, where the endometrial cells reflux through the fallopian tubes and enter the peritoneal cavity where the cells attach, thrive and form lesions (**Error! Reference source not found.**).

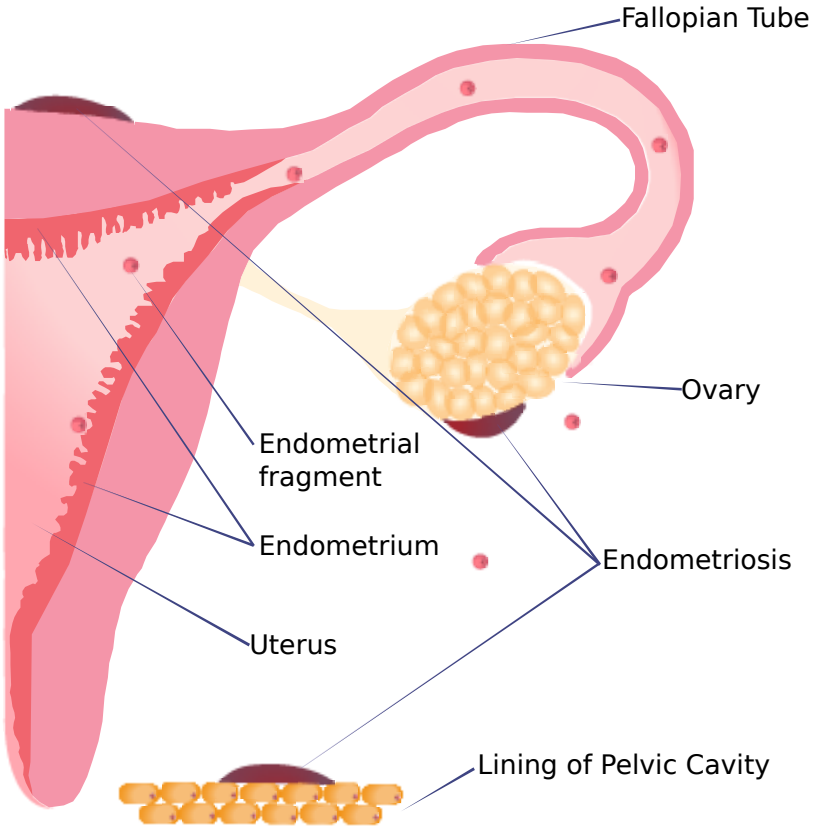



**Figure 1.3 Schematic of retrograde menstruation and endometriosis.** A diagram of half of the human uterus lined with endometrial tissue. Fragments or cells of the endometrium travel via the fallopian tube into the pelvic cavity. Endometriotic lesions are present on the outside of the uterus and its surrounding environment.

The mechanism behind what causes retrograde menstruation is unknown, with the majority of research concentrating on the frequency rather than the cause. It has been suggested however, that dysregulation of uterine peristalsis contractions, which are involved in the transportation of spermatozoa, may result in retrograde menstruation (Leyendecker et al. 2004). Retrograde menstruation is a common occurrence though, with most women experiencing it at some point, yet the majority do not develop endometriosis. What allows the cells to implant and proliferate resulting in endometriosis is unknown (Eyster et al. 2007; Vercellini et al. 2014). Several factors could be involved in this phenomenon; the immune system could fail to clear the cells or variants of single nucleotide polymorphisms (SNPs) or epigenetic alterations could favour attachment. It has been noted that endometriotic lesions themselves contain cells with stem cell properties which could contribute to the proliferation and formation of lesions (Hufnagel et al. 2015). Connected to this theory is another related theory, that says that ~5% of neonates experience a response to progesterone withdrawal following parturition, resulting in uterine bleeding which could lead to the formation of lesions that remain dormant until menarche (Brosens and Benagiano 2013).

Coelomic metaplasia is another proposed theory. This is where cells situated in situ, for example in the peritoneum, undergo metaplasia as well as transformation from one cell type to another, resulting in cells resembling endometrial epithelium and stroma.

However, this theory is only applicable to a subset of locations, ovaries and the peritoneum, as the metaplasia is suggested to occur at the germinal epithelium of the ovary and the serosa of the peritoneum (van der Linden 1996; Vinatier et al. 2001). It cannot explain the presence of endometriotic lesions on the colon, small intestines and urinary track (H. J. Lee et al. 2015). There is little evidence that supports this hypothesis, including that it cannot explain several aspects of endometriosis, such as the fact that endometriosis predominantly occurs in women of reproductive age with functioning endometrium, and how endometriosis occurs at other locations.

A third theory for the origins of endometriosis is Müllerian remnant abnormalities, yet again this only explains a subset of endometriosis locations, notably the uterus wall, fallopian tubes and deep infiltrating lesions. The Müllerian ducts develop in the foetus and result in the formation of the uterus, upper vagina and fallopian tubes. It is suggested that during the development of these organs some primordial cells may aberrantly migrate or differentiate and that these cells constitute the precursors to endometriotic lesions (Gordts, Koninckx, and Brosens 2017; van der Linden 1996).

Endometriosis shows signs of heritability with an increased prevalence of 3-15 times in first-degree relatives with endometriosis compared to the public (Kennedy, Mardon, and Barlow 1995; Matalliotakis et al. 2008; Saha et al. 2015; Stefansson et al. 2002). Twin studies with self-reported endometriosis and pathology reports, where available, concluded that heritability accounted for ~47-51% of cases (Saha et al. 2015; Treloar et al. 1999). This has led to a number of Genome Wide Association Studies (GWAS) that have associated a limited number of SNPs with endometriosis which have a small affect. However, different GWAS studies often produce conflicting or non-overlapping results, and also suggest that more SNPs would need to be identified to account for heritability (Falconer, D'Hooghe, and Fried 2007; Nyholt et al. 2012; Rogers et al. 2017). This culminates in the appreciation of complex genetic and suspected environmental factors contributing to the disease and the role they may have in the pathogenesis of endometriosis.

Many avenues of research have been conducted to try to establish other factors leading to endometriosis. These include research into environmental toxins (Rier and Foster 2003), immune or autoimmune diseases (Ahn et al. 2015; Matarese et al. 2003) and reproductive hormones (Nasu et al. 2011), although publications are often inconsistent

and conflicting. These are also all factors which could contribute to the likelihood of the attachment and formation of lesions after the initial endometrial tissue is located in an ectopic environment. With the lack of any confirmed causal link, more recently there has been an influx of studies looking into potential epigenetic contributions to the condition (Borghese et al. 2008; Dyson et al. 2014; Naqvi et al. 2014; Wu et al. 2006). The epigenome has been regarded specifically as the potential 'missing link' in endometriosis because it bridges genetics and environmental effects, and has long been overlooked as a potential factor involved in the development or progression of the disease.

## 1.3 Epigenetic Modifications

The epigenome constitutes the regulatory layer that is imposed onto the DNA sequence itself and directs chromatin compaction, genome accessibility and gene activity. The epigenome consists of multiple interconnected elements (Figure 1.4).

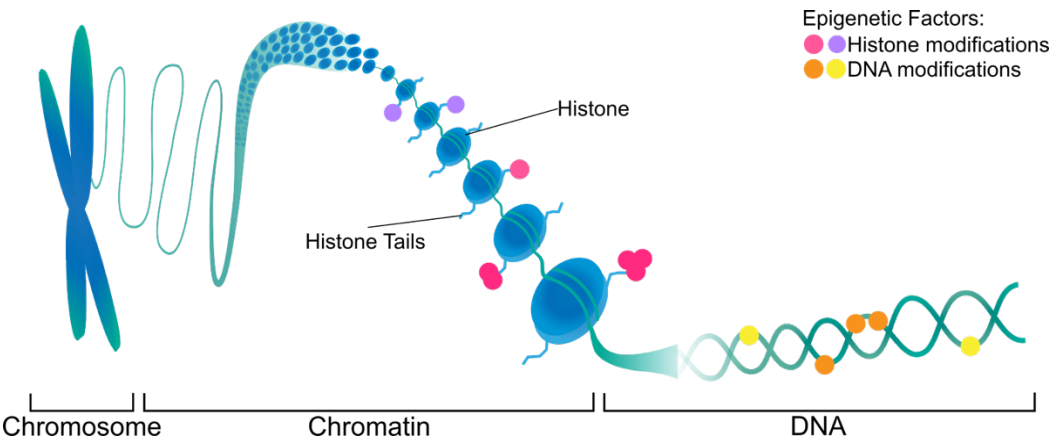



**Figure 1.4 Overview of Epigenetic modifications.** Abstract representation of epigenetic modifications to DNA and histones.

Covalent modifications to the DNA itself have been extensively studied, notably affecting the base cytosine (C) that can be methylated (5mC), hydroxymethylated (5hmC) or even undergo further modifications into carboxy- and formylcytosine. Another layer of the epigenome is provided by modifications to the histones around which the DNA is wrapped in the nucleus. Histone modifications entail a wide spectrum of biochemical

changes, including methylation, phosphorylation, acetylation, ubiquitylation and poly-ADP-ribosylation that occur at specific amino acids. Some of these modifications have been firmly linked to gene activation or repression.

### 1.3.1 Histone modifications

The chromatin structure is made up of nucleosomes which consist of 146bp DNA wrapped around histones. These histone cores are comprised of an octamer of proteins (Luger et al. 1997). The N-terminal tails of histones can be modified altering the interactions with DNA and DNA binding proteins and therefore the chromatin structure.

There have been many studies into the histone modifications that occur and the effect that they have on transcription levels of associated genes and chromatin packaging. The effect depends on the type of modification that is present and can at least in some cases be correlated either with an increase or a decrease in expression levels of the corresponding genes.

In this summary I will concentrate on three well-studied histone modifications, namely on histone 3 lysine 4 tri-methylation (H3K4me3) which is often associated with gene activation, histone 3 lysine 9 tri-methylation (H3K9me3) and histone 3 lysine 27 tri-methylation (H3K27me3) that are often associated with gene repression.

H3K4me3 is associated with euchromatin (open and transcriptionally active chromatin) and is commonly found enriched at the 5' end around the transcriptional start sites of active genes. H3K4me3 levels positively correlate with gene expression levels, although the modification does not strictly need to be present for expression to occur (Azuara et al. 2006; Heintzman et al. 2007; Roh et al. 2006). H3K4me3 can also be used to bioinformatically predict where unidentified promoters may be positioned (Heintzman et al. 2007). Mouse embryos that are deficient in H3K4 methyltransferase, MII2, are embryonic lethal prior to embryonic day (E)11.5 due to a failure to maintain appropriate expression profiles (Glaser et al. 2006). Broad H3K4me3 enrichment domains around the transcriptional start sites (TSSs) that also extend into the gene body itself have been reported for their importance to confer stability of gene expression; they are associated with active genes involved in the cells' transcriptional program that

confers cellular identity and regulates key functions of that particular cell type (Benayoun et al. 2014; Xiaoyu Liu et al. 2016; Roh et al. 2006).

H3K27me3 is deposited by the Polycomb Repressive Complex 2 (PCR2). The PRC2 complex contains several core proteins including EED, SUZ12 and EZH2. EZH2 harbours the catalytic activity of the complex and confers both di- and tri-methylation at H3K27. PCR2 and its deposition of H3K27me3 is essential for development. Mouse embryos deficient for any of these PCR2 core components fail to complete gastrulation and are embryonic lethal as follows: *Ezh2* null mice between E7-8.5, *Eed* null mice between E8.5-9.5 and *Suz12* null mice at E10.5 (Faust et al. 1995; O'Carrol et al. 2001; Pasini et al. 2004). H3K27me3 is generally associated with gene repression and thereby silences genes that often have key functions in other cell types (Roh et al. 2006). It thereby contributes to shaping and maintaining cell identity. H3K27me3 can form broad peaks, where the modification is present not only over the promoter, 1kb downstream and 100bp upstream of the TSS, but also extended to cover the gene body as well (Barski et al. 2007; Roh et al. 2006; Young et al. 2011).

Some genes can be marked by H3K27me3 as well as the functionally opposing histone modification H3K4me3. This is often found in cells that exhibit elevated plasticity such as embryonic stem cells, and affects many developmentally relevant genes. Genes that harbour this combination of marks have been classified as being "bivalent" (Azuara et al. 2006; Bernstein et al. 2006; Roh et al. 2006; Young et al. 2011). Bivalent genes are proposed to being held in a 'poised' position so that they can quickly respond to environmental signals to either be 'switched on' or to be permanently 'switched off'. Mouse embryonic stem cells (mESC) that are *Eed*-deficient, i.e. that lack a core protein of the PCR2 complex, fail to differentiate fully and exhibit higher expression levels of genes that are normally bivalently marked in the stem cell state, such as neural-specific genes (Azuara et al. 2006). Along similar lines, human ESCs (hESCs) that lack the catalytic subunit of the PRC2 complex EZH2, have reduced self-renewal potential and exhibit reduced proliferation rates in the stem cell state. When triggered to differentiate they initiate this process but fail to fully progress into mature endodermal, mesodermal and ectodermal lineages (Collinson et al. 2016). These examples demonstrate that the bivalent mark is essential to fine-tune the transition between gene suppression and activation in particular during the phase of early developmental decisions. In the case of ESCs, genes that are bivalently marked are also often linage specific and therefore need to be 'poised' to commit to a lineage under the correct environmental stimulus (Azuara

et al. 2006; Bernstein et al. 2006; Collinson et al. 2016). T-cells also have a subset of genes that are bivalently marked, these are often induced for rapid activation, as is the case in ESCs. T-cells require a swift activation or repression of genes when they come in contact with specific antigens, and bivalent chromatin domains help to facilitate this (Roh et al. 2006).

H3K9me3 has been shown to act as a transcriptionally repressive modification and is often found in co-localization with H3K27me3 and/or DNA methylation (Boros et al. 2014; Du et al. 2015; Fuks et al. 2003). Its repressive role arises from the exclusion of transcription factors being able to bind to DNA that is marked by H3K9me3, mostly due to H3K9me3 being associated with chromatin compaction (Soufi, Donahue, and Zaret 2012).Therefore, H3K9me3 is predominantly associated with heterochromatic regions. This heterochromatic DNA can contain repeat-rich regions, such as tandem-repeat satellites and retrotransposons. These are marked by H3K9me3 (as well as by DNA methylation), resulting in their repression which is an essential mechanism in the safe-guarding of the genome (Becker, Nicetto, and Zaret 2016; Nakayama et al. 2001; Peters et al. 2001). Fibroblasts lacking *Suv39h1* and *Suv39h2,* which are H3K9 methyltransferases, have high levels of chromatin mis-segregation resulting in a significant number of tetraploid and octaploid cells. This is most likely due to the depreciation of H3K9me3 at pericentric heterochromatin resulting in genomic instability (Peters et al. 2001). *Suv39h1/2* double knockout mice have reduced viability, are growth retarded and have a higher risk of tumour development (Peters et al. 2001). A knockdown of *SUV39H1* and *SUV39H2* in human fibroblasts increased their capacity to be reprogrammed to pluripotent stem cells, suggesting that H3K9me3 is required to repress pluripotent genes (Onder et al. 2012; Soufi, Donahue, and Zaret 2012). Heterochromatin can spread across chromosomes through the interaction of H3K9me3 with heterochromatin protein 1 (HP1) (Becker, Nicetto, and Zaret 2016; Nakayama et al. 2001). SUV39H1 directly interacts with DNMT3A and DNMT1, and these also interact with HP1β, suggesting that there is a direct link between H3K9me3 and DNA methylation and that both modifications form part of a reinforcing repressive chromatin state (Fuks et al. 2003). The profiles of H3K9me3 deposition differ depending on the cell-type highlighting its importance for cell identity (Becker, Nicetto, and Zaret 2016; Soufi, Donahue, and Zaret 2012).

### 1.3.2 DNA Modifications

Another epigenetic layer that shapes chromatin organisation is the epigenetic mark DNA methylation (5mC). This involves the addition of a methyl group at the 5-carbon position to a cytosine usually in the context of CpG dinucleotides. It occurs on average at ~50-80% of CpGs within the human genome (Laurent et al. 2010). The distribution of CpGs across the genome is uneven. Regions that are CpG rich - termed CpG Islands (CGIs) – are usually located within the promoter region of genes and tend to be hypomethylated. A more biologically relevant definition of CGIs, that is not merely based on GC content but on the binding of specific CpG binding proteins that contain a CxxC motif, has expanded the list of CGIs to include in particular those that are methylated in a tissue-type specific manner depending on cell lineage and cell type (Illingworth et al. 2008, 2010; F. Song et al. 2005). By contrast, promoters that do not contain CGIs have relatively high levels of CpG methylation, ~35-40%, compared to ~1-2% of CGI containing promoters (Laurent et al. 2010).

The functional role of DNA methylation has been particularly well studied in X-inactivation as well as its essential role in the regulation of imprinted regions. The phenomenon of genomic imprinting was first identified in mouse embryos which contained either two female pronuclei or two male pronuclei. These failed to complete embryogenesis suggesting that the genome contribution of both parents is required (McGrath and Solter 1984; Surani, Barton, and Norris 1984). Subsequently it has been established that through allelele-specific DNA methylation, key genes (that are often involved in regulating foetal and/or placental growth and function) are expressed in a parent-of-origin dependent manner. This relatively small group of loci are known as imprinted genes. They are often clustered together in particular genomic regions (Giannoukakis et al. 1993; Reik et al. 1987; Swain, Stewart, and Leder 1987). Insulin-like growth factor-II (IGF2) and H19 are two such imprinted genes that are present in both mice and humans. IGF2 is expressed from the paternal copy only where as H19 is expressed from the maternal copy only (Bartolomei, Zemel, and Tilghman 1991; DeChiara, Robertson, and Efstratiadis 1991; Rainier et al. 1993). DNA methylation plays a key role in establishing the monoallelic expression of imprinted genes; disruption of the DNA methylation machinery results in aberrant expression of imprinted genes (E. Li, Beard, and Jaenisch 1993).

If the expression and therefore dosage of imprinted genes is disrupted, this leads to developmental defects. In humans, imprinting disorders such as Prader–Willi syndrome, Beckwith–Wiedemann syndrome (BWS) and Silver–Russell syndrome (SRS) are well studied examples that result from the disruption of specific imprinted gene loci (Abu-Amero et al. 2010; Bliek et al. 2009; Buiting et al. 1994; Kannenberg, Urban, and Binder 2012).

DNA methylation also plays a pivotal role in maintaining genome stability through its role in silencing repetitive regions such as retrotransposons and centromeres (Jones 2012; Vilain et al. 1999). Here, it often works in conjunction with H3K9me3. Indeed, both epigenetic modifications are inter-related through the bridging function of the protein UHRF1 that binds to the trimethylated K9 mark on the one hand, and to 5mC on the other. Thereby, UHRF1 mediates DNA methylation of H3K9me3-modified regions and vice versa, creating an element of inbuilt redundancy in the surveillance and stable maintenance of the genome (Du et al. 2015; T. Li et al. 2018; Rothbart et al. 2012).

The most widely accepted role of DNA methylation is its function as repressive modification in particular in heterochromatin, but also at specific gene loci in euchromatin. Traditionally, it has been proposed that DNA methylation at promoter regions confers silencing of the associated gene. This is broadly correct and exemplified at some key loci, such as the *Elf5* transcription factor (R. K. Ng et al. 2008). However, many high-throughput studies that involve DNA methylation and gene expression profiles indeed find a fairly poor (negative) correlation between DNA methylation and gene expression on the global level (van Eijk et al. 2012; Spainhour et al. 2019). Apart from the effect of DNA methylation at gene promoters, it has been shown that actively transcribed genes are more highly methylated across their gene body. This is conferred by the corresponding enrichment of RNA polymerase II, as part of the elongation complex, along with SET2D2, a histone methyltransferase which deposits H3K36me3, both of which demarcate actively transcribed genes (Hahn et al. 2011; Yoh, Lucas, and Jones 2008). The H3K36me3 mark in turn can be bound by the *de novo* DNA methyltransferases DNMT3A and DNMT3B, linking actively transcribed genes with DNA methylation. The extent of gene body methylation correlates with higher expression, especially in the top 20% of expressed genes (Laurent et al. 2010). 5mC can also denote splicing junctions between exons (enriched) and introns (depleted) by the sharp depletion of 5mC between the transition from exons to introns (Hodges et al. 2009; Laurent et al. 2010). A general role of gene body methylation may be to keep alternative

splicing and cell type-foreign splice isoforms at bay, thereby conferring stability of gene expression.

DNA methylation is conferred by DNA methyltransferases, notably DNMT1, DNMT3A and DNMT3B. DNMT3A and DNMT3B are DNA methyltransferases that are responsible for *de novo* methylation and can be recruited by their co-factor DNMT3L which is catalytically inactive but binds to H3 when H3K4 is unmethylated (Okano et al. 1999; Ooi et al. 2007; Suetake et al. 2004). By contrast, DNMT1 is regarded as the maintenance DNA methyltransferase that confers re-methylation of hemi-methylated DNA following replication (En Li, Bestor, and Jaenisch 1992). DNA methyltransferases are key for embryonic development in mice and knockouts show developmental defects and lethal phenotypes for DNMT1, 3A and 3B (Bourc'his et al. 2016; En Li, Bestor, and Jaenisch 1992; Okano et al. 1999). DNMTs can be dysregulated in cancer and this has been linked to the frequently observed epigenetic rearrangements including global hypomethylation and gene specific hypermethylation; these epigenetic changes may contribute to the aggressive phenotype of certain cancers (Jones 2012; W. Zhang and Xu 2017).

A longstanding question in the field of epigenetics has been how DNA methylation is removed. One well-understood mechanism is the passive demethylation via replication and cell division, where the 5mC mark is not reinstated on the daughter strand resulting in a gradual dilution of 5mC. Over recent years, it has also been established that DNA methylation can be actively –albeit indirectly– removed through deamination of 5hmC involving the enzyme activation-induced cytidine deaminase (AID), which results in a base mismatch and is followed by base excision and repair (Bhutani et al. 2010). The main established route, however, involves the ten eleven translocation enzymes (TET1, TET2 and TET3), which catalyse the oxidation of the methyl group to a hydroxymethyl group resulting in DNA hydroxymethylation (5hmC). This process can continue with the oxidation of 5hmC to 5-formylcytosine (5fC) and then 5-carboxylcytosine (5caC) by TET enzymes, which then again involves base excision repair and DNA repair pathways resulting in an unmodified cytosine (C) (He et al. 2011; Ito et al. 2011; Tahiliani et al. 2009).

Of these 5mC turnover products, 5hmC has gained particular attention because of its relative enrichment in particular cell types. Overall, the levels of 5hmC are highly variable and tissue dependent, and do not necessarily correlate with the global 5mC levels (Nestor et al. 2012; Senner et al. 2012). 5hmC has been found to be relatively high

in several cell types including ESCs and Purkinje neurons (Kriaucionis and Heintz 2009; Senner et al. 2012; Tahiliani et al. 2009). 5hmC can act as a tissue-type identifier as it is able to separate out differing tissues using unsupervised hierarchical clustering after 5hmC profiling (Nestor et al. 2012).

These findings have led to the view that 5hmC is not only an intermediary turnover product of 5mC as part of the DNA demethylation cascade, but can indeed constitute an independent epigenetic mark in itself (Bachman et al. 2014). Methyl-CpG binding domain (MBD) proteins bind to 5mC and recruit factors to condense and inactivate the chromatin landscape whereas they cannot bind to 5hmC, suggesting that 5hmC has a differing function (Valinluck et al. 2004). Since 5hmC requires 5mC as a substrate, it principally can only arise where DNA methylation is, or has been, present. Yet 5hmC is globally enriched at euchromatic domains and excluded from heterochromatic domains in mESCs, i.e. at this level it exhibits an almost opposite distribution to 5mC. It is specifically enriched at promoters, CGIs as well as gene bodies (Ficz et al. 2011; Mellen et al. 2012; Nestor et al. 2012; Pastor et al. 2011). In general, high levels of 5hmC at promoters, exons as well as gene bodies are positively correlated with gene expression levels (Ficz et al. 2011; Mellen et al. 2012). The positive correlation at promoters has been shown to increase if concomitantly these sites have low levels in 5mC, suggesting that the ratio of both marks may be a better readout for gene activity than either 5mC or 5hmC alone (Ficz et al. 2011). High levels of 5hmC at promoters are also often associated with H3K4me3 as a mark of actively transcribed genes. Moreover 5hmC is enriched at H3K4me3/H3K27me3 bivalently marked promoters that are associated with 'poised' genes (Ficz et al. 2011; Pastor et al. 2011). When TET1 and TET2 are knocked down in mESCs this resulted in a number of genes being downregulated and a significant increase in 5mC levels at their promoters, suggesting that TETs may influence transcriptional regulation (Ficz et al. 2011).

It is important to note that many techniques that investigate 5mC involve bisulfite treatment which is unable to distinguish between 5mC and 5hmC. This means that when evaluating reports on 5mC it is important to consider the role that 5hmC may be playing as well. As 5mC and 5hmC are inextricably linked, with one being the substrate for the other, it is important to take both modifications into account especially as together their impact on genome regulation can be assessed more accurately.

### 1.3.3 Epigenetic modifications associated with the menstrual cycle

As discussed, the endometrium undergoes dramatic changes during the menstrual cycle. This involves morphological and transcriptional changes; therefore, it is reasonable to assume that epigenetic modifications are involved in the control of these processes. Epigenetic modifiers are expressed in the endometrium and some respond to estrogen and/or progesterone, although reports are conflicting, these will be summarised in this section (Guo 2012).

There is a smattering of reports on histone modifications in the endometrium, with H3K27me3 being the most extensively studied. There are a few reports on H3K4me3, while H3K9me3 has not been examined beyond its total global level (mono, di and tri methylation combined). The total levels of several histone modifications show no significant differences between the proliferative and secretory phase, this includes H3K4me, H3K9me and H3K27me (Monteiro et al. 2014). This result is, however, somewhat meaningless as the EpiQuick kit used in this study detected mono, di and tri methylation of these histone modifications collectively, which may have obscured differences for any precise methylation state itself. Moreover, the study combined endometrium samples from controls and from patients with endometriosis, which the authors had shown previously differed with regard to the histone modification status (Monteiro et al. 2014).

As discussed, H3K4me3 is a histone modification associated with gene expression, however it has not been extensively studied in the endometrium as a whole but it has been shown to be enriched at transcribed genes in cultured stromal cells. Decidualisation of cultured stromal cells resulted in altered H3K4me3 status at many loci with the majority exhibiting an increase in the modification, which in turn correlated with an increase in expression of the associated genes (Tamura et al. 2014).

The reports on the state of H3K27me3, a repressive histone modification, during the menstrual cycle are more inconsistent with a summary in Table 1.1 (Colón-Caraballo, Monteiro, and Flores 2015; Grimaldi et al. 2016; Monteiro et al. 2014). This could be due to the differing systems used to analyse H3K27me3, such as whole tissue lysate, immunohistochemistry and stromal cell culture. H3K27me3 global levels in whole tissue and stromal cells remained relatively consistent between the secretory and proliferative phase or after decidualisation. However, when epithelial glands were specifically

examined, this revealed a significant increase in H3K27me3 in the secretory phase (Colón-Caraballo, Monteiro, and Flores 2015; Grimaldi et al. 2016). Yet EZH2, the methyltransferase responsible for the deposition of H3K27me3, decreases or is unchanged in the secretory phase compared to the proliferative phase, depending on the study (Colón-Caraballo et al. 2018; Grimaldi et al. 2016). Inhibition of EZH2 in two epithelial cell lines and cultured stromal cells did not have an effect on the global levels of H3K27me3 (Colón-Caraballo et al. 2018; Grimaldi et al. 2016). H3K27me3 did differ at specific sites in cultured stromal cells including at the decidual markers, PRL and IGFBP1 (Grimaldi et al. 2016; Tamura et al. 2014). This suggests that H3K27me3 could be important for the transition from the proliferative to the secretory phase through its redistribution at specific loci, although the evidence for a functionally relevant role remains sparse.

**Table 1.1 H3K27me3 and EZH2 in the Endometrium.** Summary of the literature with regards to H3K27me3 and EZH2 during the menstrual cycle and in cell culture. Method used for detection: EpiQuick = specific assay kit, WB = Western Blot, IHC = Immunohistochemistry, IF= Immunofluorescence, RT-qPCR = Reverse Transcription quantitative Polymerase Chain Reaction

| | | Endometrium | | | | | Cell Culture | |
|---|---|---|---|---|---|---|---|---|
| | | Proliferative Phase vs Secretory Phase | | | | | Estrogen + Progesterone | Reference |
| | | Whole tissue | Reference | Stomal Cells | Epithelilal Cells | Reference | | |
| Total Levels | **H3K27me3** | No Change (EpiQuick) | Colón-Caraballo et al. 2015 | No Change (IHC) | Increased (IHC) | Colón-Caraballo et al. 2015 | No Change (WB and IF) | Grimaldi et al. 2016 |
| | **EZH2** | Decreased (WB) | Grimaldi et al. 2016 | No Change (IHC) | No Change (IHC) | Colón-Caraballo et al. 2018 | Decreased (WB and IF) | |
| mRNA | **EZH2** | | | | | | Decreased (RT-qPCR) | |

Histone acetylation is associated with euchromatin and most broadly linked with high expression of the corresponding genes. With regards to the endometrium, multiple histone acetylation sites fluctuate across the menstrual cycle at global total levels (Monteiro et al. 2014; Munro et al. 2010). One specific mark, H3K27ac, has been studied in cultured stromal cells and shows a significant increase at a substantial number of genes which globally correlates with their upregulation following decidualisation (Tamura et al. 2014). In particular, glycodeline, prolactin or insulin-like growth factor

expression were specifically investigated in Ishikawa cells (endometrial adenocarcinoma cells) or cultured stromal cells following histone deacetylase inhibitor treatment (TSA and/or SAHA), which resulted in their corresponding upregulation. These genes were also upregulated when the cells were decidualised, suggesting that acetylation of histones may be involved in this transition (Sakai et al. 2003; Uchida et al. 2005). Additional research is required to understand the full role that histone acetylation modifications play in regulating gene expression throughout the menstrual cycle.

DNA methylation (5mC) is an important component of the epigenome and is influenced by the cyclic hormonal changes that occur during the menstrual cycle. 5mC decreases from the proliferative to the secretory phase, in particular in epithelial cells (Ghabreau et al. 2004).

The profile of 5mC in the endometrium differs across the menstrual cycle with a number of differentially methylated regions (DMRs) and individual CpGs identified between phases (Houshdaran et al. 2014; Kukushkina et al. 2017; Saare et al. 2016). Comparing early- with mid- secretory phase, DMRs either become hypo- or hyper-methylated at specific sites, with a similar split between the two (Houshdaran et al. 2014; Saare et al. 2016). By contrast, when cultured stromal cells were decidualised in vitro, the majority of DMRs became hypo-methylated (Dyson et al. 2014). The methylation profiles differ less when closer time points of the menstrual cycle are examined, as could perhaps be expected, suggesting that the methylome changes gradually and continuously throughout the menstrual cycle (Houshdaran et al. 2014; Kukushkina et al. 2017; Saare et al. 2016). With regards to expression, the methylation changes at a number of DMRs both positively and inversely correlated with gene expression changes (i.e. methylated and repressed, hypomethylated and induced or methylated and induced, hypomethylated and repressed). These DMRs fell into interesting gene categories including extracellular matrix proteins, and factors involved in wound healing and cell adhesion, which are associated with the corresponding menstrual phase (Dyson et al. 2014; Houshdaran et al. 2014; Kukushkina et al. 2017).

Studies of 5mC have been carried out on HumanMethylation450 or HumanMethylation27 platforms which cover >450,000 and >27,000 CpGs, respectively, out of ~28 million CpGs in the human genome. They require the DNA to be bisulfite

treated, which means that calls for methylated cytosines cannot be distinguished from 5hmC, which is an intrinsic compounding factor of all these studies (Houshdaran et al. 2016; Smith and Meissner 2013; Yong, Hsu, and Chen 2016). As far as I am aware there have been no reports on 5hmC in the endometrium.

DNMT1, DNMT3A and DNMT3B, the DNA methyltransferases responsible for the deposition of methylation, are all expressed in the endometrium (Dyson et al. 2015; van Kaam et al. 2011; Vincent et al. 2011; Yamagata et al. 2009). Whether or not the levels of DNMTs change across the menstrual cycle is disputed between different papers, with an overview in Table 1.2. Immunohistochemistry reveals no changes in DNMTs between the proliferative and the secretory phase (Yamagata et al. 2009). On the mRNA level, there is a general consensus that *DNMT3A* and *DNMT3B* decrease between the proliferate and secretory phase, which would be in line with the decrease in 5mC levels reported (Ghabreau et al. 2004). However, *DNMT1* was not consistent and was shown to decrease, increase or not change between the proliferative and secretory phase (van Kaam et al. 2011; Vincent et al. 2011; Yamagata et al. 2009). Within the secretory phase all three DNMTs were either at their lowest at the mid-secretory point or did not change from early to late secretory phase (Vincent et al. 2011; Yamagata et al. 2009). Cell culture analysis on stromal cells or proliferative endometrial explants similarly show no change or a decrease in expression after exposure to estrogen and progesterone for *DNMT3A* and *DNMT3B,* with again no consistency for *DNMT1* (Dyson et al. 2015; van Kaam et al. 2011; Vincent et al. 2011; Yamagata et al. 2009). Generally, DNMT3A and 3B decrease whereas DNMT1 had inconsistent and conflicting reports both *in vivo* and *in vitro*. Overall, this shows that DNMTs are expressed in the endometrium and are likely influenced by hormones, but the exact relationship between DNMT expression, hormone responsiveness and methylation changes still remain unclear.

The epigenome is dynamic and contributes to the expression profile and chromatin organisation of endometrium as it responds to hormonal changes during the menstrual cycle. The complexity of the epigenome in the endometrium is yet to be fully evaluated. A detailed appreciation of the epigenomic changes during the menstrual cycle are important for our understanding of the tissue's role and function, particularly with regards to preparing the uterus for implantation.

**Table 1.2 DNMTs in the Endometrium.** Summary of the literature with regards to DNMT1, DNMT3A and DNMT3B during the menstrual cycle and in cell culture. Method of analysis: IHC = Immunohistochemistry, WB= Western Blot, RT-qPCR = Reverse Transcription quantitative Polymerase Chain Reaction

| | | Endometrium | | | | Cell Culture | |
|---|---|---|---|---|---|---|---|
| | | Proliferative Phase vs Secretory Phase | Reference | Within Secertory Phase | Reference | Estrogen + Progesterone | Reference |
| Protein | **DNMT1** | No Change (IHC) | Yamagata et al. 2009 | | | No Change (WB) | Dyson et al. 2015 |
| | **DNMT3A** | | | | | No Change (WB) | |
| | **DNMT3B** | | | | | Decreased after 6 days (WB) | |
| mRNA (RT-qPCR) | **DNMT1** | Decreased / No Change / Increased | Yamagata et al. 2009 / Vincent et al. 2011 / van Kaam et al. 2011 | Lowest at Mid-Secretory / No Change | Yamagata et al. 2009 / Vincent et al. 2011 | No Change / No Change / Decreased after 48 hours/ Increased after 24hours | Dyson et al. 2015 / Yamagata et al. 2009 / Vincent et al. 2011 / van Kaam et al. 2011 |
| | **DNMT3A** | Decreased / Decreased | Yamagata et al. 2009 / Vincent et al. 2011 | Lowest at Mid-Secretory / No Change | | No Change / Decreased after 8 days / or after 48hours | Dyson et al. 2015 / Yamagata et al. 2009 / Vincent et al. 2011 |
| | **DNMT3B** | Trend to Decrease / Trend to Decrease / Decreased | Yamagata et al. 2009 / van Kaam et al. 2011 / Vincent et al. 2011 | No Change / No Change | | Decreased after 6 days / after 8 days / or after 48hours | |

### 1.3.4 Epigenetic modifications associated with endometriosis

In the absence of solid genetic disease linkages with endometriosis, epigenetics is an area that has more recently been pursued for a potential role in the aetiology of this condition.

There are a few papers that report on histone modification levels in endometriosis, summarised in Table 1.3. By using an EpiQuick assay kit, global H3 acetylation as well as H3K9ac and H4K16ac are lower in ectopic lesions compared to eutopic controls (Monteiro et al. 2014). Although lower histone acetylation levels would perhaps infer lower gene expression activity, H3K4me (mono, di and tri) is globally higher in lesions compared to eutopic endometrium from women with endometriosis, which in turn is higher than eutopic endometrium from controls (Monteiro et al. 2014). Yet the repressive marker H3K9me (mono, di and tri) globally and H3K9me3 specifically is also higher in lesions compared to eutopic controls (Xishi Liu, Zhang, and Guo 2018; Monteiro et al. 2014; Q. Zhang et al. 2017). Xiaomeng et al., showed no or opposing differences for H3 acetylation, H3K4me and H3K9me although this assessment was based on mixed samples from both the proliferative and secretory phase (Xiaomeng et al. 2013). How these changes tie in with overall gene expression output or complexity of gene expression (i.e. what fraction of the genome is active) remains to be elucidated.

Perhaps the only consistent changes have been reported for H3K27methylation. H3K27me (mono, di, tri) and specifically H3K27me3 is increased in endometriosis, both in ectopic and eutopic samples from patients with endometriosis, compared to eutopic controls (Xishi Liu, Zhang, and Guo 2018; Monteiro et al. 2014; Q. Zhang et al. 2017). This is consistent with elevated EZH2 expression levels in endometriotic lesions compared to controls, as well as in endometriotic cultured epithelial cells where higher EZH2 expression levels have been linked to increased cellular migration and invasiveness (Colón-Caraballo, Monteiro, and Flores 2015; Q. Zhang et al. 2017). In vivo treatment with an EZH2 inhibitor in an endometriosis mouse model resulted in a dose-dependent reduction of H3K27me3 and a reduction in lesion weight (Q. Zhang et al. 2017). This suggests that H3K27me3 may play a role in the establishment of endometriosis, but the overarching mechanism is yet to be clarified.

**Table 1.3 Histone Modifications in Endometriosis.** Summary of the literature with regards to histone modifications and EZH2 expression comparing ectopic lesions, eutopic tissue from patients with endometriosis and eutopic tissue from control patients. Method for analysis: EpiQuick = specific assay kit, IHC = Immunohistochemistry, WB = Western Blot.

| | | Endometrium | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Eutopic Endometriosis v Lesions** | Reference | **Eutopic Control v Lesions** | Reference | **Eutopic Control v Eutopic Endometriosis** | Reference |
| **Total Levels** | **H3ac** | No Change (EpiQuick) / No Change (EpiQuick) | Monteiro et al. 2014 and/or Xiaomeng et al. 2013 | Decreased in Lesions (EpiQuick)/ No Change (EpiQuick) | Monteiro et al. 2014/ Xiaomeng et al. 2013 | Trend to decreased in Eutopic Endometriosis (EpiQuick) / No Change (EpiQuick) | Monteiro et al. 2014/ Xiaomeng et al. 2013 |
| | **H3K9ac and H4K16ac** | No Change (EpiQuick) | | Decreased in Lesions (EpiQuick) | Monteiro et al. 2014 | Decreased in Eutopic Endometriosis (EpiQuick) | Monteiro et al. 2014 |
| | **H3K4me** | Increased in L (EpiQuick) | | Increased in Lesions (EpiQuick) | Monteiro et al. 2014 | Increased in Eutopic Endometriosis (EpiQuick) | Monteiro et al. 2014 |
| | **H3K9me or H3K9me3** | No Change (EpiQuick) / Decreased in L (EpiQuick) | | Increased in Lesions (EpiQuick)/ (IHC) / (IHC) / Decreased in Lesions (EpiQuick) | Monteiro et al. 2014/ X. Liu et al. 2018 / Q. Zhang et al. 2017/ Xiaomeng et al. 2013 | No Change (EpiQuick) / Increased in Eutopic Endometriosis (IHC)/ No Change (EpiQuick) | Monteiro et al. 2014/ Q. Zhang et al. 2017/ Xiaomeng et al. 2013 |
| | **H3K27me or H3K27me3** | No Change (EpiQuick) | | Increased in L (EpiQuick) / (IHC) / (IHC) | Monteiro et al. 2014/ X. Liu et al. 2018 / Q. Zhang et al. 2017 | Increased in Eutopic Endometriosis (EpiQuick) / (IHC) | Monteiro et al. 2014/ Q. Zhang et al. 2017 |
| | **EZH2** | | | Increased in L (IHC) | Q. Zhang et al. 2017 | Increased in Eutopic Endometriosis (IHC and WB) | Q. Zhang et al. 2017 |

Multiple studies have performed DNA methylation analyses, mostly using the above-mentioned methylation arrays as well as site- specific analysis methods, and report on a number of locus-specific changes. Often, these are again highly variable and inconsistent between studies. The most established changes appear to occur at the Homeobox A10 (*HOXA10*) locus and the Progesterone Receptor (*PGR*) gene. A summary of the literature with regards to DNA methylation and endometriosis at *HOXA10, PGR* and globally are

presented in Table 1.4. *HOXA10* is highly expressed during the mid-secretory phase and is stimulated by progesterone, which peaks during this phase (Taylor et al. 1998; Taylor, Olive, and Arici 1999). In mice, *Hoxa10* knockout results in infertility due to decidualisation and implantation failure (Satokata, Benson, and Maast 1995). *HOXA10* fails to be upregulated in women with endometriosis in the mid-secretory phase (Taylor, Olive, and Arici 1999; Wu et al. 2005). In line with this lack of upregulation, the CGIs within the *HOXA10* gene are hypermethylated in eutopic tissue of women, mice and baboons with endometriosis (J. J. Kim et al. 2007; B. Lee, Du, and Taylor 2009; Wu et al. 2005). This could explain the decrease in expression of *HOXA10* in the endometrium of patients with endometriosis, even though the affected CGIs are downstream of the gene promoter but may contain critical enhancer elements. Kulp et al. found the opposite, that *HOXA10* was hypomethylated in patients with endometriosis, although it is worth noting that for previous studies the menstrual cycle was not taken into account and for this report the patients had already been treated for endometriosis (Kulp, Mamillapalli, and Taylor 2016).

Endometriosis has been associated with progesterone resistance and with lower levels of PGR expression, specifically the PGR isoform B (PGR-B) (Attia et al. 2000). Even in eutopic tissue of endometriosis patients, PGR expression levels are far more variable than in controls (Wölfler et al. 2016). The promoter of PGR-B is partially methylated in both the ectopic and eutopic endometrium from patients with endometriosis compared to eutopic controls (Wu et al. 2006). A mouse model of endometriosis (surgical implantation of endometrial tissue in ectopic locations) revealed that lesions gained methylation at *PGR* and *HOXA10* but this was suppressed if the mice were treated with a DNMT inhibitor (Y. Li et al. 2016). The increased methylation status of *PGR* is one possible mechanism for progesterone resistance in endometriosis.

Global DNA methylation profiles were able to separate out endometrium from endometriotic lesions in unsupervised hierarchical clustering approaches (Rahmioglu et al. 2017). There were a number of differentially methylated CpGs, some of which at promoters, between endometriotic lesions and eutopic tissue (Borghese et al. 2010; Rahmioglu et al. 2017). The genes associated with differential methylation were enriched for WNT signalling, angiogenesis and cadherin signalling (Rahmioglu et al. 2017). Expression data revealed that a number of differently methylated promoters inversely correlated with expression changes (Borghese et al. 2010). Yet these profiles did not detect changes at the *PGR* or *HOXA10* loci.

**Table 1.4 DNA methylation in Endometriosis.** Summary of the literature with regards to DNA methylation in Endometriosis. M= Menstrual phase, P = Proliferative phase, S= Secretory phase, ES = Early Secretory phase, MS = Mid Secretory phase, LS = Late Secretory phase, NA = Not Available

| Paper | Model System | Number of samples | Menstrual cycle | Loci or Global | Technique | BS treatment | Comments |
|---|---|---|---|---|---|---|---|
| Wu et al. 2005 | Human | Controls= 4, Endometriosis = 6 | Histology | *HOXA10* | MSP and Bisulphite sequencing | Yes | Increase in methylation in endometriosis |
| Wu et al. 2006 | | Controls= 4 (M= 2, P= 2), Endometriosis Ectopic= 8 (M=1, P=6, S=1), Endometriosis Eutopic= 6 (P=5, S=1) | Histology | *PGR-B* | MSP and Bisulphite sequencing | Yes | Heterogensous methylation in endometriosis |
| Kulp et al.2016 | | Controls = 7, Endometriosis Eutopic = 27 | NA | *HOXA10* | Sequenom MassARRAY | Yes | Decrease in methylation in endometriosis |
| Kim et al. 2007 | Baboon or Mouse | Control Eutopic baboons= 6, Endometriosis Eutopic baboons= 6 | MS, E2 levels | *HOXA10* | MSP | Yes | Increase in methylation in endometriosis |
| Lee et al. 2009 | | Control mice = 7, Surgical endometriosis mice= 8 | NA | *HOXA10* | MSP and Bisulphite sequencing | Yes | Increase in methylation in endometriosis |
| Li et al. 2016 | | Control mice =6, Surgical endometriosis ectopic mice= 6 | NA | *HOXA10* and *PGR* | EpiTect Methyl II Assay | No | Increase in methylation in endometriosis |
| Borghese et al. 2010 | Human | Endometriosis Eutopic= 20 Endometriosis Ectopic= 20 (paired) | Luteal phase, Histology | Global | MeDIP on Affymetrix GeneChip Human Promoter 1.0R | No | 229 - 108 differentially methylated regions depending on severity of endometriosis |
| Naqvi et al. 2014 | | Controls= 6, Endometriosis Eutopic= 7 | NA | Global | Infinium HumanMethylation27k | Yes | 120 differentially methylated genes |
| Houshdaran et al. 2016 | | Controls = 16 (P= 6, ES= 5, MS= 5) , Endometriosis Eutopic = 17 (P= 4, ES= 7, MS= 6) | Histology, E2 and P4 levels | Global | Infinium HumanMethylation27k | Yes | 234 differentially methylated CpGs |
| Saare et al. 2016 | | Control= 24 (M= 1, P = 3, ES= 1, MS= 17, LS= 2), Endometriosis Eutopic=31 (M= 4, P = 2, ES= 7, MS= 9, LS= 9) | Self reported history and LH surge | Global | Infinium HumanMethylation450k | Yes | 28 differentially methylated regions |
| Rahmioglu et al. 2017 | | Controls = 8, Endometriosis Eutopic= 8, Endometriosis Ectopic= 8 | Self reported, phase NA | Global | Infinium HumanMethylation450k | Yes | 3,915 differentially methylated probes between paired eutopic and ectopic |
| Dyson et al. 2014 | Cell Culture | Stromal cell culture from control and endometriosis: untreated= 6, post in vitro decidulisation= 6 | NA | Global | Infinium HumanMethylation450k | Yes | 42,248 differentially methylated CpGs |

While it is perhaps not too surprising that ectopic lesions differ in their methylation profile from eutopic control tissue, it may be more meaningful to investigate whether the eutopic endometrium of patients already has some pre-existing epigenetic changes that may predispose the cells to be more adhesive or tolerant to a foreign tissue environment. Some studies have indeed looked at methylation changes associated with eutopic endometrium from patients with or without endometriosis (Table 1.4, Houshdaran et al. 2016; Naqvi et al. 2014). These show that there are a number of differentially methylated CpGs, which are associated with expression changes of the associated gene (Houshdaran et al. 2016; Naqvi et al. 2014). These included genes related to inflammation and the immune response, cell cycle and steroid hormone response. The majority of differential methylation occurred in the mid-secretory phase compared to the early secretory or proliferative phases (Houshdaran et al. 2016). How robust and common these changes are in other patient cohorts remains to be established, as only 4 to 7 samples were used per group in this study.

Despite these changes in DNA methylation status at specific loci, no consistent change in DNMT expression levels could be detected (Borghese et al. 2010; Dyson et al. 2015; van Kaam et al. 2011; Wu et al. 2007). This goes in line with the wide spectrum of findings even across the normal menstrual cycle. At the same time, since the methylation changes involve an almost equal distribution of hypo- and hypermethylated sites, substantial global differences in DNMT or TET levels would not be expected.

As far as I am aware there are no studies on putative changes in 5hmC level or distribution in endometriosis samples that have been conducted so far. This will be an interesting avenue to investigate, alongside the screening for 5mC changes, to examine if endometriosis may result in alterations in these epigenetic modifications in eutopic tissue.

## 1.4 Project Aims

The main aim of the project was to gain a deeper understanding of the epigenomic profile of the endometrium during the secretory phase, i.e. during the critical time window for a successful implantation, in eutopic endometrium of women with endometriosis and endometrium from controls. To achieve this, I took the approach of assessing multiple epigenetic modifications in the same tissue samples, as this will enable me to gain an integrated view of any potential epigenetic changes that may be connected to the disease. Since many of the epigenetic modifications act in an interconnected manner, this approach is more powerful than the investigation of any of these modifications in isolation. The approach was aimed to establish directly comparable epigenomic profiles and to analyse them for alterations that correlate with disease state.

My specific project aims are to:

- Establish genome-wide profiles for the histone modifications H3K4me3, H3K9me3 and H3K27me3, as well as for DNA methylation and hydroxymethylation, from secretory phase-timed endometrial biopsies of control patients and patients with endometriosis.
- Determine if any of these profiles are altered in the eutopic endometrium of women with endometriosis.
- Determine if any individual modification, or the combinatorial status of different modifications at multiple sites, may be used to establish an epigenetic biomarker panel for discriminating endometriosis that hence may aid the diagnosis of endometriosis in the future.

# Chapter II

# Materials and Methods

## 2.1 Human Endometrial Biopsies and Organoids

A total of 20 human endometrial biopsies from women attending the Implantation Clinic at the University Hospitals Coventry and Warwickshire (UHCW) NHS Trust, Coventry, UK were provided by our collaborator Prof. Jan Brosens. Research was undertaken with NHS National Research Ethics Committee approval (1997/5065) and under the approved risk assessment number 15/02HT MH. Biopsies were timed between 6 to 9 days post luteal hormone surge (LH) and were obtained using a Wallach Endocell endometrial sampler. Patients had not received hormonal treatment for at least 3 months prior to undergoing the biopsy procedure. Samples were immediately sub-divided for use in multiple applications 1) snap-frozen, 2) stored in RNA-later (Sigma, R0901) and 3) formalin fixed followed by embedding in paraffin (Figure 2.1). Relevant patient information and details can be found in Table 2.1.

A total of four stromal and four epithelial cell organoids were derived from primary cell cultures established in the laboratory of Prof. Jan Brosens at the University of Warwick. Organoids were stored in RNA-later (Sigma, R0901) and frozen at -80$^{o}$C for future use.

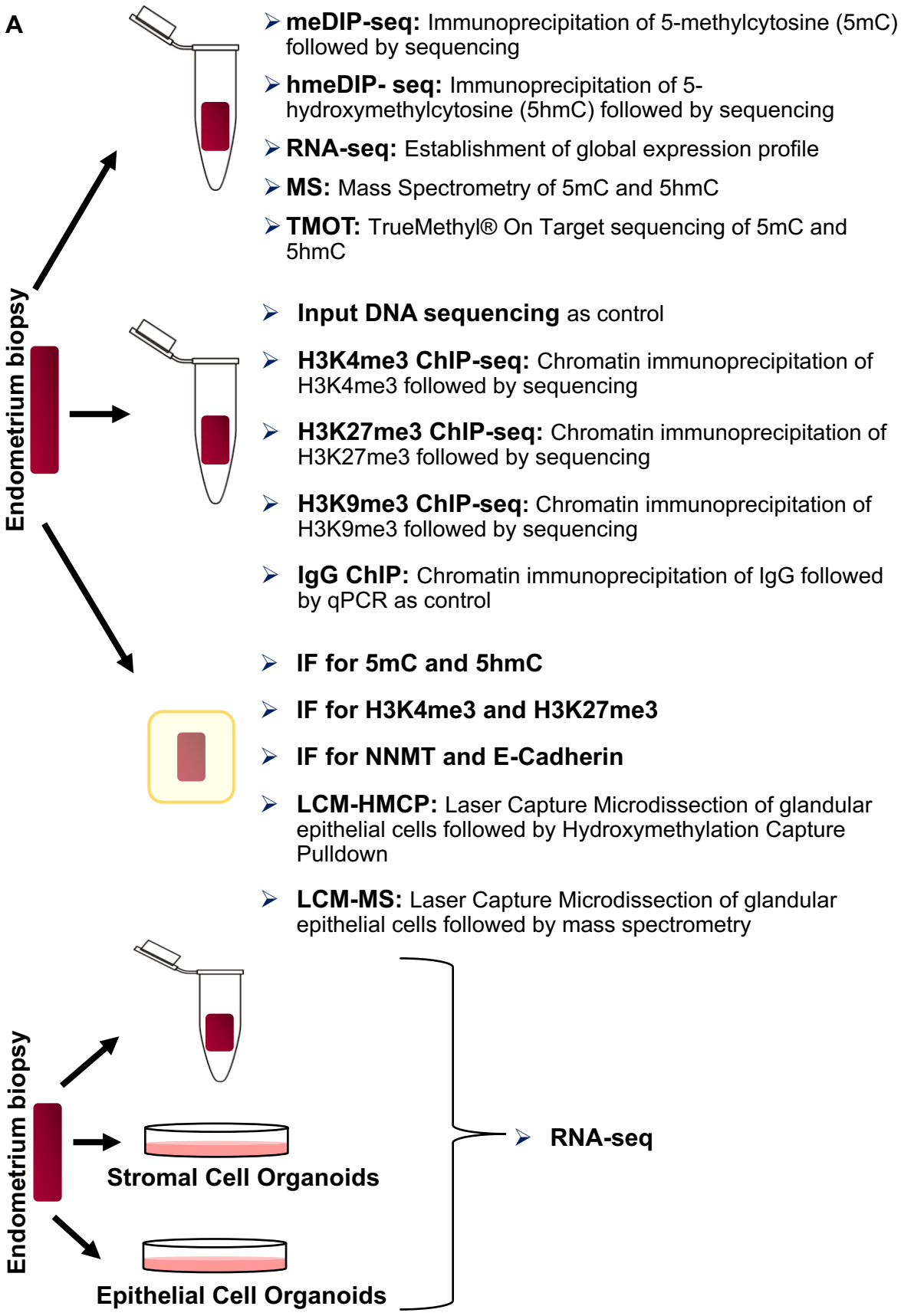



**Figure 2.1 Experimental Schematic of the workflow. A,** Endometrium biopsy and organoid experimental design schematic for global epigenetic profiling**.**

**Table 2.1 Human Endometrial Tissue Information:** Patient information was provided by Warwick Medical School. Abbreviations: (R)IVF – (Recurrent) In vitro Fertilisation, MC – Miscarriage, PCOS – Polycystic ovary syndrome, ICSI – Intracytoplasmic Sperm Injection, (R)IUI – (Recurrent) Intrauterine Insemination, PUL – Pregnancy of Unknown Location. Colour Key: green – sample used for analysis, red – sample removed due to availability (n/a) or technical reasons (No). Unpaired two-way t-tests were non-significant between the Control and the Endometriosis groups: Age = 0.6670, BMI = 0.1806, Day of biopsy = 0.3306. GW: Genome-wide.

| | Sample Information | | | | | GW analysis | | | | | | Immunofluorescence | | Laser Capture Microdissection (LCM) | | Mass Spec Whole Tissue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Age | BMI | Biopsy stage | Patient notes | Endometriosis Grade | ERA defined Receptivity | RNA-seq | meDIP/ hmeDIP | H3K4me3 | H3K9me3 | H3K27me3 | 5(h)mC | H3K4me3+ H3K27me3 | LCM-HMCP | LCM-MS | |
| CONTROL GROUP | 31 | 21.9 | LH+8 | 1x IVF(MC), 1x ICSI, PCOS | n/a | Receptive | Yes | Yes | Yes | Yes | Yes | n/a | n/a | n/a | n/a | Yes |
| | 31 | n/a | LH+9 | Fibroid myomectomy | n/a | Receptive | Yes | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes |
| | 38 | 21 | LH+6 | RIVF | n/a | Non-receptive | Yes | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes |
| | 38 | 23.3 | LH+9 | RIVF | n/a | Receptive | Yes | Yes | Yes | Yes | Yes | Yes | Yes | n/a | n/a | Yes |
| | 37 | n/a | LH+9 | 1x IVF failure | n/a | Receptive | Yes | No | Yes | No | No | Yes | Yes | n/a | n/a | No |
| | 33 | 23 | LH+7 | 1x IVF failure | n/a | Receptive | Yes | No | Yes | Yes | Yes | Yes | Yes | No | Yes | No |
| | 32 | n/a | LH+7 | 1x ICSI(MC) | n/a | Non-receptive | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes |
| | 32 | 20 | LH+8 | 2x IVF failure | n/a | Non-receptive | Yes | Yes | Yes | Yes | Yes | Yes | Yes | n/a | n/a | Yes |
| | 38 | 20 | LH+8 | RIVF | n/a | Non-receptive | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes |
| | 37 | 23 | LH+9 | 2x IVF: 1x biochemical loss, 1x no transfer; male factor | n/a | Receptive | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes |
| ENDOMETRIOSIS GROUP | 34 | 21.3 | LH+7 | Dysmenorrhea. 3x IUI, 2x IVF | Severe | Receptive | Yes | Yes | No | No | No | Yes | Yes | n/a | n/a | Yes |
| | 28 | 18 | LH+8 | | Mild | Receptive | Yes | Yes | Yes | Yes | Yes | n/a | n/a | n/a | n/a | Yes |
| | 39 | n/a | LH+8 | Adenomyosis. RIVF | Severe | Receptive | Yes | Yes | Yes | Yes | No | Yes | Yes | n/a | n/a | Yes |
| | 34 | 25 | LH+9 | Dysmenorrhea. RIUI | Mild | Receptive | Yes | Yes | Yes | Yes | Yes | No | Yes | No | Yes | Yes |
| | 33 | 22 | LH+9 | RIVF | Severe | Receptive | Yes | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes |
| | 39 | 26 | LH+6 | Dysmenorrhea, Dermoid oopherectomy. 5x RIVF | Moderate | Non-receptive | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| | 39 | 38 | LH+6 | RIVF | Severe | Non-receptive | Yes | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes |
| | 31 | n/a | LH+6 | Dysmenorrhea. RIVF | Severe | Non-receptive | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes | Yes |
| | 38 | 26 | LH+8 | RIVF, 2x PUL | Severe | Receptive | Yes | Yes | Yes | No | Yes | Yes | Yes | n/a | n/a | Yes |
| | 39 | 31 | LH+8 | RIVF | Severe | Receptive | No | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes |

## 2.2 DNA and RNA extraction

Whole tissue biopsies were stored in RNA-later (Sigma, R0901) and frozen at -80˚C. Extraction of genomic DNA and RNA was carried out using the Allprep DNA/RNA Mini Kit (Qiagen, 80204). The tissue was disrupted using a P1000 pipette followed by a 20G needle in 1ml of RLT Plus buffer with β-mercaptoethanol (β-ME) according to the manufacturer's instructions. DNA and RNA was quantified using the Nanodrop spectrophotometer (Thermo Scientific) and the quality checked by running on a 1% agarose gel.

Extraction of genomic DNA and RNA from the organoids was carried out using the Allprep DNA/RNA Micro Kit (Qiagen, 80284) following the manufacturer's instructions.

RNA extraction from HEC1A cells was performed using TRI Reagent (Sigma, T9424) following the manufacturer's instructions.

## 2.3 RNA-sequencing

mRNA was isolated from purified total RNA using OligodT Dynabeads (Thermo, 61002). PolyA$^+$ mRNA was used to generate indexed (Illumina), strand-specific libraries using the ScriptSeq V2 RNA-Library Preparation Kit (Epicentre, SSV21106) as per manufacturer's instructions, which were modified to use reaction volumes reduced to one quarter. Libraries were purified using 0.9x volume of Agencourt AMPure XP beads (Beckman Coulter A63881) and eluted in nuclease-free water. Libraries were quantified using the BioAnalyzer 2100 System (Agilent) and KAPA Library Quantification Kit (KAPA Biosystems, KK4824).

Indexed libraries were pooled and run on an Illumina HiSeq2500 instrument using a 100bp single-end sequencing program. Raw Fastq data were mapped to *Homo sapiens* GRCh38 genome assembly using TopHat by the Babraham Bioinformatics Department prior to further analysis.

## 2.4 Native histone chromatin immunoprecipitation and sequencing (ChIP-seq)

Snap-frozen whole tissue biopsies were used for ChIP-seq experiments (Figure 2.1). Histone ChIP was performed on approximately 3mm$^3$ of endometrial tissue. Digestion buffer (50mM Tris-HCL pH8, 1mM $CaCl_2$, 0.2% Triton X-100) and MNase (Sigma, N-5386-200U) was added to the tissue, incubated at 37˚C for 12 mins followed by the addition of stop buffer (110mM Tris-HCl pH8, 55mM EDTA) and placed on ice. The sample was then sonicated for 1 min on "high" setting using a BioRuptor (Diagenode). Following this, IP buffer (280mM NaCl, 1.8% Triton X-100, 0.2% SDS, 0.2% Na-Deoxycholate, 5mM EGTA) with protease inhibitor (Roche 5056489001) was added and then centrifuged at 14,000rmp for 15 mins. One tenth of the chromatin was removed as Input sample and Proteinase K added, followed by incubation at 55˚C for 1 hour in a thermo-shaker. Purification of DNA was performed using the MinElute PCR Purification Kit (Qiagen, 28004) and then run on BioAnalyzer 2100 System (Agilent) to check chromatin digestion. Equal amounts of pre-washed Protein A and Protein G Dynabeads (Thermo, 10002D; Thermo, 10004D) were added to the remaining chromatin and incubated at 4˚C for 1 hour to pre-clear the chromatin. The supernatant was transferred to clean Eppendorf tubes and split into the desired number of IPs. Appropriate antibody was added: 0.5µg H3K4me3 (Abcam, ab8580), 1µg H3K9me3 (Abcam, ab8898), 2µg H3K27me3 (Millipore, 07-449) or 2µg rabbit IgG (Diagenode, C15410206), and incubated overnight at 4˚C on a rotary mixer. Pre-washed Protein A and Protein G Dynabeads were added to each IP and incubated at 4˚C for 2-3hrs, using a rotary mixer. Beads were washed 5x with RIPA buffer (140mM NaCl, 10mM Tris-HCl pH8, 1%Triton X-100, 0.1% SDS, 0.1% Na-Deoxycholate, 1mM EDTA) plus protease inhibitor, followed by a wash with LiCl buffer (250mM LiCl, 10mM Tris pH8, 0.5%NP-40, 0.5% Na-Deoxycholate, 1mM EDTA). Beads were resuspended in 100µl TE buffer (10mM Tris-HCl, 1mM EDTA) plus 2.5µl Proteinase K and incubated at 55˚C for 1 hour in a thermo-shaker. DNA was purified with the QIAquick PCR Purification Kit (Qiagen, 28106) according to manufacturer's instructions. ChIP-qPCR was carried out on each histone modification, Input and IgG to check IP efficiency using primers for the following genes; *GAPDH*, *NANOG* and *NDN* (Table 2.2).

**Table 2.2 List of primers used.** A table of primers used for quality control of enrichment analysis for ChIPs as well as primers for RT-qPCR.

| Gene | Direction | Sequence | Purpose |
|---|---|---|---|
| GNAS | F | TCCCCGTCGAGAATGATGGC | meDIP QC |
| | R | GGGCTCCAGCTTCCCTGAAT | |
| OR4F3 | F | ATGTGGCCCTATGTAAGCCCC | meDIP QC |
| | R | TGCCAGTTGGAACAGGGAGT | |
| MUC6 | F | GTTGGTGACTGGAGAGGTGG | hmeDIP QC |
| | R | CGGTCCACTTCTCTCACCAC | |
| OCLN | F | ACTTCAGGCAGCCTCGTTAC | hmeDIP QC |
| | R | CCGCCAGTTGTGTAGTCTGT | |
| GAPDH | F | CAGGCTGGATGGAATGAAAG | ChIP QC |
| | R | AAAGGCACTCCTGGAAACCT | |
| NANOG | F | GAGATGGGCACGGAGTAGTCTT | ChIP QC |
| | R | TGTCTATCCCTCCTCCCAGGTA | |
| NDN | F | CCACTTCTTGTAGCTGCCGA | ChIP QC |
| | R | TGTGGTACGTGCTGGTCAAG | |
| GAPDH | F | CGCTGAGTACGTCGTGGAGT | RT-qPCR |
| | R | GGGCAGAGATGATGACCCTTT | |
| PAEP | F | GCTGCTCGATACTGACTACGA | RT-qPCR |
| | R | AAGCCCTGATGAATCCCTGC | |
| GPX3 | F | GGTCTGGTCATTCTGGGCTT | RT-qPCR |
| | R | TAGGGACAAAGCCTCCACCT | |
| NNMT | F | CAGAGGCCTTTGACTGGTCC | RT-qPCR |
| | R | GCTGGCTCTGAGTCACATCA | |
| GP2 | F | AAGTCCGCAGTGAAGTACCG | RT-qPCR |
| | R | CCATTCATGACACCGGGAGA | |
| KCNG1 | F | GTCAACCTCTCCGTCAGCAC | RT-qPCR |
| | R | CGATGAAGACGTTGTGGCAC | |
| GJB6 | F | ACACTGGCAGTCGTTGGAAA | RT-qPCR |
| | R | CGATGCTGGTGGAGTGTTTG | |
| FAM20C | F | TCACACGGGACAAGAAGCTC | RT-qPCR |
| | R | CTCGATCTGGTCTGGCTTCC | |
| TNC | F | ATCAGTCACCGGTTACCTGC | RT-qPCR |
| | R | CTGTGTAGTGGGTGGATGGG | |
| PLA2G4E | F | GATGCCTGGAACCTGTCACA | RT-qPCR |
| | R | GGGATTTCAGGCAGGATCGG | |
| SYT2 | F | ATGAACACAGTGGACCTCGG | RT-qPCR |
| | R | ATGCAGACAGTGAGCTTCCC | |
| DNMT1 | F | GGCTATCAGTGCACCTTCGG | RT-qPCR |
| | R | GGGAACAGAGGGAGCTTCTC | |
| DNMT3A | F | GCCAAAGAAGTGTCAGCTGC | RT-qPCR |
| | R | CCATGCTCCAGACACTCCTG | |

| | | | |
|---|---|---|---|
| DNMT3B | F | TGCCATCAAAGTTTCTGCTGC | RT-qPCR |
| | R | CAGTCCTGCAGCTCGAGTTT | |
| DNMT3L | F | CTGACCCGGGACAACTGAAG | RT-qPCR |
| | R | CGCCGTACACAAGATCGAAG | |
| TET1 | F | CTGCTTGCCTGGACTTCTGT | RT-qPCR |
| | R | GGAATAACACCCAAAGAGCGG | |
| TET2 | F | CATGCCCACAGAGACTTGCA | RT-qPCR |
| | R | TGAAGCTGCTCATCCTCAGG | |
| TET3 | F | CGACACCCTCCGGAAGTATG | RT-qPCR |
| | R | ACCAAAGGAGAAGGAGGCAC | |
| HMBS | F | AGGAGTTCAGTGCCATCATCCT | RT-qPCR |
| | R | CACAGCATACATGCATTCCTCA | |

Histone ChIP libraries for H3K4me3, H3K9me3, H3K27me3 and Inputs were generated using the NEBNext Ultra II DNA Library Prep Kit (Illumina, E7645) following the manufacturer's instructions. In brief, the ChIP DNA was added to the end repair reaction mix and incubated at 20˚C for 30 mins, and then at 65˚C for a further 30 mins. The ligation mix and adaptors were then added to the reaction and incubated at 20˚C for 15 mins. Agencourt AMPure XP beads (Beckman Coulter, A63881) were used to perform a double size selection of the libraries, by first incubating with 0.6x volume of beads, and then 0.45x volume of beads and finally eluted in TE. Libraries were then amplified using Sanger 8 base pair indexed oligos consisting of 10µM PE1.0 PCR primer and 10µM iPCR tag and purified using AMPure beads and quantified as previously outlined.

Indexed libraries were pooled for each histone ChIP/Input and sequenced using a 50bp paired-end sequencing program on an Illumina HiSeq2500 instrument. Raw Fastq data were mapped to *Homo sapiens* GRCh38 genome assembly using Bowtie2 by the Babraham Bioinformatics Department.

## 2.5 High-throughput sequencing of Methylated or Hydroxymethylated Immunoprecipitated DNA (meDIP-seq and hmeDIP-seq)

4µg of purified DNA was sonicated to 200-700bp fragments using a BioRuptor (Diagenode). Sonicated DNA was run on a 1% agarose gel to check size distribution and quantified on a Nanodrop spectrophotometer (Thermo Scientific).

The NEBNext DNA Library Prep Kit (Illumina, E6040), was used for library preparation. Manufacturer's instructions were followed with some adjustments. In brief, 3µg of sheared DNA was used in the end repair reaction and then purified using Agencourt AMPure XP beads (Beckman Coulter, A63881), without elution from the beads. dA-Tailing reaction mix was added to beads and incubated at 37˚C for 30 mins, followed by purification with 20% PEG 8000 in 1.35M NaCl. Adaptor ligation reaction mix was added to the beads and incubated at 20˚C for 15 mins, followed by purification with PEG 8000 in 1.35M NaCl and elution in warm Elution buffer (EB). Quantification of adaptor-ligated DNA was carried out using a Nanodrop spectrophotometer.

meDIP-seq or hmeDP-seq was performed on the adaptor-ligated DNA from above so that either 5mC- or 5hmC-containing fragments could be isolated for IP and further amplification. Three IPs of 500ng were set up for each sample. The samples were denatured at 99˚C for 10 mins and then placed in ice for 10 mins. IP buffer and 1.25µl 5mC antibody (Eurogentec, BI-MECY-0100) or 1µl 5hmC antibody (Active Motif, 39769) was added, incubated at 4˚C for 2hrs on a rotary mixer followed by a further 2 hours with pre-blocked Dynabeads. Dynabeads were then washed three times with IP buffer followed by incubation at 55˚C for 30 mins with digestion buffer and proteinase K. The triplicate IPs were then pooled and purified using the MinElute PCR purification Kit (Qiagen, 28004) as per manufacturer's instructions. IP efficiency was checked by qPCR of genomic regions known to be DNA methylated or hydroxymethylated, with known unmodified regions as control (Table 2.2). IPs were then amplified using the NEBNext DNA Library Prep Kit, PE PCR primer 1.0 and an iPCR tag primer was used to generate indexed libraries. The libraries were purified, quantified, sequenced and mapped as previously described in Section 2.4.

## 2.6 Cell culture and transfection

HEC1A cells, an adenocarcinoma cell line derived from the endometrium, were kindly provided by Prof. Irmgard Classen-Linke (Institute for Molecular and Cellular Anatomy, RWTH Aachen University) and are available from the American Tissue Culture Collection. HEC1A cells were cultured in standard conditions: Endometrial Epithelial Cell (EEC) media (10% steroid free fetal calf serum (CCPro, S-1-M), 90% phenol red-free DMEM/F12 base media (Life Technologies, 21041-025), 1.25x GlutaMax (Life

Technologies, 35050-038) and 1x Antibiotic-Antimycotic (Life Technologies, 15240-062)). Cells were cultured under standard condition of 5%$CO_2$ at 37˚C.

Cells were plated 24hrs before transfection, this meant the cells would be in a log growth phase which promotes a good transfection frequency. Cells in a 6-well plate were transfected with 2.5µg plasmid DNA containing a eukaryotic promoter and the human NNMT cDNA sequence (GeneScript, OHu08895D) and 5µl Lipofectamine 2000 (Life Technologies, 11668), both were prepared in Opti-MEM1 (31985-047) and mixed in a 1:1 ratio. After 24hrs the transfection media was changed to standard EEC media. After 3 days cells were fixed, see section 2.7.2.

## 2.7 Immunofluorescence (IF) staining

### 2.7.1 IF staining of paraffin-embedded human endometrial biopsies

Samples were collected as outlined in Figure 2.1 and sections were cut at 3µm. Sectioned samples were provided by Dr. Katherine Fishwick in the group of Prof. Brosens at the University of Warwick.

Sections were deparaffinised by passing through a series of organic solvents as follows; 2x Xylene 10 mins, 1x 100% Ethanol 10 mins, then 3 mins in each of the following 100% Ethanol, 70% Ethanol, 50% Ethanol, 30% Ethanol and PBS. Slides were boiled in Antigen Retrieval Buffer (10mM Tris-HCl pH9.5, 1mM EDTA, 0.05% Tween-20) and cooled. Following this, slides were washed in Wash Buffer (300mM NaCl, 50mM Tris-HCl pH7.5, 0.1% NP-40) and then blocked in PBT/BSA (PBS, 0.5% BSA, 0.1% Tween-20) for 30 mins. Slides were then incubated overnight at 4˚C in primary antibody that had been diluted in PBT/BSA. Antibody dilutions were as follows: H3K4me3 1:500 (Abcam, ab8580), H3K27me3 1:500 (Active Motif, 61017), NNMT 1:250 (Abcam, ab119758), E-Cadherin 1:400 (Cell Signalling Technology, 3195). Sections were then washed in PBS and incubated with the appropriate secondary antibody (Alexa Fluor 568 donkey anti-Mouse IgG (H+L) (Life Technologies, A10037), Alexa Fluor 488 donkey anti-Rabbit IgG (H+L) (Life Technologies, A21206) for 1 hour. Slides were washed in PBS and then counterstained with DAPI followed by a further PBS wash. Coverslips were mounted in 50% glycerol diluted in PBS, and then sealed with nail varnish.

For 5mC and 5hmC stainings the following alternative steps were added: After slides were washed with Wash Buffer they were incubated in 2N HCl for 30 mins. Slides were washed again in Wash Buffer before being blocked in PBT/BSA for 30 mins. Slides were then incubated overnight at 4˚C in primary antibody that had been diluted in PBT/BSA. Antibodies dilutions were as follows: 5mC antibody 1:250 (Eurogentec, BI-MECY-0100), 5hmC antibody 1:4000 (Active Motif, 39769). The next day, slides were washed in PBS and incubated with biotinylated antibody, goat anti-Mouse IgG, Human ads-BIOT 1:200 (SouthernBiotech, 1030-08), for 2 hours. Slides were washed in PBS followed by incubation with secondary antibodies, Streptavidin Alexa Fluor 568 conjugate (Life Technologies S11226), Alexa Fluor 488 donkey anti-Rabbit IgG (H+L) (Life Technologies, A21206), for 1 hour at room temperature (RT). Counterstaining was performed with DAPI followed by a further PBS wash, before mounting in 50% glycerol diluted in PBS and sealing with nail varnish.

### 2.7.2 IF staining of cell lines

HEC1a cells were cultured on coverslips then washed in PBS and fixed with 2% PFA in PBS for 30 mins at RT. Fixed cells were washed in PBS and permeabilised using PBS, 0.5% Triton X-100 for 30 mins at RT. This was followed by a wash in PBT (PBS, 0.05% Tween) and blocking in PBT/BSA (PBS, 0.05% Tween, 1% BSA). Coverslips were incubated with primary antibody which was diluted in PBT/BSA and then incubated with secondary antibody.

IF staining for hydroxymethylation and methylation continued with these additional steps. Following first staining coverslips were fixed with 2% PFA in PBS for 10 mins. Then they were incubated in 2N HCl, 0.05% Triton X-100 for 40 mins at 37°C. Coverslips were blocked in PBT/BSA and incubated with the primary antibody for 1 hour followed by incubation with secondary antibody. The nuclei were counter-stained with DAPI and the coverslips mounted with 50% glycerol diluted in PBS onto slides.

Antibodies dilutions were as follows: FLAG M2 1:200 (Sigma-Aldrich, F1804), NNMT 1:200 (Abcam, ab119758), H3K4me3 1:250 (Abcam, ab8580), H3K27me3 1:250 (Merck-Millipore, 07.499), 5mC antibody 1:250 (Active Motif 61255), 5hmC antibody 1:250 (Active Motif, 39769).

### 2.7.3 IF imaging and image analysis

Images were captured using a Nikon A1R or Olympus BX041 microscope and analysed with CellProfiler v2.2.0 (Carpenter et al. 2006), adapting scripts generated at the Babraham Imaging department by Dr. Hanneke Okkenhaug. In brief, nuclei positions were determined using DAPI, and then the corresponding intensities of specific epitopes measured. Where applicable, the cell type was established either by tissue context and manual selection, or by the presence of E-Cadherin staining to identify epithelial cells. Mean intensity values were exported into Excel and used for further calculations. For human biopsies, the ratio of epithelial to stromal cell fluorescence intensities was calculated per image. For HEC1A cells nuclear intensity measurements were filtered based on the mean intensity of the FLAG staining: FLAG-high >0.15, FLAG-low <0.05, and the corresponding mean intensity values of specific epitopes were used. Statistical tests were performed in GraphPad PRISM 7.00.

## 2.8 Laser Capture Microdissection (LCM)

Samples were collected as outlined in Figure 2.1 and paraffin-embedded tissue sections were cut at 10µm onto MembraneSlide 1.0 PEN (Zeiss, 41590-9041-000). Sectioned samples were provided by Dr. Katherine Fishwick in the group of Prof. Brosens at the University of Warwick.

Two to three slides were used per a sample. Slides were deparaffinised to 70% ethanol as previously described (Section 2.7.1). The Zeiss MicroBeam microscope was used to undertake the LCM, with manual selection of cells. DNA extraction was accomplished using the QIAamp DNA Micro Kit (Qiagen, 56304), as per manufacturer's instructions with the following notes: the Proteinase K step was carried out at 56˚C for 17-18 hours and the DNA was eluted in 20µl of AE Buffer which was provided in the kit. Quality and quantity of DNA was established using a Qubit 3.0 Fluorometer (Invitrogen) and/or 4200 TapeStation (Agilent).

## 2.9 Hydroxymethylation Capture Pulldown (HMCP)

Genomic DNA was isolated from samples as described above (Section 2.8) and then sonicated in 15µl microtubes (microTube-15 AFA Beads Screw-Cap PN520145) to generate 150bp fragments using the Covaris M220 Focus with the following setting; Peak Power -30, Duty Factor -20, Cycles-50. Sonication efficiency was confirmed on the TapeStation 4200 (Agilent).

The HMCP protocol was carried out during my iCASE studentship placement at Cambridge Epigenetix. The protocol was developed by and the reagents provided by Cambridge Epigenetix. In brief, sonicated DNA and spike-in control DNA were added to the End Repair and A-Tailing reaction, which was incubated at 20˚C for 30 mins followed by 65˚C for 30 mins. 3µM of NEXTflex DNA Barcodes (Bioo Scientific, NOVA-514103) and Ligation mix were added to the reaction and incubated at 20˚C for 45 mins. Following this, the DNA was purified using 0.8x magnetic bead volume and an aliquot was removed for the Input at this stage. The Input was subsequently re-indexed with ILMN indexing oligos and amplified in the same step. The Input library was then purified using 0.9x magnetic bead volume.

To the remaining adaptor-ligated DNA from above, the primers and extension buffer were added. This mix was incubated at 95˚C for 30 mins, and then allowed to cool at a rate of 0.1˚C/s to 14˚C. The extension enzyme was added and incubated at 37˚C for 30 mins followed by 75˚C for 20 mins. Following this, modification reagent 1 and enzyme were added to the reaction and incubated at 37˚C for 30 mins, then modification reagent 2 was added and retuned to incubate at 37˚C for a further 2hrs. Following the incubation, carrier DNA was added and then the DNA was purified using Micro Bio-Spin columns with Bio-Gel P30 (BioRad) that had been equilibrated with BB1 buffer. Blocked streptavidin beads were added to the purified DNA and incubated at 22˚C on a thermo-shaker for 10 mins. The supernatant was discarded and the beads were washed 3x with BB1 buffer at 22˚C on a thermo-shaker for 5 mins each, and 3x with BB4 buffer at 55˚C on a thermo-shaker for 5 mins each. The DNA was eluted from the beads in 0.1N NaOH at 22˚C on a thermo-shaker for 10 mins, and afterwards Resuspension buffer was added. The pulled-down DNA was amplified to obtain a library and then purified using 0.9x magnetic bead volume and eluted in library dilution buffer. Quantification was carried out as in Section 2.8.

Indexed libraries were pooled and sequenced using an 80bp paired-end read sequencing program on the Illumina NextSeq500 High Output Sequencer. Raw Fastq data were mapped to *Homo sapiens* GRCh38 genome assembly using bowtie2 by Cambridge Epigenetix.

## 2.10 TrueMethyl® OnTarget (TMOT)

Probes for validation were selected from a number of different analyses conducted on the meDIP- and hmeDIP-seq data. The different analyses used the following bioinformatic 'probe' sets to determine differential methylation/hydroxymethylation across different genomic features: 2kb tiling probes with 1kb overlap, Promoters (TSS -1kb +100bp), exons, CpG Islands, the MEDIPS script in R using default settings with 200bp windows (Lienhard et al. 2014). Previously published data sets were also interrogated with regards to my meDIP- and hmeDIP-seq data, resulting in the expansion of regions to be validated. These data sets were PGR binding sites (GSM1703567) (Mazur et al. 2015), FOXO1 binding sites (GSM1703607) (Vasquez et al. 2015) and differentially methylated CpGs or regions between the early and mid-secretory phase (GSE90060) (Kukushkina et al. 2017). This gave a total of 411 differentially modified regions 3 of which were present in both the hmeDIP- and meDIP-seq data.

Oligonucleotide probes were designed to overlap with the CpGs in these regions for validation. This resulted in the synthesis of 6270 oligonucleotides (two for each CpG). A custom array was ordered from CustomArray (http://www.customarrayinc.com/oligos_main.htm).

The TMOT protocol was carried out during my iCASE studentship placement at Cambridge Epigenetix and using a protocol and reagents developed by Cambridge Epigenetix.

The TrueMethyl® Array Kit was used to perform oxidative bisulphite (oxBS) and bisulphite (BS) treatment. In brief; 100µl of MMBS1 (which contains beads) was added to 500ng of genomic DNA and incubated for 20 mins at RT. The supernatant was removed and the beads were washed 3x with 80% Acetonitrile. The beads were air-dried then resuspended in 20µl of denaturing solution for 5 mins. The supernatant was

split into two clean tubes: 1xOxidation reaction mix and 1xBisulphite reaction mix. To the Oxidation reaction 1µl of Oxidant was added and incubated for 10 mins at 40˚C. Bisulphite conversion was carried out on the Oxidant and Bisulphite reactions by the addition of 30µl of Bisulphite Reagent Solution. Both were then incubated according to the following programme: 95˚C for 5 mins, 60˚C for 20 mins, 95˚C for 5 mins, 60˚C for 40 mins, 95˚C for 5 mins, 60˚C for 45 mins, then hold on 20˚C. OxBS and BS prepared DNA was cleaned up using 160µl MBBS2 beads and desulfonation buffer then eluted in 20µl of ultra-pure water and quantified with Qubit 3.0 Flurometer ssDNA (Invitrogen).

To the oxBS or BS prepared DNA, extension buffer and AD1 was added to the purified DNA and incubated at 95˚C for 1 min. The reaction was allowed to cool at a rate of 0.1˚C/s to 14˚C. Following this, additional extension buffer was added along with strand synthesis dNTPs and the extension enzyme; this mix was incubated at 37˚C for 10 mins. Extension buffer, ATP, AD2 and Ligation enzyme were added to the reaction mix and incubated at 25˚C for 5 mins. Samples were then purified using BB3 beads and eluted in ultra-pure water. After purification, enrichment primer mix, oligo 1 and oligo 2 were added and amplification was as follows: 95˚C for 3 mins, [98˚C for 20 secs, 60˚C for 20 secs, 72˚C for 1 min] x 12, 72˚C for 5 mins, then hold at 4˚C. Samples were purified using BB3 beads and eluted in 10µl of Ultra-pure water. These were then quantified with the Qubit 3.0 Fluorometer ssDNA (Invitrogen) instrument. The concentration of BS samples was correspondingly matched to the OxBS counterpart samples. The oligonucleotide probes from CutomArray along with hybridisation buffer and hybridisation additive were added to the samples and incubated at 95˚C for 3 mins and then at 51˚C overnight. Samples were purified with M-280 Streptavidin beads followed by washes with WB2, WB3 and WB4. Following the washes and with the DNA on the beads, extension buffer, extension enzyme and dNTPs was added and the samples incubated at 37˚C on a thermo-shaker for 30 mins. Samples were then washed with WB5 followed by elution in 50µl ultra-pure water. The samples were then indexed and amplified using the TMWG Index PCR mix followed by a clean-up using Agencourt AMPure XP beads (Beckman Coulter, A63881). The Libraries were quantified using the TapeStation 4200 (Agilent) and Qubit 3.0 Fluorometer (Invitrogen).

Indexed libraries were pooled with 15% Kineococcus, this is to enhance sequencing by compensating for the low cytosine base levels after oxBS and BS treatment. The libraries were sequenced using a Mid-Output 150bp paired-end sequencing program on the Illumina NextSeq500 instrument. Raw Fastq data were mapped to *Homo sapiens* GRCh38

genome assembly using the Bismark script (Krueger and Andrews 2011) developed by the Babraham Bioinformatics department.

## 2.11 RT-qPCR expression analysis

To validate the RNA-sequencing as well as to identify the expression levels of genes associated with receptivity in the HEC1A cell line, reverse transcription followed by quantitative PCR (RT-qPCR) was performed. This entailed 500ng of purified RNA being reverse transcribed (RT) by RevertAid H Minus Reverse Transcriptase (Thermo Scientific, EP0451) as per manual using a mixture of random hexamer primers (Thermo Scientific EP0451) and oligodT18 primers (Fermentas, S0132). The resulting cDNA product was diluted 1:20 using sterile water. To perform the qPCRs, 5µl of each sample was used along with the SYBR green Jump Start Taq Ready mix (Sigma, 4438) on a BioRad CFX96 or CFX384 thermocycler. Primers for gene specific analysis were intron-spanning, and the qPCRs were performed in triplicate for each sample (Table 2.2). Analysis on Ct values was carried out in Excel where they were normalised to *GAPDH*, a housekeeping gene. Statistics were performed in GraphPad PRISM 7.00.

## 2.12 Liquid chromatography- mass spectrometry

This method was used to determine total amounts of 5mC and 5hmC. This method captures all of the highly repeated sequences in the genome which are CpG-rich but owing to their repetitive nature are not included in sequence analyses because they cannot be uniquely mapped. Of note, to overcome this limitation, a specific repeat analysis was carried out on relevant data as described in the section 2.13.5.

In order to determine quantitative levels of 5mC and 5hmC in the samples, 500ng of purified DNA from whole tissue or 22ng of purified DNA from LCM-Epithelial cells was prepared for mass spectrometry using Nucleoside Digestion Mix (New England Bioscience, M0649), according to manufacturer's instructions. An additional acetonitrile precipitation was performed. Liquid chromatography- mass spectrometry was conducted in collaboration with Prof. Petra. Hajkova's group at Imperial Collage London, who have established this ultrasensitive technique of 5mC and 5hmC detection for low

cell numbers (Amouroux et al. 2016). Absolute 5mC and 5hmC amounts were carefully calibrated against standards; all samples were analysed in duplicate with statistics carried out using GraphPad PRISM 7.00.

## 2.13 Bioinformatic analysis

### 2.13.1 Expression profile analysis

RNA-sequencing was carried out on endometrial biopsies, as outlined in the workflow, (Figure 2.1) and the expression profiles were analysed as follows.

Data were visualised and analysed with the RNA-seq quantitation pipeline in SeqMonk software (http://www.bioinformatics.babraham.ac.uk). Data was normalised to RPM (reads per million mapped reads), log2-transformed and percentile normalisation to 75% distribution performed as determined by the cumulative distribution plot. Differential expression was measured using DESeq2 and/or EdgeR as stated, with a p-value threshold of 0.01, adjusted for multiple testing correction using the Benjamini-Hochberg method where stated (Love, Huber, and Anders 2014; Robinson, McCarthy, and Smyth 2009). A minimum differential change of 2.0 was applied to log2-transformed data using SeqMonk to produce a stringent list of differentially expressed genes. Heatmaps and data trees were compiled using hierarchical clustering within the SeqMonk software (http://www.bioinformatics.babraham.ac.uk).

### 2.13.2 Cell-type specific gene expression signatures

RNA-seq data generated from epithelial and stromal cell organoids (Figure 2.1) was used to generate a cell-type specific gene expression signature. Data was visualised and analysed with the RNA-seq quantitation pipeline in SeqMonk software (http://www.bioinformatics.babraham.ac.uk). Data was normalised to RPM (reads per million mapped reads), log2-transformed and percentile normalisation to 75% distribution performed if required. Expression levels above 2 in epithelial organoids and below -2.8 in stromal cell organoids were selected to establish epithelial cell signature genes. Expression levels above 2 in stromal cell organoids and below -3.5 in epithelial cell

organoids were selected to establish stromal cell signature genes. Epithelial cell signature genes were further subdivided giving rise to: *non-receptive epithelial* and *receptive epithelial* cell signatures. This was established by plotting a heatmap of the epithelial cell signature genes in the corresponding whole tissue biopsies, then using the clustering which naturally segregated the list into two to establish the receptivity effect.

### **2.13.3** Gene Ontology analysis (GO)

To evaluate if gene ontology terms were enriched in my differential lists I used the well-known databases: Database for Annotation Visualisation and Integrated Discovery (DAVID), GOrilla and REVIGO (Eden et al. 2009; Huang, Sherman, and Lempicki 2009b, 2009a; Supek et al. 2011). Bonferroni correction was applied and p values < 0.05 are reported. When conducting gene ontology on RNA-seq data, a background list of all genes containing at least 20 raw counts was used. REVIGO and R treemap script was used to create the RNA-seq treemaps (Eden et al. 2009). When conducting gene ontology on the subset of genes that resolve their bivalent to H3K4me3-only, a background list for all bivalently marked genes in the non-receptive state was used.

### **2.13.4** Histone ChIP-sequencing analysis

Analysis of the histone ChIP-seq data was conducted using chromstaR using the full multivariate approach by Dr Sarah Burge (Taudt et al. 2016). A bin size of 1000bp and a step size of 500bp was used, duplicate reads and those with a MAPQ score of lower than 10 were removed. All other parameters remained at the default chromstaR settings. The univariate fits for each sample were checked and found to be satisfactory before proceeding to the multivariate step.

For the individual analysis of the histone ChIP-seq data, MACS2 peaks were called and a FDR cut-off of 0.05 was applied (https://github.com/taoliu/MACS/). In all cases, the control file was the corresponding Input for each sample. Consensus lists of peaks were generated for each condition, i.e. the non-receptive controls, non-receptive endometriosis, receptive controls and receptive endometriosis groups. For each condition the peaks had to be present in all samples of any given group and have at least

three reads. Peaks within 300bp of each other were merged. When comparing between conditions, peaks from all groups involved were pooled to allow consistent and reliable analysis. Differential enrichment was measured using the validated approach of DESeq2 and/or EdgeR as stated with a p-value threshold of 0.01 and adjusted for multiple testing correction using the Benjamini-Hochberg method (Love, Huber, and Anders 2014; Robinson, McCarthy, and Smyth 2009). Scatterplots and wiggle plots were generated using SeqMonk software (http://www.bioinformatics.babraham.ac.uk).

### **2.13.5** Repeat element analysis

Genomic analysis usually regards only those sequences which can be uniquely mapped to the reference genome. However, specialised bespoke software has been developed at the Babraham Institute to handle repeat sequence families retaining the majority of the reads at repeats in the human genome. The repeat element genomic content was analysed by the Babraham Bioinformatics department. In brief, annotations for the human genome (GRCh38) were generated using RepeatMasker and downloaded from the UCSC genome browser web site. The repeat genomes were constructed by concatenating the sequences of each instance of the repeat into a single repeat genome for that repeat family where individual repeats were separated by NNNNN to prevent sequences from mapping over instance boundaries. For telomeric sequence the hexamer repeat (TTAGGG)n extended to 204bp was used. Adapter-trimmed sequencing reads were aligned to repeat genomes of different repeat families using Bowtie2 (default parameters), and hits were counted whereby multiple matches of a sequence were allowed.

### **2.13.6** Motif analysis

Motif analysis was conducted on the bivalent regions that resolved to H3K4me3-only, which were identified in the transition from the non-receptive to receptive state in the histone ChIP analysis.

The Fasta sequences of the regions were run through the UCSC Table Browser (https://genome.ucsc.edu/cgi- bin/hgTables) with masking of repeats to N option

applied. These sequences where then analysed for enrichment of motifs using the Analysis of Motif Enrichment (AME) tool in the MEME Suite with the HOCOMOCO v11 or JASPAR2018 database. As control, a scrambled version of the input sequences was used (McLeay and Bailey 2010).

### 2.13.7 Laser capture microdissection -hydroxymethylation pulldown analysis (LCM-HMCP)

Analysis and visualisation of the LCM-HMCP data of epithelial cells was carried out using SeqMonk software (http://www.bioinformatics.babraham.ac.uk). The data was analysed using probes for CGIs and promoters (TSS-1kb +100bp) and quantified by log2-transformation of read counts corrected to the largest data set. Wiggle plots were produced using SeqMonk software for the specified regions. Genomic features for feature distribution analysis were defined by SeqMonk for GRCh38 with promoters defined as -1kb upstream to +100bp downstream of TSSs. For the analysis of bivalently marked promoters, probes were designed to extend to +/- 5kb of the TSS and were used to create a Probe Trend plot. Statistical tests were carried out using GraphPad PRISM 7.00.

### 2.13.8 meDIP and hmeDIP analysis

The analysis and visualisation of the meDIP and hmeDIP was carried out using the SeqMonk software (http://www.bioinformatics.babraham.ac.uk). Probes were generated in a variety of formats to explore the data fully. These included; 2kb windows with a step size of 1kb, CGIs, exons and promoters (TSS -1kb +100bp). Probes were quantified by log2-transformation of read counts corrected to the largest data set and normalised using 75% distribution as required. Differentially enriched regions were established using the unpaired two-tailed statistical t-test with a p-value threshold of 0.01 and a minimal differential change of 2.0. Heatmaps, PCA plots and data trees were compiled using hierarchical clustering. Feature distribution analysis was performed as in Section 2.13.7.

### **2.13.9** True Methyl® On Target (TMOT) analysis

Bisulphite calls were generated using Bismark (Krueger and Andrews 2011) by the Babraham Bioinformatics department and the data subsequently analysed using methylKit (Akalin et al. 2012) by Dr Sarah Burge, a member of the Hemberger Lab. Samples were filtered on coverage, with a minimum low count of 10 and bases that had more than 99.9 percentile of coverage in each sample discarded to ensure quality of methylation calls. mCpG dinucleotides were present in at least 2 samples out of each group. To produce 5hmC counts, oxBS reads were subtracted from BS reads using the adjustMethylC function. R GRanges/findOverlaps was used to establish CpGs that overlapped with regions of interest and these were filtered by q-value <0.01. The MethylDiff function with a q value cutoff of 0.01 was used to determine the significant differential methylation or hydroxymethylation between the experimental combinations discussed. A region was called as "validated" if a difference occurred in the same direction. Manhattan plots were generated using ggplot in r studio, y-axis: endometriosis and non-receptive = 1, control and receptive = 0. PCA analysis was conducted using the defaults in the methylKit package.

# Chapter III

# Characterising the precise receptivity state of the endometrium

## 3.1 Introduction

Women undergo a ~28 day menstrual cycle comprising the menstrual, proliferative and secretory phases (Figure 1.1). The precise duration of each phase and indeed the total cycle length varies between women. For reproduction to succeed, the secretory phase is the most important stage and within this, the WOI is essential for a blastocyst to implant successfully. The exact timing of the WOI varies between different reports as it is difficult to determine, but an approximate time window of 6-10 days post-LH surge ("LH+6 to +10") has been used most widely (Acosta et al. 2000; Díaz-Gimeno et al. 2011). The WOI is when the endometrium is prepared for the potential implantation of an embryo, and hence synchronization of these events and exact timing of optimal receptivity are crucial for implantation to be successful.

As the endometrium is a cycling tissue it undergoes morphological and transcriptional changes (Díaz-Gimeno et al. 2011; Noyes, Hertig, and Rock 1950). This happens in a controlled manner across every ~28 day cycle, implying that fine-tuned mechanisms are at play during the menstrual cycle that control this process. Indeed, the expression profile of the endometrium has been shown to respond and adapt to hormones across the menstrual cycle to the extent that it can be used to discriminate between the phases (Díaz-Gimeno et al. 2011; Ponnampalam et al. 2004; Talbi et al. 2006). However, the majority of the expression changes do not overlap between studies, which could be due to the differing experimental set ups and techniques used (Altmäe et al. 2017). As the WOI is the most important stage of the menstrual cycle, in terms of pregnancy, this is also the most well-studied sample cohort with gene ontology terms of the expression profiles showing an enrichment for genes involved in immune response, wound healing, cell-cell adhesion junctions, synthesis of amino acids and negative control of cell proliferation (Altmäe et al. 2017; Ruiz-Alonso, Blesa, and Simon 2012; Talbi et al. 2006). These terms relate to the switch from a proliferative state to a secretory state, therefore reducing proliferation, responding to inflammation and adapting their cellular adhesion. Díaz-Gimeno et al. developed an endometrial receptivity array (ERA) to distinguish between the non-receptive and receptive state (Díaz-Gimeno et al. 2011). During the WOI, when a blastocyst can succesfully implant, the endometrium is in a receptive state, whereas the non-receptive state is any time that falls outside of the window. This ERA is now avaliable commercially and is being used to establish the receptivity state of women undergoing IVF.

Fine-regulation of gene expression patterns is conferred and facilitated by epigenetic modifications. Histone modifications have been extensively studied in general and have been shown to positively correlate with expression changes as well as repressing certain genomic regions. Trimethylation of lysine 4 on histone 3 (H3K4me3) is commonly associated with open chromatin and highly expressed genes, where this mark is highly enriched over the transcription start sites (TSSs). It is worth noting though, that transcription can still occur in the absence of H3K4me3 (Heintzman et al. 2007). H3K4me3 can also be associated with bivalent regions, which are defined as regions that have both H3K4me3 and the repressive mark H3K27me3. This bivalent chromatin state was first identified in embryonic stem cells (ESCs) and demarcates genes that are considered 'poised' for activation (Azuara et al. 2006; Bernstein et al. 2006). This means that they are not expressed yet but can either quickly be 'switched on', or permanently 'switched off', in response to environmental factors or growth factor signals. In ESCs, this unique chromatin state is usually associated with developmentally regulated gene loci (Azuara et al. 2006; Bernstein et al. 2006).

When occurring in isolation, i.e. without the co-marking by H3K4me3, H3K27me3 is associated with gene repression. In chromatin immunoprecipitation (ChIP) assays, H3K27me3 peaks can be broader, less spatially restricted with this distribution being cell type dependent (Barski et al. 2007; Young et al. 2011). H3K27me3 is often associated with silenced genes that are required for key functions of antagonistic cell types (Roh et al. 2006). H3K9me3 is also a repressive histone modification and acts via the exclusion of transcription factors through restricting their ability to bind (Soufi, Donahue, and Zaret 2012). As a repressive modification, it is often associated with heterochromatin also known as closed chromatin, which is highly compacted and transcriptionally silent (Barski et al. 2007). Heterochromatin is rich in repetitive regions, such as satellites and retrotransposons, which are marked by H3K9me3 to ensure that these regions are not expressed. Thereby, H3K9me3, in conjunction with DNA methylation, is important for the safe-guarding of the genome (Becker, Nicetto, and Zaret 2016; Nakayama et al. 2001). Histone modifications profiles are dependent on the cell type and correlate with expression profiles, playing an important role in the epigenetic control of the genome and its accessibility.

To the best of my knowledge there has been no comprehensive genome-wide study of specific histone modification on human endometrial tissue. This could provide an in-

depth understanding of the dynamics that the tissue undergoes and what occurs at this key stage of the cycle.

As a starting point of my project, I analysed a total of twenty endometrial biopsies from women attending the Implantation Clinic at the University Hospitals Coventry and Warwickshire NHS Trust, Coventry, UK. The biopsies were provided by our collaborator, Prof. Jan Brosens with the required ethics obtained. All samples were LH surge-timed to fall within the implantation window, and patients received no hormonal treatments for at least three months prior to collection. As I was not only interested in establishing the epigenetic ground state of the endometrium, but also if the epigenome was altered in eutopic endometrium of patients with endometriosis, half of the samples came from patients with this disease. Apart from having been diagnosed as suffering from endometriosis, there was no significant difference between control and endometriosis patients in terms of age, BMI or days since LH surge, or any other medical parameter to which I had access (Table 2.1). It is worth noting that the control patients had unknown fertility problems, however these are expected to be random and not related to endometriosis, and therefore are not expected to present a confounding factor in the current analysis.

A series of genome-wide analyses was carried out on each of the twenty samples to generate a comprehensive profile of DNA methylation, hydroxymethylation, histone modifications and transcriptome for each sample. The techniques included: DNA immunoprecipitation with anti-5(h)mC antibodies (meDIP-seq and hmeDIP-seq), chromatin immunoprecipitation (ChIP) of histone H3 trimethylated at lysines 4, 9 and 27 (H3K4me3 ChIP-seq, H3K9me3 ChIP-seq, H3K27me3 ChIP-seq) and RNA-sequencing (Table 2.1, Figure 2.1). This is a unique methodology and gives us a complete overview and comparable approach that has not been carried out in this depth and comprehensive nature in the endometrium before. The first key aim of my project was to establish the epigenomic ground state of the endometrium specifically at this crucial stage during the window of implantation (WOI). This has been an aspect of research that has been neglected to date despite the critical importance of this phase for a successful pregnancy.

## 3.2 Genome-wide expression profiles reveal the precise receptivity stage, splitting the samples into distinct groups

Firstly, I generated genome-wide expression profiles from the biopsies by RNA-seq. The FastQC tool, which is well validated and used extensively for sequencing QC, was used to check the quality control of the raw sequencing data, where a value above 28 is considered to be high quality. An example of the quality score across all bases is given in Figure 3.1A. Read depth was adequate and ranged between 8 and 15 million reads. The sequencing reads were mapped to the GRCh38 human reference genome using HISAT2. The bulk of the data analyses was carried out using the SeqMonk program developed by Babraham's Bioinformatic team, which is a well validated tool for this type of analysis. The RNA-seq QC plot (Figure 3.1B) produced using SeqMonk was used to check the quality of the data and determined that there were no biases within the data. The majority of reads were in genes, specifically exons, the reads were on the sense strand as expected as the ScriptSeq V2 protocol generates strand-specific libraries, and all samples measured a similar percentage of genes and mitochondrial DNA suggesting little bias (Figure 3.1B). Samples were normalised to reads per million (RPM), matched to the 75$^{th}$ percentile and log2-transformed (Figure 3.1C). Statistical analysis was carried out using DESeq2 and EdgeR with a p value threshold of 0.01 and Benjamini-Hochberg correction applied to decrease false discovery rates (Love, Huber, and Anders 2014; Robinson, McCarthy, and Smyth 2009). Both statistical tests have been validated and are widely used within the field.

Using the normalised and log2-transformed read counts of each sample as a base for global comparisons, a data tree – that visualises the relative similarity of expression profiles to each other - revealed a clear segregation of the samples into two groups (Figure 3.2A). However, and perhaps surprisingly, this clustering did not correlate with the disease status of the patients.

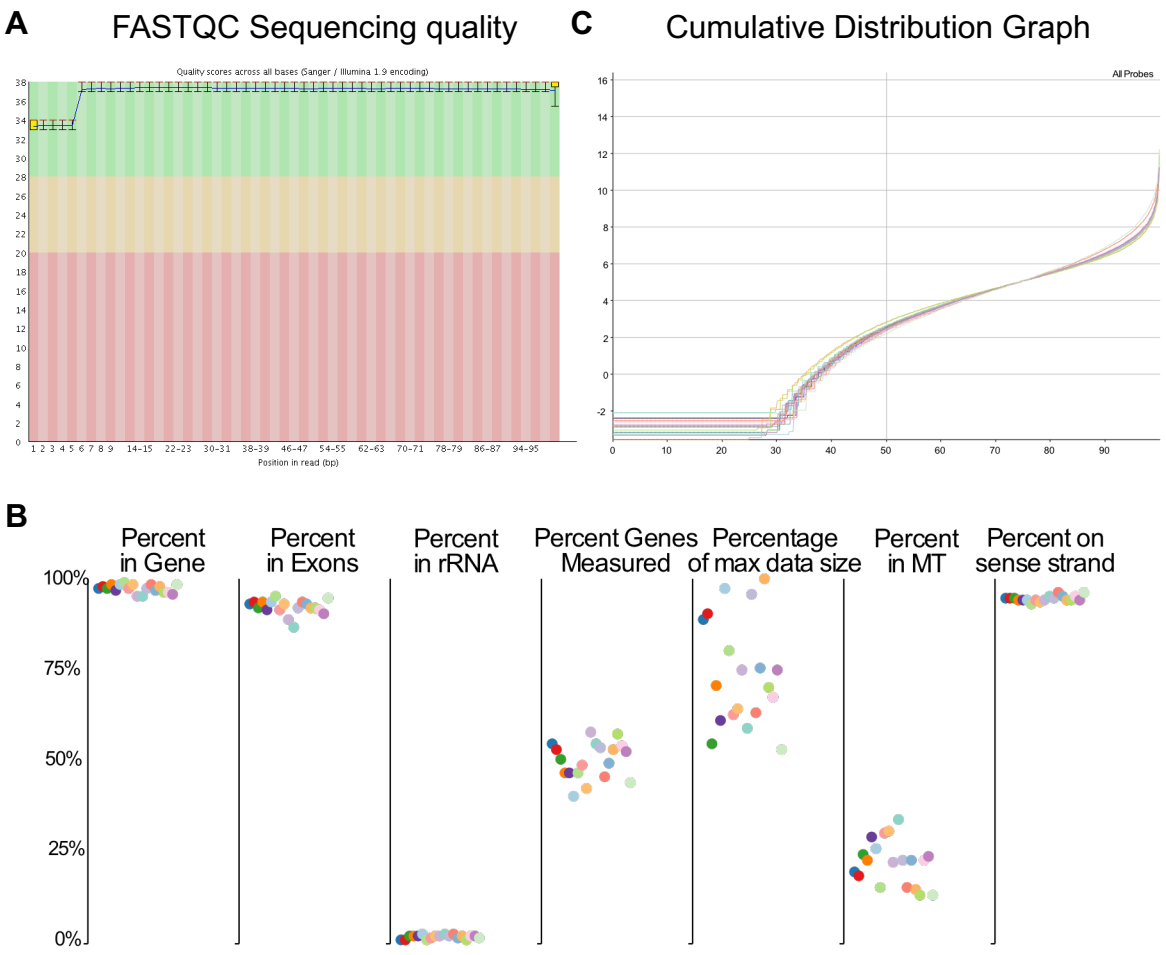



**Figure 3.1 RNA-sequencing quality control for endometrial samples. A**, An example of quality control checks on raw sequencing data, per base sequencing quality score. **B**, Mapped data quality control of the percentage of reads that fall into genes, exons, ribosomal ribonucleic acid (rRNA) and mitochondrial DNA (MT). Also showing the percentage of genes with reads, the number of reads in each sample compared to the largest data set and the percentage of reads on the sense strand. **C**, Cumulative distribution graph of RNA-sequencing data after normalisation to reads per million (RPM), matched 75% distribution and log2 transformation.

As outlined above, there is natural variation in the menstrual cycle between women. Thus, even at comparable days after the LH surge, some women may display a delayed or accelerated response to the hormone stimulus and the endometrium may therefore be outside the WOI. It was therefore possible that despite the fact that all samples were LH surge-timed to fall within the implantation window, the samples might still segregate into receptive and non-receptive cycle stages.

To examine this possibility, I used a previously published dataset termed the "Endometrial Receptivity Array (ERA )" that had specifically established gene expression signatures correlating with the WOI (Díaz-Gimeno et al. 2011). By applying this

signature to my datasets, it can be seen in both the scatterplot (Figure 3.2B) and the heatmap (Figure 3.2C) that the ERA clearly separates the samples into the same groups as established by the unbiased closest neighbour analysis displayed in the data tree (Figure 3.2A). Therefore, the precise state of receptivity can explain the majority of variation that lead to the sample segregation that was seen in the datatree. This was an important observation and is vital to consider going forward, as the precise receptivity state of the endometrium may obscure effects caused by endometriosis. For clarity, I will therefore refer to the samples that show a receptive state according to the ERA as "receptive" and the samples that fall outside of this window as "non-receptive". Most likely, these latter samples will have been delayed in reaching their optimal receptivity state. They show the expected expression patterns of genes that have previously been associated with the early or late secretory phase (i.e. high or low expression of genes associated with early secretory or late secretory phase respectively). An important aspect as this timing difference may well constitute a factor contributing to the fertility problems of these women.

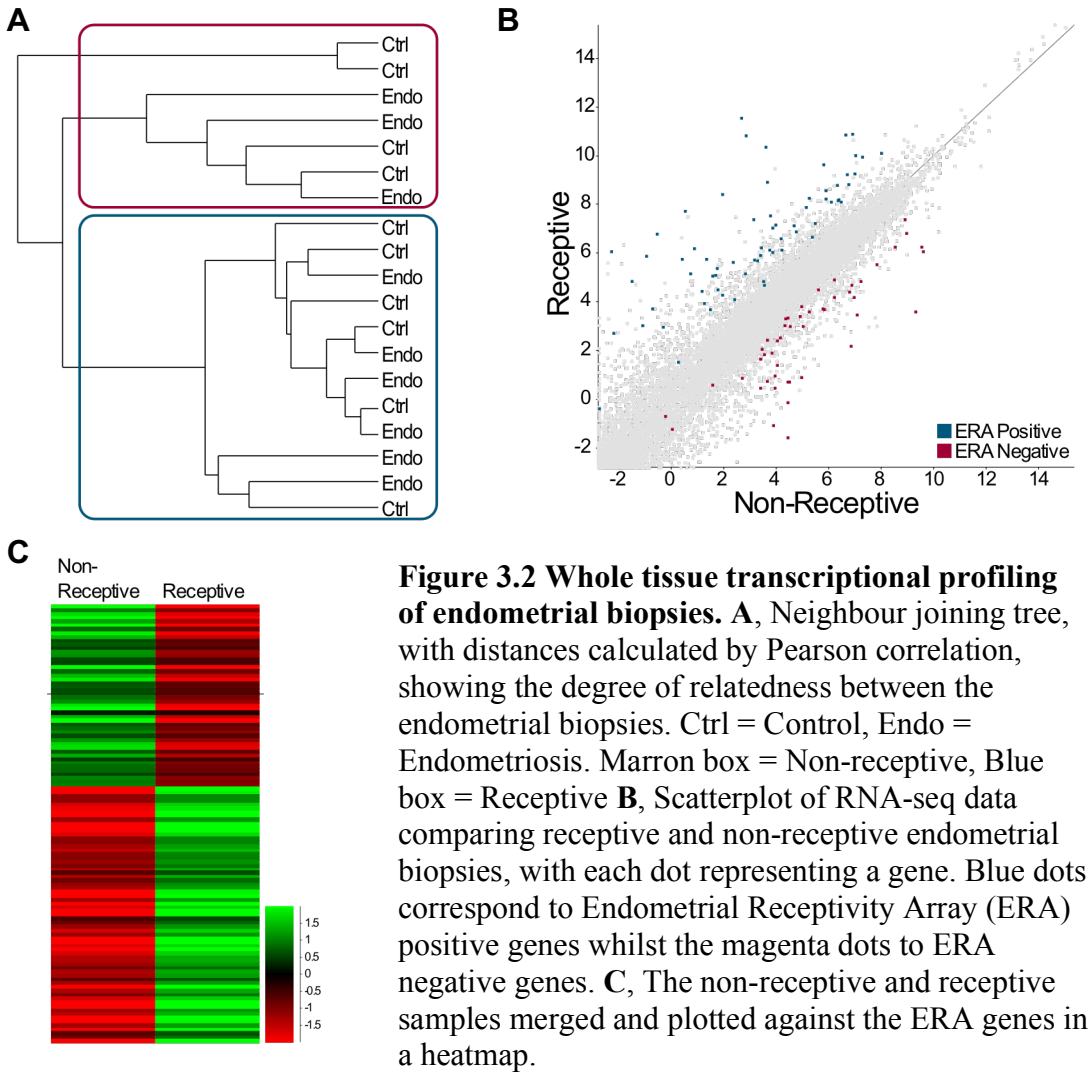



**Figure 3.2 Whole tissue transcriptional profiling of endometrial biopsies. A**, Neighbour joining tree, with distances calculated by Pearson correlation, showing the degree of relatedness between the endometrial biopsies. Ctrl = Control, Endo = Endometriosis. Marron box = Non-receptive, Blue box = Receptive **B**, Scatterplot of RNA-seq data comparing receptive and non-receptive endometrial biopsies, with each dot representing a gene. Blue dots correspond to Endometrial Receptivity Array (ERA) positive genes whilst the magenta dots to ERA negative genes. **C**, The non-receptive and receptive samples merged and plotted against the ERA genes in a heatmap.

This enhanced receptivity state assessment by global expression profiling, and the detected separation of the samples into receptive and non-receptive groups, provided us with an opportunity to investigate the epigenomic changes that occur exactly at this key tipping point of receptivity. Although expression changes during the menstrual cycle and specifically within and outside the WOI have been investigated before, the comparisons have usually been made on samples that are much farther apart, in terms of days around LH surge, than in our samples. Therefore, I have here the opportunity to investigate the exact changes that occur during a very limited time frame within, or just outside, the WOI.

To investigate further, on my detailed datasets, the transcriptional differences between the non-receptive and receptive states, I generated a whole tissue Endometrial Receptivity Transcriptome (ERT) gene set for this transition period. The ERT is very stringent as my criteria entailed that the genes had to be present in two distinct types of RNA-seq analyses, using DESeq2 and EdgeR statistics. Furthermore, for genes to be added to the ERT, I specified an expression difference of a magnitude of at least 2.0 on log2 transformed read counts, meaning that these genes are 4-fold higher or lower expressed within or outside the WOI (Figure 3.3A). These criteria established an ERT dataset consisting of 367 genes, 161 of which were up-regulated in the non-receptive state and 206 up-regulated in the receptive state (Supplementary Table 1). I randomly selected 3 genes each that were up-regulated in either the receptive or non-receptive state for independent validation by RT-qPCR (Figure 3.3B). All genes analysed were differentially expressed as expected from the RNA-seq data and therefore validated the transcriptional profiles, as well as the ERT signature gene set, for the two receptivity states.

Using this enhanced ERT signature gene set for gene ontology (GO) analysis revealed a number of interesting terms (Figure 3.3C). In particular in the non-receptive signature set, terms such as synapse assembly, negative chemotaxis (reduced movement in response to a specific chemical gradient) and cell adhesion were significantly enriched. The receptive signature gene set was enriched for biological terms including circulatory system processes, regulation of secretion and receptors. This shows that over the space of a few days, the endometrium undergoes a dramatic change in global gene expression profiles and with that a dramatic shift in its functionality.

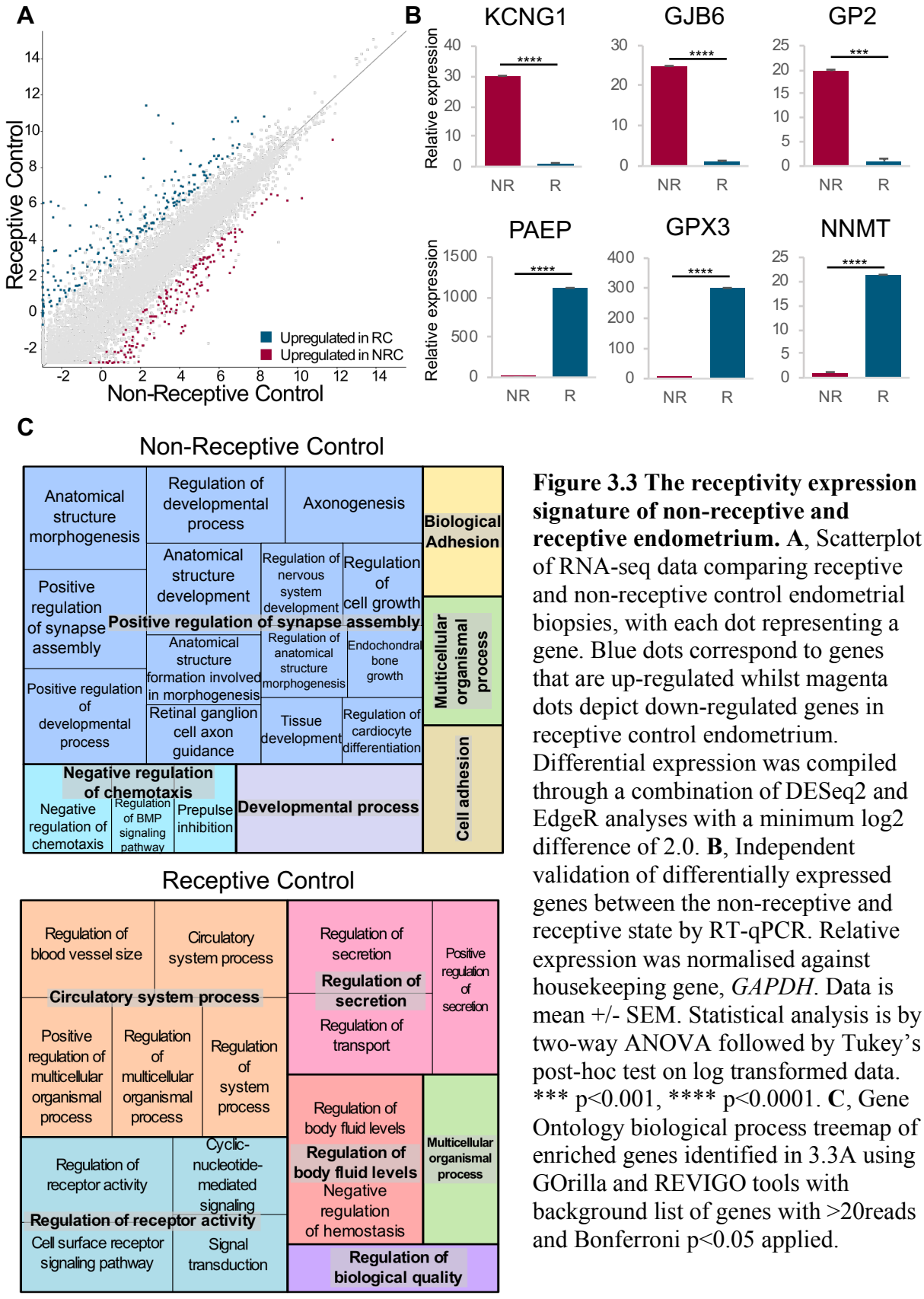



**Figure 3.3 The receptivity expression signature of non-receptive and receptive endometrium. A**, Scatterplot of RNA-seq data comparing receptive and non-receptive control endometrial biopsies, with each dot representing a gene. Blue dots correspond to genes that are up-regulated whilst magenta dots depict down-regulated genes in receptive control endometrium. Differential expression was compiled through a combination of DESeq2 and EdgeR analyses with a minimum log2 difference of 2.0. **B**, Independent validation of differentially expressed genes between the non-receptive and receptive state by RT-qPCR. Relative expression was normalised against housekeeping gene, *GAPDH*. Data is mean +/- SEM. Statistical analysis is by two-way ANOVA followed by Tukey's post-hoc test on log transformed data. *** $p<0.001$, **** $p<0.0001$. **C**, Gene Ontology biological process treemap of enriched genes identified in 3.3A using GOrilla and REVIGO tools with background list of genes with >20reads and Bonferroni $p<0.05$ applied.

## 3.3 Virtual deconvolution of cell type composition of endometrial tissue biopsies

### 3.3.1 Determination of overall proportion of the two main cell types

As previously discussed in the Introduction, the endometrium is made up of several cell types, the two main ones being epithelial cells (glandular and luminal) and stromal cells. These can clearly be distinguished even in nuclear stainings using DAPI, with the epithelial cells forming a distinct structure of characteristically columnar cells outlining glands, with stromal cells in between (Figure 3.4A).

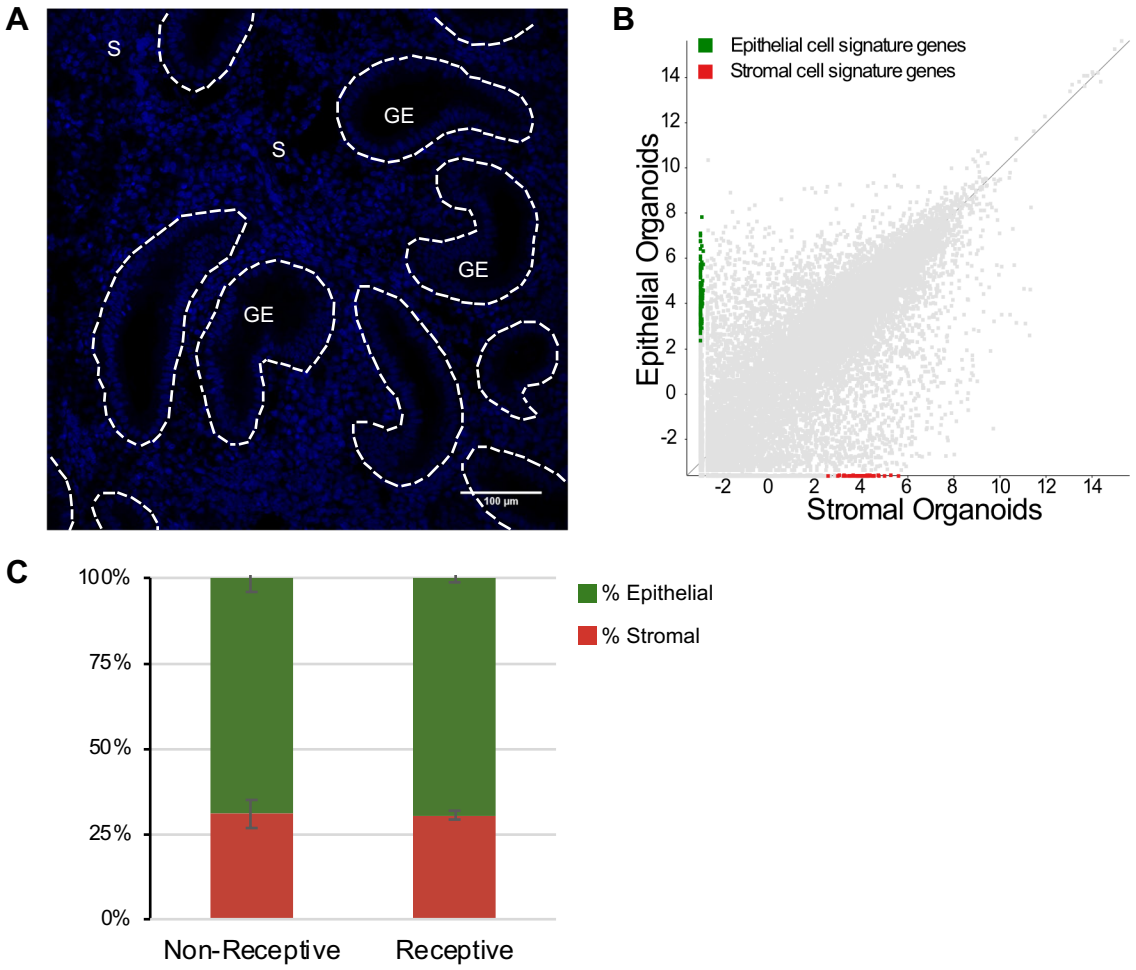



**Figure 3.4 Cell type composition of endometrial tissue with virtual deconvolution. A**, Example image of DAPI-stained endometrial tissue to highlight nuclear staining. The dotted lines demarcate glandular epithelium. S = stromal cells, GE = glandular epithelium. **B**, Scatterplot of RNA-seq data comparing stromal and epithelial cell organoids, with each dot representing a gene. Green dots correspond to genes that are only expressed in epithelial cell organoids, whilst red dots show genes expressed only in stromal cell organoids. **C**, Virtual deconvolution of approximate percentage of epithelial and stromal cells contributing to whole tissue +/- SEM. Two-way ANOVA followed by Tukey's post-hoc test showed no significant difference between both groups ($p>0.05$).

As the transcriptional profiling was carried out on whole tissue it was important to investigate the influences that these two cell types may have on the expression profiles, as differences in their tissue contribution may be an important confounding factor. To this end, I carried out RNA-seq on epithelial and stromal cell organoids established from primary tissue. Genes were identified that were highly expressed in either only stromal or epithelial cells, creating an expression signature for both cell types (Figure 3.4B, Supplementary Table 2 and 3). This now allowed the whole tissue to be virtually deconvoluted into the approximate fractions of epithelial and stromal cell contributions to the biopsy expression profile. Using this approach, I found that there was no significant difference between the non-receptive and receptive state samples in terms of contribution from epithelial or stromal cells (Figure 3.4C). This gives us confidence that the differences that arise between the two states are not due to differential tissue composition, but are biological changes that occur between the different samples and the receptivity states.

### 3.3.2 Epithelial cell signatures correlate with receptivity state

When the epithelial cell signature genes were plotted against the corresponding whole tissue from which the organoids had been derived, the genes distinctly clustered into two groups, with expression changes occurring due to receptivity (Figure 3.5A, Supplementary Table 3). This analysis revealed that the epithelial cell signature genes needed to be subdivided according to receptivity state. Therefore, the epithelial cell signature was separated into two sets of genes according to the clustering, thus giving rise to: a *non-receptive epithelial* cell signature and a *receptive epithelial* cell signature. Both epithelial cell signatures were then used to calculate approximate cell type contributions to the whole tissue samples. As can be seen in Figure 3.5B, the *non-receptive epithelial* proportion is significantly higher in the non-receptive state tissue compared to the receptive state tissue. By contrast, the *receptive epithelial* proportion is significantly higher in the receptive state tissue compared to the non-receptive state tissue. This is as expected and served as a strong validation of my RNA-seq analysis approach.

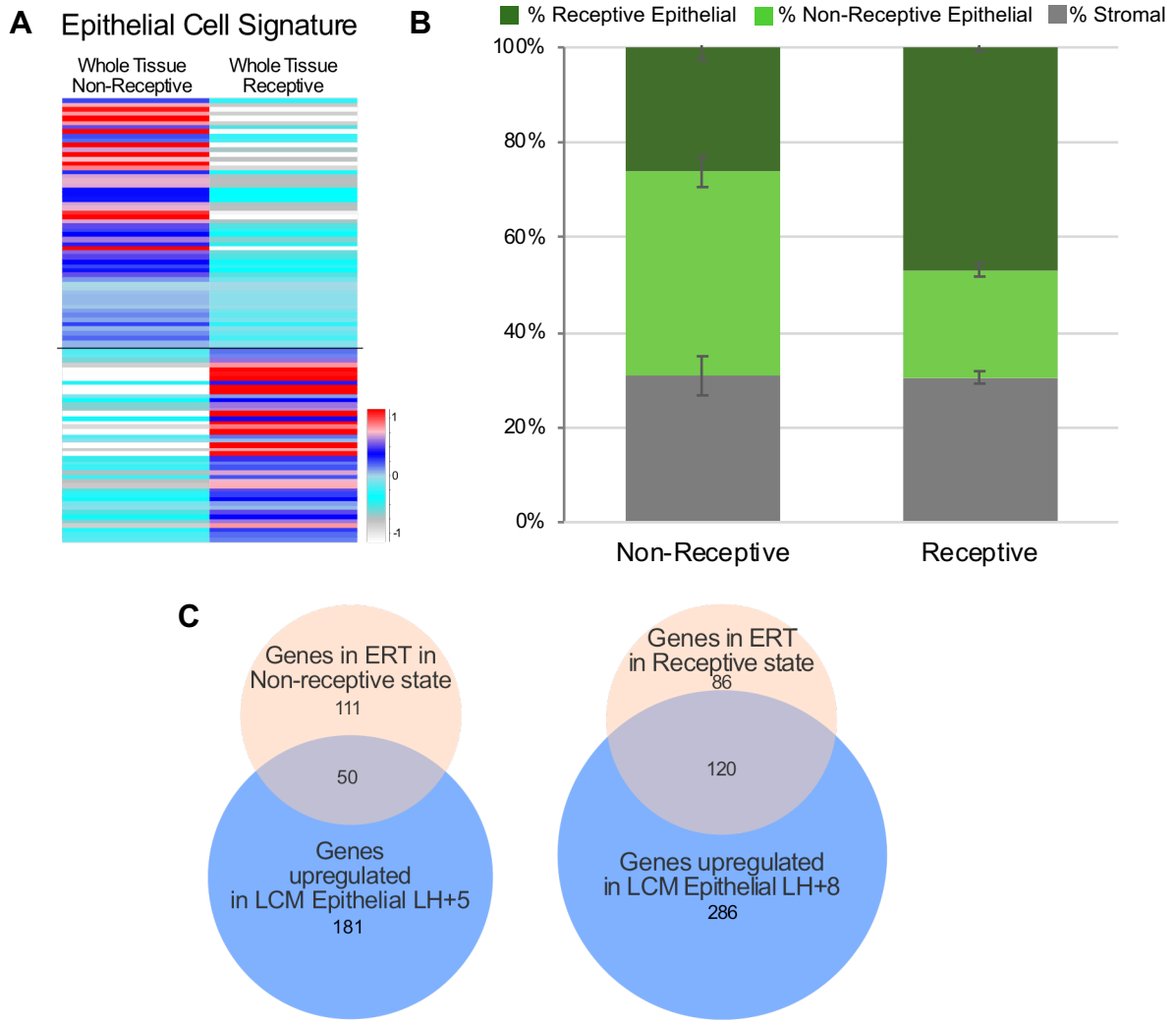



**Figure 3.5 The importance of epithelial cells in the transition to receptivity. A**, Heatmap of epithelial cell signature genes in whole tissue samples discriminates receptivity states. These whole tissue samples are grouped according to receptivity. **B**, Virtual deconvolution of approximate percentage of stromal cells, non-receptive and receptive epithelial cells contributing to whole tissue. Mean +/- SEM plotted. Two-way ANOVA followed by Tukey's post-hoc test showed significant difference of non-receptive ($p<0.0001$) and receptive ($p<0.0001$) epithelial cell signature between the non-receptive and receptive state. **C**, Venn diagram showing genes that are differentially regulated between the non-receptive and receptive state as established in the whole tissue expression analysis (ERT, orange). This is overlapped with data from Salker et al. of RNA-seq data from laser-capture microdissected epithelial cells at LH+5 and LH+8, filtered to have an expression fold change of at least 4.0 (Salker et al., 2017, blue).

In addition to establishing and using cell type-specific transcriptome datasets from organoids, I also pursued a parallel line of investigation based on data published by Salker et al. who had carried out laser capture microdissection coupled with RNA-sequencing on epithelial cells from biopsies obtained at LH+5 and LH+8. They had analysed the data using DESeq2 to determine significantly upregulated genes at either

LH+5 or LH+8, I applied an additional filter of a fold change of at least 4.0. Firstly, worth noticing is that there are 406 genes upregulated at LH+8 compared to 231 at LH+5 in LCM-epithelial cells. The LCM-epithelial data was then overlapped with the Endometrial Receptivity Transcriptome (ERT) that I established in the whole tissue (Figure 3.5C). Interestingly there is a larger proportion of the LCM-epithelial upregulated genes that overlapped in the receptive state of the ERT (58.3%) compared to the non-receptive state of the ERT (31.1%) (Figure 3.5C). This suggests that epithelial cells could be contributing to the majority of expression changes that occurs in the receptive state in whole tissue more so than in the non-receptive state.

## 3.4 Characterisation of histone modification dynamics

### 3.4.1 Global histone patterns

Having established gene expression patterns and receptivity-associated signature gene sets in all of my samples, my next aim was to determine an epigenetic baseline of the relative distribution of key histone modifications. To investigate the distribution of histone modifications in the endometrium, immunofluorescence (IF) for H3K4me3 and H3K27me3 was carried out (Figure 3.6A). H3K9me3 was not assessed by IF as it is mostly localised to heterochromatic regions and perhaps not as informative as the other two modifications for immediate gene regulation. As expected, the cellular distribution of the signals obtained was largely nuclear. To accurately determine whether there are staining intensity differences between epithelial and stromal cells, I applied a nuclear mask, determined by the DAPI signal, and measured the mean fluorescence intensities for the appropriate fluorophore of the secondary antibody used with CellProfiler software.

Since immunofluorescence stainings are prone to be variable between different sets of experiments and even between slides, my approach only compared cell types within the same image for each sample to overcome this. I then calculated the ratio of epithelial-to-stromal cell intensities in order to obtain directly comparable values. This analysis revealed that there is a trend for the mean intensity of H3K4me3 epithelial: stromal cell ratio to decrease in the receptive state, although this just misses significance ($p=0.0731$) (Figure 3.6B). There is no significant difference for H3K27me3 ($p=0.2569$) (Figure 3.6C).

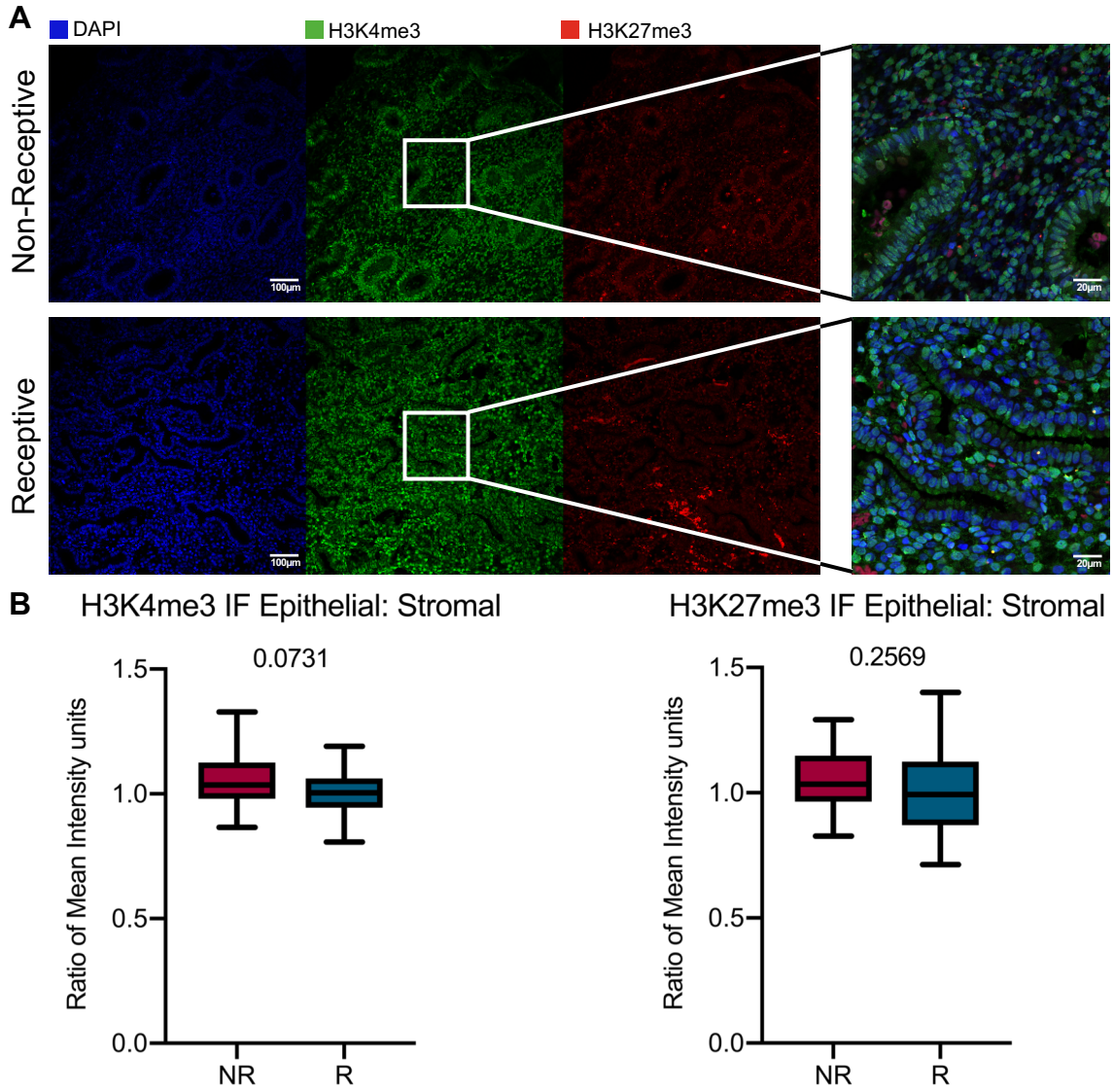



**Figure 3.6 Distribution of H3K4me3 and H3K27me3 in endometrial tissue with regards to receptivity. A**, Immunofluorescence staining of endometrial tissue in the non-receptive and receptive state for H3K4me3 (green) and H3K27me3 (red). DAPI (blue) was used as a nuclear counter-stain. Scale bar = 100µm or 20µm in zoomed merged images **B** and **C**, Epithelial: Stromal cell ratio of mean intensity units within each immunofluorescence image. Two-way ANOVA followed by Tukey's post-hoc test showed p = 0.0731 and 0.2569 for H3K4me3 and H3K27me3, respectively. Box whisker plot with minimum, median and maximum values plotted. All conditions had a minimum of 3 samples with 3 images each. NR = non-receptive, R = receptive.

To explore the role of these modifications further, Chromatin Immunoprecipitation (ChIP) for H3K4me3, H3K27me3 as well as H3K9me3 was conducted on the same biopsies from the patients that were used for the transcriptional analysis, allowing for direct comparison (Figure 2.1). This allowed the exploration of global distribution patterns and potential shifts in chromatin organization that may occur at individual loci between receptivity states or also disease status.

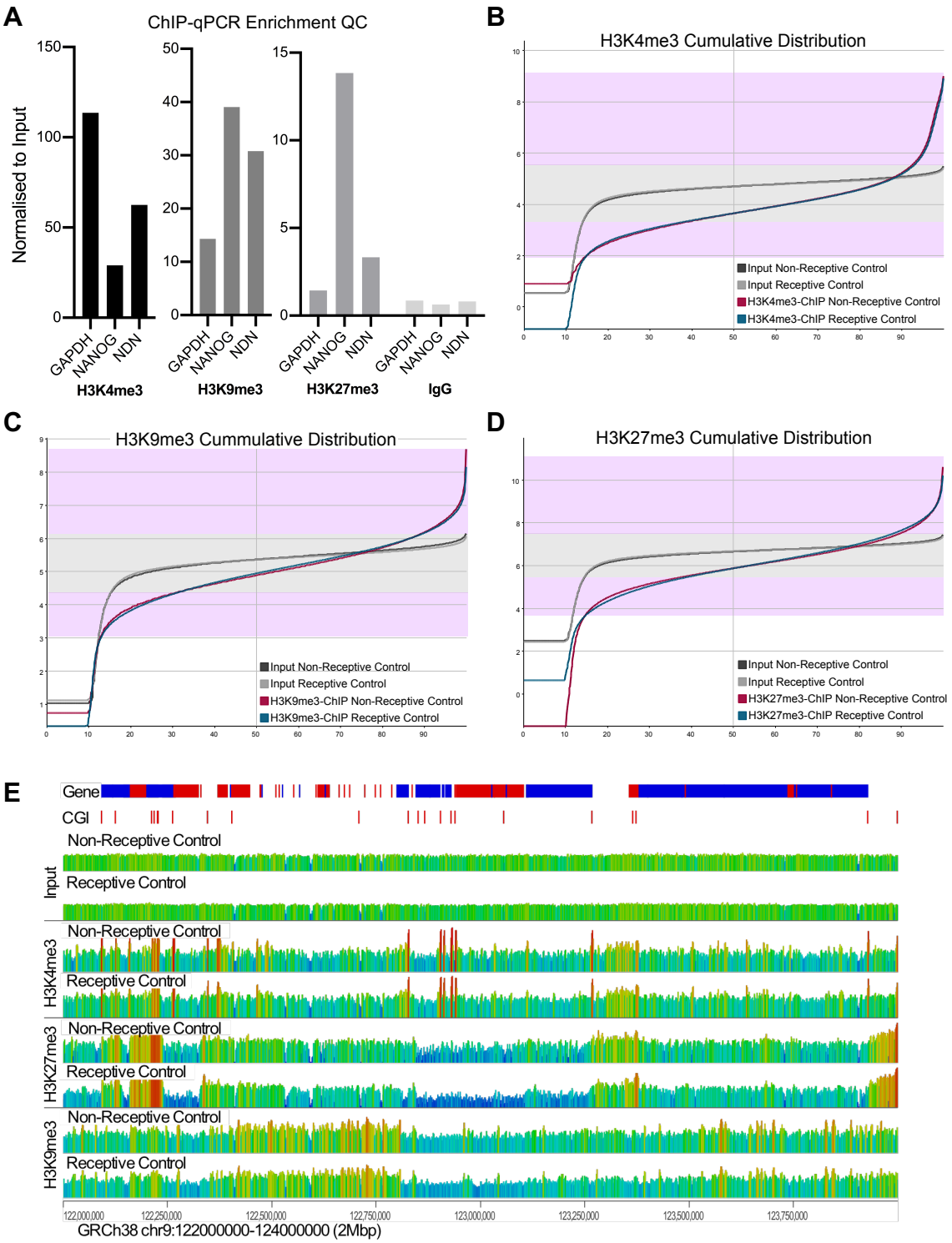



**Figure 3.7 Quality control of endometrial tissue histone-ChIP. A**, An example ChIP-qPCR of *GAPDH* (housekeeping gene), *NANOG* (repressed developmental gene) and *NDN* (imprinted gene) for histone enrichment quality control. Data are normalised to Input. **B, C** and **D**, Cumulative distribution graphs of Input and histone ChIP-seq data, using 5kb tiling probes across genome spaced 2.5kb apart. Data are corrected to largest dataset, log2 transformed and enrichment normalised to 20$^{th}$ and 90$^{th}$ percentile where applicable. Grey box illustrates inputs flat distribution. Violet boxes illustrate ChIPs dynamic distribution. **E**, 2Mb view of input and histone-ChIP-sequencing data, 2kb probes across genome every 1kb showing the enrichment of histone modifications in Chr9.

As I optimised this technique for human endometrial biopsies, extensive quality controls were carried out. Prior to sequencing ascertain of enrichment by ChIP-qPCR of *GAPDH*, a housekeeping gene, which is enriched for H3K4me3 at the TSS, of *NANOG*, a repressed developmental gene which is enriched for both H3K9me3 and H3K27me3, and lastly of *NDN*, an imprinted gene which is enriched for H3K9me3 (Figure 3.7A). All qPCRs were normalised to the Input. Reassuringly, the IgG control pulldown showed no enrichment, confirming that the ChIP procedure was successful and specific (Figure 3.7A). Furthermore, a cumulative distribution plot of the ChIP-seq data post sequencing shows that there is an enrichment of the ChIP samples with a wider dynamic range (illustrated by violet boxes) compared to the Input samples, which show no enrichment and therefore have a flatter distribution (illustrated by grey boxes) (Figure 3.7B, C, D). This can further be seen in an overview screenshot of a 2Mb window of Chr 9, which shows an enrichment in the histone-ChIP samples compared to the corresponding Inputs (Figure 3.7E). This indicates that the ChIP technique itself was successful on these limited-material samples and gives us confidence moving forward with the analysis.

In other tissues H3K4me3 is associated with transcription start sites (TSSs) of genes that are highly expressed. I plotted an alignment of H3K4me3 enrichment across TSSs (+/- 5kb) to check that it has a similar distribution and function in endometrial tissue. This was confirmed by an enrichment of H3K4me3 at genes that are highly expressed, defined in the RNA-seq as genes with expression levels greater than the median RPM. H3K4me3 enrichment decreased in conjunction with expression (Figure 3.8A). It is worth noting however, that the presence of H3K4me3 at lowly expressed genes is perhaps higher than expected. It is possible that this could be due to the fact that we are looking at a snap shot of a cycling tissue, as well as the possibility of at least some of these genes being bivalently marked (see below).

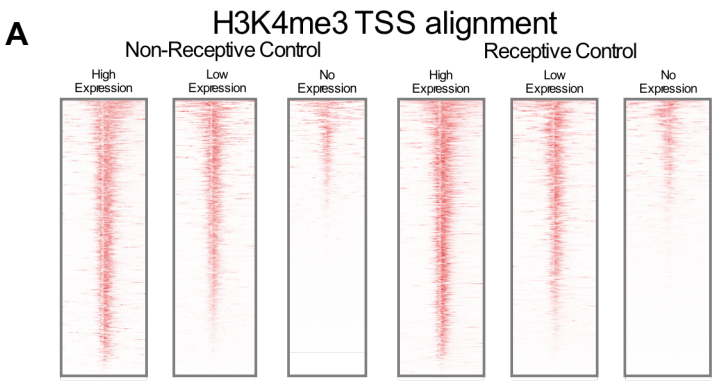



**Figure 3.8 H3K4me3 correlates with expression at TSSs**. **A**, H3K4me3 enrichment is aligned across transcription start sites +/- 5kb on either side, for non-receptive and receptive

samples. These have been further subdivided according to the level of gene expression established in the RNA-seq. High expression was defined as read count values above median RPM, low expression as between median RPM and 0 RPM, no expression as below 0 RPM.

### 3.4.2 Identification of a loss of bivalency in receptive tissue

As we wanted to integrate the complexity of the histone profiles with the progression of the menstrual cycle, we used a multivariant approach, which has previously been developed for chromatin dynamics in space and time, called chromstaR (Taudt et al. 2016). The univariant fits for each sample was checked to ensure that the peaks are called correctly and the right parameters are used. All samples that were satisfactory were used for further analysis (Table 2.1). Firstly, chromstaR analysed the frequency of each chromatin modification individually as well as in combination within each receptivity state for the control samples and this revealed no significant difference in H3K9me3-only (Figure3.9A). As H3K9me3 has been shown previously to be required for the repression of repetitive elements within the genome, this was specifically investigated. The Babraham Bioinformatics department carried out an analysis of the percentage of total reads that were present at each repetitive element by aligning the sequences using Bowtie2 to repeat genomes for each repeat family (see section 2.13.5). This confirmed that there is no significant difference between the non-receptive and receptive state in terms of H3K9me3 distribution (Figure3.9C).

In contrast to H3K9me3, I observed significant differences for H3K4me3 and H3K27me3 ($p<0.001$) in the number of peaks called between non-receptive control and receptive control samples across the genome. A particularly pronounced effect was observed at regions “bivalently” marked with both H3K4me3 and H3K27me3 ($p<0.001$) (Figure3.9A). In the non-receptive state there are 7385 bivalent peaks compared to 6131 peaks in the receptive state. A proportion of these regions in the non-receptive state lose their bivalency and resolve to another chromatin combination in the receptive state. Of the domains that resolve the majority of these resolve to H3K4me3-only (92.1%) (Figure3.9B). This indicates that bivalent regions play a key role within the cycling endometrial tissue.

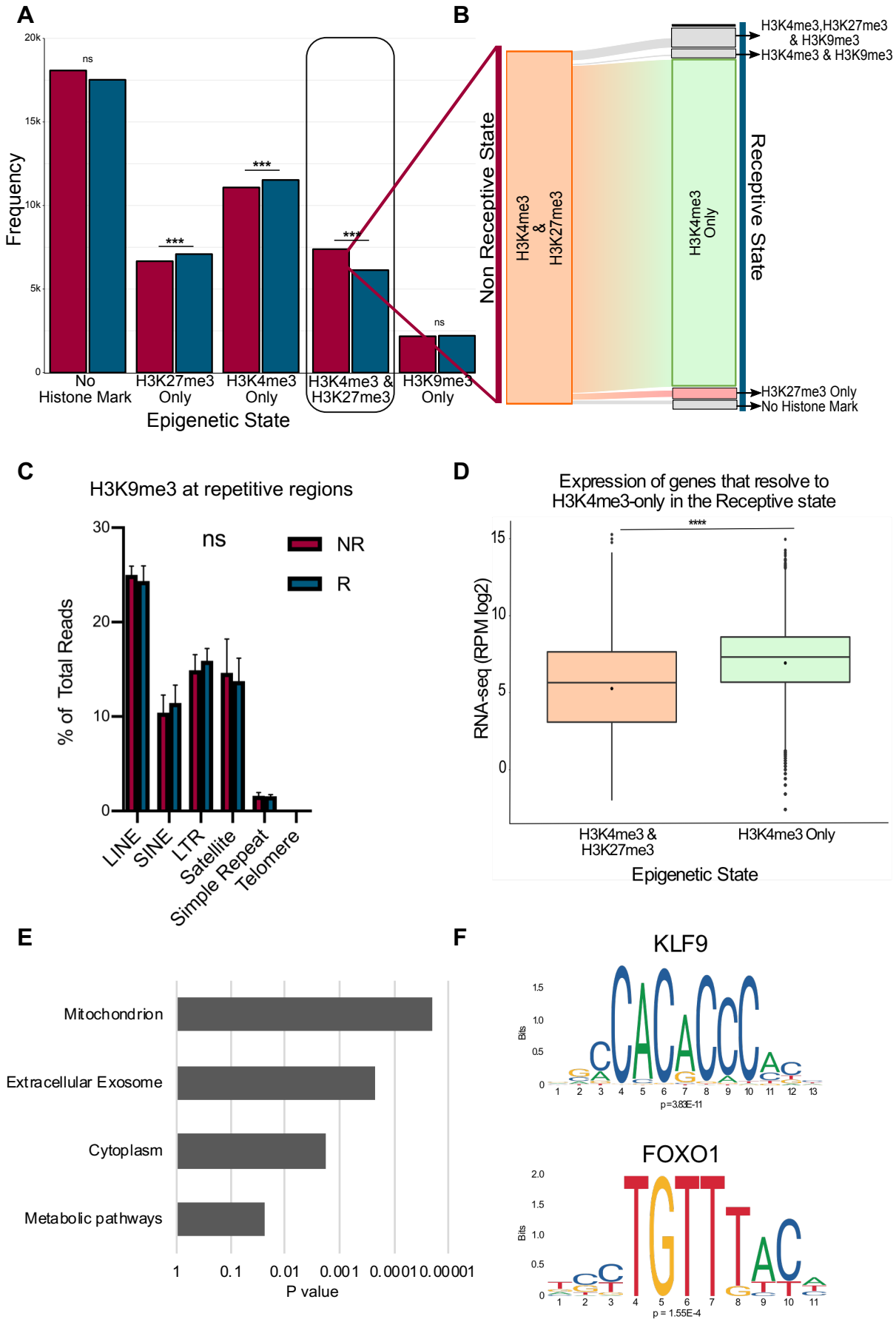



**Figure 3.9 Bivalency is key during the transition to a receptive state. A,** chromstaR analysis of peak frequency in non-receptive (maroon) and receptive (blue) state. Chi-square statistical test showing a significant difference of ***p<0.001 for H3K4me3-only,

H3K27me3-only and H3K4me3 & H3K37me3. **B**, Slankey diagram of the chromatin state destination of bivalently marked regions that resolve to another state during the transition from non-receptive to receptive state. **C**, Enrichment of H3K9me3 at various repeat elements. Analysis was conducted, by the Babraham Bioinformatics department. Two-way ANOVA shows no significant differences between the receptive and non-receptive samples at any of the elements analysed. **D**, Expression analysis of genes that are bivalently marked in the non-receptive state and resolve to H3K4me3-only in the receptive state. Wilcoxon test, ****$p<0.0001$. **E**, Gene Ontology analysis of genes that are bivalently marked in the non-receptive state and resolve to H3K4me3-only in the receptive state. This was carried out using DAVID bioinformatics tool with a background list of all bivalent genes in the non-receptive state. Bonferroni correction applied, $p<0.05$. **F**, Motif enrichment analysis of the regions that resolved to H3K4me3-only, carried out using AME tool in MEME Suite, showing enrichment in transcription factors: KLF9 p= 3.83E-11, FOXO1 p= 1.55E-4.

To determine if this resolution of bivalency to H3K4me3-only has a functional relevance, I investigated the expression levels of the corresponding genes associated with these epigenetic marks. These loci had significantly higher levels of expression in the H3K4me3-only state compared to the bivalent histone configuration in the non-receptive state (Figure3.9D). This conforms with our biological understanding of the role of bivalent domains and H3K4me3, a concept that first emerged from the analysis of ESCs where bivalent genes were found to be transcriptionally 'off' but 'poised for activation', resulting in either activation (and resolution to H3K4me3 only) or stable repression (and resolution to H3K27me3 only) upon developmental stimuli (Azuara et al. 2006; Bernstein et al. 2006). In the context of the cycling endometrium, the data shows that bivalent chromatin is also a major feature of gene regulation in this tissue, allowing genes to rapidly switch on upon entering the window of receptivity. It may also correlate with endometrium exhibiting a significant developmental plasticity.

To identify which functional gene categories were regulated by this chromatin composition, Gene Ontology (GO) analysis was carried out on the genes associated with regions that resolved to H3K4me3-only (475 genes), with a background list of all genes associated with bivalent regions, specifically including the genes that remained bivalent in the receptive controls (6523 genes), using the software DAVID (Huang, Sherman, and Lempicki 2009b, 2009a). Terms that stood out as significantly enriched included extracellular exosome (p=2.27E-04) as well as mitochondrion and metabolic pathways (p= 2.05E-05 and 0.025 respectfully) (Figure3.9E). These terms have previously been linked to the receptive endometrium and are related to the increase in secretory output of the endometrium at this timepoint (Altmäe et al. 2017; G. E. Evans et al. 2014).

Next, I used these bivalent regions that resolve to H3K4me3-only upon entry into the receptive state to carry out a Motif Analysis using the Analysis of Motif Enrichment (AME) tool in the MEME Suite package (McLeay and Bailey 2010). The rationale of this approach was to find out if any known transcription factor binding motifs were enriched at these sites. The analysis was conducted using HOCOMOCOv11 Human and JASPAR2018 core vertebrates as motif databases, and was further refined by factors having to be present in both outputs (Supplementary Table 4). A number of significantly enriched motifs were identified; of most notable interest was the enrichment of KLF9 and FOXO1 binding sites (Figure3.9F). Both these factors have been shown to be regulated by progesterone which increases and peaks during the secretory phase (Kajihara and Brosens 2013; D. Zhang et al. 2002). KLF9 null mice are also sub-fertile which could be due to lower implantation rates (Simmen et al. 2004).

Examples of regions that exhibit this epigenetic configuration and transition from bivalency to H3K4me3-only are shown for *SOD2* and *ELF3* (Figure 3.10A and B). Both these genes gain expression in the receptive state and correspondingly lose H3K27me3 at the bivalent domains, resulting in marking by H3K4me3-only (Figure 3.10A and B). Both genes also contain KLF9 and/or FOXO1 binding sites.

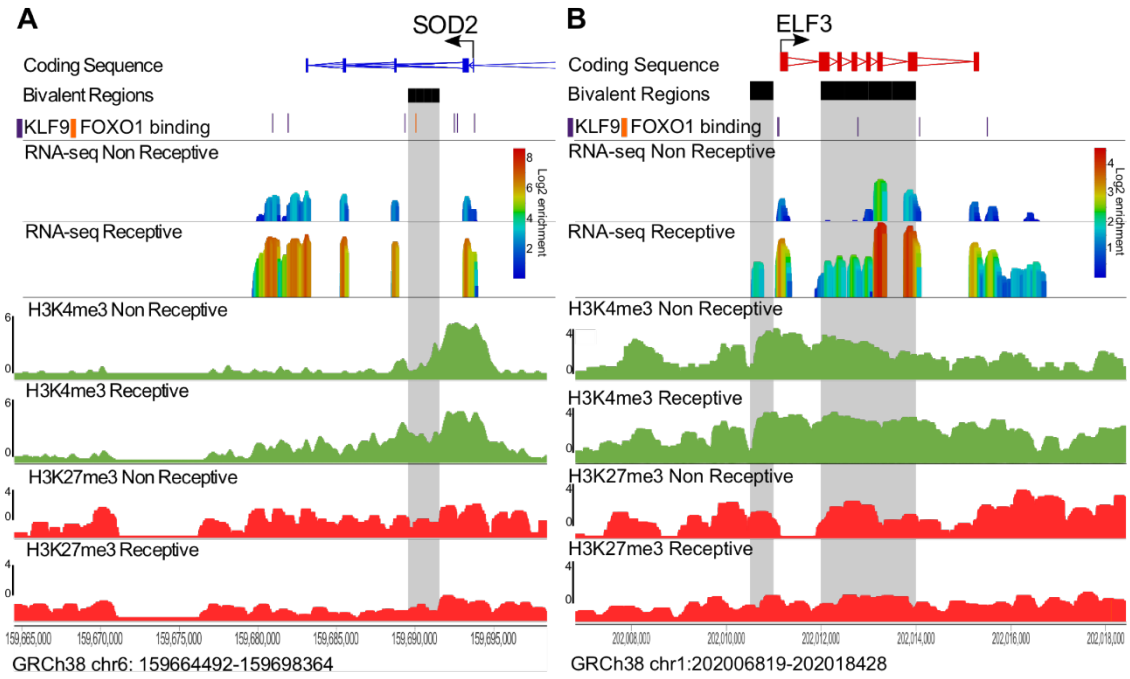



**Figure 3.10 Bivalency dynamics at specific loci. A** and **B**, Wiggle plots of gene expression and chromatin enrichment data at the *SOD2* and *ELF3* gene loci, respectively. RNA-seq (RPM log2), H3K4me3 and H3K27me3 enrichment of ChIP-seq data (normalised to largest

dataset) are shown in non-receptive and receptive state. Regions that resolve their bivalency to H3K4me3-only in the receptive state are denoted by black boxes, with greyed out areas underneath. KLF9 (purple) binding sites and FOXO1 (orange) binding sites are denoted.

ELF3 has been shown to be important for the differentiation and development of epithelial cells and specifically in the process of epithelial-to-mesenchymal transition, which partially occurs in the secretory phase (A. Y. N. Ng et al. 2002; Whitby, Salamonsen, and Evans 2018). SOD2 has previously been shown to be a target of FOXO1 in T-cells and cultured endometrial stromal cells (Choi et al. 2009; Kajihara et al. 2006). It is an important antioxidant defence enzyme that protects the endometrium against oxidative damage during receptivity in the guinea pig (Makker et al. 2006).

## 3.5 Discussion

I have established the correct and refined receptivity state of all initial endometrial biopsy samples used, and characterised their approximate cellular composition by establishing cycle stage- and cell type-characteristic signature gene sets. This is the first comprehensive epigenomic profile of endometrium during stages that specifically span the tipping point from the non-receptive to the receptive state.

The first key outcome of this chapter is the establishment of a gene expression signature (ERT) that can reliably detect the transcriptional changes within a very narrow time window of the endometrial cycle, either inside, or just outside, the receptive state of the endometrium (Figure 3.3A, Supplementary Table 1). This analysis was further advanced to determine gene sets specific to the non-receptive and receptive epithelium, which undergoes the most dramatic shift in gene expression during this time period (Figure 3.5A, Supplementary Table 3).

As I have clearly identified, establishing what precise cycle state the tissue is in via transcriptional profiles ensures that results are not obscured due to the assignment of an incorrect state. So far, to specify the endometrial cycle stage, histology using a detailed description of the differences that arise at different stages of the cycle was commonly conducted, described by Noyes in 1950 (Noyes, Hertig, and Rock 1950). Although histology was the gold standard for endometrial phase dating, more recently several papers have highlighted that it cannot be solely relied upon (Lessey 2011; Talbi

et al. 2006). Several groups have studied the changes that arise transcriptionally across the menstrual cycle as a way to accurately date the endometrium (Díaz-Gimeno et al. 2011; Ponnampalam et al. 2004; Talbi et al. 2006). These studies demonstrate that there are transcriptional changes that occur across the menstrual cycle, although there is currently no consensus list of genes that make up a signature for a specific stage. This is due to a low overlap of genes and inconsistent time points between all the studies. Other contributing factors to these discrepancies include the precise type of microarray, statistics and bioinformatics used. Probably the most important factor, however, is sample variability – number of patients, their age and prominently the day the sample was acquired and how the menstrual cycle phase was determined. Within this study the unforeseen grouping of the samples resulting in four conditions instead of two has resulted in fewer samples per a condition than intended. The reduction of the number of controls as well as the possibility of asymptomatic endometriosis being present could increase the variability within the study. This analysis has none the less expanded our knowledge of transcriptional changes that occur during the endometrial cycle and mechanisms that are in action. In order to further validate this pilot study an increase in the number of control samples would reduce the variability.

Diaz-Gimeno et al. established an endometrial receptivity array (ERA) specifically to identify the receptive state. They developed this by first selecting genes that were differentially expressed between LH+1, +3, +5 combined and LH+7 with a fold change of at least 3; resulting in 238 genes, then a second set of biopsies was used to train the predictor tool to date the endometrium and to define the signature genes specifically for the receptive state. The second set of biopsies were obtained at Day 8 to 12 of the menstrual cycle, proliferative phase, LH+1 to +5 pre-receptive and LH+7 to +8 receptive. The intersection of differentially expressed genes between proliferative vs receptive state (179 genes), and pre-receptive vs receptive state (200 genes), resulted in 134 genes that were present in both sets and defined the receptive state (Díaz-Gimeno et al. 2011). The ERA is currently being used in research as well as commercially. The ERA separated my samples according to an unsupervised data tree. As my samples are far closer in cycle timing (a few days) compared to samples spanning Day 8 to LH+8 (Day 21) in the ERA, it was interesting and important to determine what the exact expression changes beyond the ERA are in my LH surge-timed samples that fall inside, or just outside, the narrow receptivity window. This revealed striking GO terms, especially when they are grouped into related terms in the GO biological process treemap: Positive

regulation of synapse assembly as well as anatomical structure, tissue development and cellular and biological adhesion are all up-regulated in the non-receptive state (Figure 3.3C). This also aligns with GO component terms which are all involved with the membrane and the extracellular matrix (Supplementary Figure 1).

The main GO term in the non-receptive state is related to synapses which is associated with the formation and modulation of axon and nerve synapses. However, the genes that contribute to these GO terms also have functions involved in cell adhesion and cell-cell contact (LRRC4B, NRXN1, CTNNA2), cell migration and motility (ROBO2, ANK3), response to inflammation (NR4A2) and G-proteins (EPHB1, EPHB3, MAP2). These genes highlight that during the non-receptive state the endometrium is adapting its structure and cell polarity through cell-cell contacts. With the additional GO terms of biological and cellular adhesion this points to the endometrium's preparation for a potential implantation. Previous studies refer to an enrichment of cellular adhesion in the WOI. This inconsistency could be due to the differing timing of the samples with previous studies comparing LH+1 to +3 to LH+7 to +9 or the stages being established by histology, rather than ours which compared within LH+6 to +9 (Díaz-Gimeno et al. 2011; Ponnampalam et al. 2004; Ruiz-Alonso, Blesa, and Simon 2012).

Receptive samples showed an upregulation of genes involved in circulatory system processes e.g. blood vessels, secretion and receptors (Figure 3.3C). This is similar to many other GO analysis which show an increase in glandular secretions including exosomes (Altmäe et al. 2017; Díaz-Gimeno et al. 2011; Ruiz-Alonso, Blesa, and Simon 2012). Secretions (PRL, PAEP) and receptors (IL-15Rα, αvb3 integrin) have been shown to be important for the establishment of the receptive state for cell-cell communication as well as to set up for interactions with a potential embryo leading to a successful implantation (Burton et al. 2002; Franchi et al. 2008; Gellersen and Brosens 2014; Mazur et al. 2015; Talbi et al. 2006).

Not only is the correct timing of the biopsies essential but it is also important to consider the cell type composition of whole tissue samples. As previously discussed endometrial tissue is comprised of several cell types, with the two main cell types being epithelial (luminal and glandular) and stromal cells (non-decidualised and decidualised). Other cell types present in endometrium are endothelial cells (cells that form blood vessels) and immune cells. These immune cells can include macrophages, neutrophils, uNK and T cells (Gray et al. 2001; J. Y. Lee, Lee, and Lee 2011; Maybin and Critchley 2015). As my

transcriptional analysis was carried out on whole tissue, I will be analysing expression from multiple cell types, and this could potentially obscure relevant changes. Yanaihara et al. discovered when they compared laser capture microdissected epithelial and stromal cells in the mid-secretory phase of 8 patients, that a total of 28 genes were differentially expressed between the two cell types (Yanaihara et al. 2004). Also, Evans et al. showed epithelial and stromal cell types separated out independently in hierarchical clustering regardless of LH timing, suggesting that the expression differences between the cell types may be greater than the expression differences due to cycle stage (G. E. Evans et al. 2012).

To address this important point, I carried out RNA-seq on epithelial and stromal cell organoids. It is important to note that these organoids lost their precise receptivity state in culture (Supplementary Figure 2). Yet this data established an expression signature for each of the two main cell types (Figure 3.4B, Supplementary Table 2 and 3). The analysis shows that the samples contain a similar contribution of both cell types (Figure 3.4C), making them directly comparable in the various genome-wide approaches.

With regards to the stromal cell signature (44 genes), the expression of all genes contained within this set increased from the non-receptive to the receptive state. On the contrary, the epithelial cell signature (99 genes) quite clearly splits into two groups during this transition, with some genes being up-regulated (49 genes) and others down-regulated (50 genes) (Figure 3.5A). Similarly, Evans et al. showed that 53 genes are differentially regulated in stromal cells between LH+2 and LH+7, but 565 genes were differentially expressed between the same two time points in glandular epithelial cells (G. E. Evans et al. 2012). Unfortunately, these lists of genes are unavailable for direct comparison with my data. Yet this shows that epithelial cells are dramatically changing during this endometrial stage transition, more so than their stromal cell counterparts. Consequently, a large proportion of the transcriptional changes that are occurring can be attributed to epithelial cells.

Although the epithelial cell signature nicely discriminates non-receptive and receptive epithelial cells, it does not consider the differences that may arise between luminal and glandular epithelial cells. The luminal and glandular epithelial cells are a continuous monolayer. Luminal cells are positioned on top of the stromal cells facing the uterine lumen; at sites where the epithelium invaginates into the stroma, uterine glands are formed. It has been shown that uterine glands have differing expression patterns from

the luminal epithelium and are characterised, for example, by acetylated tubulin, which reflects their differing functions (G. E. Evans et al. 2014; Turco et al. 2017). Luminal cells make up a small proportion of endometrial tissue, however they are vital for a successful implantation.

It is worth noting that other expression studies were carried out on LCM cells, whereas I used organoids that were grown in culture (G. E. Evans et al. 2012, 2014; Salker et al. 2017; Yanaihara et al. 2004). The culturing of the organoids could lead to transcriptional changes that would not be present in the original tissue. Having said that, it has allowed us to still virtually deconvolute the tissue into two main cell types and has taken receptivity into account with regards to epithelial cells. For further development of cell signatures of endometrial cells, it would be an advantage to be able to discriminate between luminal and glandular epithelium. The genes that Evans et al. (2014) determined to be different between luminal and glandular epithelium in their transition from LH+2 to LH+7 were not present in my epithelial cell signature (apart from one, *MUC6*) (G. E. Evans et al. 2014). This is encouraging, as my epithelial cell signature denotes epithelial cells in general rather than being skewed towards luminal or glandular. Current large-scale efforts have also performed single-cell expression profiling from endometrial or decidual tissue. Such an approach would be desirable in the future, however, a detailed time-course analysis across the endometrial cycle would have major cost and logistical implications (such as computing power) that were prohibitive for my project.

The second key achievement described in this chapter is the establishment of genome-wide histone modification profiles from the very same samples for which transcriptome data were generated. To the best of my knowledge, no histone ChIP-seq experiments have been conducted on endometrial tissue to date. Some studies have highlighted that total levels of histone modifications may change throughout the menstrual cycle.

Colón-Caraballo et al. and Monteiro et al. established the total levels of histone modifications at different stages of the menstrual cycle (Colón-Caraballo, Monteiro, and Flores 2015; Monteiro et al. 2014). Although these studies showed no significant changes in the total levels in the whole tissue of H3K4me, H3K27me(3) or H3K9me,

Monteiro et al. did not discriminate between mono, di or tri methylation nor did either of them determine the genomic distribution pattern (Colón-Caraballo, Monteiro, and Flores 2015; Monteiro et al. 2014). Although, the glands specifically show a significant increase in H3K27me3 in the secretory phase compared to the proliferative phase (Colón-Caraballo, Monteiro, and Flores 2015). It is notable that these studies reported changes in some of these histone modifications when endometriosis was considered, and this is debated in Chapter V.

Tamura et al. showed that in culture, stromal cell decidualisation resulted in an increase of H3K4me3 and H3K27ac at a number of loci and that this correlated with an increase in expression especially if both marks were present, whereas there were only a handful of changes associated with H3K4me1 and H3K27me3 (Tamura et al. 2014). Grimaldi et al. also reported that upon decidualisation of cultured stromal cells H3K27me3 decreased at key genes, namely *PRL* and *IGFBP-1* (Grimaldi et al. 2016). EZH2, which is responsible for the deposition of H3K27me3, has been shown to remain at a consistent level or decrease between the proliferative and secretory stages of the menstrual cycle, this is contradictory to the reported increase of H3K27me3 in epithelial cells (Colón-Caraballo et al. 2018; Colón-Caraballo, Monteiro, and Flores 2015; Grimaldi et al. 2016).

I have reported no significant differences in the epithelial-to-stromal ratio for immunofluorescent staining of H3K27me3, suggesting that if the total levels do decrease or increase then this occurs in both epithelial cells and stromal cells equally. The discrepancies between the reports could be due to the differing techniques, sample selection criteria and timing of the biopsies used. The distribution of H3K4me3 between epithelial and stromal cells does show a trend to decrease in the receptive state, this suggests that there may be an increase in stromal cells and /or a corresponding decrease in epithelial cells in the total levels. This does correlate with stromal cells gaining H3K4me3 upon decidualisation, which could lead to an increase in the total levels (Tamura et al. 2014). However, this requires further evidence to substantiate and determine where the differences arise.

Firstly, I confirmed that the optimised histone ChIP-seq technique that was used on snap-frozen human endometrium was successful and that I had confidence in the data. This quality control assessment was essential to conduct as only a small amount of frozen biopsy was available. Therefore, the ChIP protocol required optimisation and validation. The majority of ChIP protocols for tissue require fixation prior to

homogenisation (mortar and pestle or Dounce). I found that the homogenisation methods resulted in the majority of material being lost. Therefore, I developed a protocol that used micrococcal nuclease (MNase) digestion, which is routinely used in native ChIP protocols to fractionate the chromatin, followed by brief sonication to disrupt any remaining tissue and obtain DNA fragments of an average size adequate for the ChIP procedure. This meant that the majority of material was available for downstream applications. ChIP-qPCR confirmed enrichment for each histone modification at selected target sites as expected (Figure 3.7A). Further confirmation was carried out post-sequencing, where the enrichment can be seen in the cumulative distribution plots as well as 2Mb overview of Chr9, and correct peak calling was checked for each sample in chromstaR. This gives us confidence in the quality of the histone ChIP-seq data, and the results that arise from this data. This method could also be implemented in different tissues when the availability of tissue in terms of frozen instead of fresh and size of tissue is limited.

Secondly, I have shown that bivalent domains, i.e. genes marked by H3K4me3 and H3K27me3, and their dynamic resolution to H3K4me3, play a key role during the important transition from the non-receptive to the receptive state of the menstrual cycle. I have also shown that these epigenetic changes correlate with higher expression levels of the associated genes in the receptive phase (Figure3.9D). The genes that resolve to H3K4me3-only are significantly enriched for functional categories including extracellular exosomes, as well as mitochondria and metabolism (Figure3.9E). All of these processes have been implicated as important in the preparation for a potential implantation. Specifically, exosomes have been suggested to be involved in cell-cell communication and may even be taken up by the blastocyst (Altmäe et al. 2017; Greening et al. 2016). Interestingly, a previous study had also highlighted that genes highly expressed at the receptive state are enriched for roles in exosomes, corroborating my general findings that extended this data to the epigenomic marking of this cohort of genes (Altmäe et al. 2017).

The same regions that resolve to H3K4me3-only are also enriched for Kruppel-like factor 9 (KLF9) and Forkhead box O1A FOXO binding sites (Figure3.9F, Supplementary Table 4) KLF9 and FOXO1 are expressed in both the non-receptive state and the receptive state (Supplementary Figure 3). As the non-receptive state is during the early secretory phase these factors could be upregulated by the increasing levels of progesterone during this period. KLF9 is a transcription factor that interacts with and is

co-regulated by progesterone receptor B, influencing its target genes (D. Zhang et al. 2002). KLF9 null females crossed with WT males produced fewer offspring than control crosses, most likely due to a reduced number of implantation sites (Simmen et al. 2004). Also, several genes in the uterus of KLF9 null mice showed partial progesterone resistance, suggesting that KLF9 may be required for their full activation (Simmen et al. 2004).

E74-like factor 3 (ELF3) is an epithelial specific transcription factor and has been shown in hepatocellular carcinoma (liver cancer) to promote epithelial-to-mesenchymal transition, a process that occurs to some extent also in epithelial cells in the endometrium in preparation for implantation (Whitby, Salamonsen, and Evans 2018; Zheng et al. 2018). The bivalent state of ELF3 resolves to H3K4me3-only in the receptive state, which correlates with increased expression levels; the ELF3 locus also contains a KLF9 binding sites (Figure 3.10B). Therefore, it is likely that ELF3 is regulated by KLF9 and could be involved in preparing epithelial cells for a potential blastocyst implantation. Further analysis is required to verify this.

FOXO1 is also a transcription factor and has been shown to open up chromatin, which in turn allows access for other transcription factors. FOXO1 is key for reproduction, it is regulated downstream of progesterone and has been shown to be up-regulated in decidualised cultured stromal cells. Knockdown of FOXO1 during decidualisation of cultured stromal cells results in a failure to up-regulate 507 genes (Takano et al. 2007). These genes include cell-cycle inhibitors, which are required to reduce proliferation after the proliferative phase of the menstrual cycle (Takano et al. 2007). FOXO1 is also essential for the resistance to oxidative cell death and when knocked down results in a failure to upregulate *SOD2*, which is an antioxidant defence enzyme (Kajihara et al. 2006). SOD2 is one of the genes that is bivalently regulated in the non-receptive state and resolves to H3K4me3-only in the receptive state; it also contains a FOXO1 binding site (Figure 3.10A). SOD2 is a mitochondrial gene, which is a GO term that was significantly up-regulated for genes that resolve to H3K4me3-only in the receptive state (Figure3.9E).

Taken together, this Chapter demonstrates an in-depth characterisation of endometrial tissue at the key transition point to receptivity during the menstrual cycle. This is the most extensive characterisation that has so far taken place, which - moreover - was conducted on the same biopsy samples. Therefore, all genome analyses can be directly

related. Epigenetics plays a key role in the function of the endometrium as it progresses through the menstrual cycle. It would be interesting to characterise in the future the epigenomic rearrangements that take place across the whole menstrual cycle and their functional implications, as this could be very informative to determine if they contribute to infertility or recurrent miscarriage due to failing implantation and decidualisation, respectively.

# Chapter IV

# Hydroxymethylation is a major hallmark of receptivity in endometrial tissue

## 4.1 Introduction

In the previous chapter, I conducted an in-depth analysis of the expression profiles and several key histone modifications. To continue to delve into the epigenomic landscape of the endometrium specifically at the key tipping point of receptivity, in this chapter I investigated DNA methylation (5mC) and hydroxymethylation (5hmC).

5mC is generally considered a silencing epigenetic mark and is most prominent at CpG sites within mammalian DNA. 5mC is established by DNMTs, *de novo* methylation via DNMT3A and DNMT3B and maintenance methylation via DNMT1 (Yong, Hsu, and Chen 2016). The removal of 5mC can either be by passive or active means. Passively, this occurs through replication and cell division when the 5mC mark is not restored on the daughter DNA strand, therefore diluting the presence of 5mC. DNA methylation can also be actively removed; this involves the TET enzymes which convert the methyl group to a hydroxymethyl group resulting in 5hmC. This can subsequently be further converted to 5-formylcytosine (5fC) and 5- carboxycytosine (5caC) by TETs, which can be removed in pathways involving base excision and DNA repair mechanisms to result in an unmodified cytosine (C) (Jones 2012; Weber et al. 2016).

The frequency of CpGs is unevenly distributed across the genome (Bird 2017). Overall, this dinucleotide is statistically under-represented, but accumulates in certain stretches of DNA termed CpG Islands (CGIs). Approximately half of all CGIs are located within the promoter region of genes and are often hypomethylated. These genes are commonly housekeeping genes that need to be continuously expressed in a broad spectrum of cell types. Promoters lacking CGIs are often (~35-40%) hypermethylated at the few CpGs that are present (Laurent et al. 2010). Moreover, CpG methylation accumulates over gene bodies of actively transcribed genes, this results due to the H3K36me3 modification that is enriched over the transcriptional unit and interacts with the *de novo* methyltransferases DNMT3A and DNMT3B (Rondelet et al. 2016).

5mC also is important in genome stability through the suppression of repetitive elements, often in conjunction with H3K9me3 (Jones 2012; Rothbart et al. 2012; Vilain et al. 1999). Another prominent role of DNA methylation is in the regulation of imprinted genes and in X-chromosome inactivation (Barlow 2011; Disteche and Berletch 2015; Ferguson-Smith and Bourc'his 2018; Paulsen and Ferguson-Smith 2001).

The relationship between 5mC and gene expression is genomic context dependent. Hypermethylation at promoters is usually associated with repression of the associated gene, whereas hypermethylation over gene bodies has been associated with expression, as mentioned above (Andresini et al. 2016; Mellen et al. 2012). Recent review articles highlight the complex and highly context-dependent relationship between DNA methylation and gene expression, and also highlight the role of DNA methylation in splicing regulation, nucleosome positioning, and the recruitment or expulsion of transcription factors (Long, Smiraglia, and Campbell 2017; Tirado-Magallanes et al. 2017).

5hmC is a turnover product of 5mC but can also serve as a stable epigenetic modification in its own right (Bachman et al. 2014). 5hmC is abundant in several cell types, including embryonic stem cells (ESCs) and Purkinje neurons. Although 5hmC production requires 5mC as a substrate, the levels of 5hmC do not necessarily correlate with 5mC and it can act as a tissue-type identifier (Kriaucionis and Heintz 2009; Nestor et al. 2012; Senner et al. 2012; Tahiliani et al. 2009). 5hmC at promoters positively correlates with gene expression levels (Ficz et al. 2011). Promoters that are marked by H3K4me3-only or that are bivalently marked by H3K4me3 and H3K27me3 in ESCs also tend to be enriched for 5hmC (Ficz et al. 2011; Pastor et al. 2011).

A number of studies have investigated whether methylation changes occur across the menstrual cycle. It has been suggested that the 5mC profiles can segregate out the differing phases of the menstrual cycle, although individual samples within the study groups did not hold up to the described change (Houshdaran et al. 2014; Kukushkina et al. 2017; Saare et al. 2016). Differentially methylated CpGs have been identified between the early and mid-secretory phase (Houshdaran et al. 2014; Kukushkina et al. 2017; Saare et al. 2016). The differential methylation was not unidirectional but occurred in both directions with a similar number of sites either gaining or losing methylation. Gene ontology terms associated with genes close to these methylation changes were related to tissue function. For example, terms enriched in the early-secretory vs mid-secretory phase transition included blood vessel morphogenesis, extracellular matrix organisation and immune response (Houshdaran et al. 2014; Kukushkina et al. 2017). Some of the differentially methylated sites were negatively correlated with transcriptional changes (i.e. hypermethylated and repressed, hypomethylated and induced), where data was available, such as *PTPRN2* which plays a role in mediating vesicle secretions and *MPP7*

which is a secreted protease that breaks down the extracellular matrix, both important processes during the receptive stage (Houshdaran et al. 2014; Kukushkina et al. 2017).

Isolated and cultured endometrial stromal cells have also been investigated for 5mC changes before and after the induction of *in vitro* decidualisation. Methylation analysis showed that the samples of the same patient source clustered together rather than grouping according to pre- and post-decidualisation state (Dyson et al. 2014). However, there were still a number of CpGs that were differentially methylated in line with the decidualisation state, with the majority being hypomethylated upon decidualisation (Dyson et al. 2014).

Overall the data suggests that methylation profiles remain relatively similar between the non-receptive and receptive state, with only a few reported changes that encompass both a gain and a loss of methylation. The finding that cultured stromal cells predominantly undergo a loss of methylation upon decidualisation may suggest that methylation changes occur in a cell type-specific manner.

DNA methylation analysis in all these studies was carried out using Illumina Infinium Human-Methylation platforms (Dyson et al. 2014; Houshdaran et al. 2014; Kukushkina et al. 2017; Saare et al. 2016). These platforms require bisulphite-converted DNA and the array covers a subset of CpGs (450k or 27k compared to 28billion CpGs in the genome). Most importantly, this method does not discriminate between 5mC and 5hmC. To the best of my knowledge, the overall distribution and potential cycle-dependent changes of 5hmC have not been investigated to date.

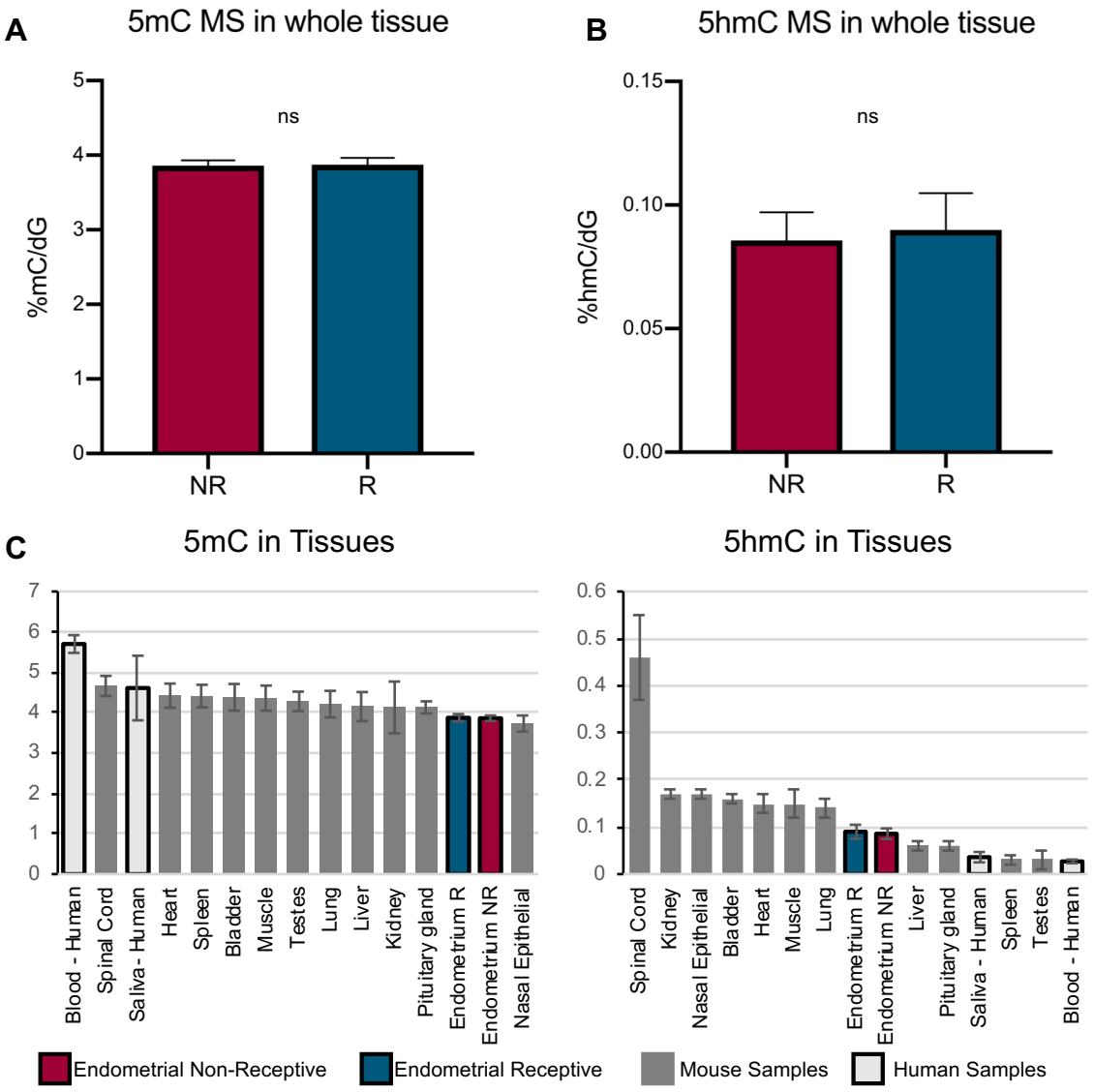
In this chapter, I investigated the genome-wide distribution of 5mC and 5hmC in the same 20 endometrial biopsy samples described in the previous chapter. I also pursued a particularly interesting candidate in functional analyses. Since it is very difficult to undertake such functional studies on primary patient-derived cells, a well-established immortalised cell line was used. There are several immortalised epithelial cell lines that are derived from endometrial cancers available including, HEC1A, Ishikawa, RL95-2 and EEC-1. Each cell line has differing characteristics, representing a non-receptive or receptive state, and exhibiting varying levels of responsiveness to hormones, adhesion and differentiation (Lessey, Vendrov, and Yuan 2003). These points need to be considered when deliberating the experiment to be undertaken and which cell line to use. Although these systems are not ideal and tissue culture itself has technical

difficulties, these epithelial cell lines still allow for mechanistic aspects of gene function in endometrial tissue to be studied.

## 4.2 Identification of cell-type specific 5hmC levels between the non-receptive and receptive state

To gain an overview of total DNA methylation levels in endometrial tissue, I first performed 5mC and 5hmC mass spectrometry on the whole tissue biopsies. This work was carried out in collaboration with Prof. Petra Hajkova's group at Imperial College London, whose lab has established an ultrasensitive liquid chromatography-mass spectrometry method for the reliable absolute quantification of 5mC and 5hmC from very small cell numbers (Amouroux et al. 2016). This analysis revealed that the total 5mC levels were 3.88% mC/dG and 3.87% mC/dG for non-receptive and receptive samples, respectively (Figure 4.1A). Total 5hmC levels were 0.086% hmC/dG and 0.09% hmC/dG for the non-receptive and receptive state, respectively (Figure 4.1B). I concluded that, overall, the total levels of 5mC and 5hmC do not differ significantly between the non-receptive and receptive states of the endometrium.

To get an understanding of how these levels compare with other tissues or even species, I used 5mC and 5hmC quantification data from the mouse and the relatively few available human data points. Comparatively, the endometrium has similar 5mC levels to a range of other tissues, including heart, kidney and nasal epithelium that contain on average ~4-5% total 5mC (Globisch et al. 2010; Godderis et al. 2015) (Figure 4.1C). The 5hmC levels fall within the medium range, which includes the mouse pituitary gland and liver containing ~0.05-0.10% total 5hmC (Figure 4.1C) (Globisch et al. 2010; Godderis et al. 2015). Interestingly, the two human data points of blood and saliva have a very low ratio of 5hmC: 5mC (<0.008), whereas the endometrial tissue has a higher ratio of 5hmC: 5mC (~0.02) (Figure 4.1D). This ratio discrepancy is caused by both, lower 5hmC as well as higher 5mC levels in human blood and saliva compared to endometrium (Figure 4.1C). Given that 5hmC is commonly associated with cell plasticity and stem cells, this may mean that comparatively, endometrium exhibits a higher degree of cellular plasticity than other human tissues (Koh et al. 2011).



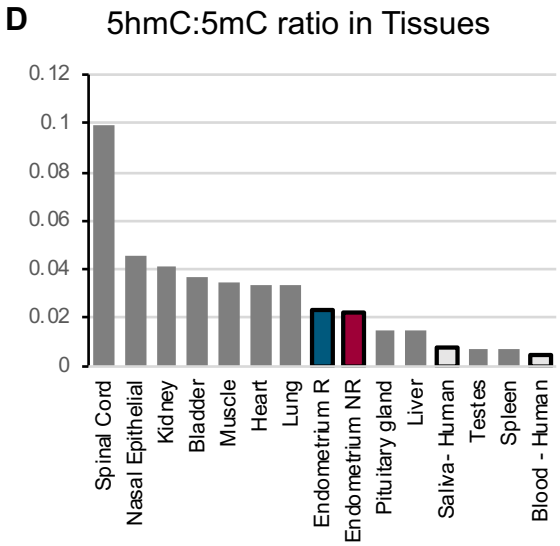




**Figure 4.1 5mC and 5hmC mass spectrometry of the endometrium. A** and **B**, Mass spectrometry of 5mC and 5hmC of endometrial biopsies in the non-receptive and receptive state. Two-way ANOVA followed by Tukey's post-hoc test showed no significant difference, p = 0.8801 and 0.3723 respectfully. NR = non-receptive (n=7), R = receptive (n=11) **C**, Mass spectrometry of mouse and human tissue (Globisch *et al.*, 2010; Godderis *et al.*, 2015), including my endometrial samples, for 5mC and 5hmC. Mean +/- SD plotted. **D**, The ratio of 5hmC:5mC in mouse and human tissue (Globisch *et al.*, 2010; Godderis *et al.*, 2015) along with endometrium. Mean plotted.

As shown in Chapter III, endometrial tissue is not homogeneous and consists of two main cell types, epithelial and stromal cells (Figure 3.4A). To determine if the distribution of DNA methylation changes between the two cell types, I conducted immunofluorescence stainings (IF) using antibodies against 5mC and 5hmC on the endometrial biopsy samples. Whereas 5mC levels did not exhibit any significant differences in the ratio of epithelial: stromal cells between the receptive and non-receptive stages (Supplementary Figure 4), the relative abundance of 5hmC was remarkably different depending on cell type and cycle stage. Specifically, a significant difference in the distribution of 5hmC was observed between the non-receptive and receptive state (Figure 4.2A). Visually, epithelial cells in receptive samples exhibited far higher 5hmC levels than epithelial cells in non-receptive samples (Figure 4.2A). This was specifically evident in the image overlay with the DAPI channel that made endometrial glands stand out as far less green (for 5hmC staining) than surrounding stromal cells.

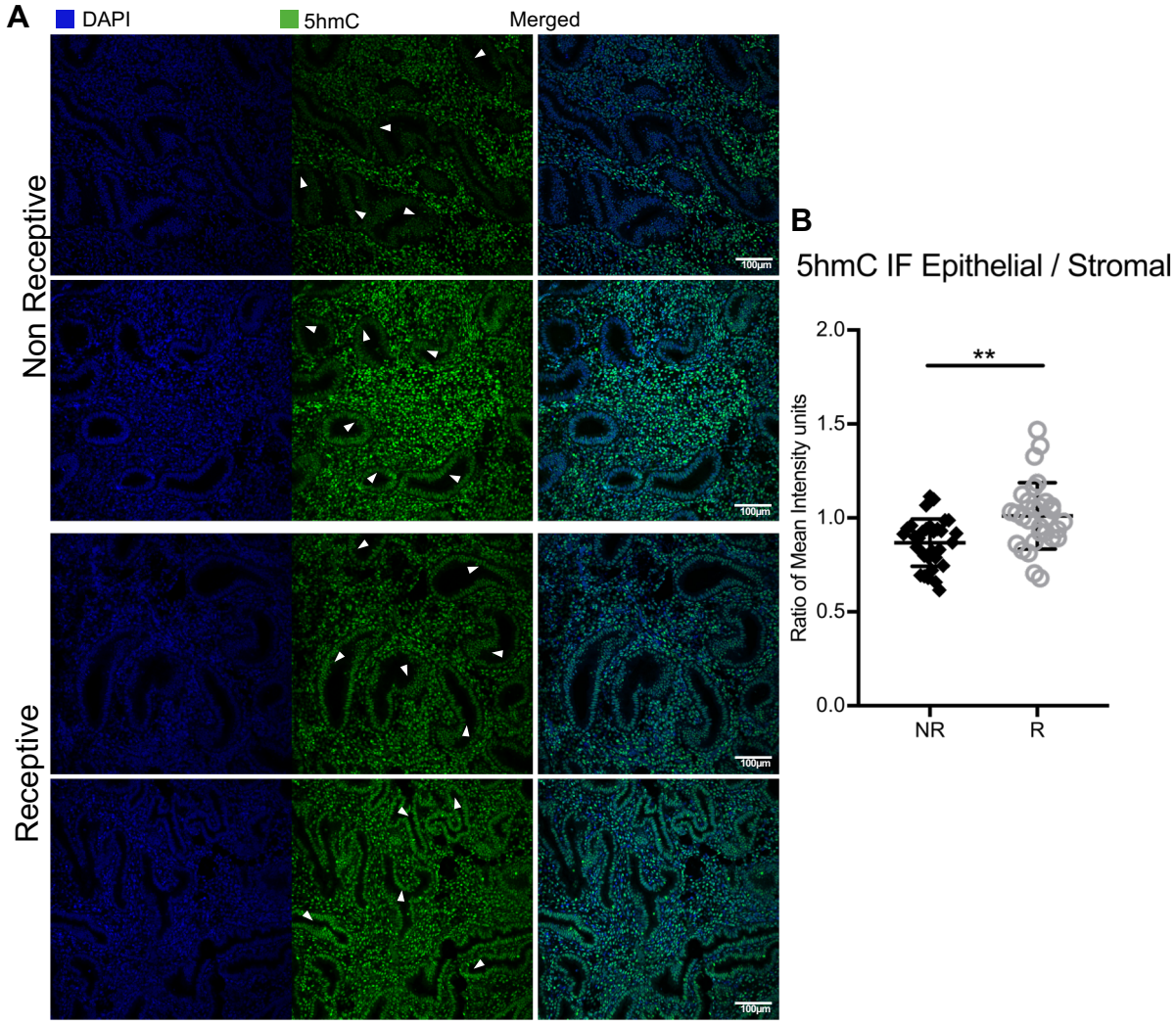



**Figure 4.2 5hmC distribution in non-receptive and receptive endometrium. A**, Immunofluorescence staining of endometrial tissue in the non-receptive and receptive state for 5hmC (green) with DAPI (blue) as nuclear counter-stain. Epithelial cells are denoted by

white arrows. Scale bar = 100μm **B**, The epithelial: stromal cell ratio of the mean fluorescence intensity units for 5hmC (individual data points and mean +/- SD plotted). Non-receptive samples: n=6, at least 2 images per biological replicate; Receptive samples: n=8, at least 2 images per biological replicate. Two-way ANOVA followed by Tukey's post-hoc test p= 0.0010.

To gain a quantitative, unbiased measure of this change, I calculated the nuclear mean intensities of 5hmC in epithelial cells over stromal cells using the CellProfiler and Excel software. Because this value was determined for every image, using the intensity measures of immediately adjacent cells, it accounted for possible overall staining intensity variations between different tissue sections and rounds of staining. Importantly, this epithelial-to-stromal ratio of 5hmC intensities was significantly higher in receptive than in non-receptive samples (Figure 4.2B).

The dramatic change in 5hmC was an unexpected discovery. To validate this observation, I used laser capture microdissection (LCM) to isolate the epithelial cells directly from these sections (Figure 4.3A). Glands from two to three sections were pooled, DNA isolated and used again for 5mC and 5hmC quantification by mass spectrometry. Obviously, due to the LCM procedure, the amount of DNA retrieved was very limited therefore 5hmC quantities extremely low. Nevertheless, the mass spectrometry procedure revealed a trend for increased amounts of 5hmC in epithelial cells in the receptive state compared to the non-receptive state (Figure 4.3B). There is a higher variability of 5hmC in the receptive state, although all samples tend to be higher, this could reflect the transient nature of 5hmC. Moreover, it also indicated a corresponding decrease in 5mC (Figure 4.3C). Although this change in 5mC was not observed by IF *per se*, the liquid chromatography-mass spectrometry procedure is far more sensitive and can detect even small changes. A decrease in 5mC in conjunction with higher 5hmC levels may be expected from a higher turnover of 5mC into 5hmC. Overall, the mass spectrometric analysis largely confirmed the IF results that showed a dramatic increase in 5hmC in the receptive state specifically in epithelial cells. This finding opened up new avenues in terms of the role of the epigenome throughout the menstrual cycle and in particular during the WOI.

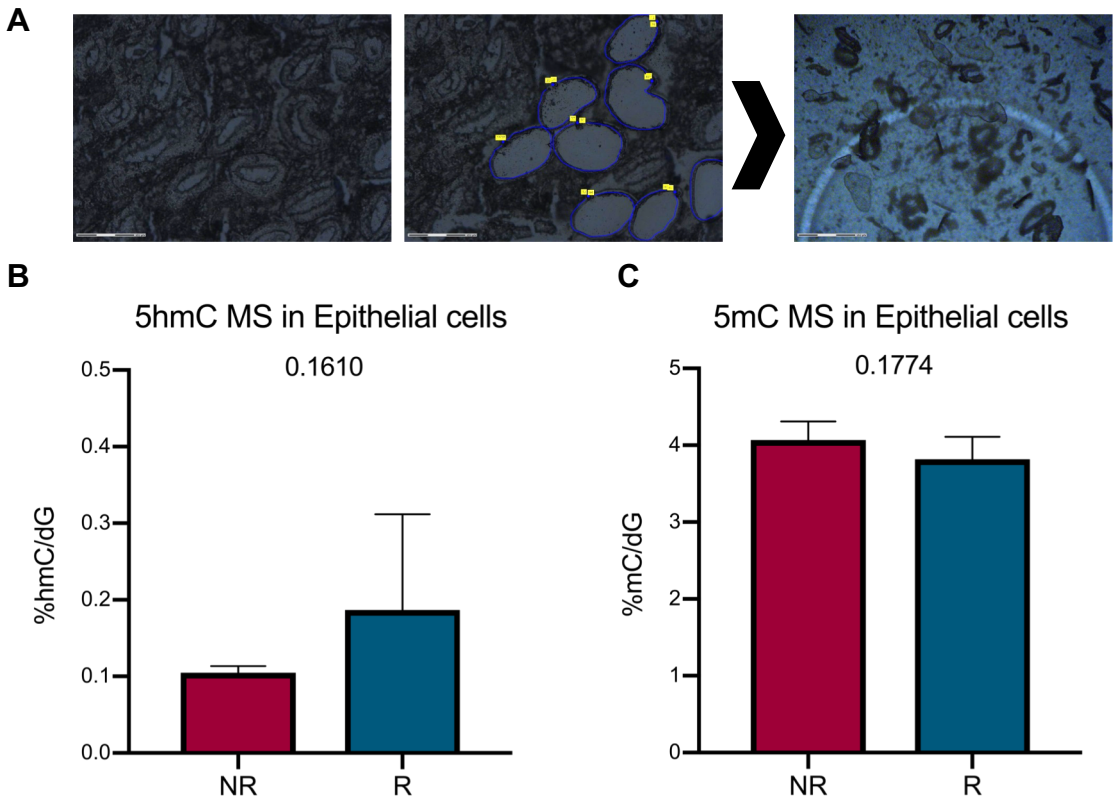



**Figure 4.3 Laser capture microdissection coupled with mass spectrometry. A**, Images from section used for laser capture microdissection (LCM) of epithelial cell in the endometrium in the non-receptive and receptive state. Left image: Prior to LCM. Middle image: After LCM, epithelial cells laser-cut out. Right image: Epithelial cells catapulted into the lid of a 0.5ml Eppendorf tube. **B** and **C**, Mass spectrometry of LCM epithelial cells in the NR (non-receptive) and R (receptive) state to determine absolute quantities of 5hmC and 5mC. n= 6 for both non-receptive and receptive samples. Two-way ANOVA followed by Tukey's post-hoc test showed p = 0.1610 and 0.1744 for 5hmC and 5mC, respectively. Mean+/-SD plotted.

## 4.3 The role of 5hmC in Epithelial cells of the Endometrium

As mentioned in the previous section, I performed DNA methylation and hydroxymethylation profiling on all of the whole tissue biopsies. These data will be described in the next chapter. Here, because of the cell type specificity of the 5hmC changes, I will specifically focus on a specialised procedure of 5hmC mapping that I performed as part of my industrial placement at Cambridge Epigenetix (CEGX). This proprietary procedure, termed hydroxymethylation capture pull-down sequencing (HMCP), was chosen because it is particularly powerful to determine 5hmC profiles from very low input material.

HMCP is a technique developed by CEGX to look specifically at 5hmC distribution across the genome, it does not require oxidative bisulphite conversion and has been optimised for very low amounts of DNA. Therefore, in collaboration with CEGX, I performed HMCP on endometrial epithelial cells isolated by LCM from paraffin sections of 6 of the patients, 3 of which had been determined to be in the non-receptive state and 3 in the receptive state. All these samples were from control patients and will be referred to as non-receptive (NR) and receptive (R) going forward for clarity and simplicity.

The protocol for HMCP contains spike-in controls of DNA that is either methylated, hydroxymethylated or unmodified. Quantification of the enrichment within each HMCP sample is calculated using the ratio of 5hmC control reads to the sum of 5mC and C control reads. One receptive sample had low levels of enrichment, 58.19 compared to ~120 to 360, and was therefore removed from the analysis. Cumulative distribution plots confirmed high enrichment compared to inputs in the remaining samples, so I am confident in the analysis going forward (Supplementary Figure 5A).

Assessing this data first for the overall distribution of 5hmC across various genomic features on a global level showed similar enrichment of 5hmC in both the non-receptive and receptive state (Figure 4.4A). However, using promoter regions, defined as -1kb to +100bp around the TSS (Figure 4.4B) for unsupervised clustering, the receptive samples segregated away from the non-receptive samples. The same separation was achieved when the unsupervised clustering was performed with CGIs (Supplementary Figure 5B). This suggests that the genomic distribution of 5hmC in epithelial cells is distinct for the receptive state and could be used as a marker of this timing.

It has previously been shown that, in general, 5hmC enrichment at promoters correlates with an increase in gene expression (Ficz et al. 2011). Indeed, this is the case also in the endometrium. 5hmC was enriched at promoters of genes that are more highly expressed in the receptive state, with the non-receptive state showing the same trend (Figure 4.4C and D). This correlation with expression is far greater in the receptive state, suggesting that 5hmC may play a more prominent role in the 'switching on' of genes in this state compared to the non-receptive state.

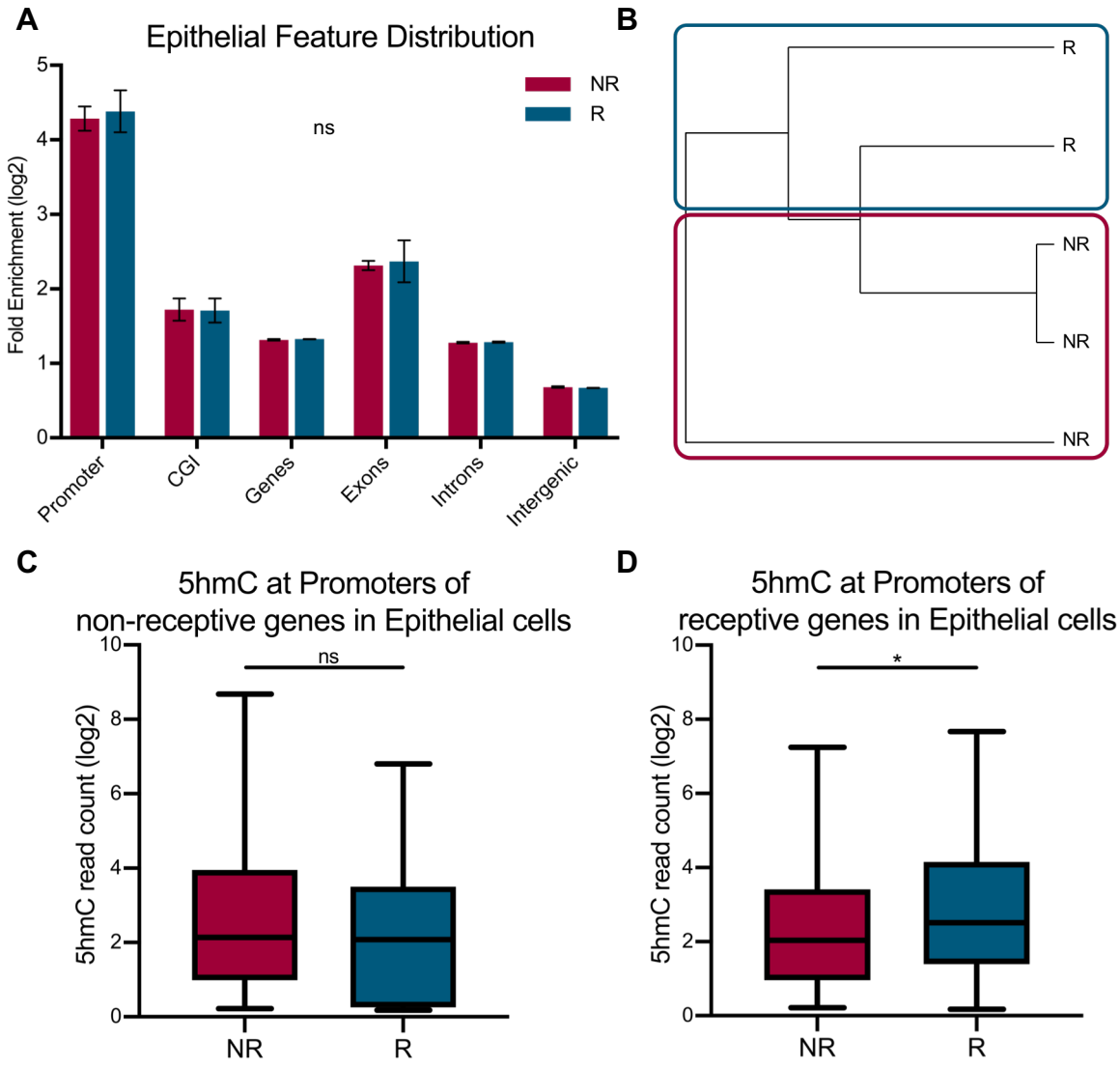



**Figure 4.4 DNA hydroxymethylation analysis of laser microdissected endometrial epithelial cells. A**, The relative 5hmC enrichment of different genomic features. Features were determined using GRCh38 in SeqMonk with enrichment calculated by: log2% reads / % genome. Unpaired two-tailed t-test was performed for each feature, no significant differences were observed. **B**, Neighbour joining tree, with distances calculated by Pearson correlation, showing the degree of relatedness between epithelial cells in either the non-receptive or receptive state. **C** and **D**, LCM-HMCP read count at promoters (-1kb to +100bp around TSS) of genes specific for the non-receptive (C) or receptive (D) state, as determined by my endometrial receptivity (ERT) signature gene set described in Chapter III. * p = 0.0150 (unpaired t-test).

As expected from this positive correlation, genes that are highly expressed in the respective state are also highly enriched for 5hmC. Progesterone Associated Endometrial Protein (*PAEP*) and Nicotinade N-Methylatranserase (*NNMT*) have been identified in this and previous studies as being key genes that are upregulated in the receptive state (Figure 3.3B, Díaz-Gimeno *et al.*, 2011; Ruiz-Alonso, Blesa and Simon,

2012). Here, I show that *PAEP* and *NNMT* also exhibit a significant increase in 5hmC at their promoters (Figure 4.5A and B). Furthermore, 5hmC levels are increased across large regions of the *PAEP* gene, including an alternate transcription start site (Figure 4.5A).

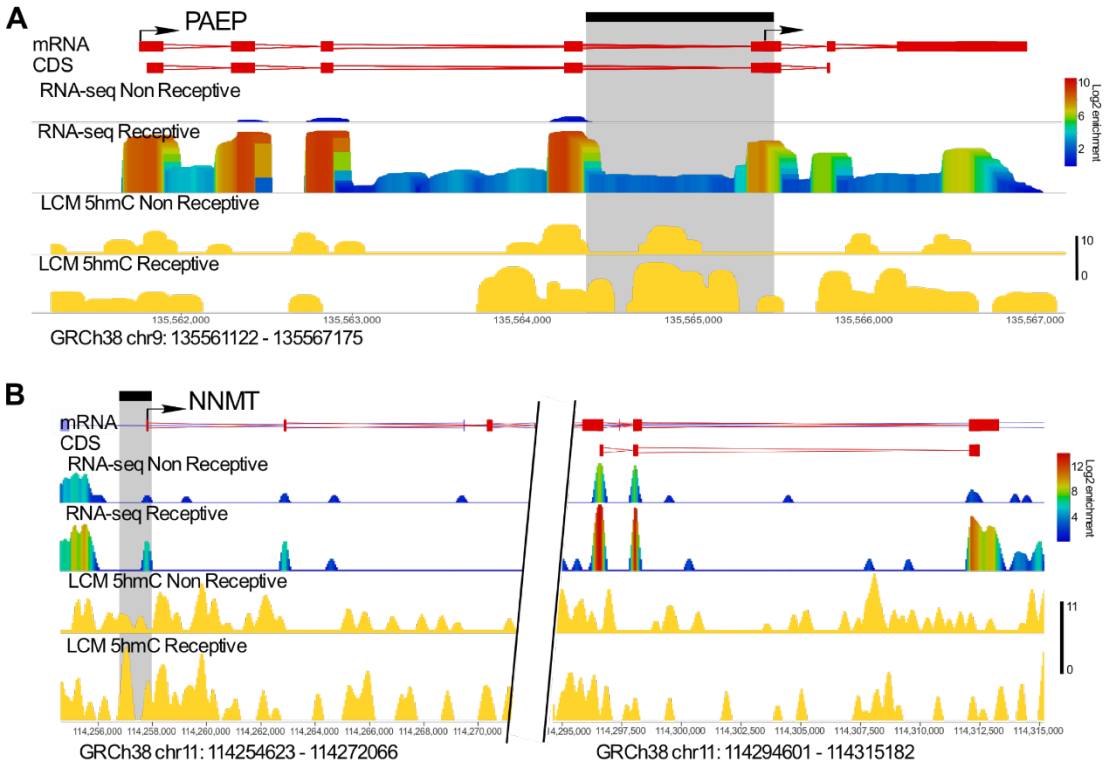



**Figure 4.5 5hmC is enriched at promoters of specific genes. A** and **B**, Wiggle plots for *PAEP* (A) and *NNMT* (B). RNA-seq tracks of whole tissue (RPM log2-transformed) and LCM-HMCP of epithelial cells (log2-transformed) for non-receptive and receptive states. Promoters (-1kb to +100bp around TSS) are highlighted by black box, with greyed out area underneath. mRNA and coding sequence (CDS) displayed. Tracks are merged results of non-receptive n=3 (RNA-seq or corresponding LCM 5hmC), receptive n=2 (RNA-seq or corresponding LCM 5hmC) independent biological replicates.

5hmC in mouse and human ESCs has been shown to co-localise with bivalent regions (Ficz et al. 2011; Pastor et al. 2011). Therefore, I investigated the relationship between bivalent gene loci and 5hmC in my samples. In non-receptive epithelial cells, bivalently marked promoters harbour a similar level of 5hmC as all promoters across the genome. However, bivalent promoters that resolve to H3K4me3-only in the receptive state have a significantly higher level of 5hmC enrichment in the receptive state compared to all promoters ($p<0.0001$, Mann-Whitney test) (Figure 4.6A). The bivalent promoters were further subdivided into promoters that remained bivalent or resolved to H3K4me3-only in the receptive state. I then plotted the 5hmC enrichment profiles across a region of

+/-5kb surrounding the TSSs (Figure 4.6B). Promoters that resolved to H3K4me3-only in the receptive state were significantly enriched for 5hmC compared to promoters that remained bivalent ($p<0.0001$, Mann-Whitney test). What is intriguing to note though, is that promoters that are going to resolve to H3K4me3-only were already significantly enriched for 5hmC in non-receptive epithelial cells compared to ones that remain bivalent, prior to their activation ($p<0.0001$, Mann-Whitney test) (Figure 4.6B). The enrichment of 5hmC at a subset of bivalent promoters in the non-receptive state precedes, and may even be required for, their resolution to H3K4me3-only upon transition to the receptive state.

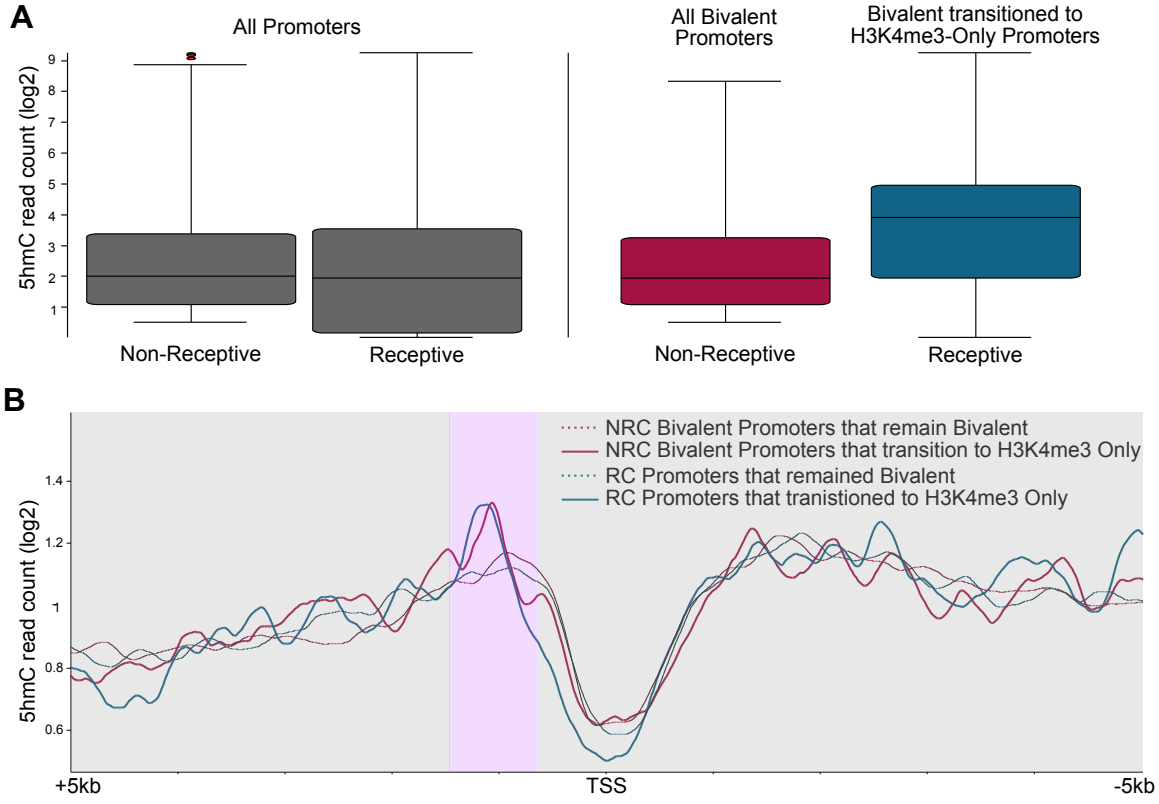



**Figure 4.6 5hmC at bivalently marked promoters. A**, LCM-HMCP enrichment at all promoters in non-receptive and receptive epithelial cells (grey boxes), bivalent promoters in non-receptive epithelial cells (maroon box) and bivalent promoters that resolved to H3K4me3-only in receptive epithelial cells (blue box). Middle line is the median, boxes edges are the 25$^{th}$ and 75$^{th}$ percentile, whiskers = median +/- the interquartile (25-75%) range x2, dots indicate outliers of this range. **B**, LCM-HMCP epithelial trend plot +/- 5kb of TSS. Dotted lines are promoters that remained bivalent in the receptive state. Solid lines are promoters that resolved to H3K4me3-only in the receptive state. Non-receptive= maroon, Receptive= blue. Region of interest is illustrated by violet box.

## 4.4 Investigating mechanisms underlying the 5hmC dynamics

As 5hmC plays such an apparent key role in epithelial cells during the transition from the non-receptive to the receptive state, I investigated possible candidate genes that could be involved in the dynamic regulation of this epigenetic modification. The most obvious ones are the genes involved in the deposition and removal of 5mC and its derivatives, notably the DNA methyltransferases (DNMTs) and the members of the ten eleven translocation (TETs) family of enzymes. However, the expression of all these factors, i.e. *DNMT1*, *DNMT3A*, *DNMT3B*, *DNMT3L*, *TET1*, *TET2* and *TET3*, remained unchanged between the non-receptive and receptive state in the RNA-seq expression data of the whole tissue (Figure 4.7A).

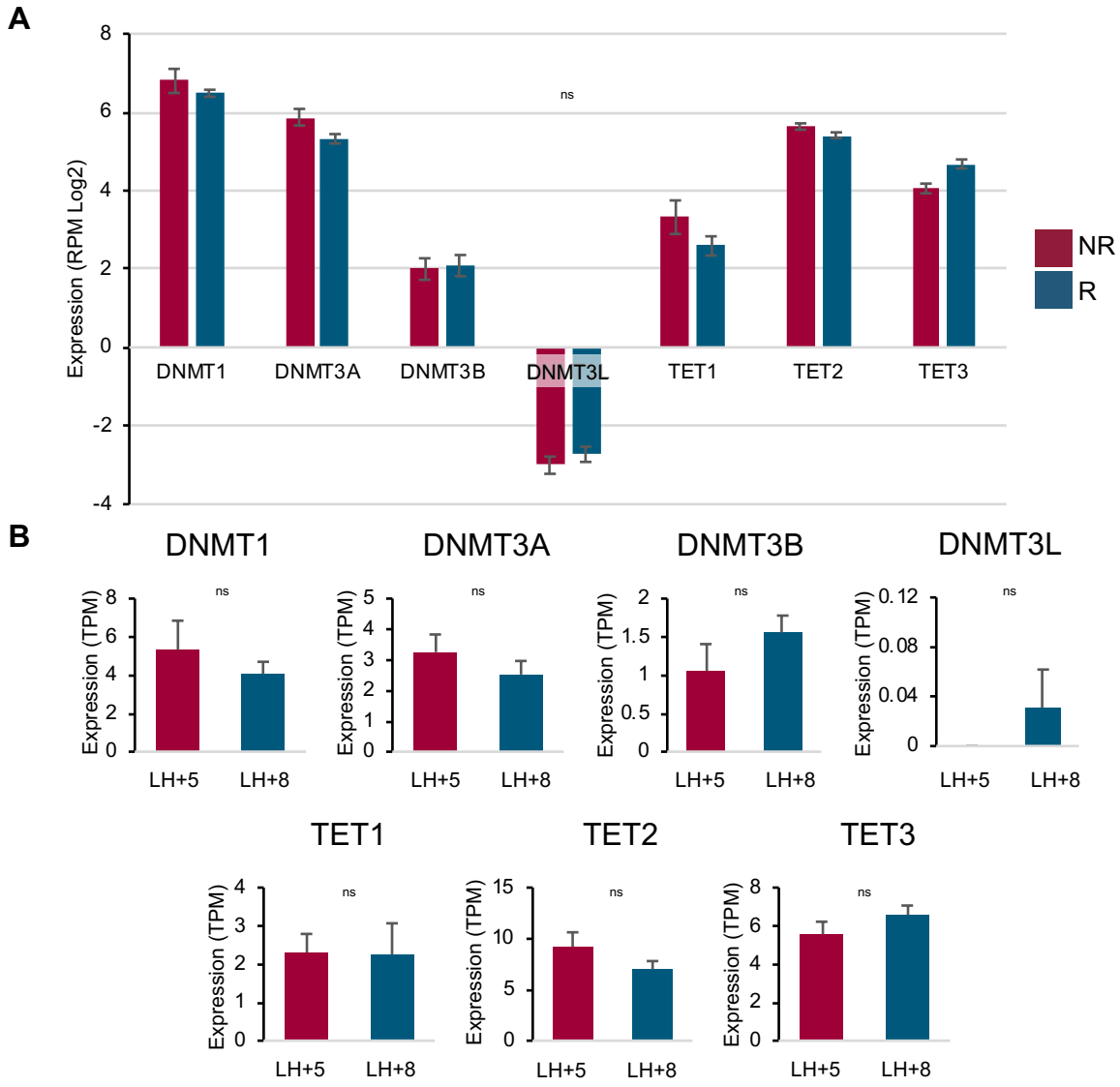



**Figure 4.7 DNA methylation and hydroxymethylation modifiers. A**, RNA-seq expression levels (RPM log2) in whole endometrial tissue of *DNMT1*, *DNMT3A*, *DNMT3B*, *DNMT3L*,

*TET1*, *TET2* and *TET3*. DESeq2 and EdgeR statistical tests with Benjamini-Hochberg correction applied showed no significant differences. Plots depict mean +/- SEM. NR = non-receptive state (n=4), R= receptive state (n=6) **B**, RNA-seq data (TPM log2) of isolated epithelial cells at LH+5 (n= 3) and LH+8 (n= 3) (Salker et al. 2017) of *DNMT1*, *DNMT3A*, *DNMT3B*, *DNMT3L*, *TET1*, *TET2* and *TET3*. DESeq2 with Benjamini-Hochberg correction applied showed no significant differences.

As the phenomenon of cycle-specific 5hmC changes occurs specifically in epithelial cells, I also analysed published RNA-seq data from laser capture microdissected epithelial cells (LH+5 and LH+8 RNA-seq data) for potential differences in expression of these epigenetic modifiers (Salker et al. 2017). This data supported the whole tissue analysis insofar as there were no significant changes in the expression of *DNMTs* or *TETs* (Figure 4.7B). Although it is possible that the protein levels and/or post translational modifications maybe altered.

There are several other important genes that have been shown to be involved in the turnover and dynamics of 5mC, 5hmC, 5fC and 5caC, e.g. *UHRF1*, *MBD2*, *POLB*, *XRCC1*, *APEX1*, *MPG*, *OGG1*, *SMUG1*, *TDG*, *NEIL1*, and *PARP1,* which are involved in the base excision and DNA repair pathways leading to 5mC turnover (Y.-J. Kim and Wilson III 2013). Expression of these genes did not differ significantly either between the receptive and non-receptive states, as determined by the whole-tissue or epithelial cell-specific RNA-seq data (Supplementary Table 5).

## 4.5 NNMT as a potential candidate regulating 5hmC levels

In the absence of expression changes of the traditional modifiers of DNA methylation dynamics, I looked further afield for possible mechanisms that may be responsible for the rather dramatic changes in 5hmC levels in epithelial cells during the menstrual cycle. A candidate for these dynamic changes in 5hmC levels was the Nicotinamide N-Methyltransferase (*NNMT)* gene. As previously established, *NNMT* is highly upregulated in the receptive state compared to the non-receptive state in my ERT signature gene set (Figure 3.3B, Figure 4.8A). *NNMT* is expressed in both epithelial and stromal cell organoids, but its expression is notably higher in epithelial organoids (Figure 4.8C), suggesting it may play a role in epithelial cells specifically. Analysis of the Salker et al

epithelial LCM data coupled with RNA-seq data of LH+5 and LH+8 confirmed the upregulation of *NNMT* in epithelial cells (Figure 4.8B).

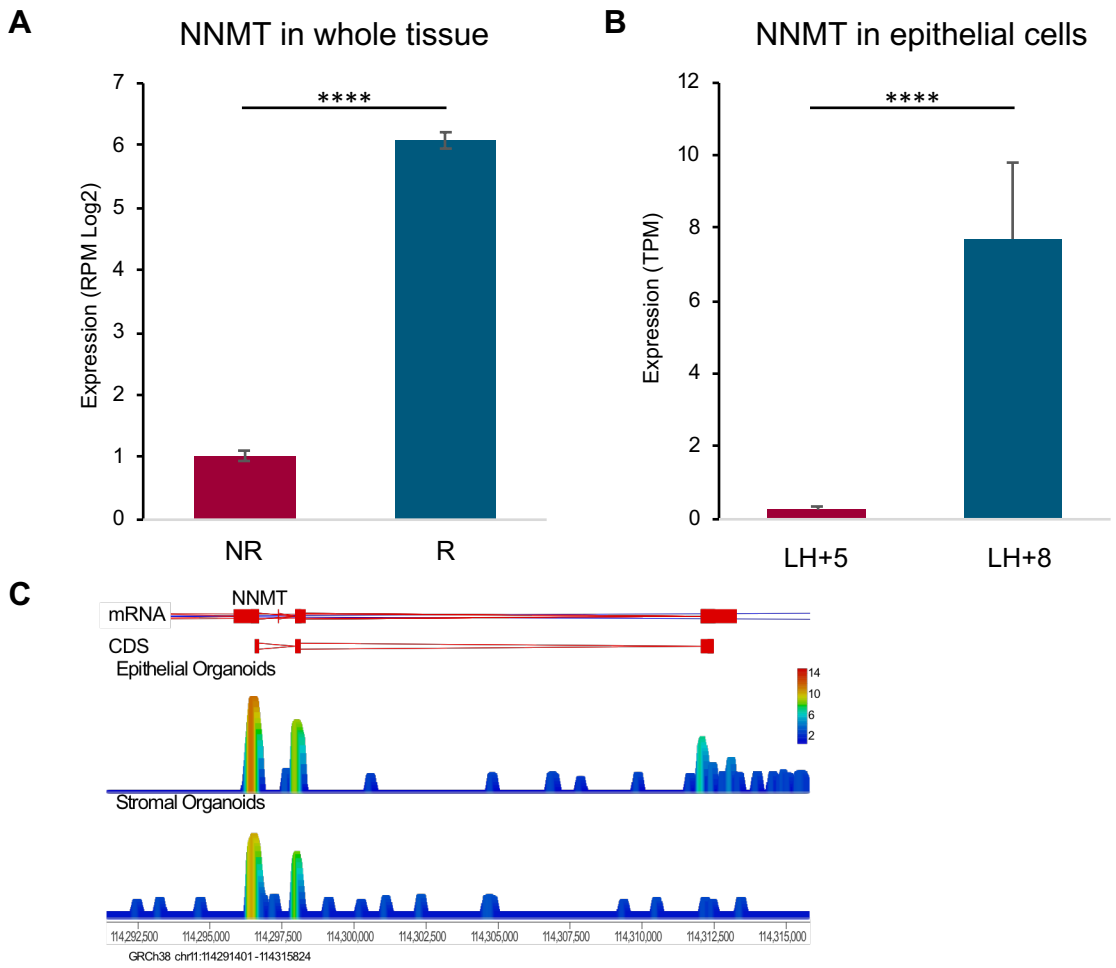



**Figure 4.8 *NNMT* expression in the endometrium. A**, The expression of *NNMT* in whole tissue endometrium in the non-receptive and receptive state. RNA-seq data analysed with DESeq2 and Benjamini-Hochberg correction applied. **** p< 0.0001. NR= non-receptive (n=4), R= receptive (n=6). **B**, *NNMT* expression in isolated epithelial cells at LH+5 (n=3) and LH+8 (n=3) of the endometrial cycle (Salker et al. 2017). RNA-seq data analysed with DESeq2 and Benjamini-Hochberg correction applied **** p< 0.0001. **C**, Wiggle plot of RNA-seq data from epithelial (n=4) and stromal cell (n=4) organoids across the *NNMT* coding sequence (CDS).

NNMT was first identified in the liver where it acts as a clearance enzyme for nicotinamide (Vitamin B3, NAM) as well as for xenobiotic and other drugs (Ellinger and Coulson 1944). As a clearance enzyme it modifies NAM via its methyltransferase activity, adding a methyl group to produce 1-methylnicotinamide (1MNA), which is then excreted. NNMT contains a methyltransferase site as well as multiple S-adenosylmethionine (SAM) binding sites (Figure 4.9A, https://www.ebi.ac.uk/interpro/protein/P40261). SAM is the main universal methyl

donor in cells and supplies the methyl group to NNMT, DNMTs and histone methyltransferases (Struck et al. 2012). When NNMT increases within a cell it has been reported that it acts as a methyl sink, meaning that methyl groups are not available to other methyltransferases, including DNMTs (Jung et al. 2017; Ulanovskaya, Zuhl, and Cravatt 2013). NNMT has also been implicated in histone methylation in human ESCs. It is highly expressed in the naïve state of pluripotency and downregulated on transition to the primed state, with a corresponding increase in H3K27me3 and H3K9me3 (Sperber et al. 2015). Together, these observations suggest that NNMT may act as a modulator of both DNA and histone methylation dynamics.

NNMT has been consistently identified in other studies as being upregulated in the WOI, however, it has not been investigated for its role in the endometrium and in particular in epithelial cells (Altmäe et al. 2017; Díaz-Gimeno et al. 2011; Talbi et al. 2006). NNMT has been studied in cancer; it is markedly over-expressed in several cancers, including ovarian, glioblastoma and colon cancers, to the extent that it has been suggested as an early detection serum marker (Jung et al. 2017; Roeßler et al. 2005; Ulanovskaya, Zuhl, and Cravatt 2013). NNMT knock-down in cancer cell lines results in a reduction in their ability to proliferate and migrate (Ulanovskaya, Zuhl, and Cravatt 2013). Interestingly, it has also been shown to be regulated by STAT3, and knock-down of *STAT3* resulted in reduced *NNMT* levels (Tomida et al. 2008). This could be relevant in the endometrium as in the mouse, STAT3 is essential in uterine epithelial cells for implantation to occur. Indeed *Stat3* null mice are infertile due to a failure of embryo attachment (Pawar et al. 2013).

With the dramatic changes of 5hmC observed with the transition to a receptive state whilst DNA methylation modifiers remained seemingly unaffected, I hypothesised that NNMT may act as a modulator of DNA methylation dynamics in the endometrium through its role as a methyl sink. Higher levels of NNMT may reduce the availability of methyl groups for DNMTs to deposit DNA methylation. As a consequence, with TET levels remaining consistent, 5mC would still be turned over to 5hmC potentially resulting in higher levels of 5hmC (Figure 4.9B). Although a direct connection between NNMT and 5hmC has not yet been described, its stark upregulation in the receptive state made it a candidate that was well worth investigating.

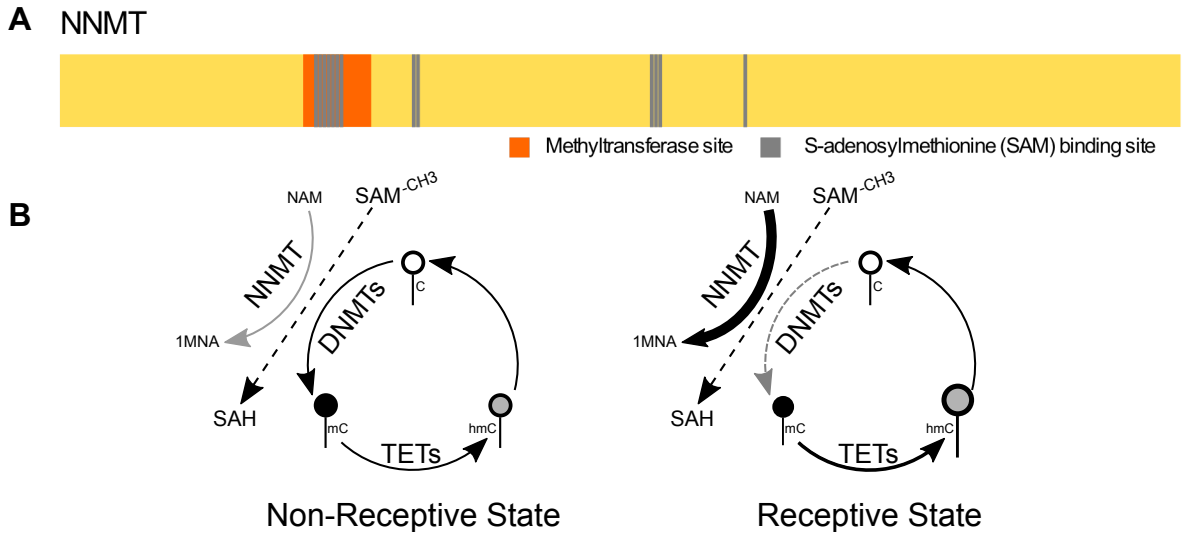
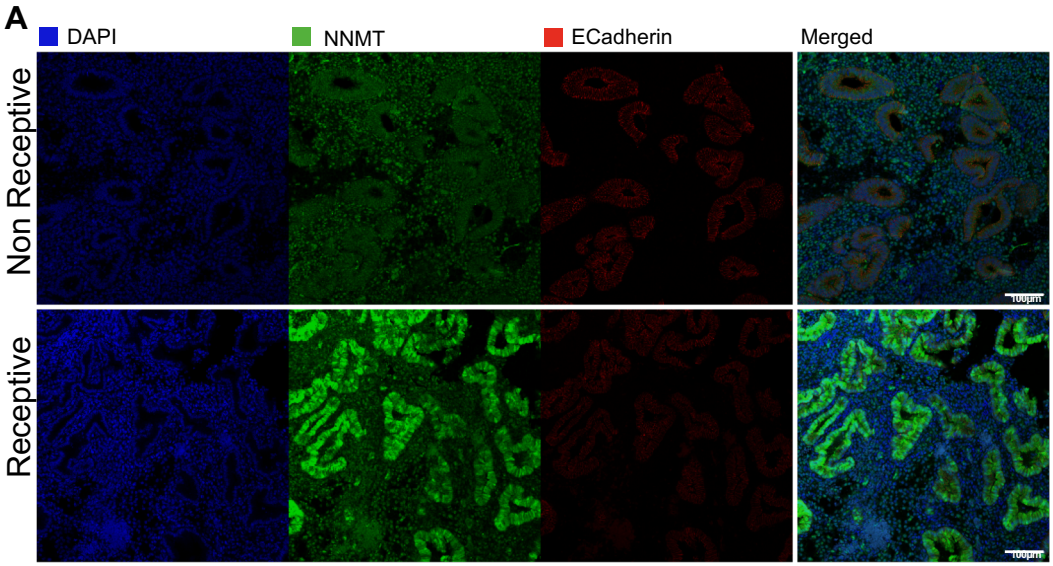



**Figure 4.9 NNMT. A**, NNMT protein structure with methyltransferase site (orange) and S-adenosylmethionine binding sites (grey) marked. **B**, Hypothesised mechanism of NNMT function within epithelial cells in the endometrium. A potential consequence of the profound up-regulation of *NNMT* in the receptive phase may be a decrease in 5mC and a relative increase in 5hmC levels.

To confirm that the upregulation of *NNMT* mRNA is also present at the protein level, I conducted IF on the paraffin embedded biopsies that corresponded to the same samples used for RNA-seq analysis. Sections were stained for NNMT and E-Cadherin. E-Cadherin allowed epithelial cells to be identified with ease for semi-automated analysis using the Cell Profiler software.

The IF staining showed that NNMT was particularly highly expressed in epithelial cells in receptive samples (Figure 4.10A). When quantifying the cellular staining intensities of NNMT in epithelial and stromal cells, it was evident from the epithelial: stromal ratio that NNMT levels were significantly increased in the receptive state (Figure 4.10B). NNMT increased in both the nucleus and cytoplasm.



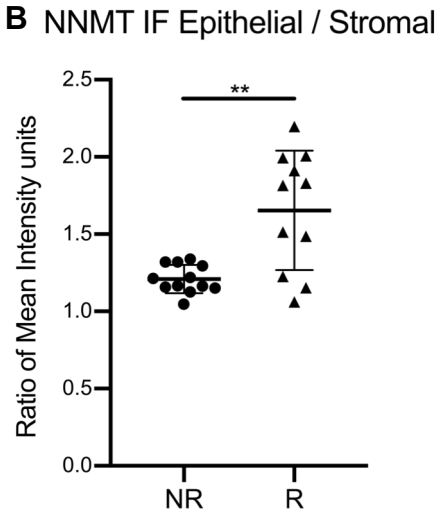




**Figure 4.10 NNMT protein levels in endometrium.**
**A**, Immunofluorescence images of NNMT in non-receptive and receptive endometrium. DAPI (blue), NNMT (green), E-Cadherin (red). E-Cadherin denotes epithelial cells. Scale bar = 100μm **B**, Calculation of the epithelial: stromal cell ratios of mean intensity units within each image. NR (n=4 with 2-3 images per a biological replicate), R (n=4 with 2-3 images per a biological replicate). Mann-Whitney two-tailed test **p = 0.0086.

To test the hypothesis that elevated NNMT levels may affect 5hmC dynamics, I overexpressed NNMT in HEC1A cells (kindly provided by Prof. Irmgard Classen-Linke at RWTH Aachen University). HEC1A cells are an endometrial epithelial cell line established from an adenocarcinoma. They have been used in several studies to represent endometrial epithelial cells in the non-receptive state (Bhagwat et al. 2013; Tamm et al. 2009). Initially, I measured the endogenous levels of 2 non-receptive genes, *KCNG1* and *GJB6*, 2 receptive genes, *PAEP* and *NNMT*, genes involved in the deposition and removal of 5mC and its derivatives, *DNMT1*, *DNMT3A*, *DNMT3B*, *DNMT3L*, *TET1*, *TET2* and *TET3* as well as housekeeping genes *GAPDH* and *HMBS* by RT-qPCR. The receptivity-associated genes including *NNMT* were very lowly expressed with DNMTs (expect *DNMT3L*), TETs and the housekeeping gene *HMBS* being highly expressed. Confirming that these cells are an appropriate model to study *NNMT* over-expression and its potential role as a modulator of DNA methylation dynamics (Figure 4.11A).

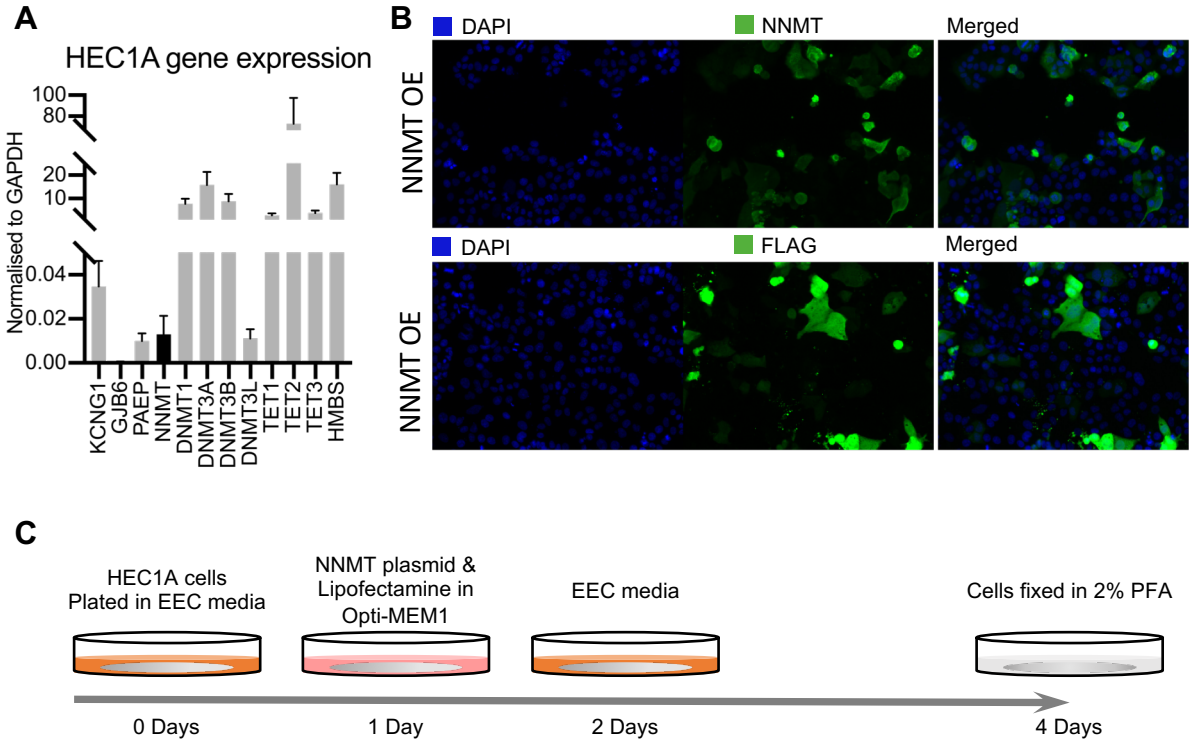



**Figure 4.11 Endogenous *NNMT* expression in HEC1A cells and proof of *NNMT* over-expression upon transfection. A**, RT-qPCR of non-receptive genes, *KCNG1* and *GJB6*, receptive genes, *PAEP* and *NNMT* (black bar), genes involved in the deposition and removal of 5mC and its derivatives, *DNMT1*, *DNMT3A*, *DNMT3B*, *DNMT3L*, *TET1*, *TET2* and *TET3* and housekeeping gene *HMBS* in HEC1A cells. Normalised to *GAPDH*.
**B**, Immunofluorescence images of HEC1A cells transfected with a FLAG-tagged NNMT expression construct. DAPI (blue), NNMT or FLAG (green). **C**, Experimental design of NNMT over-expression in HEC1A cells.

Transient transfection experiments to overexpress *NNMT* were carried out and the levels of 5mC, 5hmC, H3K4me3 and H3K27me3 measured in the transfected cells, compared to un-transfected cells. The experimental design was set up as shown in Figure 4.11C. Cells were transfected with the *NNMT* expression construct, and fixed 3 days after transfection, for use in immunostainings. This allowed for a direct comparison within the experiment between cells that were over-expressing (OE) *NNMT* and cells that had not incorporated the plasmid and, therefore, were not OE *NNMT*. The GenScript NNMT expression plasmid (NM_006169.2) that I used to transfect HEC1A cells also contained a C-terminal FLAG-tag to allow identification of transfected cells by anti-FLAG staining. To confirm that the transfected HEC1A cells were over-expressing NNMT, I stained the cells for NNMT or FLAG (Figure 4.11B). This indicated that ~15-20% of the cells were indeed over-expressing NNMT; the surrounding non-transfected cells could be used as internal controls.

HEC1A over-expressing cells were stained for FLAG and for 5mC or 5hmc. Analysis of cells that were low or negative for FLAG ("FLAG-Low") compared to "FLAG-High" cells revealed no significant differences in the levels of 5mC or 5hmC between both groups (Figure 4.12A and B). This suggests that NNMT is not acting as a modulator of DNA methylation, at least in the context of transient transfection in the cell culture system utilised here.

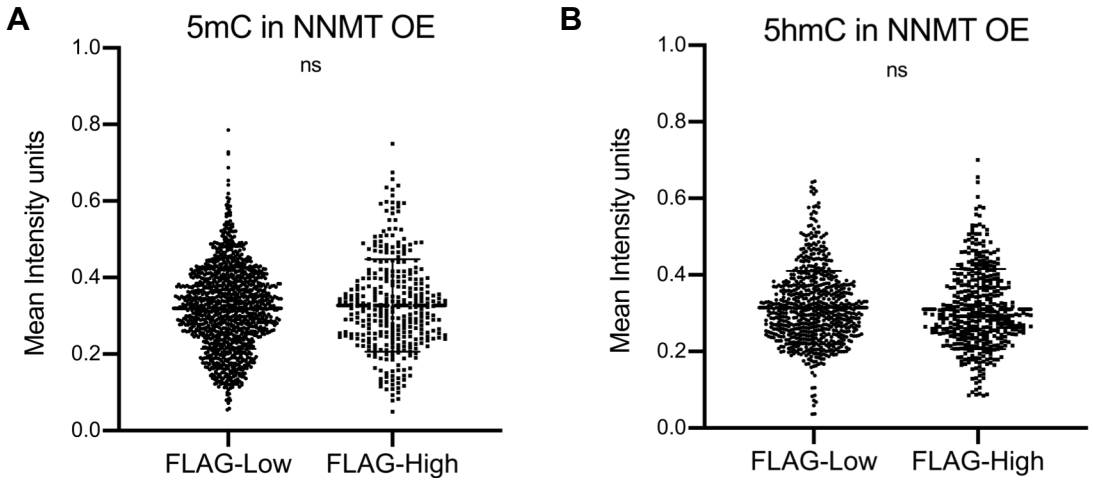



**Figure 4.12 5mC and 5hmC in HEC1A cells over-expressing NNMT. A**, The mean intensity units of 5mC in FLAG-Low (n=1236) and FLAG-High (n=332) HEC1A cells. Mann-Whitney test p=0.8541. **B**, The mean intensity units of 5hmC in FLAG-Low (n=687) and FLAG-High (n=415) HEC1A cells. Mann-Whitney test p=0.3324.

As NNMT has also been shown to play a role in histone methylation levels through its role in depleting the availability of the methyl donor, I also investigated the levels of H3K4me3 and H3K27me3 in FLAG-Low and FLAG-High HEC1A cells (Figure 4.13A). Higher NNMT levels would be expected to deplete more substrate and lead to lower amounts of histone methylation. However, I found that H3K4me3 and H3K27me3 levels were significantly increased in NNMT transfected cells (Figure 4.13B and C). This suggests that NNMT may be operating in a different way than we had hypothesised but is still affecting the deposition of methylation within the cell. This is considered in the discussion.

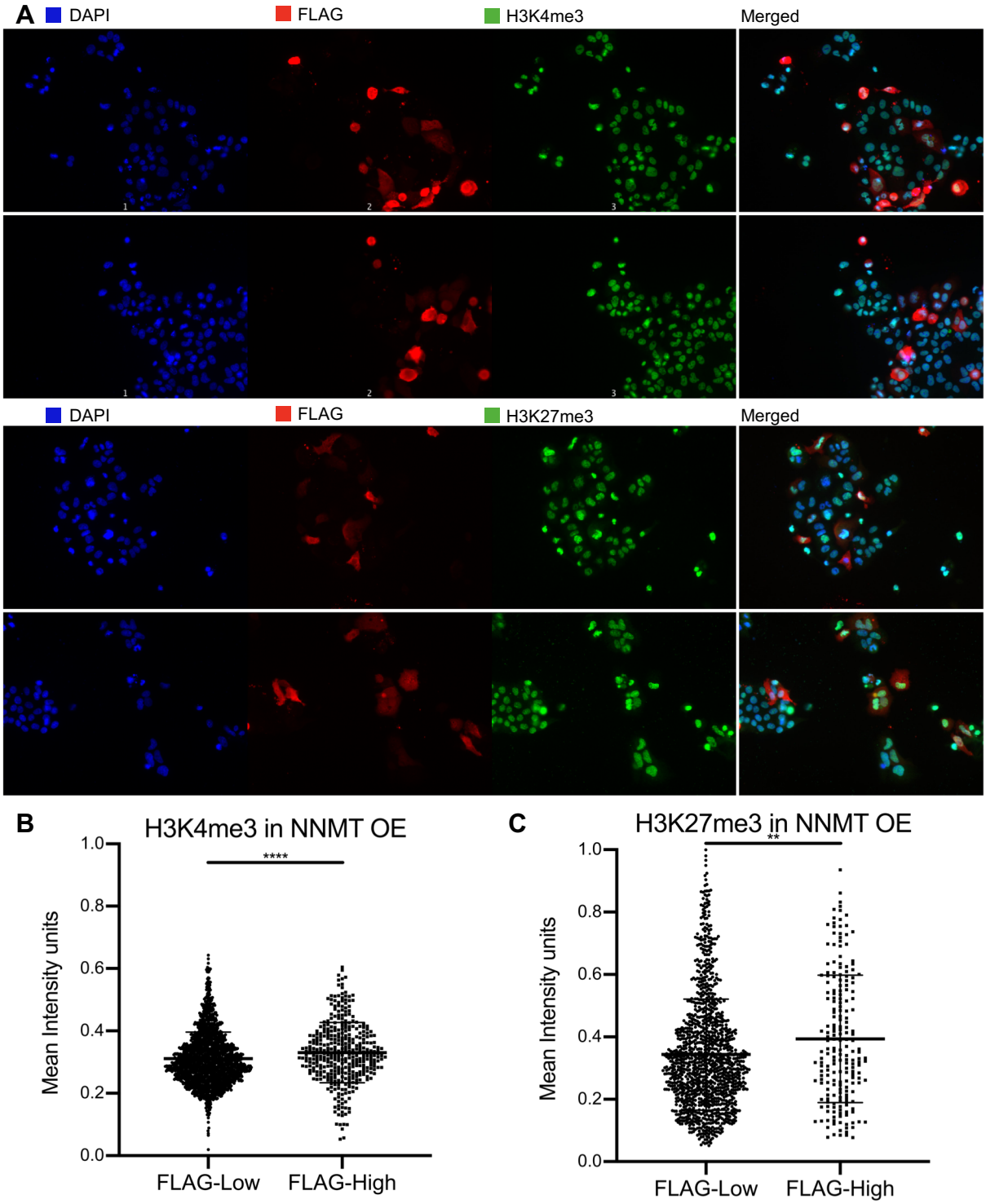



**Figure 4.13 H3K4me3 and H3K27me3 in HEC1A cells OE NNMT.**
**A**, Immunofluorescence images of HEC1A cells transfected with a FLAG-tagged NNMT expression construct. DAPI (blue), FLAG (red), H3K4me3 or H3K27me3 (green). **B**, Mean intensity units of H3K4me3 in FLAG-Low (n=1714) and FLAG-High (n=373) HEC1A cells. Mann-Whitney test **** $p<0.0001$. **C**, Mean intensity units of H3K27me3 in FLAG-Low (n=1233) and FLAG-High (n=203) HEC1A cells. Mann-Whitney test ** $p<0.01$.

Although it is worth nothing that HEC1A cells do not express genes associated with either of the receptivity states as well as the cell line being established from an adenocarcinoma in a 71-year old women. This could result in the cells having a disrupted epigenome and responding abnormally to the increase in NNMT. This is further expanded upon in the discussion.

## 4.6 Discussion

In this chapter, I established the amount of 5mC and 5hmC in endometrial tissue in the non-receptive and receptive state (Figure 4.1A and B). Comparing this to data available from other tissues we can see that the endometrial 5mC levels are similar to other tissues (Figure 4.1C). The 5hmC levels are far more variable across different tissues, with endometrium tissue falling within the medium range (Figure 4.1C). When considering the other human data, the endometrium has a higher ratio of 5hmC: 5mC compared to the available human blood or saliva data (Figure 4.1D). This could point to 5hmC playing an important role specifically in the endometrium. 5hmC levels in the human brain and in mouse Purkinje neurons are very high (~0.6% of CpGs), and 5hmC has been shown to be dynamically modified at loci during neurodevelopment and ageing associating it with cellular plasticity (Kriaucionis and Heintz 2009; W. Li and Liu 2011; Szulwach et al. 2011). In stem cells, which by nature are a highly dynamic and plastic cell type, 5hmC is also high (~0.08% 5hmC/C) and seemingly crucial for this plasticity; TET ablation renders ESCs more prone to differentiation, and 5hmC levels decline quickly with ESC differentiation (Ficz et al. 2011; Koh et al. 2011; Pastor et al. 2011; Senner et al. 2012). Epithelial cells were assessed separately for absolute 5hmC amounts, this revealed a large increase of this modification in the receptive state which was obscured in the whole tissue samples (Figure 4.3B). 5hmC levels in endometrial epithelial cells were absolutely higher than in whole tissue samples, with the non-receptive and receptive mean value in epithelial cells being 0.11% and 0.19%, respectively, compared to 0.08% and 0.09% 5hmC/dG in the whole tissue samples. Interestingly, available data from other epithelial tissues in the mouse also showed much higher 5hmC: 5mC ratios compared to non-epithelial tissues. This could indicate that 5hmC may play an important role in epithelial cells in general. Further analysis of human tissue is required to be able to substantiate this.

A remarkable discovery described in this chapter, is the extent of 5hmC fluctuation in endometrial epithelial cells between the non-receptive and the receptive state, specifically the accumulation of this modification with receptivity. This was an unexpected finding that was clearly evidenced by IF and confirmed by the LCM-MS data of isolated epithelial cells. This denotes that 5hmC is undergoing a dramatic change in epithelial cells at the tipping point between the non-receptive and receptive state, which prompted me to look into this phenomenon in greater detail.

The development of the highly sensitive HMCP method by Cambridge Epigenetix allowed me to interrogate the genome-wide distribution of 5hmC in epithelial cells isolated by LCM for the first time. Considering the quantity (~13ng) and quality of DNA recovered after LCM this can be regarded as quite an achievement. Using these profiles for unsupervised hierarchical clustering grouped the receptive samples together, and away from the non-receptive samples. Considering how close these samples are with regards to cycle timing, this is a rather notable observation. Houshdaran et al. and Saare et al. showed that 5mC was able to discriminate between the phases of the menstrual cycle in general but was unable to discriminate within the secretory phase, with the early and mid-secretory phases clustering together (Houshdaran et al. 2014; Saare et al. 2016). This shows that the different cell types in a complex tissue sample could be obscuring changes that are happening in individual cell types during the menstrual cycle. Also, this data supports the notion that 5hmC itself is an important epigenetic mark in its own right and not just a by-product of 5mC turnover, and as such should be investigated as its own entity.

5hmC was enriched at the promoter regions of key genes that are dramatically upregulated in the receptive state, such as *PAEP* and *NNMT* (Figure 4.5A and B). These genes have been shown to be upregulated with receptivity in other studies as well as my own (Figure 3.3A and B, Díaz-Gimeno *et al.*, 2011; Altmäe *et al.*, 2017; Burmenskaya *et al.*, 2017). In particular *PAEP*, also known as glycodelin, gains 5hmC in the receptive state compared to the non-receptive state at the promoter as well as across the whole gene body. PAEP is a glycoprotein that is secreted by epithelial cells, it has several isoforms and it plays a role not only in the initiation of pregnancy but also during pregnancy. It provides histiotrophic support before the establishment of the placenta via its uptake by syncytiotrophoblast (Burton et al. 2002).

Further to this, 5hmC is highly enriched at bivalently marked genes that resolve to H3K4me3-only in the receptive state. Remarkably, it is enriched at these bivalent loci already in the non-receptive state prior to their activation, predicting which subset of bivalent loci will resolve to H3K4me3-only. Further investigations will need to investigate whether indeed 5hmC may be required for these loci to resolve to an active chromatin configuration, which would imply a highly important role of 5hmC during the endometrial cycle. By contrast, 5hmC is lowly enriched at bivalent genes that remain bivalent during this timeframe. The genes that resolve to H3K4me3-only were shown in Chapter III (Figure3.9E) to be implicated in extracellular exosomes, mitochondria and the metabolic pathway adding further weight to the conclusion that the epigenome is involved in the setup of the secretory function of epithelial cells during the receptive state.

In general, the increase in 5hmC levels in receptive state endometrial epithelium may be correlated with the high transcriptional demand and ultimately secretory output of these cells in this phase in preparation for their role in histiotrophic nutrition. Whether or not 5hmC is indeed required to facilitate this transcriptional upregulation remains to be studied, for example in knockouts or knockdowns of the TET proteins specifically in endometrial epithelium. However, the correlation of 5hmC enrichment with increased gene expression levels, specifically in the receptive phase, is in line with such a general role of 5hmC, which has been corroborated in other tissues (Ficz et al. 2011; Szulwach et al. 2011).

Due to the limited availability of sections from paraffin-embedded samples which had been timed for their receptivity as well as the extensive time required to perform laser capture microdissections, the number of samples that I was able to analyse by HMCP was limited. To further validate these findings, the addition of further samples would be advantageous.

To understand how the dramatic change in 5hmC levels may come about, I firstly looked at genes responsible for the addition and removal of DNA methylation and hydroxymethylation. The RNA-seq data shows that there is no significant difference in the expression of the various DNMTs or TETs between the non-receptive and receptive state. Data of epithelial cells at LH+5 (earlier than my non-receptive samples) and LH+7 showed no significant changes in genes involved in this process either (Salker et al. 2017). This is also in agreement with Vincent et al. who showed that there were no

significant differences in *DNMT1*, *DNMT3A*, and *DNMT3B* mRNA levels between the early, mid and late secretory phase (Vincent et al. 2011). Yet another study showed *DNMT1* and *DNMT3A* had their lowest expression levels in the mid-secretory phase, whereas *DNMT3B* remained unchanged (Yamagata et al. 2009). It is worth noting that these studies used histology for dating the endometrium, which has been suggested by several reports not to be reliable (Díaz-Gimeno et al. 2011; Gibson et al. 1991; Talbi et al. 2006). In addition to these studies on primary tissue, endometrial explants and cultured stromal cells have also been assessed and showed that *DNMT3A* and *DNMT3B* were downregulated or unchanged after the addition of estrogen and progesterone, whereas *DNMT1* showed inconsistent changes (Dyson et al. 2015; Vincent et al. 2011; Yamagata et al. 2009). There is considerable variability, between different study designs and samples. Importantly, even in studies that reported minor mRNA expression level differences, the protein levels of DNMTs did not change during the secretory phase of the menstrual cycle (Yamagata et al. 2009). Although there is disagreement between papers, which could arise due to variance in clinical studies, systems and experimental design, the majority of data suggest that there is no change in DNMT or TET enzyme expression specifically at the time period where I observed a dramatic shift in 5hmC. This suggests that other factor/s are responsible for the substantial change in 5hmC levels in epithelial cells.

Nicotinamide N-Methyltransferase (NNMT) presented itself as a good candidate to be involved as a dynamic methylation modulator of 5mC and 5hmC in epithelial cells. NNMT has been shown to be highly upregulated in the receptive state in multiple analyses, including specifically in epithelial cells (Figure 4.8A and B and C, Díaz-Gimeno *et al.*, 2011; Altmäe *et al.*, 2017). It also gains 5hmC at its promoter in the receptive state. The traditional role of NNMT is to catalyse the reaction of Nicotinamide (NAM) to 1-methylnicotinamide (1MNA), acting as a vitamin B clearance enzyme (Ellinger and Coulson 1944). The structure of NNMT reveals a methyltransferase site and multiple SAM binding sites (Figure 4.9A). SAM is a universal methyl donor and adds a methyl group to NAM in the process of its clearance. The role of NNMT beyond NAM clearance has more recently been investigated, specifically in cancer. NNMT has been shown to be upregulated in several cancer tissues, and cancers are known to often exhibit a dysregulated epigenome (Jung et al. 2017; Ramsden et al. 2017). Knockdown of *NNMT* in cancer cells results in reduced proliferation and/or metastasis suggesting that NNMT contributes to the aggressiveness of cancer. NNMT is also regulated by STAT3 in

hepatocellular carcinoma cells. STAT3 is essential in mice uterine epithelial cells for embryo attachment and subsequent implantation to occur (Pawar et al. 2013; Tomida et al. 2008). The mechanism by which NNMT affects these traits of cancer has not been fully elucidated but multiple studies have suggested that NNMT could act as a methyl sink, reducing the availability of SAM to other methyltransferases (Jung et al. 2017; Ulanovskaya, Zuhl, and Cravatt 2013). Ulanovskaya et al. showed that NNMT over-expression in renal carcinoma cells causes a decrease of H3K9me2 and H3K27me3, and conversely an increase in these modifications when NNMT is knocked down in ovarian carcinoma cells. This also resulted in expression changes of several genes including *SNAI2* and *CNTN1* (Ulanovskaya, Zuhl, and Cravatt 2013). However, they also showed that 5mC levels remained globally unchanged, while 5hmC levels were not studied (Ulanovskaya, Zuhl, and Cravatt 2013). By contrast, in proneural and mesenchymal cultured cells, depletion of NNMT resulted in a ~30-50% increase in 5mC levels as well as an upregulation of DNMTs, and the reverse occurred upon NNMT over-expression (Jung et al. 2017). In ESCs, NNMT is known to be highly expressed in the naive state and down-regulated in the primed state, which corresponds with an increase in H3K27me3 and H3K9me3 (Sperber et al. 2015). Overall, these data suggest that NNMT influences certain methylation pathways via regulating access to the methyl donor SAM or via the downregulation of particular methyltransferases. The exact mechanism, and what other proteins may be involved to influence the dynamics of methylation deposition, remains unknown. The relationship between NNMT and DNA hydroxymethylation has not been explored.

I hypothesised that the increase in *NNMT* levels in receptive epithelial cells may act as a methylation modulator by reducing the availability of SAM and consequently methyl groups so that DNA methylation would not be restored while conversion rates to 5hmC remain unchanged, resulting in a net increase in 5hmC over 5mC levels. Even if this would not necessarily explain an absolute increase in 5hmC levels in endometrial epithelial cells, it was an hypothesis which at least could explain a shift in relative 5hmC: 5mC ratios. Upon over-expression of *NNMT* in the epithelial cell line HEC1A, I determined that 5mC and 5hmC levels were not affected. However, the levels of histone methylation, notably the H3K4me3 and H3K27me3 marks, significantly increased.

This result is contrary to expectations from the previously published data, suggesting that a different mechanism may be at play in endometrial cells. How precisely the higher H3K4me3 and H3K27me3 levels may come to pass remains speculation. However, one

possibility is that NNMT over-expression indirectly alters histone methylation levels by disrupting the NAD salvage cycle. NNMT converts nicotinamide (NAM) to 1-methylnicotinamide, which permanently removes NAM from the NAD salvage cycle. This would result in a decrease in the substrate for NAD+ potentially lowering the levels of NAD+. This in turn could have an effect on histone methyltransferases. SIRT1 is a NAD+ dependant deacetylase and it has been shown to deacetylate MLL, a K4 methyltransferase. Therefore, the reduction of NAD+ availability would lead to an increase in acetylated MLL which in turn would lead to higher levels of H3K4me3. Indeed Mouse Embryonic Fibroblasts (MEFs) that are SIRT1 null or treated with a SIRT1 specific inhibitor (EX527) have higher levels of H3K4me3 (Aguilar-Arnal et al. 2016). Further to this, NAD+ is also required for the poly(ADP-ribosyl)ation (PARylation) of EZH2, a K27 methyltransferase. EZH2 PARylation results in the dissociation of the PRC2 complex, which in turn leads to a reduction in H3K27me3 levels (Yamaguchi et al. 2018). Therefore, the increased levels of NNMT could reduce the availability of NAD+, which could lead to an increase in acetylated MLL and PCR2 stability and increased H3K4me3 and H3K27me3, respectively. Alternatively, NNMT may interact with different proteins which are also upregulated in the receptive state and this in turn may determine the availability of methyl groups to different methyltransferases.

Although HEC1A cells are a useful tool to investigate epithelial endometrium, they do have their drawbacks. They have been used to represent the non-receptive state in previous studies, however they do not always respond to hormones and I show that they do not express genes associated with the non-receptive or receptive state (Figure 4.11A, Lessey, 2003; Tamm *et al.*, 2009; Bhagwat *et al.*, 2013). Therefore, HEC1A cells are not necessarily representative of a particular timepoint in the menstrual cycle and may not respond to NNMT as non-receptive epithelial cells may. HEC1A cells are adenocarcinoma cells from a 71 year-old women, therefore they may already have a disrupted epigenetic landscape due to the cancer origin and the individual's age. For instance, cancer tissues have globally lower levels of 5hmC compared to their healthy counterparts (W. Li and Liu 2011). Cell culture in itself brings its own challenges that may induce further epigenetic rearrangements including alteration in DNA methylation and hydroxymethylation levels when compared to the corresponding *in vivo* tissue, as has been shown in trophoblast, bone marrow stem cells and breast cancer cell lines (Bentivegna et al. 2016; Hamadneh et al. 2018; Senner et al. 2012). Also, the cross talk

and interactions with other cell types cannot be recapitulated in the cell line culture model.

To further investigate the role of NNMT in epithelial cells a different system could be utilised. Turco et al. has recently developed epithelial organoids that upon decidualisation show an increase in *NNMT* mRNA levels in all three biological replicates of organoids (Turco et al. 2017). This may be a system that more closely resembles *in vivo* functionality of epithelial cells.

# Chapter V

# Epigenetic changes associated with endometriosis

## 5.1 Introduction

Endometriosis is a condition where endometrial tissue is found ectopically, predominantly in the peritoneum, ovaries and rectovaginal septum, where it forms lesions (Vercellini et al. 2014). Endometriosis affects approximately 10% of women of reproductive age (Nnoaham et al. 2011). The symptoms experienced by women vary greatly; they include chronic pelvic pain, infertility, dyspareunia, and dysmenorrhea. However, endometriosis can also be asymptomatic (Guo 2009a; Holt and Weiss 2000). The American Society for Reproductive Medicine 1996 has classified endometriosis into four clinical stages. It assigned a scoring system based on location, size and depth which results in a classification of Stages I to IV; minimal, mild, moderate and severe, respectively (Canis et al. 1997). Minimal and mild endometriosis usually corresponds to superficial adhesions, moderate and severe to endometriomas and more severe adhesions (Canis et al. 1997; Vercellini et al. 2014). The stages do not necessarily correspond to symptoms or fertility (Canis et al. 1997; Vercellini et al. 2014). Studies that have assessed the presence of endometriosis in women undergoing sterilization and that were asymptomatic show a wide range of prevalence rates, from <2% to ~20%, with patients mostly presenting with minimal endometriosis (Moen and Stokstad 2002; Tissot et al. 2017). Currently there is no cure for endometriosis and only few treatment options available.

Endometriosis is very difficult to diagnose due to the varying symptoms that are experienced by women (Guo 2009b; Holt and Weiss 2000). It takes on average 7 visits to a primary doctor before getting referred, usually gaining a definitive diagnosis on average after 6.7 years (Nnoaham et al. 2011). The gold standard for an unambiguous diagnosis of endometriosis is laparoscopic surgery followed by histology (Holt and Weiss 2000). There has been a push in the scientific community to establish a diagnostic test that would speed up the process for women to get a conclusive verdict.

Many avenues of research have been conducted to try to establish a causal link to endometriosis. These include research into environmental toxins (Rier and Foster 2003), immune or autoimmune diseases (Ahn et al. 2015; Matarese et al. 2003), reproductive hormones (Nasu et al. 2011) and genetic factors, although publications are often inconsistent and conflicting (Falconer, D'Hooghe, and Fried 2007). With the lack of any confirmed causal link, more recently there has been an influx of studies looking into

potential epigenetic contributions to the condition (Borghese et al. 2008; Dyson et al. 2014; Naqvi et al. 2014; Wu et al. 2006).

To pursue an epigenetic causal link and a potential diagnostic tool for endometriosis, we analysed 20 endometrial biopsies from women attending the infertility clinic at the University of Warwick, which were provided by Prof. Jan Brosens. These biopsies were used for analysis in Chapter III and Chapter IV, and consist of 10 control biopsies and 10 biopsies from women with confirmed endometriosis (stages II to IV). All patients were free from any hormone treatments at least three months prior to the biopsy been taken. Samples were LH surge-timed to fall within the implantation window. As discussed in Chapter III, analysis of the transcriptomic data, Endometrial Receptivity Transcriptome ("ERT"), along with the use of the endometrial receptivity array (ERA) allowed me to establish the precise receptivity state of the biopsies (Díaz-Gimeno et al. 2011). A full description of all the samples along with the established receptivity state can be found in Chapter II– Materials and Methods, Table 2.1.

Here, I have carried out a most comprehensive epigenetic analysis of eutopic samples from patients with endometriosis. The analyses performed included expression profiles, as well as genome-wide profiles of several histone modifications as well as DNA methylation and hydroxymethylation, see Table 2.1 and Figure 2.1. All analyses were carried out on the same biopsies allowing for direct comparison between datasets, which is invaluable to the interpretation of results.

There have been several reports that have analysed transcriptional changes associated with endometriosis. These have all been carried out using microarrays and/or qPCRs to asses expression levels. The high frequency of endometriosis in infertile patients and with endometriosis often remaining undetected could explain why so few loci have been identified. The majority of transcriptional changes that have been described are as a function of the menstrual phase that occur during the transition from the proliferative phase to the secretory phase (Burney et al. 2007; Sherwin et al. 2008). There has been one report carried out on 229 women, which suggested that there is a strong genetic effect on gene expression in the endometrium throughout the menstrual cycle. However, there are no transcriptional differences in eutopic tissue associated with patients with endometriosis when the menstrual phase is considered (Fung et al. 2018).

As discussed in Chapter III, histone modifications play a key role with regards to transcriptional regulation within the endometrium. Global H3 acetylation, H3K9ac and

H4K16ac tend to be lower in the eutopic tissue of patients with endometriosis compared to controls and are significantly reduced in endometriotic lesions (Monteiro et al. 2014). Global histone methylation levels tend to increase in patients with endometriosis for H3K9me (mono, di and tri combined) and are significantly increased for H3K4me (mono, di and tri combined) and H3K27me (mono, di and tri combined) (Monteiro et al. 2014). This was confirmed by Liu et al. using more specific and methylation-state specific antibodies for H3K9me3, H3K27me3 and EZH2 (the protein that deposits H3K27me3), which were significantly increased in endometriotic lesions (Xishi Liu, Zhang, and Guo 2018). In contrast to these results, Xiaomeng et al. showed that the abundance of H3 acetylation remained unchanged in eutopic and ectopic tissue from endometriosis patients compared with controls, and H3K9me (mono, di and tri combined) was significantly lower in ectopic tissue but was tendentially higher in eutopic tissue from endometriosis patients (Xiaomeng et al. 2013). Overall, these data, albeit conflicting, suggest that histone methylation levels may be dysregulated in patients with endometriosis, and a more precise resolution of mono, di and tri methylation along with receptivity stage would lead to a greater understanding.

As discussed previously, DNA methylation and hydroxymethylation act as a functional layer regulating gene expression, with the profiles changing throughout the menstrual cycle. It has been shown that DNA methylation patterns are altered in the presence of endometriosis by several studies. There have been a number of reports where specific loci were analysed for DNA methylation changes depending on endometriosis, and these have shown that some promoters are hypomethylated whereas others are hypermethylated in endometriosis. Wu et al. showed that the progesterone receptor B (*PGR-B*) promoter is hypermethylated in endometriosis and that this could cause and explain the progesterone resistant phenotype that is commonly associated with endometriosis (Burney et al. 2007; Wu et al. 2006). Similarly, the *HOXA10* promoter has been shown to be hypermethylated in the eutopic tissue of mice and baboons who have induced endometriosis, compared to controls (J. J. Kim et al. 2007; B. Lee, Du, and Taylor 2009). HOXA10 is a progesterone-regulated gene that is a common marker of the decidualisation state. In contrast, however, human endometrial samples showed hypomethylation at the promoter of *HOXA10* compared to controls (Kulp, Mamillapalli, and Taylor 2016; Taylor et al. 1998). Xue et al. identified that *ESR2* is hypomethylated in endometriotic stromal cells in one paper and that *SF-1* is hypomethylated in another paper (Xue, Lin, Cheng, et al. 2007; Xue, Lin, Yin, et al. 2007).

Several articles have also compared global profiles of DNA methylation levels between eutopic control and ectopic endometriotic endometrial tissue. They found a number of regions that were differentially methylated (Borghese et al. 2010; Dyson et al. 2014; Rahmioglu et al. 2017). They did not address, however, if the eutopic tissue from patients with endometriosis may already show alterations compared to control patients. This shortfall was addressed in a study by Naqvi et al. who compared eutopic tissue between patients with endometriosis and controls. However, this study lacked information on what menstrual phase the samples were taken in, which as we know is an important consideration as DNA methylation changes throughout the cycle (Chapter 4, Naqvi et al. 2014; Saare et al. 2016). Yamagata et al. profiled DNA methylation in cultured endometrial stromal cells from eutopic tissue of patients with or without endometriosis and found only few differences (Yamagata et al. 2014). It is worth noting that the culture conditions themselves could have disrupted the methylation profiles. An additional drawback is that the majority of previous studies have used methylation arrays which do not cover the whole genome and require a bisulphite conversion step, which is not always 100% efficient. Although the arrays are widely used, deep sequencing is required to try to compensate for this shortfall. Also, the bisulphite conversion means that there is no discrimination between 5hmC and 5mC and is instead interpreted as one mark, methyl-cytosine, perhaps a more significant shortcoming of the array method. As I have shown earlier in Chapter IV, 5hmC plays an important role in the transition from pre-receptivity to the receptive state. Therefore, it is chiefly important to determine if this epigenetic modification is affected by endometriosis. To fully understand if DNA methylation and hydroxymethylation are altered in eutopic endometrial tissue of endometriosis patients, I here undertook genome-wide 5mC and 5hmC immunoprecipitation followed by sequencing (meDIP-seq and hmeDIP-seq), in combination with a re-analysis of the transcriptional and histone modification profiles described earlier that were now analysed with the specific emphasis on disease state.

## 5.2 Transcriptional changes associated with endometriosis

The establishment of the fine discrimination between the non-receptive and receptive state in Chapter III highlighted that it was important to take receptivity into consideration when comparing transcriptional profiles of samples from endometriosis

and control patients, as the precise receptivity state could overshadow transcriptional changes associated with endometriosis. DESeq2 and EdgeR analyses were performed between controls and endometriosis samples with multiple testing corrections applied, separating samples into separate groups that had been determined to be in the non-receptive and receptive states. This resulted in four groups that were compared in a pair-wise manner: non-receptive controls (NRC) and non-receptive endometriosis (NRE), and receptive controls (RC) and receptive endometriosis (RE). The controls and endometriosis samples were highly correlated in the receptive state (R=0.997) and there were no significant gene expression differences (Figure 5.1A). Within the non-receptive state, the controls and endometriosis samples were still highly correlated (R= 0.982), however the spread of gene expression levels was greater and there were a number of genes that were significantly differentially expressed between control and endometriosis samples (Figure 5.1B).

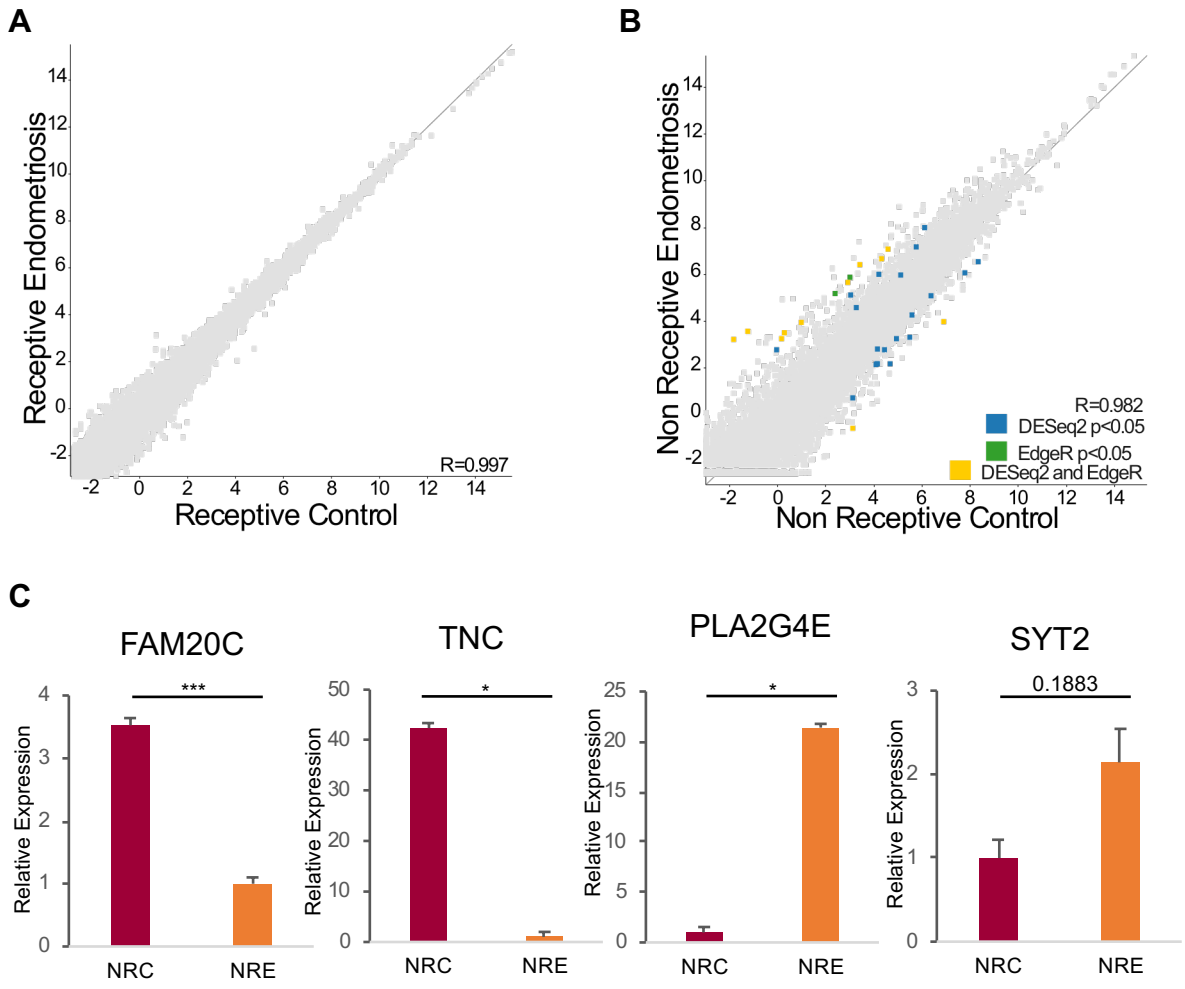



**Figure 5.1 Differential expression associated with endometriosis. A**, Scatterplot comparing RNA-seq data from receptive controls with receptive endometriosis samples. DESeq2 and EdgeR analyses showed no significant differences. **B**, Scatterplot comparing non-receptive controls with non-receptive endometriosis samples. DESeq2 and EdgeR analyses with

multiple testing correction applied. DESeq2 $p<0.05$= blue, EdgeR $p<0.05$ = green, DESeq2 and EdgeR = yellow. **C**, RT-qPCR of genes that were differentially expressed between control and endometriosis samples in the non-receptive state. Data was log2-transformed prior to statistical analysis. Unpaired two-tailed t-test was performed. ***$p<0.001$, *$p<0.05$. Mean +/- SEM plotted. Receptive Control (n=6), Receptive Endometriosis (n=6), NRC= non-receptive control (n=4), NRE = non-receptive endometriosis (n=3).

32 and 13 genes were significantly differentially expressed between control and endometriosis samples in the non-receptive state according to DESeq2 and EdgeR, respectively. Eleven genes were jointly present in both statistical tests. Four genes were randomly selected for independent validation by RT-qPCR, two of which were up-regulated and two down-regulated according to the RNA-seq data. All four genes were confirmed as being differentially expressed between controls and endometriosis samples by RT-qPCR, three of which were significant, verifying the results of the RNA-seq data (Figure 5.1C).

The eleven high-confidence dysregulated genes are all membrane-bound or have cell membrane-associated functions; these include *LRP4*, *BTNL9*, *CNTN4*, *PLA2G4E*, *NRK*, *KCNB1*, *SLC16A6*, *SYT2*, *TNC*, *THEM4* and *TMED6*. SYT2, LRP4, KCNB1 are ion channels; SLC16A6, PLA2G4E, TMED6 are membrane transporters, BTNL9 is a membrane glycoprotein that binds to immune cells and CNTN4 and TNC have roles in cell adhesion. This implies a particular dysregulation of genes involved in membrane biology with endometriosis.

## 5.3 Characterisation of histone modifications with respect to endometriosis

As changes in certain histone modifications have been associated with endometriosis at a total level, I carried out analysis of key histone modifications in eutopic endometrium from control patients and patients with endometriosis. Firstly, I conducted immunofluorescence staining experiments for H3K4me3 and H3K27me3 to compare the epithelial-to-stromal cell staining intensity ratio between control and endometriosis patient samples (Figure 5.2A,B and C).

I found that the spatial distribution patterns and intensity ratios of these two histone modifications were not significantly different between control and endometriosis samples within each receptivity state (Figure 5.2B and C).

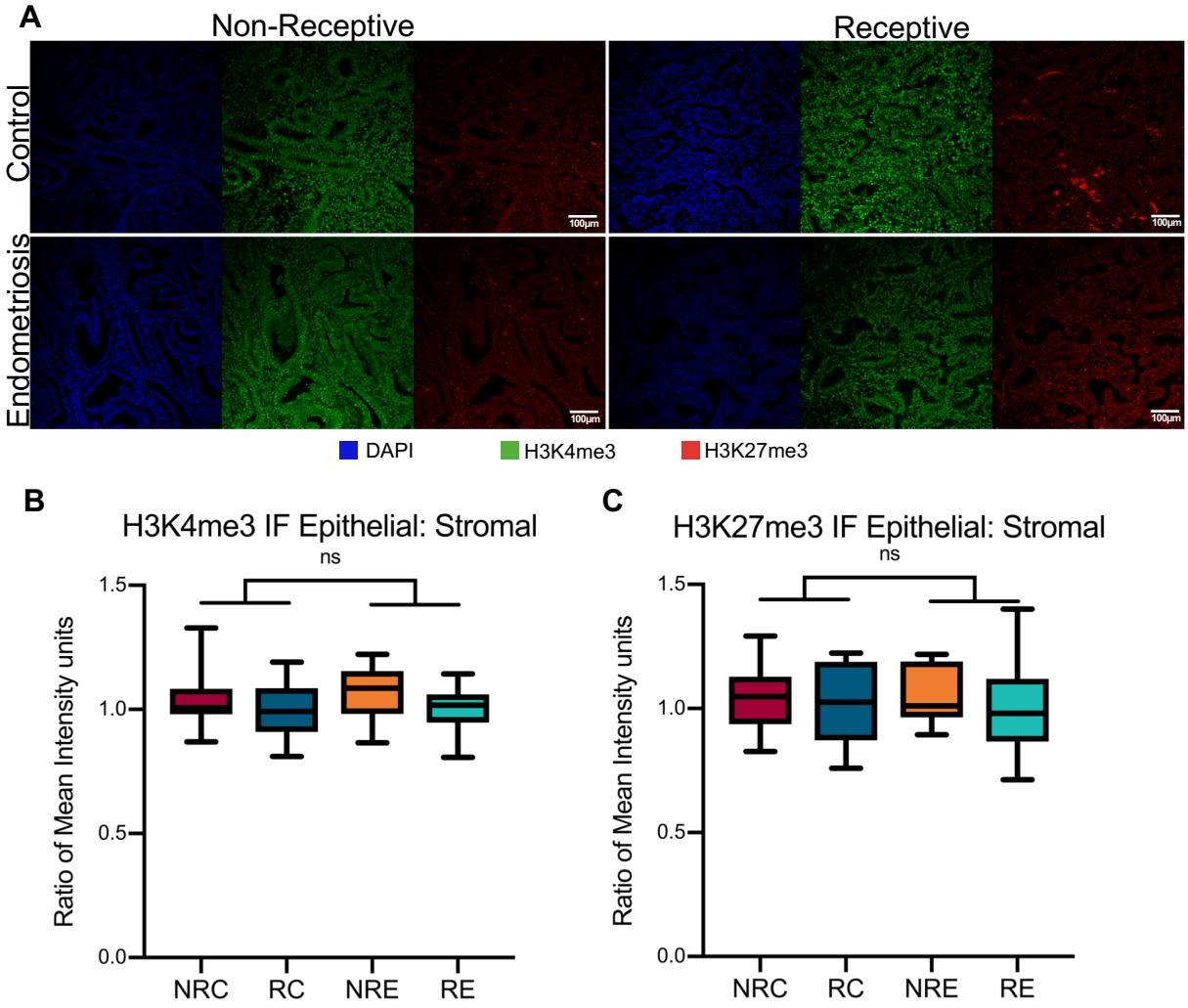



**Figure 5.2 H3K4me3 and H3K27me3 immunofluorescence in endometrium.**
**A**, Immunofluorescence images of control and endometriosis samples in the non-receptive and receptive states. DAPI (blue) was used as nuclear counter-stain, H3K4me3 (green) and H3K27me3 (red). Scale bar = 100µm **B** and **C**, Analysis of the epithelial: stromal cell ratio of mean intensity units within each immunofluorescence image for H3K4me3 and H3K27me3. Two-way ANOVA followed by Tukey's post-hoc test showed no significant differences. NRC= non-receptive control, NRE= non-receptive endometriosis, RC= receptive control, RE= receptive endometriosis. All conditions had a minimum of 3 biological replicates with 3 images each. Box whisker plot depicting the box as 25$^{th}$ percentile, medium and 75$^{th}$ percentile and the whiskers as the minimum and maximum values.

Even if the global ratio of histone modification levels was not affected, it does not rule out that their distribution at specific loci may be altered by endometriosis, and this is an avenue I pursued in the next step. For this reason, I carried out chromatin immunoprecipitation for the histone marks H3K4me3, H3K27me3 and H3K9me3 followed by high-throughput sequencing (ChIP-seq) on pieces of the same biopsies used before. Extensive quality control was undertaken pre- and post- sequencing as described in Chapter III, Figure 3.7. Enrichment was determined bioinformatically using the

MACS2 peak calling method as it can detect narrow and broad peaks with a false discovery rate (FDR) cut-off of 0.05, with the corresponding input used as the background for each sample (https://github.com/taoliu/MACS/). MACS2 peaks were called for individual samples. Consensus peaks present within each condition (non-receptive control (NRC), non-receptive endometriosis (NRE), receptive controls (RC), receptive endometriosis (RE)) were used for analysis. After peak calling DESeq2 and EdgeR analysis was conducted with multiple testing corrections, which is an approved method to determine differential enrichment values in ChIP-seq datasets (Anders and Huber 2011; Love, Huber, and Anders 2014; Robinson, McCarthy, and Smyth 2009; Shpargel et al. 2014). This was conducted between control and endometriosis samples in each receptivity state (i.e. NRC vs NRE, RC vs RE). Histone modification profiles were highly correlated between controls and endometriosis samples, ranging between R= 0.853 and 0.988 in all receptivity states (Figure 5.3A to F). Within the non-receptive state, the variability in enrichment levels across all peaks, visualised by the “spread” of data points, was greater than in the receptive state (Figure 5.3A to F). There was no significant difference between control and endometriosis samples for H3K4me3 or H3K9me3. However, for H3K27me3, 7 peaks were significantly differentially enriched between control and endometriosis tissues in the receptive state (Figure 5.3F).

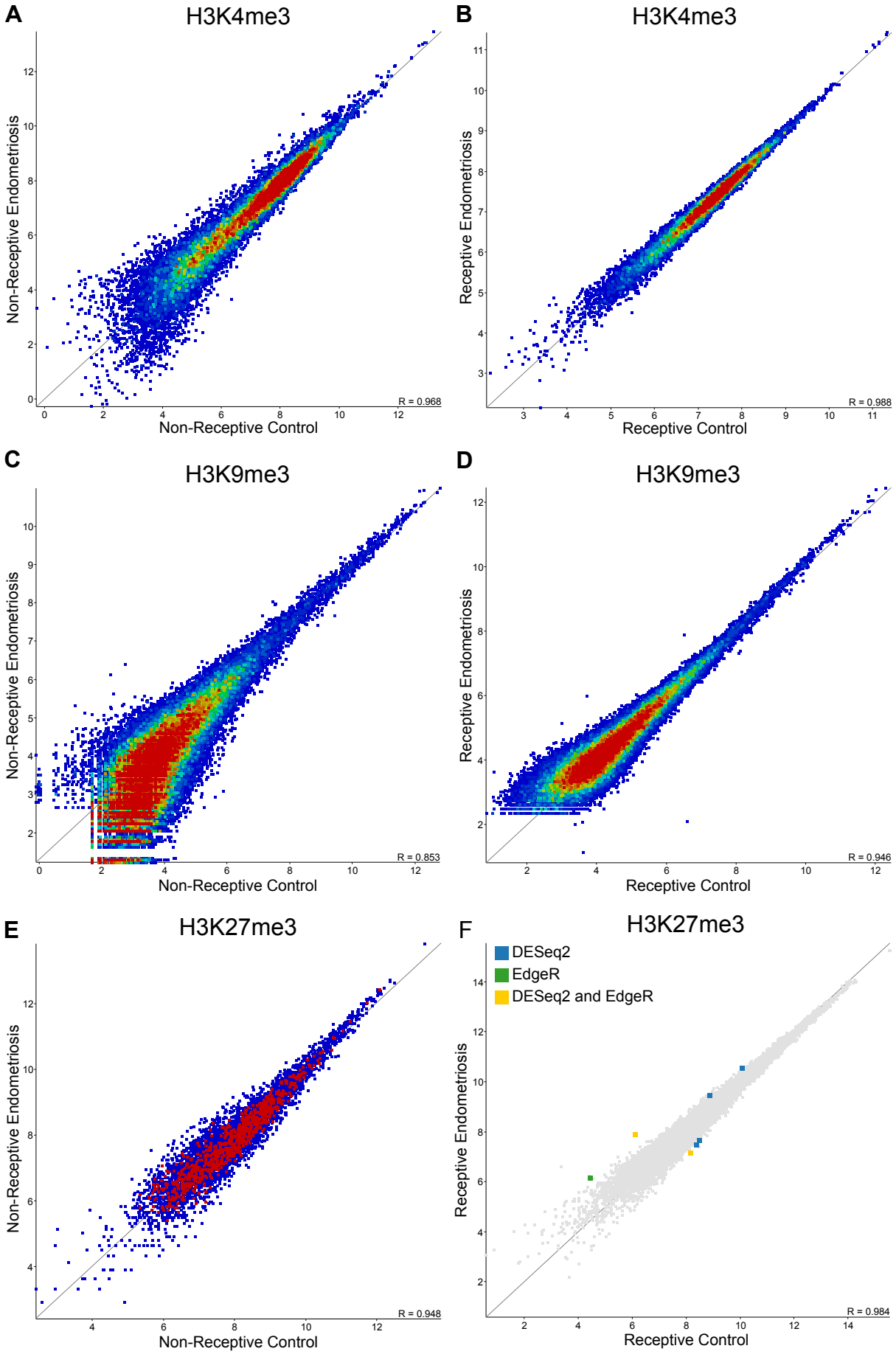



**Figure 5.3 Histone methylation MACS2 peak profiles. A, B, C, D, E** and **F**, Scatterplots of H3K4me3 (A, B), H3K9me3 (C, D) and H3K27me3 (E, F) with controls compared to

endometriosis samples in both non-receptive and receptive state. Corrected to largest dataset, log2-transformed and enrichment-normalised to the 20th and 90th percentile where applicable. DESeq2 and EdgeR statistics applied with no significant differences for H3K4me3 (A, B), H3K9me3 (C, D) or H3K27me3 in the non-receptive state (E). Significant statistical differences highlighted for H3K27me3 in the receptive state (F), DESeq2 = blue, EdgeR = green, DESeq2 and EdgeR = yellow.

On a log2 scale, the significantly differentially enriched H3K27me3 peaks that were higher in the receptive endometriosis samples had values of 1.6 and 0.6, and those enriched in the receptive controls had values between -1.0 and -0.9, meaning that there was a relative enrichment of between 1.2- to 3.2-fold at these peaks. The closest gene within 2kb was identified for each peak and is listed in Table 5.1. The associated genes of all these regions were not differentially expressed in my samples, which agrees with the analysis of expression being highly correlated between control and endometriosis samples in the receptive state (Figure 5.1A). These genes have not previously been associated with endometriosis.

**Table 5.1 Genomic regions and nearby genes associated with differentially enriched H3K27me3 peaks between control and endometriosis samples in the receptive state.** Corrected DESeq2 and EdgeR p values, log2-fold change: positive values indicate enrichment in endometriosis, negative values indicate enrichment in controls.

| Gene | Full Name | Chr | Start | End | DESeq2 P-value | EdgeR P-value | Log2 Fold Change |
|---|---|---|---|---|---|---|---|
| **C3orf70** | Chromosome 3 open reading frame 70 | 3 | 185150193 | 185151639 | 0.024 | 0.039 | -1.021 |
| **TMEM176A/B** | Transmembrane protein 176A / B | 7 | 150799083 | 150802731 | 0.049 | | 0.623 |
| **null** | n/a | 11 | 133831336 | 133832961 | 0.024 | 0.011 | 1.628 |
| **RP11-271M24.2** | No description | 13 | 25295131 | 25298442 | 0.049 | | -0.851 |
| **BAHCC1** | Bromo Adjacent Homology domain and coiled-coil containing 1 | 17 | 81407110 | 81421519 | 0.049 | | 0.574 |
| **RP11-308D16.2** | No description | X | 137022230 | 137025182 | 0.049 | | -0.945 |
| **null** | n/a | 9 | 134291440 | 134291736 | | 0.011 | 1.531 |

This is a noteworthy finding, in particular the TMEM176A/TMEM176B locus (the TSSs of these paralogs are overlapping and promoter assignment is ambiguous), which has an enrichment of H3K27me3 over the TSSs, although expression levels are not significantly different (Table 5.1, Figure 5.4A). TMEM176B has been shown to suppress inflammation. Therefore, differing signals in an ectopic environment could potentially lead to an attenuated expression of this gene in endometriosis, due to pre-existing higher H3K27me3 levels at the TSS, resulting in enhanced inflammation. Increased inflammation is associated with endometriosis and in particular with the diseases progression (Bruner-Tran et al. 2013; Segovia et al. 2019).

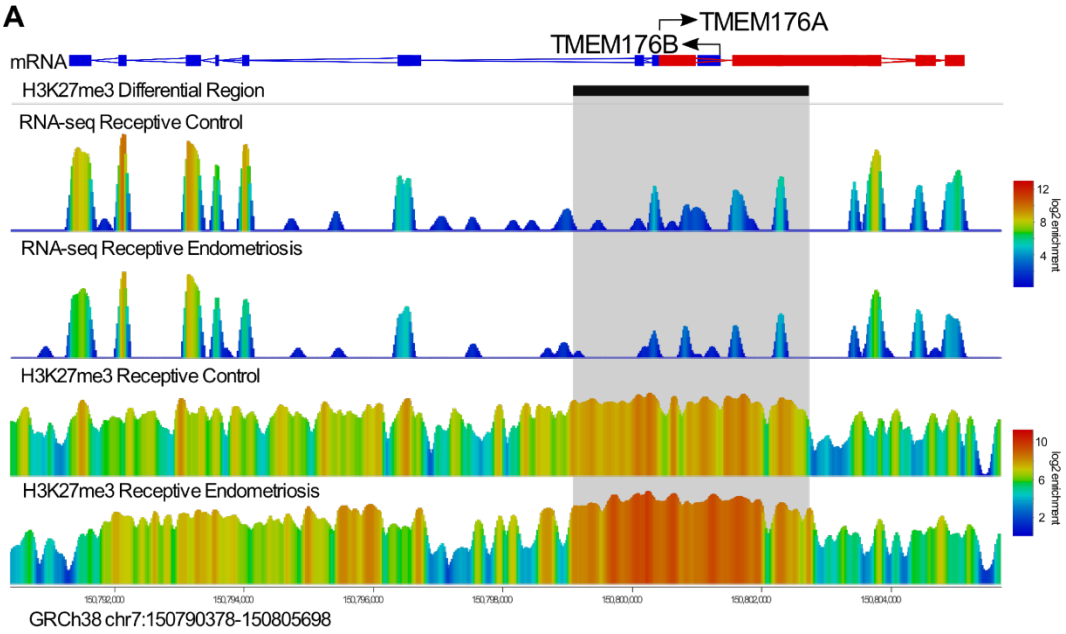



**Figure 5.4 Differential H3K27me3 enrichment at *TMEM176A/B*. A**, Wiggle plot across the *TMEM176A/B* locus. RNA-seq and H3K27me3 tracks for control (n=4) and endometriosis (n=2) samples in the receptive state. MACS2 peak of differential H3K27me3 enrichment marked by black box with greyed out area underneath.

As H3K9me3 is associated with chromatin compaction and involved in maintaining the stable repression of repeat elements, the Babraham Bioinformatics department carried out an analysis of enrichment at repetitive regions for H3K9me3 on my data (Figure 5.5A). Two-way ANOVA was performed and this showed no significant difference between controls and endometriosis. This suggests that H3K9me3 is not disrupted in the eutopic endometrium of patients with endometriosis in my samples.

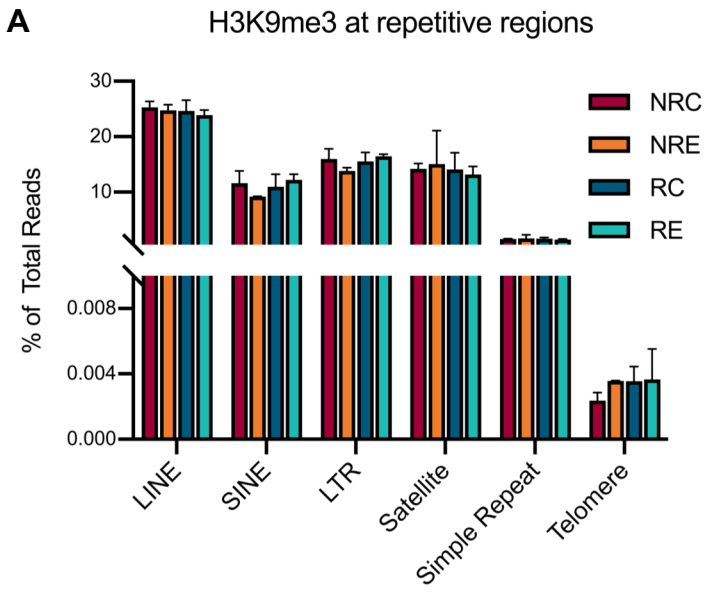



**Figure 5.5 H3K9me3 repetitive analysis. A**, H3K9me3 analysis of repetitive regions. Two-way ANOVA followed by Tukey's post-hoc test conducted revealed no significant differences. Mean +/- SD. NRC= non-receptive control (n=2), NRE= non-receptive endometriosis (n=2), RC= receptive control (n=5), RE= receptive endometriosis (n=3).

Several genes have been suggested to be involved in endometriosis by GWAS studies. Notable examples of these encompass *WNT4*, *VEZT* and *GREB1,* which I therefore examined directly for the presence of differential histone modification within 5 kb of the gene locus, but found no differences (Albertsen and Ward 2016; Nyholt et al. 2012; Saha et al. 2015).

In a converse approach, I examined the eleven genes that I had identified as differentially expressed between control and endometriosis samples specifically in the non-receptive state. However, these were not obviously associated with differential histone modifications. This suggests that any potential differences in histone modifications, such as an increase in H3K4me3 enrichment at genes that are more highly expressed in either endometriosis or control tissue, are below the detection range for differential ChIP-seq enrichment. Alternatively the variability between samples is too high for detection of such changes as significant. Indeed, it could also be that the histone marks remain unchanged and the transcriptional output is determined by other means, such as the abundance and/or availability of critical transcription factors.

Overall, the detailed analysis of histone modification profiles revealed no dramatic differences globally, nor on the locus-specific level, that correlated with the disease state of endometriosis. The only exception to this were a handful of locus-specific differences in H3K27me3 between control and endometriosis samples in the receptive state, in particular at *TMEM176A/B,* which warrants further investigation.

## 5.4 DNA hydroxymethylation dynamics are altered in endometriosis

### 5.4.1 Global distribution of DNA methylation and hydroxymethylation

There have been reports of 5mC differences that have been associated with endometriosis but the absolute levels of 5mC have not been established, nor has 5hmC been investigated in this way. In collaboration with Prof. Petra Hajkova at Imperial College London, whose lab has established an ultrasensitive liquid chromatography-mass spectrometry method for the reliable absolute quantification of 5mC and 5hmC, we established 5mC and 5hmC levels in eutopic tissue from control and endometriosis patients (Amouroux et al. 2016). A two-way ANOVA with Tukey's post hoc test, taking into consideration both the receptivity and disease state, revealed no significant differences in total levels of 5mC or 5hmC related to endometriosis (Figure 5.6A and B). However, it was notable that the increase in 5hmC levels from the non-receptive to the receptive state (described in Chapter IV) appeared markedly blunted in endometriosis samples (Figure 5.6B), even if the relatively high standard deviation and small sample number in each of the four groups did not allow for further statistical testing of this trend.

To further evaluate whether 5mC and 5hmC levels are impacted by the disease state at a cellular level, I carried out IF stainings for these modifications on paraffin sections for all of these tissue samples. The epithelial: stromal cell ratio of 5mC staining intensities tends to be lower in the disease state compared to the controls, just outside the significance threshold, with $p=0.0632$ (Figure 5.6C). The 5hmC ratio is not affected in the disease state, suggesting that there may be a slight decrease in the total level of 5hmC rather than a change in the distribution between the two main cell types in the endometrium (Figure 5.6D).

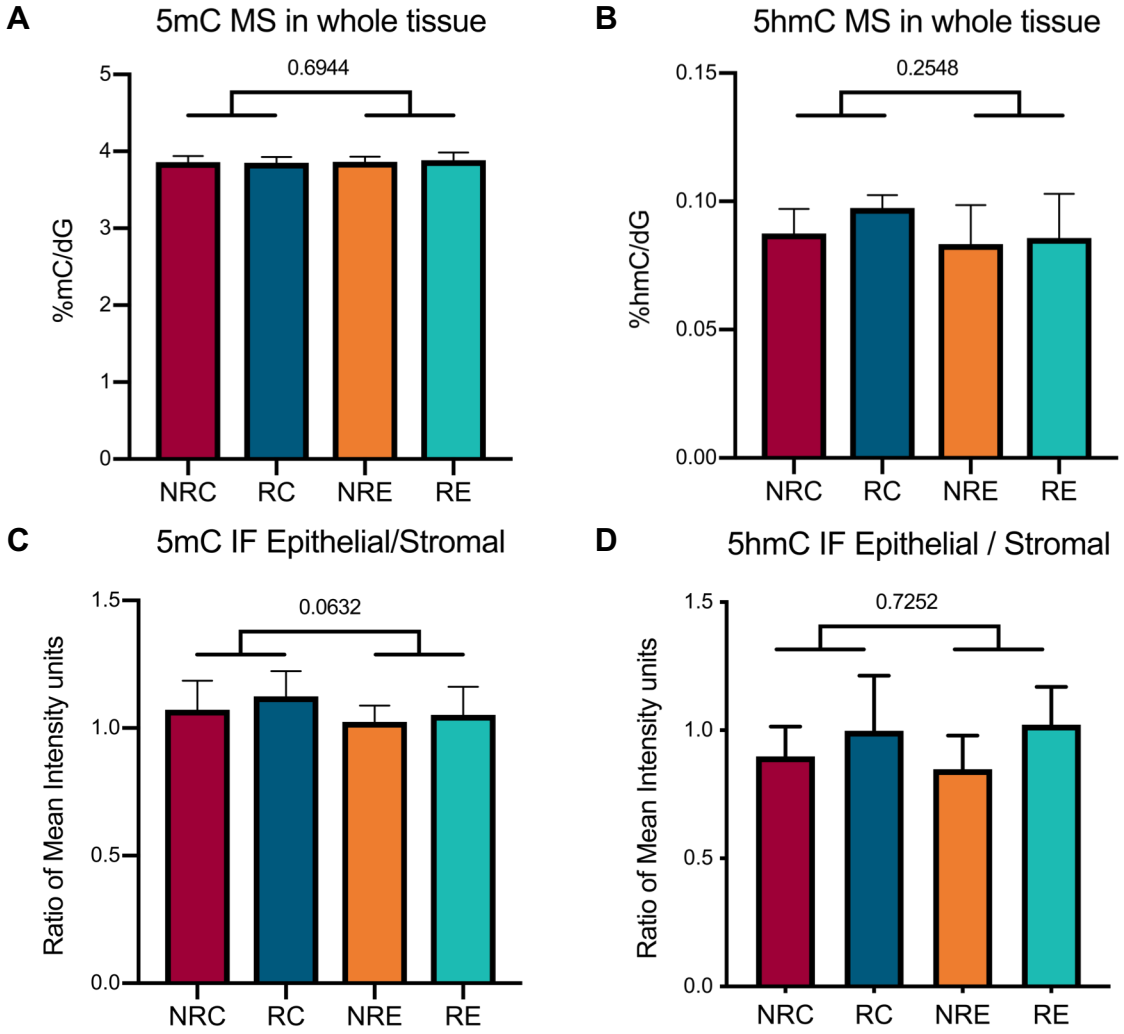



**Figure 5.6 5mC and 5hmC levels in control and endometriosis samples. A** and **B**, Mass spectrometry of 5mC (A) and 5hmC (B) of endometrial biopsies from control and endometriosis patients, plotted according to receptivity and disease state. **C** and **D**, Analysis of immunofluorescence stainings of 5mC (C) and 5hmC (D), displayed as the epithelial: stromal cell ratio of mean intensity units in the non-receptive and receptive state. Two-way ANOVA followed by Tukey's post-hoc test showed no significant differences. NRC= non-receptive control, NRE = non-receptive endometriosis, RC = receptive control, RE = receptive endometriosis.

As before, I followed up on these global and histological analyses with the establishment of genome-wide distribution profiles by meDIP-seq and hmeDIP-seq. Unlike the methylation arrays that have been reported in the literature previously, this method allows us to specifically look at 5mC and 5hmC separately, identifying regions that are enriched for either DNA methylation or hydroxymethylation, respectively.

To explore the data bioinformatically, I first created 2kb tiling probes spaced 1kb apart to cover the entire genome (Figure 5.7A to D).

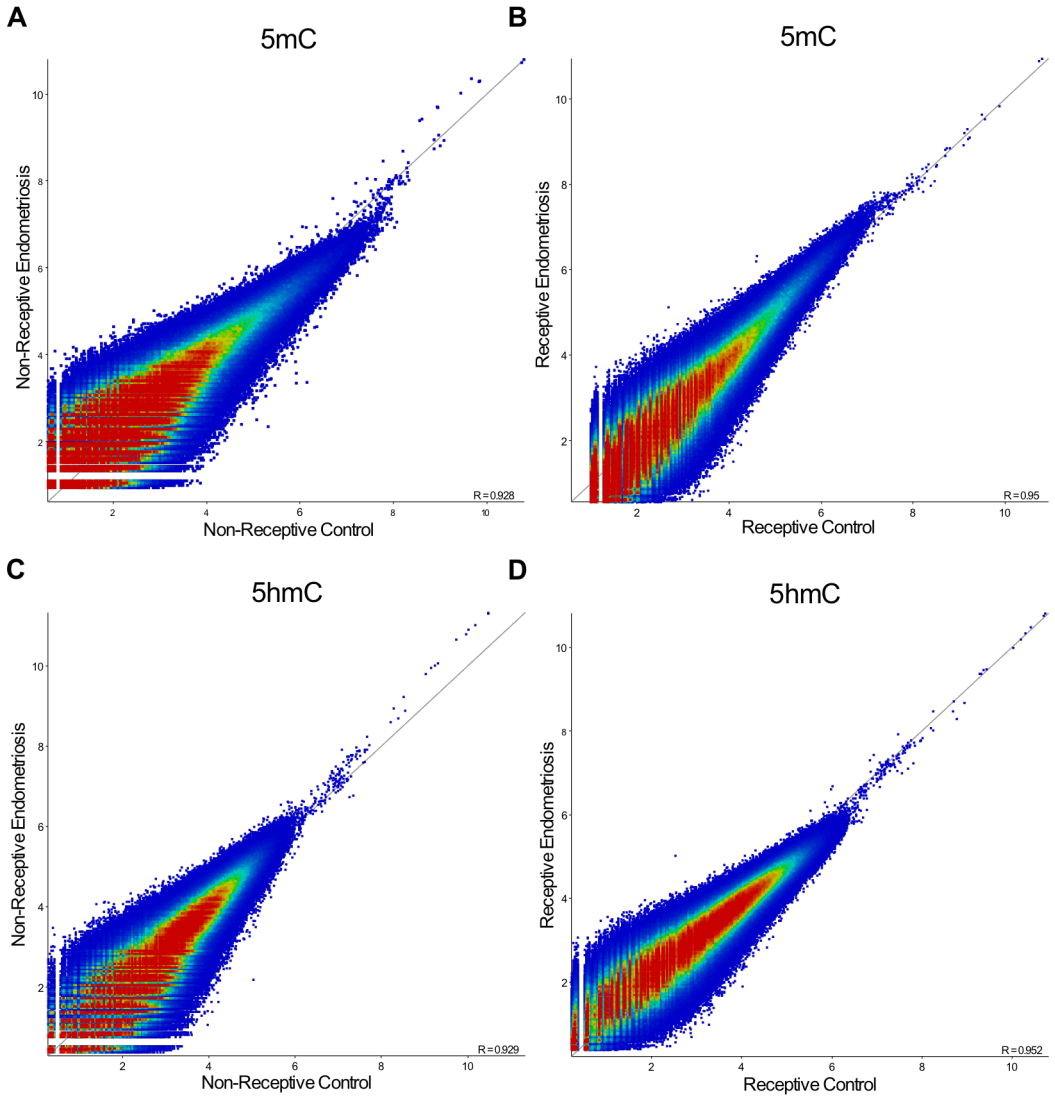



**Figure 5.7 5mC and 5hmC correlation between control and endometriosis samples. A** and **B**, Scatterplots of meDIP-seq data analysed with 2kb running window probes spaced 1kb apart, comparing controls and endometriosis samples in the non-receptive (A) and receptive (B) state. **C** and **D**, Scatterplots of hmeDIP-seq data using the same type of analysis in the non-receptive (C) and receptive (D) state. Non-receptive control (n=4), Non-receptive endometriosis (n=3), Receptive control (n=4), Receptive endometriosis (n=7).

Globally the 5mC and 5hmC profiles are similar between control and endometriosis samples with a correlation between R= 0.928 and 0.952 (Figure 5.7A to D). Since the comparisons are between the same receptivity state and between tissues of approximately similar cell type composition (see Chapter III, Figure 3.4C), a high level of correlation is expected.

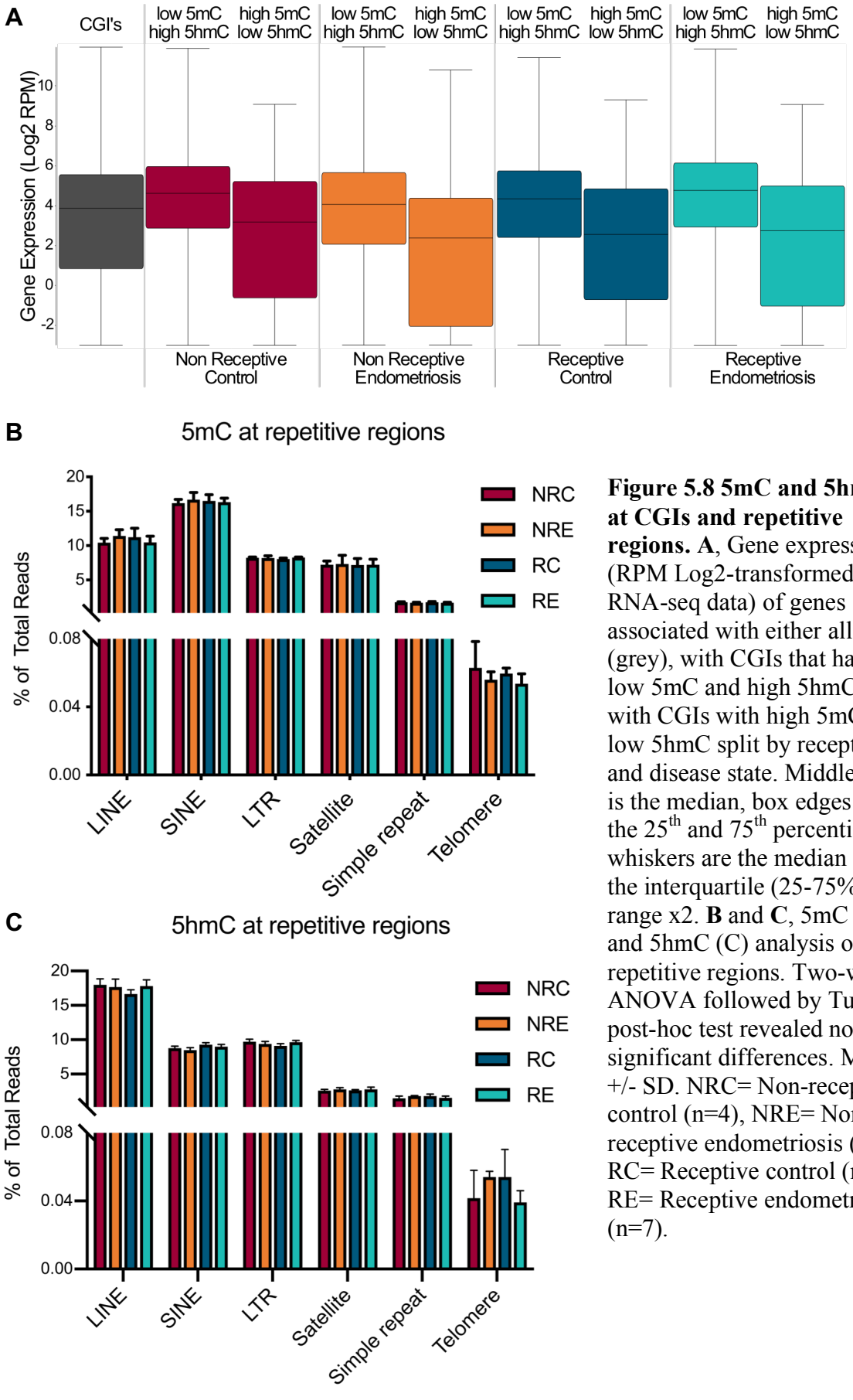



**Figure 5.8 5mC and 5hmC at CGIs and repetitive regions. A**, Gene expression (RPM Log2-transformed RNA-seq data) of genes associated with either all CGIs (grey), with CGIs that have low 5mC and high 5hmC, or with CGIs with high 5mC and low 5hmC split by receptivity and disease state. Middle line is the median, box edges are the 25$^{th}$ and 75$^{th}$ percentile, whiskers are the median +/- the interquartile (25-75%) range x2. **B** and **C**, 5mC (B) and 5hmC (C) analysis of repetitive regions. Two-way ANOVA followed by Tukey's post-hoc test revealed no significant differences. Mean +/- SD. NRC= Non-receptive control (n=4), NRE= Non-receptive endometriosis (n=3), RC= Receptive control (n=4), RE= Receptive endometriosis (n=7).

As anticipated, high levels of 5hmC and low levels of 5mC at CGIs correlated with higher expression levels of the associated genes. Correspondingly low 5hmC and high 5mC levels correlated with lower expression (Figure 5.8A). This is the same relationship that has been shown in other tissue types, suggesting that 5mC and 5hmC play roles that are similar in the endometrium as in other cell types in regulating chromatin accessibility and gene activity.

A widely accepted role for DNA methylation is its function as a repressive modification in heterochromatin where it provides genomic stability though silencing of repetitive regions (Jones 2012; Vilain et al. 1999). This function is often exerted in conjunction with H3K9me3, with the two modifications in 'crosstalk' with each other through UHRF1, a protein that can bind to both H3K9me3 and 5mC (T. Li et al. 2018; Rothbart et al. 2012). To assess whether DNA methylation or hydroxymethylation levels were altered at repeat elements, 5mC and 5hmC enrichment analysis of repetitive regions was carried out by the Babraham Bioinformatics department on my (h)meDIP-seq data, but no significant differences were found (Figure 5.8B and C).

To identify if 5mC and 5hmC deposition is more generally altered by endometriosis, I investigated enrichment levels across multiple genomic features, including promoters (TSS -1kb +100bp), CGIs, gene bodies, exons, introns and in intergenic regions (Figure 5.9A to C). I performed a two-way ANOVA, testing three parameters, namely disease state, receptivity state and the interaction between the two. The interaction null hypothesis is that there is no interaction between the disease and receptivity states, i.e. the difference associated with disease state is consistent for the non-receptive and receptive state. If the interaction p value is significant, the following null hypothesis tests on the disease factor alone and receptivity factor alone are unlikely to be significant, i.e. the disease state has different effects in the non-receptive and receptive state thus the overall disease state will unlikely be significant due to the variation. This revealed significant interactions in the 5hmC hmeDIP data for genes, introns and intergenic regions with $p<0.05$ (Figure 5.9B). This is even more striking when the 5hmC and 5mC data were combined, with a significant interaction for promoters, CGIs, intergenic regions with $p<0.01$, introns with a $p<0.05$ and with exons $p= 0.0574$ (Figure 5.9C). There is no significant difference between controls and endometriosis most likely due to variation levels.

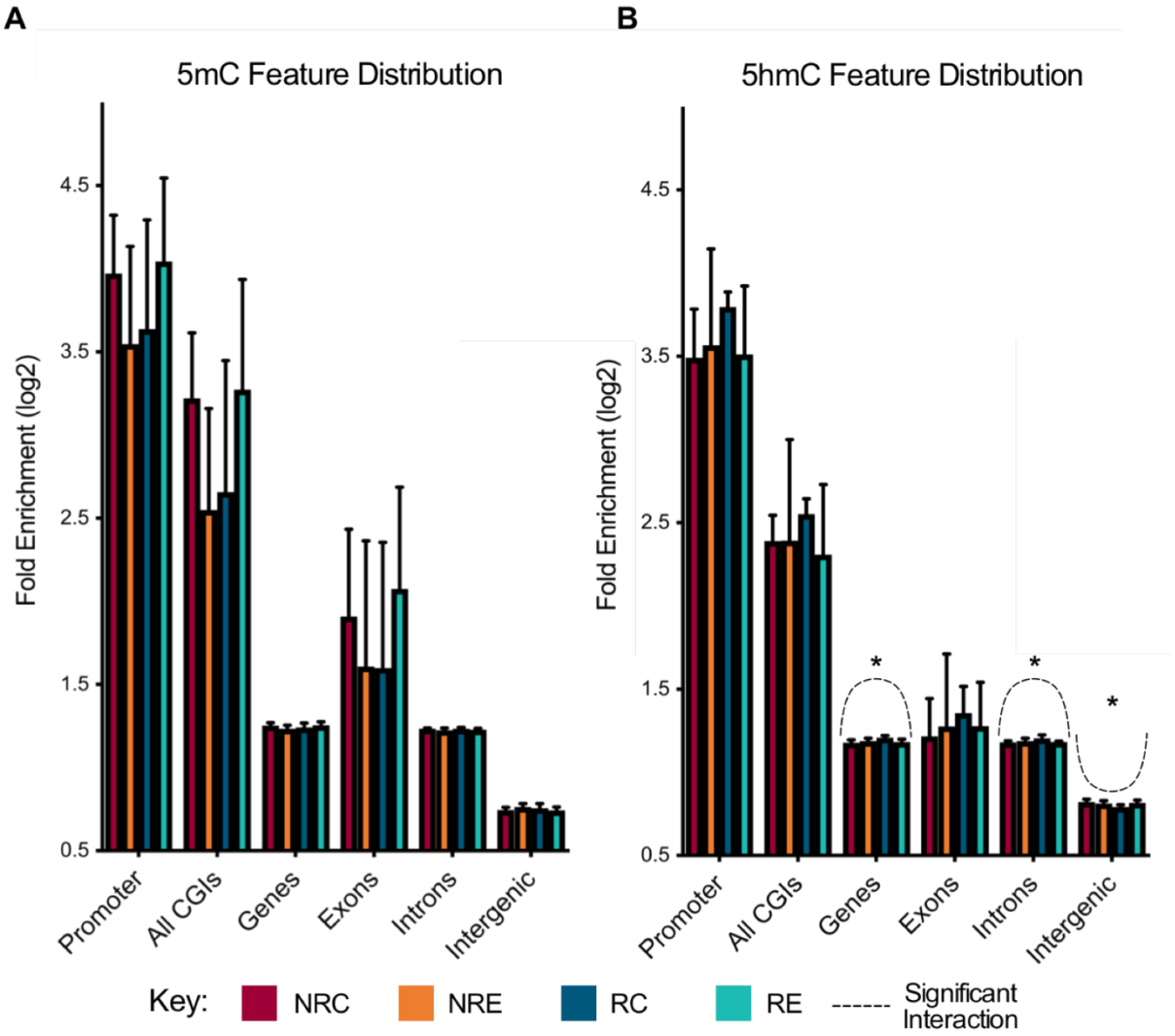



**Figure 5.9 5mC and 5hmC Feature Distribution. A**, **B** and **C**, The relative enrichment of different genomic features. Features were determined using GRCh38 in SeqMonk with enrichment calculated by: log2% reads / % genome for 5mC (A), 5hmC (B) and the ratio of 5hmC:5mC (C). Mean +/- SD. Two-way ANOVA followed by Tukey's post-hoc test revealed a significant interaction marked by asterisks, with the dotted arcs visualising the enrichment pattern in the different sample groups. **p<0.01, *p<0.05. NRC= non-receptive control (n=4), NRE = non-receptive endometriosis (n=3), RC = receptive controls (RC=4), RE = receptive endometriosis (n=7).

This pattern entails that the receptive control samples have a higher level of 5hmC enrichment compared to the non-receptive controls at genes and introns (Figure 5.9B). Promoters, CGIs and exons exhibit the same dynamics even more obviously, but the higher variability between samples interfered with statistical significance. This is even more apparent in the 5hmC: 5mC ratio where all features have a significant (or close to significant) interaction (Figure 5.9C). Therefore, when comparing control samples, all features except intergenic regions display a higher 5hmC: 5mC ratio in the receptive as opposed to the non-receptive state (compare marron and blue bars in Figure 5.9C). By contrast, these dynamics are blunted or even reversed in the endometriosis samples (orange and turquoise bars in Figure 5.9C). These data correlate with the evidence from the mass spectrometric quantification where the increase in 5hmC from non-receptive to receptive samples appeared blunted (Figure 5.6B), even if this was not as obvious in the IF data (Figure 5.6D). This may mean that the abrupt gene activation that sets in with the acquisition of receptivity and correlates with higher 5hmC levels in glandular and luminal epithelial cells (Chapter IV) may be less pronounced, or is temporally less well orchestrated, in endometriosis.

Intergenic regions exhibited the opposite change, with 5hmC enrichment being lower in receptive compared to non-receptive control samples. Again, in endometriosis, these changes are less evident. The complementary patterns are reflected in the 5mC features, although there is no significant interaction due to the high standard deviations.

### 5.4.2 Locus-specific changes in DNA methylation and hydroxymethylation reveal epigenetic de-regulation of the Hippo signalling pathway

To follow up on the altered 5mC and 5hmC feature distribution, I investigated locus-specific changes that may occur across the genome using 2kb tiling probes. I conducted an unpaired, two-tailed t-test between control and endometriosis samples and then applied at least a value difference of 2 to the log2 transformed data (meaning ≥4-fold enrichment/depletion). This highlighted a number of regions that were differentially methylated or hydroxymethylated between controls and endometriosis samples (Figure 5.10A to D). Gene ontology (GO) analysis was conducted on the top 500 differentially hydroxymethylated or methylated regions found within 2kb of a gene, using DAVID (https://david.ncifcrf.gov/) with Bonferroni correction applied (Figure 5.10E).

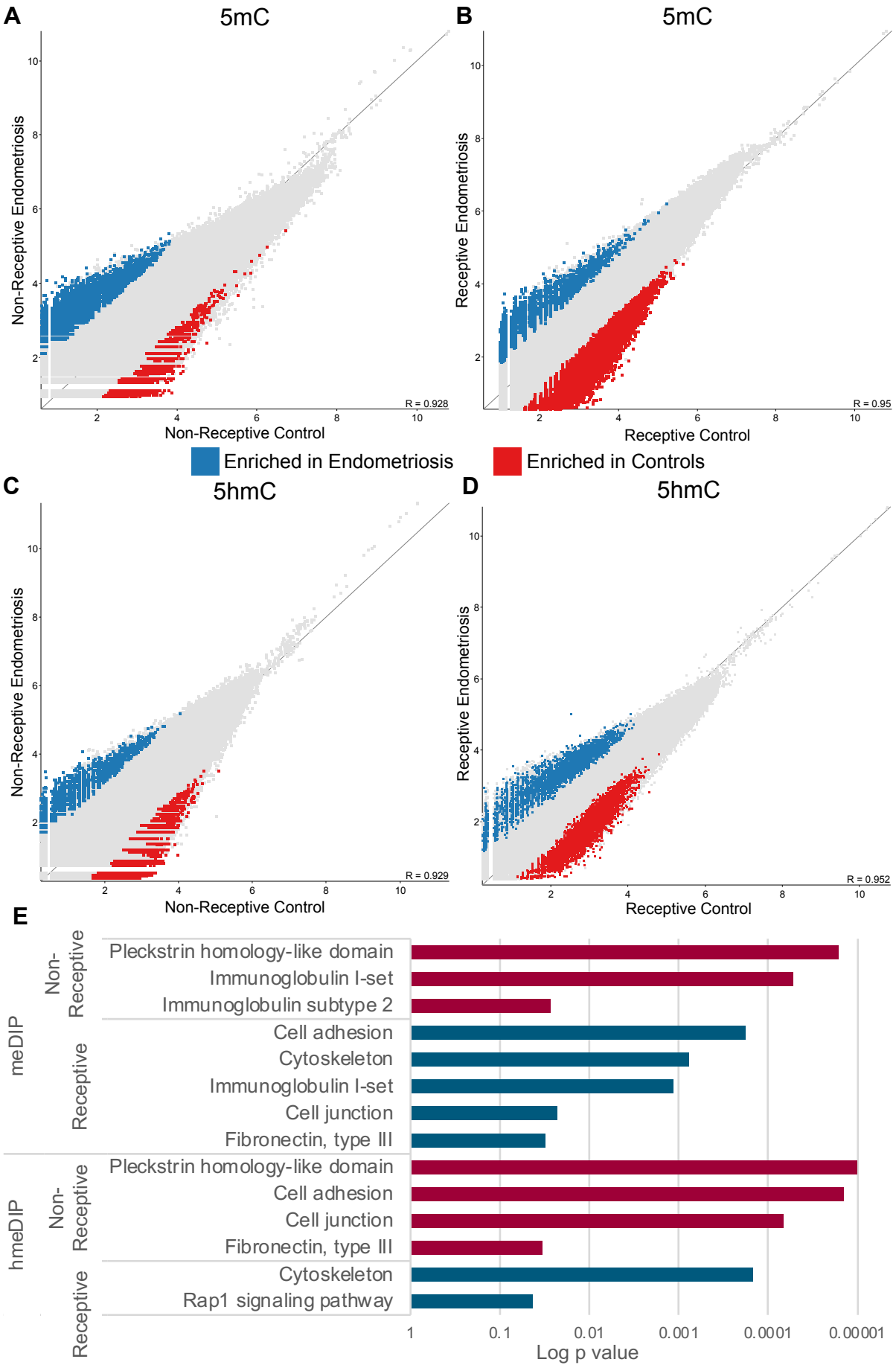



**Figure 5.10 Differentially methylated and hydroxymethylated regions in endometriosis.** **A** and **B**, Scatterplot of meDIP-seq data analysed with 2kb running window probes spaced 1kb apart, comparing controls and endometriosis samples in the non-receptive (A) and

receptive (B) state. Differentially methylated probes are highlighted. **C** and **D**, Scatterplot of hmeDIP-seq data analysed in the same way in the non-receptive (C) and receptive (D) state, again with differentially hydroxymethylated probes highlighted. Differential enrichment was determined by unpaired two-tailed t-test and a log2-fold change of at least 2. **E**, Gene Ontology analysis of genes within 2kb of the top 500 differentially marked regions. This was carried out using the DAVID bioinformatics tool with Bonferroni correction $p<0.05$ applied**.**

A number of GO terms are significantly enriched and include relevant and interesting pathways. Of note is the enrichment of the terms 'cell adhesion', 'fibronectin' and 'immunoglobulin' with the former two present in both the meDIP and hmeDIP analysis and the latter present in both receptivity states in the meDIP analysis. These processes have previously been implicated in endometriosis. When scrutinizing the list of differentially modified loci, I noted that a number of genes associated with these regions were involved in the extended Hippo signalling pathway. These genes were associated with one or several significantly differentially methylated or hydroxymethylated 2kb regions (DMRs or DhMRs) in either the non-receptive or receptive state (Figure 5.11A).

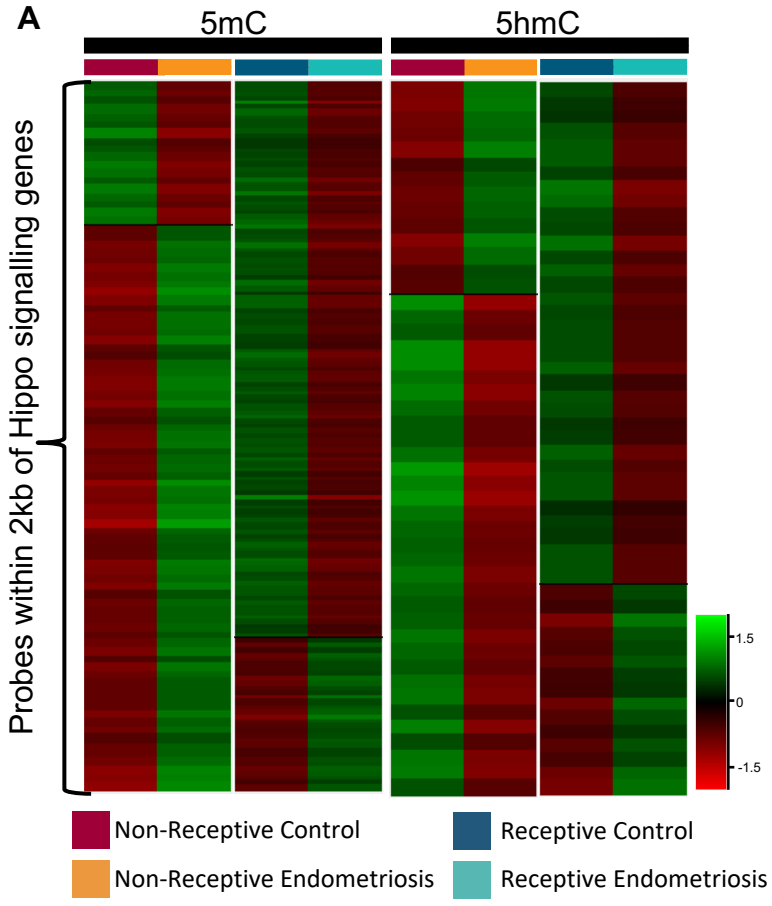



**Figure 5.11 Heatmap of DMRs and DhMRs associated with Hippo signalling genes. A**, Heatmap of 2kb probes differentially enriched for 5mC or 5hmC that are associated with the genes in the Hippo signalling pathway. Samples pooled for each group: Non-receptive control (n=4), Non-receptive endometriosis (n=3), Receptive control (n=4), Receptive endometriosis (n=7).

I used the KEGG pathway hsa04390 as the standardised list of genes involved in the Hippo signalling pathway. The genes with DMRs (orange box) or DhMRs (yellow box) associated with genes in this list are highlighted on the Hippo pathway map and are significantly enriched in the meDIP analysis with the corresponding hmeDIP data just missing significance (p= 0.0528) (Figure 5.12A and B).

A full list of Hippo pathway genes and genes associated with D(h)MRs can be found in the appendix (Supplementary Table 6). There are a number of genes that are present in both the non-receptive and receptive state as well as genes that had D(h)MRs in only one state (Figure 5.12C). There is no obvious correlation between the Hippo genes that are differentially methylated and/ or hydroxymethylated with expression levels (Supplementary Figure 6 and 7). This, however, is not surprising as the differentially marked regions occurred anywhere in the gene and some genes have multiple, and sometimes 'antagonistic' (i.e. both 5mC and 5hmC enriched) peaks. Also, there are very few expression differences that arose when comparing control and endometriosis samples ( Figure 5.1A and B).

Overall, this data revealed that the Hippo signalling pathway components are subject to epigenetic changes at significant and higher levels than other intra- or intercellular pathways in endometriosis. Even though there was no clear-cut correlation with regards to the direction of 5mC and 5hmC enrichment, this data reveals a higher susceptibility of this pathway to epigenetic change in endometriosis. The Hippo pathway incorporates cell positional and cell-cell contact information to direct cell proliferation, polarity and adhesion. These are processes of key relevance to the patho-aetiology of endometriosis.

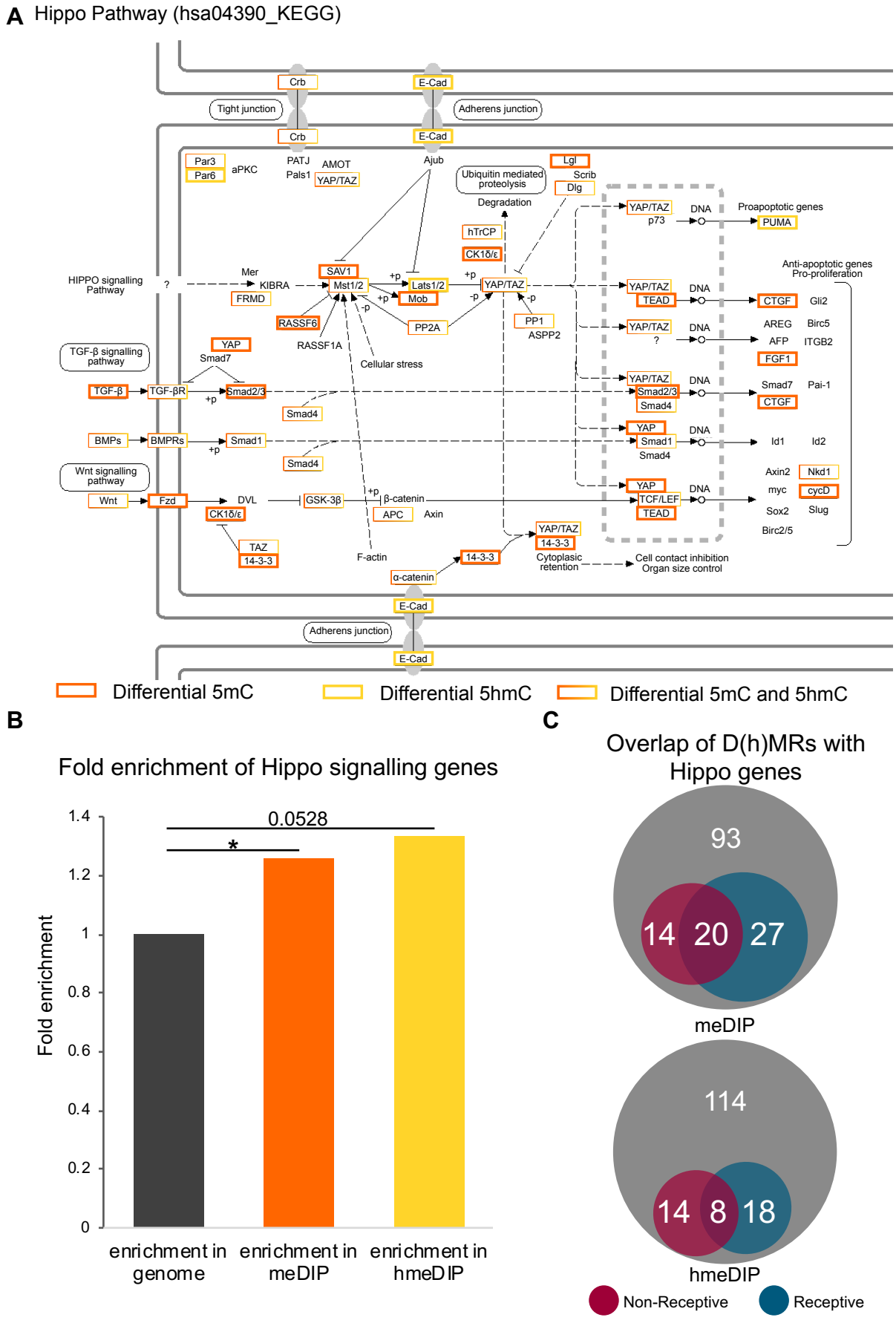



**Figure 5.12 Differential 5mC and 5hmC associated with the Hippo Pathway. A**, Diagram of the Hippo pathway (hsa04390) from KEGG (https://www.genome.jp/dbget-bin/www_bget?pathway+hsa04390), with genes close to or overlapping differentially

enriched epigenetic marks colour-coded as follows: DMRs (orange), DhMRs (yellow), and DMRs and DhMRs (orange and yellow). **B**, Enrichment of Hippo signalling genes associated with 2kb regions that were differentially marked by 5mC or 5hmC. Chi-square test with Yates correction applied *p<0.05. **C**, The overlap of differentially marked 2kb regions with Hippo signalling genes, defined by KEGG.

### 5.4.3 Validation of an epigenetic signature for endometriosis

Establishing a diagnostic test that is less invasive than laparoscopic surgery for endometriosis would be highly desirable. As I have shown that the 5mC and 5hmC patterns are disrupted in the eutopic endometrium of women with endometriosis, this opens up a new avenue to investigate and to potentially identify an epigenetic signature and putative biomarker panel of the disease.

I explored the meDIP-seq and hmeDIP-seq data using different probe types to select a list of DMRs and DhMRs, which included a number of regions that overlap genes involved in the Hippo signalling pathway. When used for principal component analysis, these probes successfully separated out the samples according to receptivity state as well as disease state in the meDIP and hmeDIP data (Figure 5.13B and C).

The differentially enriched probes used to generate this collection of candidate regions consisted of 411 loci that were combined from the analysis of the 2kb tiling probes, CGIs, Promoters (define as -1kb to +100 bp surrounding TSSs) and Exon probes (Figure 5.13A). I also analysed the data using the MEDIPS package in R with default settings (Lienhard et al. 2014), with window size defined as 200bp. This identified a total 4 regions, 2 of which were next to each other and present in both meDIP and hmeDIP. I also integrated FOXO1 and PGR binding sites along with DMRs identified in previous studies. The PGR (GSM1703567) and FOXO1 (GSM1703607) binding sites were identified in cultured human endometrial stromal cells from primary biopsies taken during the proliferative phase. These factors play a key role in the preparation for implantation (Mazur et al. 2015; Vasquez et al. 2015). Kukushkina et al. data (GSE90060) identified DMRs and CpGs that were differently methylated between LH+2 and LH+8 (Kukushkina et al. 2017). This allowed me to assess if my meDIP- and hmeDIP-data also showed differential enrichment of DNA methylatation or hydroxymethylation at these sites resulting in the inclusion of additional probes to be validated (Figure 5.13A).

**A**

| Region Identity | meDIP | hmeDIP |
|---|---|---|
| 2kb | 76 | 25 |
| CGI | 37 | 44 |
| Promoters | 19 | 21 |
| Exons | 70 | 27 |
| MEDIPs | 1 | 3 |
| FOXO1 binding sites | 16 | 29 |
| PGR binding sites | 11 | 9 |
| Previously Identified DMRs or CpGs | 6 | 17 |
| **Total** | **236** | **175** |
| | | **411** |

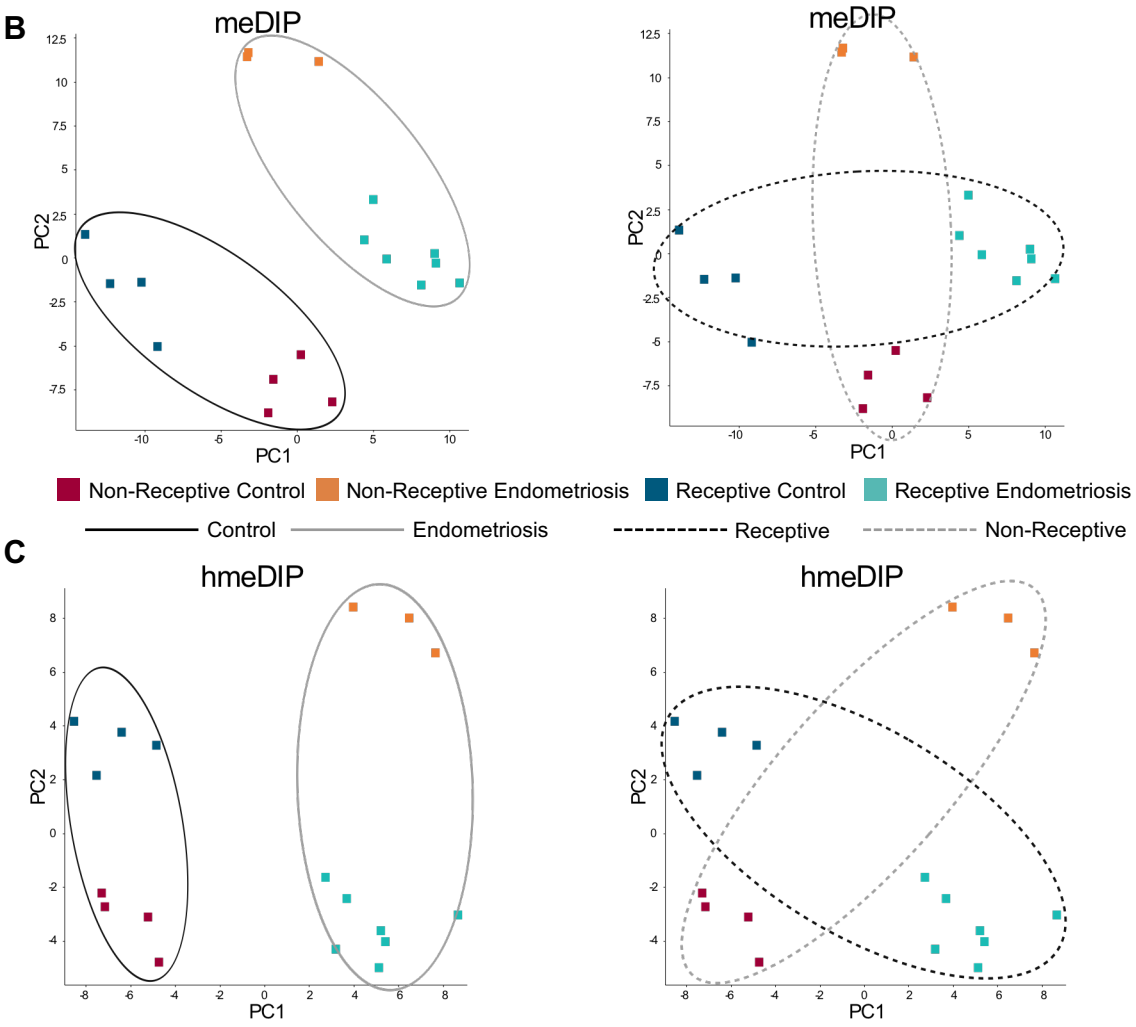



**Figure 5.13 Candidate regions identified by the meDIP- and hmeDIP-seq analysis selected for independent validation. A**, Distribution of regions selected for validation across the genome. FOXO1 data = GSM1703607, PGR data = GSM1703567, Previously Identified DMRs or CpGs = GSE90060. **B** and **C**, Principal component analysis (PCA) plots of the meDIP (B) or hmeDIP (C) regions selected for validation. Each plot is highlighted for either the disease state or receptivity state. Solid lines denote control (dark grey) or endometriosis (light grey). Dotted lines denote receptive state (dark grey) or non-receptive state (light grey). Non-receptive control (n=4), Non-receptive endometriosis (n=3), Receptive control (n=4), Receptive endometriosis (n=7).

Oligonucleotide probes were designed to cover all the CpGs within the 411 regions identified using a method established by CEGX. I then used a protocol developed by CEGX called TrueMethyl® On Target (TMOT), which is a commercially available technique, to validate the DNA methylation and hydroxymethylation patterns at a CpG base-specific level. The protocol utilises bisulphite (BS) and oxidative bisulphite (OxBS) conversion followed by sequencing (Figure 5.14A and B). BS treatment results in the conversion of unmodified cytosines to a thymine whilst methylated and hydroxymethylated cytosines are protected from this conversion, and so are still read as a 'C' in the sequencing. Therefore, the combined location of either 5mC or 5hmC can be established (Figure 5.14A and B). OxBS conversion results in 5hmC being first oxidised and subsequently being converted to a thymine whereas 5mC remains protected and is read as a C in the sequencing. Therefore, the location of the 'true' 5mC can be established. In order to determine the 5hmC positions, the oxBS data is bioinformatically subtracted from the BS data (Figure 5.14A and B). With my relatively small collection of specific target loci the sequencing depth per probe will be much higher, which makes this a superior technique to whole genome analyses to confirm the 5mC and 5hmC modification state of candidate regions.

Calls for 5mC bases were produced using the Bismark programme for both the oxBS and BS datasets by the Babraham Bioinformatic department (Krueger and Andrews 2011). To produce the 5hmC count the adjustMethylC function within the MethylKit package was used to subtract oxBS reads from BS resulting in 5hmC reads. The MethylKit package in R was used to call methylation differences between experimental conditions (Akalin et al. 2012). The Q value is a confidence metric for how reliable the methylation call is across multiple samples. The number of CpGs that pass this test could be increased with the addition of more samples.

A

| Base | C | 5mC | 5hmC |
|---|---|---|---|
| Sequencing | C | C | C |
| BS Conversion and Sequencing | T | C | C |
| OxBS Conversion and Sequencing | T | C | T |

B

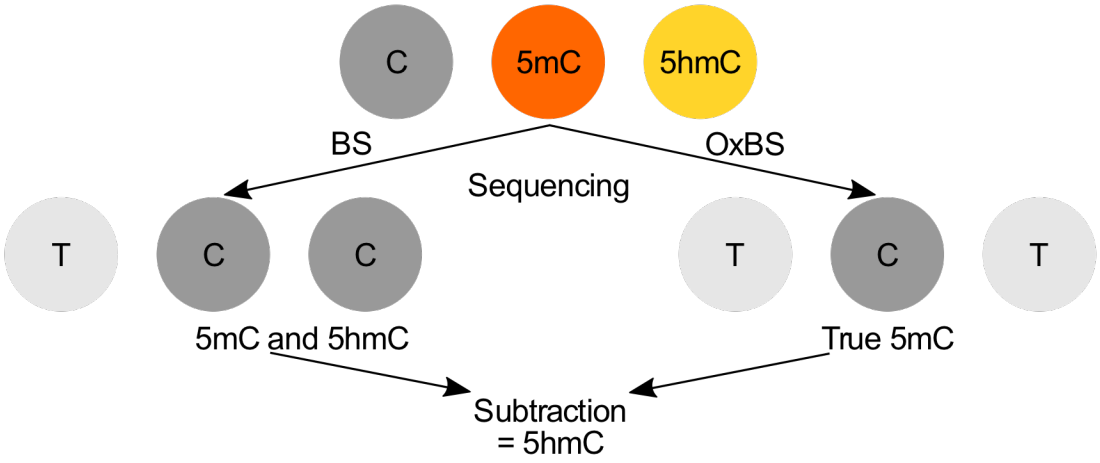



**Figure 5.14 OxBS and BS sequencing. A**, The base calling after oxBS or BS conversion and subsequent sequencing. **B**, Schematic of the sequencing read out and the establishment of 5mC and 5hmC modifications.

Overall a substantial number of differentially modified regions were validated by TMOT (Figure5.15A). The percentage of 5mC CpGs that validated was high for all conditions, averaging at 74.2%. Validation for 5hmC CpGs was lower than 5mC, averaging at 47.9% (Figure5.15A), although this is expected and is most likely due to the technical noise in the 5hmC datasets as debated in the discussion. This can be further seen in the Manhattan plots which show the validated CpGs in green and the CpGs that did not validate in red (Figure5.15B to I).

**A**

| | State | MethylKit CpGs that overlap with h/meDIPS peaks | Of which q<0.01 | Of which validate | %validated |
|---|---|---|---|---|---|
| **5mC** | C vs E | 243 | 89 | 80 | 90% |
| | NRC vs RC | 1000 | 93 | 55 | 59% |
| | NRC vs NRE | 12 | 3 | 2 | 67% |
| | RC vs RE | 2014 | 36 | 27 | 75% |
| **5hmC** | C vs E | 5 | 5 | 2 | 40% |
| | NRC vs RC | 984 | 205 | 91 | 44% |
| | NRC vs NRE | 984 | 137 | 96 | 70% |
| | RC vs RE | 2788 | 377 | 158 | 42% |

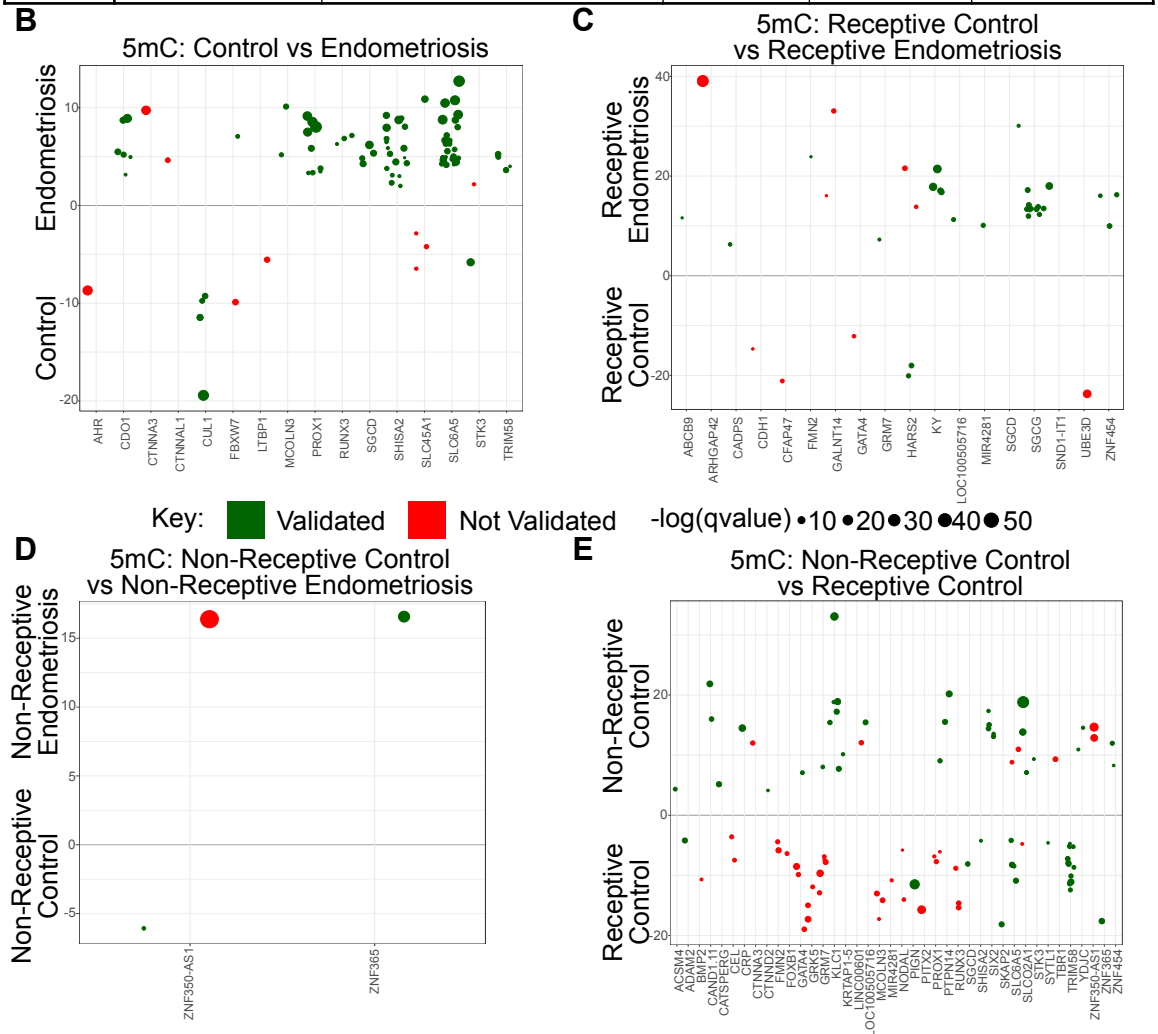



**Figure 5.15 TMOT validation of selected candidate regions. A,** The table shows the number of CpGs that are present in the regions that are to be validated (in column two). The number of CpGs that have a q value <0.01 (confidence in the methylation call across samples) are shown (in column three), followed (in column four) by the number of CpGs that validated and (in column five) the percentage of CpGs that validated. **B**, **C**, **D**, and **E**, Manhattan plots of the CpGs that had q values <0.01 with the y-axis showing the % of differential methylation between conditions and the size of the dot positively correlating with the significance of the differential methylation. In these graphs, each vertical line reflects one particular gene locus that may contain multiple CpG sites. 5mC or 5hmC "validation" is considered for each CpG. Validated = green, Not Validated = red. Figure continued on following page.

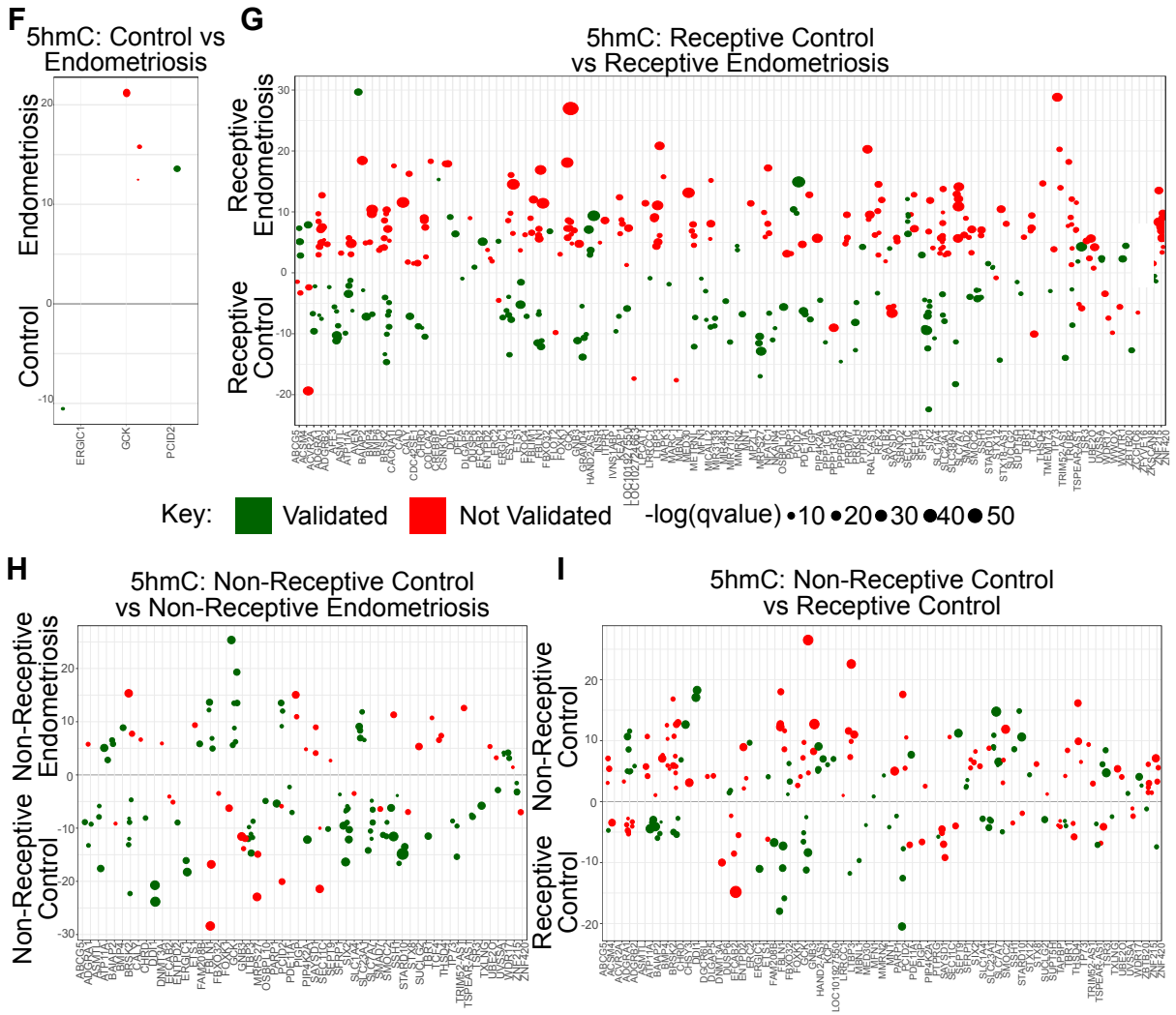



**Figure 5.15 TMOT validation continued. F**, **G**, **H** and **I**, Manhattan plots of the CpGs for hydroxymethylation using the same analysis as in B to E. Validated = green, Not Validated = red.

It is worth noting that for some regions that contained multiple CpGs, all individual CpG dinucleotides validated with the expected modification whereas in other instances only one or two CpGs within a gene/region validated. The latter situation could be due to a technical reason with regards to the hmeDIP and meDIP protocol and is further explored in the discussion. The majority of CpGs that did not validate are distributed randomly between the datasets, however when comparing 5hmC control and endometriosis samples in the receptive state the CpGs that did not validate are predominately present in the endometriosis samples (Figure5.15G). The same is true for the 5hmC comparison of controls between the receptive and non-receptive state, with CpGs predominately not validating in the non-receptive control samples (Figure5.15I). A possible explanation for this could be the differing techniques used. hmeDIP-seq is an enrichment technique, making it difficult to control for differences in total 5hmC levels between samples after

normalisation is applied to account for differences in total read count. When samples are normalised by total read count, samples with more reads ("sample A") are proportionally reduced in size over the entire genome to be comparable to samples with fewer reads ("sample B"). This means that regions where 5hmC is in fact equal in A and B they appear to be lower in A, as A was 'downsized' during normalisation. Therefore, differences in total 5hmC levels may yield false positive results after normalisation for read count. In contrast, oxBS and BS-sequencing gives an approximation of the 'absolute' levels of 5hmC at an individual CpG, and therefore is not subject to distortion by the normalisation processes. The mass spectrometry data suggests that 5hmC increases in the receptive state compared to the non-receptive state and that this increase may be blunted in endometriosis (Figure 5.6B). Therefore, the CpGs that did not validate in these circumstances could be due to this technical artefact in the technique. This highlights the need for validating candidate regions identified, and adds weight to the importance of establishing total 5mC and 5hmC levels to reveal epigenetic differences between sample groups, in my case between disease and receptivity states.

Despite these cases of non-validated CpG sites, the significance of those that did validate is highlighted in a principle component analysis (PCA), using only the CpGs that validated, which discriminates not only between receptivity states but also between disease states (Figure 5.16A and B). This is similar to the PCA plots derived from the original probe set before validation (Figure 5.13B and C), giving me confidence that the epigenetic state of a sufficient number of probes was confirmed to still achieve the same separation of sample groups. This is a highly promising dataset and upon validation of further independent samples could lead to the establishment of a set of loci that can serve as biomarker for both endometrial receptivity and endometriosis, which is currently lacking and vitally needed within the field.

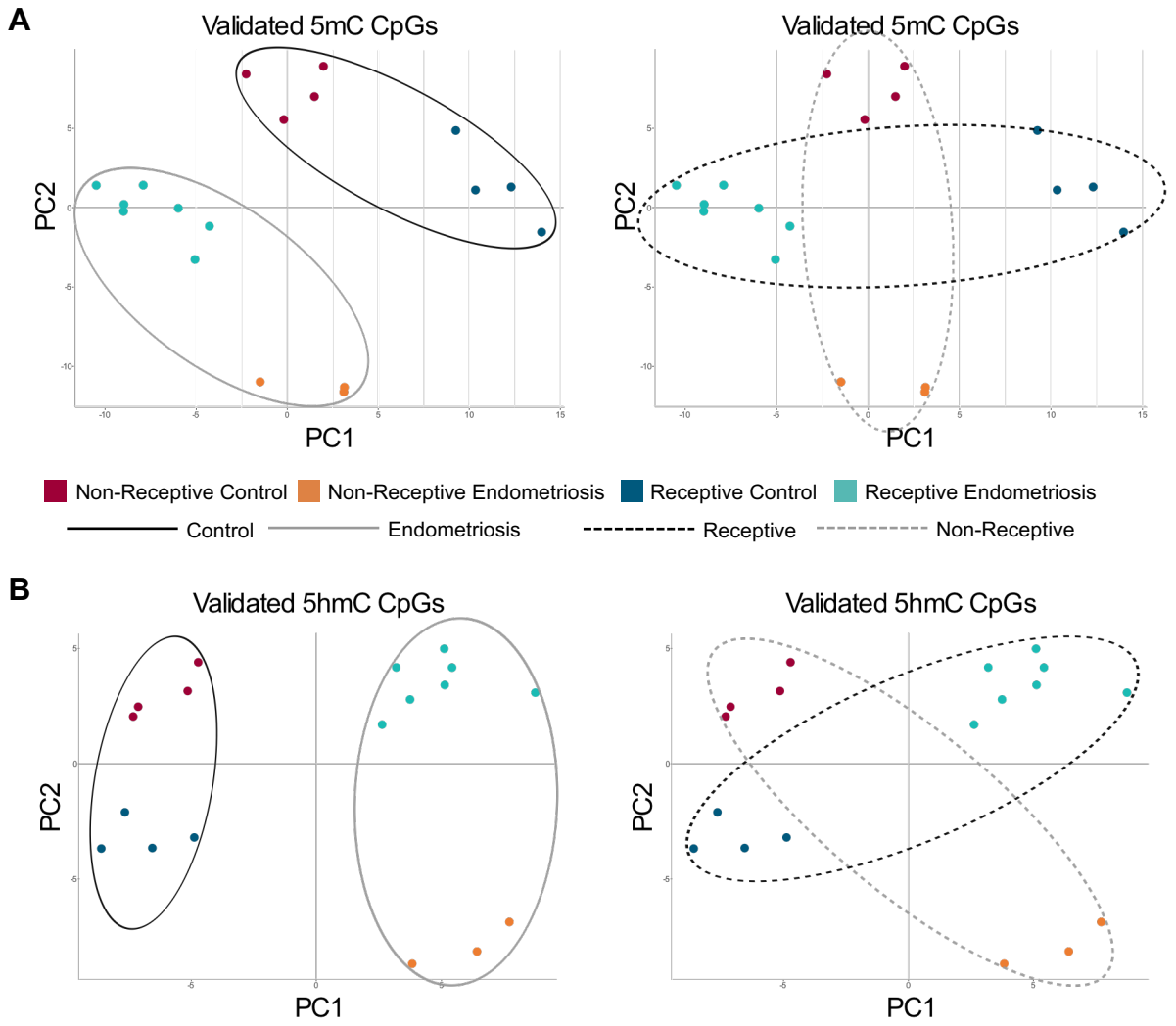



**Figure 5.16 CpGs that validated for differential 5mC and 5hmC enrichment. A** and **B**, PCA plots of the validated individual 5mC CpGs (A) or validated individual 5hmC CpGs (B) from the TMOT validation analysis. Each plot is highlighted for either the disease state or receptivity state. Solid lines denote control (dark grey) or endometriosis (light grey). Dotted lines denote receptive state (dark grey) or non-receptive state (light grey). Non-receptive control (n=4), Non-receptive endometriosis (n=3), Receptive control (n=4), Receptive endometriosis (n=7).

## 5.5 Discussion

Endometriosis is a complex disease and the precise mechanism of how it develops remains unknown. The main theory in the field is that endometriosis arises from retrograde menstruation, where endometrial tissue exits the uterus via the fallopian tubes and fragments of tissue or cells adhere and grow in ectopic positions, forming lesions (Sampson 1927). This theory would mean that ectopic lesions are formed from what was originally eutopic tissue. Therefore, the eutopic tissue itself could already be

epigenetically altered which consequently allows cells to adhere and proliferate more effectively in such an ectopic location.

There are no transcriptional differences associated with endometriosis in the receptive state in my dataset. However, there are a handful of genes that are differentially expressed in the endometrium from patients with endometriosis in the non-receptive state (Figure 5.1B). These genes are involved in processes that have been suggested to be involved in endometriosis, including cell adhesion, immunological functions and Epidermal Growth Factor (EGF) signalling. Amongst the genes detected in my data, *TNC*, *PLA2G4E* and *THEM4* have also been shown in other studies to be dysregulated in endometriosis (Afshar et al. 2013; Flores et al. 2007; Irungu et al. 2019), which is a great corroboration of my data and the general validity of these genes as targets in endometriosis. Whilst *LRP4*, *TMED6* and *SLC16A6* have been shown to change expression during the menstrual cycle (Chan et al. 2013; Díaz-Gimeno et al. 2011), these genes had not been identified as de-regulated with endometriosis before. These results also agree with the majority of reports, which showed that the expression profiles altered are of genes that affect the transition from the proliferative to the secretory phase. This could be due to a delay the endometrium responding to progesterone, as progesterone resistance is commonly observed in endometriosis (Aghajanova et al. 2011; Attia et al. 2000; Burney et al. 2007; Rahmioglu et al. 2017). As such it seems logical that these genes stand out as differentially expressed between control and endometriosis samples; it aligns with the fact that some of the genes that were dysregulated have been associated with receptivity and therefore, their induction may be delayed in patients with endometriosis.

However, these results conflict with Fung et al. 2018, who suggest that there are no alterations in the gene expression profiles that are due to endometriosis when adjusted for the menstrual cycle (Fung et al. 2018). They did report that the levels of *HOXA10* were more variable in patients with endometriosis in the late secretory phase compared to controls. The discrepancies between studies could be due to how the menstrual phase was determined as well as the fact that different microarrays were used. It is difficult to delineate if the endometriosis samples are simply delayed in their receptivity or if they have aberrant timing. This leads to the assumption that the expression differences are due to endometriosis. These reports including my own show only few genes that are altered and of these, the expression level changes are relatively low. Further comprehensive analysis using a larger sample size of endometrial biopsies which are

precisely timed for their receptivity is required, as has also been suggested by Rahmioglu et al. (Rahmioglu et al. 2017).

My findings reported in this chapter also show that key histone methylation marks remain mostly unaffected by endometriosis in eutopic tissue, with no significant differences observed for H3K4me3 and H3K9me3 when receptivity is considered. However, there are seven regions that are significantly differentially enriched for H3K27me3 in my endometriosis samples within the receptive state. These sites and/or the nearest genes have not been associated with endometriosis previously and some peaks are in intergenic regions. TMEM176A/B is particularly noteworthy as the TSSs have higher levels of H3K27me3 in receptive endometriosis samples. This gene inhibits the inflammation response and TMEM176B null mice fail to recruit neutrophils (Segovia et al. 2019). A potential consequence of the enrichment of H3K27me3 over the TSSs of TMEM176A and B could be that it attenuates their expression when the tissue is in an ectopic environment. This in turn could lead to an activation of the inflammation response, which is associated with endometriosis and its progression (Bruner-Tran et al. 2013; Burney et al. 2007; Y. Li et al. 2016). These regions including TMEM176A and B will need to be further verified and validated in additional samples along with mechanistic studies to assess the functional consequences that these differential marks may impose.

There are several papers that show that histone methylation is altered in endometriotic lesions. These studies generally measured the total levels of a particular modification in tissues. They did not necessarily discriminate between mono, di or tri methylation at specific H3 lysine residues, or the menstrual cycle phase. EZH2 (methyltransferase for H3K27me3) immunostaining has shown to be upregulated in eutopic and ectopic tissue in patients with endometriosis compared to controls and another study showed that it was significantly upregulated in eutopic tissue from patients with endometriosis between the proliferative and secretory phase, whereas controls showed no difference between phases (Colón-Caraballo et al. 2018; Q. Zhang et al. 2017).

Histone methylation differences that arise in ectopic lesions may be a consequence of the tissue's relocation and response to the differing environmental signals. Alternatively, it is possible that the differing histone methylation marks pre-existed and facilitated the formation of lesions by promoting tissue adherence and proliferation. However, here I show that there are no overarching changes in H3K4me3, H3K9me3 or H3K27me3 and

that genes that are differentially expressed in the non-receptive state do not exhibit corresponding differences in their histone status. Overall, I conclude that changes to the epigenetic ground state at the histone methylation level are not a prevailing factor in the aetiology of endometriosis.

As discussed in Chapter IV, the actual levels of 5mC and 5hmC have not previously been established in endometrial tissue, let alone compared to eutopic tissue from patients with endometriosis. As the comparison of global 5mC and 5hmC is between tissues of the same receptive state with approximately similar tissue composition, a high correlation is to be expected. The combinatorial status of DNA methylation and hydroxymethylation levels at promoters and CGIs correlates with transcriptional activity of the associated genes as has been previously shown in other tissues, notably in embryonic stem cells (Ficz et al. 2011; Mellen et al. 2012; R. K. Ng et al. 2008). When considering other genomic features, it is noteworthy that Borghese et al. suggested that subtelomers are hypermethylated at some chromosomes in ectopic endometriotic tissue. I do not see such an effect at telomeres, which may be due to the fact that our analysis compared eutopic and not ectopic tissue (Borghese et al. 2010).

The significant interaction and pattern of 5hmC and 5mC in the feature distribution analysis suggests that overall the distribution and turnover of epigenetic modifications to DNA (i.e. 5mC and 5hmC) are affected in eutopic tissues of patients with endometriosis. In fact Wu et al. showed that DNMT3A is over-expressed in eutopic tissue from patients with endometriosis and that all three DNMTs are over-expressed in ectopic tissue compared to controls (Wu et al. 2007), although this is not apparent in my samples. This disruption to the epigenetic configuration assessed by genomic feature distribution can be seen at specific loci that are associated with a number of gene ontology terms connected with endometriosis. This could imply that the epigenetic profiles of 5mC and 5hmC are already altered in eutopic tissue of patients with endometriosis and that these disruptions may lead to, or facilitate, cellular adherence and formation of lesions. This is congruent with the observed differential methylation and/or hydroxymethylation of many genes involved in the Hippo signalling pathway. Differential peaks are present in both the non-receptive and receptive state suggesting that the maintenance and propagation of these marks is not accomplished correctly. The Hippo pathway is involved in cellular functions, which have been implicated in the

progression of endometriosis; these include cell polarity and adhesion as well as apoptosis and proliferation (Yu and Guan 2013). This data strongly resonate with those by Song et al., who showed that the Hippo pathway, and in particular YAP, was significantly up-regulated in ectopic lesions and showed a greater variance in eutopic tissue of patients with endometriosis compared to controls (Y. Song et al. 2016). They also showed that over-expression of YAP in endometrial stromal cells promoted proliferation and reduced apoptosis, key process involved in endometriosis (Y. Song et al. 2016). Although we observe no expression changes of the Hippo genes, our sample size did not allow to determine levels of variance once disease and receptivity state had been accounted for (i.e. the *n* in each of the four sample groups was too small). However, even if there is no consistent effect on the mRNA level, there could still be an effect at the protein level or on the level of subcellular protein localisation, which is a key regulatory mechanism in the Hippo signalling pathway, for example on the levels and localisation of TAZ as shown in decidualised stromal cells (Strakova, Reed, and Ihnatovych 2010).

There have been reports of locus-specific hyper- or hypomethylation at *HOXA10*, *PGR*, *ESR2* and *SF-1* (also known as *NR5A1*) in endometriosis (J. J. Kim et al. 2007; Kulp, Mamillapalli, and Taylor 2016; B. Lee, Du, and Taylor 2009; Wu et al. 2006; Xue, Lin, Cheng, et al. 2007; Xue, Lin, Yin, et al. 2007). In our datasets, no significant differences in methylation or hydroxymethylation levels were detected at the promoters of *HOXA10*, *ESR2* and *SF-1*. This is in conflict with previous reports. However, this discrepancy could be due to several reasons. The *HOXA10* methylation analysis was carried out in different model systems (baboons and mouse) with induced endometriosis. In human studies, where this locus has been reported, the menstrual phase in which the sample was taken was not recorded (J. J. Kim et al. 2007; Kulp, Mamillapalli, and Taylor 2016; B. Lee, Du, and Taylor 2009). *ESR2* and *SF-1* methylation was determined by comparing ectopic and eutopic cultured endometrial stromal cells (Xue, Lin, Cheng, et al. 2007; Xue, Lin, Yin, et al. 2007). In my data, the *PGR-B* promoter does show an increase in methylation levels in receptive endometriosis samples compared to receptive controls. However, this differential enrichment does not reach statistical significance, most likely due to the substantial variability between individual samples. This is in line with the data reported by Wu et al., who showed a significant increase in *PGR-B* methylation in endometrial lesions and a trend for an increase in methylation in some but not all of the eutopic tissues (p=0.076)(Wu et al. 2006). It is worth noting that the samples in the Wu et al.

report were taken at various points in the menstrual cycle, with the majority being in the menstrual phase (Wu et al. 2006).

My data did not reveal a striking correlation between differentially methylated or hydroxymethylated loci and gene expression levels of the associated genes. This is not surprising as many other studies have observed a relatively poor (inverse) correlation between DNA methylation and gene expression (van Eijk et al. 2012; Spainhour et al. 2019). Taking the combined status of DNA methylation and hydroxymethylation into account is a much better predictor of transcriptional activity, a correlation I observed also in the endometriosis samples in this study (Mellen et al. 2012). A further interesting hypothesis is that the epigenetic differences in DNA methylation and hydroxymethylation may not impact on gene expression levels when the tissue is in a eutopic environment, but may predispose the cells to be more responsive to environmental cues, such as a different growth factor or extracellular matrix environment, when the tissue is in an ectopic location. Thereby, even subtle, preexisting epigenetic alterations may give cells an advantage in an ectopic location, making changes to their expression profiles more permissible, and ultimately resulting in enhanced cell adherence and proliferation leading to the establishment of endometriosis. The significant enrichment of the Hippo pathway that I observed is particularly intriguing in this context.

As we have established that a number of regions are differentially methylated and/or hydroxymethylated with regards to endometriosis along with the fact that there is no diagnostic biomarker for endometriosis to date, this opened up a new possibility. DNA methylation has been suggested as a biomarker for endometriosis. There are some DNA methylation test that are currently in clinical trials for cancer (Mikeska and Craig 2014; Nasu et al. 2011).

To determine if our data could lend itself to the identification of an epigenetic signature of endometriosis, we used a wide variety of approaches including overlapping our data with existing DNA methylation profiles, and with PGR and FOXO1 binding site profiles (Kukushkina et al. 2017; Mazur et al. 2015; Vasquez et al. 2015). In collaboration with CEGX, we verified a large fraction of the initial cohort of 411 genomic sites (the starting candidate set of putative ‘signature’ loci) using the TrueMethyl® On Target (TMOT) protocol. The validation rate for 5hmC is lower than 5mC, this is expected and most

likely due to technical noise. The technique itself has several sources of noise: the biochemistry is not 100% efficient (conversion of C to U) and as hydroxymethylation accounts for <1% of the genome, an inefficient conversion rate could result in high noise level. As oxBS uses an additional biochemistry step to BS this could result in differing degradation levels of the DNA (bisulphite treatment results in high levels of DNA degradation), which would consequently result in differing noise levels between the data sets. The bioinformatic technique to determine 5hmC involves the subtraction of these two datasets (with differing noise levels) and therefore this results in an estimation of the 5hmC levels. CpG sites that did not exhibit the expected (hydroxy)methylation state in any given probe set may be explained by the fact that the meDIP and hmeDIP techniques will pull down genomic DNA fragments that contain at least one modified CpG residue. Therefore, neighbouring CpGs may not necessarily exhibit the same modification state, but may seemingly appear as 'negative' validations. This highlights the importance of validating candidate regions by another method. TMOT has allowed us to establish 5mC and 5hmC modification profiles for each probe at base resolution and enabled the identification and validation of specific changes at each CpG interrogated.

TMOT analysis has resulted in a number of validated differentially methylated and/or hydroxymethylated CpG's within probes that can remarkably distinguish between endometrial receptivity as well as disease state. This is a promising result that could lead to a breakthrough in the diagnosis of endometriosis, as it is less invasive to acquire an endometrial biopsy than to undergo laparoscopic surgery. It has been suggested by another study that >500 samples are required to robustly detect methylation changes (Rahmioglu et al. 2017). Therefore, before over-stating its significance, it will be imperative to confirm the epigenetic signature biomarker set on a much larger cohort of endometrial biopsies to establish its robustness and validity as a potential stand-alone diagnostic tool.

# Chapter VI

# Discussion

## 6.1 Receptivity and its foremost importance

### 6.1.1 Expression profiles that define the immediate window of receptivity

The endometrium is a highly dynamic tissue that undergoes dramatic cyclic changes in response to steroid hormones. These involve proliferation, differentiation, then tissue degradation followed by regeneration if no pregnancy occurs.

The endometrium is vital for pregnancy as it is the tissue into which the embryo implants and supplies nutritional support to the fetus throughout pregnancy. It has been highlighted that the synchronised timing between the preparation of the endometrium and embryo maturation is essential for a successful implantation to occur (Valdes, Schutt, and Simon 2017). Indeed, a lack of synchronisation could be a contributing reason for infertility and certainly needs to be considered carefully during in vitro fertilisation (IVF) treatments. There is only a narrow time window when the endometrium is in a receptive state and this is indicated by the dramatic change in the expression profile that occurs rapidly over a short period of time. Diaz-Gimeno et al. developed the Endometrial Receptivity Array (ERA), which measures the expression of 134 genes from an endometrial biopsy and which can be used to determine the receptivity state (Díaz-Gimeno et al. 2011). The ERA is now available commercially at IGENOMIX and is used in the clinic in support of IVF treatment.

One of the noteworthy outcomes of my project was that I identified an Endometrial Receptivity Transcriptome (ERT) signature that can identify the receptive state within a narrow time window of LH+6 to +9 whereas the ERA compares LH+7 to +8 (receptive) to Day 8 to 12 (proliferative) and LH+1 to +5 (non-receptive). Considering the close timing of the biopsies I identified a transcriptional signature that shows a dramatic shift in the expression of multiple genes. Altmäe et al. carried out a meta-analysis of the limited previous expression analyses available and found that only 39 genes overlapped between these studies assessed (35 upregulated, 4 downregulated). What is encouraging is that these 39 genes highly overlap with my ERT; 20/35 (57%) of the upregulated and 3/4 (75%) of the downregulated genes (Altmäe et al. 2017). It has previously been suggested that histology itself is not a reliable form of endometrial stage identification due to its lack of reproducibility and consistency between evaluators (Gibson et al. 1991). Even in conjunction with measuring the hormonal timing of the LH surge this is still not sufficient to ensure correct staging of the endometrium, most likely

due to the natural variation of the menstrual cycle between women. Interestingly, of my LH+6 to +9 timed biopsies, 60-70% fell into the refined receptive state. This affected both the control and endometriosis groups equally but it may explain fertility issues experienced by all patients. This did result in the reduction in the number of samples per a condition i.e. controls and endometriosis are further subdivided into non-receptive and receptive. To further validate the ERT that was established, additional control samples would be required to reduce variability and enhance sensitivity.

This pilot study highlights the vital importance of accurately and correctly determining the receptive state not only in a clinical setting but also in a research setting. Many reports on the endometrium do not consider the general menstrual cycle phase, let alone the exact receptivity state. This lack of phase identification likely obscures many analyses well beyond the expression profile, resulting in concealing the dynamics of this ever-changing tissue. The ERT established in this study once further validation has been conducted will be instrumental for researchers who are interested in examining this significant time window for all types of analyses to ensure the correct comparisons and conclusions are being drawn.

### 6.1.2 Overview of epigenetics and the endometrium

The epigenome is a layer of communication that is imposed on the underlying DNA sequence and directs chromatin compaction, genome accessibility as well as gene activity. It is very complex and consists of multiple interconnected elements. These include covalent modifications to the DNA base cytosine (C) resulting in methylcytosine (5mC), hydroxymethylcytosine (5hmC), formylcytosine (5fC) or carboxycytosine (5caC). The most extensively studied cytosine modification is 5mC, which is commonly referred to as DNA methylation. The most notable function of DNA methylation is at heterochromatin where it acts as a repressor. It has also been associated with a repressive function at the promoters of specific genes, although globally this negative correlation is fairly poor (van Eijk et al. 2012; R. K. Ng et al. 2008; Spainhour et al. 2019). Histone modifications provide another layer of control through the modification of their N-terminal tails, which impacts the interactions between DNA and DNA binding proteins.

We and others have already established that there is an extensive change in the expression profile throughout the menstrual cycle and importantly during the transition from the non-receptive to the receptive state. Hormones provide an external signal for the endometrium to progress through the menstrual cycle, but how these changes are underpinned by the epigenome has not been fully established.

To investigate this in considerable detail, I carried out global profiling of H3K4me3, H3K9me3, H3K27me3, 5mC and 5hmC by immunoprecipitation followed by sequencing (ChIP-seq or (h)meDIP-seq or HMCP). This is undeniably the most extensive epigenetic profiling of the endometrium that has been carried out on the same tissue and it allows for direct cross-comparisons. To conduct the histone ChIP-seq I needed to optimise the protocol so that it was compatible with limited frozen material, and I subsequently confirmed that it resulted in adequate enrichment of the histone modifications of interest.

H3K9me3 is a repressive modification that often co-localises with 5mC and chromatin compaction. Knockdown of the H3K9 methyltransferases (SUV39H1 and SUV39H2) in fibroblasts increased their capacity to be reprogrammed to pluripotent stem cells, through their decreased capacity to repress pluripotency genes (Onder et al. 2012; Soufi, Donahue, and Zaret 2012). Total levels of H3K9me (mono, di and tri combined) have previously been shown to remain similar between the proliferative and secretory phase (Monteiro et al. 2014). I have confirmed that the global profile of H3K9me3 specifically is not significantly altered between the non-receptive and receptive state. This suggests that H3K9me3 is not majorly involved in the extensive transition that this tissue undergoes during the secretory phase. The dynamics and putative roles of H3K4me3 and H3K27me3 within the endometrium are evaluated in detail below (Section 6.1.3).

5mC has previously been studied with regards to the menstrual cycle in the endometrium. These studies have highlighted that a number of CpGs are differentially methylated throughout the menstrual cycle, with closer timepoints resulting in fewer differences (Houshdaran et al. 2014; Saare et al. 2016). The genes associated with these differences exert known functions in the menstrual phase they are expressed in (Dyson et al. 2014; Houshdaran et al. 2014; Kukushkina et al. 2017). To unambiguously distinguish 5mC from 5hmC and to assess DNA methylation and hydroxymethylation profiles genome-wide, I carried out meDIP- and hmeDIP -seq on all endometrial biopsies that formed part of this project. I identified multiple regions that were differentially

methylated or hydroxymethylated between the non-receptive and receptive state. Several of these sites overlapped with sites identified by Kukushkina et al. 2017. I included a number of these in the site-specific (TMOT) validation analysis along with other regions that I identified. This resulted in the demarcation of samples according to their receptivity state. This is further discussed in greater detail in Section 6.3. Although some of the differentially modified sites overlapped and were subsequently validated, overall there is little overlap between all the studies including my own. This could be due to several factors including the techniques involved, which either identified regions or individual CpGs as differentially modified target sites. Also, bisulphite sequencing does not allow the distinction between 5mC and 5hmC and may be a further reason for the discrepancy between this study and others. Bisulphite-based methylation analyses likely obscure both 5mC and 5hmC changes that may arise, and reduce the likelihood of overlap between studies that use differing techniques. Other factors include the exact timing of the samples. The sample numbers per phase in each study including this study are relatively small, and therefore the normal heterogeneity between women could obscure true changes by reducing the likelihood of any measured differences reaching significance. This brings into focus that 5mC differences do arise throughout the menstrual cycle. However, a systematic approach is required to allow for a more comprehensive overlap between studies. This would include a consistent and reliable method of menstrual cycle timing as well as an increase in the number of samples assessed per timepoint. It has been suggested that over 500 samples would be necessary to be confident in the 5mC changes that occur in the endometrium, indeed sample number is a limiting factor in all studies (Rahmioglu et al. 2017).

Throughout my thesis I have identified several significant observations that are of particular importance in the endometrium with regards to epigenetic modifications and I shall highlight and discuss them in the next sections.

### 6.1.3 The identification of bivalently marked genes in the endometrium

Histone modifications play a key role in coordinating the expression of genes by providing critical organisational instructions to the chromatin. The role of histone

modifications has been extensively studied in other cell types, specifically in embryonic stem cells (ESCs). H3K4me3 at TSSs is associated with open chromatin and expression of the corresponding genes. By contrast, enrichment of H3K27me3 at TSS has the opposite effect since, it is associated with chromatin condensation and transcriptional repression. Co-occupancy of the opposing H3K4me3 and H3K27me3 marks at TSSs was first established in ESCs and was shown to hold genes in a 'poised' state so that expression could be quickly up- or down-regulated in response to external or internal signals (Azuara et al. 2006; Bernstein et al. 2006; Roh et al. 2006; Young et al. 2011).

Histone modifications in general have not been in the focus of endometrial research and the few available studies yielded conflicting results. The most extensively studied histone modification is H3K27me3 (Colón-Caraballo, Monteiro, and Flores 2015; Grimaldi et al. 2016; Monteiro et al. 2014). Total levels of H3K27me3 remain unchanged between the proliferative and secretory phase, although when epithelial cells were specifically examined, a significant increase was observed in the receptive phase (Colón-Caraballo, Monteiro, and Flores 2015; Grimaldi et al. 2016). The immunofluorescence analysis of H3K27me3 ratio of epithelial-to-stromal cells in my samples agreed with the former studies, that the global levels remain unchanged between the non-receptive and receptive state. I did not observe an increase in epithelial H3K27me3 levels in the receptive and non-receptive samples within the narrow window of LH+6 to +9 that I assessed. However, H3K27me3 has been shown to differ in a locus-dependant manner in cultured cells upon decidualisation (Grimaldi et al. 2016; Tamura et al. 2014). Similarly, total H3K4me levels are not significantly changed between the proliferative and secretory phase (Monteiro et al. 2014). Contrary to these data, analysis of H3K4me3 epithelial-to-stromal cell ratios in my endometrial samples suggests that there may be an increase in H3K4me3 in stromal cells in the receptive state. This requires further investigation to quantify these differences by independent methods and to determine that it does indeed arise from stromal cells. In line with my results, cultured stromal cells have been shown to gain in H3K4me3 with decidualisation that correlated with gene up-regulation (Tamura et al. 2014).

To the best of my knowledge, ChIP-seq studies for specific histone modifications are not available for human endometrial tissue. This leaves a gaping hole in our understanding of what the epigenetic alterations involved in the orchestration of the extensive expression changes in the endometrium. Here, I report on the identification of a cohort of bivalently marked gene loci in the endometrium that seem to be of particular

importance for the establishment of the receptivity state. It is worth bearing in mind that number of samples used to identify these regions was limited, however several of the genes associated with these regions have previously been identified, as highlighted below, adding weight to these findings. With the addition of supplementary samples this would add further evidence and validate these findings. There is a substantial loss of bivalent regions between the non-receptive and the receptive state that transition to H3K4me3-only in the receptive state. As in other tissues, these bivalently marked genes are held 'poised' for activation in the non-receptive state, and the majority of the regions that lose their bivalency are 'switched on' in the receptive state, thereby driving the unidirectional development of the tissue.

These bivalently marked loci that resolve to H3K4me3-only in the receptive state are associated with genes involved in the preparation for a potential implantation. They fall into gene categories involved in extracellular exosomes, mitochondrion and metabolic pathways and are significantly enriched for KLF9 and FOXO1 binding motifs, all of which have previously been shown to be essential in the receptive state (Altmäe et al. 2017; Greening et al. 2016; Kajihara et al. 2006; Kajihara and Brosens 2013; Simmen et al. 2004; Tamura et al. 2014; D. Zhang et al. 2002). Extracellular exosomes have been shown to be crucial as they are released from epithelial glands and interact with the blastocyst to facilitate implantation (Greening et al. 2016). H3K4me3 was previously shown to increase at *FOXO1* along with other genes involved in the metabolic pathway in decidualised cultured stromal cells (Tamura et al. 2014). Of note are *SOD2* and *ELF3,* which are both bivalently marked in the non-receptive state and resolve to H3K4me3-only in the receptive state, and contain FOXO1 and/or KLK9 binding sites. They have both been identified to be up-regulated in the receptive state in this study as well as in previous studies (Bhagwat et al. 2013; Chan et al. 2013; Díaz-Gimeno et al. 2011; Talbi et al. 2006). ELF3 is a specific epithelial transcription factor that promotes epithelial-to-mesenchymal transition, which is a process that occurs to some extent in the endometrium and is important in the in preparation for implantation (Whitby, Salamonsen, and Evans 2018; Zheng et al. 2018). SOD2 is a mitochondrial gene that is down-regulated in women, who fail to achieve pregnancy (Tapia et al. 2008). It is also an antioxidant defence enzyme that is upregulated by FOXO1 during the decidualisation of cultured stromal cells as part of the resistance to oxidative cell death (Kajihara et al. 2006).

KLF9 and FOXO1 are transcription factors that are positively regulated by progesterone, which increases in the secretory phase. Therefore, progesterone hormone signalling increases KLF9 and FOXO1 levels and could thereby lead to an enhanced binding of these transcription factors to bivalently marked genes, leading to a switch from them being 'poised' to being 'switched on'.

A key aspect of bivalent domains in ESCs is their heightened link to cellular potency and plasticity (Azuara et al. 2006; Bernstein et al. 2006). The presence and resolution of bivalent domains in the endometrium is a good molecular read out for the endometrium being a plastic tissue, perhaps reflecting its cyclic nature of monthly regeneration. Contributing to this developmental plasticity are the endometrial stem/progenitor cells. Several studies have identified endometrial stem cells, specifically endometrial mesenchymal stem cells and more recently a stem cell-like behaviour of epithelial cells, when cultured as endometrial gland organoids. These glandular organoid cells are capable of self-renewal, proliferation and remain hormone responsive (Gargett, Schwab, and Deane 2015; Masuda et al. 2012; Turco et al. 2017). Mesenchymal stem cells are found in perivascular locations and are multipotent, giving rise to several cell types including endothelial, myogenic and fibroblast cells (Gargett, Schwab, and Deane 2015; Masuda et al. 2012). Endometrial gland organoids can be grown long-term and can recapitulate the molecular characteristics of their in vivo precursor cells and can undergo differentiation into both luminal and glandular epithelium (Turco et al. 2017). This suggests that the endometrium is a developmentally plastic tissue and that bivalent domains may play role in the tissue's responsiveness and development.

The mid-secretory phase is part way through the menstrual cycle and the endometrium has not yet reached its determined end point, which is either to respond to a pregnancy or to start shedding during menstruation, and the bivalency of histone marks reflects this. I hypothesise that multiple genes are held in a 'poised' position throughout the menstrual cycle and that this epigenetic mode of endowing genes with the capacity to respond rapidly and co-ordinately is particularly important as the endometrium develops in response to differing hormonal signals, resulting in a cascade of genes 'switching on'. To test this hypothesis and to further understand the extent to which bivalency contributes to the changes in gene expression, it is important to track the bivalency and expression changes across the entire menstrual cycle.

In general, I believe that my findings of the fine-tuned epigenetic gene regulation are important to better understand the orchestration and tight regulation of the expression profile during the menstrual cycle.

### 6.1.4 The most dramatic changes occur in glandular and luminal epithelial cells

Epithelial cells undergo structural and biochemical changes during the secretory phase that are key for implantation to occur. In response to hormones, the epithelial cells become less adherent to each other to allow a blastocyst to penetrate the barrier and embed, whilst at the same time becoming adhesive to the blastocyst itself to assist this process (Whitby, Salamonsen, and Evans 2018). Concomitantly, epithelial cells secrete a number of products that contribute to the uterine luminal fluid. This could be further investigated by carrying out uterine flushing's followed by mass spectrometry when the endometrium is in a receptive state. This fluid is suggested to be in a 'dialogue' with the blastocyst and facilitates a successful implantation. To orchestrate these changes a number of processes are in action and work harmoniously together. I show that the transcriptional profile of epithelial cells undergoes a dramatic shift during this narrow time window between the non-receptive and receptive state. Importantly, epithelial cells indeed contribute to the majority of transcriptional changes that arise in the whole tissue, highlighting their essential role.

The analysis of the epithelial cells underlines the need to combine not only the precise cycle state of the tissue, but the cell type and tissue composition as well. In many other tissues the expression profile of the various cell types of which the tissue is composed remain fairly static, and therefore tissue composition is not much of a concern when comparing different specimens. However, in case of the endometrium, expression profiles change dramatically and differentially in the different cell types during the menstrual cycle, making the availability of cell type-specific signatures across the cycle of particular importance.

My epithelial and stromal cell gene expression signature sets are highly instructive for researchers to take into consideration the proportion of expression contributed by these two cell types within whole tissue. To extend this work it will be interesting to establish a luminal and glandular epithelial cell signature separately, which have been shown

previously to have differing expression profiles in the secretory phase (G. E. Evans et al. 2014). Either this could be achieved through laser capture microdissection or through the use of flow cytometry to separate out the cell type followed by sequencing. This could be expanded in the future beyond the secretory phase to the entire menstrual cycle so that an applicable cell signature could be available.

### 6.1.5 Epithelial cells undergo dramatic changes in hydroxymethylation levels

One of the most remarkable discoveries was the dramatic fluctuation of 5hmC between the non-receptive and receptive state in epithelial cells. This was initially identified in the immunofluorescence stainings, which highlighted a significant increase in the epithelial-to-stromal cell ratio of 5hmC signal in the transition to a receptive state. In these experiments, the calculation of the staining intensity ratios was introduced to obtain fully unbiased data. Therefore, while staining intensities may vary between individual experiments and even between different tissue sections within the same experiment, the comparative staining intensity of two cell types that exist in immediate proximity within the tissue is a reliable read-out of any changes that may occur. Therefore, I can confidently say that the glandular epithelium was the cell type that consistently exhibited higher levels of 5hmC in samples that were in the receptive state. This was additionally corroborated by the liquid chromatography-mass spectrometry data on laser captured microdissected epithelial cells, which showed an increase in 5hmC levels in the receptive state. With additional samples to be analysed in the future, this difference is likely to reach significance.

To further investigate the changes in 5hmC levels and distribution specifically in endometrial epithelial cells, laser capture microdissection (LCM) on the epithelial cells followed by a new sequencing technique was carried out at Cambridge Epigenetix: Hydroxymethylation Capture Pulldown (HMCP). The fact that I obtained good enrichment values and genome coverage in these experiments is a great achievement considering that only 13ng of sonicated DNA from paraffin embedded LCM epithelial cells was used per HMCP experiment.

This shift in 5hmC positively correlates with the increase in expression of genes that are upregulated in the receptive state, suggesting that 5hmC plays a significant role in the

regulation of gene expression at this stage in the menstrual cycle. 5hmC is also prominently enriched at genes that are bivalently marked and resolve to H3K4me3-only in the receptive state. This has been shown in ESCs previously, and here I show that 5hmC shows similar dynamics at genes marked in this way in epithelial cells (Ficz et al. 2011; Pastor et al. 2011). Of note, bivalently marked genes that remain bivalent do not show this increase in 5hmC upstream of the TSSs. This suggests that the addition of 5hmC to bivalently marked genes may be required prior to their resolution to H3K4me3-only and concomitant transcriptional up-regulation. Perhaps the bivalently marked genes that remain bivalent in the receptive state require the addition of 5hmC before they are 'switched on' at a later stage (Figure 6.1). To be able to fully address this point, an extensive study across the menstrual cycle assessing a combination of gene expression, 5hmC, H3K4me3 and H3K27me3 profiles would be required. In any case, the ability to predict the subset of bivalently marked genes that will become activated by their 5hmC state is a highly interesting finding that adds 5hmC as a key epigenetic marker to the bivalency concept.

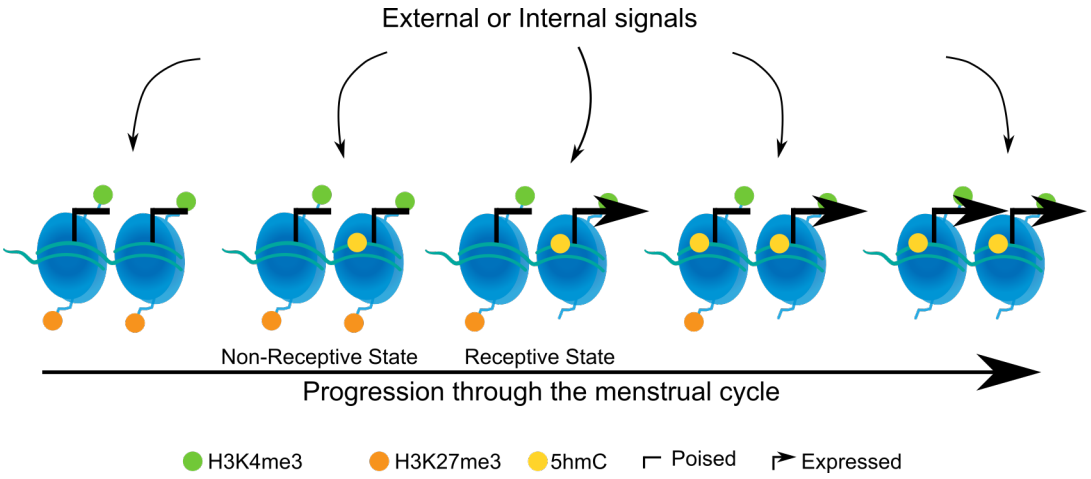



**Figure 6.1 Hypothetical diagram of the role of bivalency and 5hmC in the Endometrium.** As the endometrium progresses through the menstrual cycle it receives multiple external and internal signals. These signals may result in the addition of 5hmC at bivalent regions which in turn demarcates these loci towards resolution to H3K4me3-only and the induction of gene expression. This process could continue as the endometrium receives additional signals resulting in remaining bivalent regions gaining 5hmC and resolving to H3K4me3. Since endometrial epithelial cells exhibit the most dramatic dynamics in 5hmC levels, this regulation may specifically pertain to this cell type.

Overall, this highlights the central role that 5hmC plays in the regulation of gene expression in epithelial cells and their transition to the receptive state. It also supports 5hmC is a functional epigenetic modification in its own right and not just a turnover product of 5mC.

The potential mechanisms that lead to the 5hmC dynamics in the endometrium are currently unknown. TETs, which oxidise 5mC to 5hmC, remain unchanged at the mRNA level during the time window from a non-receptive to a receptive state in my data. It remains a possibility that protein levels of the TETs change, or that their stability or catalytic activity is affected by post-translational modifications, which will require further investigation. Similarly, DNMT mRNA expression profiles did not exhibit any major changes, with the same remaining possibilities of their potential regulation downstream of transcription that would require future investigations (Estève et al. 2009; Goyal et al. 2007; Ling et al. 2004).

In the absence of any notable changes in these 'classical' pathways of the DNA methylation machinery, I widened my search and noted the enzyme NNMT as one of the most differentially regulated genes between the receptivity states. NNMT provides another potential hypothesis for introducing 5mC/5hmC changes as it may act as a modulator of DNA methylation dynamics. NNMT is a methyltransferase, which is dramatically upregulated in the receptive state. It has been reported that NNMT acts as a methyl sink by binding to the main methyl donor, SAM, so that it is unavailable for other methyltransferases (Ulanovskaya, Zuhl, and Cravatt 2013). I hypothesised that NNMT may act as a methyl sink in the receptive state of the endometrium, thereby reducing the availability of methyl groups for DNMTs to deposit and maintain DNA methylation. In the unchanged presence of TETs, 5mC would continue to be converted into 5hmC, resulting in a net increase in 5hmC levels.

I pursued to test this hypothesis in cultured epithelial cells by over expressing NNMT, however, this did not affect 5mC or 5hmC levels. The role of NNMT is still in need of further analysis due to its potential role in increasing the histone H3 modifications H3K4me3 and H3K27me3, perhaps through the dynamics of NAD+ as discussed in Chapter IV. It would be informative to investigate NNMT in a different model system as this cell line is from a cancer patient and therefore may not be a true representation. For example, the newly developed epithelial organoids may better recapitulate the *in vivo* behaviour of these cells and retain the hormone responsiveness of these cells would be

an avenue to facilitate our understanding of the role of NNMTs in endometrial epithelial cells (Turco et al. 2017).

Another potential hypothesis for the increase of 5hmC in epithelial cells in the receptive state is the changing metabolic environment. Specifically, TET enzymes depend on α-ketoglutarate and $Fe^{2+}$ as co-factors, the former a product of the Krebs cycle and the latter requiring a reducing environment. The secretory phase is associated with an increase in secretions from glandular epithelium that contribute to the uterine fluid. The production of the secretions is due to in an increase in metabolic processes and activity. As a consequence of these biochemical reactions there is an increase in reactive oxygen species that are counteracted in the tissue by a concomitant up-regulation of antioxidants, resulting in a reducing environment. For example, I and others have shown that superoxide dismutase 2 (SOD2), which is an antioxidant defence enzyme, is highly upregulated in the receptive state, most likely to counteract oxidative stress (Matsuoka et al. 2010; Talbi et al. 2006). A reducing environment would in turn maintain the levels of $Fe^{2+}$ which is required for the activity of TETs, while the availability of oxygen has been shown not to be rate-limiting for TET activity (Laukka et al. 2016). Vitamin C, which is a reducing agent, has been shown to increase 5hmC levels by increasing the levels of $Fe^{2+}$ in embryonic fibroblast (Minor et al. 2013). It is possible that a similar scenario is occurring during the receptive phase; although the expression levels of TETs remain unchanged their enzymatic activity is higher due to the availability of $Fe^{2+}$, which would lead to increased levels of 5hmC. This hypothesis would need further testing in the future by assessing the metabolome of epithelial glands comparatively in the non-receptive and receptive states.

Epithelial cells are essential for a successful implantation and require a highly controlled expression profile to orchestrate this synchronised process. I have established that 5hmC and histone modifications, in particular those establishing the so-called bivalent state, play an essential role in this process.

## 6.2 The dynamic alterations of the epigenetic profile in endometriosis

### 6.2.1 Transcriptional differences associated with endometriosis

Endometriosis is a debilitating condition that affects 10% of women of reproductive age (Nnoaham et al. 2011). Understanding the condition in as much detail as possible will be invaluable in progressing the diagnosis and treatment options that are currently limited. Endometriosis is defined by the presence of endometrial tissue located ectopically where it forms adhesions, this can lead to infertility, chronic pelvic pain, and dysmenorrhea among other symptoms.

There have been a number of studies that have analysed the differences that arise in the ectopic lesions themselves. This is invaluable research and has added great understanding to the field. It has also highlighted the large degree of variation that occurs and the difficulty that this imposes when studying ectopic lesions. To delve deeper into the potential origins of endometriosis, I was interested in examining the potential differences that might exist between the eutopic endometrium from women with endometriosis compared to the endometrium from controls, and that may make this tissue more susceptible to attachment or survival outside the womb.

As established in Chapter III, it is key in all analyses to establish the correct menstrual phase in which the biopsy is taken. The LH-surge and/or histology alone has been shown to be insufficient to correctly identify the menstrual phase. The number of samples in this study is comparable to similar studies into endometriosis, however previous studies often do not take the menstrual phase into account, let alone establish the correct stage by expression profiling. Irrespective of whether eutopic or ectopic tissue has been studied, this could potentially hinder many of the analyses conducted as both tissues are responsive to hormones and, therefore, are constantly transitioning.

The ERT that I established will be of great use to many researchers that are studying the important transitioning point between the non-receptive and receptive state. I established the receptivity state of the endometriosis samples to be sure that any expression differences that arise are indeed due to endometriosis rather than due to timing within the menstrual cycle. As a matter of fact, I found no differentially expressed genes between receptive controls (RC) and receptive endometriosis (RE) samples. However, eleven genes were differentially expressed between non-receptive controls (NRC) and non-receptive endometriosis (NRE) biopsies, identified by two well-

established independent types of transcriptomic analysis using the EdgeR and DESeq2 tests. *TNC*, *PLA2G4E* and *THEM4* were confirmed by previous studies to be dysregulated in endometriosis, corroborating my data. I also highlighted a number of additional genes that were dysregulated at this stage with some of them already been associated with the menstrual cycle.

Overall, this agrees with previous reports that the expression profile in the endometrium is disrupted with the largest disruption occurring during the transition from the proliferative to the secretory phase (Afshar et al. 2013; Flores et al. 2007; Irungu et al. 2019). This also aligns with the established finding that the endometrium of endometriosis patients often exhibits an attenuated response to progesterone, i.e. the hormone is highly expressed and key to the regulation of the endometrium during the secretory phase (Aghajanova et al. 2011; Burney et al. 2007; Wei et al. 2010). Due to the close alignment of endometrial cycle stages that I was able to make through the expression data, I can be confident that the transcriptional changes are indeed related to the disease state, and not a discrepancy in cycle stages. There has been a comprehensive study that has looked at expression changes across the menstrual cycle (mainly late proliferative to mid secretory phase) in controls and endometriosis patients and mapped expression quantitative trait loci (eQTLs) (Fung et al. 2018). This study highlights the importance of considering the menstrual cycle when establishing expression changes, they found expression differences associated with endometriosis, however these did not pass multiple testing corrections. The *HOXA10* gene, however was significantly variable in cases compared to controls suggesting that expression maybe less well orchestrated. Two eQTLs at genes (*VEZT* and *LINC00339*) that have previously been associated with endometriosis where identified, suggesting a genetic effect on the pathology of endometriosis (Fung et al. 2018). However, this previous data do highlight the fact that the endometrium of endometriosis patients may frequently be delayed due to its attenuated response to progesterone. The delineation of expression difference arising due to the timing of the endometrium or due to the endometrium being delayed in patients with endometriosis is difficult to determine and requires further investigation by additional samples and accurate determination of precise stage.

There have been recommendations in the field to reduce the high levels of heterogeneity with regards to endometrial biopsies by the World Endometriosis Research Foundation (WERF), who have developed standardised protocols to collect, process and store samples (Fassbender et al. 2014). This may increase the availability of control samples

for studies allowing for human heterogeneity and decreasing the overall variability within studies. This should decrease the variability that arises between studies allowing for cross- comparisons between studies as well as collaboration between facilities.

This study has highlighted the vital importance of correctly establishing the stage of the endometrial biopsy so that it does not obscure changes associated with endometriosis. When the menstrual cycle is considered it is evident that there a number of differentially regulated genes within the non-receptive state. To further tease apart the effects of the menstrual cycle and endometriosis on the endometrium, a further comprehensive analysis of a larger sample size spanning the menstrual cycle as a whole with precise timing is required, as has also been suggested by Rahmioglu et al. (Rahmioglu et al. 2017).

### 6.2.2 Endometriosis is associated with altered epigenetic profiles in the endometrium

The most recognised theory for the pathogenesis of endometriosis is that put forward by Sampson, which involves retrograde menstruation (Sampson 1927). This scenario provides an explanation of how endometrial cells end up outside the uterus through reflux during menstruation. However, retrograde menstruation occurs at a higher rate than endometriosis, and therefore, it has been suggested that additional conditions in endometriosis patients arise that favour the attachment and proliferation of endometrial cells. There have been recent reports on the role that epigenetics may play in the aetiology of endometriosis by priming the endometrium for its attachment and proliferation in an ectopic environment. The World Endometriosis Research Foundation (WERF) updated its recommendations in 2017 to include the following: “Studies should be undertaken on all aspects of epigenetic regulation of endometriosis.”, which emphasises the need for epigenetic research (Rogers et al. 2017).

I carried out the first comprehensive epigenetic profiling on endometrium from patients with endometriosis and controls in a pilot study. This included the histone modifications H3K4me3, H3K9me3 and H3K27me3, and the DNA modifications 5mC and 5hmC, in all cases using immunoprecipitation followed by sequencing. All of these experiments were carried out on the same biopsies along with RNA-sequencing so that direct comparisons can be made between datasets. These data provide unique insights into the epigenetic

alterations that may be associated with eutopic endometrium in patients with endometriosis, as well as the role of the epigenome during the establishment of the receptive state. Although the number of samples used in this study are comparable to previous studies it would have been an advantage to have additional samples to reduce variability and increase statistical significance.

The majority of reports on histone modifications only measured the total levels within the tissue and did not necessarily discriminate between mono, di or tri methylation at H3 lysine residues. They show that overall levels of H3K4me, H3K27me(3) and EZH2 (methyltransferase of H3K27me3) are higher in eutopic tissue from women with endometriosis compared to eutopic controls, and these differences usually increased in the ectopic lesions (Monteiro et al. 2014; Zhang et al. 2017). I identified that the eutopic endometrium histone profiles of H3K4me3, H3K9me3 and to a large extent H3K27me3 remain unaffected by endometriosis. Therefore, even if global levels of some of these modifications may change, their genome-wide distribution remained largely unaltered (with a few exceptions for H3K27me3), at least as far as I can tell from the set of pilot samples. It was important that the precise receptivity state was considered to make sure that differences that arose were due to the disease state rather than discrepancies in receptivity state. I identified seven regions that were significantly different in H3K27me3 between controls and endometriosis in the receptive state, four enriched and three depleted in endometriosis. Interestingly, the genes associated with these regions have not been identified before in the context of endometriosis. In particular TMEM176A/B, has been shown to repress inflammation and as a consequences its transcriptional repression could lead to enhanced inflammation, which is a characteristic of endometriosis (Bruner-Tran et al. 2013; Segovia et al. 2019). This is a noteworthy finding that requires further validation by additional samples and mechanistic analysis to understand the functional consequences of altered H3K27me3 enrichment at these loci.

### 6.2.3 Changes in covalent DNA modifications often affect Hippo pathway components

This study further shows that 5mC and 5hmC is disrupted at a global and locus-specific level in the eutopic tissue of women with endometriosis. These differentially methylated and/or hydroxymethylated regions are associated with genes that have been shown to be involved in endometriosis, such as cell adhesion and cell junction. Another pathway that stood out was the extended Hippo signalling cascade, that was enriched in both the differentially methylated and hydroxymethylated probe sets. The Hippo pathway was initially identified as essential for organ size control, and its dysregulation contributes to tumorigenesis. It is involved in many cellular aspects that are suspected to play a role in the establishment of endometriosis, in particular the balance between proliferation and apoptosis as well as cell adhesion (Figure 6.2, Yu and Guan 2013). Indeed, the Hippo pathway has been previously suggested as involved in the pathogenesis of endometriosis in an analysis of endometrial stromal cells (Y. Song et al. 2016). YAP1 and TAZ (WWTR1) are transcriptional co-activators and core components of the Hippo signalling pathway. YAP1 knockdown in endometrial stromal cells resulted in reduced proliferation and an increase in apoptosis. Mice treated with verteporfin, which disrupts the YAP1-TEAD complex, resulted in a significant reduction in endometriotic lesion size compared to controls (Y. Song et al. 2016). The over-expression of TAZ in Ishikawa cells resulted in an increase in migration and wound closure (Romero-Pérez et al. 2015). These data would suggest that the Hippo pathway is over-stimulated in endometriosis patients, giving cells an advantage in adhering and proliferating in an ectopic location. The Hippo pathway has also been shown to be involved in endometrial cancers, specifically aggressive and high-grade types, which have higher levels of TAZ and it is located within the nucleus of the cell (Romero-Pérez et al. 2015).

This study adds additional weight to this overall argument by identifying differential 5mC and 5hmC marks at multiple genes that are involved in the Hippo pathway. Although these changes did not correlate with any consistent alterations on the expression of the Hippo genes, there could still be an effect at the protein level or on the subcellular localisation, which is a key regulatory mechanisms of Hippo pathway activity (Romero-Pérez et al. 2015; Strakova et al. 2010; Yu and Guan 2013). In any case, this study shows that this pathway is a target of increased epigenetic changes in

endometriosis patients, which may ultimately lead to an altered activity of the pathway, even if perhaps by different routes in different patients. The involvement of the Hippo pathway could be investigated in primary stromal cell cultures and epithelial organoids to assess the proliferation, apoptosis as well as migration in knock-out and over expression experiments.

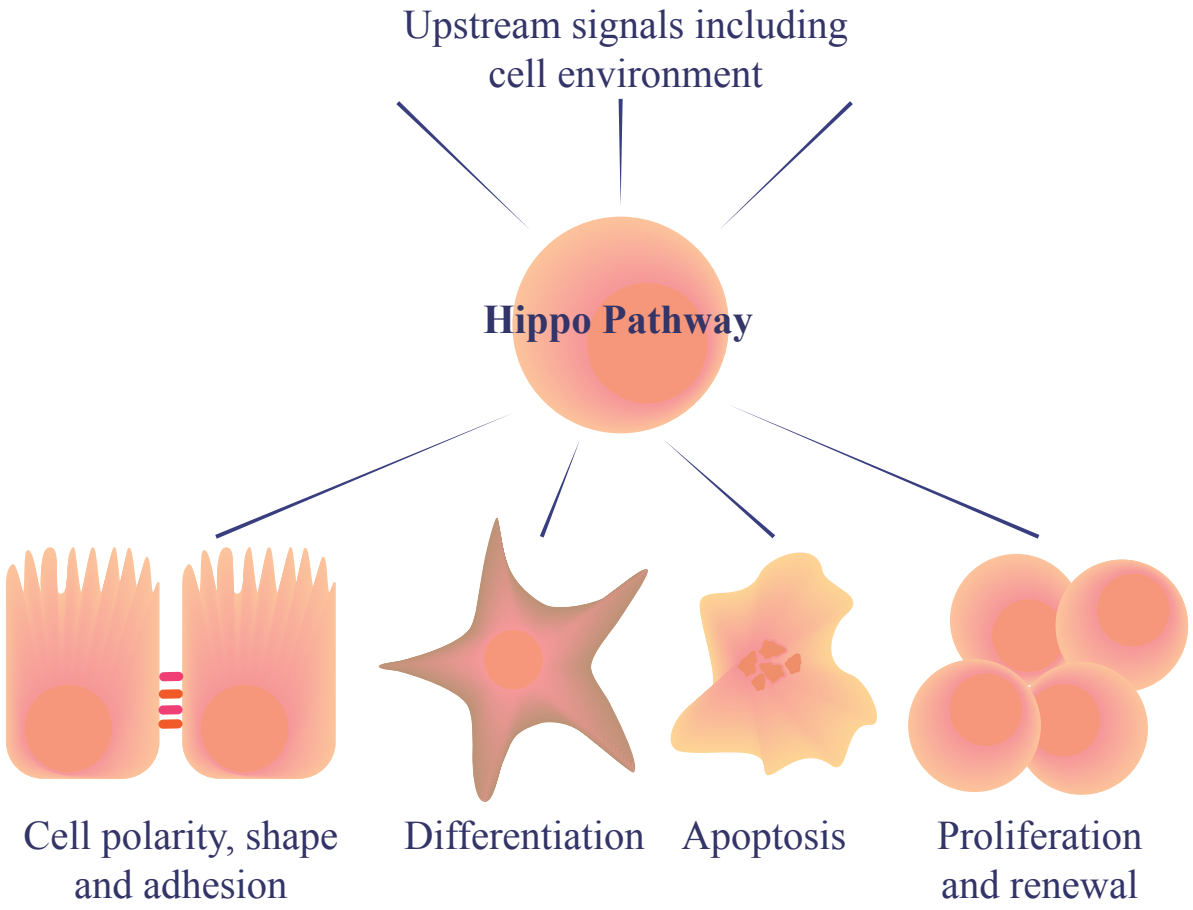



**Figure 6.2 Hippo signalling pathway.** The Hippo signalling pathway affects several cellular functions including, cell polarity, shape, adhesion, differentiation, apoptosis, proliferation and renewal.

In general, even beyond the Hippo pathway there is little correlation between differentially 5mC- and 5hmC-modified regions and gene expression changes in the datasets. This is not unexpected as other studies have shown a poor negative and sometimes positive correlation between DNA methylation and gene expression (van Eijk et al. 2012; Spainhour et al. 2019). The majority of genome-wide methylation studies on endometriosis show relatively few correlations with expression, and this occurred in both a positive and negative direction (Borghese et al. 2010; Dyson et al. 2014;

Houshdaran et al. 2014). Locus-specific and genome-wide methylation reports often do not asses the menstrual cycle stage or if it is considered, the methods used are often inaccurate. Therefore, correlations that may have been drawn between methylation and expression changes could be due to the variation in menstrual cycle stage. These data emphasise that endometriosis cannot be assessed as an isolated condition but that the menstrual cycle is pivotal to the current status of eutopic endometrium.

I hypothesise that the differential 5mC and 5hmC in the endometrium, in particular at Hippo signalling genes, may be kept in check by additional gene regulatory controls, such as histone modifications. However, when the tissue ends up in an ectopic location by retrograde menstruation, the alterations in 5mC and 5hmC may predispose the affected genes for dysregulation triggered by the differing environmental cues, for example extracellular matrix environment or growth factors. Ultimately, these pre-existing epigenetic alterations may give these cells an advantage and enhance their ability to adhere and proliferate leading to the formation of lesions.

## 6.3 5mC and 5hmC biomarker cohorts for endometriosis and receptivity

One of the major hurdles within the field of endometriosis is the length of time it takes for a positive diagnosis, which currently is on average 6.7 years (Nnoaham et al. 2011). The delay in diagnosis not only brings with it a substantial economic and mental burden, but endometriosis is also often a progressive disease and is usually more advanced with a delayed diagnosis. The development for a biomarker for the detection of endometriosis is one of the recommendations that have been put forward by the WERF (Rogers et al. 2017). It is unlikely that a single biomarker for either disease or receptivity state will be found due to the complexity and variability between individuals. Therefore, to progress this field a combinatorial set of biomarkers is likely required.

DNA methylation (5mC) biomarkers are being developed in other fields and have shown promising results. In disease states such as cancer, 5mC biomarkers are being investigated for diagnostic use of the disease as well as for predicting clinical outcomes. 5mC is often dysregulated in cancer and has been suggested to be involved with the disruption or dysfunction of expression which could lead to the formation of cancer. Indeed, there are several biomarkers that are in clinical trials, some of which are

commercially available for colorectal, lung, prostate and glioblastoma cancer (Mikeska and Craig 2014). DNA methylation biomarkers have also been developed for predicting the chronological age of mice and humans. The biomarkers consist of a number of CpGs where the likelihood of their methylation differs throughout life and can be used to predict the age of the individual (Horvath 2015; Stubbs et al. 2017).

In order to potentially develop an epigenetic biomarker set for endometriosis, I carried out a pilot study to establish the 5mC and 5hmC genomic profile in the endometrium of women with endometriosis and controls. This is the first time that 5hmC has been analysed in endometriosis (and in the endometrium in general), and I have identified that 5hmC likely plays a significant role within the tissue. It is an important first step for the field to have information on both inter-dependent marks available to better define the receptivity state and endometrial disease conditions.

Multiple DNA methylation profiles of endometrium and endometriotic lesions have previously been reported, however, the majority of studies used methylation arrays, e.g. the HumanMethylation450 and HumanMethylation27 arrays, which cover <450,000 and >27,000 CpGs, respectively, out of the ~28million CpGs in the human genome. Significantly, bisulphite (BS) treatment needs to be performed prior to hybridizing the DNA to the array. BS treatment is very harsh resulting in many breaks in the DNA leading to its degradation. The sodium metabisulphite used for the treatment converts unmodified cytosines as well as 5fC and 5caC to uracil, which is subsequently read as a thymine in the following PCR amplification rounds and sequencing. However, BS treatment does not discriminate between 5mC and 5hmC resulting in the same sequencing readout for both modifications, namely cytosine. As the majority of reports so far have used BS treatment-based methods for the establishment of DNA methylation profiles, they cannot distinguish between 5mC or 5hmC. This could result in the obscuring of differences that may have occurred but are disguised due to the inability to distinguish 5mC and 5hmC. Several studies have used meDIP for their analysis of 5mC which would circumvent this issue with BS treatment. However, these studies had several drawbacks, the meDIP analysis was either meDIP-on-Chip (Affymetrix GeneChip Human Promoter 1.0R Array), or meDIP-qPCR, conducted on cultured cells and / or compared differing types of endometriotic lesions rather than eutopic tissue (Borghese et al. 2010; Lucas et al. 2016; Moreno-Moya et al. 2014). This may also be a potential reason why some loci have been identified in multiple previous studies as differentially methylated, notably *PGR* and *HOXA10*, but these genes did not stand out in my analysis.

It is possible that they are differentially modified by a mix of 5mC and 5hmC between control and endometriosis samples, and therefore do not pass statistical threshold levels in my datasets. Alternatively, this discrepancy to previous studies may arise from the variability between different patient samples, endometriosis model system or location and timing of biopsies. It has indeed been noted that an estimated >500 samples would need to be investigated to gain solid insights into consistent 5mC changes in endometriosis that hold up on a population-wide scale (Rahmioglu et al. 2017).

There are newer methods, notably the so-called oxidative bisulphite (oxBS) treatment, that includes an additional oxidation step in a parallel sample to help distinguish 5mC from 5hmC. This treatment first oxidizes 5hmC further to 5fC, which is followed by the conventional BS treatment, resulting in only the original 5mC being read as cytosine, giving a 'true' methylation reading (Booth et al. 2012; Lunnon et al. 2016; Raiber et al. 2017). By bioinformatically subtracting the oxBS from the BS sequencing this will give the 5hmC read out.

I have used 5mC and 5hmC immunoprecipitation-based sequencing, which uses antibodies against the corresponding modification allowing for the separation of 5mC and 5hmC. This method identifies regions of DNA that contain either 5mC or 5hmC, but does not resolve these modifications at base resolution. Therefore, I validated a number of regions at a CpG base resolution level using the newly developed methodology TrueMethyl® On Target (TMOT) at Cambridge Epigenetix. This method does use BS but I also implemented oxBS treatment in parallel, which now allows us to distinguish between 5mC and 5hmC.

I employed TMOT to validate differentially modified candidate regions that were able to separate the samples according to their receptivity state as well as according to their disease state. TMOT uses oligonucleotides that have been designed specifically for the regions of interest, resulting in less sequencing required to gain a high read depth across these regions. TMOT validated a number of differentially methylated and hydroxymethylated regions at base resolution level. The validated CpGs confirmed the meDIP and hmeDIP analysis and separated out the eutopic endometrium according to receptivity and disease state, as can be rather impressively seen in the PCA plots. This result holds great promise for future research and with further development and refinement it could potentially be used as a diagnostic test.

This 5mC- and 5hmC-based biomarker set could be the basis for an answer to the conundrum of a diagnostic test for endometriosis. Not only could it be used as a diagnostic tool for endometriosis, it could also be used in an infertility setting. It has been raised that assisted reproductive technologies (ART) still has relatively low success rates for pregnancy, some of which can be attributed to the endometrium being in an inadequate receptivity state (Diedrich et al. 2007). In particular women who are infertile have a much higher prevalence of endometriosis with about 47% being affected (Meuleman et al. 2009). Therefore, this 5mC and 5hmC biomarker set could in the future be used within the infertility setting to tackle these two vital features; the determination of receptivity state and endometriosis.

To further develop this epigenetic biomarker set, it would require testing on additional samples from patients with a variety of conditions and over the menstrual cycle to identify the sensitivity, specificity and robustness of the identified loci across a much larger population cohort. Changes in 5mC have been associated with ageing and therefore this biomarker could potentially be influenced by the age of the patient. This is an important aspect that needs to be considered and accounted for, with adjustments to the genomic probe set as needed. Further development of this could also involve establishing if the epigenetic biomarker set can be used directly with menstrual blood. This would be another interesting prospect as it would negate the need for biopsies, and is material that is readily available each month.

## 6.4 Conclusions and implications

Understanding the mechanisms that are involved in the establishment of the receptive state is an important step forward in reproductive biology. The receptive state within the secretory phase of the menstrual cycle is the most vital for a successful pregnancy to occur. Therefore, it is essential to fully unravel the mechanisms that are involved in the establishment and control of this state if we are to understand what is required for implantation to occur. I have identified an Endometrial Receptivity Transcriptome (ERT) that can be used to distinguish biopsies at the key tipping point of receptivity between being non-receptive to attaining the receptive state. This is essential for all analyses that are carried out when investigating the receptive state, as it has been highlighted by other studies that histology and LH surge are insufficient and inadequate in

distinguishing the correct state (Díaz-Gimeno et al. 2011; Gibson et al. 1991; Talbi et al. 2006). Assigning the incorrect state to biopsies will lead to the obscuring of results or the potential drawing of incorrect conclusions.

In this thesis, I have identified the role that bivalency and 5hmC plays in the establishment of the receptive expression profile. I identified regions that are bivalent in the non-receptive state which transitioned to H3K4me3-only in the receptive state, corresponding to an up-regulation in the expression of the associated genes. I also demonstrated a sudden increase in 5hmC in epithelial cells which correlates with an increase in expression of associated genes. This has deepened our understanding of the epigenetic changes that occur, and likely are required, to achieve a receptive state. These insights warrant further investigation to expand our knowledge into the functionality of these modifications and how precisely they modulate and fine-tune gene expression patterns to help establish the correct state and timing of receptivity.

Endometriosis is a debilitating condition that occurs in ~10% of women of reproductive age (Nnoaham et al. 2011). The personal and societal burden is immense, which is exacerbated by the long delay in diagnosis as well as the limited knowledge of its pathogenesis. In this work I have shown that the 5mC and 5hmC profiles are altered in the endometrium of women with endometriosis, in particular at genes involved in the Hippo pathway. These alterations may potentially predispose these cells to attach and form lesions outside the womb leading to endometriosis in these women. I also showed that the differential enrichment of 5mC and 5hmC at a cohort of genomic sites could be used as a diagnostic test for endometriosis as well as for identifying the receptivity state. With further comprehensive mechanistic and validation studies this could progress the understanding of the origins of endometriosis as well as provide a much-needed faster diagnostic route.

In conclusion, the in-depth analysis of the endometrium has highlighted important epigenetic alterations that occur in both around the window of receptivity and disease state of endometriosis. I hope this promotes further investigation and exploration into the role that epigenetics plays within the endometrium, as this could lead to an increased understanding with regards to the orchestration of the endometrium during the receptive window and the pathogenesis of endometriosis.

## 6.5 Future directions

In this thesis I have identified a number of cell type- and receptivity state-defining signatures that emphasise the importance of assigning the precise state to the endometrium to ensure correct analysis. This prompts the question of how the endometrial signatures change not only during the implantation window but across the whole menstrual cycle. To address this question as well as to validate and expand on these signature gene sets, single cell expression analysis could be carried out on samples taken from the entire course of the menstrual cycle. This could lead to the generation of cell type- and menstrual cycle phase-specific signatures that could be used to correctly determine the precise stage of endometrial samples beyond the implantation window as well as helping to unravel the complex transcriptional changes that occur throughout the menstrual cycle. Moreover, such approaches would undoubtedly identify sub-populations within the epithelium and stroma that remain unidentified or uncharacterised to date, but that may play important roles in the transitions from an unreceptive to a receptive endometrium.

A significant discovery in this study is the role that the bivalent chromatin state plays in the establishment of the receptive state and the potential function it exerts in the transcriptional orchestration of the menstrual cycle. Thus, the co-occupancy by both H3K4me3 and H3K27me3 identified a number of gene loci that were bivalently marked in this way and overlapped with transcription factor bind sites for FOXO1 and/or KLF9. To further elucidate if these transcription factors are present at these sites and potentially signify the associated genes for imminent activation, ChIP-seq for KLF9 and FOXO1 could be conducted. This work could also be expanded to include the evolution and resolution of bivalently marked genes throughout the menstrual cycle by expanding the histone ChIP analysis to include samples from across the menstrual cycle.

One of my most striking observations was the finding that 5hmC dramatically increases as endometrial epithelial cells transition from a non-receptive to a receptive state. Although the mRNA levels of the *TET* enzymes remained unchanged in this study it would be interesting to investigate the protein levels via Western blot as well as to identify whether the regulation of protein levels reflects the 5hmC increase. In addition, post-transcriptional modifications of the TET proteins may impact on protein stability and hence may be associated with the increase in 5hmC. As this increase in 5hmC

specifically occurs in epithelial cells, epithelial organoids could provide a model to investigate the levels and regulation of 5hmC in the presence of steroid hormones. Alongside this, such *in vitro* models would allow for the manipulation of genes to study the functional consequences in a timely manner. Notably, the effect of TET knock-downs and the provision of a reducing environment on the ability for the organoids to reach a repetitive state would provide useful insights into the functionality of 5hmC in endometrial epithelial cells during the transition to receptivity.

From the 5mC- and 5hmC-profiling analysis in endometriosis I have identified that the Hippo signalling pathway is disrupted in endometriosis patients. This work could be built upon by investigating the mechanism in which the Hippo pathway may contribute to the aetiology of endometriosis. Key experiments along these lines would include to assess proliferation, apoptosis as well as migration in primary stromal cells and epithelial organoids alongside knock-out and over expression experiments of Hippo pathway genes.

A specific focus of my project was the potential for 5mC and 5hmC to serve as a biomarker for endometriosis as well as for the receptivity state of the endometrium at a cohort of genomic loci. My data is very promising, as I identified a number of loci that collectively can distinguish between endometriosis and controls as well as between receptive and non-receptive states. While this is undoubtedly an immensely exciting result and prospect, there are still a number of substantial hurdles that need to be overcome for these data to be clinically relevant. For one, in order for such an epigenetic receptivity test to be useful in the clinic in an IVF context, it would need to be able to be conducted in a time frame that allows adjustments to be made in real time. Most significantly, though, the biomarker set of loci will require comprehensive validation with a substantial number of additional samples, as well as refinement in terms of which specific sites and locus combinations provide the most meaningful results, before it is pursued further.

As noted throughout the thesis, the number of patient samples used in this study is similar to other studies, however this is still a limiting factor especially when considering the extent of heterogeneity of the human genome and epigenome. Moreover, the prevalence of undiagnosed endometriosis is a confounding factor that needs to be kept in mind and again, demands substantial future validation experiments on a large number of additional patient and control samples. Going forward, to validate

and expand on the analysis conducted in this study, the assessment of significant numbers of additional samples by 5mC and 5hmC profiling, either genome-wide or on the collated set of regions I identified, will be imperative and will add significant weight to these findings.

# References

Abu-Amero, Sayeda et al. 2010. “Epigenetic Signatures of Silver - Russell Syndrome.” *Journal of Medical Genetics* 47(3): 150–54.

Acosta, Anibal A. et al. 2000. “Endometrial Dating and Determination of the Window of Implantation in Healthy Fertile Women.” *Fertility and Sterility* 73(4): 788–98.

Afshar, Yalda et al. 2013. “Changes in Eutopic Endometrial Gene Expression During the Progression of Experimental Endometriosis in the Baboon, Papio Anubis1.” *Biology of Reproduction* 88(2): 1–9.

Aghajanova, L et al. 2011. “Unique Transcriptome, Pathways, and Networks in the Human Endometrial Fibroblast Response to Progesterone in Endometriosis.” *Biology of Reproduction* 84(4): 801–15.

Aguilar-Arnal, Lorena, Sayako Katada, Ricardo Orozco-Solis, and Paolo Sassone-Corsi. 2016. “NAD+ -SIRT1 Control of H3K4 Trimethylation through Circadian Deacetylation of MLL1M.” *Nature structural & molecular biology* 22(4): 312–18.

Ahn, Soo Hyun et al. 2015. “Pathophysiology and Immune Dysfunction in Endometriosis.” *BioMed Research International* 2015: 1–12.

Akalin, Altuna et al. 2012. “MethylKit : A Comprehensive R Package for the Analysis of Genome-Wide DNA Methylation Profiles.” *Genome Biology* 13(10): R87.

Albertsen, Hans M., and Kenneth Ward. 2016. “Genes Linked to Endometriosis by GWAS Are Integral to Cytoskeleton Regulation and Suggests That Mesothelial Barrier Homeostasis Is a Factor in the Pathogenesis of Endometriosis.” *Reproductive Sciences* 24(6): 193371911666084.

Altmäe, Signe et al. 2017. “Meta-Signature of Human Endometrial Receptivity: A Meta-Analysis and Validation Study of Transcriptomic Biomarkers.” *Scientific Reports* 7(1): 1–15.

Amouroux, Rachel et al. 2016. “De Novo DNA Methylation Drives 5hmC Accumulation in Mouse Zygotes.” *Nature Cell Biology* 18(2): 225–33.

Anders, Simon, and Wolfgang Huber. 2011. “Differential Expression Analysis for Sequence Count Data.” *Genome biology* 11(10): R106.

Andresini, Oriella et al. 2016. “A Cross-Talk between DNA Methylation and H3 Lysine 9 Dimethylation at the KvDMR1 Region Controls the Induction of Cdkn1c in Muscle

Cells." *Epigenetics* 11(11): 791–803.

Attia, George R et al. 2000. "Progesterone Receptor Isoform A But Not B Is Expressed in Endometriosis." *The Journal of Clinical Endocrinology & Metabolism* 85(8): 2897–2902.

Azuara, Véronique et al. 2006. "Chromatin Signatures of Pluripotent Cell Lines." *Nature cell biology* 8(5).

Bachman, M et al. 2014. "5-Hydroxymethylcytosine Is a Predominantly Stable DNA Modification." *Nature Chemistry* 6(12): 1049–55.

Bakos, O., Ö Lundkvist, and T. Bergh. 1993. "Transvaginal Sonographic Evaluation of Endometrial Growth and Texture in Spontaneous Ovulatory Cycles-a Descriptive Study." *Human Reproduction* 8(6): 799–806.

Barlow, Denise P. 2011. "Genomic Imprinting: A Mammalian Epigenetic Discovery Model." *Annual Review of Genetics* 45(1): 379–403.

Barski, Artem et al. 2007. "High-Resolution Profiling of Histone Methylations in the Human Genome." *Cell* 129(4): 823–37.

Bartolomei, Marisa S., Sharon Zemel, and Shirley M. Tilghman. 1991. "Parental Imprinting of the Human H19 Gene." *Nature* 351: 153–55.

Bassil, Rawad et al. 2018. "Does the Endometrial Receptivity Array Really Provide Personalized Embryo Transfer?" *Journal of Assisted Reproduction and Genetics* 35(7): 1301–5.

Becker, Justin S, Dario Nicetto, and Kenneth S Zaret. 2016. "H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes." *Trends in Genetics* 32(1): 29–41.

Benayoun, Bérénice A. et al. 2014. "H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency." *Cell* 158(3): 673–88.

Bentin-Ley, U. et al. 1999. "Presence of Uterine Pinopodes at the Embryo-Endometrial Interface during Human Implantation in Vitro." *Human Reproduction* 14(2): 515–20.

Bentivegna, Angela et al. 2016. "The Effect of Culture on Human Bone Marrow Mesenchymal Stem Cells: Focus on DNA Methylation Profiles." *Stem Cells International* 2016.

Bernstein, Bradley E et al. 2006. "A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells." *Cell* 125(2): 315–26.

Bhagwat, Sonali R. et al. 2013. "Endometrial Receptivity: A Revisit to Functional Genomics Studies on Human Endometrium and Creation of HGEx-ERdb." *PLoS ONE* 8(3).

Bhutani, Nidhi et al. 2010. "Reprogramming towards Pluripotency Requires AID-Dependent DNA Demethylation." *Nature* 463(7284): 1042–47.

Bird, Adrian P. 2017. "DNA Methylation and the Frequency of CpG in Animal DNA." *Nucleic Acids Research* 8(7): 1499–1504.

Bliek, Jet et al. 2009. "Hypomethylation at Multiple Maternally Methylated Imprinted Regions Including PLAGL1 and GNAS Loci in Beckwith-Wiedemann Syndrome." *European Journal of Human Genetics* 17(5): 611–19.

Booth, Michael J et al. 2012. "Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution." *Science* 336(6083): 934–37.

Borghese, Bruno et al. 2008. "Gene Expression Profile for Ectopic versus Eutopic Endometrium Provides New Insights into Endometriosis Oncogenic Potential." *Molecular Endocrinology* 22(11): 2557–62.

———. 2010. "Genome-Wide Profiling of Methylated Promoters in Endometriosis Reveals a Subtelomeric Location of Hypermethylation." *Molecular Endocrinology* 24(9): 1872–85.

Boros, Joanna et al. 2014. "Polycomb Repressive Complex 2 and H3K27me3 Cooperate with H3K9 Methylation to Maintain Heterochromatin Protein 1α at Chromatin." *Molecular and cellular biology* 34(19): 3662–74.

Bourc'his, Déborah et al. 2016. "Dnmt3L and the Establishment of Maternal Genomic Imprints." *Science* 294(5551): 2536–39.

Brosens, Ivo, and Giuseppe Benagiano. 2013. "Is Neonatal Uterine Bleeding Involved in the Pathogenesis of Endometriosis as a Source of Stem Cells?" *Fertility and Sterility* 100(3): 622–23.

Bruner-Tran, K. L. et al. 2013. "Medical Management of Endometriosis: Emerging Evidence Linking Inflammation to Disease Pathophysiology." *Minerva Ginecologica*

65(2): 199–213.

Buck, V. U., B. Gellersen, R. E. Leube, and I. Classen-Linke. 2015. “Interaction of Human Trophoblast Cells with Gland-like Endometrial Spheroids: A Model System for Trophoblast Invasion.” *Human Reproduction* 30(4): 906–16.

Buiting, Karin et al. 1994. “Detection of Aberrant Dna Methylation in Unique Prader - Willi Syndrome Patients and Its Diagnostic Implications.” *Human Molecular Genetics* 3(6): 893–95.

Burmenskaya, O. V. et al. 2017. “Transcription Profile Analysis of the Endometrium Revealed Molecular Markers of the Personalized ‘Window of Implantation’ during in Vitro Fertilization.” *Gynecological Endocrinology* 33: 22–27.

Burney, Richard O. et al. 2007. “Gene Expression Analysis of Endometrium Reveals Progesterone Resistance and Candidate Susceptibility Genes in Women with Endometriosis.” *Endocrinology* 148(8): 3814–26.

Burns, Katherine A. et al. 2012. “Role of Estrogen Receptor Signaling Required for Endometriosis-like Lesion Establishment in a Mouse Model.” *Endocrinology* 153(8): 3960–71.

Burton, Graham J et al. 2002. “Uterine Glands Provide Histiotrophic Nutrition for the Human Fetus during the First Trimester of Pregnancy.” *The Journal of Clinical Endocrinology & Metabolism* 87(6): 2954–59.

Canis, M. et al. 1997. “Revised American Society for Reproductive Medicine Classification of Endometriosis: 1996.” *Fertility and Sterility* 67(5): 817–21.

Carpenter, Anne E et al. 2006. “CellProfiler: Image Analysis Software for Identifying and Quantifying Cell Phenotypes.” *Genome Biology* 7(10).

Chan, Crystal et al. 2013. “Discovery of Biomarkers of Endometrial Receptivity through a Minimally Invasive Approach : A Validation Study with Implications for Assisted Reproduction.” *Fertility and Sterility* 100(3): 810–817.e8.

Charnock-Jones, D S, A M Sharkey, P Fenwick, and S K Smith. 1994. “Leukaemia Inhibitory Factor MRNA Concentration Peaks in Human Endometrium at the Time of Implantation and the Blastocyst Contains MRNA for the Receptor at This Time.” *Journal of Reproduction and Fertility* 101: 421–26.

Choi, Juhyun et al. 2009. "Mst1-FoxO Signaling Protects Naïve T Lymphocytes from Cellular Oxidative Stress in Mice." *PLoS ONE* 4(11): 1–10.

Collinson, Adam et al. 2016. "Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and Differentiation Defects in Human Embryonic Stem Cells." *Cell Reports* 17(10): 2700–2714.

Colón-Caraballo, Mariano et al. 2018. "Effects of Histone Methyltransferase Inhibition in Endometriosis." *Biology of reproduction* 0(0): 1–15.

Colón-Caraballo, Mariano, Janice B Monteiro, and Idhaliz Flores. 2015. "H3K27me3 Is an Epigenetic Mark of Relevance in Endometriosis." *Reproductive sciences* 22(9): 1134–42.

Cornillie, F. et al. 1991. "Lysosomal Enzymes in the Human Endometrium: A Biochemical Study in Untreated and Levonorgestrel-Treated Women." *Contraception* 43(4): 387–400.

DeChiara, Thomas M., Elizabeth J. Robertson, and Argiris Efstratiadis. 1991. "Parental Imprinting of the Mouse Insulin-like Growth Factor II Gene." *Cell* 64(4): 849–59.

Díaz-Gimeno, Patricia et al. 2011. "A Genomic Diagnostic Tool for Human Endometrial Receptivity Based on the Transcriptomic Signature." *Fertility and Sterility* 95(1): 50–60.

Diedrich, Klaus, B. C.J.M. Fauser, P. Devroey, and G. Griesinger. 2007. "The Role of the Endometrium and Embryo in Human Implantation." *Human Reproduction Update* 13(4): 365–77.

Dimitriadis, Evdokia et al. 2006. "Interleukin-11, IL-11 Receptorα and Leukemia Inhibitory Factor Are Dysregulated in Endometrium of Infertile Women with Endometriosis during the Implantation Window." *Journal of Reproductive Immunology* 69(1): 53–64.

———. 2010. "Local Regulation of Implantation at the Human Fetal-Maternal Interface." *International Journal of Developmental Biology* 54(2–3): 313–22.

Disteche, Christine M., and Joel B. Berletch. 2015. "X-Chromosome Inactivation and Escape." *Journal of Genetics* 94(4): 591–99.

Du, Jiamu, Lianna M. Johnson, Steven E. Jacobsen, and Dinshaw J. Patel. 2015. "DNA

Methylation Pathways and Their Crosstalk with Histone Methylation." *Nature reviews. Molecular cell biology* 16: 519–32.

Dunselman, G. A.J. et al. 2014. "ESHRE Guideline: Management of Women with Endometriosis." *Human Reproduction* 29(3): 400–412.

Dyson, Matthew T et al. 2014. "Genome-Wide DNA Methylation Analysis Predicts an Epigenetic Switch for GATA Factor Expression in Endometriosis." *PLoS genetics* 10(3): e1004158.

———. 2015. "Aberrant Expression and Localization of Deoxyribonucleic Acid Methyltransferase 3B in Endometriotic Stromal Cells." *Fertility and sterility*.

Eden, Eran et al. 2009. "GOrilla: A Tool for Discovery and Visualization of Enriched GO Terms in Ranked Gene Lists." *BMC Bioinformatics* 10: 1–7.

van Eijk, Kristel R et al. 2012. "Genetic Analysis of DNA Methylation and Gene Expression Levels in Whole Blood of Healthy Human Subjects." *BMC Genomics* 13(1): 636.

Ellinger, P, and R A Coulson. 1944. "The Urinary Elimination of Nicotinamide Methochloride by Man." *Biochemical Journal* 38: 265–70.

Emera, Deena, Roberto Romero, and Günter Wagner. 2012. "The Evolution of Menstruation: A New Model for Genetic Assimilation." *BioEssays* 34(1): 26–35.

Enciso, M. et al. 2018. "Development of a New Comprehensive and Reliable Endometrial Receptivity Map (ER Map/ER Grade) Based on RT-QPCR Gene Expression Analysis." *Human Reproduction* 33(2): 220–28.

Estève, Pierre Olivier et al. 2009. "Regulation of DNMT1 Stability through SET7-Mediated Lysine Methylation in Mammalian Cells." *Proceedings of the National Academy of Sciences of the United States of America* 106(13): 5076–81.

Evans, Gloria E. et al. 2012. "Gene and Protein Expression Signature of Endometrial Glandular and Stromal Compartments during the Window of Implantation." *Fertility and Sterility* 97(6).

———. 2014. "In the Secretory Endometria of Women, Luminal Epithelia Exhibit Gene and Protein Expressions That Differ from Those of Glandular Epithelia." *Fertility and Sterility* 102(1): 307–17.

Evans, Gloria E et al. 2018. "Does the Endometrial Gene Expression of Fertile Women

Vary within and between Cycles?" *Human Reproduction* 33(3): 452–63.

Evans, Jemma et al. 2016. "Fertile Ground: Human Endometrial Programming and Lessons in Health and Disease." *Nature Reviews Endocrinology* 12(11): 654–67.

Evans, Jemma, and Lois A. Salamonsen. 2012. "Inflammation, Leukocytes and Menstruation." *Reviews in Endocrine and Metabolic Disorders* 13(4): 277–88.

———. 2014. "Decidualized Human Endometrial Stromal Cells Are Sensors of Hormone Withdrawal in the Menstrual Inflammatory Cascade1." *Biology of Reproduction* 90(1): 1–12.

Eyster, Kathleen M, Olga Klinkova, Vanessa Kennedy, and Keith A Hansen. 2007. "Whole Genome Deoxyribonucleic Acid Microarray Analysis of Gene Expression in Ectopic versus Eutopic Endometrium." *Fertility and sterility* 88(6): 1505–33.

Falconer, Henrik, Thomas D'Hooghe, and Gabriel Fried. 2007. "Endometriosis and Genetic Polymorphisms." *Obstetrical & gynecological survey* 62(November 2015): 616–28.

Fassbender, Amelie et al. 2014. "World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: IV. Tissue Collection, Processing, and Storage in Endometriosis Research." *Fertility and Sterility* 102(5): 1244–53.

Faust, C., A. Schumacher, B. Holdener, and T. Magnuson. 1995. "The Eed Mutation Disrupts Anterior Mesoderm Production in Mice." *Development* 121(2): 273–85.

Ferguson-Smith, Anne C., and Deborah Bourc'his. 2018. "The Discovery and Importance of Genomic Imprinting." *eLife* 7: 1–5.

Ficz, Gabriella et al. 2011. "Dynamic Regulation of 5-Hydroxymethylcytosine in Mouse ES Cells and during Differentiation." *Nature* 473(7347): 398–404.

Filant, Justyna, and Thomas E. Spencer. 2014. "Uterine Glands: Biological Roles in Conceptus Implantation, Uterine Receptivity and Decidualization." *International Journal of Developmental Biology* 58(2–4): 107–16.

Flores, Idhaliz et al. 2007. "Molecular Profiling of Experimental Endometriosis Identified Gene Expression Patterns in Common with Human Disease." *Fertility and Sterility* 87(5): 1180–99.

Franchi, Anahi et al. 2008. “Expression of Immunomodulatory Genes, Their Protein Products and Specific Ligands/Receptors during the Window of Implantation in the Human Endometrium.” *Molecular Human Reproduction* 14(7): 413–21.

Fuks, François, Paul J. Hurd, Rachel Deplus, and Tony Kouzarides. 2003. “The DNA Methyltransferases Associate with HP1 and the SUV39H1 Histone Methyltransferase.” *Nucleic Acids Research* 31(9): 2305–12.

Fung, Jenny N et al. 2018. “Genetic Regulation of Disease Risk and Endometrial Gene Expression Highlights Potential Target Genes for Endometriosis and Polycystic Ovarian Syndrome.” *Scientific Reports* 8(1): 1–19.

Gargett, C. E., K. E. Schwab, and J. a. Deane. 2015. “Endometrial Stem/Progenitor Cells: The First 10 Years.” *Human Reproduction Update* 0(0): 1–27.

Garry, R., R. Hart, K. A. Karthigasu, and C. Burke. 2009. “A Re-Appraisal of the Morphological Changes within the Endometrium during Menstruation: A Hysteroscopic, Histological and Scanning Electron Microscopic Study.” *Human Reproduction* 24(6): 1393–1401.

Gaynor, Louise M., and Francesco Colucci. 2017. “Uterine Natural Killer Cells: Functional Distinctions and Influence on Pregnancy in Humans and Mice.” *Frontiers in Immunology* 8(APR).

Gellersen, Birgit, and Jan J. Brosens. 2014. “Cyclic Decidualization of the Human Endometrium in Reproductive Health and Failure.” *Endocrine Reviews* 35(6): 851–905.

Ghabreau, Lina et al. 2004. “Correlation between the DNA Global Methylation Status and Progesterone Receptor Expression in Normal Endometrium, Endometrioid Adenocarcinoma and Precursors.” *Virchows Archiv* 445(2): 129–34.

Giannoukakis, Nick et al. 1993. “Parental Genomic Imprinting of the Human IGF2 Gene.” *Nature Genetics* 4(1): 98–101.

Gibson, Mark et al. 1991. “Error in Histologic Dating of Secretory Endometrium: Variance Component Analysis.” *Fertility and Sterility* 56(2): 242–47.

Glaser, Stefan et al. 2006. “Multiple Epigenetic Maintenance Factors Implicated by the Loss of Mll2 in Mouse Development.” *Development* 133(8): 1423–32.

Globisch, Daniel et al. 2010. “Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates.” *PLoS ONE* 5(12): 1–9.

Godderis, Lode et al. 2015. “Global Methylation and Hydroxymethylation in DNA from Blood and Saliva in Healthy Volunteers.” *BioMed Research International* 2015: 8.

Gordts, Stephan, Philippe Koninckx, and Ivo Brosens. 2017. “Pathogenesis of Deep Endometriosis.” *Fertility and Sterility* (20).

Goyal, Rachna et al. 2007. “Phosphorylation of Serine-515 Activates the Mammalian Maintenance Methyltransferase Dnmt1.” *Epigenetics* 2(3): 155–60.

Gray, C Allison et al. 2001. “Developmental Biology of Uterine Glands.” *Biology of Reproduction* 65: 1311–23.

Greening, David W et al. 2016. “Human Endometrial Exosomes Contain Hormone-Specific Cargo Modulating Trophoblast Adhesive Capacity : Insights into Endometrial-Embryo Interactions 1.” *Biology of Reproduction* 94(2): 1–15.

Grimaldi, Giulia et al. 2016. “Down-Regulation of the Histone Methyltransferase EZH2 Contributes to the Epigenetic Programming of Decidualizing Human Endometrial Stromal Cells.” *Molecular Endocrinology* 25(11): 1892–1903.

Guidice, Linda C. 2010. “Clinical Practice: Endometriosis.” *The New England Journal of Medicine* 362(25): 2389–98.

Guo, Sun Wei. 2009a. “Epigenetics of Endometriosis.” *Molecular human reproduction* 15(10): 587–607.

———. 2009b. “Recurrence of Endometriosis and Its Control.” *Human Reproduction Update* 15(4): 441–61.

———. 2012. “The Endometrial Epigenome and Its Response to Steroid Hormones.” *Molecular and Cellular Endocrinology* 358(2): 185–96.

Hahn, Maria A. et al. 2011. “Relationship between Gene Body DNA Methylation and Intragenic H3K9me3 and H3K36me3 Chromatin Marks.” *PLoS ONE* 6(4).

Hamadneh, Lama et al. 2018. “Culturing Conditions Highly Affect DNA Methylation and Gene Expression Levels in MCF7 Breast Cancer Cell Line.” *In Vitro Cellular and Developmental Biology - Animal* 54(5): 331–34.

Hannan, Natalie J. et al. 2010. "2D-DiGE Analysis of the Human Endometrial Secretome Reveals Differences between Receptive and Nonreceptive States in Fertile and Infertile Women." *Journal of Proteome Research* 9(12): 6256–64.

———. 2011. "Analysis of Fertility-Related Soluble Mediators in Human Uterine Fluid Identifies VEGF as a Key Regulator of Embryo Implantation." *Endocrinology* 152(12): 4948–56.

Hashimoto, Tomoko et al. 2017. "Efficacy of the Endometrial Receptivity Array for Repeated Implantation Failure in Japan: A Retrospective, Two-Centers Study." *Reproductive Medicine and Biology* 16(3): 290–96.

He, Yu-fei et al. 2011. "Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA." *Science* 333(6047): 1303–7.

Heintzman, Nathaniel D et al. 2007. "Distinct and Predictive Chromatin Signatures of Transcriptional Promoters and Enhancers in the Human Genome." *Nature Genetics* 39(3): 311–18.

Hodges, Emily et al. 2009. "High Definition Profiling of Mammalian DNA Methylation by Array Capture and Single Molecule Bisulfite Sequencing." *Genome Research* 19(9): 1593–1605.

Hoff, J. D., M. E. Quigley, and S. S.C. Yen. 1983. "Hormonal Dynamics at Midcycle: A Reevaluation." *Journal of Clinical Endocrinology and Metabolism* 57(4): 792–96.

Holt, V L, and N S Weiss. 2000. "Recommendations for the Design of Epidemiologic Studies of Endometriosis." *Epidemiology* 11(6): 654–59.

Horvath, Steve. 2015. "DNA Methylation Age of Human Tissues and Cell Types." *Genome Biology* 16(1): 96.

Houshdaran, Sahar et al. 2016. "Aberrant Endometrial DNA Methylome and Associated Gene Expression in Endometriosis." *Biology of Reproduction* 95(5): 93.

Houshdaran, Sahar, Zara Zelenko, Juan C. Irwin, and Linda C. Giudice. 2014. "Human Endometrial DNA Methylome Is Cycle-Dependent and Is Associated With Gene Expression Regulation." *Molecular Endocrinology* 28(7): 1118–35.

Huang, Da Wei, Brad T. Sherman, and Richard A. Lempicki. 2009a. "Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large

Gene Lists." *Nucleic Acids Research* 37(1): 1–13.

———. 2009b. "Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources." *Nature Protocols* 4(1): 44–57.

Hufnagel, Demetra et al. 2015. "The Role of Stem Cells in the Etiology and Pathophysiology of Endometriosis." *Seminars in Reproductive Medicine* 33(5): 333–40.

Illingworth, Robert S. et al. 2008. "A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci." *PLoS Biology* 6(1): 0037–0051.

———. 2010. "Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome." *PLoS Genetics* 6(9).

Irungu, Stella et al. 2019. "Discovery of Non-Invasive Biomarkers for the Diagnosis of Endometriosis." *Clinical Proteomics* 16(1): 1–16.

Ito, Shinsuke et al. 2011. "Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine." *Science* 333(6047): 1300–1303.

Jones, Peter A. 2012. "Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and Beyond." *Nature reviews. Genetics* 13(7): 484–92.

Jung, Jinkyu et al. 2017. "Nicotinamide Metabolism Regulates Glioblastoma Stem Cell Maintenance." *Journal of Clinical Investigation* 2(10): 1–23.

van Kaam, Kim J A F et al. 2011. "Deoxyribonucleic Acid Methyltransferases and Methyl-CpG-Binding Domain Proteins in Human Endometrium and Endometriosis." *Fertility and Sterility* 95(4): 1421–27.

Kajihara, Takeshi et al. 2006. "Differential Expression of FOXO1 and FOXO3a Confers Resistance to Oxidative Cell Death upon Endometrial Decidualization." *Molecular Endocrinology* 20(10): 2444–55.

Kajihara, Takeshi, and Jan J Brosens. 2013. "The Role of FOXO1 in the Decidual Transformation of the Endometrium and Early Pregnancy." *Medical Molecular Morphology* 46: 61–68.

Kannenberg, K., C. Urban, and G. Binder. 2012. "Increased Incidence of Aberrant DNA Methylation within Diverse Imprinted Gene Loci Outside of IGF2/H19 in Silver-Russell Syndrome." *Clinical Genetics* 81(4): 366–77.

Kennedy, Stephen, Helen Mardon, and David Barlow. 1995. "Familial Endometriosis." *Journal of Assisted Reproduction and Genetics* 12(1): 32–34.

Khan, Khaleque Newaz et al. 2008. "Immunopathogenesis of Pelvic Endometriosis: Role of Hepatocyte Growth Factor, Macrophages and Ovarian Steroids." *American Journal of Reproductive Immunology* 60(5): 383–404.

Kim, J Julie et al. 2007. "Altered Expression of HOXA10 in Endometriosis: Potential Role in Decidualization." *Molecular human reproduction* 13(5): 323–32.

Kim, Yun-Jeong, and David M Wilson III. 2013. "Overview of Base Excision Repair Biochemistry COPING WITH ENDOGENOUS DNA DAMAGE." *Current Molecular Pharmacology* 5(1): 3–13.

Koh, Kian Peng et al. 2011. "Tet1 and Tet2 Regulate 5-Hydroxymethylcytosine Production and Cell Lineage Specification in Mouse Embryonic Stem Cells." *Cell Stem Cell* 8(2): 200–213.

Kriaucionis, Skirmantas, and Nathaniel Heintz. 2009. "The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain." *Science* 324(5929): 929–30.

Krueger, Felix, and Simon R Andrews. 2011. "Bismark : A Flexible Aligner and Methylation Caller for Bisulfite-Seq Applications." *Bioinformatics* 27(11): 1571–72.

Kukushkina, Viktorija et al. 2017. "DNA Methylation Changes in Endometrium during the Transition from Pre-Receptive to Receptive Phase." *Scientific Reports* (May): 1–11.

Kulp, J. L., R. Mamillapalli, and H. S. Taylor. 2016. "Aberrant HOXA10 Methylation in Patients With Common Gynecologic Disorders: Implications for Reproductive Outcomes." *Reproductive Sciences* 23(4): 455–63.

Laukka, Tuomas et al. 2016. "Fumarate and Succinate Regulate Expression of Hypoxia-Inducible Genes via TET Enzymes." *Journal of Biological Chemistry* 291(8): 4256–65.

Laurent, Louise et al. 2010. "Dynamic Changes in the Human Methylome during Differentiation." *Genome Research* 20: 320–31.

Lee, Banghyun, Hongling Du, and Hugh S Taylor. 2009. "Experimental Murine Endometriosis Induces DNA Methylation and Altered Gene Expression in Eutopic Endometrium." *Biology of Reproduction* 80(1): 79–85.

Lee, Hyun Ju et al. 2015. "Various Anatomic Locations of Surgically Proven Endometriosis: A Single-Center Experience." *Obstetrics & Gynecology Science* 58(1): 53.

Lee, Ji Yeong, Millina Lee, and Sung Ki Lee. 2011. "Role of Endometrial Immune Cells in Implantation." *Clinical and Ecperimental Reproductive Medicine* 38(3): 119–25.

Lessey, Bruce A. 2011. "Assessment of Endometrial Receptivity." *Fertility and Sterility* 96(3): 522–29.

Lessey, Bruce A, Aleksandr E. Vendrov, and Lingwen Yuan. 2003. "Endometrial Cancer Cells as Models to Study Uterine Receptivity." In *Cell and Molecular Biology of Endometrial Carcinoma*, , 267–79.

Levy, Adrian R et al. 2011. "Economic Burden of Surgically Confirmed Endometriosis in Canada." *Journal of obstetrics and gynaecology Canada* 33(8): 830–37.

Leyendecker, Gerhard et al. 2004. "Uterine Peristaltic Activity and the Development of Endometriosis." *Annals of the New York Academy of Sciences* 1034: 338–55.

Li, E., C. Beard, and R. Jaenisch. 1993. "Role for DNA Methylation in Genomic Imprinting." *Nature* 366: 362–65.

Li, En, Timothy H. Bestor, and Rudolf Jaenisch. 1992. "Targeted Mutation of the DNA Methyltransferase Gene Results in Embryonic Lethality." *Cell* 69(6): 915–26.

Li, Tao et al. 2018. "Structural and Mechanistic Insights into UHRF1-Mediated DNMT1 Activation in the Maintenance DNA Methylation." *Nucleic Acids Research* 46(6): 3218–31.

Li, Weiwei, and Min Liu. 2011. "Distribution of 5-Hydroxymethylcytosine in Different Human Tissues." *Journal of Nucleic Acids* 2011.

Li, Yanfen et al. 2016. "Progesterone Alleviates Endometriosis via Inhibition of Uterine Cell Proliferation, Inflammation and Angiogenesis in an Immunocompetent Mouse Model." *Plos One* 11(10): e0165347.

Lienhard, Matthias et al. 2014. "Sequence Analysis MEDIPS : Genome-Wide Differential Coverage Analysis of Sequencing Data Derived from DNA Enrichment Experiments." *Bioinformatics* 30(2): 284–86.

van der Linden, Paul J.Q. 1996. "Theories on the Pathogenesis of Endometriosis."

*International journal of reproductive medicine* 11: 53–65.

Ling, Yan et al. 2004. "Modification of de Novo DNA Methyltransferase 3a (Dnmt3a) by SUMO-1 Modulates Its Interaction with Histone Deacetylases (HDACs) and Its Capacity to Repress Transcription." *Nucleic Acids Research* 32(2): 598–610.

Liu, Xiaoyu et al. 2016. "Distinct Features of H3K4me3 and H3K27me3 Chromatin Domains in Pre-Implantation Embryos." *Nature* 537(7621): 558–62.

Liu, Xishi, Qi Zhang, and Sun Wei Guo. 2018. "Histological and Immunohistochemical Characterization of the Similarity and Difference Between Ovarian Endometriomas and Deep Infiltrating Endometriosis." *Reproductive Sciences* 25(3): 329–40.

Long, Mark D., Dominic J. Smiraglia, and Moray J. Campbell. 2017. "The Genomic Impact of DNA CpG Methylation on Gene Expression; Relationships in Prostate Cancer." *Biomolecules* 7(1): 1–20.

Love, Michael I, Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome biology* 15(12): 550.

Lucas, Emma S et al. 2016. "Loss of Endometrial Plasticity in Recurrent Pregnancy Loss." *Stem Cells* 34(2): 346–56.

Luger, Karolin et al. 1997. "Crystal Structure of the Nucleosome Core Particle at 2.8 A Resolution." *Nature* 389: 251–60.

Lunnon, Katie et al. 2016. "Variation in 5-Hydroxymethylcytosine across Human Cortex and Cerebellum." *Genome Biology* 17(27).

Macklon, N. S., and B. C J M Fauser. 2001. "Follicle-Stimulating Hormone and Advanced Follicle Development in the Human." *Archives of Medical Research* 32(6): 595–600.

Makker, A., F. W. Bansode, V. M.L. Srivastava, and M. M. Singh. 2006. "Antioxidant Defense System during Endometrial Receptivity in the Guinea Pig: Effect of Ormeloxifene, a Selective Estrogen Receptor Modulator." *Journal of Endocrinology* 188(1): 121–34.

Masuda, Hirotaka et al. 2012. "A Novel Marker of Human Endometrial Mesenchymal Stem-like Cells." *Cell transplantation* 21(10): 2201–14.

Matalliotakis, Ioannis M. et al. 2008. "Familial Aggregation of Endometriosis in the Yale

Series." *Archives of Gynecology and Obstetrics* 278(6): 507–11.

Matarese, Giuseppe, Giuseppe De Placido, Yorgos Nikas, and Carlo Alviggi. 2003. "Pathogenesis of Endometriosis: Natural Immunity Dysfunction or Autoimmune Disease?" *Trends in Molecular Medicine* 9(5): 223–28.

Matsuoka, Aki et al. 2010. "Progesterone Increases Manganese Superoxide Dismutase Expression via a CAMP-Dependent Signaling Mediated by Noncanonical Wnt5a Pathway in Human Endometrial Stromal Cells." *Journal of Clinical Endocrinology and Metabolism* 95(11): 291–99.

Maybin, Jacqueline A, and Hilary O D Critchley. 2015. "Menstrual Physiology : Implications for Endometrial Pathology and Beyond." *Human Reproduction Update* 21(6): 748–61.

Mazur, Erik C. et al. 2015. "Progesterone Receptor Transcriptome and Cistrome in Decidualized Human Endometrial Stromal Cells." *Endocrinology* 156(6): 2239–53.

McGrath, James, and Davor Solter. 1984. "Completion of Mouse Embryogenesis Requires Both the Maternal and Paternal Genomes." *Cell* 37(1): 179–83.

McLeay, Robert C., and Timothy L. Bailey. 2010. "Motif Enrichment Analysis: A Unified Framework and an Evaluation on ChIP Data." *BMC Bioinformatics* 11.

Mellen, Marian et al. 2012. "MeCP2 Binds to 5hmc Enriched within Active Genes and Accessible Chromatin in the Nervous System." *Cell* 151(7): 1417–1430.

Meuleman, Christel et al. 2009. "High Prevalence of Endometriosis in Infertile Women with Normal Ovulation and Normospermic Partners." *Fertility and Sterility* 92(1): 68–74.

Mikeska, Thomas, and Jeffrey M. Craig. 2014. "DNA Methylation Biomarkers: Cancer and Beyond." *Genes* 5(3): 821–64.

Minor, Emily A., Brenda L. Court, Juan I. Young, and Gaofeng Wang. 2013. "Ascorbate Induces Ten-Eleven Translocation (Tet) Methylcytosine Dioxygenase-Mediated Generation of 5-Hydroxymethylcytosine." *Journal of Biological Chemistry* 288(19): 13669–74.

Mo, Bilan et al. 2006. "ECC-1 Cells: A Well-Differentiated Steroid-Responsive Endometrial Cell Line with Characteristics of Luminal Epithelium1." *Biology of*

*Reproduction* 75(3): 387–94.

Moen, Mette Haase, and Trine Stokstad. 2002. "A Long-Term Follow-up Study of Women with Asymptomatic Endometriosis Diagnosed Incidentally at Sterilization." *Fertility and Sterility* 78(4): 773–76.

Monteiro, Janice B et al. 2014. "Endometriosis Is Characterized by a Distinct Pattern of Histone 3 and Histone 4 Lysine Modifications." *Reproductive Sciences* 21(3): 305–18.

Moreno-Moya, Juan Manuel et al. 2014. "The Transcriptomic and Proteomic Effects of Ectopic Overexpression of MiR-30d in Human Endometrial Epithelial Cells." *Molecular Human Reproduction* 20(6): 550–66.

Munro, S. K., C. M. Farquhar, M. D. Mitchell, and a. P. Ponnampalam. 2010. "Epigenetic Regulation of Endometrium during the Menstrual Cycle." *Molecular Human Reproduction* 16(5): 297–310.

Murray, Michael J. et al. 2004. "A Critical Analysis of the Accuracy, Reproducibility, and Clinical Utility of Histologic Endometrial Dating in Fertile Women." *Fertility and Sterility* 81(5): 1333–43.

Nakayama, J. et al. 2001. "Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly." *Science* 292(5514): 110–13.

Naqvi, Hanyia, Ysabel Ilagan, Graciela Krikun, and Hugh S Taylor. 2014. "Altered Genome-Wide Methylation in Endometriosis." *Reproductive sciences* 21(10): 1237–43.

Nasu, Kaei et al. 2011. "Aberrant DNA Methylation Status of Endometriosis: Epigenetics as the Pathogenesis, Biomarker and Therapeutic Target." *The journal of obstetrics and gynaecology research* 37(7): 683–95.

Nestor, Colm E. et al. 2012. "Tissue Type Is a Major Modifier of the 5-Hydroxymethylcytosine Content of Human Genes." *Genome Research* 22(3): 467–77.

Ng, Annie Y N et al. 2002. "Inactivation of the Transcription Factor Elf3 in Mice Results in Dysmorphogenesis and Altered Differentiation of Intestinal Epithelium." *Gastroenterology* 122(5): 1455–66.

Ng, Ray Kit et al. 2008. "Epigenetic Restriction of Embryonic Cell Lineage Fate by

Methylation of Elf5." *Nature Cell Biology* 10(11): 1280–90.

Nisolle, Michelle, and Jacques Donnez. 1997. "Peritoneal Endometriosis, Ovarian Endometriosis, and Adenomyotic Nodules of the Rectovaginal Septum Are Three Different Entities." *Fertility and Sterility* 68(4): 585–96.

Nnoaham, Kelechi E et al. 2011. "Impact of Endometriosis on Quality of Life and Work Productivity : A Multicenter Study across Ten Countries." *Fertility and sterility* 96(2): 366–73.

Nothnick, Warren B et al. 2016. "Serum MiR-451a Levels Are Significantly Elevated in Women With Endometriosis and Recapitulated in Baboons (Papio Anubis) With Experimentally-Induced Disease." *Reproductive sciences (Thousand Oaks, Calif.)*.

Noyes, R. W., A. T. Hertig, and J. Rock. 1950. "Dating the Endometrial Biopsy." *Fertility and Sterility* 1(1): 3–25.

Nyholt, Dale R et al. 2012. "Genome-Wide Association Meta-Analysis Identifies New Endometriosis Risk Loci." *Nature Genetics* 44(12): 1355–59.

O'Carrol, Donal et al. 2001. "The Polycomb-Group Gene Ezh2 Is Required for Early Mouse Development." *Molecular and cellular biology* 21(13): 4330–36.

Okada, Hidetaka, Tomoko Tsuzuki, and Hiromi Murata. 2018. "Decidualization of the Human Endometrium." *Reproductive Medicine and Biology* 17(3): 220–27.

Okano, Masaki, Daphne W Bell, Daniel A Haber, and En Li. 1999. "DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development." *Cell* 99: 247–57.

Onder, Tamer T. et al. 2012. "Chromatin-Modifying Enzymes as Modulators of Reprogramming." *Nature* 483(7391): 598–602.

Ooi, Steen K.T. et al. 2007. "DNMT3L Connects Unmethylated Lysine 4 of Histone H3 to de Novo Methylation of DNA." *Nature* 448(7154): 714–17.

Pasini, Diego et al. 2004. "Suz12 Is Essential for Mouse Development and for EZH2 Histone Methyltransferase Activity." *EMBO Journal* 23(20): 4061–71.

Pastor, William A. et al. 2011. "Genome-Wide Mapping of 5-Hydroxymethylcytosine in Embryonic Stem Cells." *Nature* 473(7347): 394–97.

Paulsen, Martina, and Anne C. Ferguson-Smith. 2001. "DNA Methylation in Genomic Imprinting, Development, and Disease." *Journal of Pathology* 195(1): 97–110.

Pawar, Sandeep et al. 2013. "STAT3 Regulates Uterine Epithelial Remodeling and Epithelial-Stromal Crosstalk During Implantation." *Molecular Endocrinology* 27(12): 1996–2012.

Peters, Antoine H F M et al. 2001. "Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability Ment of Transcription Factors and Chromatin Remodeling." *Cell* 107: 323–37.

Ponnampalam, Anna P. et al. 2004. "Molecular Classification of Human Endometrial Cycle Stages by Transcriptional Profiling." *Molecular Human Reproduction* 10(12): 879–93.

Rahmioglu, Nilufer et al. 2017. "Variability of Genome-Wide DNA Methylation and MRNA Expression Profiles in Reproductive and Endocrine Disease Related Tissues." *Epigenetics* 12(10): 897–908.

Raiber, Eun Ang et al. 2017. "Base Resolution Maps Reveal the Importance of 5-Hydroxymethylcytosine in a Human Glioblastoma." *npj Genomic Medicine* 2(6): 1–6.

Raine-Fenning, Nicholas J. et al. 2004. "Defining Endometrial Growth during the Menstrual Cycle with Three-Dimensional Ultrasound." *BJOG: An International Journal of Obstetrics and Gynaecology* 111(9): 944–49.

Rainier, Shirley et al. 1993. "Relaxation of Imprinted Genes in Human Cancer." *Nature* 362(6422): 747–49.

Raju, Gottumukkala Achyuta Rama et al. 2013. "Luteinizing Hormone and Follicle Stimulating Hormone Synergy: A Review of Role in Controlled Ovarian Hyper-Stimulation." *Journal of Human Reproductive Sciences* 6(4): 227–34.

Ramsden, David B, Rosemary H Waring, David J Barlow, and Richard B Parsons. 2017. "Nicotinamide N -Methyltransferase in Health and Cancer." *International Journal of Tryptophan Research* 10: 1–19.

Rashid, Najwa A., Sujata Lalitkumar, Parameswaran G. Lalitkumar, and Kristina Gemzell-Danielsson. 2011. "Endometrial Receptivity and Human Embryo Implantation." *American Journal of Reproductive Immunology* 66: 23–30.

Reik, Wolf et al. 1987. “Genomic Imprinting Determines Methylation of Parental Alleles in Transgenic Mice.” *Nature* 328: 248–51.

Rier, Sherry, and Warren G. Foster. 2003. “Environmental Dioxins and Endometriosis.” *Seminars in Reproductive Medicine* 21: 145–53.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2009. “EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data.” *Bioinformatics* 26(1): 139–40.

Roeßler, Markus et al. 2005. “Identification of Nicotinamide N-Methyltransferase as a Novel Serum Tumor Marker for Colorectal Cancer.” *Clinical Cancer Research* 11(18): 6550–57.

Rogers, Peter A.W. et al. 2017. “Research Priorities for Endometriosis: Recommendations from a Global Consortium of Investigators in Endometriosis.” *Reproductive Sciences* 24(2): 202–26.

Roh, T.-Y., S. Cuddapah, K. Cui, and K. Zhao. 2006. “The Genomic Landscape of Histone Modifications in Human T Cells.” *Proceedings of the National Academy of Sciences* 103(43): 15782–87.

Romero-Pérez, Laura et al. 2015. “A Role for the Transducer of the Hippo Pathway, TAZ, in the Development of Aggressive Types of Endometrial Cancer.” *Modern pathology* 28(11): 1492–1503.

Rondelet, Grégoire, Thomas Dal Maso, Luc Willems, and Johan Wouters. 2016. “Structural Basis for Recognition of Histone H3K36me3 Nucleosome by Human de Novo DNA Methyltransferases 3A and 3B.” *Journal of Structural Biology* 194(3): 357–67.

Rothbart, Scott B. et al. 2012. “Association of UHRF1 with H3K9 Methylation Directs the Maintenance of DNA Methylation.” *Nature structural & molecular biology* 19: 1155–60.

Ruiz-Alonso, Maria, David Blesa, and Carlos Simon. 2012. “The Genomics of the Human Endometrium.” *Biochimica et Biophysica Acta - Molecular Basis of Disease* 1822(12): 1931–42.

Saare, Merli et al. 2016. “The Influence of Menstrual Cycle and Endometriosis on Endometrial Methylome.” *Clinical Epigenetics*: 1–10.

Saha, Rama et al. 2015. “Heritability of Endometriosis.” *Fertility and Sterility* 104(4): 947–52.

Sakai, Nozomi et al. 2003. “Involvement of Histone Acetylation in Ovarian Steroid-Induced Decidualization of Human Endometrial Stromal Cells.” *The Journal of biological chemistry* 278(19): 16675–82.

Salker, Madhuri S. et al. 2017. “Loss of Endometrial Sodium Glucose Cotransporter SGLT1 Is Detrimental to Embryo Survival and Fetal Growth in Pregnancy.” *Scientific Reports* 7(1): 1–10.

Sampson, John A. 1927. “Metastatic or Embolic Endometriosis, Due to the Menstrual Dissemination of Endometrial Tissue into the Venous Circulation*.” *The American journal of pathology* 3(2): 93–110.43.

Satokata, Ichiro, Gall Benson, and Richard Maast. 1995. “Sexually Dimorphic Sterility Phenotypes in Hoxa10-Deficient Mice.” *Nature* 374: 460–63.

Scotchie, Jessica G. et al. 2009. “Proteomic Analysis of the Luteal Endometrial Secretome.” *Reproductive Sciences* 16(9): 883–93.

Segovia, Mercedes et al. 2019. “Targeting TMEM176B Enhances Antitumor Immunity and Augments the Efficacy of Immune Checkpoint Blockers by Unleashing Inflammasome Activation.” *Cancer Cell* 35(5): 767–81.

Senner, Claire E. et al. 2012. “DNA Methylation Profiles Define Stem Cell Identity and Reveal a Tight Embryonic-Extraembryonic Lineage Boundary.” *Stem Cells* 30(12): 2732–45.

Shen, Fanghua, Xishi Liu, Jian-Guo Geng, and Sun Wei Guo. 2009. “Increased Immunoreactivity to SLIT/ROBO1 in Ovarian Endometriomas: A Likely Constituent Biomarker for Recurrence.” *The American journal of pathology* 175(2): 479–88.

Sherwin, J. R A et al. 2008. “Global Gene Analysis of Late Secretory Phase, Eutopic Endometrium Does Not Provide the Basis for a Minimally Invasive Test of Endometriosis.” *Human Reproduction* 23(5): 1063–68.

———. 2010. “The Endometrial Response to Chorionic Gonadotropin Is Blunted in a Baboon Model of Endometriosis.” *Endocrinology* 151(10): 4982–93.

Shpargel, Karl B et al. 2014. “KDM6 Demethylase Independent Loss of Histone H3 Lysine

27 Trimethylation during Early Embryonic Development." *PLoS genetics* 10(8): e1004507.

Simmen, Rosalia C M et al. 2004. "Subfertility , Uterine Hypoplasia , and Partial Progesterone Resistance in Mice Lacking the Kruppel-like Factor 9 / Basic Transcription Element-Binding Protein-1 ( Bteb1 ) Gene *." *The Journal of biological chemistry* 279(28): 29286–94.

Smith, Zachary D, and Alexander Meissner. 2013. "DNA Methylation: Roles in Mammalian Development." *Nature Reviews Genetics* 14(3): 204–20.

Song, F. et al. 2005. "Association of Tissue-Specific Differentially Methylated Regions (TDMs) with Differential Gene Expression." *Proceedings of the National Academy of Sciences* 102(9): 3336–41.

Song, Yong et al. 2016. "Activated Hippo/Yes-Associated Protein Pathway Promotes Cell Proliferation and Anti-Apoptosis in Endometrial Stromal Cells of Endometriosis." *The Journal of Clinical Endocrinology & Metabolism* 101(4): 2016–1120.

Soufi, Abdenour, Greg Donahue, and Kenneth S Zaret. 2012. "Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome." *Cell* 151(5): 994–1004.

Sourial, Samer, Nicola Tempest, and Dharani K Hapangama. 2014. "Theories on the Pathogenesis of Endometriosis." *International Journal of Reproductive Medicine* 2014: 1–9.

Spainhour, John CG, Hong Seo Lim, Soojin V Yi, and Peng Qiu. 2019. "Correlation Patterns Between DNA Methylation and Gene Expression in The Cancer Genome Atlas." *Cancer Informatics* 18: 1–11.

Sperber, Henrik et al. 2015. "The Metabolome Regulates the Epigenetic Landscape during Naïve to Primed Human Embryonic Stem Cell Transition." *Nature Cell Biology* 17(12): 1523–35.

Stefansson, H. et al. 2002. "Genetic Factors Contribute to the Risk of Developing Endometriosis." *Human Reproduction* 17(3): 555–59.

Strakova, Zuzana, Szilvia Kruss, Kirsten Morris, and Jen Reed. 2010. "Members of the Hippo Pathway Are Regulated in the Uterus During the Menstrual Cycle." *Biology of Reproduction* 83: 363–363.

Strakova, Zuzana, Jennifer Reed, and Ivanna Ihnatovych. 2010. "Human Transcriptional Coactivator with PDZ-Binding Motif (TAZ) Is Downregulated during Decidualization." *Biology of reproduction* 82(6): 1112–18.

Struck, Anna Winona, Mark L. Thompson, Lu Shin Wong, and Jason Micklefield. 2012. "S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications." *ChemBioChem* 13(18): 2642–55.

Stubbs, Thomas M et al. 2017. "Multi-Tissue DNA Methylation Age Predictor In Mouse." *Genome Biology* 18(68): 1–14.

Suetake, Isao et al. 2004. "DNMT3L Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction." *Journal of Biological Chemistry* 279(26): 27816–23.

Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. "Revigo Summarizes and Visualizes Long Lists of Gene Ontology Terms." *PLoS ONE* 6(7).

Surani, M A H, S C Barton, and M L Norris. 1984. "Development of Reconstituted Mouse Eggs Suggests Imprinting of the Genome during Gametogenesis." *Nature* 308: 548–50.

Swain, Judith L., Timothy A. Stewart, and Philip Leder. 1987. "Parental Legacy Determines Methylation and Expression of an Autosomal Transgene: A Molecular Mechanism for Parental Imprinting." *Cell* 50(5): 719–27.

Szulwach, Keith E et al. 2011. "5-HmC–Mediated Epigenetic Dynamics during Postnatal Neurodevelopment and Aging." *Nature Neuroscience* 14(12): 1607–16.

Tahiliani, Mamta et al. 2009. "Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1." *Science* 15: 930–35.

Takano, Masashi et al. 2007. "Transcriptional Cross Talk between the Forkhead Transcription Factor Forkhead Box O1A and the Progesterone Receptor Coordinates Cell Cycle Regulation and Differentiation in Human Endometrial Stromal Cells." *Molecular Endocrinology* 21(10): 2334–49.

Talbi, S. et al. 2006. "Molecular Phenotyping of Human Endometrium Distinguishes Menstrual Cycle Phases and Underlying Biological Processes in Normo-Ovulatory

Women." *Endocrinology* 147(3): 1097–1121.

Tamm, Karin, Miia Rõõm, Andres Salumets, and Madis Metsis. 2009. "Genes Targeted by the Estrogen and Progesterone Receptors in the Human Endometrial Cell Lines HEC1A and RL95-2." *Reproductive Biology and Endocrinology* 7: 1–12.

Tamura, I et al. 2014. "Genome-Wide Analysis of Histone Modifications in Human Endometrial Stromal Cells." *Molecular Endocrinology* 28(10): 1656–69.

Tan, J. et al. 2018. "The Role of the Endometrial Receptivity Array (ERA) in Patients Who Have Failed Euploid Embryo Transfers." *Journal of Assisted Reproduction and Genetics* 35(4): 683–92.

Tapia, Alejandro et al. 2008. "Differences in the Endometrial Transcript Profile during the Receptive Period between Women Who Were Refractory to Implantation and Those Who Achieved Pregnancy." *Human Reproduction* 23(2): 340–51.

Taudt, Aaron et al. 2016. "ChromstaR: Tracking Combinatorial Chromatin State Dynamics in Space and Time." *bioRxiv*: 038612.

Taylor, H. S., D Olive, and A Arici. 1999. "HOXA10 Gene Expression Is Altered in the Endometrium of Patients with Endometriosis." *Journal of the Society for Gynecologic Investigation* 5(5): 111A–111A.

Taylor, H S, Aydin Arici, David Olive, and Peter Igarashi. 1998. "HOXA10 Is Expressed in Response to Sex Steroids at the Time of Implantation in the Human Endometrium." *The Journal of clinical investigation* 101(7): 1379–84.

Teklenburg, Gijs et al. 2010. "Natural Selection of Human Embryos: Decidualizing Endometrial Stromal Cells Serve as Sensors of Embryo Quality upon Implantation." *PLoS ONE* 5(4): 2–7.

Thie, Michael, and Hans Werner Denker. 2002. "In Vitro Studies on Endometrial Adhesiveness for Trophoblast: Cellular Dynamics in Uterine Epithelial Cells." *Cells Tissues Organs* 172(3): 237–52.

Tirado-Magallanes, Roberto et al. 2017. "Whole Genome DNA Methylation: Beyond Genes Silencing." *Oncotarget* 8(3): 5629–37.

Tissot, M. et al. 2017. "Clinical Presentation of Endometriosis Identified at Interval Laparoscopic Tubal Sterilization: Prospective Series of 465 Cases." *Journal of*

*Gynecology Obstetrics and Human Reproduction* 46(8): 647–50.

Tomida, Mikio et al. 2008. "Stat3 Up-Regulates Expression of Nicotinamide N-Methyltransferase in Human Cancer Cells." *Journal of Cancer Research and Clinical Oncology* 134(5): 551–59.

Treloar, Susan A., Daniel T. O'Connor, Vivienne M. O'Connor, and Nicholas G. Martin. 1999. "Genetic Influences on Endometriosis in an Australian Twin Sample." *Fertility and Sterility* 71(4): 701–10.

Turco, Margherita Y. et al. 2017. "Long-Term, Hormone-Responsive Organoid Cultures of Human Endometrium in a Chemically Defined Medium." *Nature Cell Biology* 19(5): 568–77.

Uchida, Hiroshi et al. 2005. "Histone Deacetylase Inhibitors Induce Differentiation of Human Endometrial Adenocarcinoma Cells through Up-Regulation of Glycodelin." *Endocrinology* 146(12): 5365–73.

Ulanovskaya, Olesya A., Andrea M. Zuhl, and Benjamin F. Cravatt. 2013. "NNMT Promotes Epigenetic Remodeling in Cancer by Creating a Metabolic Methylation Sink." *Nature Chemical Biology* 18(3): 386–92.

Valdes, Cecilia T., Amy Schutt, and Carlos Simon. 2017. "Implantation Failure of Endometrial Origin: It Is Not Pathology, but Our Failure to Synchronize the Developing Embryo with a Receptive Endometrium." *Fertility and Sterility* 108(1): 15–18.

Valinluck, Victoria et al. 2004. "Oxidative Damage to Methyl-CpG Sequences Inhibits the Binding of the Methyl-CpG Binding Domain (MBD) of Methyl-CpG Binding Protein 2 (MeCP2)." *Nucleic Acids Research* 32(14): 4100–4108.

Vasquez, Yasmin M. et al. 2015. "FOXO1 Is Required for Binding of PR on *IRF4* , Novel Transcriptional Regulator of Endometrial Stromal Decidualization." *Molecular Endocrinology* 29(3): 421–33.

Vercellini, Paolo et al. 2009. "Surgery for Endometriosis-Associated Infertility: A Pragmatic Approach." *Human Reproduction* 24(2): 254–69.

Vercellini, Paolo, Paola Viganò, Edgardo Somigliana, and Luigi Fedele. 2014. "Endometriosis: Pathogenesis and Treatment." *Nature Reviews Endocrinology* 10(5): 261–75.

Vilain, Armelle, Nicolas Vogt, Bernard Dutrillaux, and Bernard Malfoy. 1999. "DNA Methylation and Chromosome Instability in Breast Cancer Cell Lines." *FEBS Letters* 460(2): 231–34.

Vilella, Felipe et al. 2015. "Hsa-MiR-30d, Secreted by the Human Endometrium, Is Taken up by the Pre-Implantation Embryo and Might Modify Its Transcriptome." *Development* 142(18): 3210–21.

Vinatier, D., G. Orazi, M. Cosson, and P. Dufour. 2001. "Theories of Endometriosis." *European Journal of Obstetrics Gynecology and Reproductive Biology* 96: 21–34.

Vincent, Zoe L, Cindy M Farquhar, Murray D Mitchell, and Anna P Ponnampalam. 2011. "Expression and Regulation of DNA Methyltransferases in Human Endometrium." *Fertility and sterility* 95(4): 2–6.

Weber, Alain R. et al. 2016. "Biochemical Reconstitution of TET1-TDG-BER-Dependent Active DNA Demethylation Reveals a Highly Coordinated Mechanism." *Nature Communications* 7: 1–13.

Wei, Qingxiang et al. 2010. "Reduced Expression of Biomarkers Associated with the Implantation Window in Women with Endometriosis." *Fertility and Sterility* 91(5): 1686–91.

Whitby, Sarah, Lois A Salamonsen, and Jemma Evans. 2018. "The Endometrial Polarity Paradox: Differential Regulation of Polarity Within Secretory-Phase Human Endometrium." *Endocrinology* 159(1): 506–18.

Wölfler, Monika Martina et al. 2016. "Altered Expression of Progesterone Receptor Isoforms A and B in Human Eutopic Endometrium in Endometriosis Patients." *Annals of Anatomy* 206: 1–6.

Wu, Yan et al. 2005. "Aberrant Methylation at HOXA10 May Be Responsible for Its Aberrant Expression in the Endometrium of Patients with Endometriosis." *American Journal of Obstetrics and Gynecology* 193: 371–80.

———. 2006. "Promoter Hypermethylation of Progesterone Receptor Isoform B (PR-B) in Endometriosis." *Epigenetics* 1(2): 106–11.

———. 2007. "Aberrant Expression of Deoxyribonucleic Acid Methyltransferases DNMT1, DNMT3A, and DNMT3B in Women with Endometriosis." *Fertility and sterility* 87(1): 24–32.

Xiaomeng, Xia, Zhao Ming, Ma Jiezhi, and Fang Xiaoling. 2013. "Aberrant Histone Acetylation and Methylation Levels in Woman with Endometriosis." *Archives of gynecology and obstetrics* 287(3): 487–94.

Xue, Qing, Zhihong Lin, You-Hong Cheng, et al. 2007. "Promoter Methylation Regulates Estrogen Receptor 2 in Human Endometrium and Endometriosis." *Biology of reproduction* 77(4): 681–87.

Xue, Qing, Zhihong Lin, Ping Yin, et al. 2007. "Transcriptional Activation of Steroidogenic Factor-1 by Hypomethylation of the 5′ CpG Island in Endometriosis." *Journal of Clinical Endocrinology and Metabolism* 92(8): 3261–67.

Yamagata, Yoshiaki et al. 2009. "DNA Methyltransferase Expression in the Human Endometrium: Down- Regulation by Progesterone and Estrogen." *Human Reproduction* 24(5): 1126–32.

———. 2014. "Genome-Wide DNA Methylation Profiling in Cultured Eutopic and Ectopic Endometrial Stromal Cells." *PloS one* 9(1): e83612.

Yamaguchi, Hirohito et al. 2018. "EZH2 Contributes to the Response to PARP Inhibitors through Its PARP-Mediated Poly-ADP Ribosylation in Breast Cancer." *Oncogene* 37(2): 208–17.

Yanaihara, Atsushi et al. 2004. "Comparison in Gene Expression of Secretory Human Endometrium Using Laser Microdissection." *Reproductive biology and endocrinology* 2(1): 66.

Yoh, Sunnie M., Joseph S. Lucas, and Katherine A. Jones. 2008. "The Iws1:Spt6:CTD Complex Controls Cotranscriptional MRNA Biosynthesis and HYPB/Setd2-Mediated Histone H3K36 Methylation." *Genes and Development* 22(24): 3422–34.

Yong, Wai-Shin, Fei-Man Hsu, and Pao-Yang Chen. 2016. "Profiling Genome-Wide DNA Methylation." *Epigenetics & chromatin* 9: 26.

Young, Matthew D. et al. 2011. "ChIP-Seq Analysis Reveals Distinct H3K27me3 Profiles That Correlate with Transcriptional Activity." *Nucleic Acids Research* 39(17): 7415–27.

Yu, Fa Xing, and Kun Liang Guan. 2013. "The Hippo Pathway: Regulators and Regulations." *Genes and Development* 27(4): 355–71.

Zhang, Daying et al. 2002. “Direct Interaction of the Kruppel-like Family (KLF) Member, BTEB1, and PR Mediated Progesterone- Responsive Gene Expression in Endometrial.” *Endocrinology* 143(1): 62–73.

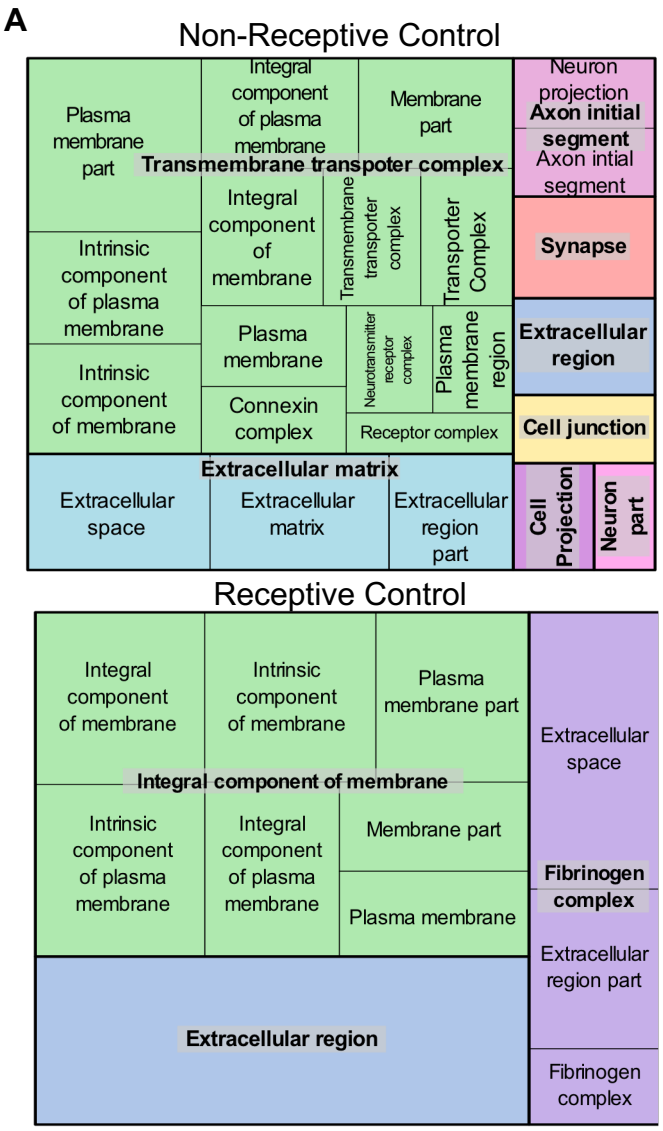
Zhang, Qi et al. 2017. “Enhancer of Zeste Homolog 2 (EZH2) Induces Epithelial-Mesenchymal Transition in Endometriosis.” *Scientific Reports* 7(1): 6804.

Zhang, Wu, and Jie Xu. 2017. “DNA Methyltransferases and Their Roles in Tumorigenesis.” *Biomarker Research* 5(1): 1–8.

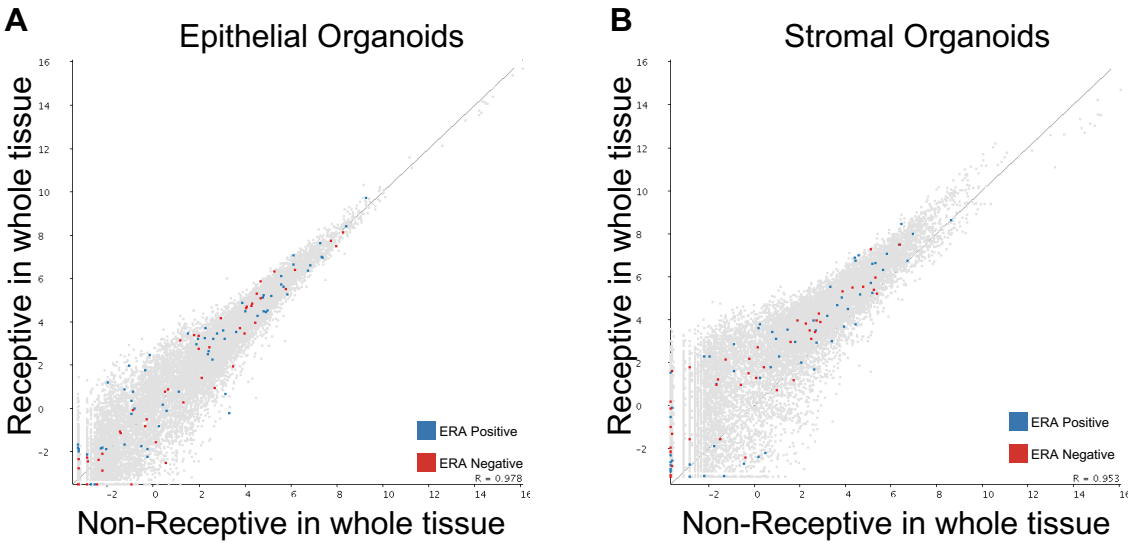
Zheng, Longbo et al. 2018. “ELF3 Promotes Epithelial – Mesenchymal Transition by Protecting ZEB1 from MiR-141-3p-Mediated Silencing in Hepatocellular Carcinoma.” *Cell Death and Disease* 9(3): 387.

# Appendix

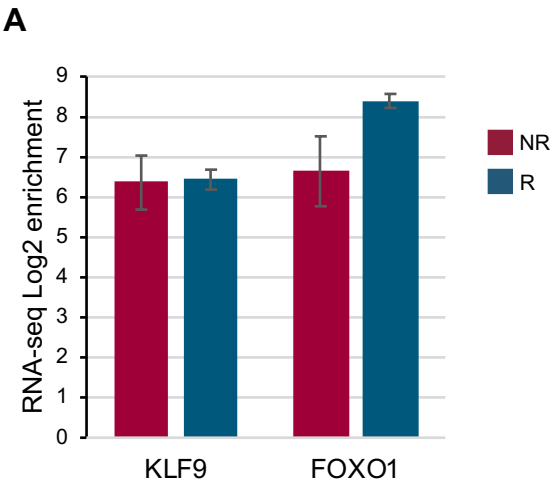
## Supplementary Figures



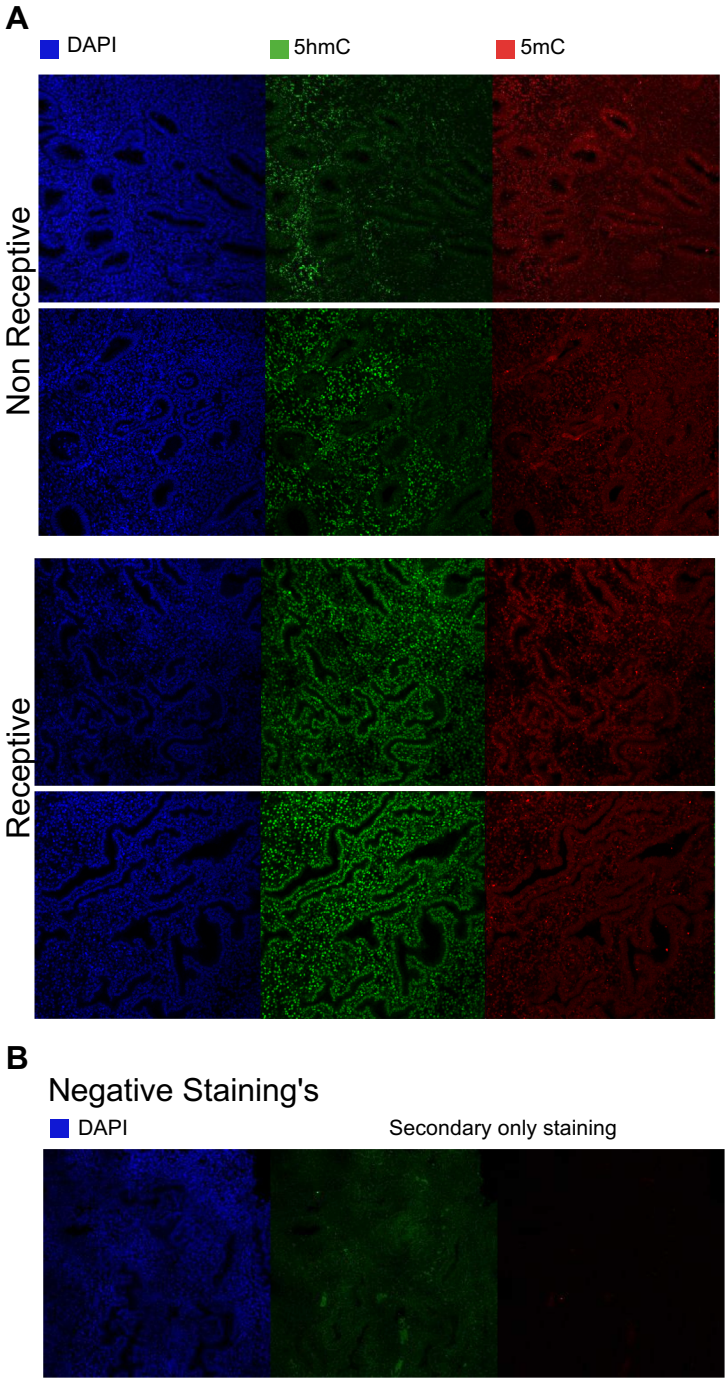

**Supplementary Figure 1 Component Gene Ontology analysis of the receptivity signature. A**, Gene Ontology component treemap of enriched genes identified in Figure 3.3A using GOrilla and REVIGO tools with background list of genes with >20reads and Bonferroni $p<0.05$ applied.



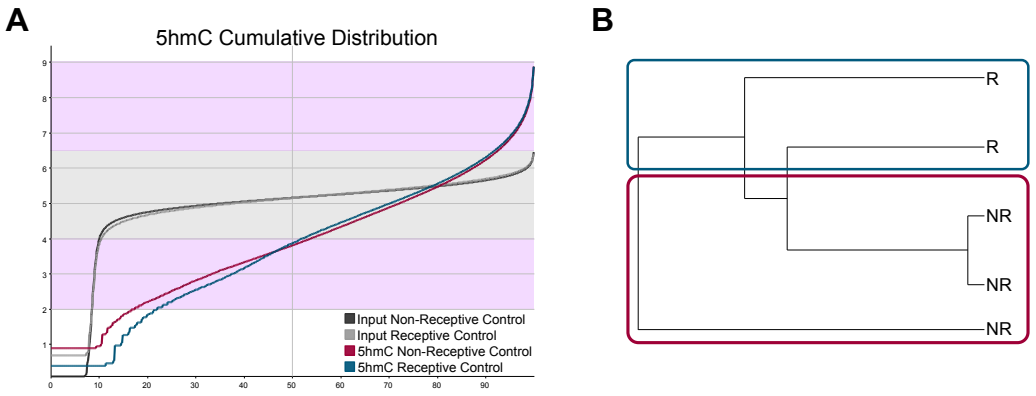

**Supplementary Figure 2 Receptivity of the organoids. A** and **B**, Scatterplot of RNA-seq data comparing epithelial organoids (A) or stromal organoids (B) with the receptivity state established using the corresponding endometrial whole tissue. Each dot represents a gene with blue dots corresponding to the positive genes and marron dots corresponding to the negative genes in the endometrial receptivity array.



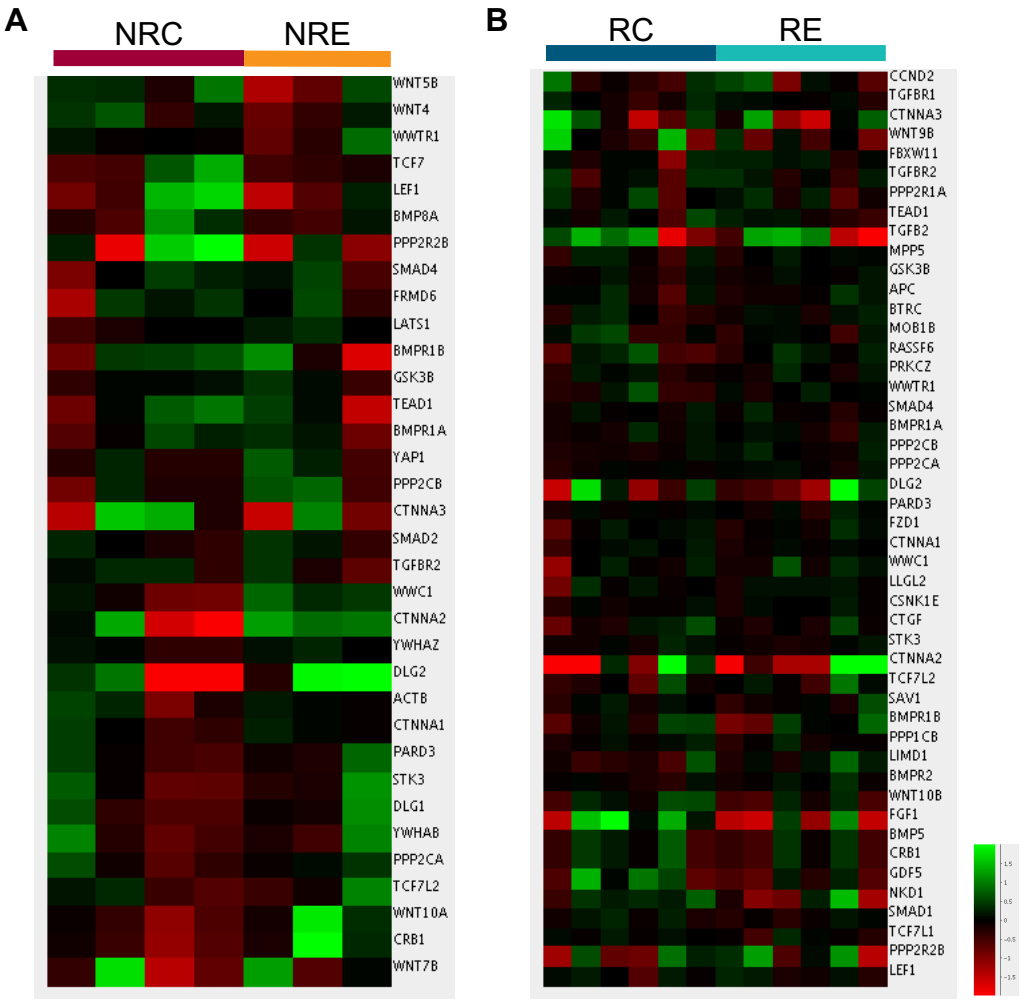

**Supplementary Figure 3 KLF9 and FOXO1 Expression. A**, RNA-seq expression levels (RPM log2) in whole endometrial tissue of *KLF9* and *FOXO1*. Plots depict mean +/- SEM. NR = non-receptive state (n=4), R= receptive state (n=6).




**Supplementary Figure 4 Endometrial immunofluorescent images for 5mC, 5hmC and negative stainings. A**, Immunofluorescence staining of endometrial tissue in the non-receptive and receptive state with 5hmC (green), 5mC (red) and DAPI nuclear counterstain (blue). **B**, Negative stainings of the endometrium with only secondaries and DAPI nuclear counterstain (blue).

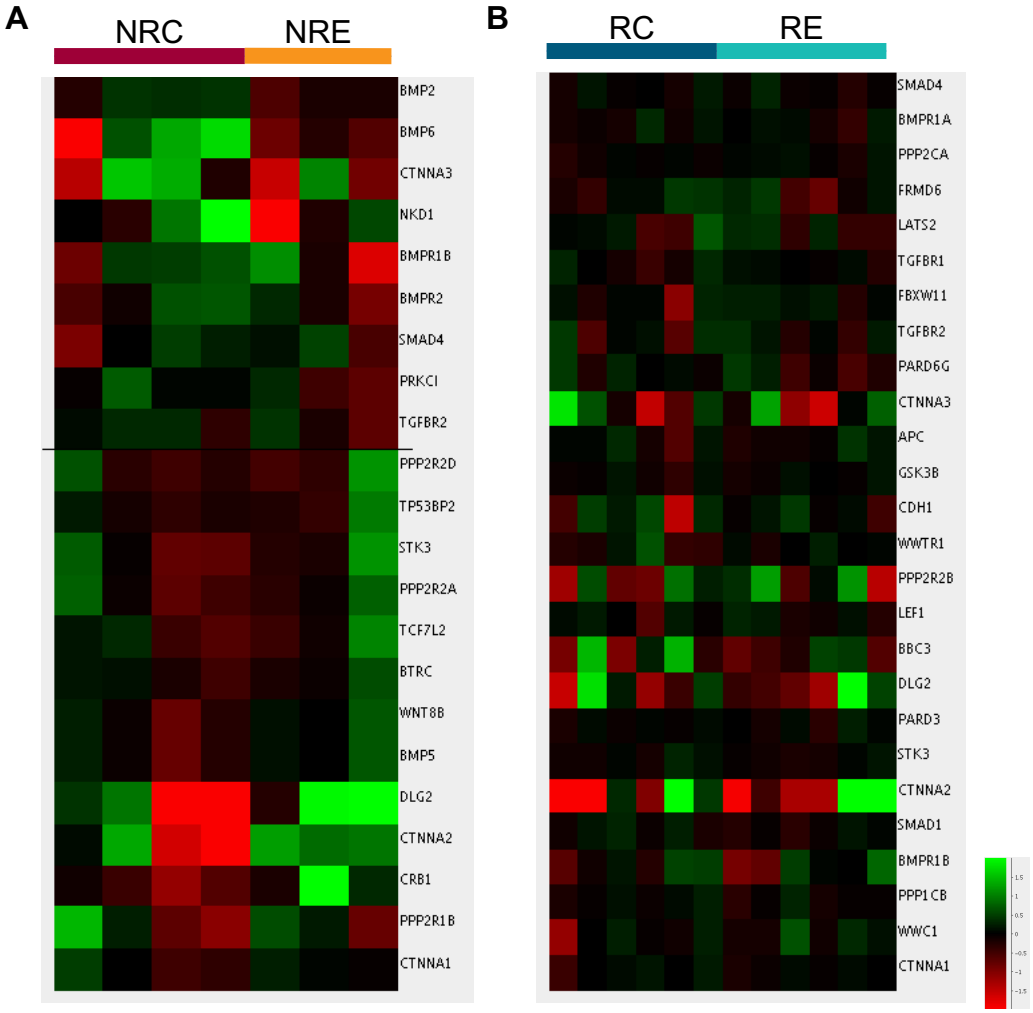



**Supplementary Figure 5 5hmC of LCM endometrial epithelial cells A**, Cumulative distribution graph of Input and HMCP data, using 5kb tiling probes across genome spaced 2.5kb apart. Data are corrected to largest dataset, log2 transformed. Grey box illustrates inputs flat distribution. Violet boxes illustrate HMCPs dynamic distribution. **B**, Neighbour joining tree, with distances calculated by Pearson correlation, showing the degree of relatedness between epithelial cells in either the non-receptive or receptive state at CpG Islands.




**Supplementary Figure 6 The expression of differentially methylated Hippo genes. A** and **B**, Heatmap of the expression of hippo genes that were differentially methylated between controls and endometriosis samples in the non-receptive (A) or receptive (B) state. NRC =

non-receptive control, NRE = non-receptive endometriosis, RC = receptive control, RE = receptive endometriosis.




**Supplementary Figure 7 The expression of differentially hydroxymethylated Hippo genes. A** and **B**, Heatmap of the expression of hippo genes that were differentially hydroxymethylated between controls and endometriosis samples in the non-receptive (A) or receptive (B) state. NRC = non-receptive control, NRE = non-receptive endometriosis, RC = receptive control, RE = receptive endometriosis.

## Supplementary Tables

**Supplementary Table 1 Endometrial Receptivity Transcriptome (ERT) genes.** List of the genes in the ERT signature that was developed using non-receptive and receptive endometrial whole tissue samples. Genes marked as non-receptive are upregulated in the non-receptive state and genes marked receptive are upregulated in the receptive state.

| Gene | Chr | Start | End | Upregulated in: | Description |
|---|---|---|---|---|---|
| RBP7 | 1 | 9997216 | 10016020 | non-receptive | retinol binding protein 7, cellular |
| MFAP2 | 1 | 16974502 | 16980835 | non-receptive | microfibrillar-associated protein 2 |
| ALPL | 1 | 21509372 | 21578412 | non-receptive | alkaline phosphatase, liver/bone/kidney |
| FNDC5 | 1 | 32862268 | 32872480 | non-receptive | fibronectin type III domain containing 5 |
| GJA4 | 1 | 34792998 | 34795747 | non-receptive | gap junction protein, alpha 4, 37kDa |
| DAB1 | 1 | 56994778 | 57424057 | non-receptive | Dab, reelin signal transducer, homolog 1 (Drosophila) |
| MCOLN3 | 1 | 85018082 | 85048470 | non-receptive | mucolipin 3 |
| NTNG1 | 1 | 107140007 | 107483458 | non-receptive | netrin G1 |
| PHGDH | 1 | 119711887 | 119744215 | non-receptive | phosphoglycerate dehydrogenase |
| ECM1 | 1 | 150508076 | 150513789 | non-receptive | extracellular matrix protein 1 |
| CRABP2 | 1 | 156699606 | 156705816 | non-receptive | cellular retinoic acid binding protein 2 |
| ATP1B1 | 1 | 169105697 | 169132722 | non-receptive | ATPase, Na+/K+ transporting, beta 1 polypeptide |
| VASH2 | 1 | 212950520 | 212991585 | non-receptive | vasohibin 2 |
| GREM2 | 1 | 240489573 | 240612149 | non-receptive | gremlin 2, DAN family BMP antagonist |
| RGS7 | 1 | 240775514 | 241357230 | non-receptive | regulator of G-protein signalling 7 |
| KMO | 1 | 241532134 | 241595642 | non-receptive | kynurenine 3-monooxygenase (kynurenine 3- hydroxylase) |
| NRXN1 | 2 | 49918505 | 51032536 | non-receptive | neurexin 1 |
| CTNNA2 | 2 | 79185231 | 80648861 | non-receptive | catenin (cadherin-associated protein), alpha 2 |
| PROM2 | 2 | 95274453 | 95291308 | non-receptive | prominin 2 |

| | | | | | |
|---|---|---|---|---|---|
| NR4A2 | 2 | 156324432 | 156342348 | non-receptive | nuclear receptor subfamily 4, group A, member 2 |
| ERMN | 2 | 157318625 | 157327713 | non-receptive | ermin, ERM-like protein |
| B3GALT1 | 2 | 167868948 | 167874041 | non-receptive | UDP-Gal:betaGlcNAc beta 1,3- galactosyltransferase, polypeptide 1 |
| MAP2 | 2 | 209424058 | 209734118 | non-receptive | microtubule-associated protein 2 |
| LRRN1 | 3 | 3799437 | 3847703 | non-receptive | leucine rich repeat neuronal 1 |
| PFKFB4 | 3 | 48517684 | 48561172 | non-receptive | 6-phosphofructo-2-kinase/fructose-2,6- biphosphatase 4 |
| SEMA3G | 3 | 52433053 | 52445085 | non-receptive | sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3G |
| PRICKLE2- | 3 | 64067964 | 64103131 | non-receptive | PRICKLE2 antisense RNA 1 |
| ROBO2 | 3 | 75906695 | 77649964 | non-receptive | roundabout, axon guidance receptor, homolog 2 (Drosophila) |
| STXBP5L | 3 | 120908072 | 121424761 | non-receptive | syntaxin binding protein 5-like |
| SLC15A2 | 3 | 121894089 | 121944102 | non-receptive | solute carrier family 15 (oligopeptide transporter), member 2 |
| PODXL2 | 3 | 127629181 | 127672809 | non-receptive | podocalyxin-like 2 |
| COL6A6 | 3 | 130560334 | 130678155 | non-receptive | collagen, type VI, alpha 6 |
| RAB6B | 3 | 133824239 | 133895836 | non-receptive | RAB6B, member RAS oncogene family |
| EPHB1 | 3 | 134615365 | 135260467 | non-receptive | EPH receptor B1 |
| VEPH1 | 3 | 157259742 | 157533619 | non-receptive | ventricular zone expressed PH domain- containing 1 |
| EPHB3 | 3 | 184561784 | 184582409 | non-receptive | EPH receptor B3 |
| FGF12 | 3 | 192139395 | 192767764 | non-receptive | fibroblast growth factor 12 |
| FGFR3 | 4 | 1793307 | 1808872 | non-receptive | fibroblast growth factor receptor 3 |
| PPP2R2C | 4 | 6320578 | 6563600 | non-receptive | protein phosphatase 2, regulatory subunit B, gamma |
| ERVMER34-1 | 4 | 52742552 | 52751640 | non-receptive | endogenous retrovirus group MER34, member 1 |
| RASL11B | 4 | 52862290 | 52866835 | non-receptive | RAS-like, family 11, member B |
| EPHA5 | 4 | 65319563 | 65670495 | non-receptive | EPH receptor A5 |
| FRAS1 | 4 | 78057570 | 78544269 | non-receptive | Fraser extracellular matrix complex subunit 1 |
| SCD5 | 4 | 82629539 | 82798857 | non-receptive | stearoyl-CoA desaturase 5 |

| | | | | | |
|---|---|---|---|---|---|
| BMPR1B-AS1 | 4 | 94743800 | 94757533 | non-receptive | BMPR1B antisense RNA 1 (head to head) |
| PCDH10 | 4 | 133149315 | 133194700 | non-receptive | protocadherin 10 |
| PABPC4L | 4 | 134196333 | 134201748 | non-receptive | poly(A) binding protein, cytoplasmic 4-like |
| LRAT | 4 | 154626961 | 154753118 | non-receptive | lecithin retinol acyltransferase (phosphatidylcholine--retinol O-acyltransferase) |
| SORBS2 | 4 | 185585444 | 185956652 | non-receptive | sorbin and SH3 domain containing 2 |
| PLEKHG4B | 5 | 140258 | 189970 | non-receptive | pleckstrin homology domain containing, family G (with RhoGef domain) member 4B |
| ADAMTS16 | 5 | 5140330 | 5320304 | non-receptive | ADAM metallopeptidase with thrombospondin type 1 motif, 16 |
| HCN1 | 5 | 45259247 | 45696151 | non-receptive | hyperpolarization activated cyclic nucleotide-gated potassium channel1 |
| RGS7BP | 5 | 64506299 | 64612312 | non-receptive | regulator of G-protein signaling 7 binding protein |
| FAM169A | 5 | 74777574 | 74866951 | non-receptive | family with sequence similarity 169, member A |
| THBS4 | 5 | 80035348 | 80083287 | non-receptive | thrombospondin 4 |
| SLC27A6 | 5 | 128538013 | 129033642 | non-receptive | solute carrier family 27 (fatty acid transporter), member 6 |
| NRG2 | 5 | 139846779 | 140043299 | non-receptive | neuregulin 2 |
| SH3RF2 | 5 | 145936579 | 146081791 | non-receptive | SH3 domain containing ring finger 2 |
| SGCD | 5 | 155870344 | 156767788 | non-receptive | sarcoglycan, delta (35kDa dystrophin- associated glycoprotein) |
| KCNIP1 | 5 | 170353487 | 170736632 | non-receptive | Kv channel interacting protein 1 |
| MSX2 | 5 | 174724533 | 174730893 | non-receptive | msh homeobox 2 |
| NEDD9 | 6 | 11183298 | 11382348 | non-receptive | neural precursor cell expressed, developmentally down-regulated 9 |
| HLA-DOB | 6 | 32812763 | 32817048 | non-receptive | major histocompatibility complex, class II, DO beta |
| SPDEF | 6 | 34537802 | 34556333 | non-receptive | SAM pointed domain containing ETS transcription factor |
| HEY2 | 6 | 125747664 | 125761269 | non-receptive | hes-related family bHLH transcription factor with YRPW motif 2 |
| EPB41L2 | 6 | 130839347 | 131063322 | non-receptive | erythrocyte membrane protein band 4.1-like 2 |
| FNDC1 | 6 | 159169397 | 159272109 | non-receptive | fibronectin type III domain containing 1 |
| PRR15 | 7 | 29563811 | 29567295 | non-receptive | proline rich 15 |

| | | | | | |
|---|---|---|---|---|---|
| WIPF3 | 7 | 29806486 | 29917066 | non-receptive | WAS/WASL interacting protein family, member 3 |
| ADCYAP1R1 | 7 | 31052461 | 31111479 | non-receptive | adenylate cyclase activating polypeptide 1 (pituitary) receptor type I |
| SFRP4 | 7 | 37905932 | 38025695 | non-receptive | secreted frizzled-related protein 4 |
| VWC2 | 7 | 49773661 | 49921950 | non-receptive | von Willebrand factor C domain containing 2 |
| SEMA3C | 7 | 80742538 | 80922359 | non-receptive | sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C |
| SEMA3D | 7 | 84995553 | 85186855 | non-receptive | sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3D |
| COL1A2 | 7 | 94394561 | 94431232 | non-receptive | collagen, type I, alpha 2 |
| DYNC1I1 | 7 | 95772506 | 96110071 | non-receptive | dynein, cytoplasmic 1, intermediate chain 1 |
| LRRC17 | 7 | 102912991 | 102944949 | non-receptive | leucine rich repeat containing 17 |
| NRCAM | 7 | 108147626 | 108456717 | non-receptive | neuronal cell adhesion molecule |
| ATP6V0A4 | 7 | 138706295 | 138798196 | non-receptive | ATPase, H+ transporting, lysosomal V0 subunit a4 |
| TMEM213 | 7 | 138797952 | 138838101 | non-receptive | transmembrane protein 213 |
| SOX7 | 8 | 10723768 | 10730512 | non-receptive | SRY (sex determining region Y)-box 7 |
| NEFM | 8 | 24913012 | 24919098 | non-receptive | neurofilament, medium polypeptide |
| NEFL | 8 | 24950955 | 24957110 | non-receptive | neurofilament, light polypeptide |
| SFRP1 | 8 | 41261958 | 41309497 | non-receptive | secreted frizzled-related protein 1 |
| SOX17 | 8 | 54457935 | 54460888 | non-receptive | SRY (sex determining region Y)-box 17 |
| CRISPLD1 | 8 | 74984515 | 75034558 | non-receptive | cysteine-rich secretory protein LCCL domain containing 1 |
| SLC26A7 | 8 | 91209495 | 91398152 | non-receptive | solute carrier family 26 (anion exchanger), member 7 |
| PIP5K1B | 9 | 68705414 | 69009175 | non-receptive | phosphatidylinositol-4-phosphate 5-kinase, type I, beta |
| FBP1 | 9 | 94603133 | 94640249 | non-receptive | fructose-1,6-bisphosphatase 1 |
| GALNT12 | 9 | 98807699 | 98850081 | non-receptive | polypeptide N- acetylgalactosaminyltransferase 12 |
| GRIN3A | 9 | 101569353 | 101738580 | non-receptive | glutamate receptor, ionotropic, N-methyl-D- aspartate 3A |
| FRRS1L | 9 | 109130293 | 109167291 | non-receptive | ferric-chelate reductase 1-like |

| | | | | | |
|---|---|---|---|---|---|
| RNF183 | 9 | 113297093 | 113303271 | non-receptive | ring finger protein 183 |
| OLFM1 | 9 | 135075422 | 135121179 | non-receptive | olfactomedin 1 |
| SAPCD2 | 9 | 137062124 | 137070588 | non-receptive | suppressor APC domain containing 2 |
| FAM13C | 10 | 59246129 | 59363181 | non-receptive | family with sequence similarity 13, member C |
| ANK3 | 10 | 60026298 | 60733490 | non-receptive | ankyrin 3, node of Ranvier (ankyrin G) |
| TMEM26 | 10 | 61406647 | 61453450 | non-receptive | transmembrane protein 26 |
| UNC5B | 10 | 71212570 | 71302864 | non-receptive | unc-5 homolog B (C. elegans) |
| KAZALD1 | 10 | 101061841 | 101065592 | non-receptive | Kazal-type serine peptidase inhibitor domain 1 |
| DMBT1 | 10 | 122560665 | 122643736 | non-receptive | deleted in malignant brain tumors 1 |
| ADAM12 | 10 | 126012381 | 126388455 | non-receptive | ADAM metallopeptidase domain 12 |
| B4GALNT4 | 11 | 369804 | 382116 | non-receptive | beta-1,4-N-acetyl-galactosaminyl transferase 4 |
| GAS2 | 11 | 22625642 | 22813055 | non-receptive | growth arrest-specific 2 |
| MPPED2 | 11 | 30384493 | 30586872 | non-receptive | metallophosphoesterase domain containing 2 |
| SLC43A1 | 11 | 57484534 | 57515786 | non-receptive | solute carrier family 43 (amino acid system L transporter), member 1 |
| RCOR2 | 11 | 63911221 | 63916844 | non-receptive | REST corepressor 2 |
| ANO1 | 11 | 70078302 | 70189528 | non-receptive | anoctamin 1, calcium activated chloride channel |
| PLEKHB1 | 11 | 73646178 | 73662819 | non-receptive | pleckstrin homology domain containing, family B (evectins) member 1 |
| TENM4 | 11 | 78652831 | 79440651 | non-receptive | teneurin transmembrane protein 4 |
| DLG2 | 11 | 83455012 | 85627270 | non-receptive | discs, large homolog 2 (Drosophila) |
| LOH12CR2 | 12 | 12355406 | 12357067 | non-receptive | loss of heterozygosity, 12, chromosomal region 2 (non-protein coding) |
| AQP5 | 12 | 49961870 | 49965681 | non-receptive | aquaporin 5 |
| SLC4A8 | 12 | 51391317 | 51515763 | non-receptive | solute carrier family 4, sodium bicarbonate cotransporter, member 8 |
| CSRP2 | 12 | 76858715 | 76879060 | non-receptive | cysteine and glycine-rich protein 2 |
| BTBD11 | 12 | 107318413 | 107659642 | non-receptive | BTB (POZ) domain containing 11 |
| GJB2 | 13 | 20187470 | 20192898 | non-receptive | gap junction protein, beta 2, 26kDa |
| GJB6 | 13 | 20221971 | 20232395 | non-receptive | gap junction protein, beta 6, 30kDa |

| | | | | | |
|---|---|---|---|---|---|
| CCNA1 | 13 | 36431830 | 36442882 | non-receptive | cyclin A1 |
| SIAH3 | 13 | 45777243 | 45851736 | non-receptive | siah E3 ubiquitin protein ligase family member 3 |
| SLITRK6 | 13 | 85792790 | 85799488 | non-receptive | SLIT and NTRK-like family, member 6 |
| STXBP6 | 14 | 24809656 | 25050297 | non-receptive | syntaxin binding protein 6 (amisyn) |
| LINC00645 | 14 | 27612588 | 27639636 | non-receptive | long intergenic non-protein coding RNA 645 |
| AKAP6 | 14 | 32329273 | 32837681 | non-receptive | A kinase (PRKA) anchor protein 6 |
| NPAS3 | 14 | 32934933 | 33804176 | non-receptive | neuronal PAS domain protein 3 |
| LINC00639 | 14 | 38749339 | 38948273 | non-receptive | long intergenic non-protein coding RNA 639 |
| DIO2 | 14 | 80197527 | 80387757 | non-receptive | deiodinase, iodothyronine, type II |
| SLC24A4 | 14 | 92322581 | 92501483 | non-receptive | solute carrier family 24 (sodium/potassium/calcium exchanger), member 4 |
| SERPINA5 | 14 | 94561442 | 94593120 | non-receptive | serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 5 |
| GREM1 | 15 | 32717974 | 32745107 | non-receptive | gremlin 1, DAN family BMP antagonist |
| SORD | 15 | 45023104 | 45077185 | non-receptive | sorbitol dehydrogenase |
| SLC47A1 | 17 | 19533462 | 19579034 | non-receptive | solute carrier family 47 (multidrug and toxin extrusion), member 1 |
| PRR15L | 17 | 47951967 | 47957878 | non-receptive | proline rich 15-like |
| KCNJ2 | 17 | 70168673 | 70180048 | non-receptive | potassium inwardly-rectifying channel, subfamily J, member 2 |
| SDK2 | 17 | 73334384 | 73644089 | non-receptive | sidekick cell adhesion molecule 2 |
| CLUL1 | 18 | 596988 | 650334 | non-receptive | clusterin-like 1 (retinal) |
| LAMA1 | 18 | 6941744 | 7117814 | non-receptive | laminin, alpha 1 |
| SERPINB5 | 18 | 63476761 | 63505085 | non-receptive | serpin peptidase inhibitor, clade B (ovalbumin), member 5 |
| CBLN2 | 18 | 72536680 | 72638517 | non-receptive | cerebellin 2 precursor |
| OLFM2 | 19 | 9853718 | 9936552 | non-receptive | olfactomedin 2 |
| CRLF1 | 19 | 18593237 | 18606850 | non-receptive | cytokine receptor-like factor 1 |
| C5AR2 | 19 | 47332147 | 47347327 | non-receptive | complement component 5a receptor 2 |

| | | | | | |
|---|---|---|---|---|---|
| DBP | 19 | 48630030 | 48637438 | non-receptive | D site of albumin promoter (albumin D-box) binding protein |
| LRRC4B | 19 | 50516892 | 50568045 | non-receptive | leucine rich repeat containing 4B |
| KLK4 | 19 | 50906352 | 50910738 | non-receptive | kallikrein-related peptidase 4 |
| RASSF2 | 20 | 4780023 | 4823645 | non-receptive | Ras association (RalGDS/AF-6) domain family member 2 |
| PAK7 | 20 | 9537389 | 9839041 | non-receptive | p21 protein (Cdc42/Rac)-activated kinase 7 |
| SYNDIG1 | 20 | 24469199 | 24666616 | non-receptive | synapse differentiation inducing 1 |
| KCNG1 | 20 | 51003656 | 51023129 | non-receptive | potassium voltage-gated channel, subfamily G, member 1 |
| CBLN4 | 20 | 55997440 | 56005472 | non-receptive | cerebellin 4 precursor |
| RBBP8NL | 20 | 62410237 | 62427533 | non-receptive | RBBP8 N-terminal like |
| MMP11 | 22 | 23772819 | 23784316 | non-receptive | matrix metallopeptidase 11 (stromelysin 3) |
| PHF21B | 22 | 44881162 | 45009999 | non-receptive | PHD finger protein 21B |
| MXRA5 | X | 3308565 | 3346641 | non-receptive | matrix-remodelling associated 5 |
| EGFL6 | X | 13569605 | 13633575 | non-receptive | EGF-like-domain, multiple 6 |
| HMGN5 | X | 81113701 | 81201942 | non-receptive | high mobility group nucleosome binding domain 5 |
| PCDH19 | X | 100291644 | 100410273 | non-receptive | protocadherin 19 |
| GLA | X | 101397803 | 101407925 | non-receptive | galactosidase, alpha |
| ARMCX4 | X | 101488044 | 101495811 | non-receptive | armadillo repeat containing, X-linked 4 |
| TCEAL5 | X | 103273691 | 103276872 | non-receptive | transcription elongation factor A (SII)-like 5 |
| AFF2 | X | 148500619 | 149000663 | non-receptive | AF4/FMR2 family, member 2 |
| MEGF6 | 1 | 3489920 | 3611495 | receptive | multiple EGF-like-domains 6 |
| DHRS3 | 1 | 12567910 | 12617731 | receptive | dehydrogenase/reductase (SDR family) member 3 |
| CDA | 1 | 20588948 | 20618908 | receptive | cytidine deaminase |
| HTR1D | 1 | 23191895 | 23194729 | receptive | 5-hydroxytryptamine (serotonin) receptor 1D, G protein-coupled |
| NCMAP | 1 | 24556111 | 24609325 | receptive | noncompact myelin associated protein |
| TRNP1 | 1 | 26993707 | 27000898 | receptive | TMF1-regulated nuclear protein 1 |
| TTC39A | 1 | 51287258 | 51345116 | receptive | tetratricopeptide repeat domain 39A |

| | | | | | |
|---|---|---|---|---|---|
| DNAJC6 | 1 | 65254462 | 65415869 | receptive | DnaJ (Hsp40) homolog, subfamily C, member 6 |
| GADD45A | 1 | 67685061 | 67688338 | receptive | growth arrest and DNA-damage-inducible, alpha |
| CHI3L2 | 1 | 111200771 | 111243440 | receptive | chitinase 3-like 2 |
| NPR1 | 1 | 153678637 | 153693992 | receptive | natriuretic peptide receptor 1 |
| EFNA1 | 1 | 155127460 | 155134857 | receptive | ephrin-A1 |
| ACKR1 | 1 | 159203307 | 159206500 | receptive | atypical chemokine receptor 1 (Duffy blood group) |
| CRP | 1 | 159712289 | 159714589 | receptive | C-reactive protein, pentraxin-related |
| C4BPA | 1 | 207104262 | 207144972 | receptive | complement component 4 binding protein, alpha |
| CD55 | 1 | 207321508 | 207360966 | receptive | CD55 molecule, decay accelerating factor for complement (Cromer blood group) |
| LAMB3 | 1 | 209614870 | 209652466 | receptive | laminin, beta 3 |
| G0S2 | 1 | 209675420 | 209676388 | receptive | G0/G1 switch 2 |
| TGFB2 | 1 | 218346235 | 218444619 | receptive | transforming growth factor, beta 2 |
| CAPN8 | 1 | 223538008 | 223665734 | receptive | calpain 8 |
| LEFTY1 | 1 | 225886282 | 225911382 | receptive | left-right determination factor 1 |
| CYS1 | 2 | 10056780 | 10080411 | receptive | cystin 1 |
| CAPN13 | 2 | 30722771 | 30807446 | receptive | calpain 13 |
| LINC01320 | 2 | 34677579 | 34738231 | receptive | long intergenic non-protein coding RNA 1320 |
| GNLY | 2 | 85694291 | 85698854 | receptive | granulysin |
| MALL | 2 | 110083870 | 110116566 | receptive | mal, T-cell differentiation protein-like |
| PAX8 | 2 | 113215997 | 113278950 | receptive | paired box 8 |
| INHBB | 2 | 120346143 | 120351808 | receptive | inhibin, beta B |
| DPP4 | 2 | 161992241 | 162074542 | receptive | dipeptidyl-peptidase 4 |
| NABP1 | 2 | 191678136 | 191696659 | receptive | nucleic acid binding protein 1 |
| AOX1 | 2 | 200585868 | 200677064 | receptive | aldehyde oxidase 1 |
| TUBA4A | 2 | 219249711 | 219278170 | receptive | tubulin, alpha 4a |

| | | | | | |
|---|---|---|---|---|---|
| DNER | 2 | 229357629 | 229714558 | receptive | delta/notch-like EGF repeat containing |
| B3GNT7 | 2 | 231395543 | 231401164 | receptive | UDP-GlcNAc:betaGal beta-1,3-N- acetylglucosaminyltransferase 7 |
| THRB | 3 | 24117160 | 24495282 | receptive | thyroid hormone receptor, beta |
| SLC6A20 | 3 | 45755450 | 45796535 | receptive | solute carrier family 6 (proline IMINO transporter), member 20 |
| LINC01212 | 3 | 69999577 | 70015318 | receptive | long intergenic non-protein coding RNA 1212 |
| EIF4E3 | 3 | 71675416 | 71754773 | receptive | eukaryotic translation initiation factor 4E family member 3 |
| NFKBIZ | 3 | 101827991 | 101861022 | receptive | nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, zeta |
| ZPLD1 | 3 | 102099244 | 102479841 | receptive | zona pellucida-like domain containing 1 |
| GRAMD1C | 3 | 113838418 | 113947174 | receptive | GRAM domain containing 1C |
| CP | 3 | 149162417 | 149222055 | receptive | ceruloplasmin (ferroxidase) |
| RARRES1 | 3 | 158696892 | 158732696 | receptive | retinoic acid receptor responder (tazarotene induced) 1 |
| THPO | 3 | 184371935 | 184378144 | receptive | thrombopoietin |
| ST6GAL1 | 3 | 186930485 | 187078553 | receptive | ST6 beta-galactosamide alpha-2,6- sialyltranferase 1 |
| BCL6 | 3 | 187721377 | 187745727 | receptive | B-cell CLL/lymphoma 6 |
| CLDN1 | 3 | 190305701 | 190322475 | receptive | claudin 1 |
| PPARGC1A | 4 | 23792021 | 23890077 | receptive | peroxisome proliferator-activated receptor gamma, coactivator 1 alpha |
| FGA | 4 | 154583126 | 154590766 | receptive | fibrinogen alpha chain |
| FGG | 4 | 154604134 | 154612967 | receptive | fibrinogen gamma chain |
| GUCY1B3 | 4 | 155758992 | 155807591 | receptive | guanylate cyclase 1, soluble, beta 3 |
| PTGER4 | 5 | 40679498 | 40693735 | receptive | prostaglandin E receptor 4 (subtype EP4) |
| GZMA | 5 | 55102648 | 55110252 | receptive | granzyme A (granzyme 1, cytotoxic T- lymphocyte-associated serine esterase 3) |
| IL6ST | 5 | 55935095 | 55994993 | receptive | interleukin 6 signal transducer |
| ANKRD55 | 5 | 56099678 | 56233359 | receptive | ankyrin repeat domain 55 |
| ELOVL7 | 5 | 60751791 | 60844389 | receptive | ELOVL fatty acid elongase 7 |

| | | | | | |
|---|---|---|---|---|---|
| SLCO4C1 | 5 | 102233986 | 102296549 | receptive | solute carrier organic anion transporter family, member 4C1 |
| CXCL14 | 5 | 135570679 | 135579279 | receptive | chemokine (C-X-C motif) ligand 14 |
| SPINK1 | 5 | 147824568 | 147831786 | receptive | serine peptidase inhibitor, Kazal type 1 |
| GPX3 | 5 | 151020438 | 151028993 | receptive | glutathione peroxidase 3 (plasma) |
| HAVCR1 | 5 | 157029413 | 157059119 | receptive | hepatitis A virus cellular receptor 1 |
| GFOD1 | 6 | 13357830 | 13487662 | receptive | glucose-fructose oxidoreductase domain containing 1 |
| PRL | 6 | 22287244 | 22297501 | receptive | prolactin |
| RNF39 | 6 | 30070266 | 30075887 | receptive | ring finger protein 39 |
| VWA7 | 6 | 31765590 | 31777294 | receptive | von Willebrand factor A domain containing 7 |
| SLC44A4 | 6 | 31863192 | 31879046 | receptive | solute carrier family 44, member 4 |
| UNC5CL | 6 | 41026911 | 41039217 | receptive | unc-5 homolog C (C. elegans)-like |
| GPR110 | 6 | 46997703 | 47042363 | receptive | G protein-coupled receptor 110 |
| VNN1 | 6 | 132681590 | 132714049 | receptive | vanin 1 |
| ALDH8A1 | 6 | 134917390 | 134950122 | receptive | aldehyde dehydrogenase 8 family, member A1 |
| MAP3K5 | 6 | 136557047 | 136792518 | receptive | mitogen-activated protein kinase kinase kinase 5 |
| CITED2 | 6 | 139371807 | 139374620 | receptive | Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2 |
| PPP1R14C | 6 | 150143076 | 150250357 | receptive | protein phosphatase 1, regulatory (inhibitor) subunit 14C |
| SOD2 | 6 | 159669057 | 159762529 | receptive | superoxide dismutase 2, mitochondrial |
| THBS2 | 6 | 169215780 | 169254044 | receptive | thrombospondin 2 |
| SCIN | 7 | 12570686 | 12653603 | receptive | scinderin |
| SNX10 | 7 | 26291895 | 26374329 | receptive | sorting nexin 10 |
| ADCY1 | 7 | 45574140 | 45723116 | receptive | adenylate cyclase 1 (brain) |
| ABCA13 | 7 | 48171460 | 48647496 | receptive | ATP-binding cassette, sub-family A (ABC1), member 13 |
| CLDN4 | 7 | 73827744 | 73832693 | receptive | claudin 4 |
| FGL2 | 7 | 77193371 | 77199826 | receptive | fibrinogen-like 2 |

| | | | | | |
|---|---|---|---|---|---|
| CYP3A5 | 7 | 99648194 | 99679998 | receptive | cytochrome P450, family 3, subfamily A, polypeptide 5 |
| MET | 7 | 116672390 | 116798386 | receptive | MET proto-oncogene, receptor tyrosine kinase |
| CPA4 | 7 | 130293134 | 130324180 | receptive | carboxypeptidase A4 |
| TMEM140 | 7 | 135148072 | 135166215 | receptive | transmembrane protein 140 |
| NOS3 | 7 | 150990995 | 151014588 | receptive | nitric oxide synthase 3 (endothelial cell) |
| MTMR7 | 8 | 17298030 | 17413327 | receptive | myotubularin related protein 7 |
| SLC7A2 | 8 | 17497088 | 17570573 | receptive | solute carrier family 7 (cationic amino acid transporter, y+ system), member 2 |
| STC1 | 8 | 23841915 | 23854807 | receptive | stanniocalcin 1 |
| IDO1 | 8 | 39902275 | 39928431 | receptive | indoleamine 2,3-dioxygenase 1 |
| CEBPD | 8 | 47736909 | 47739086 | receptive | CCAAT/enhancer binding protein (C/EBP), delta |
| RGS20 | 8 | 53851808 | 53959303 | receptive | regulator of G-protein signaling 20 |
| FAM84B | 8 | 126552442 | 126558393 | receptive | family with sequence similarity 84, member B |
| DOCK8 | 9 | 214865 | 465259 | receptive | dedicator of cytokinesis 8 |
| SLC1A1 | 9 | 4490444 | 4587469 | receptive | solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1 |
| LURAP1L | 9 | 12775021 | 12822131 | receptive | leucine rich adaptor protein 1-like |
| AQP3 | 9 | 33441154 | 33447611 | receptive | aquaporin 3 (Gill blood group) |
| TMEM252 | 9 | 68536580 | 68540867 | receptive | transmembrane protein 252 |
| GAS1 | 9 | 86944363 | 86947189 | receptive | growth arrest-specific 1 |
| DAPK1 | 9 | 87497228 | 87708633 | receptive | death-associated protein kinase 1 |
| PAEP | 9 | 135561758 | 135566955 | receptive | progestagen-associated endometrial protein |
| LINC01502 | 9 | 135574935 | 135587112 | receptive | long intergenic non-protein coding RNA 1502 |
| KLF6 | 10 | 3775996 | 3785281 | receptive | Kruppel-like factor 6 |
| PHYHIPL | 10 | 59176590 | 59247774 | receptive | phytanoyl-CoA 2-hydroxylase interacting protein-like |
| ARID5B | 10 | 61901300 | 62096944 | receptive | AT rich interactive domain 5B (MRF1-like) |

| | | | | | |
|---|---|---|---|---|---|
| TSPAN15 | 10 | 69451473 | 69507669 | receptive | tetraspanin 15 |
| RBP4 | 10 | 93591687 | 93601744 | receptive | retinol binding protein 4, plasma |
| COL17A1 | 10 | 104031286 | 104086002 | receptive | collagen, type XVII, alpha 1 |
| SORCS1 | 10 | 106573663 | 107164534 | receptive | sortilin-related VPS10 domain containing receptor 1 |
| ADRA2A | 10 | 111077163 | 111080907 | receptive | adrenoceptor alpha 2A |
| HABP2 | 10 | 113550837 | 113589602 | receptive | hyaluronan binding protein 2 |
| SLC18A2 | 10 | 117241093 | 117279430 | receptive | solute carrier family 18 (vesicular monoamine transporter), member 2 |
| ACADSB | 10 | 123008979 | 123058311 | receptive | acyl-CoA dehydrogenase, short/branched chain |
| SERPING1 | 11 | 57597387 | 57614853 | receptive | serpin peptidase inhibitor, clade G (C1 inhibitor), member 1 |
| RARRES3 | 11 | 63536816 | 63546462 | receptive | retinoic acid receptor responder (tazarotene induced) 3 |
| PLA2G16 | 11 | 63573195 | 63614469 | receptive | phospholipase A2, group XVI |
| CTSW | 11 | 65879809 | 65883741 | receptive | cathepsin W |
| NNMT | 11 | 114257831 | 114313285 | receptive | nicotinamide N-methyltransferase |
| TAGLN | 11 | 117199321 | 117204782 | receptive | transgelin |
| SLC6A12 | 12 | 190077 | 214120 | receptive | solute carrier family 6 (neurotransmitter transporter), member 12 |
| B4GALNT3 | 12 | 460364 | 563509 | receptive | beta-1,4-N-acetyl-galactosaminyl transferase 3 |
| CLEC4E | 12 | 8533305 | 8540963 | receptive | C-type lectin domain family 4, member E |
| MFAP5 | 12 | 8637346 | 8662884 | receptive | microfibrillar associated protein 5 |
| RIMKLB | 12 | 8697177 | 8783085 | receptive | ribosomal modification protein rimK-like family member B |
| ERP27 | 12 | 14914035 | 14939082 | receptive | endoplasmic reticulum protein 27 |
| KRT7 | 12 | 52233114 | 52248921 | receptive | keratin 7 |
| AVIL | 12 | 57797882 | 57818704 | receptive | advillin |
| PTPRR | 12 | 70638073 | 70920843 | receptive | protein tyrosine phosphatase, receptor type, R |
| TSPAN8 | 12 | 71125085 | 71441898 | receptive | tetraspanin 8 |
| PPFIA2 | 12 | 81257975 | 81759553 | receptive | protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 2 |

| | | | | | |
|---|---|---|---|---|---|
| HAL | 12 | 95972662 | 95996365 | receptive | histidine ammonia-lyase |
| HNF1A-AS1 | 12 | 120941728 | 120980965 | receptive | HNF1A antisense RNA 1 |
| SLC15A4 | 12 | 128793191 | 128823983 | receptive | solute carrier family 15 (oligopeptide transporter), member 4 |
| GLT1D1 | 12 | 128853494 | 128984968 | receptive | glycosyltransferase 1 domain containing 1 |
| PCDH17 | 13 | 57631812 | 57729311 | receptive | protocadherin 17 |
| EDNRB | 13 | 77895481 | 77919768 | receptive | endothelin receptor type B |
| SLC15A1 | 13 | 98683801 | 98752654 | receptive | solute carrier family 15 (oligopeptide transporter), member 1 |
| RNASE4 | 14 | 20684100 | 20701215 | receptive | ribonuclease, RNase A family, 4 |
| ANG | 14 | 20684177 | 20694186 | receptive | angiogenin, ribonuclease, RNase A family, 5 |
| PTGER2 | 14 | 52314305 | 52328606 | receptive | prostaglandin E receptor 2 (subtype EP2), 53kDa |
| KCNH5 | 14 | 62699454 | 63101866 | receptive | potassium voltage-gated channel, subfamily H (eag-related), member 5 |
| SYNE2 | 14 | 63852983 | 64226433 | receptive | spectrin repeat containing, nuclear envelope 2 |
| PAPLN | 14 | 73237497 | 73274640 | receptive | papilin, proteoglycan-like sulfated glycoprotein |
| TMEM63C | 14 | 77116568 | 77259495 | receptive | transmembrane protein 63C |
| CATSPERB | 14 | 91580774 | 91780707 | receptive | catsper channel auxiliary subunit beta |
| TNFAIP2 | 14 | 103123461 | 103137439 | receptive | tumor necrosis factor, alpha-induced protein 2 |
| BCL2L10 | 15 | 52109263 | 52112775 | receptive | BCL2-like 10 (apoptosis facilitator) |
| LIPC | 15 | 58410569 | 58569843 | receptive | lipase, hepatic |
| C2CD4B | 15 | 62163535 | 62165283 | receptive | C2 calcium-dependent domain containing 4B |
| ST20 | 15 | 79898840 | 79923702 | receptive | suppressor of tumorigenicity 20 |
| BCL2A1 | 15 | 79960889 | 79971446 | receptive | BCL2-related protein A1 |
| FAM169B | 15 | 98437162 | 98514382 | receptive | family with sequence similarity 169, member B |
| TMC5 | 16 | 19410539 | 19499113 | receptive | transmembrane channel-like 5 |
| IL4R | 16 | 27313668 | 27364778 | receptive | interleukin 4 receptor |
| GSG1L | 16 | 27787535 | 28063509 | receptive | GSG1-like |
| NUPR1 | 16 | 28532708 | 28539174 | receptive | nuclear protein, transcriptional regulator, 1 |

| | | | | | |
|---|---|---|---|---|---|
| IRX3 | 16 | 54283304 | 54286763 | receptive | iroquois homeobox 3 |
| IRX5 | 16 | 54930862 | 54934485 | receptive | iroquois homeobox 5 |
| MT2A | 16 | 56608199 | 56609497 | receptive | metallothionein 2A |
| MT1E | 16 | 56625475 | 56627112 | receptive | metallothionein 1E |
| CDH3 | 16 | 68644836 | 68722616 | receptive | cadherin 3, type 1, P-cadherin (placental) |
| HAS3 | 16 | 69105564 | 69118719 | receptive | hyaluronan synthase 3 |
| CDYL2 | 16 | 80597906 | 80804329 | receptive | chromodomain protein, Y-like 2 |
| ALOX12 | 17 | 6996065 | 7010736 | receptive | arachidonate 12-lipoxygenase |
| ACAP1 | 17 | 7336529 | 7351478 | receptive | ArfGAP with coiled-coil, ankyrin repeat and PH domains 1 |
| CD68 | 17 | 7579467 | 7582113 | receptive | CD68 molecule |
| DDX52 | 17 | 37609778 | 37643464 | receptive | DEAD (Asp-Glu-Ala-Asp) box polypeptide 52 |
| HNF1B | 17 | 37686432 | 37745247 | receptive | HNF1 homeobox B |
| GRB7 | 17 | 39737927 | 39747291 | receptive | growth factor receptor-bound protein 7 |
| GAST | 17 | 41712326 | 41715969 | receptive | gastrin |
| HAP1 | 17 | 41717742 | 41734644 | receptive | huntingtin-associated protein 1 |
| MPP3 | 17 | 43800799 | 43833170 | receptive | membrane protein, palmitoylated 3 (MAGUK p55 subfamily member 3) |
| ITGB3 | 17 | 47253846 | 47344292 | receptive | integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61) |
| B4GALNT2 | 17 | 49132460 | 49169989 | receptive | beta-1,4-N-acetyl-galactosaminyl transferase 2 |
| ITGA3 | 17 | 50056110 | 50090481 | receptive | integrin, alpha 3 (antigen CD49C, alpha 3 subunit of VLA-3 receptor) |
| TMEM92 | 17 | 50271406 | 50281485 | receptive | transmembrane protein 92 |
| ABCC3 | 17 | 50634777 | 50692252 | receptive | ATP-binding cassette, sub-family C (CFTR/MRP), member 3 |
| ERN1 | 17 | 64039142 | 64130819 | receptive | endoplasmic reticulum to nucleus signaling 1 |
| ABCA6 | 17 | 69078702 | 69141888 | receptive | ATP-binding cassette, sub-family A (ABC1), member 6 |
| SOCS3 | 17 | 78356778 | 78360077 | receptive | suppressor of cytokine signaling 3 |
| DLGAP1-AS3 | 18 | 3878180 | 3897069 | receptive | DLGAP1 antisense RNA 3 |
| IMPA2 | 18 | 11981037 | 12030883 | receptive | inositol(myo)-1(or 4)-monophosphatase 2 |

| | | | | | |
|---|---|---|---|---|---|
| GATA6 | 18 | 22169443 | 22202528 | receptive | GATA binding protein 6 |
| ALPK2 | 18 | 58481247 | 58628957 | receptive | alpha-kinase 2 |
| C3 | 19 | 6677704 | 6730562 | receptive | complement component 3 |
| COMP | 19 | 18782773 | 18791314 | receptive | cartilage oligomeric matrix protein |
| RHPN2 | 19 | 32978593 | 33064888 | receptive | rhophilin, Rho GTPase binding protein 2 |
| FXYD3 | 19 | 35115879 | 35124324 | receptive | FXYD domain containing ion transport regulator 3 |
| PSG9 | 19 | 43211791 | 43269530 | receptive | pregnancy specific beta-1-glycoprotein 9 |
| LYPD3 | 19 | 43460787 | 43465660 | receptive | LY6/PLAUR domain containing 3 |
| KIR2DL4 | 19 | 54803535 | 54814517 | receptive | killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 4 |
| THBD | 20 | 23045633 | 23049741 | receptive | thrombomodulin |
| LBP | 20 | 38346356 | 38377023 | receptive | lipopolysaccharide binding protein |
| SLPI | 20 | 45252239 | 45254564 | receptive | secretory leukocyte peptidase inhibitor |
| KCNB1 | 20 | 49363877 | 49482647 | receptive | potassium voltage-gated channel, Shab- related subfamily, member 1 |
| CYP24A1 | 20 | 54153449 | 54173973 | receptive | cytochrome P450, family 24, subfamily A, polypeptide 1 |
| SLCO4A1 | 20 | 62642445 | 62672295 | receptive | solute carrier organic anion transporter family, member 4A1 |
| TMEM1 | 21 | 5155499 | 5163695 | receptive | TMEM1 protein {ECO:0000313\|EMBL:BAA21136.1} |
| SGSM1 | 22 | 24806169 | 24927578 | receptive | small G protein signaling modulator 1 |
| LIF | 22 | 30240447 | 30246851 | receptive | leukemia inhibitory factor |
| WNT7B | 22 | 45920362 | 45977129 | receptive | wingless-type MMTV integration site family, member 7B |
| TYMP | 22 | 50525752 | 50530085 | receptive | thymidine phosphorylase |
| CSF2RA | X | 1268800 | 1310381 | receptive | colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage) |
| MAOA | X | 43654907 | 43746824 | receptive | monoamine oxidase A |
| LINC00890 | X | 111511662 | 111522399 | receptive | long intergenic non-protein coding RNA 890 |
| GABRE | X | 151953124 | 151974679 | receptive | gamma-aminobutyric acid (GABA) A receptor, epsilon |

**Supplementary Table 2 Stromal cell signature genes.** List of genes in the stromal cell signature that was developed from cultured stromal cells.

| Gene | Chr | Start | End | Description |
|---|---|---|---|---|
| COL24A1 | 1 | 85729233 | 86156943 | collagen, type XXIV, alpha 1 |
| DPT | 1 | 168695459 | 168729264 | dermatopontin |
| AFF3 | 2 | 99545419 | 100142739 | AF4/FMR2 family, member 3 |
| HOXD10 | 2 | 176108790 | 176119942 | homeobox D10 |
| HOXD9 | 2 | 176122720 | 176124937 | homeobox D9 |
| ZNF804A | 2 | 184598366 | 184939492 | zinc finger protein 804A |
| LINC01391 | 3 | 138935189 | 138944020 | long intergenic non-protein coding RNA 1391 |
| FOXL2 | 3 | 138944224 | 138947140 | forkhead box L2 |
| SLC9A9 | 3 | 143265222 | 143848531 | solute carrier family 9, subfamily A (NHE9, cation proton antiporter 9), member 9 |
| HTRA3 | 4 | 8269765 | 8307111 | HtrA serine peptidase 3 |
| GPR78 | 4 | 8558725 | 8619761 | G protein-coupled receptor 78 |
| CPZ | 4 | 8592660 | 8619759 | carboxypeptidase Z |
| C1QTNF7 | 4 | 15339818 | 15446166 | C1q and tumor necrosis factor related protein 7 |
| ADAMTS12 | 5 | 33523535 | 33892192 | ADAM metallopeptidase with thrombospondin type 1 motif, 12 |
| CASC15 | 6 | 21664772 | 22214505 | cancer susceptibility candidate 15 (non- protein coding) |
| TREM1 | 6 | 41267926 | 41286719 | triggering receptor expressed on myeloid cells 1 |
| HOXA5 | 7 | 27141052 | 27143668 | homeobox A5 |
| HOXA9 | 7 | 27162435 | 27175180 | homeobox A9 |
| HOXA10-HOXA9 | 7 | 27163535 | 27180013 | Uncharacterized protein {ECO:0000313\|Ensembl:ENSP00000421799} |
| TRBC2 | 7 | 142670777 | 142802748 | T cell receptor beta constant 2 |
| RUNX1T1 | 8 | 91954967 | 92103286 | runt-related transcription factor 1; translocated to, 1 (cyclin D-related) |
| HAS2 | 8 | 121612116 | 121641390 | hyaluronan synthase 2 |
| TEK | 9 | 27109141 | 27230175 | TEK tyrosine kinase, endothelial |
| TMEM215 | 9 | 32783499 | 32787399 | transmembrane protein 215 |

| | | | | |
|---|---|---|---|---|
| VSTM4 | 10 | 49014245 | 49115509 | V-set and transmembrane domain containing 4 |
| HPSE2 | 10 | 98457077 | 99235862 | heparanase 2 (inactive) |
| DPYSL4 | 10 | 132186900 | 132205776 | dihydropyrimidinase-like 4 |
| SLC1A2 | 11 | 35251206 | 35420063 | solute carrier family 1 (glial high affinity glutamate transporter), member 2 |
| NOX4 | 11 | 89324356 | 89498187 | NADPH oxidase 4 |
| CASP1 | 11 | 105025443 | 105035250 | caspase 1, apoptosis-related cysteine peptidase |
| KIF5A | 12 | 57549998 | 57586632 | kinesin family member 5A |
| ATP8A2 | 13 | 25372071 | 26025851 | ATPase, aminophospholipid transporter, class I, type 8A, member 2 |
| RP11-404P21.8 | 14 | 96204844 | 96263929 | Uncharacterized protein {ECO:0000313\|Ensembl:ENSP00000450984} |
| CACNA1H | 16 | 1153241 | 1221771 | calcium channel, voltage-dependent, T type, alpha 1H subunit |
| FOXF1 | 16 | 86510527 | 86515418 | forkhead box F1 |
| TBX2 | 17 | 61399896 | 61409466 | T-box 2 |
| PIEZO2 | 18 | 10666483 | 11148762 | piezo-type mechanosensitive ion channel component 2 |
| FBXO17 | 19 | 38941401 | 38975910 | F-box protein 17 |
| ADAM33 | 20 | 3667965 | 3682246 | ADAM metallopeptidase domain 33 |
| KIAA1755 | 20 | 38210488 | 38260772 | KIAA1755 |
| RP11-524D16__A.3 | X | 100673330 | 100673981 | No description |
| MUM1L1 | X | 106168305 | 106208956 | melanoma associated antigen (mutated) 1-like 1 |
| LINC00890 | X | 111511662 | 111522399 | long intergenic non-protein coding RNA 890 |
| FHL1 | X | 136147400 | 136211359 | four and a half LIM domains 1 |

**Supplementary Table 3 Epithelial cell signature genes.** List of genes in the epithelial cell signature that was developed from epithelial cell organoids. Genes marked as non-receptive are upregulated in the non-receptive state and genes marked receptive are upregulated in the receptive state.

| Gene | Chr | Start | End | Upregulated in: | Description |
|---|---|---|---|---|---|
| SYTL1 | 1 | 27342020 | 27353937 | non-receptive | synaptotagmin-like 1 |
| GJB3 | 1 | 34781189 | 34786369 | non-receptive | gap junction protein, beta 3, 31kDa |
| PCSK9 | 1 | 55039548 | 55064852 | non-receptive | proprotein convertase subtilisin/kexin type 9 |
| DENND2D | 1 | 111187174 | 111204535 | non-receptive | DENN/MADD domain containing 2D |
| RAB25 | 1 | 156061160 | 156070514 | non-receptive | RAB25, member RAS oncogene family |
| TSTD1 | 1 | 161037631 | 161038990 | non-receptive | thiosulfate sulfurtransferase (rhodanese)- like domain containing 1 |
| SLC26A9 | 1 | 205913048 | 205943460 | non-receptive | solute carrier family 26 (anion exchanger), member 9 |
| GRHL1 | 2 | 9951693 | 10002277 | non-receptive | grainyhead-like 1 (Drosophila) |
| LYPD6B | 2 | 149038107 | 149215262 | non-receptive | LY6/PLAUR domain containing 6B |
| LY75 | 2 | 159803355 | 159904749 | non-receptive | lymphocyte antigen 75 |
| NUP210 | 3 | 13316235 | 13420309 | non-receptive | nucleoporin 210kDa |
| WNT7A | 3 | 13816258 | 13880121 | non-receptive | wingless-type MMTV integration site family, member 7A |
| FAM107A | 3 | 58564117 | 58627610 | non-receptive | family with sequence similarity 107, member A |
| MECOM | 3 | 169083499 | 169663618 | non-receptive | MDS1 and EVI1 complex locus |
| FGFBP1 | 4 | 15935569 | 15938740 | non-receptive | fibroblast growth factor binding protein 1 |
| C4orf19 | 4 | 37453941 | 37623495 | non-receptive | chromosome 4 open reading frame 19 |
| CXCL5 | 4 | 73995642 | 73998779 | non-receptive | chemokine (C-X-C motif) ligand 5 |
| ANXA3 | 4 | 78551519 | 78610451 | non-receptive | annexin A3 |
| SH3RF2 | 5 | 145936579 | 146081791 | non-receptive | SH3 domain containing ring finger 2 |
| FAT2 | 5 | 151504093 | 151568944 | non-receptive | FAT atypical cadherin 2 |
| GCNT2 | 6 | 10492223 | 10629368 | non-receptive | glucosaminyl (N-acetyl) transferase 2, I- branching enzyme |
| CLIC5 | 6 | 45898451 | 46080395 | non-receptive | chloride intracellular channel 5 |
| ENPP4 | 6 | 46129993 | 46146699 | non-receptive | ectonucleotide pyrophosphatase/phosphodiesterase 4 (putative) |

| | | | | | |
|---|---|---|---|---|---|
| DLX6-AS1 | 7 | 96955141 | 97014065 | non-receptive | DLX6 antisense RNA 1 |
| CFTR | 7 | 117465784 | 117715971 | non-receptive | cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) |
| SOX17 | 8 | 54457935 | 54460888 | non-receptive | SRY (sex determining region Y)-box 17 |
| CALB1 | 8 | 90058608 | 90095475 | non-receptive | calbindin 1, 28kDa |
| GRHL2 | 8 | 101492432 | 101669726 | non-receptive | grainyhead-like 2 (Drosophila) |
| SYK | 9 | 90801787 | 90898549 | non-receptive | spleen tyrosine kinase |
| PTPN3 | 9 | 109375466 | 109498313 | non-receptive | protein tyrosine phosphatase, non-receptor type 3 |
| MUC6 | 11 | 1012821 | 1036706 | non-receptive | mucin 6, oligomeric mucus/gel-forming |
| MUC5AC | 11 | 1157953 | 1201138 | non-receptive | mucin 5AC, oligomeric mucus/gel-forming |
| CDC42BPG | 11 | 64823387 | 64844569 | non-receptive | CDC42 binding protein kinase gamma (DMPK- like) |
| AQP5 | 12 | 49961870 | 49965681 | non-receptive | aquaporin 5 |
| NOS1 | 12 | 117208142 | 117452170 | non-receptive | nitric oxide synthase 1 (neuronal) |
| MCF2L | 13 | 112894378 | 113099739 | non-receptive | MCF.2 cell line derived transforming sequence-like |
| ERN2 | 16 | 23690326 | 23713500 | non-receptive | endoplasmic reticulum to nucleus signaling 2 |
| MARVELD3 | 16 | 71626161 | 71642114 | non-receptive | MARVEL domain containing 3 |
| PPP1R1B | 17 | 39626740 | 39636626 | non-receptive | protein phosphatase 1, regulatory (inhibitor) subunit 1B |
| HOXB8 | 17 | 48611377 | 48614939 | non-receptive | homeobox B8 |
| HOXB9 | 17 | 48621159 | 48626356 | non-receptive | homeobox B9 |
| RNF43 | 17 | 58352500 | 58417595 | non-receptive | ring finger protein 43 |
| EVPL | 17 | 76004502 | 76027452 | non-receptive | envoplakin |
| LIPG | 18 | 49560699 | 49599182 | non-receptive | lipase, endothelial |
| ONECUT3 | 19 | 1752373 | 1780988 | non-receptive | one cut homeobox 3 |
| TJP3 | 19 | 3708109 | 3750813 | non-receptive | tight junction protein 3 |
| CEACAM5 | 19 | 41576273 | 41729798 | non-receptive | carcinoembryonic antigen-related cell adhesion molecule 5 |
| CEACAM6 | 19 | 41750977 | 41772208 | non-receptive | carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen) |

| | | | | | |
|---|---|---|---|---|---|
| FAM83F | 22 | 39994949 | 40043529 | non-receptive | family with sequence similarity 83, member F |
| SLC6A14 | X | 116436622 | 116461458 | non-receptive | solute carrier family 6 (amino acid transporter), member 14 |
| SMPDL3B | 1 | 27934993 | 27959157 | receptive | sphingomyelin phosphodiesterase, acid-like 3B |
| TACSTD2 | 1 | 58575423 | 58577773 | receptive | tumor-associated calcium signal transducer 2 |
| ERICH3 | 1 | 74568111 | 74673738 | receptive | glutamate-rich 3 |
| S100A14 | 1 | 153614255 | 153616986 | receptive | S100 calcium binding protein A14 |
| C1orf116 | 1 | 207018521 | 207032756 | receptive | chromosome 1 open reading frame 116 |
| CAPN13 | 2 | 30722771 | 30820542 | receptive | calpain 13 |
| TGFA | 2 | 70447280 | 70554193 | receptive | transforming growth factor, alpha |
| RAB17 | 2 | 237574322 | 237601614 | receptive | RAB17, member RAS oncogene family |
| SLC6A20 | 3 | 45755450 | 45796535 | receptive | solute carrier family 6 (proline IMINO transporter), member 20 |
| RASSF6 | 4 | 73571550 | 73620631 | receptive | Ras association (RalGDS/AF-6) domain family member 6 |
| FGF18 | 5 | 171419656 | 171457623 | receptive | fibroblast growth factor 18 |
| RNF144B | 6 | 18387350 | 18468874 | receptive | ring finger protein 144B |
| SLC44A4 | 6 | 31863192 | 31879046 | receptive | solute carrier family 44, member 4 |
| KCNK5 | 6 | 39188973 | 39229450 | receptive | potassium channel, subfamily K, member 5 |
| MACC1 | 7 | 20134655 | 20217404 | receptive | metastasis associated in colon cancer 1 |
| RAPGEF5 | 7 | 22118238 | 22357144 | receptive | Rap guanine nucleotide exchange factor (GEF) 5 |
| CPVL | 7 | 28995231 | 29195451 | receptive | carboxypeptidase, vitellogenic-like |
| SBSPON | 8 | 73064540 | 73124088 | receptive | somatomedin B and thrombospondin, type 1 domain containing |
| AQP3 | 9 | 33441154 | 33447611 | receptive | aquaporin 3 (Gill blood group) |
| LCN2 | 9 | 128149071 | 128153455 | receptive | lipocalin 2 |
| HKDC1 | 10 | 69220303 | 69267559 | receptive | hexokinase domain containing 1 |
| COL17A1 | 10 | 104031286 | 104086002 | receptive | collagen, type XVII, alpha 1 |
| PKP3 | 11 | 392614 | 404908 | receptive | plakophilin 3 |
| IL18 | 11 | 112143251 | 112164117 | receptive | interleukin 18 |

| | | | | | |
|---|---|---|---|---|---|
| MPZL3 | 11 | 118226690 | 118252350 | receptive | myelin protein zero-like 3 |
| TMEM45B | 11 | 129815819 | 129860003 | receptive | transmembrane protein 45B |
| GLB1L2 | 11 | 134331874 | 134378341 | receptive | galactosidase, beta 1-like 2 |
| RASAL1 | 12 | 113098819 | 113136239 | receptive | RAS protein activator like 1 (GAP1 like) |
| TESC | 12 | 117038923 | 117099479 | receptive | tescalcin |
| TMEM233 | 12 | 119593459 | 119643066 | receptive | transmembrane protein 233 |
| SCEL | 13 | 77535674 | 77645263 | receptive | sciellin |
| GPX2 | 14 | 64939152 | 64942905 | receptive | glutathione peroxidase 2 (gastrointestinal) |
| SERPINA1 | 14 | 94376747 | 94390693 | receptive | serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1 |
| SLC27A2 | 15 | 50182196 | 50236395 | receptive | solute carrier family 27 (fatty acid transporter), member 2 |
| PRSS33 | 16 | 2783953 | 2787948 | receptive | protease, serine, 33 |
| TOX3 | 16 | 52438005 | 52547802 | receptive | TOX high mobility group box family member 3 |
| CDH3 | 16 | 68636189 | 68722616 | receptive | cadherin 3, type 1, P-cadherin (placental) |
| CHST6 | 16 | 75472052 | 75495384 | receptive | carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 6 |
| ATP2C2 | 16 | 84368527 | 84464187 | receptive | ATPase, Ca++ transporting, type 2C, member 2 |
| TNS4 | 17 | 40475828 | 40501597 | receptive | tensin 4 |
| SIGLEC15 | 18 | 45825512 | 45844080 | receptive | sialic acid binding Ig-like lectin 15 |
| MISP | 19 | 751126 | 764319 | receptive | mitotic spindle positioning |
| LRG1 | 19 | 4536409 | 4540474 | receptive | leucine-rich alpha-2-glycoprotein 1 |
| CTD-2396E7.11 | 19 | 6469465 | 6470152 | receptive | No description |
| SH2D3A | 19 | 6752160 | 6767588 | receptive | SH2 domain containing 3A |
| MUC16 | 19 | 8848844 | 8981342 | receptive | mucin 16, cell surface associated |
| B3GNT3 | 19 | 17794828 | 17813082 | receptive | UDP-GlcNAc:betaGal beta-1,3-N- acetylglucosaminyltransferase 3 |
| FOXA2 | 20 | 22581005 | 22585455 | receptive | forkhead box A2 |
| PRODH | 22 | 18912777 | 18936553 | receptive | proline dehydrogenase (oxidase) 1 |

**Supplementary Table 4 Enriched Motif IDs for bivalent regions that resolved to H3K4me3-only in the receptive state.** Analysis was carried out using JASPAR2018 and HOMCOMOCOv11 motif databases. Analysis was further refined by factors having to be present in both datasets. Adjusted p-value for each motif and database displayed.

| Motif ID | Adjusted p-value | | Motif ID | Adjusted p-value | |
|---|---|---|---|---|---|
| | JASPAR2018 | HOCOMOCOv11 | | JASPAR2018 | HOCOMOCOv11 |
| E2F6 | 1.20E-02 | 4.09E-10 | KLF4 | 3.94E-03 | 3.78E-07 |
| EHF | 2.58E-04 | 5.68E-14 | KLF5 | 7.22E-09 | 9.47E-17 |
| ELF1 | 1.04E-02 | 2.71E-11 | KLF9 | 3.83E-11 | 3.17E-03 |
| ELF5 | 6.02E-10 | 2.64E-15 | MEF2A | 3.98E-05 | 2.56E-03 |
| ERG | 1.34E-03 | 5.50E-21 | MEF2B | 1.32E-02 | 8.42E-03 |
| ETS1 | 3.80E-04 | 4.14E-16 | MEF2C | 1.09E-03 | 4.47E-03 |
| FOXA1 | 3.00E-04 | 6.02E-05 | MEF2D | 8.27E-03 | 2.83E-04 |
| FOXA2 | 2.64E-05 | 3.83E-04 | MYOD1 | 7.97E-03 | 5.94E-04 |
| FOXJ3 | 2.90E-03 | 5.29E-06 | MYOG | 9.50E-04 | 4.35E-03 |
| FOXK1 | 3.96E-08 | 1.35E-08 | SOX3 | 3.64E-07 | 7.57E-06 |
| FOXO1 | 1.55E-04 | 4.09E-07 | SP1 | 1.45E-08 | 1.47E-09 |
| FOXP1 | 1.68E-04 | 4.82E-09 | SP2 | 1.98E-08 | 1.09E-13 |
| FOXP2 | 3.92E-07 | 3.08E-05 | SP3 | 2.47E-04 | 1.30E-13 |
| GABPA | 5.08E-05 | 3.81E-05 | SP4 | 1.45E-03 | 4.04E-20 |
| HIF1A | 1.39E-02 | 3.12E-03 | SPIB | 1.56E-11 | 4.06E-04 |
| IRF1 | 4.40E-10 | 9.45E-10 | STAT1 | 9.34E-03 | 2.79E-03 |
| IRF3 | 1.73E-05 | 1.84E-27 | STAT4 | 1.21E-04 | 2.07E-06 |
| IRF7 | 1.39E-02 | 6.05E-07 | TBP | 3.03E-03 | 9.83E-03 |
| KLF1 | 7.93E-04 | 6.44E-12 | ZEB1 | 2.52E-04 | 2.62E-06 |

**Supplementary Table 5 Additional DNA modifiers expression in the endometrium.** DESeq2 and/or EdgeR corrected p values for genes involved in the base excision repair pathway. Expression of genes in LCM coupled with RNA-seq of epithelial cells between LH+5 and LH+8 (Salker et al. 2017). As well as the expression in my whole tissue RNA-sequencing data comparing non-receptive with receptive samples.

| | LCM RNA-seq Epithelial cells LH+5 vs LH+8 | My RNA-seq data from Whole Tissue Samples | |
|---|---|---|---|
| Gene | DESeq2 p value | DESeq2 | EdgeR |
| UHRF1 | 1 | 0.808 | 1 |
| MBD2 | 0.728 | 0.289 | 0.356 |
| POLB | 1 | 0.543 | 0.907 |
| XRCC1 | 1 | 0.417 | 0.611 |
| APEX1 | 1 | 0.149 | 0.509 |
| MPG | 1 | 0.292 | 0.722 |
| OGG1 | 1 | 0.867 | 1 |
| SMUG1 | 1 | 0.877 | 1 |
| TDG | 1 | 0.945 | 1 |
| NEIL1 | 1 | 0.813 | 1 |
| PARP1 | 1 | 0.415 | 0.794 |

**Supplementary Table 6 Extended Hippo pathway genes and associated differential 5mC or 5hmC genes.** A list of KEGG Hippo signalling pathway genes (column one and two, https://www.genome.jp/dbget-bin/www_bget?pathway+hsa04390). The genes associated with differential 5mC between non-receptive control and endometriosis samples (column three) and between receptive control and endometriosis samples (column four). Genes associated with differential 5hmC between non-receptive control and endometriosis samples (column five), and between receptive control and endometriosis samples (column six).

| KEGG Hippo Pathway | | meDIP NRC v | meDIP RC v RE | hmeDIP NRC v | hmeDIP RC v RE |
|---|---|---|---|---|---|
| ACTB | MYC | ACTB | APC | BMP2 | APC |
| ACTG1 | NF2 | BMP8A | BMP5 | BMP5 | BBC3 |
| AFP | NKD1 | BMPR1A | BMPR1A | BMP6 | BMPR1A |
| AJUBA | PARD3 | BMPR1B | BMPR1B | BMPR1B | BMPR1B |
| AMH | PARD6A | CRB1 | BMPR2 | BMPR2 | CDH1 |
| AMOT | PARD6B | CTNNA1 | BTRC | BTRC | CTNNA1 |
| APC | PARD6G | CTNNA2 | CCND2 | CRB1 | CTNNA2 |
| APC2 | PATJ | CTNNA3 | CRB1 | CTNNA1 | CTNNA3 |
| AREG | PPP1CA | DLG1 | CSNK1E | CTNNA2 | DLG2 |
| AXIN1 | PPP1CB | DLG2 | CTGF | CTNNA3 | FBXW11 |
| AXIN2 | PPP1CC | FRMD6 | CTNNA1 | DLG2 | FRMD6 |

| | | | | | |
|---|---|---|---|---|---|
| BBC3 | PPP2CA | GSK3B | CTNNA2 | NKD1 | GSK3B |
| BIRC2 | PPP2CB | LATS1 | CTNNA3 | PPP2R1B | LATS2 |
| BIRC5 | PPP2R1A | LEF1 | DLG2 | PPP2R2A | LEF1 |
| BMP2 | PPP2R1B | PARD3 | FBXW11 | PPP2R2D | PARD3 |
| BMP4 | PPP2R2A | PPP2CA | FGF1 | PRKCI | PARD6G |
| BMP5 | PPP2R2B | PPP2CB | FZD1 | SMAD4 | PPP1CB |
| BMP6 | PPP2R2C | PPP2R2B | GDF5 | STK3 | PPP2CA |
| BMP7 | PPP2R2D | SMAD2 | GSK3B | TCF7L2 | PPP2R2B |
| BMP8A | PRKCI | SMAD4 | LEF1 | TGFBR2 | SMAD1 |
| BMP8B | PRKCZ | STK3 | LIMD1 | TP53BP2 | SMAD4 |
| BMPR1A | RASSF1 | TCF7 | LLGL2 | WNT8B | STK3 |
| BMPR1B | RASSF6 | TCF7L2 | MOB1B | | TGFBR1 |
| BMPR2 | SAV1 | TEAD1 | MPP5 | | TGFBR2 |
| BTRC | SCRIB | TGFBR2 | NKD1 | | WWC1 |
| CCND1 | SERPINE1 | WNT10A | PARD3 | | WWTR1 |
| CCND2 | SMAD1 | WNT4 | PPP1CB | | |
| CCND3 | SMAD2 | WNT5B | PPP2CA | | |
| CDH1 | SMAD3 | WNT7B | PPP2CB | | |
| CRB1 | SMAD4 | WWC1 | PPP2R1A | | |
| CRB2 | SMAD7 | WWTR1 | PPP2R2B | | |
| CSNK1D | SNAI2 | YAP1 | PRKCZ | | |
| CSNK1E | SOX2 | YWHAB | RASSF6 | | |
| CTGF | STK3 | YWHAZ | SAV1 | | |
| CTNNA1 | TCF7 | | SMAD1 | | |
| CTNNA2 | TCF7L1 | | SMAD4 | | |
| CTNNA3 | TCF7L2 | | STK3 | | |
| CTNNB1 | TEAD1 | | TCF7L1 | | |
| DLG1 | TEAD2 | | TCF7L2 | | |
| DLG2 | TEAD3 | | TEAD1 | | |
| DLG3 | TEAD4 | | TGFB2 | | |
| DLG4 | TGFB1 | | TGFBR1 | | |
| DVL1 | TGFB2 | | TGFBR2 | | |
| DVL2 | TGFB3 | | WNT10B | | |
| DVL3 | TGFBR1 | | WNT9B | | |
| FBXW11 | TGFBR2 | | WWC1 | | |
| FGF1 | TP53BP2 | | WWTR1 | | |
| FRMD1 | TP73 | | | | |
| FRMD6 | WNT1 | | | | |
| FZD1 | WNT10A | | | | |
| FZD10 | WNT10B | | | | |
| FZD2 | WNT11 | | | | |
| FZD3 | WNT16 | | | | |

| | | | | | |
|---|---|---|---|---|---|
| FZD4 | WNT2 | | | | |
| FZD5 | WNT2B | | | | |
| FZD6 | WNT3 | | | | |
| FZD7 | WNT3A | | | | |
| FZD8 | WNT4 | | | | |
| FZD9 | WNT5A | | | | |
| GDF5 | WNT5B | | | | |
| GDF6 | WNT6 | | | | |
| GDF7 | WNT7A | | | | |
| GLI2 | WNT7B | | | | |
| GSK3B | WNT8A | | | | |
| ID1 | WNT8B | | | | |
| ID2 | WNT9A | | | | |
| ITGB2 | WNT9B | | | | |
| LATS1 | WTIP | | | | |
| LATS2 | WWC1 | | | | |
| LEF1 | WWTR1 | | | | |
| LIMD1 | YAP1 | | | | |
| LLGL1 | YWHAB | | | | |
| LLGL2 | YWHAE | | | | |
| CSNK1E | YWHAG | | | | |
| MOB1A | YWHAH | | | | |
| MOB1B | YWHAQ | | | | |
| MPP5 | YWHAZ | | | | |