Exploring endometrial cellular heterogeneity and its role in endometriosis using single-cell transcriptomics

Magda Marečková

Kellogg College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2023

Abstract

The human endometrium, the inner mucosal lining of the uterus, undergoes cycles of shedding, regeneration, growth, and differentiation on a monthly basis in response to steroid hormones. In endometriosis, endometrial-like cells grow outside of the uterus, and are associated with debilitating chronic pain and subfertility that can have a substantial negative impact on quality of life. Yet, it is not fully understood how the endometrium achieves its immense regenerative capacity and whether the cellular make-up of the endometrium in health and endometriosis differs. Studies profiling the endometrium at the single-cell level have so far analysed only a limited number of cells and samples, making it difficult to disentangle the inherent heterogeneity of this dynamic tissue. In this thesis I aimed to generate a comprehensive cellular map of the endometrium in a large set of donors with/without endometriosis both during natural menstrual cycles and when taking exogenous hormonal therapy. I hypothesised that by analysing a large set of endometrial biopsies we can uncover novel cell populations and disease-specific transcriptomic signatures.

The single-cell map presented in this thesis consists of ~600,000 whole cells and nuclei from 90 individuals, and was compiled by integrating data newly generated during my studies with two previously published datasets. Focusing on the epithelial and mesenchymal cell lineages, I defined novel cell populations that appear around the time of ovulation and window of implantation. Moreover, I observed cell populations specific to taking various forms of exogenous hormonal therapy. Additionally, I utilised the *in vivo* endometrial cellular map to dissect the cellular heterogeneity of *in vitro* endometrial organoids and organoid-stromal cell co-cultures. Analysing ~100,000 cells, I described a population of cells expressing *MUC5B*, *PAEP,* and *TFF3* as a feature present in organoids from donors with endometriosis. Altogether, the cellular maps of the endometrium *in vivo* and *in vitro* presented in this thesis have the potential to serve as a great resource to further study endometrial function and associated pathologies, as well as guide the development of novel clinical *in vitro* models of the endometrium.

Acknowledgements

For a collaborative and large-scale project as the one described in this thesis, many individuals were involved and provided invaluable support, guidance and hours of their time and work to make it a reality. The below is an expression of my gratitude to you all, including:

Patients/donors
First and foremost, the work presented in this thesis would have not been possible without the many courageous individuals who decided to take part in our clinical studies and donated their time and tissue samples for endometriosis research. My most sincere thank you goes to you all.

Clinical staff and research nurses
Patient recruitment, sample collection and processing are crucial for what we researchers do in the lab. Our tissue access is possible thanks to the hard work of a team of dedicated clinicians and nurses who while treating the patients also collect specimens for research. A huge thank you goes to the clinical team and our wonderful research nurses Carol Hubbard, Kelly Barrett, and Lisa Buck at the John Radcliffe Hospital in Oxford, who collected most of the samples I processed for my studies. A big thank you also goes to Dr Ee Von Woon and Dr Vicky Male at the Chelsea and Westminster Hospital in London and the clinical team at TFP Oxford Fertility who collected further endometrial biopsies during their clinics.

DPhil supervisors
I would like to thank my supervisory team for their support over the years, never saying no to any of my ideas and activities both as part of and outside of the DPhil work. My thank you goes to Professor Christian Becker, Dr Rebecca Dragovic, Dr Karin Hellner, Dr Jen Southcombe, Dr Roser Vento-Tormo and Professor Krina Zondervan. Krina and Christian, thank you for sharing your expert knowledge and wisdom about everything endometriosis-related. Bec and Jen, thank you for always being there, willing to help, listen and encourage me, especially when it was not easy. Bec, I am also grateful for your guidance on best teaching practices and both yours and Jen's help and advice on anything and everything lab-related. Roser, I cannot thank you enough for making me part of your research group, the time I spent working with you all at the Sanger and everything you have involved me in, from grant to manuscript writing to giving talks on your behalf to helping establish collaborations across Europe. Thank you for being a great mentor, valuing my input and all the opportunities for growth and learning!

Members of the EndoCare group
A big thank you goes to all past and present members of the EndoCare group. I would especially like to thank Dr Kurtis Garbutt, who on 100s of occasions dealt with my data, sample and all sorts of requests, Michał Krassowski for sharing genotyping data that was a final product of many weeks of his hard work and always being happy to help, Dr Cecilia Cheuk for being a very friendly and wise senior DPhil student, helping us newbies a lot when we first started, and thank you also to Dr Ana Kisovar for dealing with my samples while I was working at the Sanger, but also for being a good friend. The most special thank you goes to Dr Danille Perro. Danielle, *ma chérie*, getting through *ze*(the) DPhil would have been impossible without you. Thank you for all your support, friendship from day one of the DPhil journey and being a wonderful person to work and run a podcast series with!

Members of the Vento lab and beyond
To the members of the Vento team - thank you for your warm welcome to the group and making me feel like I had always been a lab member. Your kindness, willingness to help, friendship, sharing knowledge and feedback has been inspiring and I am extremely grateful for it. I have learnt a lot from you all.

A huge thank you for everything data analysis related (which was a bit of a mess) and beyond is for Dr Luz García-Alonso. Dr Regina Hoo, without your initial guidance, I would not have grown a single organoid, so thank you very much for that. I'm also grateful for your many wise pieces of advice in other lab and work/life-related issues - they have been invaluable. Dr Carmen Sancho-Serra, honestly Mentxu, what would I have done without all your help regarding everything that ever needed to be done in the lab? The list is so long that I could easily write a whole thesis about it. Thank you SO much! It was such a pleasure to work with you. I truly enjoyed 'nuclei Thursdays' with you and Dr Agnes Oszlanczi and thank you both immensely for processing the many samples. Thank you Dr Elena Prigmore for optimising the nuclei extraction protocol for the described sample processing and Dr Tarryn Porter for lab-work support.

Thank you to Dr Iva Kelava for being a great office and conference buddy on top of sharing your expert organoid, work-life balance and career advice, as well as looking after my 'babies' when I went away. Cecilia Icoresi-Mazzeo, it was always a pleasure to see you around the lab. Thank you so much for happily tackling the mysterious samples and the many questions I had and being a wonderful and soulful person to talk with. My thank you also goes to Dr Kenny Roberts and Dr Fani Memi, not quite officially Vento lab members, but very kind and caring people to work with, who make both the workplace and time 'outside' delightful. Lastly, to my dear fellow PhD students, Valentina Lorenzi and Elías Ruiz-Morales, you two have been such amazing colleagues, but most importantly friends. Thank you very much for making my life in Cambridge so bright and enjoyable. I'm grateful to have met you. Vale, I will happily remember our walks and chats about everything and anything, including new crazy scientific hypotheses and FemTech. Elías, my favourite TC buddy, spending long hours with cells was undoubtedly more fun with you around, as were the bus journeys, formals, and many get-togethers. Thank you both for always being there for me (even when I was a bit of a biohazard full of covid) and being so caring!

Staff at the Wellcome Sanger Institute and University of Oxford
My sincere thank you goes to all the members of staff at both institutes that ensure the smooth running of everything in the background, so we do not need to worry about as many things - there are hundreds of people to thank. However, specifically, I want to thank Louise Lattimore and Rebecca Chaplin for their help with organising and arranging so many things over the years. Dr Loren Gibson, Dr Erica Ballabio and Dr Rocco Coppejans for dealing with the many MTA amendments, ethical approvals and making sure everything was up to the standards needed. Heather Stanley, thank you for arranging all the couriers and shipments - there were many! Cellular genetics wet-lab support and informatics teams, as well as Sanger Core Sequencing pipeline, thank you for the crucial support with sample processing, sequencing library preparation and data processing. Everything would have taken 10x longer without the work of your teams.

Collaborators
I would also like to thank Professor Linda Griffith and her team and Dr Juan Gnecco for sharing the hydrogel they developed with us, so I could test it in our lab. Thank you for making it happen, sharing your knowledge, and the many interesting discussions about organoids and the endometrium in general.

Funders
The work in this thesis was made possible thanks to the Oxford Medical Research Council Doctoral Training Programme (MRC-DTP) and the Nuffield Department of Women's and Reproductive Health scholarship that I was awarded. On top of the scholarship, I am also grateful for the supplementary funding I was awarded for extra training, doing a science communication internship and the 6-month transition period following thesis submission through the Oxford MRC-DTP. Thank you also to Kellogg College for travel awards to attend conferences and the Goodger and Schorstein Research Scholarship for covering my stipend. Thank you is also warranted to the Human Uterus Cell Atlas Project funded by the European Union's Horizon 2020 Progamme and the John Fell Fund from the University of Oxford, both of which funded most of the experimental work presented in this thesis.

My family
Mami, tati a Zuzi - Vám trom patrí veľká vďaka za podporu počas celého štúdia, či osobne alebo virtuálne, za rady do života a pripomenutia perspektívny a toho na čom ozaj v živote záleží. Zuzi, tebe patrí vďaka aj za všetku grafickú a vizuálnu pomoc, a využitie tvojho kreatívneho ducha a potenciálu na nádherné stvárnenie endometriózy, bolesti a všetkého, čo s tým súvisí v našom edukačnom videu. Ďakujem aj babkám Božke a Zuzke, a dedkom Antonovi a Jánovi, ktorí vždy hovorili, že študovať je dôležité a podporovali ma, ako vedeli. Dúfam a verím, že ich prvý doktor(ne-doktor) v rodine poteší a tých, čo medzi nami už nie sú by potešil.

Finally, *çok teşekkür ederim* to my partner Tolga. Thank you for (*w*)always being there for me and never complaining. Somehow you put up with me while I was writing my MSc thesis in Australia, happily moved across the globe to witness the whole DPhil journey and made sure that while writing the DPhil thesis I remained sane and well fed. Words cannot express how grateful I am to have you in my life!

Table of contents

Abbreviations

BSA	bovine serum albumin
cAMP	cyclic adenosine monophosphate
dS	decidualised stromal cells
E2	oestrogen
EEO	endometrial epithelial organoids
EEO-ES	endometrial epithelial organoids-endometrial stromal cells
EMT	epithelial mesenchymal transition
ENDOX	Endometriosis Oxford study
ERA	endometrial receptivity array
eS	endometrial stromal cells
FACS	fluorescent-activated cell sorting
FBS	foetal bovine serum
FENOX	Fibroids and Endometriosis Oxford study
Fib	fibroblast
GWAS	genome-wide association study
MPA	medroxyprogesterone acetate
OCT	optimal cutting temperature
P4	progesterone
PBS	phosphate buffered saline
PRL	prolactin
Prolif	proliferative
PV	perivascular cells
rASRM	revised American Society for Reproductive Medicine
scRNAseq	single-cell RNA-sequencing
SNP	single nucleotide polymorphism
snucRNAseq	single-nucleus RNA-sequencing
Stim	stimulated
UMAP	uniform manifold approximation and projection
Unstim	unstimulated
WSI	Wellcome Sanger Institute

Throughout this thesis, I use the term ‘woman/women’ to refer to individuals whose sex assigned at birth was female, regardless of their gender. Historically, studies mentioned in this thesis did not distinguish between sex and gender when studying the endometrium and endometriosis in individuals born with an uterus. The lack of distinction between sex and gender makes it therefore difficult to use inclusive language in this thesis. However, I would like to acknowledge and stress its importance and hope for more research and use of inclusive language in the field of endometrial and endometriosis research in the future.

Chapter 1: Introduction

1.1 Human endometrium in health

The human endometrium is the innermost mucosal lining of the uterus that provides the site for embryo implantation and nutrition to the embryo during the first weeks of pregnancy. Thus, this tissue is crucial for our species preservation. The endometrium is a highly dynamic tissue that undergoes cycles of shedding, repair, and differentiation in preparation for pregnancy. Under the influence of steroid hormones, the endometrium proliferates and undergoes extensive morphological changes, growing as much as 5-10 mm in just 7 days every month and is capable of full regeneration without scarring upon parturition[1]. In the absence of pregnancy, the superficial (functional) layer of the endometrium sheds and repairs in the next menstrual cycle. It is estimated that during a woman's reproductive years, the endometrium undergoes around 400 cycles of such extensive proliferation, shedding and repair[2].

1.1.1 Endometrial morphology

Morphologically, during the reproductive years, the endometrium is composed of two layers: the upper *functionalis* layer (adjacent to the uterine cavity) and the bottom *basalis* layer (adjacent to the myometrium) (Figure 1.1). The *functionalis* layer sheds during menstruation and is believed to regenerate from the *basalis*[3–7]. At the cellular level, the endometrium has a heterogeneous architecture with a single layer of pseudostratified columnar epithelial cells lining the uterine cavity (luminal epithelial cells), while tubular glands extend from the surface all the way to

the endometrial-myometrial junction (glandular epithelial cells) where a network of complex, horizontally interconnected *basalis* glands is observed[8–10](Figure 1.1). Ciliated epithelial cells are also present and their numbers change in response to steroid hormones across the cycle[11]. Stromal, fibroblast, perivascular as well as endothelial cells provide support and structural integrity, including rich vasculature within the tissue. Moreover, a wide array of immune cells play crucial roles in endometrial shedding and repair, as well as embryo implantation and their numbers change substantially across the cycle and in pregnancy[12,13]. In addition, across the cycle, stromal cells undergo substantial morphological and functional transformation, known as decidualisation, in order to secrete factors required for embryo implantation and induction of immunotolerance in the maternal immune system[14]. Despite many years of research, how the endometrium achieves its immense regenerative capacity remains only incompletely understood, and we are still in search for specific markers, better molecular characterisation and *in situ* location of putative endometrial stem/progenitor cells[15].

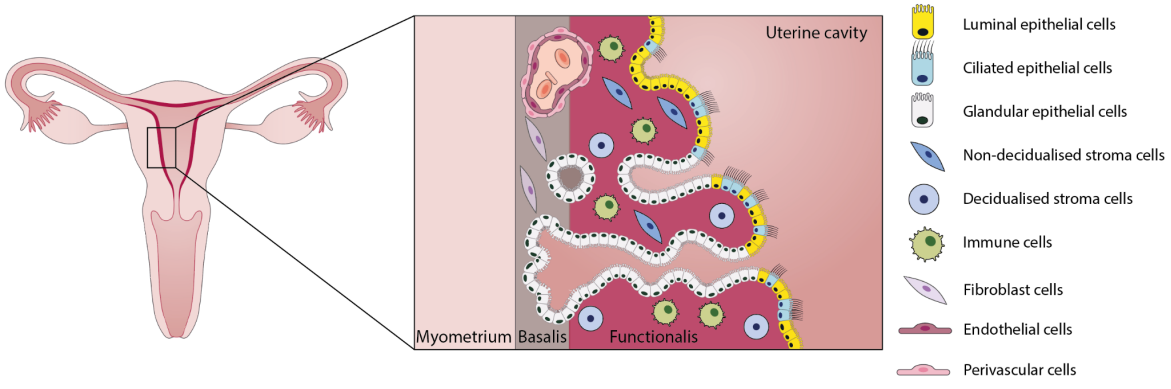


Figure 1.1: Endometrial cellular morphology. This figure shows the relative locations of the *basalis* and *functionalis* layers of the endometrium in relation to the myometrium and the uterine cavity. Cellular composition, including the location of the different types of epithelial, stromal and immune cells is also shown. *The original figure was commissioned from Antonio García, scientific illustrator at Bio-Graphics and is adapted here.*

1.1.2 The menstrual cycle

As described above, the *functionalis* layer of the human endometrium undergoes extensive morphological changes during the menstrual cycle, which conventionally lasts 28 days[16]. The menstrual cycle can be divided into 3 major phases: menstrual (day 0 - 5), proliferative (day 5 - 14) and secretory (day 14 - 28). However, the length of the cycle and its different phases vary greatly among and within women[17–19]. Broadly speaking, the cycle begins with the menstrual phase during which the *functionalis* layer sheds and repairs rapidly while the breakdown is still underway[20]. Re-epithelialisation of the endometrial surface is completed within 48 hours after shedding[21], followed by the proliferative phase characterised by prominent expansion of the epithelial glands, stroma and vasculature under the influence of rising oestrogen levels produced by the ovaries[22]. As shown in Figure 1.2, initially, the glands appear short and straight, swiftly elongating and acquiring more complex, tortuous morphology towards the end of the proliferative phase[23]. After ovulation, progesterone is secreted by the corpus luteum within the ovary and its levels start to rise. The endometrial cells respond by stopping their growth and becoming more mature. They differentiate into secretory cells able to produce secretions needed for embryo implantation and histiotrophic nutrition[24]. These effects are mediated by the response of both the stromal and epithelial cells to progesterone in a direct manner, as well as paracrine signalling and cross-talk between these cell types. In the absence of a pregnancy, regression of the corpus luteum leads to a drop in progesterone production and its levels drop, causing the *functionalis* to shed, initiating the menstrual phase of the cycle again[25].

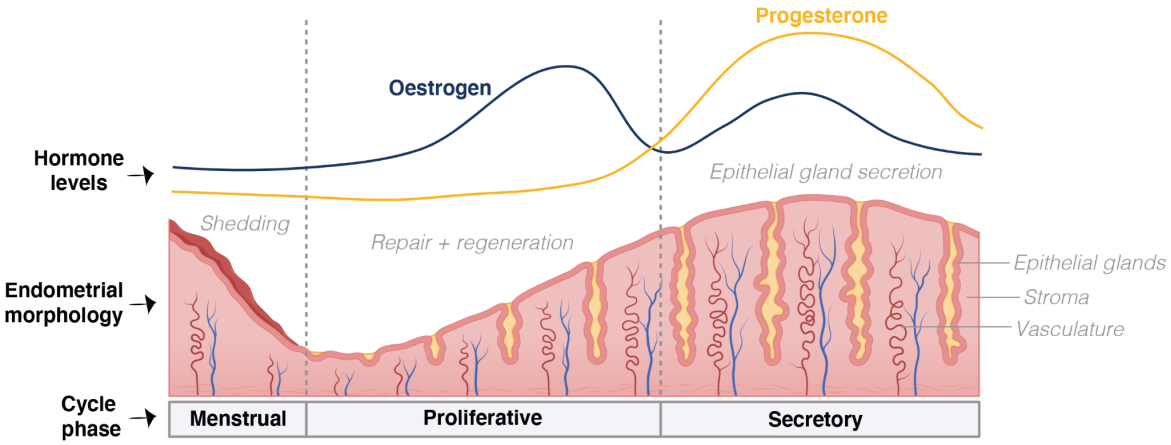


Figure 1.2: Schematic overview of the endometrium across the menstrual cycle. The endometrium sheds during the menstrual phase of the cycle, followed by repair and regeneration during the proliferative phase, under the influence of rising oestrogen levels. During the secretory phase, as progesterone levels rise, the epithelial glands start to release secretions needed for embryo implantation. *Figure created in BioRender using the available presets for endometrium and thus the structure of the basalis glands is not accurately represented (see Figure 1.1).*

1.2 Human endometrium in disease

The previous section described the physiologically normal functioning of the endometrium across the menstrual cycle. Yet, we still lack full understanding of how the crosstalk between the endometrial, immune and vascular cells under the influence of steroid hormones ensure the menstrual cycle proceeds as it should. Better understanding of the processes driving endometrial regeneration in health could help us understand the common pathologies affecting the uterus, including menstrual disorders (such as heavy and abnormal uterine bleeding), uterine fibroids, endometrial polyps as well as endometriosis and adenomyosis. The next sections discuss what is known about the role of the human endometrium in disease, focusing on endometriosis - a common condition described by the presence of endometrial-like cells (also known as lesions) outside of their normal location, the lining of the uterus.

1.2.1 Endometriosis and its pathogenesis

In ~10% of women of child bearing age[26] (~190 million women worldwide), cells resembling the endometrium are found outside of the uterus[27]. This condition is called endometriosis and is characterised as a chronic inflammatory disease. Endometriosis causes subfertility, debilitating chronic and cyclical pain[28] and is estimated to cost the UK economy £8.2 billion a year in healthcare and loss of work[29]. Despite its high prevalence and economic costs, global studies have shown an average time between onset of symptoms and first diagnosis of 8-12 years[30,31].

Clinically, endometriosis affects every individual differently and symptoms can often be mistaken for other diseases, e.g. gastrointestinal issues, urinary tract infection, cancer, etc.[32]. The lack of reliable diagnostic tools as well as disease treatment is strongly associated with poor understanding of the pathogenesis of endometriosis. Many theories exist for what causes the disease (reviewed in[33]) but the most widely accepted theory for endometriosis pathogenesis is the one of 'retrograde menstruation.' During retrograde menstruation, the shed endometrial tissue and blood travels up the Fallopian tubes into the peritoneal cavity instead of leaving the body through the vagina. The shed tissue is believed to attach to the peritoneal lining and organs, forming endometriotic lesions[34] (Figure 1.3). However, retrograde menstruation occurs in 90% of all women[35] and only 10% of women develop endometriosis. It is not clear what makes some women more susceptible to develop endometriosis than others. Different properties of the shed tissue in women with endometriosis, the lack of appropriate immune cell clearance of the shed tissue as well as higher peritoneal cell receptivity are thought to be involved, reviewed in[27,33]. Yet, there are cases in which retrograde menstruation cannot explain the presence

of endometriosis and different theories for its origin have been postulated. For example, endometriosis in women with Müllerian duct defects where there is no reflux of menstrual debris has been suggested to occur as a result of coelomic metaplasia, i.e. the transformation of peritoneal mesothelium into endometrial epithelium[36]. In the case of extrapelvic endometriosis, lymphatic and vascular metastasis, i.e the transport of endometrial cells through lymphatic and blood vessels has been proposed as another theory for endometriosis origin[37]. From a genetics perspective, collaborative genome-wide association studies (GWAS) have identified variants that increase the risk of developing endometriosis[38]; however, the contribution of these factors to disease development or to endometrial biology is still unclear.

1.2.2 Endometriosis - current diagnosis and treatment

Historically, the only way to reliably diagnose endometriosis was by performing an invasive surgery (laparoscopy) and histologically verifying the presence of endometrial-like cells in the specimens excised[39,40]. Recently, the development of more accurate imaging technologies and careful consideration of patients' medical history has enabled a less invasive route for endometriosis diagnosis, particularly for ovarian and deep disease; however, it is neither sensitive nor specific enough for superficial endometriosis[41–43]. The current diagnostic delay of 8-12 years often has crippling effects on the individuals affected, as well as their partners and families. For many years, people with endometriosis can suffer from dysmenorrhea, dyspareunia, dyschezia, chronic pelvic pain and subfertility, all of which can have a significant impact on their quality of life[44].

Current endometriosis treatment is limited to (repeated) surgery, hormonal and analgesic drugs with morbidity and many side-effects to alleviate and manage the patients' symptoms[45,46]. Surgical removal of the lesions is thought to improve pain symptoms and help increase chances of conceiving. However, the first line treatment is usually prescribing hormonal therapy (such as the pill) to manage pelvic pain[47]. Taking oral contraceptives or progestogens can be effective for many, but does not always work long-term or sufficiently for many others that require second line hormonal treatment (such as gonadotrophin releasing hormone (GnRH) analogues)[48–50]. GnRH analogues are very potent drugs that inhibit ovulation, resulting in oestrogen levels comparable with menopause and as such have severe side-effects[51]. Yet, it is unclear whether the endometriotic lesions' growth rate and progression patterns change when patients take these hormones, even though their pain symptoms are alleviated[45]. Furthermore, we also lack knowledge about the effects these hormones have on the endometrium and how this might differ based on the type and dose of hormones taken. Given millions of women world-wide take exogenous hormones either for contraceptive reasons or as treatment options for a multitude of reproductive conditions, studying the molecular changes exogenous hormones have on the endometrium in health and disease states is long overdue.

1.2.3 Endometriotic lesions and disease staging

Endometriotic lesions found within the abdominal cavity have a myriad of different features and appearances, but are generally grouped under 3 main categories: peritoneal (or superficial) lesions, ovarian lesions (known as endometriomas) and deep endometriosis lesions (Figure 1.3)[52]. The common feature for all lesions is the presence of endometrial-like epithelial and/or endometrial-like stromal cells[53]. The

establishment, maintenance and response to treatment of the different lesions is not well understood, nor whether and how endometriosis progresses.

To systematically describe the heterogeneous surgical presentation of endometriosis, the extent of disease is classified based on the number, size and location of the endometriotic lesions, as well as pelvic adhesions identified during surgery[54]. This most-widely used staging approach is referred to as the revised American Society for Reproductive Medicine (rASRM) staging system[55], and it classifies patients into four stages: stage I (minimal), stage II (mild), stage III (moderate) and stage IV (severe) disease. The stage is calculated based on an arbitrary points system, where different surgical findings are assigned different numbers of points which are then added up (Figure 1.3). Unfortunately, minimal correlation has been observed between this staging system and severity of symptoms, disease prognosis or progression. Moreover, the staging system mainly takes into account superficial lesions and thus fails to capture deep endometriosis, including the localisation and size of deep endometriotic lesions[56]. As discussed, endometriosis as a disease, including its aetiology and pathophysiology, still remains enigmatic. This lack of knowledge hinders diagnosis without delays and efficient treatment for endometriosis. The development of non-invasive diagnosis methods and non-hormonal targeted treatment methods is currently a key priority for both women and clinicians.

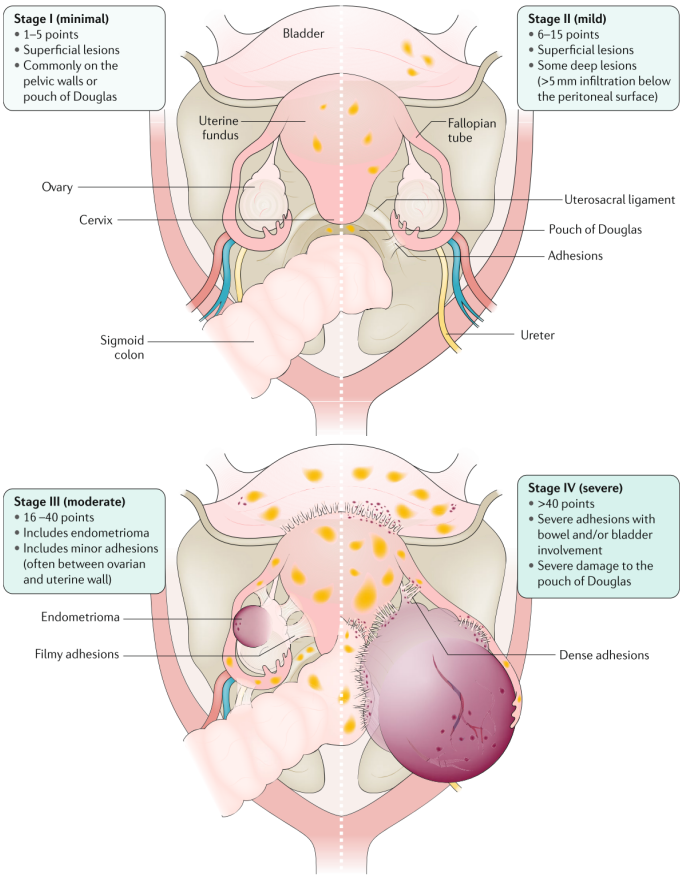


Figure 1.3: Overview of endometriotic lesions and endometriosis staging. The types of lesions found, and points assigned for each of the four endometriosis stages (I - IV) during surgery. © *Zondervan et al. (2018), Endometriosis. Nat Rev Dis Primers 4, 9*[57]*. With permission from Springer Nature, licence number 5425580956020.*

1.2.4 Other common uterine and endometrial pathologies

Apart from endometriosis, there are many other common endometrial and uterine pathologies, visualised in Figure 1.4. Even though these conditions are considered benign, they have a significant impact on the quality of life of those affected. Adenomyosis, for example, affects ~20% of women[58] and is often the reason why women undergo hysterectomy[59]. The aberrant growth of the endometrial epithelial and stromal cells into the myometrium can cause debilitating pain, heavy bleeding, and subfertility[60]. Another condition mostly localised to the myometrium and associated with heavy bleeding and significant pain is uterine fibroids. Fibroids are abnormal growths of muscle and connective tissue diagnosed in around 70-80% of

women in their reproductive years[61]. Fibroids can grow in different locations of the uterus and are classified accordingly.

With regards to endometrial pathologies, hyperplasia is a condition in which the endometrium proliferates irregularly, resulting in an increase in the gland-to-stroma-ratio[62]. Endometrial hyperplasia often presents with abnormal uterine bleeding, and if left untreated has the propensity to progress to endometrial cancer[63]. Similarly, endometrial polyps are associated with abnormal bleeding, and in some cases can progress to malignancy[64,65]. Polyps are outgrowths of endometrial glands, stroma and blood vessels that protrude into the uterine cavity and are reported to be the most common pathological finding in the uterus[66]. Yet, their exact incidence is not known. Likely, the true prevalence of all these endometrial/uterine conditions described is underestimated and unknown. However, it has been reported that there is substantial comorbidity between the different conditions, with endometriosis and adenomyosis as well as endometriosis and fibroids and adenomyosis and fibroids being co-diagnosed to a significant degree[67,68]. A better understanding of these conditions and their effects on uterine and endometrial function is needed.

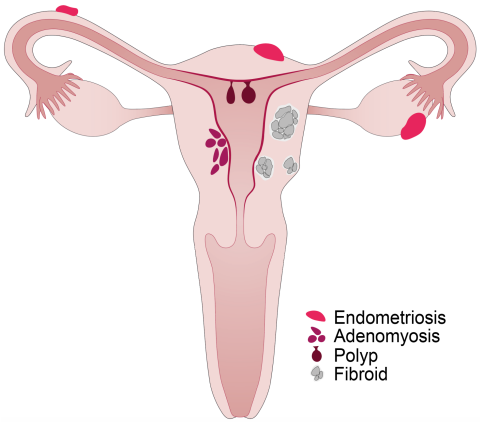


Figure 1.4 Common endometrial and uterine pathologies. A schematic overview of endometriosis, adenomyosis, endometrial polyps and uterine fibroids in relation to the reproductive tract. *The original figure was commissioned from Antonio García, scientific illustrator at Bio-Graphics and is adapted here.*

1.2.5 Effect of exogenous hormones on endometrial morphology

Exogenous hormones are routinely prescribed as a treatment for the pathologies described above and include commonly used contraceptives. A plethora of contraceptive options exists, with different hormonal formulations and modes of administration having different effects on the endometrium and its morphology. For example, the combined oral contraceptive pill (COCP) composed of a combination of oestrogen and progestin is taken orally for 21 days, followed by a 7-day tablet-free period during which bleeding occurs. COCP suppresses ovulation and induces the presence of narrow, straight epithelial glands with very little mitotic activity surrounded by pseudodecidualised stroma[69,70].

In contrast, progestin-containing contraceptives are inconsistent in inhibiting ovulation and mostly work by thinning the endometrium and thickening the cervical mucus[71]. They can be administered orally, as subdermal implants or as an intrauterine device (IUD) and their effects on the endometrium vary based on the duration of treatment, dose administered and partly on the type of progestin used. The progestogen-only pill (POP) contains a progestin called levonorgestrel and is taken continuously, with women experiencing light bleeds which may initially be irregular. POP stops endometrial growth, induces variable stromal pseudodecidualisation and with prolonged use can cause glandular atrophy, abundant stromal pseudodecidualisation and presence of numerous thin-walled blood vessels[70,72,73]. With regards to progestin implants, atrophic, weakly proliferative and pseudodecidualisation changes have been described[69]. Similar changes together with infiltration of immune cells are also detected when the

progestin-releasing IUD (also known as the Mirena coil) is used and in some women, ovulation may be inhibited[74–76].

Lastly, as described above for the treatment of endometriosis, administration of GnRH analogues, such as goserelin acetate (drug name Zoladex) leads to suppression of ovulation and ovarian hormone secretion causing endometrial atrophy[77].

Taken together, the effects of exogenous hormones on both ovarian hormone production and endometrial morphology are varied and require careful consideration and extensive characterisation of the available clinical data and endometrial phenotypes in order to be able to group patients in a meaningful way and describe the morphological and transcriptomic response of endometrial cells to exogenous hormones.

1.3 Human endometrium *in vitro*

Menstruation is limited to ~1.6% of known eutherian species, including humans, Old World primates, elephant shrews, a few species of bats and the spiny mouse[78,79]. Meaning, no readily available animal models exist that can accurately mimic the dynamic cyclical changes of the human endometrium. This lack of appropriate models to study the endometrium in both health and disease poses a significant challenge, limiting us to characterising biopsies of the endometrium collected during clinical studies. This brings several difficulties, such as: (i) collecting samples only at a particular time/phase of the menstrual cycle, (ii) studying mostly superficial biopsies of the endometrium and not the deeper *basalis* layer and the myometrium

unless women undergo hysterectomy, and (iii) not being able to perform perturbations on these samples by for example manipulating them genetically or pharmacologically.

To overcome some of these hurdles, *in vitro* model systems of the human endometrium were established and have been reviewed in detail[80–82]. Historically, cells have been cultured using conventional two-dimensional (2D) cultures of the two major cell types: epithelial and stromal cells. However, growing epithelial cells long-term as 2D monolayers has been hard to achieve and only possible by culturing immortalised cell lines or cancer cell lines that no longer accurately represent the naive cellular states of the endometrium as they are karyotypically and physiologically abnormal[83]. In contrast, stromal cells grow readily as 2D monolayers (Figure 1.5). However, their morphology and properties are different to those in their three-dimensional (3D) *in vivo* environment and are also altered when cultured long-term. Recent years have seen the development of 3D culture systems, including the endometrial epithelial organoids[84,85], co-cultures of epithelial organoids and stromal cells[86,87] and a vascularised organ-on-a-chip system[88] attempting to recreate the complex physiological processes of the endometrium *in vitro*. These systems are opening new frontiers for the study of the endometrium in both health and disease. The following sections describe in more detail the 3D organoid and co-cultures systems that I used and profiled during my studies.

1.3.1 Endometrial epithelial organoids

Organoids are self-assembled multicellular 3D structures that recapitulate important aspects of a particular organ's epithelium development and function[89]. Organoids have been established for a multitude of human organs[90], including those of the

reproductive tract[91]. The establishment of endometrial epithelial organoids was for the first time reported in 2017 by two groups[84,85] that embedded endometrial epithelial cells in Matrigel and cultured them in chemically-defined media (Figure 1.5). These organoids could be expanded long-term and were reported to be genetically stable, as well as hormone-responsive, which is a key feature of the endometrium *in vivo*. To date, endometrial organoids have been derived from healthy endometrium, hyperplastic endometrium, pregnant decidua, menstrual fluid, endometrium from pathologies such as Lynch syndrome, cancer and endometriosis, as well as from the cancer tumours and endometriotic lesions themselves[84,85,92,93]. Importantly, organoids from pathologies maintained the disease phenotype *in vitro* and in the case of cancer organoids showed a patient-specific drug response[92]. However, the endometrium is made of more than just epithelial cells, prompting researchers to engineer systems in which multiple cell types, such as immune, stromal and endothelial cells, can be co-cultured and all thrive.

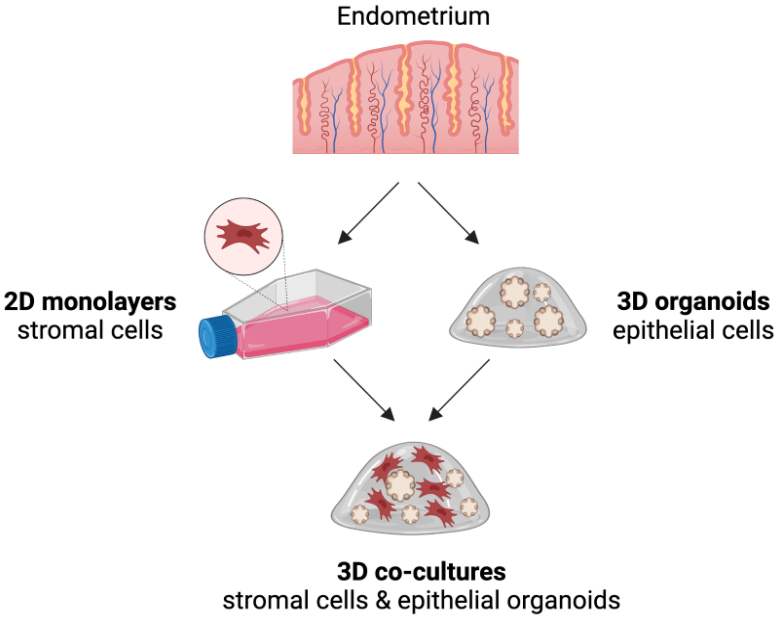


Figure 1.5: Overview of *in vitro* endometrial cell models. Endometrial biopsies are separated into stromal and epithelial cells and grown *in vitro* as either 2D monolayers (stromal cells) in flasks or 3D organoids (epithelial cells) in Matrigel domes. Once expanded, the stromal cells and epithelial organoids can be grown together in co-culture systems. *Figure created in BioRender.*

1.3.2 Endometrial epithelial and stromal cell co-cultures

Co-culturing of epithelial and stromal cells has been reported using various approaches, including scaffold-free cultures[94], porous collagen scaffolds seeded with cells[95] and embedding epithelial organoids and stromal cells in synthetic hydrogels[86,87]. In the scaffold-free cultures, cells self-assemble into spheroids made of a core of stromal cells, surrounded by a layer of epithelial cells. In the porous collagen scaffolds, stromal cells are first allowed to adhere to the scaffold, followed by seeding with epithelial organoids initially grown in Matrigel. A similar approach of assembling already established epithelial organoids and stromal cells is also utilised by embedding these inside synthetic hydrogels[86,87](Figure 1.5). The major difference is that the synthetic hydrogels are chemically fully-defined and are not of animal origin (unlike the collagen scaffold approach) and their stiffness can be fine-tuned to match that of the endometrium *in vivo*[86].

Whilst all the different models and approaches were reported to generate cells that respond to hormones, they all likely capture slightly different aspects of endometrial biology and require deeper characterisation and further optimisation. That includes understanding which subtypes of stromal and epithelial cells are in the cell culture (e.g. glandular vs luminal epithelium). All of the described co-culture systems were found to disintegrate in culture after a short period of time, and long-term maintenance of these systems remains a challenge. In my studies, I used one of the hydrogel systems in which organoids can be cultured for up to 15 days[86]. Its properties are described in more detail in Chapter 5. Briefly, the main advantages of using synthetic hydrogels for the co-culture include the possibility to: (i) functionalise them with peptides that will support the growth of the different cell

types, (ii) fine-tune their composition and stiffness to accommodate multiple cell types, and (iii) keep their composition reproducible across all batches. This is in contrast to using Matrigel, which varies in its composition batch-to-batch given its animal cancer cell line origin. Moreover, the animal-free origin of the synthetic hydrogels makes them more suitable for clinical and translational applications further down the line. The synthetic approaches therefore provide more reproducible and robust options for endometrial cell culture and are likely to be utilised and adjusted extensively over the coming years.

In summary, being able to derive patient-specific long-term propagating cellular models for many individuals opens up new avenues to study endometrial function, disease pathology as well as drug response in a clinically-relevant personalised manner. However, in order for these systems to be useful for translational applications, we need to be sure they model the processes of interest accurately and reliably. The next section describes how profiling the endometrium at the single-cell level, both *in vivo* and *in vitro,* can provide powerful insights into not only the cellular composition and function of the endometrium, but also how we can model, evaluate, and screen these *in vitro* models in a high-throughput manner.

1.4 Endometrium in health, disease, and *in vitro* at the single-cell level

As described at the beginning of this chapter, the human endometrium is a highly dynamic and heterogeneous tissue, made of many different cell types. In order to understand what cell types it is made of and their role in normal endometrial function as well as disease, the endometrium has to be studied at the single-cell level. When I started my DPhil studies, there was a lack of published research work on

endometrial cellular heterogeneity, with only one publication profiling ~70 cultured endometrial stromal cells available[96]. Since then, multiple manuscripts profiling the endometrial cells in health and disease *in vivo* and cultured *in vitro* have emerged and are described below. Here, I first describe the basics of single-cell genomics technologies and discuss in more detail the three major studies analysing the human endometrium at the single-cell level published in the last two years[97–99].

1.4.1 Single-cell genomics technologies

No two cells in the body are the same even though they contain the same genetic information. Each cell utilises the genome in a different way to fulfil its specialised function based on its unique spatio-temporal context. The way cells respond to their environment can be detected by analysing their gene expression using RNA sequencing. Assessing the messenger RNA (mRNA) content of cells (i.e. the transcriptome) at the single-cell level, however, was not possible until 2009[100]. Historically, gene expression studies characterised tissues using bulk genomics technologies, obtaining the average gene expression across multiple cell populations. These bulk RNA-sequencing approaches provided important insights into many biological processes; however, could not capture cell-specific gene expression signatures and identities of multiple cell types in parallel. To overcome this obstacle, single-cell RNA-sequencing (scRNAseq) was developed followed by many other single-cell genomics technologies profiling further modalities, such as the genome[101–105], epigenome[106–109] and proteome[110–112] at cellular resolution. Since their development, scRNAseq and single-cell genomics have revolutionised our understanding of cell identity and tissue composition.

1.4.1.1 Single-cell transcriptomics

Single-cell transcriptomics, the most commonly profiled single-cell modality, was developed to analyse the cellular composition of a blastomere[100], and has since evolved rapidly - it is now possible to analyse transcriptomes from tens of thousands of cells in parallel[113]. Such detailed gene expression analyses have led to the discovery of many new cell states and rare cell populations[114–117]. A plethora of protocols exist for the isolation and capture of mRNA from single cells, but by far the most popular method today relies on using microfluidic devices to encapsulate single cells with barcoded gel beads inside microdroplets[118,119]. Once both a cell and a bead are inside a droplet, the cell is lysed and its mRNA molecules are labelled with a barcode that is unique to each bead, ensuring no two cells' contents are labelled with the same barcode. A commonly used commercial platform utilising such a droplet-based approach is the Chromium platform from 10x Genomics, also used in this thesis. Alternative methods include plate-based single-cell genomics (e.g. Smart-seq3[120]), where cells are physically separated into 96 or 384 well plates, one cell per well, prior to cell lysis and mRNA capture.

The advantages/disadvantages of both methods have been reviewed in detail[121,122], thus I only describe a few key differences here. In contrast to droplet-based technologies, plate-based methods do not have a size limitation and can be utilised to analyse large cells that would not fit inside the microdroplets. Moreover, they allow for full-length transcript capture and sequencing, while the 10x Genomics droplet-based protocols only capture either the 3' or 5' ends of polyadenylated transcripts. However, only a limited number of cells can be profiled using plate-based methods, while tens of thousands of cells can be analysed using the droplet-

based approaches. In both cases, once the cells' mRNA contents have been captured, the mRNA is reverse transcribed into complementary DNA (cDNA). The cDNA is then multiplied by polymerase chain reaction (PCR) to increase the minute mRNA yields from single cells and used to construct sequencing libraries. These libraries are most commonly sequenced using short-read sequencing platforms, and the single-cell transcriptomic data is aligned to the genome reference of choice and analysed.

In order to obtain high quality single-cell transcriptomic data, cell viability has to be high upon tissue dissociation into single cells. Initially, scRNAseq was therefore performed exclusively on freshly collected samples and later also cryopreserved ones. This limited the samples one could process and analyse. With improvements in protocols and technologies, it became possible to also profile snap-frozen samples where isolating viable single cells is technically challenging. In this case, RNA sequencing of mRNA from single nuclei (snucRNAseq) can provide a useful alternative[123], and opens up the exciting possibility of exploring frozen archival tissues stored in biobanks. However, single-nuclei data has its own unique characteristics, such as lower input material than whole cells where both the nuclear and cytoplasmic RNA is analysed. Moreover, there are differences between the type and proportions of RNA found in the nucleus vs cytoplasm[124] and higher proportion of ambient RNA in snucRNAseq data released from the cytoplasm during the nuclei extraction protocol. Nonetheless, it has been shown that single nuclei contain enough informative transcripts to identify the various cell types in the brain[125,126] as well as the endometrium[98].

1.4.1.2 Spatial transcriptomics and emerging technologies

While the above described single-cell methods have provided remarkable insights into cellular diversity, they rely on tissue dissociation, meaning information about the cells' spatial context is lost. Spatial transcriptomics allows one to locate specific cell types *in situ*, within intact tissue sections. Recently, many spatial transcriptomics technologies with distinct sensitivity and readout throughput have been developed, pushing the boundaries of resolution and field of view these methods can achieve[127]. Spatial transcriptomics technologies have been reviewed in detail recently[128,129], but broadly can be divided into: (i) sequencing-based, which label transcripts with their positional information prior to next-generation sequencing and (ii) imaging-based, which either rely on (a) *in situ*-sequencing that directly amplifies and sequences the transcripts inside the tissue or (b) *in situ* hybridisation that uses pre-designed probes that hybridise to the targets inside tissues[130]. Spatial transcriptomics technologies hold a great promise for better understanding of tissue architecture and intercellular communication, as well as how it is altered in diseases. Moreover, technologies providing access to multiple modalities from the same cell are helping to give a more holistic view of the cell. For example, measurements of both the genome and transcriptome modalities have been used at the single-cell level[131], as well as measurements of both the transcriptome and protein expression[132]. In the future, spatial profiling of all of the modalities (genome, epigenome, transcriptome and proteome) from a single cell will likely be feasible. Such spatial multi-omics data could provide a more complete and mechanistic view of each and every cell within their tissue microenvironment.

1.4.2 Human endometrium at the single-cell level

The last two years have seen an increase in the number of publications profiling human endometrial cells at the single-cell level. The focus of these has varied greatly. Some studies characterised cultured endometrial stromal[96,133] and perivascular[134] cells showing that indeed culturing cells *in vitro* alters their transcriptomic profiles when compared to the *in vivo* cells[96,134]. In one of the studies, stromal cells were grown from both healthy donors and donors with endometriosis. The authors reported an aberrant differentiation pathway in one of the stromal fibroblast populations in women with endometriosis, suggesting that this could increase the capacity of these cells to initiate endometriotic lesion formation and growth[133]. Validation of these findings in women with endometriosis in an *in vivo* context is now warranted given cell culturing can introduce artefacts and phenotypes that may not occur outside of the *in vitro* context.

Other studies have profiled the endometrium in a set of other pathological conditions, including thin endometrium[135,136], adenomyosis[137], recurrent pregnancy loss and implantation failure[138–140] as well as endometrial cancer[141,142]. In most of these studies, the sample size was low, ranging from 1 to 3 patients per condition studied and low numbers of cells analysed. Oftentimes the samples came either from the same menstrual phase of the cycle or the phase was not considered. Given these limitations, it is difficult to judge the reported results and dysregulated pathways and phenotypes. For this reason, these papers are not discussed here in detail.

1.4.2.1 Endometrium across the menstrual cycle

Publications described in detail include the only two studies looking at the endometrium in health across the menstrual cycle[97,98]. Wang *et al*.[97] focused on characterising the transcriptomic changes of the *functionalis* layer of the endometrium - the layer that sheds and regenerates across the menstrual cycle. The authors studied the main dynamic changes of gene expression across the cycle in the epithelial and stromal cell compartments. They identified sudden and abrupt changes in gene expression in epithelial cells (upregulation of *PAEP*, *GPX3* and *CXCL14*) marking the opening of the window of implantation (WOI), while the changes observed in the stromal cells (upregulation of *FOXO1*, *IL15*) were more continuous and observed already earlier on in the cycle[97]. Most of their results were; however, based on characterising only ~2,000 cells from 19 donors across the cycle. They generated a further dataset of ~70,000 cells predominantly from secretory phase samples to confirm their findings for the WOI markers but the data was not integrated with the previous dataset to corroborate other findings. This dataset is discussed in more detail in Chapters 2 and 3 of this thesis.

The second study of the endometrium across the cycle, performed single-cell and spatial transcriptomics on full-thickness endometrial biopsies from organ donors, as well as single-cell transcriptomics on biopsies of the *functionalis* layer in living donors[98]. Obtaining the spatial information allowed the authors to give spatial coordinates to the cell states obtained and characterise the cell-cell communication processes mediating epithelial identity in the three main endometrial layers: (i) the *basalis* layer, enriched in endometrial progenitor cells; (ii) the *functionalis* layer, enriched in glandular secretory cells, and (iii) the luminal layer, enriched in ciliated

and luminal cells. Garcia-Alonso *et al.* showed that the luminal epithelium is defined by high WNT and low NOTCH expression, while the opposite is true for the glandular epithelium, and postulated this has an impact on epithelial cell identity. Using endometrial epithelial organoids, the authors demonstrated that the low NOTCH environment characteristic of the lumen is essential to maintain ciliated identity while the low WNT environment present in the glands is key to acquire a glandular secretory cell fate. Moreover, this study described further epithelial and mesenchymal subsets across the menstrual cycle, and these are described in more detail in Chapters 2 and 3.

1.4.2.1 Endometrium in endometriosis

Recently, a few studies have started looking into the cellular composition of the eutopic endometrium from patients with endometriosis. Mostly, though, these studies focused on describing the cellular composition of the ectopic endometriotic lesions themselves[99,143,144], and have identified distinct cell type composition and expression differences between peritoneal lesions and ovarian endometriomas, specifically in their angiogenic and immune microenvironments[99,144]. With regards to the endometrium from healthy women and those with endometriosis, studies have reported dysregulation of the stromal and immune compartments in the endometrium of women with endometriosis to various degrees[99,143,145–147]. For example, comparing the endometrium of 3 healthy controls and 9 endometriosis cases (with most donors on hormones at sample collection), the Tan *et al*. study[99] reported dramatic changes in the cellular composition of the endometrium - in endometriosis, the proportion of epithelial cells was reduced and replaced by stroma and lymphocytes. Moreover, an increase in cell-cycle-related gene expression and proliferation of fibroblasts in endometriosis and specifically, higher expression of

osteoglycin (*OGN*) was reported. Heterogeneity of these findings across individuals was reported as well as the existence of a new epithelial cell population, termed MUC5B$^+$, found both in eutopic endometrium and ectopic lesions. MUC5B$^+$ cells specifically expressed *RUNX3*, *TFF3* and *SAA1*, hinting at their role in epithelial repair.

In contrast, Ma *et al*.[143] reported the endometrium of controls and cases from the proliferative phase of the cycle (*n* = 3 for each group) to be generally similar, observing only some differences in endometrial fibroblasts. Fibroblasts from donors with endometriosis were reported to be enriched in functions including gland development and tissue morphogenesis, while functions of fibroblasts from controls were related to ATP and metabolic processes[143]. In another study, significantly reduced expression of *TYRO3* in stromal cells of endometriosis cases was reported based on the analyses of proteomics data and scRNAseq data from 3 controls and 3 endometriosis cases[147], suggesting dysregulated cell-cell interaction between immune and stromal cells. A dysregulated immune and stromal cell phenotype was also observed when analysing the menstrual effluent from controls (*n* = 9) and endometriosis cases (*n* = 11). It revealed that stromal cells from endometriosis cases were enriched in cells expressing pro-inflammatory and senescent phenotypes and lacked a subset of decidualised stromal cells that were abundant in controls[145]. Moreover, a striking reduction of uterine natural killer cells (uNK cells) in endometriosis cases was reported. This was not seen in a different study utilising scRNAseq data to deconvolute bulk RNA-sequencing datasets where, on the contrary, an increase in uNK cells was observed in endometriosis cases (especially stage III-IV) when compared to controls[146].

Taken together, all of the above studies have attempted to shed light on the cellular composition of the endometrium in health and endometriosis; however, with conflicting findings. Likely, this is due to the small number of samples profiled. Given the inherent heterogeneity of the endometrium both inter- and intra-individually, larger sets of samples need to be profiled to understand whether and what differences exist between the endometrial cells of those with and without endometriosis.

1.4.3 Endometrial epithelial organoids at the single-cell level

When it comes to studying the endometrial epithelial cells *in vitro*, three studies have characterised endometrial organoids using scRNAseq and studied the cells' response to hormonal stimulation as well as differentiation to different lineages[98,148,149]. Even though the studies varied greatly in experimental set-up, they were in agreement in observing an increase in secretory (expressing *PAEP*) and ciliated cells (expressing *FOXJ1*) upon stimulation with hormones that mimic the menstrual cycle[98,148]. The most detailed scRNAseq analysis utilised an automated classifier based on logistic regression that was trained on *in vivo* endometrium data to quantitatively annotate the *in vitro* organoid data. This analysis showed that endometrial organoids' early response to hormonal stimulation closely matches that of *in vivo* epithelial cells. However, the final differentiated glandular and secretory cells were a bit more difficult to recapitulate *in vitro*. In addition, combining clonal organoids (i.e. organoids derived from a single cell) and trajectory reconstructions, the authors demonstrated that, at least *in vitro*, the secretory and ciliated lineages arise from one common epithelial progenitor[98].

In the Tan *et al*. study into endometriosis[99], organoids were generated for both the endometrium as well the endometriotic lesions and profiled using scRNAseq. The authors did not compare the transcriptomic profiles of organoids derived from the endometrium of healthy controls to those with endometriosis. Instead, they showed that organoids derived from the for the first time identified MUC5B$^+$ cells grew in larger numbers and to significantly bigger sizes than those from MUC5B$^-$ cells. Overall, they reported a large proportion of the cells in culture (~70%) to express *MUC5B*. Considering marker gene expression, these cells were similar to the MUC5B$^+$ cells identified in their *in vivo* data, suggesting that indeed the MUC5B$^+$ cells are likely progenitors able to expand long-term *in vitro*. With regards to the co-cultures of organoids and stromal cells, one study[87] profiled these using scRNAseq and showed that stimulation of the model system with hormones led to decidualisation of the cells, closely resembling the mid-luteal phase endometrium.

In summary, *in vitro* models of the endometrium can play an important role in enhancing our understanding of the endometrium in health and disease, as these systems can be manipulated both genetically and pharmacologically. Single-cell genomics can be used to evaluate and improve endometrial *in vitro* systems by identifying key regulators and factors critical for successful recapitulation of a certain physiological or disease process *in vitro*, and also for high-throughput screening readout of these systems upon manipulation[150]. Comparing single-cell *in vivo* reference organ atlases with *in vitro* model ones can serve as a powerful tool to assess and enhance the systems' quality and guide their development by revealing any aberrant gene expression or missing cell types/states.

1.5 Hypotheses and aims

As described in this chapter, single-cell and spatial transcriptomics have provided extraordinary resolution for studying the dynamic function and cellular heterogeneity of the human endometrium in health, with some studies starting to provide insights into endometrial pathologies. In addition, endometrial epithelial organoids and their co-culture with stromal cells have opened new avenues to mechanistically dissect the underlying molecular processes important for endometrial regeneration and disease development. However, to truly understand the mechanisms inherent to disease, and to translate the findings obtained through using these cutting-edge approaches into therapeutic target discoveries, we need to analyse an adequate number of patients and tissue samples and link detailed clinical metadata with their transcriptomic signatures. To maximise biological insights, leveraging advanced computational methods is key in the analysis of these 'big data' datasets. Finally, even though the *in vitro* model systems described here have been used for some years, they urgently require further characterisation and benchmarking to understand what biological processes they can/cannot recapitulate, and which model systems are best suited for the research questions asked. This thesis aims to address some of the gaps in our knowledge described in the context of the human endometrium in donors with and without endometriosis.

Overall hypothesis: The human endometrium is made of many different cell types and states, the majority of which can be described by their unique transcriptomic signatures. The abundance of these cell states across the menstrual cycle and/or their transcriptomic signatures are altered in those with endometriosis when compared to those without endometriosis.

Overall aim: Characterise the transcriptomic signatures of endometrial cell states driving the regeneration and differentiation of the human endometrium, pinpointing any differences specific to endometriosis. To address this aim, I use single-cell transcriptomics in a large set of endometrial samples from women with/without endometriosis across the menstrual cycle.

More specifically, this thesis is divided into three experimental chapters (Chapters 3-5), addressing the following 3 main research questions:

Question 1: *Can a harmonised single-cell transcriptomic map of the endometrium be built by combining multiple datasets and data sources? (Chapters 3 and 4)*

> ***1.1:*** Are endometrial cell populations easily and consistently identifiable based on their transcriptomic signatures across multiple datasets?
>
> ***1.2:*** Can my newly generated dataset of the endometrium be integrated and analysed together with previously published scRNAseq datasets of the endometrium?
>
> ***1.3:*** Is it possible to profile snap-frozen endometrial biopsies previously banked using snucRNAseq?
>
> ***1.4:*** Can the different transcriptomic data sources (scRNAseq and snucRNAseq) be integrated and analysed together?

Question 2: *What is the transcriptome-defined cellular composition of the human endometrium across the menstrual cycle? (Chapters 3 and 4)*

2.1: Can new cellular states/heterogeneity be uncovered by analysing a large set of endometrial biopsies?

2.2: Are there differences in the cellular composition of the endometrium across the menstrual cycle in donors with and without endometriosis?

2.3: What are the cellular and transcriptomic signatures of the endometrium from donors taking exogenous hormonal therapy?

Question 3: *Do in vitro model systems of the endometrium capture the cellular heterogeneity seen in in vivo endometrium? (Chapter 5)*

3.1: What is the cellular heterogeneity of endometrial epithelial organoids and endometrial epithelial organoids co-cultured with stromal cells?

3.2: How does the presence/absence of endometriosis change the transcriptomic profiles of endometrial cells *in vitro?*

3.3: Does taking exogenous hormones at sample collection influence the transcriptomic profiles of these cells *in vitro?*

3.4: Which *in vivo* endometrial cell states are recapitulated in the presence of oestrogen and progesterone in the culture media in both the endometrial epithelial organoids and organoid-stromal cell co-cultures?

3.5: How accurately are *in vivo* endometrial cell states recapitulated in the two different *in vitro* systems?

Chapter 2: Datasets and samples analysed

2.1 General overview of datasets analysed

In this thesis, apart from generating and analysing my own data, I have also examined and re-analysed datasets from three recently published studies looking at the endometrium[97–99]. Here, I outline the characteristics and specifics of each of these studies and their set of samples. Briefly, all studies (including mine) profiled the endometrium of reproductive-age women and focused on collecting superficial endometrial biopsies. In the case of the Garcia-Alonso *et al*.[98] study, two full thickness endometrial samples were also collected (see section 2.3.1). Where possible, the menstrual phase was assigned by a pathologist, and considered for data analysis and interpretation purposes. Moreover, use of hormonal treatment (including which type) was also reported. Two of the studies, Wang *et al*.[97] and Garcia-Alonso *et al*.[98], only analysed samples from individuals without reported endometrial/uterine pathologies, while Tan *et al*.[99] also profiled the endometrium of women with endometriosis. In the case of the Tan *et al*. study, the endometriosis stage was assigned using the rASRM staging system. In my study, from now on referred to as the Mareckova *et al.* study, I have generated data for the most comprehensive set of samples, including the endometrium of individuals without endometriosis (i.e. controls), those with endometriosis and those with other endometrial pathologies, as well as of those taking hormonal therapy. Figure 2.1 gives an overview of the number of samples analysed, including which datasets are re-analysed and presented in which chapters of this thesis.

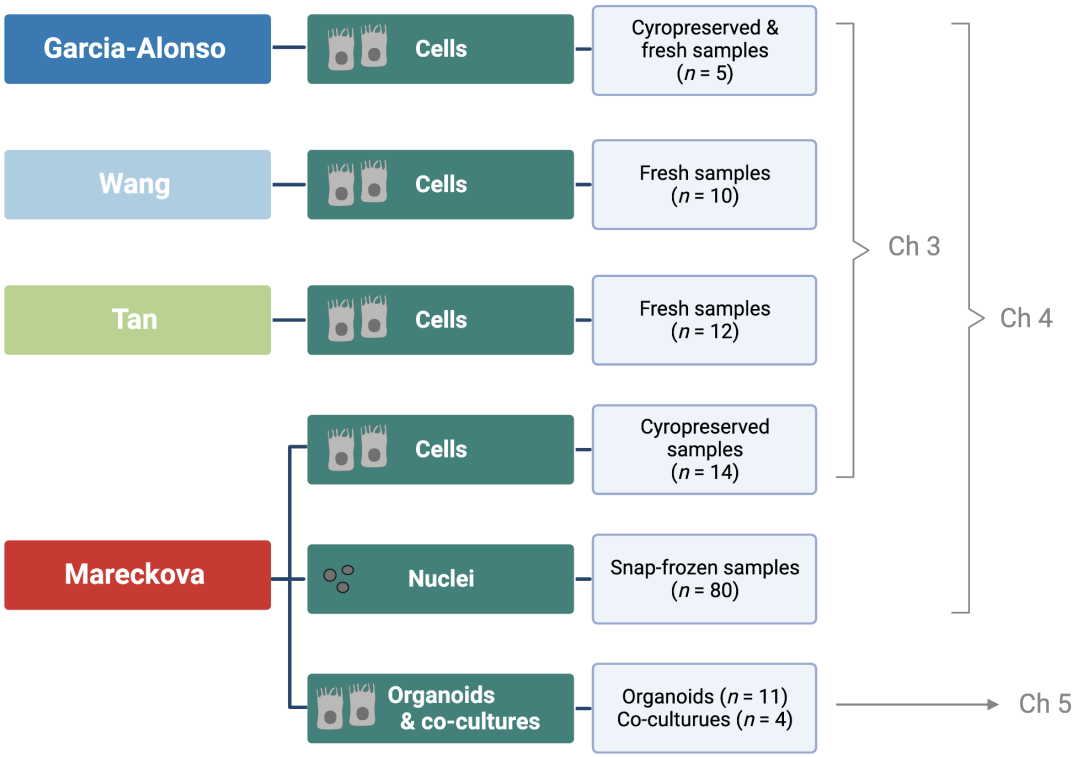


Figure 2.1: Overview of datasets' use in thesis chapters. Schematic representation of the different entities profiled and presented in this thesis (endometrial cells, nuclei, organoids or co-cultures) for each dataset analysed and the number of unique samples processed. Chapters (Ch) in which the data are presented are indicated.

2.2 Newly generated data: the Mareckova dataset

2.2.1 Ethical approvals for sample collection

Tissue samples used for my DPhil project - the Mareckova *et al.* study, were collected under three studies: (i) Endometriosis Oxford (ENDOX), (ii) Fibroids and Endometriosis Oxford (FENOX) and (iii) Sanger Human Cell Atlasing Project. Both ENDOX (REC: 09/H0604/58) and FENOX (REC: 17/SC/0664) obtained ethical approvals from the Central University Research Ethics Committee, University of Oxford. Yorkshire & The Humber - Leeds East Research Ethics Committee approved the Sanger Human Cell Atlasing Project (REC: 19/YH/0441). One sample was also collected under the Immunology of Subfertility study (REC: 08/H0606/94)

approved by the Oxford Research Ethics Committee C. In all instances, written informed consent was provided by study participants prior to obtaining tissue samples and phenotypic data.

2.2.2 Donor inclusion/exclusion criteria and phenotypic data

Only individuals aged 18-50 years (*n* = 89) were recruited for my study (Figure 2.2). Patients taking part in the ENDOX and FENOX studies were undergoing laparoscopic surgery for suspected endometriosis or infertility reasons at the John Radcliffe Hospital, Oxford. At the beginning of surgery, a pipelle biopsy of the endometrium was taken and the presence/absence of endometriosis, including endometriosis stage (rASRM stages I-IV) assigned upon surgical evaluation during the laparoscopy (*n* = 86 donors). Three additional control samples (i.e. samples from donors without endometriosis) came from the Sanger Cell Atlasing Project study (*n* = 2) and Immunology of Subfertility study (*n* = 1). Absence of endometriosis was determined based on the clinical and medical history of the patients. For the Sanger Cell Atlasing Project, patients attended a coil clinic at the Chelsea and Westminster Hospital, London. During the coil insertion procedure, a biopsy of the endometrium was taken in an outpatient setting. For the Immunology and Subfertility study, patients were undergoing *in vitro* fertilisation at TFP Oxford Fertility, Oxford and a luteal phase endometrial biopsy was taken in an outpatient setting one cycle before the patient became pregnant and had a live birth.

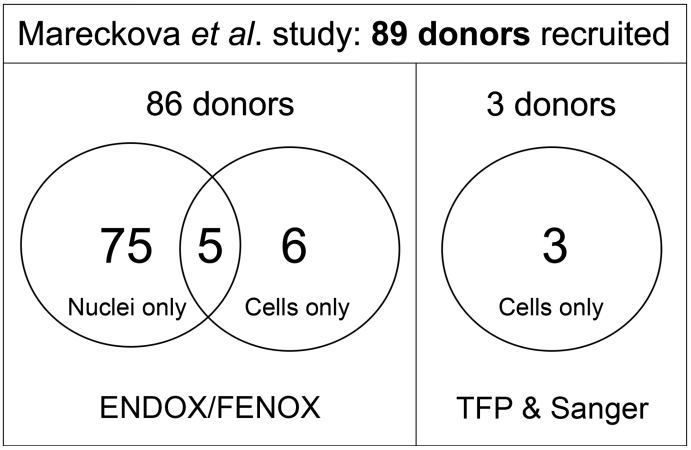


Figure 2.2: Overview of donors recruited for the Mareckova *et al*. study. In total, 89 donors were recruited, with 86 donors recruited as part of the ENDOX/FENOX studies and 3 at TFP Oxford Fertility and as part of the Sanger Cell Atlasing Project study. Shown is the number of donor samples used for extracting and profiling cells or nuclei or both.

While 89 donors were recruited and their samples processed, in this thesis I only present data for 75 donors for reasons described in Chapter 4. Out of the 75 donors, 58 donors were recruited during their natural menstrual cycles (i.e. not taking hormonal therapy at least 3 months prior to sample acquisition), but we also collected the endometrium of patients taking hormones (*n* = 17). A summary of donors' phenotypic data is shown in Table 2.1, highlighting that our donors were of similar age, irrespective of being controls (*n* = 22) or endometriosis cases *(n* = 53). Moreover, we collected endometrial biopsies from across the menstrual cycle and from all endometriosis stages (I-IV). However, in our patient set there is a bias towards having more samples from patients with endometriosis stages I/II (*n* = 34) vs endometriosis stages III/IV (*n* = 19).

Furthermore, we also collected data about donors having other benign uterine pathologies at recruitment (i.e. fibroids, polyps, adenomyosis, hyperplasia) and aimed to exclude patients with these pathologies. However, in some cases, later histological evaluations revealed the presence of these pathologies after the

sequencing data had already been generated. Table 2.1 summarises which of these pathologies are present in the Mareckova *et al*. study. Information about other uterine/endometrial pathologies was not provided by the rest of the studies. Donors with endometrial cancer were not recruited.

Table 2.1: Summary of donor phenotypic data for all datasets analysed.

	Study					
	Mareckova		Garcia-Alonso	Tan		Wang
	Control	Endoms	Control	Control	Endoms	Control
Donor #	22	53	5	3	9	10
Age^	32.5 ± 0.8	32.2 ± 0.8	27.8 ± 2.0	33.3 ± 5.9	35.8 ± 1.9	18 - 34
Menstrual cycle group						
Menstrual	1	4	-	-	1	-
Proliferative	6	15	2	1	-	2
Secretory	9	21	3	-	-	8
Hormones	5	12	-	2	8	-
Unknown	1	1	-	-	-	-
Endometriosis stage (rASRM)						
I/II	-	34	-	-	1	-
III/IV	-	19	-	-	8	-
Other endometrial/uterine pathology						
Adeno	-	3	-	-	-	-
Adeno + Polyps	-	1	-	-	-	-
Adeno + Fibroids	-	2	-	-	-	-
Fibroids	-	6	-	-	-	-
Fibroids + Polyps	1	-	-	-	-	-
Hyperplasia	-	1	-	-	-	-
Polyps	2	-	-	-	-	-

Abbreviations: Endoms, endometriosis; Adeno, adenomyosis; rASRM, revised American Society for Reproductive Medicine staging system.
Superscripts: ^ age is reported as the mean ± standard error of the mean.

2.2.3 Sample collection and processing

Superficial biopsies of the endometrium were collected using the Pipelle® sampling device and immediately transferred into ice-cold phosphate buffered saline (PBS) solution (Gibco, 10010023). The endometrial tissue was then cut into smaller pieces and either moved into a cryovial and snap-frozen on dry ice (for single-nuclei extraction and processing) or moved into ice-cold HypoThermosol®FRS solution (Sigma-Aldrich, H4416) and stored at 4°C until further processing (either to be digested fresh or cryopreserved and digested later for single-cell processing). Where possible and sample size allowed, a small piece of tissue was also embedded in optimal cutting temperature (OCT) compound (ThermoFisher Scientific, 23730571) inside a cryomold and rapidly frozen in dry ice/isopentane slurry for histological evaluation.

2.2.4 Menstrual phase assignment

For those samples where snap-frozen tissue was available, OCT blocks were sectioned at 10 μM thickness and haematoxylin and eosin-stained following standard protocols. Menstrual phase was assigned based on histological evaluation[151] of the stained slides by Dr Kezia Gaitskell and Dr Slaveya Yancheva. Where this was not possible, the menstrual phase was assigned based on the transcriptomic data and cellular profiles of the samples.

2.3 Published data: the Garcia-Alonso, Tan & Wang datasets

In the Mareckova *et al*. study, I generated both scRNAseq and snucRNAseq data, while the previously published studies by Tan *et al*. and Wang *et al*. had only generated scRNAseq data from endometrial biopsies. The Garcia-Alonso *et al*. study generated both scRNAseq (n = 5) and snucRNAseq (n = 4) data but the

snucRNAseq data was from the proliferative phase only and used for validation purposes of the scRNAseq data. For the analyses in this thesis, I only considered scRNAseq data generated for endometrial biopsies that were processed using the 10x Genomics Chromium platform and kits to reduce further technical discrepancies. All of the studies used fresh tissue samples, while in my study I also profiled cryopreserved samples, as I have shown that cryopreservation does not influence the cells' transcriptomic profiles (Appendix 1). The main characteristics and sample sets for each of the studies are highlighted in the below sections and Table 2.1 summarises the main differences in patient phenotypic data. Due to scarcity of information on ethnicity, i.e. only reported by Tan *et al*. and not available for most of the Mareckova *et al.* study donors, this information is not included in the table. Moreover, information on BMI, fertility status and medications taken was not available for most individuals, but would be of importance to consider in future endometrial studies. With regards to age, the mean age of the donors is comparable between all of the studies (Table 2.1).

2.3.1 The Garcia-Alonso dataset

In this study, data was generated for both superficial endometrial biopsies as well as endometrium and myometrium from deceased organ donors across the menstrual cycle. The biopsies from the organ donor programme were processed in a way that allowed the endometrium and myometrium to be separated and cells from each loaded as a separate 10x reaction. Considering the rest of the datasets used superficial biopsies only, in this thesis I only re-analysed the endometrial fraction of the biopsies from the organ donor programme. The collection, storing and subsequent enzymatic digestion protocol (2-step digestion with collagenase V and trypsin at 37°C) are almost identical to what I did for my study's scRNAseq

samples (described in detail in sections 3.3.1 and 3.3.2). In the Garcia-Alonso dataset, two samples are available from the proliferative phase, and three from the secretory phase.

2.3.2 The Tan dataset

The focus of the Tan *et al.* study was to describe single cell expression differences in endometrium from healthy controls (*n* = 3) vs endometriosis patients (*n* = 9), as well as profile endometriotic lesions. Most of the donors were taking hormonal therapy at the time of sample collection (*n* = 10), apart from one control and one endometriosis patient (Table 2.1). The sample digestion protocol differed significantly when compared to the rest of the studies: the collected samples were digested at 6°C using a protease and DNAse I solution and a Miltenyi GentleMACS Dissociator. Cells obtained were also live/dead sorted using FACS to enrich viable cells. In this thesis, only the data for endometrial biopsies are re-analysed, excluding the endometriotic lesions. While Tan *et al.* analysed endometrium of controls and endometriosis cases, the numbers were limited and there was a higher percentage of samples from endometriosis stages III/IV (89%) (Table 2.1).

2.3.3 The Wang dataset

The Wang *et al.* study aimed to describe the cellular heterogeneity of the endometrium across the cycle and in their initial dataset profiled 19 biopsies across the menstrual cycle using the C1 Fluidigm platform. This approach yielded only ~2,000 cells in total and the authors therefore generated a further validation dataset of ~70,000 cells from 10 biopsies using the 10x Genomics Chromium platform and kits. It is only the 10x validation dataset that is re-analysed in this thesis. Most of the samples collected were from donors in the secretory phase of the menstrual cycle

(*n* = 8), and 2 samples are from the proliferative phase (Table 2.1). Unlike the other studies, Wang *et al*. do not provide the exact age on a per donor basis, but rather as a range of ages (18-34 years) for the whole patient cohort. With regards to sample processing, the Wang *et al*. study also used their study specific protocol. Briefly, fresh biopsies were digested in a two-step fashion: (i) overnight at 4°C in a collagenase A1 solution followed by (ii) digestion of filtered, undigested tissue pieces in TypLE at 37°C. Both fractions were then enriched for live cells using MACS® and mixed at a 1:1 ratio before loading the 10x chip for scRNAseq.

2.4 Conclusions

In summary, all of the previously published datasets described above looked at a limited number of samples and cells when analysing the human endometrium across the menstrual cycle and in case of the Tan *et al*. study, in health vs endometriosis. Analysing a larger number of samples is particularly important as the endometrium is highly variable and dynamic, both inter- and intra- individually. The complexity is further increased when talking about diseases, such as endometriosis, which has a very heterogeneous presentation. Taking and not taking various types of exogenous hormonal therapy also greatly affects the endometrium and needs to be studied in more detail. In order to address these gaps, for my study I have profiled and generated data for a large set of individuals without endometriosis, with endometriosis (all stages), have natural cycles and take hormonal therapy. Having such a diverse dataset allowed me to build a more comprehensive map of the endometrium in health, disease and under hormonal influence and to utilise the previously published datasets by integrating them with my data. The next chapters deal with the different analyses performed on all of these datasets and how they

have helped me to better define the cellular heterogeneity of the human endometrium.

Throughout the next chapters I refer to the cellular heterogeneity of the human endometrium even though most of the samples analysed and presented in this thesis are superficial biopsies collected from living donors (i.e. predominantly sampling the *functionalis* layer) and only two samples from the Garcia-Alonso *et al.* dataset include full-thickness endometrium (i.e. both the *functionalis* and *basalis* layers). I therefore wish to acknowledge that technically-speaking, it could be more accurate to refer to the cellular heterogeneity of the *functionalis* of the human endometrium. However, it would not fully encompass all of the samples presented in this thesis and the fact that even when superficial endometrial biopsies are collected, depending on the thickness of the *functionalis*, the clinic/clinician collecting these samples and sampling device used, more than just the *functionalis* might be sampled and analysed. This is supported by the observation of cells found within the *basalis* layer also detected in some of the superficial biopsies presented in Chapters 3 and 4.

Chapter 3: Single-cell map of the human endometrium

3.1 Introduction

As described in Chapter 1 and Chapter 2, the endometrium is highly heterogeneous, both inter- and intra-individually. This is true both across the menstrual cycle and one's lifetime. To capture the heterogeneity and cell states present, large cohorts of patients need to be characterised. Previous studies have started profiling the endometrium at the single-cell level but so far the number of cells and samples analysed have been limited (Table 3.1). Integrating datasets from multiple sources offers the possibility to build more comprehensive and cohesive maps of the endometrium, achieving higher statistical power for downstream analyses. Undeniably though, datasets from multiple sources have their inherent characteristics and technical variations that need to be considered. This chapter deals with the integration of three previously generated datasets with data newly generated as part of my project.

3.1.1 Data integration

Multiple factors such as library size, experimental batches/runs, individual donor samples and different tissue processing protocols can introduce strong batch-effects that need to be accounted for when analysing scRNAseq datasets. Otherwise, these technical factors can mask the real biological signals in the data. Computationally, batch-effects can be corrected for, meaning any technical nuisances can be removed while maintaining the biological variation of interest. Such batch-effects correction is especially important when analysing and merging datasets from multiple labs/sources as is the case in this thesis. A plethora of tools exists for fulfilling this task, reviewed in[152] but here we utilised the single-cell

Variational Inference (*scVI*) method[153] for data integration. *scVI* is a fully probabilistic method that uses neural networks to model the expression of each gene in each cell while accounting for batch effects. *scVI* is a machine learning method which is particularly suitable for integrating large datasets with multiple sources of batch effects, as is the case with the datasets to be combined here (please, see below and Chapters 2 and 4 for more information).

In this chapter, *scVI* was used to capture the commonalities and differences between 4 datasets: (i) the newly generated Mareckova *et al.* dataset containing mostly proliferative phase samples, samples from controls and endometriosis cases as well as donors on hormones, (ii) the Garcia-Alonso *et al.* dataset with control samples from multiple tissue sources - superficial and whole uterus biopsies, (iii) the Wang *et al.* dataset with majority of samples from the secretory phase, and (iv) the Tan *et al*. dataset with most samples from donors taking exogenous hormones. All of these variations were important to consider and correct for, including the use of different protocols in each study for tissue dissociation as described in Chapter 2, while avoiding data overcorrection.

3.1.2 Overview of epithelial and mesenchymal cell states identified previously

As summarised in Table 3.1, the identification and characterisation of endometrial cells has been done to various levels of granularity. The first cellular map published by Wang *et al*. in 2020 identified only 3 groups of epithelial cells (luminal, glandular, ciliated) and 2 groups of mesenchymal cells (stromal fibroblasts and uterine smooth muscle cells). As described in Chapter 2, the main analyses for cell type identification were performed on a small subset of cells (~2,000 cells profiled using the C1 Fluidigm platform) and the larger dataset (~70,000 cells profiled using the

10x Chromium platform) was mainly used to validate the initial findings and was not explored in detail. Below only the 10x data is reported and considered.

Table 3.1: Overview of endometrial cells analysed & identified in multiple datasets.

	Publication		
	Wang et al.	**Garcia-Alonso et al.**	**Tan et al.**
Cells^	~70,000 (10x)	~47,000 (10x)	~44,000 (10x)
Donors	10	5	12
Epithelial cell states*	*~37,000 cells* Ciliated Luminal Glandular	*~3,000 cells* Ciliated Ciliated_LGR5 Pre-ciliated SOX9+ proliferative SOX9+LGR5- SOX9+LGR5+ Lumenal 1 Lumenal 2 Glandular Glandular-secretory	*~14,000 cells* Ciliated MUC5B+ Lumenal Lumenal 1 Lumenal 2 Glandular-proliferative Glandular early-secret Mid-secretory TP63+KRT5+
Mesenchymal cell states*	*~17,000 cells* Stromal fibroblasts uSMCs	*~29,000 cells* eS dS Fibroblast_C7 uSMCs PV_STEAP4 PV_MYH11	*~13,000 cells* eF1 eF2 eF3 dS1 dS2 PV_STEAP4 PV_MYH11

Abbreviations: 10x, 10x Genomic Chromium Platform & kits; dS, decidualised stromal cells; eF, endometrial fibroblasts; eS, endometrial stromal cells; PV, perivascular cells; uSMC, uterine smooth muscle cells.

Superscripts:

^ the total number of endometrial cells analysed as reported in the original manuscripts and/or checking the numbers when downloading their data.

* the number of cells assigned as high QC cells belonging to either the epithelial or mesenchymal cell lineage when integrated and analysed by me.

The second map of the endometrium published by Garcia-Alonso *et al*. in 2021 identified 10 different epithelial and 6 mesenchymal cell states. Note, that the numbers analysed in Table 3.1 seem small, because only newly generated data for each study is presented. In the Garcia-Alonso *et al*. study, the authors also integrated their data with the Wang *et al*. dataset and generated snucRNAseq data for 4 further proliferative phase samples to confirm the identity of the cells described. Subtypes of ciliated, glandular, lumenal and *SOX9*$^{+}$ cells were defined based on their marker gene expression, location within the tissue and presence in different phases of the menstrual cycle[6,98,154]. Of particular relevance is the identification of two epithelial progenitors: *SOX9*$^{+}$*LGR5*$^{+}$ cells located in the lumen, and *SOX9*$^{+}$*LGR5*$^{-}$ cells located in the *basalis* layer of the endometrium, close to the myometrium. Similarly, 2 subtypes of endometrial stromal cells, a new type of fibroblast cells located in the *basalis* and two groups of perivascular cells were identified (Table 3.1). In the Tan *et al.* cellular map of the endometrium in endometriosis cases and controls, with donors predominantly taking hormonal therapy, subtypes of lumenal and glandular cells were defined, as well as ciliated and the newly discovered MUC5B^{+} cells. Moreover, in the mesenchymal lineage, 3 types of endometrial fibroblasts, two types of decidualised stroma and the two types of perivascular cells previously defined by Garcia-Alonso *et al*. were described.

Overall, it seems that when it comes to cell annotations, there is a consensus in being able to detect lumenal/luminal (in this thesis I use the word luminal to describe my data), glandular and ciliated cells across all of the datasets. It is less clear how the finer annotations and subgroups of cells compare across the datasets. In the mesenchymal lineage, there are discrepancies in nomenclature when likely the same cell state is called either endometrial stromal cells (Garcia-Alonso *et al*.) or

fibroblasts (Wang *et al*. and Tan *et al*.). Moreover, multiple subtypes of stromal/fibroblasts and decidualised stromal cells were only reported in the Tan *et al*. dataset. Taken together, it is clear that a harmonised map of these datasets using the same cell state markers and nomenclature is currently lacking and needed to understand the transcriptomic profiles of the endometrium in health, endometriosis and under the influence of exogenous hormones. This chapter describes the generation of such a map, combining the 3 datasets discussed with further data generated as part of my DPhil project.

3.2 Aims

1. To integrate and assess the integration of previously published datasets of the human endometrium with the data generated as part of my project.
2. To annotate the cell clusters identified and define a reference list of markers of any newly discovered endometrial cell populations in the epithelial and mesenchymal cell lineages.
3. Describe any trends observed in the data and generate hypotheses to test when it comes to the cellular composition of the endometrium in health, endometriosis and under exogenous hormones.

3.3 Materials & Methods

3.3.1 Sample processing

In a Petri dish on ice, tissue originally collected in HypoThermosol®FRS solution was covered in ice-cold RPMI 1640 Medium (RPMI) (ThermoFisher Scientific, 11875093) containing 10% (v/v) heat-inactivated foetal-bovine-serum (FBS) (Sigma-Aldrich, F7524) (RPMI/FBS) and minced mechanically with scalpels (size 24). The sample was centrifuged (500 x g, 3 min), supernatant discarded and tissue re-suspended in ice-cold CryoStor®cell cryopreservation medium (Sigma-Aldrich, C2874). Samples were frozen in cryovials as per the manufacturer's instructions using Corning® CoolCell® and stored at -80°C.

3.3.2 Sample digestion

Frozen samples were thawed at 37°C, quickly transferred to a 15 ml tube and topped-up with 13 ml of ice cold RPMI/FBS. Samples were centrifuged (500 x g, 5 min, 4°C) and the supernatant discarded. The tissue was enzymatically digested on a MACSMix rotator (set to 16 rpm speed) at 37°C in pre-warmed RPMI/FBS containing Collagenase V (Sigma-Aldrich, C9263), and DNAse I (Roche, 11284932001) with final concentrations of 1 mg/ml and 0.1 mg/ml, respectively. Digested tissue was centrifuged (500 x g, 5 min), resuspended in 10 ml of PBS and passed through a 40 µm cell strainer (BD Biosciences, 352340), generating the collagenase fraction, enriched in stromal and immune cells (Figure 3.1). The filter was back-washed with PBS into a 50 ml tube and centrifuged (500 x g, 5 min). Supernatant was discarded and any undigested tissue within the pellet was incubated with 0.25% (v/v) trypsin-EDTA (Sigma-Aldrich, T3924) and DNAse I (0.1 mg/ml) at 37°C for 15 min on a MACSMix rotator. The digestion process was

stopped by adding RPMI/FBS and samples centrifuged (500 x g, 5 min). This step yielded the trypsin fraction. The collagenase fraction was centrifuged (500 x g, 5 min) and resuspended in 2 ml of red-blood-cell (RBC) lysis buffer (eBioscience, 00-4300) for 5-10 min at room temperature. After incubation, the samples were centrifuged (500 x g, 5 min), the RBC buffer discarded and both fractions (collagenase and trypsin) resuspended in 0.04% bovine serum albumin (BSA) (Sigma-Aldrich, A9418) in PBS (v/v). The generated single-cell suspensions were stored on ice until further processing.

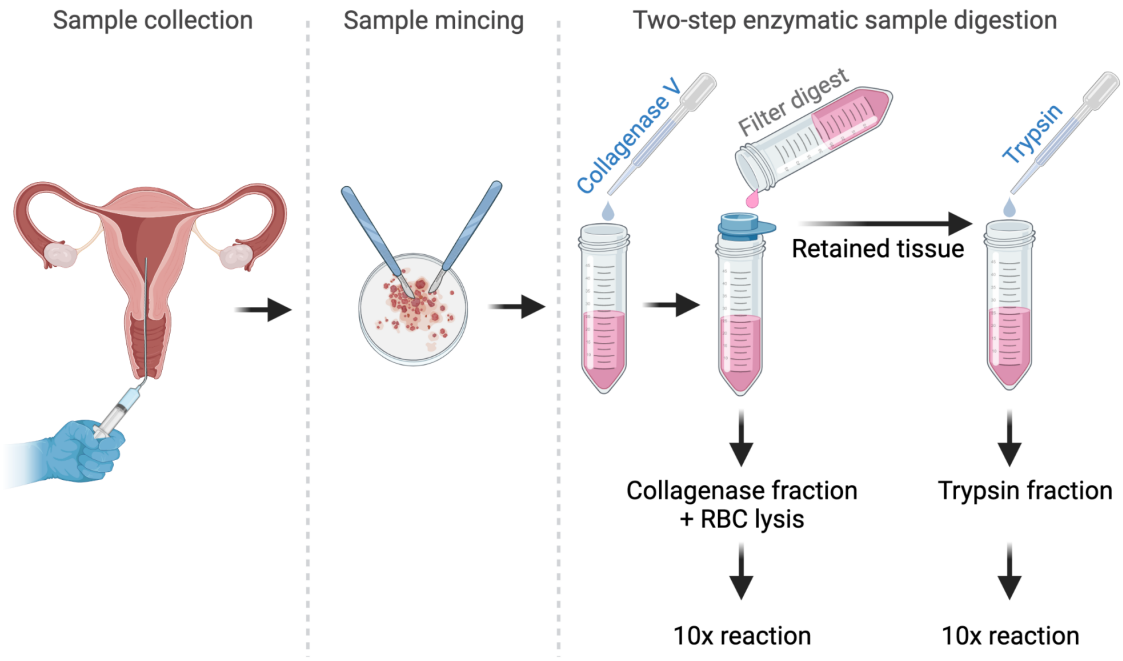


Figure 3.1: Schematic representation of sample processing and digestion for scRNAseq. Upon sample collection, the sample was minced using scalpels and frozen. The thawed cut-up sample was digested in two steps: 1) with collagenase V, yielding the collagenase fraction upon filtering that was further treated with RBC lysis buffer and then loaded as a separate 10x reaction, 2) undigested tissue retained on the filter was digested with trypsin, yielding the trypsin fraction that is loaded as a separate 10x reaction. *Abbreviations:* RBC, red blood cell.

3.3.3 Single-cell RNA-sequencing

3.3.3.1 Cell counting & live/dead sorting

Cells were counted using a haemocytometer and trypan blue to determine their viability prior to loading the Chromium Next GEM Chip G (10x Genomics, 1000120). In the case of two samples (donor IDs: FX1125 and FX1176), cells from the collagenase fraction were live/dead sorted prior to loading to enrich for live cells. Sorting was carried out by Dr Helen Ferry at the Experimental Medicine Division Flow Cytometry Facility, University of Oxford. The nuclear stain DAPI (4′,6-diamidino-2-phenylindole) was used to visualise and distinguish live/dead cells and debris.

3.3.3.2 Cell loading and target recovery

Given the endometrium was digested in two steps, the collagenase and trypsin fractions were loaded into two separate 10x reactions (see Figure 3.1). In the majority of cases, two donor samples per 10x reaction were pooled, only mixing the collagenase fractions with collagenase fractions and trypsin with trypsin. Cells were loaded at volumes and concentrations required to achieve an equal target cell recovery per donor. For the collagenase fraction this was at 8,000 - 10,000 viable cells per donor, while for the trypsin fraction it was at 4,000 - 6,000 cells. After a couple of initial trials, I found these cell numbers to be most suitable for loading without clogging the 10x chip.

3.3.3.3 Multiplexing donors - TotalSeq™ antibodies

In order to demultiplex the pooled samples, in most cases we used the different donors' genotypic data to assign each cell to a specific donor (see 3.3.3 Data analysis section for more details). When genotypic data was not available, prior to

pooling, cells were labelled using the TotalSeq™-B antibodies from Biolegend (see Table 3.2). The oligo-conjugated antibodies bind ubiquitously expressed cell surface markers and so can label each cell with a unique oligo 'barcode' that will be sample/donor specific. The first paper publishing the method describes the technology in more detail[155]. Following the manufacturer's protocol, 1-2 million cells were resuspended in 50 µl Cell Staining Buffer (Biolegend, 420201) and incubated for 10 minutes at 4°C with 5 µl of Human TruStain FcX™ blocking reagent (Biolegend, 422302). The antibodies were centrifuged (14,000 x g, 10 min, 4°C) before adding 1 µg of each antibody to the cell suspensions. After incubation (30 min, 4°C), 3.5 ml of the Cell Staining Buffer was added and samples centrifuged (400 x g, 5 min, 4°C) to wash the cells. In total, cells were washed 3 times before resuspending them in 0.04% (v/v) BSA/PBS and filtering through a 40 µm Flowmi® cell strainer (Scienceware, BAH136800040). Cells were counted to determine their concentration and viability prior to loading the 10x chip.

Table 3.2: TotalSeq™ antibodies used to label cells prior to pooling.

Antibody	Lot number	Product number
Hashtag 1	B309306	394631
Hashtag 2	B294183	394633
Hashtag 3	B294188	394635
Hashtag 4	B301069	394637
Hashtag 5	B296944	394639

3.3.3.4 10x Genomics libraries and sequencing

The 10x Genomics v3 chemistry libraries (10x Genomics, CG000183) were prepared as per the manufacturer's protocols by either Sanger Cellular Genetics or Sanger pipelines. Libraries were sequenced to a minimum coverage of 20,000 raw reads per cell on Illumina HiSeq 4000 and NovaSeq 6000 platforms, using the sequencing formats: read 1: 28 cycles; i7 index: 8 cycles, i5 index: 0 cycles; read 2: 91 cycles.

3.3.3 Data analysis

The upstream bioinformatics analyses in this thesis (genotyping assignment, read alignment, data integration, clustering and defining high QC cells and cell states) were performed by Dr Luz Garcia-Alonso, Principal Bioinformatician in the Vento-Tormo lab at the Wellcome Sanger Institute (WSI), with contributions from Dr Louis-François Handfield, Postdoctoral Researcher at the WSI. I participated in the analyses through regular meetings and discussions with Dr Garcia-Alonso and Dr Handfield, designing the analysis, interpreting the results and discussion of the next steps. All figures shown were generated by me, using the finalised and processed data objects.

3.3.3.1 Obtaining and re-analysing previously published datasets

FASTQ files for the previously published Wang *et al*. dataset (Gene Expression Omnibus accession number GSE111976) and Garcia-Alonso *et al*. dataset (ArrayExpress accession number E-MTAB-10287) were downloaded and re-aligned using the same genome reference and *CellRanger* pipeline as described in section 3.3.3.2. For the Tan *et al*. dataset (Gene Expression Omnibus accession number GSE179640), only the cell-by-gene count matrices generated by the authors using

CellRanger v3.1.0 and reference genome GRCh38.p13 (GRCh38 10x Genomics reference 3.0.0) were downloaded and used for analyses. From the Garcia-Alonso *et al.* dataset, only scRNAseq data for the endometrium-enriched fraction of the two samples from the organ donor programme were included in the analyses. For the Tan *et al*. dataset, only scRNAseq data generated for eutopic endometrium was re-analysed. All downstream analyses were performed as outlined below.

3.3.3.2 Data alignment, quantification and donor deconvolution

The raw sequencing data was aligned and quantified using the *CellRanger* v5.0.2 pipeline[156] to obtain gene expression matrices of raw counts for downstream analyses. A reference genome to which reads were aligned was based on the 10x Genomics' GRCh38 1.2.0 release, constructed from Ensembl 84, with gene_biotypes filtered to protein_coding, lincRNA, antisense and the various IG and TR genes and pseudogenes. Reads were required to align within exonic regions to be counted. Libraries in which cells from multiple donors were pooled together were demultiplexed using *Souporcell[157],* which assigns single cells into a number of groups that correspond to the number of expected genotypes that share coherent sets of Single Nucleotide Polymorphisms (SNPs) from transcriptomic reads. To assign each cell to a specific donor, a custom script was used to (i) genotype the barcodes using the aligned reads from the scRNAseq data and (ii) link the barcode variants to the genotypic variant file produced by genotyping the different donors using microarrays. The microarray genotypic data was generated previously as part of the ENDOX and FENOX studies and that data was processed and provided by Mr. Michał Krassowski. The correspondence of the genotypes was assessed by comparing the genotypic variant files produced from single cells and microarrays using the *vcftools* suite and selecting the assignment with the highest genotype

similarity. In the case of libraries in which cells from multiple donors were first labelled with TotalSeq™-B antibodies prior to pooling them, the deconvolution was performed using *Scanpy*'s *HashSolo*[158].

3.3.3.3 Quality control filtering and doublet identification

All of the downstream analyses were performed using the standard *Scanpy*[159] pipeline and its in-built functions, with some modifications that are highlighted throughout the text. To filter out non-viable cells, cells with < 1000 detected genes and for which total mitochondrial gene expression exceeded 20% were removed. Genes that were expressed in < 3 cells were also removed. The next step was to identify and remove genes associated with cell stage progression, as they have high expression and can impede proper cell type annotations further downstream. Marker genes for G2/M and S phase of the cell cycle listed inside the *Seurat* package[160] were excluded in a data-driven manner as described previously[161]. To identify droplets containing more than one cell, two approaches were used. First, doublets were identified by *Souporcell*, which calls doublets based on the presence of a mixture of SNPs from different genotypes inside single droplets when multiple samples are pooled. These doublets were discarded. The second approach estimates a doublet score with *Scrublet*[162], which quantifies the presence of mixed transcriptomic signatures. These doublets were not excluded from the initial analyses, but were instead used to flag clusters with high doublet scores that were then removed.

3.3.3.4 Batch effect correction, data integration, dimensionality reduction

All data was analysed on a per sample basis prior to batch correction to evaluate their quality. Then, data was integrated into one joint manifold using *scVI*[153]

correcting for batch, genotype and library. The training was done on 32 latent variables and 2 layers of neural networks to achieve satisfactory data integration. These parameters were needed when integrating the scRNAseq dataset generated as part of the Mareckova *et al*. study with the other previously published datasets. The resulting latent variables were used for neighbour identification required for Leiden clustering[163] and Uniform Manifold Approximation and Projection (UMAP) visualisation[164] from the *Scanpy* package. UMAP was used for dimensionality reduction of the data followed by cell clustering with the Leiden algorithm. The identified clusters were annotated as described below. Clusters with no distinctive markers and that had overall lower numbers of genes expressed or higher percentages of mitochondrial gene expression were flagged as low QC and excluded from downstream analyses. In addition to the general analysis for the main cell lineages, the epithelial and mesenchymal cells were subsequently re-analysed and clustered as described above, and re-annotated using the strategy described in the next section.

3.3.3.5 Annotation of single-cell data

Identification, labelling and naming of the cell clusters identified was done by manual inspection of marker genes and interpretation of these based on literature. To aid the cluster annotation, the Garcia-Alonso *et al.* annotations were visualised onto the main cell lineages UMAP and the UMAPs generated for the separately re-analysed epithelial and mesenchymal cells. To identify marker genes specific to any given cluster, we used the Term Frequency – Inverse Document Frequency (TF-IDF) approach from the *SoupX* R package v.1.5.0[165]. TF-IDF had originally been used in the field of text analysis, but has been gaining popularity in scRNAseq analyses, allowing one to identify what the most specific genes for a given cluster are[166].

3.4 Results

3.4.1 Integration of the Garcia-Alonso, Mareckova, Wang and Tan datasets

In this chapter, I integrated the scRNAseq data generated for my study (*n* = 14) with the previously published scRNAseq datasets (Garcia-Alonso, Wang and Tan) to build a harmonised map of the human endometrium across the menstrual cycle. Chapter 2 of this thesis details the specificities of each of these datasets while here the focus is on assessing the integration of all the samples summarised in Table 3.3. Satisfactory integration of the data was achieved using *scVI* (see section 3.3.3.4 for details) and the integrated UMAP of ~320,000 cells is shown in Figure 3.2A. The aim was to integrate the datasets in a way that does not remove important biological features of each of the datasets by collapsing everything together. It is clear from Figure 3.2A that complete mixing of the cells from each dataset was not achieved; however, this can be explained by the inherent biological biases in each of the datasets. For example, the Mareckova *et al*. dataset is mostly made of proliferative phase samples, while the Wang *et al*. dataset is made of mostly secretory phase samples and the Tan *et al*. dataset contains biopsies from most donors taking hormonal therapy. Therefore, the integration shown here manages to capture all of these biologically important differences but also captures commonalities across the datasets.

Table 3.3: Overview of the datasets and sets of samples integrated.

	Dataset					
	Mareckova		Garcia-Alonso	Tan		Wang
	Control	Endoms*	Control	Control	Endoms*	Control
Donor #	5	9	5	3	9	10
Age^	32.5 ± 0.76	32.1 ± 1.15	27.8 ± 2.03	33.3 ± 5.9	35.8 ± 1.9	18 - 34
Menstrual cycle group						
Menstrual	-	-	-	-	1	-
Proliferative	3	5	2	1	-	2
Secretory	2	2	3	-	-	8
Hormones	-	2	-	2	8	-
Unknown	-	-	-	-	-	-
Endometriosis stage (rASRM)						
I/II	-	6	-	-	1	-
III/IV	-	3	-	-	8	-

Abbreviations: rASRM, revised American Society for Reproductive Medicine staging system.
Superscripts:
* Endoms stands for Endometriosis.
^ Age is reported as the mean ± standard error of the mean.

3.4.1.1 Main cell lineage identified

Using previously described canonical markers[97,98], 6 major cell lineages were identified: epithelial (*EPCAM*+), stromal (*PDGRFA*+), immune (*PTPRC*+), endothelial (*CDH5*+), perivascular (PV) (*RGS5*+), and uterine smooth muscle cells (uSMCs) (*ACTG2*+). Doublets and low QC cells were also identified and are visualised in Figure 3.2B. Contribution of each donor to the cell clusters identified was also assessed and highlighted a donor-specific cluster of uSMCs and some PV cells predominantly coming from donor A30 (Figure 3.2C). Given this particular sample came from the organ-donor programme where both the full-thickness

endometrium and myometrium were sampled, it is likely that the endometrium and myometrium were not separated completely and thus the myometrial uSMCs and a proportion of PV cells were detected in this donor sample only.

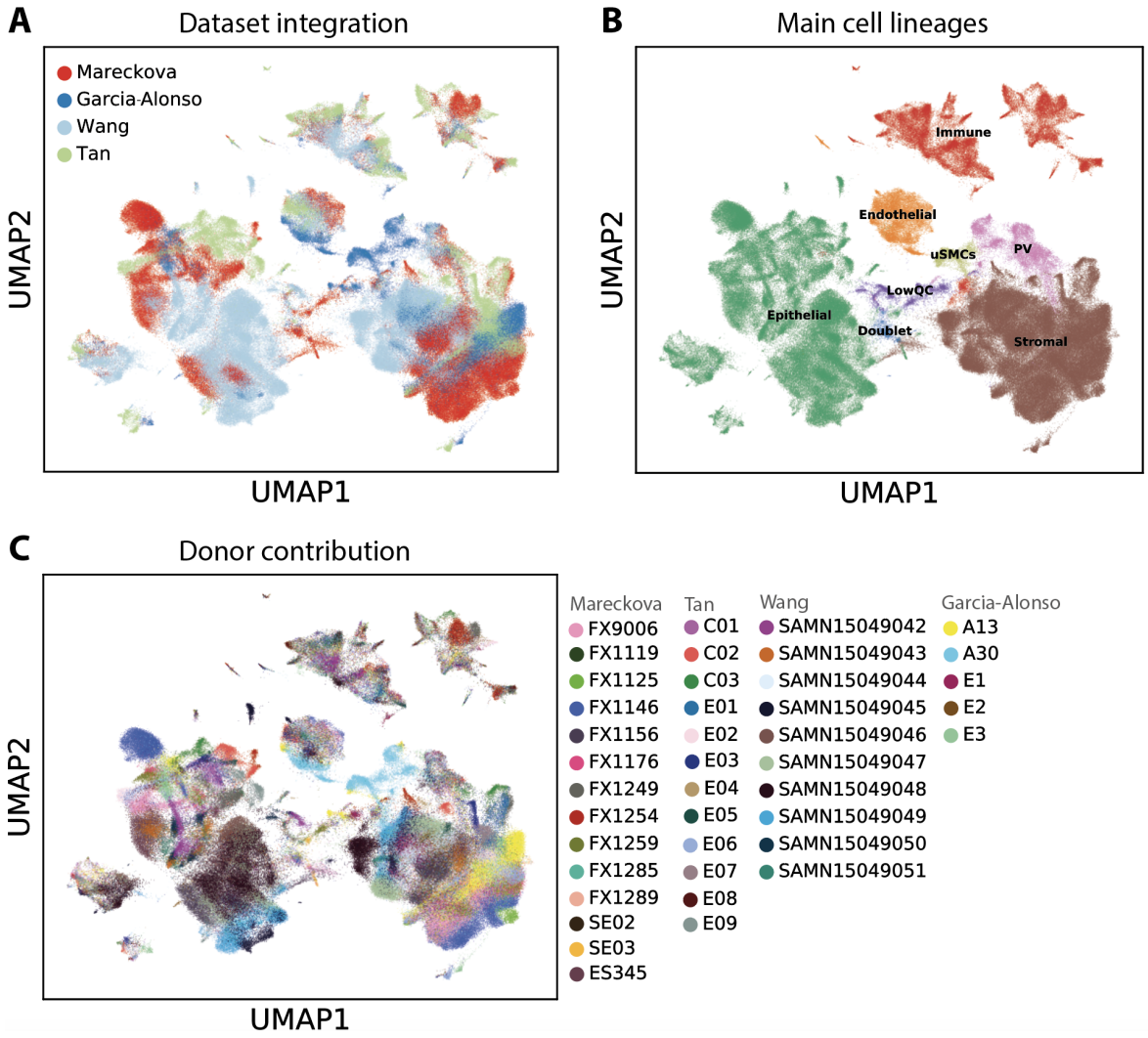


Figure 3.2: Unified map of endometrial cells from multiple datasets. UMAP projections of the four scRNAseq datasets integrated, coloured by the different datasets (A), main cell lineages identified (B) and individual donor samples (C). *Abbreviations*: PV, perivascular; QC, quality control; scRNAseq, single-cell RNA-sequencing; UMAP, uniform manifold approximation and projection; uSMCs, uterine smooth muscle cells.

Next, I looked at more fine-grained annotations of the cell clusters, identifying multiple subtypes for each of the main lineages. In this thesis, I only present the sub-clustering of the epithelial cell lineage (section 3.4.1.2) as this is where I

observed most differences between the datasets, prompting me to reassess the inclusion of the Tan dataset when building a unified cellular map of the human endometrium across the menstrual cycle.

3.4.1.2 Zooming-in on the epithelial cell lineage

Analysing the epithelial cell lineage on its own revealed multiple cell states within this lineage shown in Figure 3.3. The cells can be broadly divided into 4 main groups: those coming from the secretory phase of the cycle (secretory), cells from the proliferative phase (proliferative), cells from donors on hormones (hormones) and ciliated cells (ciliated). Furthermore, two small clusters of cells expressing cervical markers (i.e. *TP63*, *KRT5*) termed cervix and cells previously described as MUC5B cells defined by the expression of *MUC5B, TPP3 and SAA1*[99] were also identified. Doublets and low QC cells are also shown in Figure 3.3A. Using previously defined markers[98] and label transfer, I could identify further subtypes of the 4 main groups of cells and identified novel cell states, specifically the TENM2 and glandular_secretory_FGF7 states which are discussed in more detail in section 3.4.2.3. Here the focus is on describing the differences between the datasets integrated. Figure 3.3B shows the contribution of each dataset to the various cell states identified, showing that for example, the cervix and MUC5B cells predominantly came from the Tan dataset.

Next, I focused on examining the cellular composition of each donor sample and observed that the two samples from donors not on hormones from the Tan *et al*. dataset did not show the same patterns of expression observed for samples not on hormones in the Mareckova, Garcia-Alonso and Wang datasets (Figure 3.3C). Based on the metadata provided, the two Tan dataset samples (E01 and C03) were

assigned as being from the menstrual and proliferative phases, respectively. However, cells from these donors predominantly clustered with cells from donors on hormones and only to a small extent with cells characteristic of the proliferative phase. This was in disagreement with the proliferative phase samples from the Wang, Garcia-Alonso and Mareckova datasets that all had similar cellular profiles (cellular profiles of representative proliferative phase samples from each dataset are shown in Figure 3.3C). The Tan samples from donors not on hormones were transcriptomically similar to those of a donor having a hormonal intrauterine device (IUD) fitted from the Mareckova *et al*. dataset (Mareckova-Hormones-FX1249) defined as the Hormones_II cell state and also to the Hormones_I cell state characteristic of the rest of the Tan samples from donors on hormones.

Based on the metadata provided by the authors, all of the Tan dataset donors on hormones were taking similar hormonal therapy (oestrogen combined with progestin), while the samples from the Mareckova dataset were on progestin only, either taken orally (donor FX1254) or as the fitted IUD (donor FX1249). The different hormonal regiments taken likely explain the observed heterogeneity within the Hormones group of cells (cell states Hormones_I, Hormones_II and Hormones_III) and require further investigation. However, it is not clear why the Tan dataset samples from donors not on hormones also belonged to these cell states. Given these discrepancies, for the purpose of this thesis, I decided to exclude the Tan *et al*. dataset from further analyses.

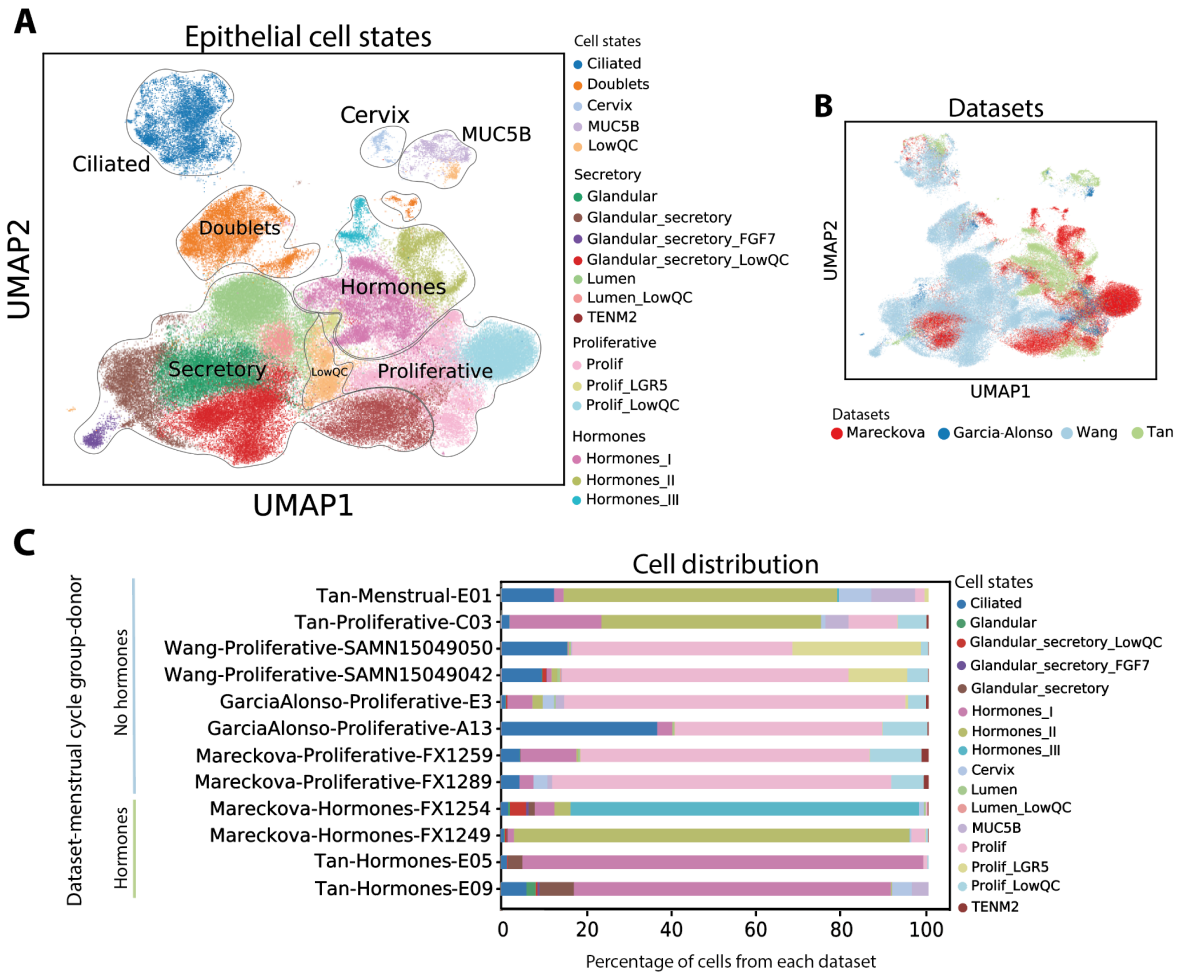


Figure 3.3: Endometrial epithelial cells. Shown are UMAP projections of the epithelial cell lineage from the four scRNAseq datasets integrated, coloured by cell states identified (A) and the different datasets (B). In (C) the distribution of the epithelial cell states is shown on a sample basis. Representative samples from each dataset of donors with/without hormonal therapy are shown. Two exemplary samples from the proliferative phase of the menstrual cycle are shown for the Wang, Garcia-Alonso and Mareckova datasets. For the Mareckova and Tan datasets, samples from donors on hormones are also shown. *Abbreviations*: Prolif, proliferative; QC, quality control; scRNAseq, single-cell RNA-sequencing; UMAP, uniform manifold approximation and projection.

3.4.2 Updated single-cell map of the human endometrium: integration of the Mareckova, Garcia-Alonso and Wang datasets

3.4.2.1 Datasets' QC metrics

This section focuses on assembling a single-cell map of the endometrium by collating the Mareckova, Garcia-Alonso and Wang datasets. ~280,000 cells were integrated prior to filtering (Mareckova: 124,238, Garcia-Alonso: 51,117, Wang:

104,962). Overall, the QC metrics for these datasets were comparable: the number of genes, mRNA molecules and percentage of mitochondrial genes per cell showed similar distribution with only a slightly higher frequency of doublets observed in the Mareckova dataset (Figure 3.4). The cell annotation of the main lineages and more fine-grained annotation of the epithelial and mesenchymal cells is described in the next sections.

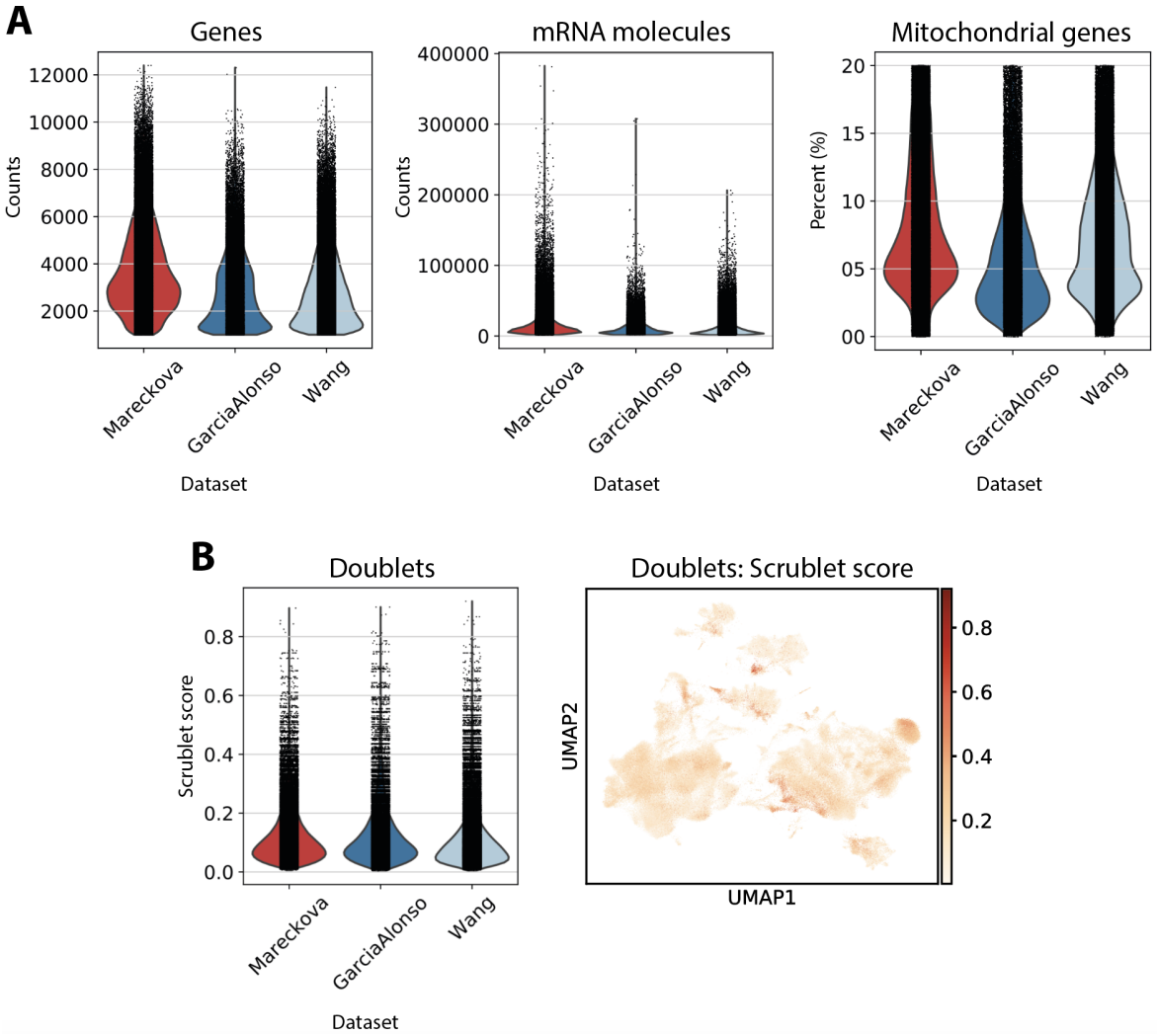


Figure 3.4: Dataset quality control metrics plots. The violin plots in (A) show the data distribution for number of genes, number of mRNA molecules and percentage of mitochondrial gene expression on a per cell basis for each of the three datasets. (B) shows the distribution of doublets for each dataset, with a UMAP projection of the *Scrublet* score used to identify and visualise doublets in the data. *Abbreviations*: UMAP, uniform manifold approximation and projection.

3.4.2.2 Main cell lineages

The main cell lineages identified included epithelial, stromal, immune, endothelial, PV and uSMCs - visualised in Figure 3.5A together with doublets and low QC cells. The integration of the 3 datasets preserved their biological differences while capturing their similarities (Figure 3.5B) such as the differences in cells' transcriptomic profiles dependent on the phase of the menstrual cycle and hormonal therapy taken (Figure 3.5C). No menstrual phase samples were analysed, only samples from the proliferative and secretory phases and two samples from donors on hormones. These included taking progestins either orally (Hormones_Desogestrel) or as the fitted IUD (Hormones_Mirena coil) and cells from these donors clustered separately (Figure 3.5C).

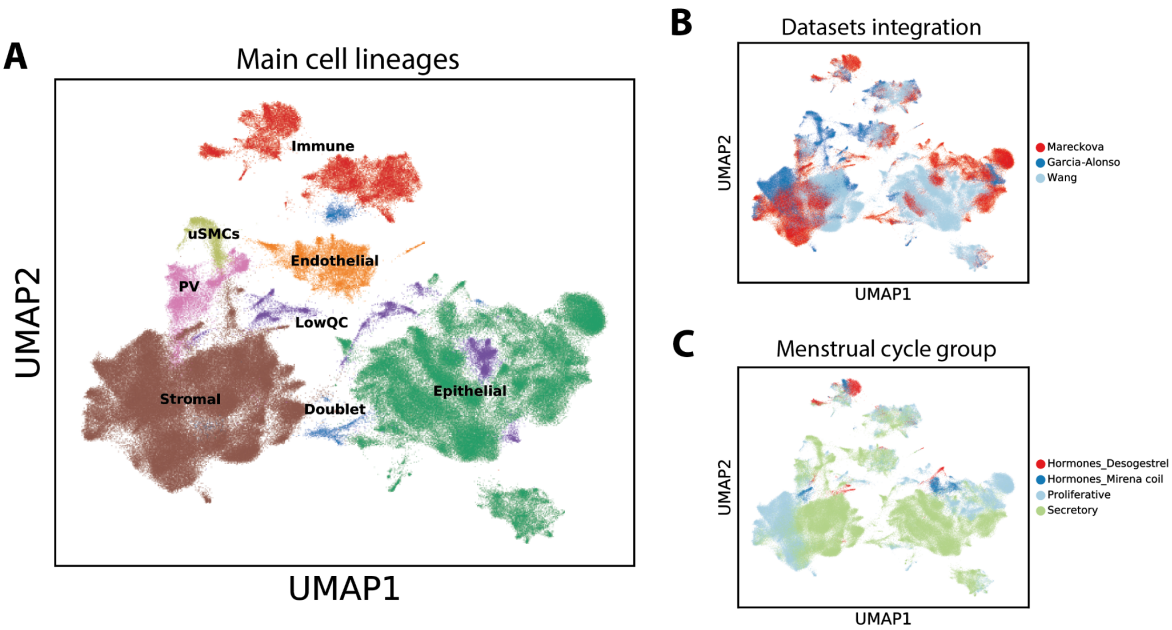


Figure 3.5: UMAP visualisation of the main cell lineages. Shown are UMAP projections of the three scRNAseq datasets integrated, coloured by main cell lineages identified (A), the different datasets (B), and menstrual cycle phase and hormonal therapy taken (C). *Abbreviations*: PV, perivascular; QC, quality control; scRNAseq, single-cell RNA-sequencing; UMAP, uniform manifold approximation and projection; uSMCs, uterine smooth muscle cells.

The proportion of cells coming from each dataset is shown in Figure 3.6A and highlights that the uSMCs came from the Garcia-Alonso dataset only where two samples from the organ donor programme were included as described in earlier sections. Otherwise, all datasets contributed cells to each of the main cell lineages identified and the 3 main/canonical marker genes for each of the lineages used to confirm their identity are presented in Figure 3.6B.

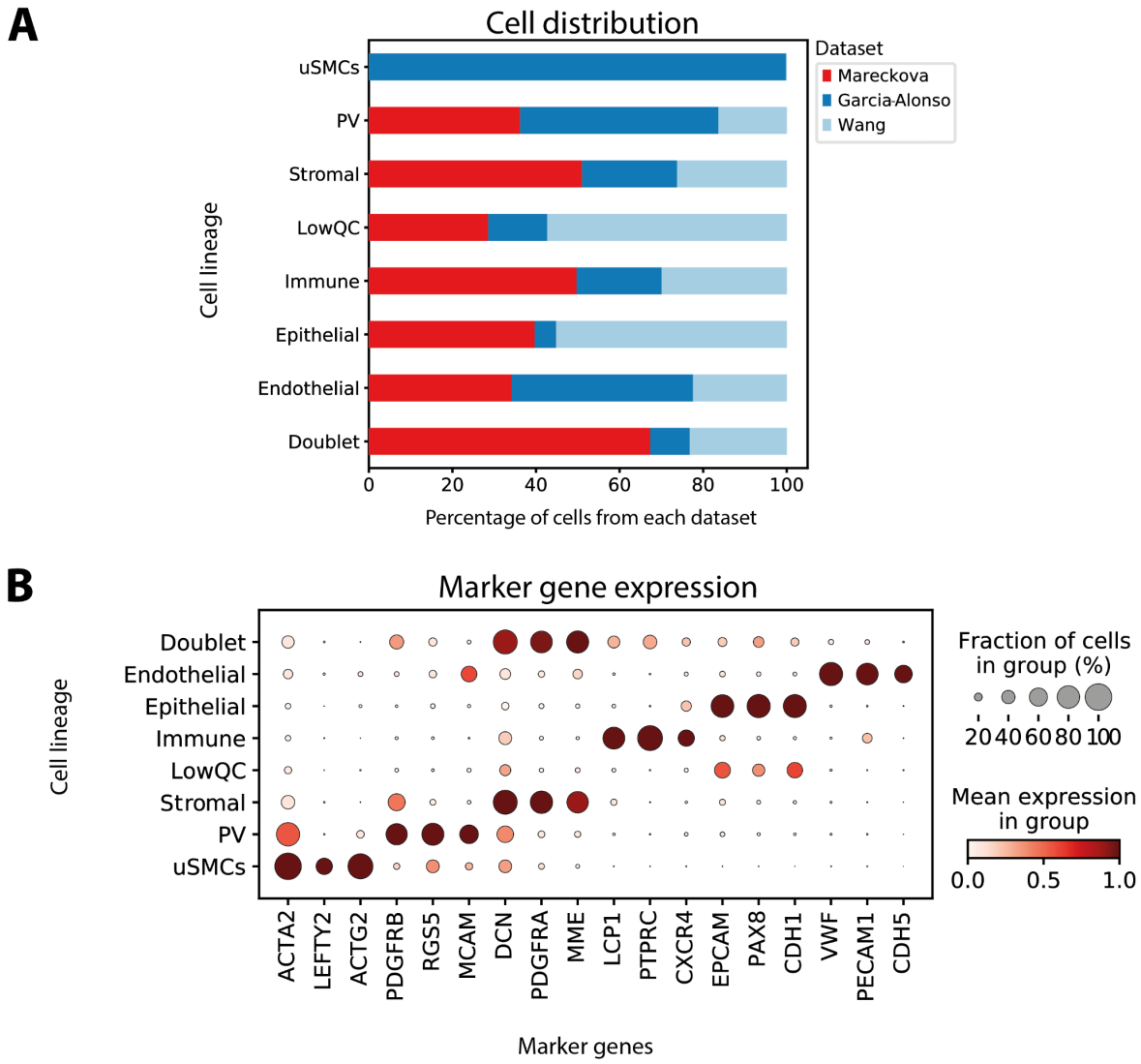


Figure 3.6: Main cell lineages - dataset contribution and marker gene expression. The contribution of each of the three datasets (Mareckova, Garcia-Alonso and Wang) to the cell lineages on the y-axis is expressed as the percentage of all cells assigned to that lineage (A). The dotplot in (B) shows the variance-scaled, log-transformed expression of 3 marker genes (x-axis) characteristic for each of the lineages (y-axis). *Abbreviations*: PV, perivascular; QC, quality control; uSMCs, uterine smooth muscle cells.

Having integrated the datasets and identified the main lineages successfully, the focus of this thesis was to annotate and characterise the epithelial and mesenchymal cell lineages in more detail (see the next sections). Detailed annotation of these two lineages was necessary for defining and understanding the cellular composition of the *in vitro* endometrial models that I generated as part of my DPhil studies. These included epithelial cell organoids and co-cultures of the epithelial organoids with stromal cells. Chapter 5 discusses the utility of the comprehensive cell annotation in understanding which endometrial cell states are present *in vivo* and which *in vitro*.

3.4.2.3 The epithelial cell lineage

The final UMAP for the epithelial cell lineage after removing low QC cells and doublets is shown in Figure 3.7A. Contribution of each dataset, menstrual cycle phase and donor to each cell state is visualised in Figure 3.7B-D. Overall, I defined 15 different cell states within the epithelial cell lineage using a combination of clustering analyses, label transfer of previously published cell state markers and manual inspection of TF-IDF results for cluster-specific marker gene expression. As described for the initial integration with the Tan *et al.* dataset, the cell states can be grouped into 4 main groups - ciliated, proliferative, secretory and hormones - and two smaller clusters of MUC5B and cervix cell states (Figure 3.7A). Cell contamination from the cervix is not surprising given the sampling device has to pass through the cervix to obtain a biopsy of the endometrium. Most of the 128 cervical cells detected came from two donors: E3 (Garcia-Alonso dataset) and SAMN15049044 (Wang dataset). Similarly, with regards to the MUC5B population (total of 270 cells), these cells were detected in a few samples but in very small numbers (< 5 cells per donor). The majority of MUC5B cells came from two donors:

SAMN15049044 (Wang dataset) and E3 (Garcia-Alonso dataset) contributing 132 and 100 cells, respectively.

The ciliated group

Three ciliated cell states were defined as described previously[98], expressing the following genes: ciliated (*DNAH11, PIFO, OMG*), pre-ciliated (*CCNO*, *NEK2*, *CDC20B*), ciliated_LGR5 (*LEFTY1, LGR5, PTGS1*) as shown in Figure 3.7E. Taking into account the phase of the menstrual cycle, during the proliferative phase, the pre-ciliated state is most abundant while the ciliated and ciliated_LGR5 states are characteristic of the secretory phase (Figure 3.8). The data also suggests that taking hormonal therapy reduces the number of all ciliated cell states, but requires further validation as only two samples from donors on hormones were analysed (Figure 3.8).

The proliferative group

Within the proliferative group of cells, I defined two cell states, specifically the proliferative (prolif) and prolif_LGR5 cell states. The prolif cell state expresses *IHH* and *GREM2* while prolif_LGR5 expresses *LGR5* and *WNT7A*. The prolif cell state is highly abundant during the proliferative phase, and its numbers decrease for samples in early secretory phase (Figure 3.8). The prolif_LGR5 cells also appear abundant up until the early secretory phase when their numbers start to decline with both the prolif and prolif_LGR5 cells being absent from mid-secretory phase onwards (Figure 3.8B).

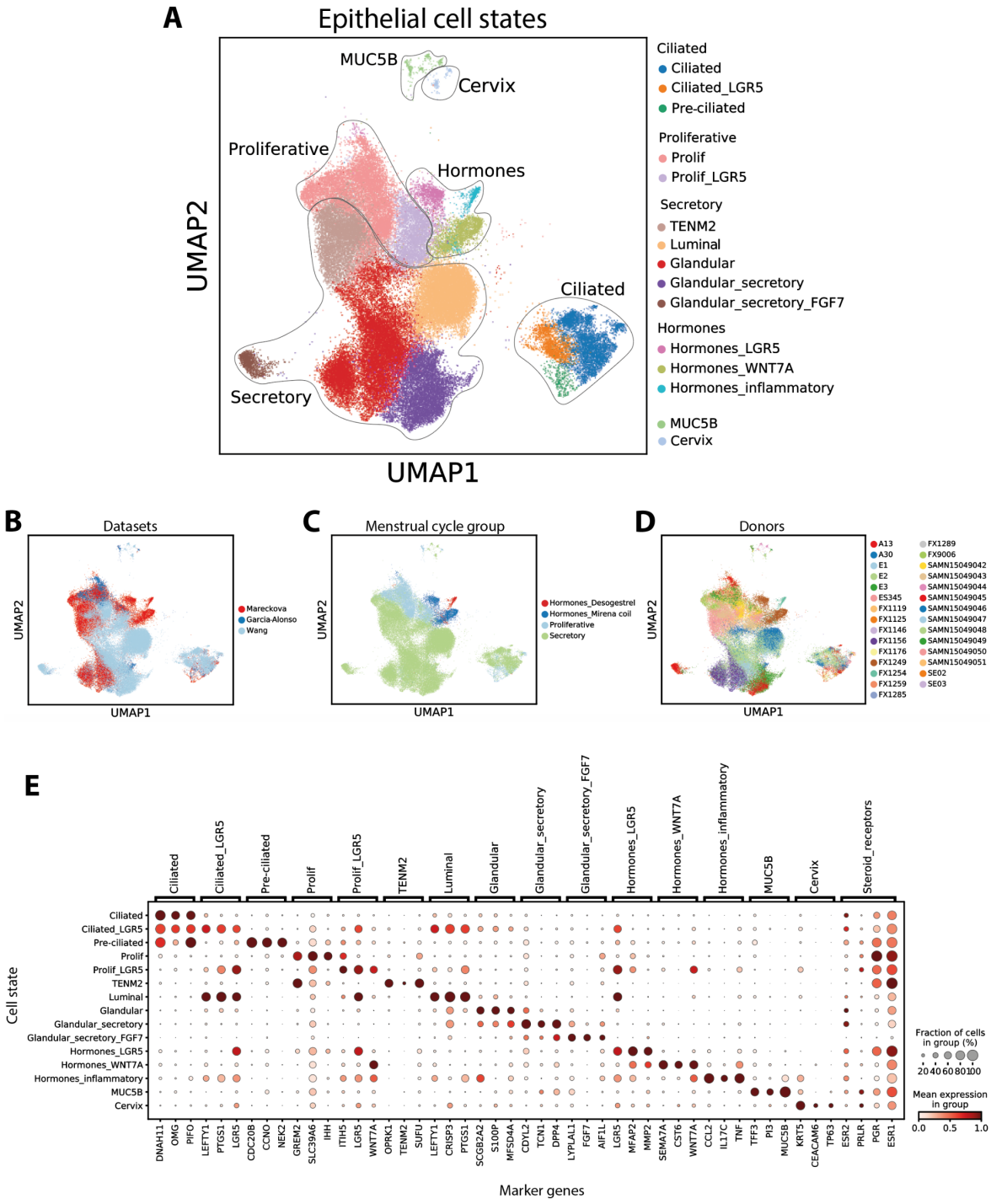


Figure 3.7: Epithelial cell states identified. Shown are UMAP projections of the three scRNAseq datasets integrated, coloured by epithelial cell states identified (A), the different datasets (B), menstrual cycle phase and hormonal therapy taken (C) and individual donor samples (D). The dotplot in (E) shows the variance-scaled, log-transformed expression of 3 marker genes (x-axis) characteristic for each of the cell states (y-axis). Expression of steroid hormones per cell lineage is also shown. *Abbreviations*: Prolif, proliferative; scRNAseq, single-cell RNA-sequencing; UMAP, uniform manifold approximation and projection.

The secretory group

Five cell states were defined in the secretory group, including two novel cell states. The glandular, glandular_secretory and luminal populations have been described previously[98]. I observed that the early to mid-secretory phase samples seemed to be enriched in the luminal population, expressing markers such as *CRISP3*, *LEFTY1* and *PTGS1*. The top markers identified as cluster-specific for the luminal cell state showed high expression in this group of cells, but I also noted a similar expression pattern at much lower levels in the ciliated_LGR5 cell state suggesting these cell states may be related. The two novel cell states identified and their 3 top marker genes were: TENM2 (*TENM2, OPRK1, SUFU*) and glandular_secretory_FGF7 (*FGF7, AIF1L, LYPLAL1*). While some TENM2 cells could be observed in the proliferative phase samples, most cells came from early to early/mid secretory phase samples (Figure 3.8). Out of the top 30 marker genes identified by TF-IDF, 8 were previously reported as differentially expressed in the receptive endometrium (Appendix 2). The glandular_secretory_FGF7 cells were confined to late secretory phase samples from two donors - one from the Garcia-Alonso dataset (E1) and one from the Wang dataset (SAMN15049045) as highlighted in Figure 3.8B. Moreover, this cell state lacked the expression of steroid hormone receptors when compared to the rest of the epithelial cell states (Figure 3.7E).

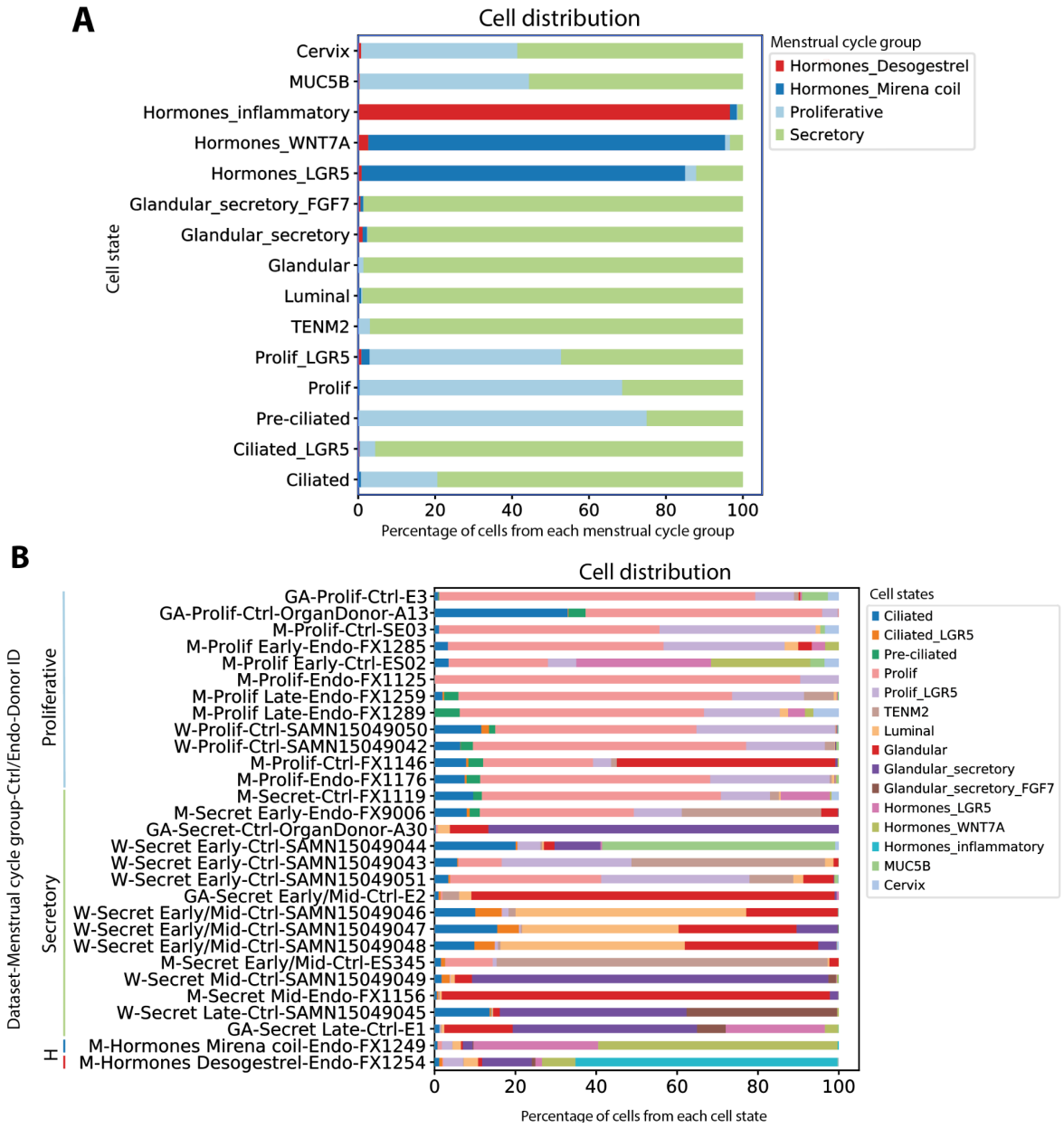


Figure 3.8: Epithelial cell states distribution across the menstrual cycle and donors. The contribution of cell from each of the menstrual cycle groups (Hormones_Desogestrel, Hormones_Mirena coil, Proliferative or Secretory) to the cell lineages on the y-axis is expressed as the percentage of all cells assigned to that lineage (A). In (B) the distribution of the epithelial cell states is shown on a sample basis. Each sample name contains the dataset it came from (GA for Garcia-Alonso, M for Mareckova and W for Wang), followed by menstrual cycle phase, information about being either a control or endometriosis case, indication of samples coming from whole uterus biopsies (OrganDonor) and the donor ID. *Abbreviations*: Ctrl, control; Endo, endometriosis; H, hormones; Prolif, proliferative; Secret, secretory.

The hormones group

Cell states in the hormones group were divided into 3 types: hormones_LGR5 and hormones_WNT7A cell states from a donor on progestins in the form of the Mirena coil, and hormones_inflammatory from a donor taking the progestin pill Desogestrel orally (Figure 3.8). All showed a unique expression of markers with three such markers per cell state plotted in Figure 3.7E. Since only one sample per each hormonal regiment was analysed, I cannot rule out these cell states are donor-specific and will therefore need to validate and re-assess the transcriptomic signatures of cells from donors on hormones in a larger set of samples.

Low QC samples

An important point to note is that not all samples followed the general trend of cell state distribution for the menstrual cycle phase they were assigned to. Specifically, there were samples that overall had many low QC cells and thus in some cases only < 100 epithelial cells remained for analyses. These included donors SE02, SE03, FX1146, FX1285 and FX1289 from the Mareckova dataset. The value of having and keeping these samples for follow-up downstream analyses needs to be further investigated.

Cellular heterogeneity in donors with and without endometriosis

Next, I explored and compared the cellular composition of donors with and without endometriosis. Given the lower number of donors with endometriosis in my integrated map as compared to controls and the bias of the controls being mostly from the secretory phase of the cycle, here, I describe only the patterns observed that need additional validation. Overall, the cellular composition of the endometrium

during the proliferative phase did not seem to differ between controls and cases (Figure 3.9). However, it can be noted that there is a higher percentage of ciliated and glandular cell states in healthy controls, while the TENM2 population is reduced. There appear to be more distinct changes in the secretory phase, where in endometriosis cases the ciliated_LGR5, luminal and glandular_secretory_FGF7 cell states are almost absent. However, this is most likely due to the lack of samples in the endometriosis cohort from the early/mid secretory and late secretory phases, in which these cells were defined. Dividing the menstrual cycle into proliferative and secretory phases only is not sufficient for capturing such changes. However, this was done to be able to group a larger number of samples per menstrual phase group for both cases and controls when comparing between them. The number of samples would be insufficient if split based on the more fine-grained annotation of the menstrual cycle and requires a larger dataset.

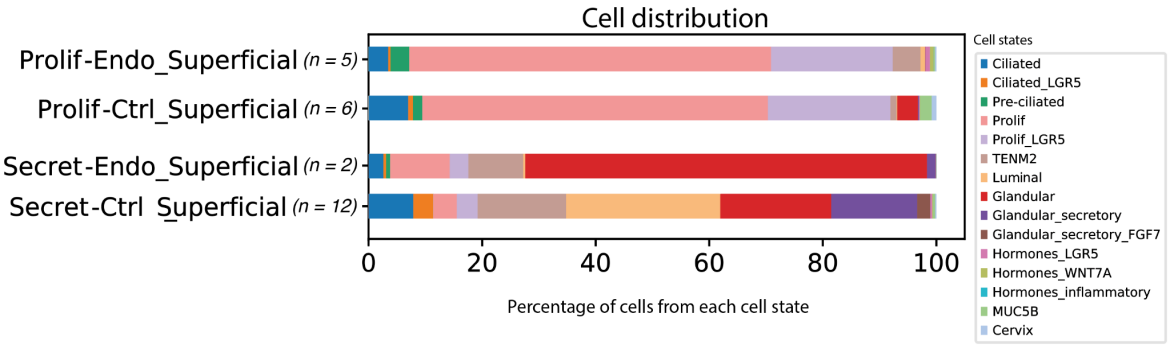


Figure 3.9: Epithelial cell state distribution in endometriosis. The cell state composition of samples from either endometriosis cases or controls without endometriosis is shown both for samples from the proliferative and secretory phases. Only superficial biopsies of the endometrium were considered - the 2 donor samples obtained as part of the organ donor programme are not shown. *Abbreviations*: Ctrl, control; Endo, endometriosis; Prolif, proliferative; Secret, secretory.

3.4.2.4 The mesenchymal cell lineage

The final UMAP for the mesenchymal cell lineage after removing low QC cells, doublets and uSMCs is shown in Figure 3.10A. Contribution of each dataset, menstrual cycle phase and donor to each cell state is visualised in Figure 3.10B-D. Overall, I identified 15 different cell states that can be grouped into 5 main groups: fibroblasts, perivascular, stromal non-decidualised, stromal intermediate and stromal decidualised cells.

The fibroblasts group

Within the fibroblasts group, the previously described fibroblast C7 (Fib_C7)[98] was detected in samples from the organ donor programme in the Garcia-Alonso dataset (Figure 3.10B). In the original publication, Fib_C7 was found at the endometrial-myometrial junction, thus the superficial biopsies used in this thesis likely fail to sample this region and capture the Fib_C7 cell state.

The perivascular group

In the perivascular group, I identified 4 cell states. Previously, PV cells specific to the myometrium (PV_MYH11) and those specific to the endometrium (PV_STEAP4) were described[98,99] and I also detected them here. To distinguish their myometrial (m) and endometrial (e) origin, I call them mPV_MYH11 and ePV_STEAP4 cells in this thesis. The mPV_MYH11 state came from samples from the organ donor programme where parts of the myometrium were also sampled. Interestingly, I identified two novel endometrial PV cell states that I called ePV_MMP11 and ePV_AOC3. The marker genes for ePV_MMP11 were *MMP11, CYGB*, and *PAG1* as shown in Figure 3.10E.

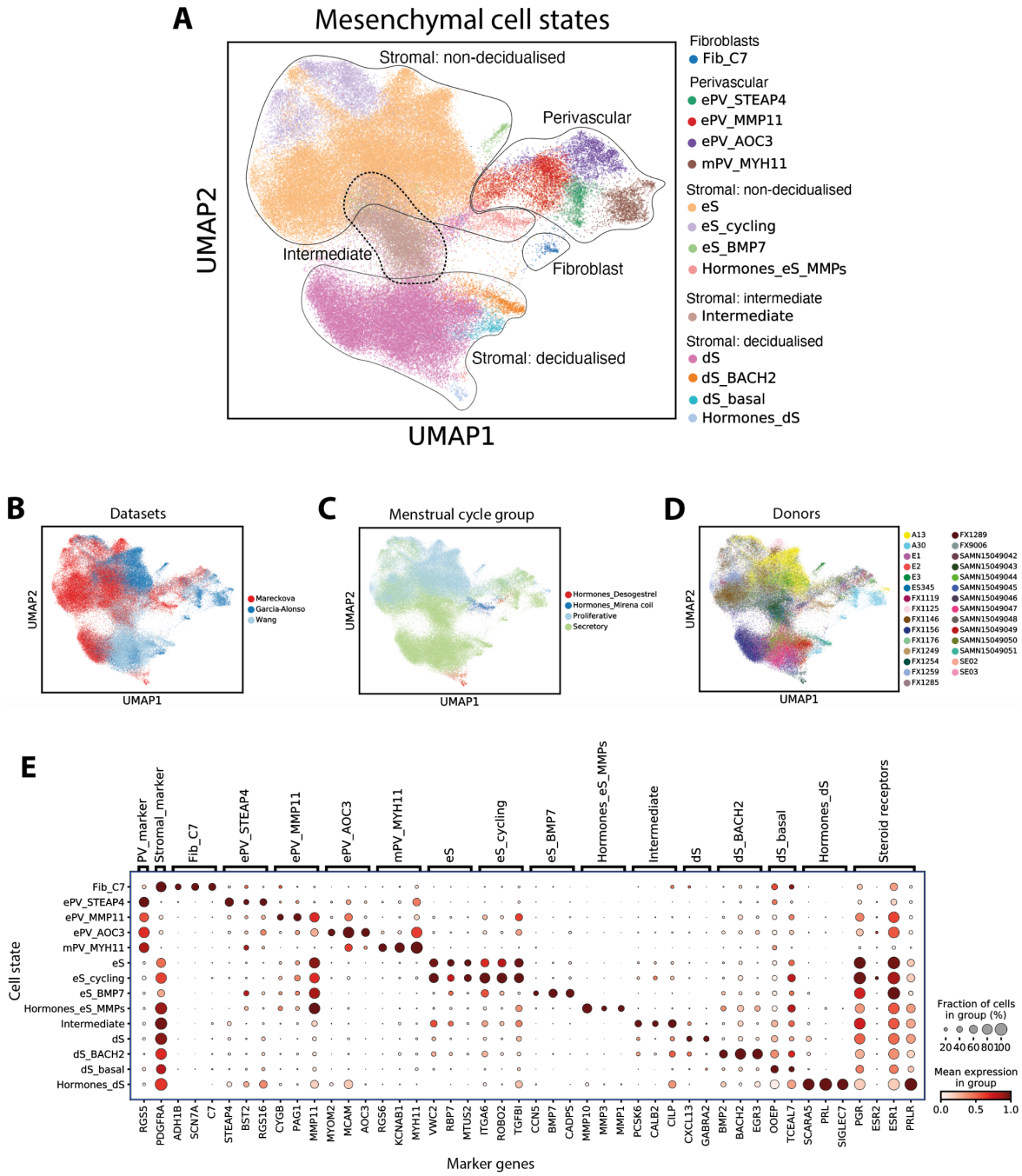


Figure 3.10: Mesenchymal cell states identified. Shown are UMAP projections of the three scRNAseq datasets integrated, coloured by mesenchymal cell states identified (A), the different datasets (B), menstrual cycle phase and hormonal therapy taken (C) and individual donor samples (D). The dotplot in (E) shows the variance-scaled, log-transformed expression of 3 marker genes (x-axis) characteristic for each of the cell states (y-axis). Expression of steroid hormone receptors, general PV and stromal markers per cell lineage is also shown. *Abbreviations*: dS, decidualised stromal cells; ePV, endometrial perivascular cells; eS, endometrial stromal cells; Fib, fibroblast; MMPs, matrix metalloproteinases; mPV, myometrial perivascular cells; PV, perivascular cells; scRNAseq, single-cell RNA-sequencing; UMAP, uniform manifold approximation and projection.

The marker genes of the ePV_MMP11 cells state were also lowly expressed in cells from the stromal non-decidualised group. This suggests that there may be a relationship between the ePV_MMP11 cells and the proliferative phase endometrial stromal cells. Perivascular cells have been proposed to be involved in endometrial regeneration[167–169] thus these cells should be studied further. Nonetheless, the ePV_MMP11 cells specifically expressed PV markers (e.g. *RGS5*) and not stromal cell markers (e.g. *PDGFRA*). This extrapolates also to the second novel cell state, ePV_AOC3 with cells characterised by the expression of *AOC3*, *MCAM*, and *MYOM2*. *MCAM* (also known as CD146) was previously used as a marker of endometrial mesenchymal stem cells in combination with *PDGFRB* expression[167]. *MCAM* was expressed across all perivascular cells, but ePV_AOC3 showed the highest expression of this marker. Expression of putative stem cell markers, such as *PDGFRB, SUSD2, THY1* across the PV cells is shown in Appendix 3.

Stromal cells: the intermediate group

Shifting to the stromal cells, my analyses revealed substantial heterogeneity within this group of cells. Apart from the non-decidualised (eS) and decidualised groups (dS) described previously[97,98], I also identified a novel cell state sitting in between these two extremes and termed it the intermediate cell state. The 3 most specific markers for this cell state were *CILP*, *CALB2*, and *PCSK6* as shown in Figure 3.10E. It appeared that the intermediate state is to a small extent present during the proliferative phase already (Figure 3.11) with cell numbers increasing during early to early/mid secretory phase (i.e. around the time of window of implantation), followed by a decrease in their numbers in the late secretory phase of the cycle.

Stromal cells: the non-decidualised group

I noted cellular heterogeneity both within the non-decidualised and decidualised groups. In the non-decidualised group, the previously described endometrial stromal (eS) cells[98] and their cycling version undergoing mitosis (eS_cycling), were identified. Accordingly, these cells were found present in large numbers during the proliferative phase of the menstrual cycle (Figure 3.11). Furthermore, a small cluster of cells uniquely expressing *BMP7, CCN5* and *CADPS* was also identified and termed eS_BMP7. Multiple donors contributed to this cluster; however, most of the cells came from one donor (E3) from the Garcia-Alonso dataset (Figure 3.11B). Given a similar cell state was described previously[170] for a limited number of samples (*n* = 3) from the proliferative phase, this cell state requires further investigation. The final cell state within the non-decidualised group is the hormones_eS_MMPs cell state expressing a set of matrix metalloproteinases (MMPs), such as *MMP1, MMP3,* and *MMP10* (Figure 3.10E). Cells from donors on hormones contributed to this cell state, indicating that the high MMP expression is likely a transcriptomics signature induced by the presence of the IUD Mirena coil (Figure 3.11).

Stromal cells: the decidualised group

Within the decidualised group, in correspondence with the previously defined decidualised stromal cells (dS), I also detected this dS cell population that is characteristic of the secretory phase of the cycle (Figure 3.11). Two markers uniquely expressed by dS cells were *CXCL13* and *GABRA2.* Other novel cell states identified included the dS_BACH2 population with expression of *BACH2*, *BMP2* and *EGR3* defining this group of cells present in multiple donors during the secretory phase. However, a large number of these cells were contributed by only one donor

- SAMN15049051 from the Wang dataset (Figure 3.11B). Similarly, the cell state termed dS_basal was largely made of cells from one donor in the Garcia-Alonso dataset where the whole uterus was collected. The name dS_basal therefore represents the hypothesised location of this cell state - the *basalis* layer unlikely to be sampled when taking superficial biopsies. However, given the cell state was present in small numbers in other donors as well, I decided to explore its existence and marker gene expression (*OOEP, TCEAL7*) further in the next chapter. Hormones_dS is the last cell state identified in my unified map of the mesenchymal cell lineage. Cells in this cluster were exclusively found in the one donor taking progestin (pill name Desogestrel) orally. The transcriptomic signature of this cell state included the expression of genes such as *SIGLEC7, SCARA5* and *PRL* (Figure 3.11).

Cellular heterogeneity in donors with and without endometriosis

Next, I wanted to explore the presence of different cell states in the endometrium of women with and without endometriosis and understand if there are any differences between these groups. Similarly to what was described above for the epithelial cells states, I only considered the data from superficial biopsies and split it into proliferative and secretory phases only, not the finer annotation of the menstrual cycle as this was not always available and the sample size was not large enough for such detailed comparisons. The cell state distribution is shown in Figure 3.12 and suggests that during the proliferative phase the cell distribution is comparable between controls and cases with only the eS_BMP7 cell state detected at a larger scale in controls.

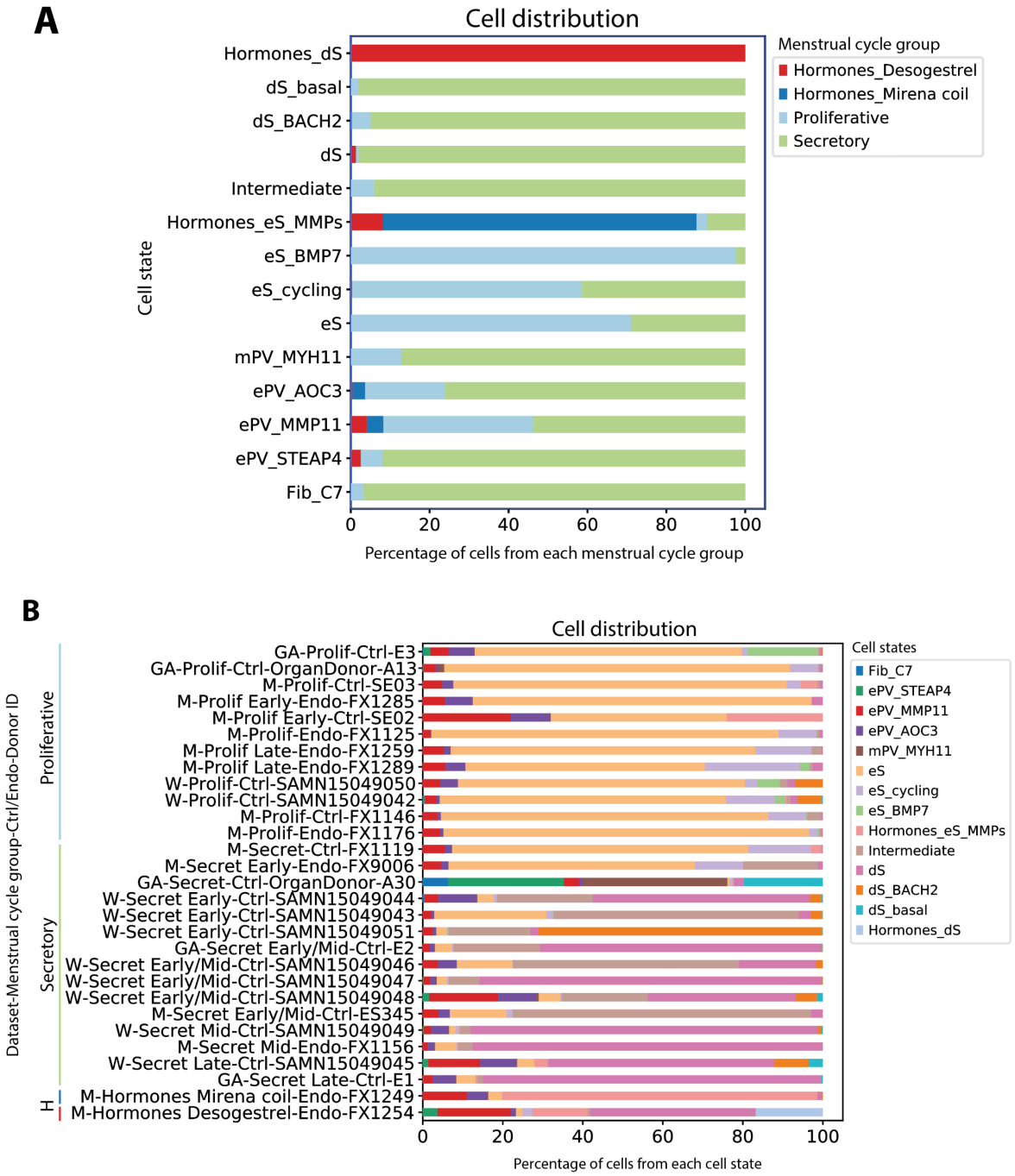


Figure 3.11: Mesenchymal cell state distribution across the menstrual cycle and donors. The contribution of cells from each of the menstrual cycle groups (Hormones_Desogestrel, Hormones_Mirena coil, Proliferative or Secretory) to the cell lineages on the y-axis is expressed as the percentage of all cells assigned to that lineage (A). In (B) the distribution of the mesenchymal cell states is shown on a sample basis. Each sample name contains the dataset it came from (GA for Garcia-Alonso, M for Mareckova and W for Wang), followed by menstrual cycle phase, information about being either a control or endometriosis case, indication of samples coming from whole uterus biopsies (OrganDonor) and the donor ID. *Abbreviations*: Ctrl, control; dS, decidualised stromal cells; ePV, endometrial perivascular cells; eS, endometrial stromal cells; Endo, endometriosis; Fib, fibroblast; H, hormones; MMPs, matrix metalloproteinases; mPV, myometrial perivascular cells; Prolif, proliferative; Secret, secretory.

During the secretory phase, in endometriosis cases the proportion of the eS cell state was larger and proportion of dS cells smaller when compared to controls. This could suggest reduced decidualisation of the stromal cells in endometriosis cases, but could also be explained by the lack of samples from later secretory phase stages in this group. Nevertheless, the cellular composition of the endometrium in both health and endometriosis needs to be studied in a larger cohort of samples. It will also be of interest to confirm whether the insertion of the IUD Mirena coil leads to more of a proliferative phase phenotype with the hormones_eS_MMPs signature being dominant (Figure 3.12), while taking Desogestrel orally induces more of a decidualised phenotype with both dS and hormones_dS cells present at larger proportions.

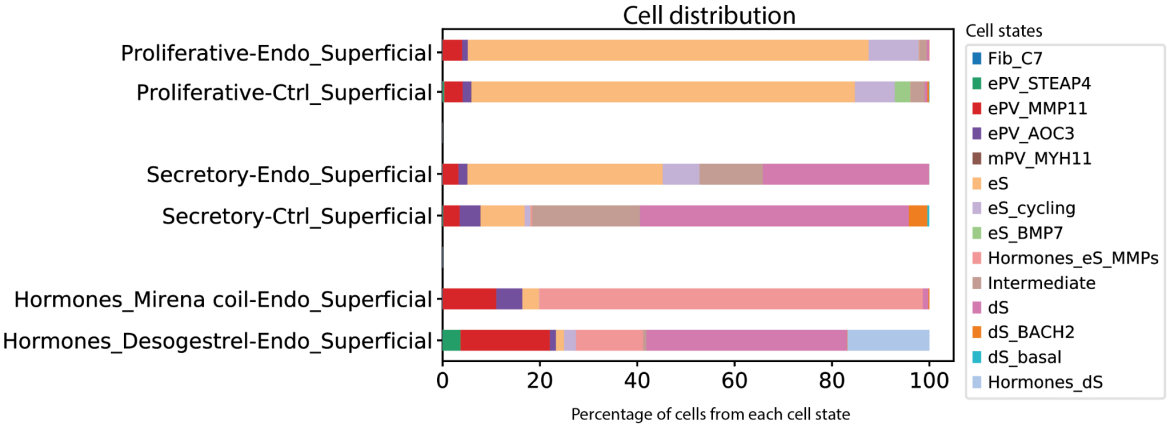


Figure 3.12: Mesenchymal cell state distribution in endometriosis and under hormonal therapy. The cell state composition of samples from either endometriosis cases (Endo) or controls without endometriosis (Ctrl) is shown both for samples from the proliferative and secretory phases. Cellular composition of samples from donors on hormones is also shown. Only superficial biopsies of the endometrium are included in these plots. *Abbreviations*: Ctrl, control; dS, decidualised stromal cells; ePV, endometrial perivascular cells; eS, endometrial stromal cells; Endo, endometriosis; Fib, fibroblast; MMPs, matrix metalloproteinases; mPV, myometrial perivascular cells.

3.5 Discussion

In summary, in this chapter I aimed to generate a harmonised map of endometrial cell states across the menstrual cycle and under the influence of exogenous hormones using 3 previously published datasets and my own data. In the end, I had to exclude the Tan dataset from further analyses as its control samples did not match the patterns of expression observed across the controls of the other datasets. Annotating the integrated map built using my own data and the Garcia-Alonso and Wang datasets led to the identification of multiple novel cell states in the epithelial and mesenchymal cell lineages, suggesting that larger cohort sizes are needed to dissect the heterogeneity of endometrial cells. Distribution of the cell states identified was examined in the context of the endometrium from women with and without endometriosis. Moreover, I also described the transcriptomic signatures associated with taking two different types of hormonal therapy. The findings are discussed in more detail below.

3.5.1 Exclusion of the Tan dataset

The Tan dataset was excluded from the harmonised map as the samples from donors not taking hormonal therapy exhibited transcriptomic signatures similar to those from donors on hormones in the Mareckova dataset. Multiple factors could be responsible for the discrepancies observed. Firstly, as described in Chapter 2, the Tan dataset samples were digested into single cells at 6°C using a protease and DNAse I solution, which was different to the digestion with collagenases used in the other studies. The digestion protocols and enzymes used can have a large impact on the transcriptomic profiles of the obtained cells, as shown for example for the gut mucosa[171]. It is therefore plausible that these different protocols could have

introduced strong artefacts in the data, separating the cells in the clustering analyses and needs to be investigated. To understand the differences in gene expression and cell type composition between the different digestion protocols used, in the future, samples should be digested using both protocols in parallel and compared. Secondly, the observed discrepancies could also be the result of the data not being aligned to the same reference genome. Due to time constraints, the Tan dataset could not be re-aligned to the same reference as used for the Mareckova, Garcia-Alonso and Wang datasets. I am therefore currently waiting for the re-alignment of the Tan data to match the rest of the datasets and will be repeating the integration and analyses to explore the genome reference as a potential source of the discrepancies observed.

3.5.2 Novel epithelial cell states

When it comes to the integrated map of the epithelial cells, I identified many cell states that have been described previously by Garcia-Alonso *et al.* but also two novel cell states: the TENM2 and glandular_secretory_FGF7 populations. The TENM2 cell state was observed to peak in its numbers at around the time of the window of implantation, suggesting that this cell population may play a role in endometrial receptivity. Indeed, the TENM2 cells highly and uniquely express the Opioid Receptor Kappa 1 (*OPRK1*) gene and 7 other genes listed in Appendix 2 which have been found to be differentially expressed in the endometrium during the receptive phase[172,173]. Furthermore, the other marker genes reported in this thesis included Teneurin Transmembrane Protein 2 (*TENM2*) and Suppressor of fusing (*SUFU*). *TENM2* has been described mostly in the context of neural development, playing an important role in calcium-mediated signalling and enabling cell-cell adhesion[174,175]. While *SUFU* is a negative regulator of hedgehog and beta-catenin

signalling[176,177]. In the context of the endometrium, specifically endometrial receptivity, neither *TENM2* nor *SUFU* have been studied in detail. However, given inhibition of WNT/beta-catenin signalling is needed for stopping the oestrogen induced proliferation and inducing cell differentiation needed for embryo implantation[178,179], one can speculate that increased *SUFU* expression in these cells could play a role in promoting such differentiation by negatively regulating beta-catenin signalling. Upregulation of *SUFU* expression has been found to suppress the growth of endometrial cancer cells *in vitro* and *in vivo*[180], further supporting the role of *SUFU* expression in inhibiting endometrial cell proliferation.

With regards to the glandular_secretory_FGF7 cell state, the markers identified fibroblast growth factor 7 (*FGF7),* Allograft Inflammatory Factor 1 Like (*AIF1L*) and Lysophospholipase like 1 (*LYPLAL1*) have been studied to very little, if any, extent in the human endometrium. *FGF7* is thought to be expressed by mesenchymal and not epithelial cells, but has been shown to be expressed in the luminal epithelial cells of porcine endometrium[181]. Moreover, studies of mouse and rat homologs of *FGF7* have suggested that it has a role in the morphogenesis of epithelium, re-epithelialisation of wounds and acts as a strong chemoattractant[182,183]. Whether *FGF7* could have such a role in the human endometrium is to be determined, but given its expression is most prominent during the late secretory phase, one can hypothesise that indeed these cells could play a role in the re-epithelisation of the endometrium, which would be required at the onset of menstruation.

Lastly, in my re-analysis of the data, I observed a small number of cervical cells and MUC5B cells coming from the Garcia-Alonso and Wang datasets that the authors

did not describe in their publications. Likely, this was due to the low numbers detected and contribution from only a couple of donors, prompting the authors to disregard these cells as donor-specific clusters. The cervical cells were likely discarded as contamination. The MUC5B population was only described in detail by Tan *et al*., identifying this cell population as a likely progenitor cell type involved in endometrial regeneration. There are multiple reasons for why the MUC5B population might have been enriched in the Tan dataset, including the use of the different digestion protocol as mentioned above. The cold digestion protocol could have enriched for this particular cell population. Moreover, the actual sampling of the endometrium might have differed between clinics and/or clinicians, and thus might have led to enrichment for slightly different regions of the endometrium in the different datasets. Another reason could also be the use of exogenous hormones that alter the properties of the endometrium, making it possible to sample these cells more efficiently or even increasing their numbers. This cell population requires further investigation, including its exact location within the endometrium and understanding the conditions that favour its survival and extraction for scRNAseq analyses to be able to draw any conclusions about this cell state.

3.5.3 Novel mesenchymal cell states

The identified stromal cells in the mesenchymal cell lineage consisted of cells previously described (eS, dS, eS_cycling)[98] and eS_BMP7 cells characteristic of the proliferative phase and dS_BACH2 and dS_basal states from the secretory phase. Each of these cell populations was relatively low in the numbers detected and the number of donors these cells were found in. Therefore, it cannot be excluded that some of these cell states may be donor-specific. Nonetheless, I defined their marker genes and will explore their existence in a larger set of samples

in Chapter 4. The eS_BMP7 cell state expressed a set of genes previously described for a subpopulation of proliferative phase endometrial stromal cells termed BMP7+[170]. These BMP7+ cells were suggested to be involved in myofibroblast differentiation, epithelial mesenchymal transition (EMT) and TGFβ1-WNT signalling[170], which is in line with the marker gene expression, such as Bone Morphogenetic Protein 7 (*BMP7),* Cellular Communication Network Factor 5 (*CCN5*) and Calcium Dependent Secretion Activator (*CADPS*) reported in this thesis. It is likely these cells are important for the remodelling and tissue repair that occurs during the proliferative phase after menstruation. Given the previous publication looked at endometrial biopsies from day 7 since the onset of menstrual bleeding, it may be that these cells are detected only in the early proliferative phase and thus were not detected in many samples presented here.

With regards to the new intermediate cell state, its increase in numbers during the early to early/mid secretory phase and expression of proprotein convertase 6 (*PCSK6*), both hint at its role during the window of implantation and stromal cell decidualisation. Previous studies in mice showed that blocking *PCSK6* expression led to complete inhibition of implantation and pregnancy[184] and *PCSK6* expression was shown to be important in mouse stromal cell decidualisation[185]. In humans, however, at the protein level, using immunohistochemical analyses, high expression of *PCSK6* was reported in stromal cells during the late secretory phase, while the early secretory phase was characterised by high *PCSK6* expression in the epithelial and not stromal cells[184]. *In vitro* analyses of human endometrial stromal cells showed that in order for *PCSK6* expression to increase, the cells needed to be stimulated with both oestrogen and progesterone and without *PCSK6*, the stromal

cells would not become decidualised[186]. Taken together, it is therefore likely that the intermediate cell state identified in this thesis marks the transition from proliferative phase stromal cells to their decidualised counterparts. It would now be of interest to understand where these cells are located within the tissue and what factors trigger their transition and transcriptomic phenotype.

My analyses of the perivascular cells revealed a much more heterogeneous population of these cells than described previously. Apart from the already known ePV_STEAP4 and mPV_MYH11 cell states, I also identified two cell states that I called ePV_MMP11 and ePV_AOC3 and hypothesised that these cells may be involved in the regeneration of the stromal cells based on the marker gene expression gradient. Perivascular cells have long been suggested to play a role in the monthly regeneration of the endometrium, with *MCAM* (known as CD146) in combination with *PDGFRB* expression proposed as putative mesenchymal stem cell markers[167]. ePV_AOC3 expressed *MCAM* at the highest level out of all the PV cells identified and ePV_MMP11 had the highest expression of *PDGFRB*. With regards to other putative stem cell markers, *SUSD2* and *THY1*, *SUSD2* was mainly detected in the ePV_AOC3 cell state while *THY1* was more highly expressed in ePV_MMP1 cells. Taken together, it seems likely that these cells are indeed the putative progenitor cells described previously to have stem cell like properties, but their existence and exact location need to be validated. Furthermore, their potential to differentiate into stromal cells also needs to be tested. Lastly, based on marker gene expression, including high expression of *MMP11*, the ePV_MMP11 cell state is transcriptomically similar to the previously described PV2 cells in the pregnant decidua localised to the smooth muscle media of spiral arteries[187]. ePV_MMP1 and

ePV_AOC3 populations were likely not defined in the previous endometrial datasets, including the pregnant decidua, as not enough cells were analysed to achieve the resolution needed to distinguish these cell states from the rest of the PV cells. These cell populations now require further investigation in the non-pregnant endometrium.

3.5.4 Endometrial cells in endometriosis and controls

When it comes to comparing the cellular composition of endometrial biopsies between endometriosis cases and controls, no significant changes were observed in the epithelial and mesenchymal cell lineages in the proliferative phase. Looking at the secretory phase samples could suggest a reduction in the decidualisation response in endometriosis cases. This would be in line with previously published work on progesterone resistance in the endometrium of women with endometriosis[188,189] reporting incomplete transition of the endometrium from the proliferative to secretory phase, with cell exhibiting enhanced survival and lack of progesterone-induced differentiation. However, the differences observed in this chapter could also be attributed to the uneven distribution of samples across the secretory phase when it comes to cases and controls. Twelve samples from women without endometriosis came from the secretory phase while only two samples from endometriosis cases were from the secretory phase. It was clear that the cellular composition of the endometrium varied greatly across the menstrual cycle with large differences observed within the proliferative and secretory phases themselves. Thus ideally, both the proliferative and secretory phases should be split into further subgroups of early and late proliferative phase and early, mid and late secretory phase to account for these differences. However, such assignment based on histology can be challenging and in this case was not available for all samples.

Moreover, a limited number of samples was analysed, making it impossible to have enough samples for both cases and controls in all the different menstrual cycle phase subgroups. A larger cohort of samples across all of these subgroups is needed to understand when and if the endometrium of endometriosis cases exhibits a dysregulated response to either oestrogen or progesterone.

3.5.5 Endometrial cells under the influence of exogenous hormones

According to a 2019 United Nation's study, globally 407 million women in their reproductive years rely on taking exogenous hormones as a means to prevent unwanted pregnancy[190]. Moreover, many women take hormonal contraception for other reasons, including as a treatment option for various reproductive pathologies, such as heavy menstrual bleeding, endometriosis, fibroids, etc. Even with such a large number of women taking exogenous hormones world-wide, understanding how these hormones influence the transcriptomic profiles of endometrial cells is limited. Many different forms and types of exogenous hormones at various doses can be taken, which makes studying these effects challenging. For example, one microarray study looked at the effect of three different contraceptive methods on the endometrial transcriptome and revealed that the progestin-releasing IUD they studied caused significant alterations in genes regulating immune and inflammatory pathways[191]. Here, after the exclusion of the Tan dataset, I could only describe the transcriptomic profiles of endometrial samples from two donors taking progestins in two different forms: one as an IUD (brand name Mirena coil) and the other as an oral pill (brand name Desogestrel). My data showed that the progestin-releasing IUD caused more of a 'proliferative' phase phenotype, not an inflammatory one as described in the microarray study. The epithelial cells expressed *WNT7A* and *LGR5*, both observed in the proliferative phase endometrium and the mesenchymal cells

also clustered closely with the eS cells characteristic of the proliferative phase, and uniquely expressed a set of MMPs. On the other hand, taking Desogestrel orally induced more of an inflammatory response in the epithelial cells and decidualised/secretory phenotype in the mesenchymal cell lineage. Given the small number of samples studied, these findings require further validation in a larger set of samples before drawing any definitive conclusions.

3.5.6 A summary of future directions

In this chapter, I presented new hypotheses and speculations about the role and existence of the cell states identified. All of these so far descriptive and exploratory findings need to be validated. To do so, I have designed small molecule fluorescent *in situ* hybridization (smFISH) panels of markers for each of the new cell states identified and will stain endometrial tissue sections for these. Visualising the panels of markers will help quantify the presence of these populations in the tissue and where they are located *in situ.* Uncovering the spatiotemporal dynamics of the newly defined cell states will be of crucial importance to better understand the changes the endometrium undergoes throughout the menstrual cycle. Moreover, here I have only presented the more granular annotations of the epithelial and mesenchymal cell lineages. The next steps will include looking at annotating the immune and endothelial cells in more detail and comparing the immune cell profiles of the endometrium in health and endometriosis. The immune cell compartment has been reported to be altered in endometriosis[145,146] and thus warrants further exploration. Furthermore, trajectory analyses (using *Velocyto*) as well as cell-cell communication analyses (using CellPhone DB) will be performed to describe the likely relationships between the cell states identified and the cellular interactions that may lead to the phenotypes observed and reveal the likely functional roles of these cells.

3.6 Conclusions

Taken together, in this chapter I presented the data newly generated as part of my project and its integration with previously published datasets. My aim was to build a harmonised map of these datasets and to annotate the epithelial and mesenchymal cell lineages more finely. The cell annotation was done in order to be used as a reference for annotating the snucRNAseq data presented in Chapter 4 and to annotate the organoid and co-culture models of the endometrium presented in Chapter 5. Moreover, I set out to explore what effects exogenous hormones and endometriosis have on the cellular composition of the endometrium. In conclusion, I succeeded at completing my aims of building the cellular map and annotating the cell populations, but given I could not include a key dataset with samples from donors on hormones with and without endometriosis, I could only generate hypotheses for what effects exogenous hormones and endometriosis may have on the endometrium. These hypotheses and aims are further explored in the following chapter.

Chapter 4: Single-cell and single-nucleus map of the endometrium in endometriosis and controls

4.1 Introduction

When the field of single-cell transcriptomics was established, the standard procedure involved analysing freshly collected samples to obtain viable, high-quality cells. This meant that in the clinical setting, where often only one sample is available on a given day, each sample is processed separately and considered its own unique batch. To overcome this issue, cryopreserved samples started to be profiled and the emerging studies reported that the transcriptomic profiles of cells are not significantly affected by cryopreservation[192–194]. At the beginning of my DPhil, I therefore set out to prospectively collect fresh samples, and after showing that also for the endometrium cryopreservation does not alter the cells' transcriptome (Appendix 1), I started banking cryopreserved samples for later scRNAseq analyses to be performed across a smaller number of batches. Unfortunately, these efforts were halted by the then ongoing Covid-19 pandemic which brought with it cancellations of surgeries for endometriosis (lasting over one year) and our department's laboratory closure (for a six-month period).

However, the field of single-cell genomics is one of the fastest growing ones with new protocols and technologies developed at an extraordinary rate[113,195]. The development of protocols for extracting single nuclei from snap-frozen samples and their optimisation reported for the endometrium[98] meant that instead of relying on fresh samples and scRNAseq, I could look into generating snucRNAseq data for frozen banked samples. Reviewing sample and donor metadata for a large

collection of endometrial biopsies collected and snap-frozen as part of the ENDOX and FENOX studies, I identified 80 samples suitable for snucRNAseq processing based on my inclusion/exclusion criteria described in Chapter 2.

4.1.1 Features of scRNAseq and snucRNAseq data

As briefly highlighted in Chapter 1, both snucRNAseq and scRNAseq data have their unique features/challenges. For example, scRNAseq relies on dissociating tissues into single cells using various enzymes, which has been reported to induce changes in gene expression[196]. The changes have been observed in mitochondrial and stress response genes, introducing artifactual transcriptomic signatures[197,198]. However, these can be corrected computationally when performing data analysis. Moreover, in scRNAseq data, there is a bias towards profiling cells which can be released easily upon tissue dissociation while those more difficult to dissociate may remain uncharacterised. The choice of enzymes and temperature at which tissues are dissociated can also have a large effect[197,199]. In contrast, stress-induced gene expression is not detected in snucRNAseq data and a reduced dissociation bias in detected cell types has been reported[200]. On the other hand, snucRNAseq data suffers from a higher proportion of ambient RNA coming from the cytoplasm of cells from which the nuclei are extracted, and a lower number of transcripts detected[125,201]. To achieve high-resolution cell type identification, aligning snucRNAseq reads to introns is required[125], but even then, some markers are not detected in snucRNAseq data, while they may be abundant and used as canonical markers in scRNAseq datasets[126]. Nonetheless, many bioinformatics tools exist that are able to take these unique features/challenges of scRNAseq and snucRNAseq data into consideration, allowing us to extract information and biological insights from a variety of sample sources[202].

4.1.2 Analysing the human endometrium using snucRNAseq data

For the human endometrium, snucRNAseq data have been generated and published for only 4 samples from the proliferative phase[98]. The authors used these samples to validate their findings from a larger cohort of scRNAseq data, but did not integrate these two data sources. With regards to cell type identification, they provided both general and finer annotations, and cosine distance values as a quantitative measure of similarity between the identified cell states in the snucRNAseq and scRNAseq datasets. In this chapter, I explore whether snucRNAseq and scRNAseq datasets can be integrated successfully and be annotated based on the marker genes I defined in Chapter 3, both at the broader and high-resolution levels. Moreover, I validate the existence of the cell states identified in Chapter 3 in a large cohort of samples covering the endometrium in those with/without endometriosis and under exogenous hormones.

4.2 Aims

1. Generate snucRNAseq data for a large cohort of donors with and without endometriosis during natural cycles and while taking hormonal therapy.
2. Integrate scRNAseq datasets from Chapter 3 with the newly generated snucRNAseq data.
3. Understand the differences and commonalities between the two sources of RNA sequencing data - single cells and single nuclei.
4. Annotate the integrated cells/nuclei map, focusing on the epithelial and mesenchymal lineages.
5. Evaluate the existence of epithelial and mesenchymal cell states identified in Chapter 3 in a larger set of samples.

4.3 Materials & Methods

4.3.1 Sample processing

Snap-frozen endometrial pipelle biopsies were removed from cryovials and embedded in OCT for cryosectioning, storing them at -80°C overnight. The following day, the OCT blocks were left inside the cryostat for ~1 h to equilibrate to the chamber temperature of -20°C. The blocks were trimmed until reaching the tissue, when the first 10 µm thick sections for morphological assessment under a light microscope started to be collected. Three sections were placed on SuperFrost® Plus slides (ThermoFisher, 12312148) before cutting and collecting 50 µm thick sections for nuclei extraction. Depending on tissue size, between 10 to 20 sections were placed into a 7 ml Dounce tissue grinder (Sigma-Aldrich, D9063-1SET) on dry-ice and a further three 10 µm thick sections were placed on slides and stored at -80°C for later histological staining.

4.3.2 Single-nuclei extraction

Tissue collected in the Dounce tissue grinder was placed on ice inside a class II safety cabinet and incubated with 3 ml of homogenisation buffer (see Table 4.1 for buffer composition) for 5 min. To help dissolve the OCT, the suspension was gently mixed with a 2 ml aspiration pipette half-way through the incubation. The tissue was then homogenised by 10-20 strokes of both pestle A and B. The number of strokes was sample-dependent - homogenisation with each pestle was performed until no resistance and tissue changes were observed. Each pestle was washed with 500 µl of the homogenisation buffer and the homogenate filtered through a 40 µm cell strainer into a new 50 ml tube. The sample was then centrifuged using the following setting: 500 x g, 6 min, 4°C, acceleration set at 0 and deceleration set to 3. After

removing the supernatant, 500 µl of wash buffer (see Table 4.2 for buffer composition) was added to the cell pellet and incubated for 2 min on ice. The nuclei pellet was gently resuspended using wide-bore tips to avoid damaging the nuclei, and the yield checked using a haemocytometer and trypan blue. Next, the nuclei suspension was transferred to a 1.5 ml tube and washed twice by adding 1 ml of the wash buffer and centrifugation (500 x g, 3 min, 4°C). The supernatant was removed and nuclei resuspended in 200 µl of the wash buffer (volume was nuclei yield-dependent). To remove debris and clumps, the nuclei suspension was filtered twice through the 40 µm Flowmi® cell strainers and nuclei counted using a haemocytometer and trypan blue.

Table 4.1: Homogenisation buffer composition.

Reagent	Supplier	Product number	Final concentration
Sucrose	Sigma	S0389-1KG	250 mM
2 M KCL	Invitrogen	AM9640G	25 mM
1 M $MgCl_2$	Invitrogen	AM9530G	5 mM
1 M Tris buffer	Invitrogen	AM9855G	10 mM
1 M DTT	Sigma	646563	1 µM
100 x Protease inhibitor	Roche	11697498001	1 x
RNasin Plus	Promega	N261B	0.4 U/µl
SUPERasesIn	Invitrogen	AM2694	0.2 U/µl
Triton X-100	Sigma	93443-100ml	0.05 %
Nuclease free water	Ambion	AM9937	NA

Abbreviations: DTT, Dithiothreitol.

Table 4.2: Wash buffer composition.

Reagent	Supplier	Product number	Final concentration
BSA	Invitrogen	AM2618	2 %
RNasin Plus	Promega	N261B	0.2 U/µl
PBS	Gibco	10010-023	1 x

Abbreviations: BSA, bovine serum albumin; PBS, phosphate buffered saline.

The above-described protocol was optimised for single-nuclei extraction from endometrial biopsies by Dr Elena Prigmore, Senior Staff Scientist at the Wellcome Sanger Institute (WSI). Sample processing for snucRNAseq was carried out by Dr Agnes Oszlanczi (Technical Specialist, WSI), Dr Maria del Carmen Sancho-Serra (Technical Specialist, WSI) and myself. To minimise the number of batches in which the nuclei were extracted and sequenced, in most cases we processed 4 samples (from cases and controls) at the same time, distinguishing between donors based on the genotyping data available.

4.3.3 Single-nuclei RNA-sequencing

In most of the cases, two donor samples per 10x reaction were pooled and nuclei were loaded at volumes and concentrations required to achieve an equal target nuclei recovery per donor: 5,000 - 7,000 nuclei/donor. The 10x Genomics dual index v3.1 libraries (10x Genomics, 1000268) were prepared and sequenced as described in Chapter 3, section 3.3.3.4.

4.3.4 Data analysis - pipeline modifications for snucRNAseq data

Data analysis was performed following the same computational pipelines as described in Chapter 3, section 3.3.3 but with a few modifications. First, the sequencing reads for single nuclei were aligned to whole transcript regions (including both exonic and intronic regions). This was to obtain a higher number of transcripts detected from nuclei, as they contain a higher proportion of pre-mRNA. Second, more stringent filtering was applied to identify high QC nuclei. Nuclei were removed if < 1,000 genes and > 3% mitochondrial gene expression was detected per nucleus. Nuclei with UMI counts > 150,000 and < 3,000 were also filtered out. Thirdly, count matrices from single nuclei were denoised from ambient RNA prior to data integration using the *DecontX* tool[203] on a per sample basis.

Following denoising, all samples were integrated into a single manifold, together with the rest of the scRNAseq datasets (Garcia-Alonso, Wang & Mareckova) using *scVI* as described in Chapter 3. Modifications included correcting for data source (cells vs nuclei) apart from sample/donor and dataset origin as described in Chapter 3. To achieve satisfactory integration, we performed and evaluated multiple versions of integration with a different number of neural network layers and chose 4 layers of neural networks to integrate the data.

The downstream analyses were performed as described in section 3.3.3. To help guide cell state annotation, we imported the annotations obtained from the analyses in Chapter 3 on single cells and projected these onto the UMAPs of the integrated manifold of scRNAseq and snucRNAseq data. Lastly, to assess the similarity of scRNAseq and snucRNAseq data, I calculated the average gene expression per cell state (restricted to 3,000 highly variable genes identified) in both the scRNAseq

and snucRNAseq data separately. Using the *pairwise_distances* function from the *sklearn.metrics* package I calculated the Euclidean distances between the average gene expression values for cells and nuclei for each cell state identified.

4.4 Results

In total, we processed 80 samples for snucRNAseq. Here, I present data for 66 samples due to: (i) not obtaining sufficient nuclei yields for 3 samples, (ii) removing libraries for 4 samples due to low QC metrics, and (iii) not receiving sequencing data in time for 7 samples prior to writing this thesis. Table 4.3 shows an overview of the metadata for the 66 samples presented in this chapter.

Table 4.3: Metadata for samples used for snucRNAseq.

	Mareckova_Nuclei	
	Control	Endoms*
Donor #	20	46
Age^	32.55 ± 1.58	33.00 ± 0.95
Menstrual cycle group		
Menstrual	1	4
Proliferative	5	10
Secretory	8	21
Hormones	5	10
Unknown	1	1
Endometriosis stage (rASRM)		
I/II	-	29
III/IV	-	17

Abbreviations: rASRM, revised American Society for Reproductive Medicine staging system.
Superscripts:
* Endoms stands for Endometriosis.
^ Age is reported as the mean ± standard error of the mean.

4.4.1 Integration of scRNAseq and snucRNAseq data for 5 donor samples

To assess whether integrating scRNAseq and snucRNAseq data is feasible, and what the features of each of the data sources are, I generated both scRNAseq and snucRNAseq data for the same samples from 6 donors. One of the snucRNAseq libraries failed the QC review process and thus I removed all data from this donor from the initial scRNAseq/snucRNAseq comparison analysis. The remaining samples from 5 donors yielded high QC data for 45,464 cells and 34,232 nuclei. Successful integration was achieved using *scVI* with 4 layers of neural networks, which was higher than what was used for integrating scRNAseq data across the multiple datasets in Chapter 3. The integrated UMAP of the data and the main cell/nuclei states identified are shown in Figure 4.1. The cells/nuclei states were identified and assigned based on canonical marker gene expression as described in Chapter 3 and included endothelial, epithelial, immune, stromal and PV cells. Moreover, ciliated epithelial cells were also identified as a distinct cluster already in this broad classification (Figure 4.1A).

Next, I assessed the contribution of scRNAseq (cells) and snucRNAseq (nuclei) data to the main cells/nuclei states identified. As shown in Figure 4.1B, mostly, there appeared to be a good overlap between the cells and nuclei data. However, I observed a lower recovery of nuclei in the PV and endothelial lineages as compared to cells, contributing ~10% of all PV and endothelial cells/nuclei identified. Moreover, a slightly lower percentage of nuclei was also detected in the stromal cell population, while the opposite was true for the epithelial lineage with nuclei contributing ~65% of all cells/nuclei assigned as epithelial. The contribution of cells and nuclei to the immune and ciliated epithelial populations were comparable given the lower total number of nuclei analysed. Lastly, with regards to the number of genes and mRNA

molecules detected per cell/nucleus, the cells and nuclei datasets were comparable, with slightly lower numbers detected in nuclei (Figure 4.1C). The most striking difference between the datasets was observed in the percentage of mitochondrial genes expressed per cell/nucleus, with the nuclei data characterised by extremely low expression of mitochondrial genes (Figure 4.1C).

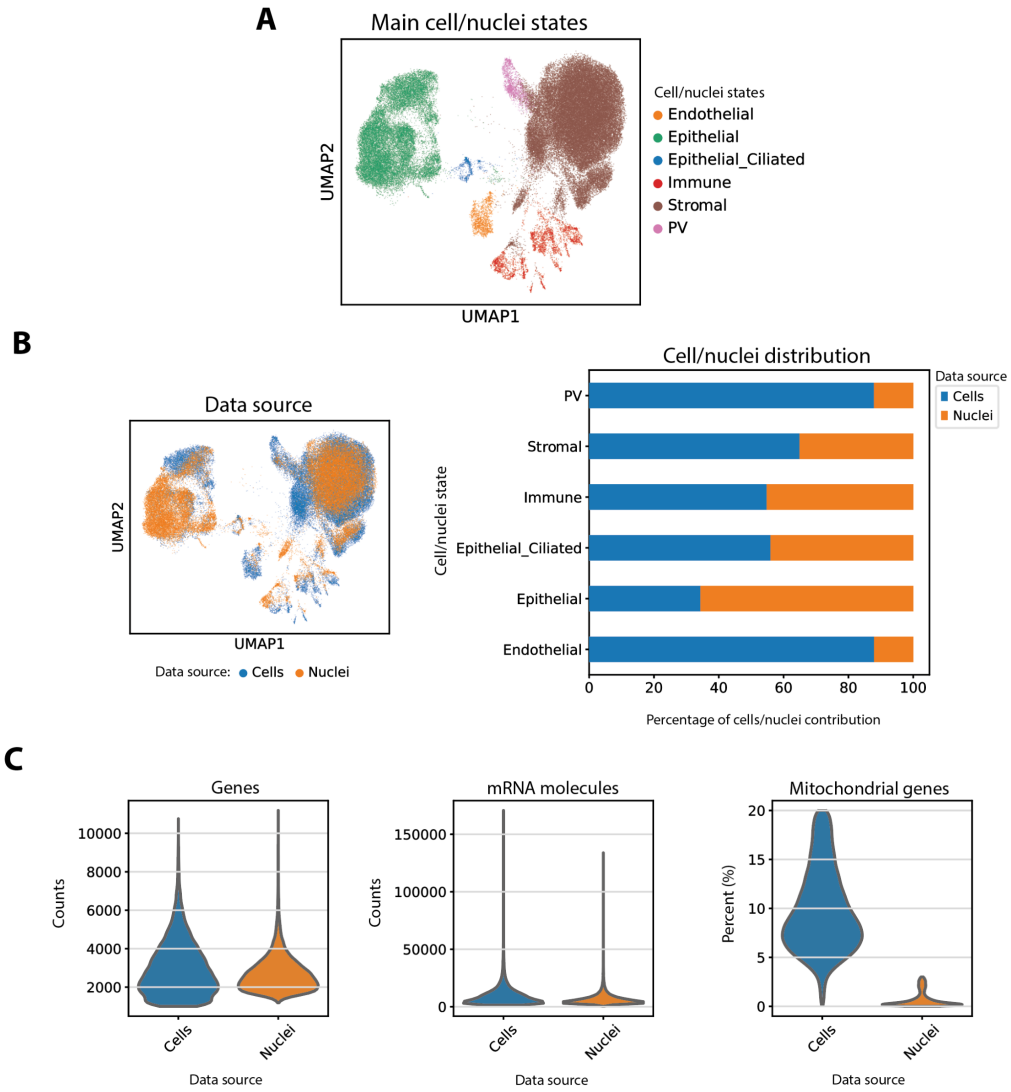


Figure 4.1: Unified map of scRNAseq and snucRNAseq data for 5 donors. Shown are UMAP projections of the integrated scRNAseq data (cells) and snucRNAseq data (nuclei), coloured by the main cells/nuclei states identified (A) and the data source (B). A bar plot visualisation of the percentual contribution of cells and nuclei to each cells/nuclei state identified is also shown. The violin plots in (C) show the data distribution for number of genes, number of mRNA molecules and percentage of mitochondrial gene expression on a per cell/nucleus basis for each of the data sources. *Abbreviations*: PV, perivascular; snucRNAseq, single-nucleus RNA-sequencing; scRNAseq, single-cell RNA-sequencing; UMAP, uniform manifold approximation and projection.

To check the averaged data for all the donor samples above was not confounded by an outlier sample, I next considered the data for each donor separately (Figure 4.2). It confirmed the above results about detecting more epithelial cells in the nuclei data when compared to cells data and also the low numbers of endothelial and PV cells in the nuclei data across all donors. However, heterogeneity was observed across the donors: e.g. donors FX1146 and FX1125 had very few cells annotated as epithelial cells in samples processed for scRNAseq. For both of these donors a large proportion of cells expressing epithelial markers was removed during the QC steps as they were regarded as low QC for further downstream analyses (data not shown).

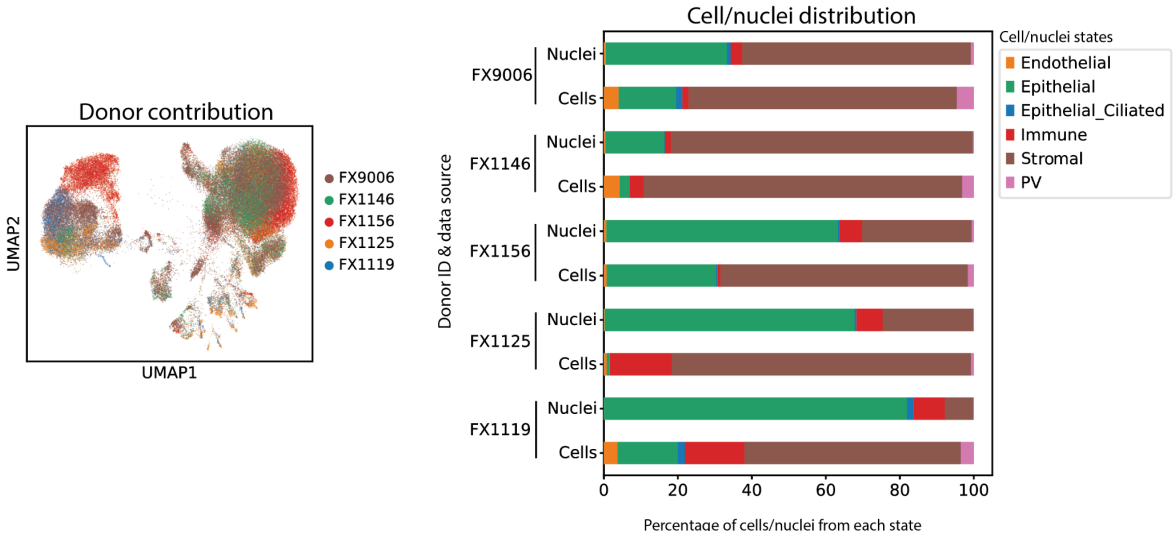


Figure 4.2: Sample-specific distribution of main cells/nuclei states. Shown is an UMAP projection of the integrated scRNAseq data (cells) and snucRNAseq data (nuclei), coloured by donor ID (left). To the right is a bar plot visualisation of the percentual contribution of cells and nuclei to each cells/nuclei state identified on a donor basis either in the cells or nuclei data. *Abbreviations*: PV, perivascular; snucRNAseq, single-nucleus RNA-sequencing; scRNAseq, single-cell RNA-sequencing; UMAP, uniform manifold approximation and projection.

4.4.2 Integration of scRNAseq and snucRNAseq data for all datasets

Understanding the features of both the scRNAseq and snucRNAseq datasets, we next integrated the three scRNAseq datasets described in Chapter 3 (in this chapter referred to as Garcia-Alonso_Cells, Mareckova_Cells and Wang_Cells) together with snucRNAseq data for the 66 samples described above in Table 4.3 (here referred to as Mareckova_Nuclei). In this chapter, I present the integrated map of 600,056 high QC single cells and nuclei from 90 donors (Figure 4.3), consisting of 254,650 cells from 29 donors and 345,406 nuclei from 66 donors (see Table 4.4). The total number of individual donors profiled stands at 90 due to the overlap of having both scRNAseq and snucRNAseq data for 5 donors as described above. In the following sections, I describe the findings obtained from analysing the main cells/nuclei states identified before zooming-in on the epithelial and mesenchymal lineages, and describe what effect exogenous hormones and endometriosis have on the endometrium's transcriptomic profile.

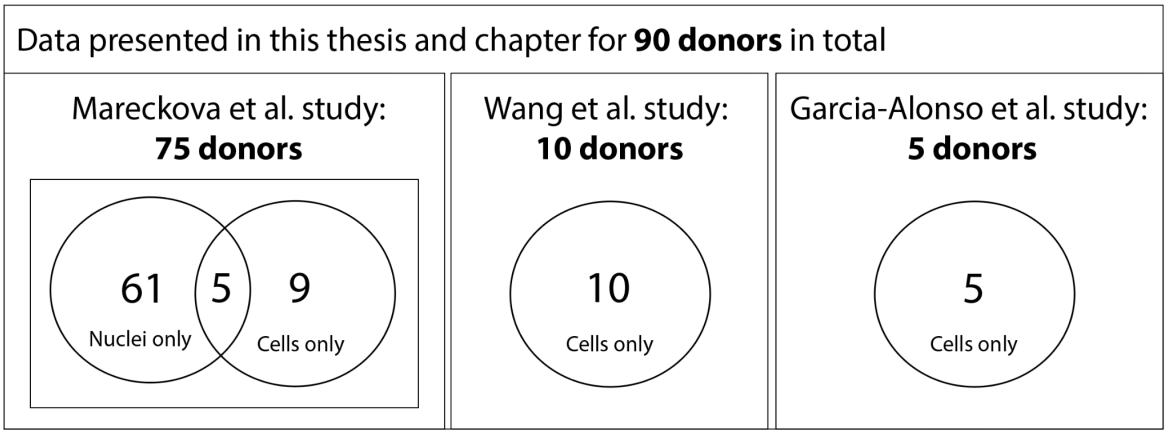


Figure 4.3: Overview of donor samples analysed. Number of samples analysed per each study is shown and whether these were used for scRNAseq (Cells only) or snucRNAseq (Nuclei only) or both as shown for the Mareckova *et al.* study.

Table 4.4: Overview of samples used for integration in Chapter 4.

	Controls	Endoms*
Data source		
scRNAseq (cells)	18	6
snucRNAseq (nuclei)	17	44
scRNAseq & snucRNAseq	2	3
Menstrual cycle group		
Menstrual	1	4
Proliferative	10	15
Secretory	20	21
Hormones	5	12
Unknown	1	1

Superscripts:
* Endoms denotes Endometriosis.

4.4.2.1 Main cells/nuclei states identified

The broad annotation of the integrated cells and nuclei was performed the same way as described in Chapter 3 and for the 5 donors above. Upon identification, the Fib_C7 and uSMCs populations were removed from further analyses as these were present only in the samples from the organ donor programme as described in Chapter 3. The main cells/nuclei states identified included endothelial, immune (myeloid and lymphoid subgroups), stromal, PV and epithelial cells/nuclei (Figure 4.4). Ciliated epithelial cells/nuclei formed a distinct cluster as did the epithelial MUC5B cells/nuclei. Both the ciliated and MUC5B populations were assigned their identity based on the marker genes described in Chapter 3. The detection of a large number of MUC5B cells/nuclei (~27,000) was in contrast to the low numbers (~270)

detected in the single-cell map of the endometrium in Chapter 3. As shown in Figure 4.3, the MUC5B population was predominantly found in the nuclei dataset (~98% of all cells/nuclei assigned as MUC5B) and is described in more detail in section 4.4.2.2. With regards to the integration and overlap for endothelial, PV, immune, and ciliated epithelial cells/nuclei there seemed to be a good agreement between cells and nuclei assigned the same identity. However, the mixing of cells and nuclei was less successful in the epithelial and stromal lineages (Figure 4.4B) and is discussed in more detail in the next sections. Lastly, when it comes to the proportions of PV and endothelial cells/nuclei, the recovery of these was lower in the nuclei dataset, confirming the findings from the initial analysis of the 5 donor samples described in section 4.4.1.

Next, I further explored the contribution of each of the datasets in the harmonised map of scRNAseq and snucRNAseq data (Figure 4.4C). I observed that even though only 5 donors were included in the Garcia-Alonso_Cells dataset, a large proportion of the PV and endothelial cells/nuclei were contributed by this dataset (~40% and ~35%, respectively) when compared to the rest of the datasets (Figure 4.4C). The Wang_Cells dataset on the other hand was characterised by a large number of ciliated epithelial cells, with this dataset contributing ~50% of all cells/nuclei assigned this identity in the integrated map. Lastly, a signature of the Mareckova_Nuclei dataset was the detection of the majority of MUC5B cells with ~98% of MUC5B cells/nuclei found in this dataset.

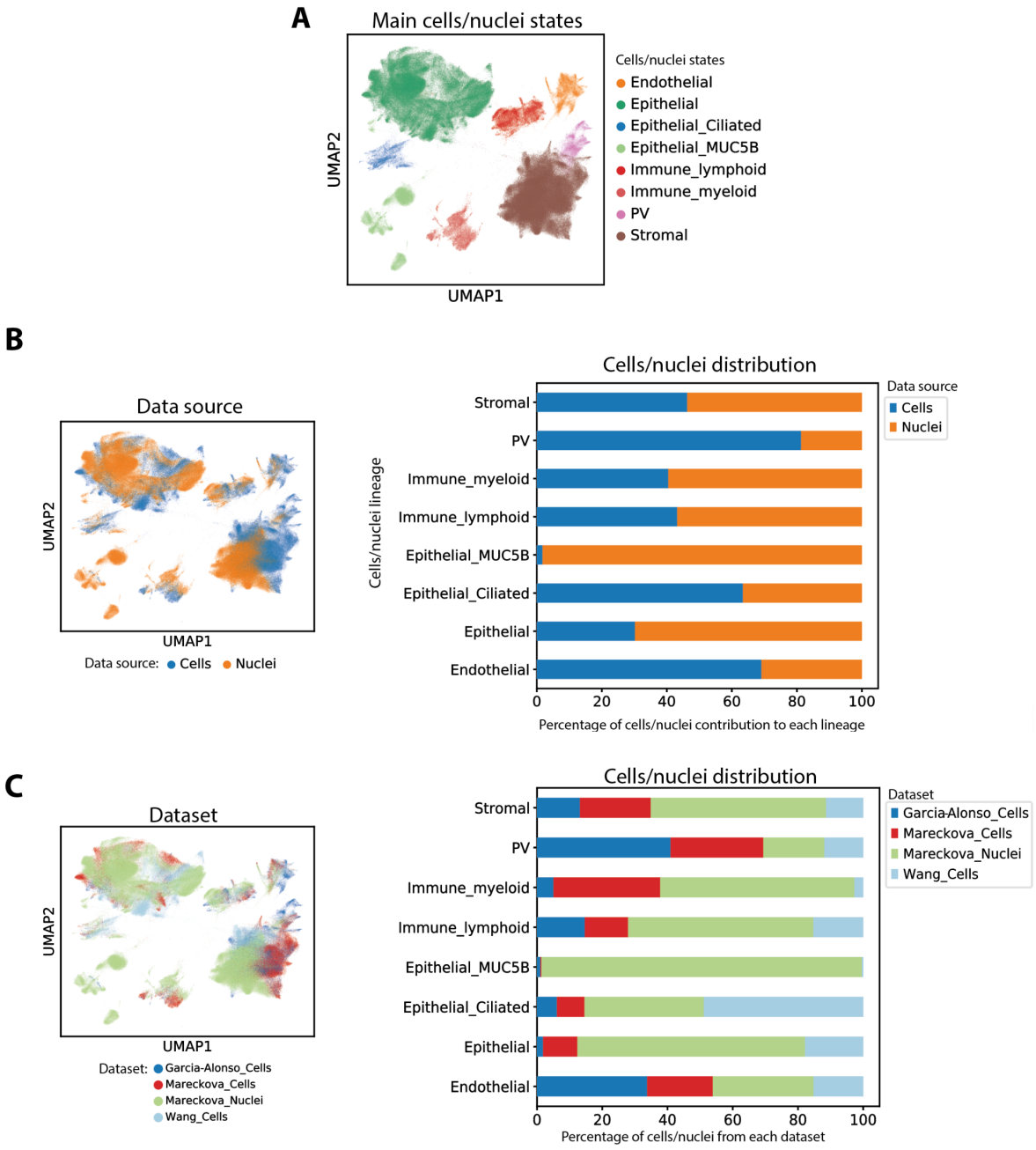


Figure 4.4: Unified map of scRNAseq and snucRNAseq data for 90 donors. Shown are UMAP projections of the integrated scRNAseq data (cells) and snucRNAseq data (nuclei), coloured by the main cell/nuclei states identified (A), the data source (B) and the dataset of origin (C). A bar plot visualisation of the percentual contribution of cells and nuclei to each lineage identified is shown in (B) while in (C) it is the percentual contribution of each dataset to the lineages identified. *Abbreviations*: PV, perivascular; snucRNAseq, single-nucleus RNA-sequencing; scRNAseq, single-cell RNA-sequencing; UMAP, uniform manifold approximation and projection.

Furthermore, I wanted to explore the gene expression patterns of scRNAseq and snucRNAseq data across the main cells/nuclei states identified. Overall, I observed a good resemblance of the average gene expression for the states identified when comparing these states between cells and nuclei data (Figure 4.5). Less similarity was observed for immune_myeloid cells and the epithelial lineage (including the ciliated epithelial cells). Of note is that all of the snucRNAseq data were denoised using *DecontX* (see Methods) to remove ambient RNA contamination prior to the nuclei and cells data integration. As *DecontX* adjusts the raw count matrix based on the model it fits to remove the ambient RNA, these analyses need to be repeated and adjusted to take the adjustment into account.

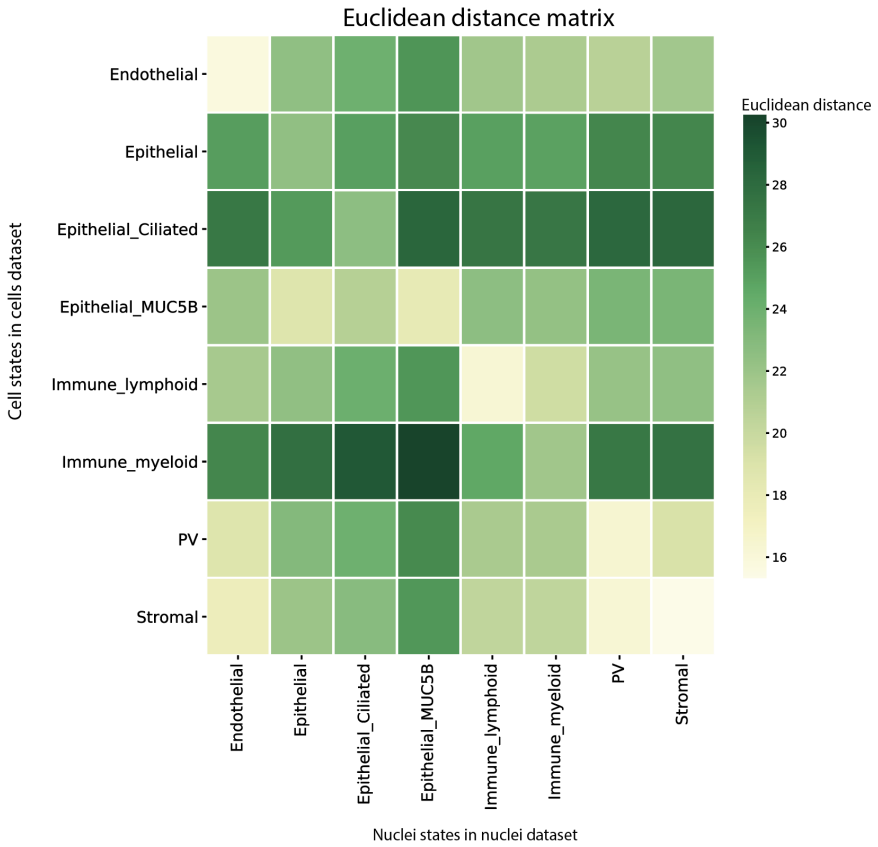


Figure 4.5: Cells-nuclei average gene expression similarity matrix. Shown is a correlation matrix of Euclidean distances between cells and nuclei, considering the main states identified in each dataset. Values for each state were obtained by averaging the gene expression (restricted to the 3,000 highly variable genes) of all cells and all nuclei assigned to that particular state. Pairwise Euclidean distance was calculated between these average gene expression values and is visualised. The shorter the distance, the more similarity is observed. *Abbreviations:* PV, perivascular.

4.4.2.2 The epithelial cell lineage

A total of 232,150 high QC cells/nuclei (26% cells, 74% nuclei) belonging to the epithelial lineage are shown in Figure 4.6 with the main cells/nuclei states annotated and marker gene expression shown in Figure 4.7. The next sections cover the identified cells/nuclei states in more detail and how these compare to those annotated in Chapter 3. Moreover, the trends observed across the menstrual cycle, under exogenous hormones and in endometriosis are described.

4.2.2.2.1 Epithelial cells/nuclei states identified

The ciliated group

Similar to those described in Chapter 3, three subtypes of ciliated cells/nuclei were identified: pre-ciliated, ciliated and ciliated_LGR5 states. Overall, the percentual contribution of nuclei to the ciliated and ciliated_LGR5 states (~20-30%) was lower than expected given the total number of cells and nuclei analysed (Figure 4.6B). Considering the contribution of each dataset to these cells/nuclei states showed that the majority of the ciliated and ciliated_LGR5 states (> 55%) came from the Wang_Cells dataset profiling secretory phase samples around the time of the window of implantation (Figure 4.6C). The pre-ciliated state was made of ~70% nuclei. The states of the ciliated group could be identified using the marker genes described in Chapter 3, with the exception of *NEK2* (pre-ciliated marker reported in the Chapter 3) not detected in this cells/nuclei integration (Figure 4.7).

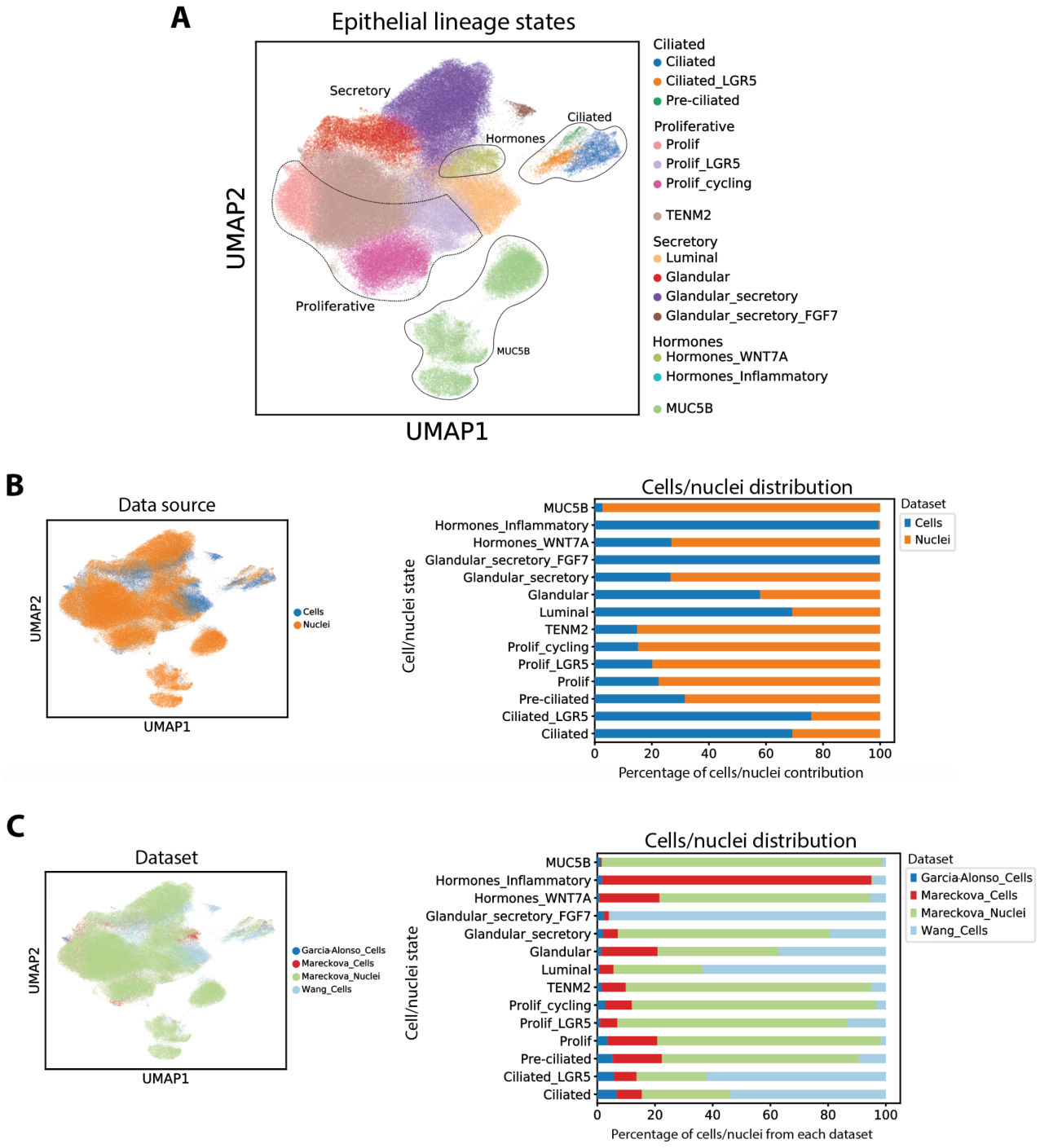


Figure 4.6: Epithelial lineage cells/nuclei states identified. Shown are UMAP projections of the scRNAseq and snucRNAseq datasets integrated, coloured by epithelial cells/nuclei states identified (A), the different data sources (B), and the datasets integrated (C). In (B, right) the contribution of cells and nuclei to each epithelial cells/nuclei state identified is shown and in (C, right) the contribution of each dataset is quantified. *Abbreviations*: Prolif, proliferative; scRNAseq, single-cell RNA-sequencing; snucRNAseq, single-nucleus RNA-sequencing; UMAP, uniform manifold approximation and projection.

The proliferative group

Apart from the defined proliferative and proliferative_LGR5 cells/nuclei states described in Chapter 3, here I further defined the proliferative_cycling state with nuclei making > 80% of this cells/nuclei state (Figure 4.6B). Identified marker genes through TF-IDF included *CENPP*, *MND1* and *MELK* (Figure 4.7B).

The TENM2 population

Overall, this population was abundant in the snucRNAseq data, with nuclei contributing over 80% (Figure 4.6B). The marker gene expression; however, slightly differed in the snucRNAseq data, with lower expression of the *OPRK1* marker gene in nuclei when compared to cells (Figure 4.7C). Moreover, in Chapter 3, the TENM2 population was uniquely identified in secretory phase samples, while in this larger dataset the population was detected in both the proliferative and secretory phases (described in more detail in section 4.2.2.2.2). Therefore, in this chapter it is not grouped under the secretory group. Based on marker gene expression, the TENM2 population shares expression of some marker genes with the proliferative and proliferative_cycling cells/nuclei states, suggesting it could be a transitioning cell state, transitioning towards the glandular state with which it also shares some markers (Figure 4.7).

The secretory group

Apart from the glandular_secretory_FGF7 cell state, the luminal, glandular and glandular_secretory states could also be identified in the nuclei data (Figure 4.6B). The glandular_secretory_FGF7 cell state came from the cells datasets only, with ~95% from the Wang dataset (Figure 4.6C). Thus, this cell state was not kept for

further analyses. The majority of the luminal and glandular cells/nuclei states were of scRNAseq origin (~60-70%) and again a Wang dataset-specific feature: ~60% of all luminal and ~40% of all glandular cells/nuclei were contributed by this dataset (Figure 4.6C). In contrast, the glandular_secretory cells/nuclei state was found in both the cells and nuclei datasets, with nuclei contributing ~70%. All cells/nuclei states in the secretory group were identified based on marker gene expression, with nuclei data having lower expression of the markers defined in Chapter 3 (Figure 4.7B & C). Interestingly, the glandular_secretory state highly expressed *ESR2* in the nuclei dataset but not in the cells' dataset (Figure 4.7B & C).

The MUC5B population

The MUC5B population was mostly detected in the nuclei dataset (i.e. Mareckova_Nuclei) with only ~2% contribution from the cells datasets (Figure 4.6B & C). I observed heterogeneity within this population, with multiple small clusters expressing MUC5B identified (Figure 4.6A). Donor-wise, the number of cells/nuclei detected also varied greatly. Most donors contributed only a limited number of cells/nuclei to this MUC5B population (between 1 to 100 cells/nuclei per donor), while 6 donors contributed ~90% of all cells/nuclei of this identity (data not shown). The marker gene expression also varied when compared to the markers defined in Chapter 3. The new MUC5B specific markers identified included *PTPRD*, *TGFBR3* and *DKK2* (Figure 4.7).

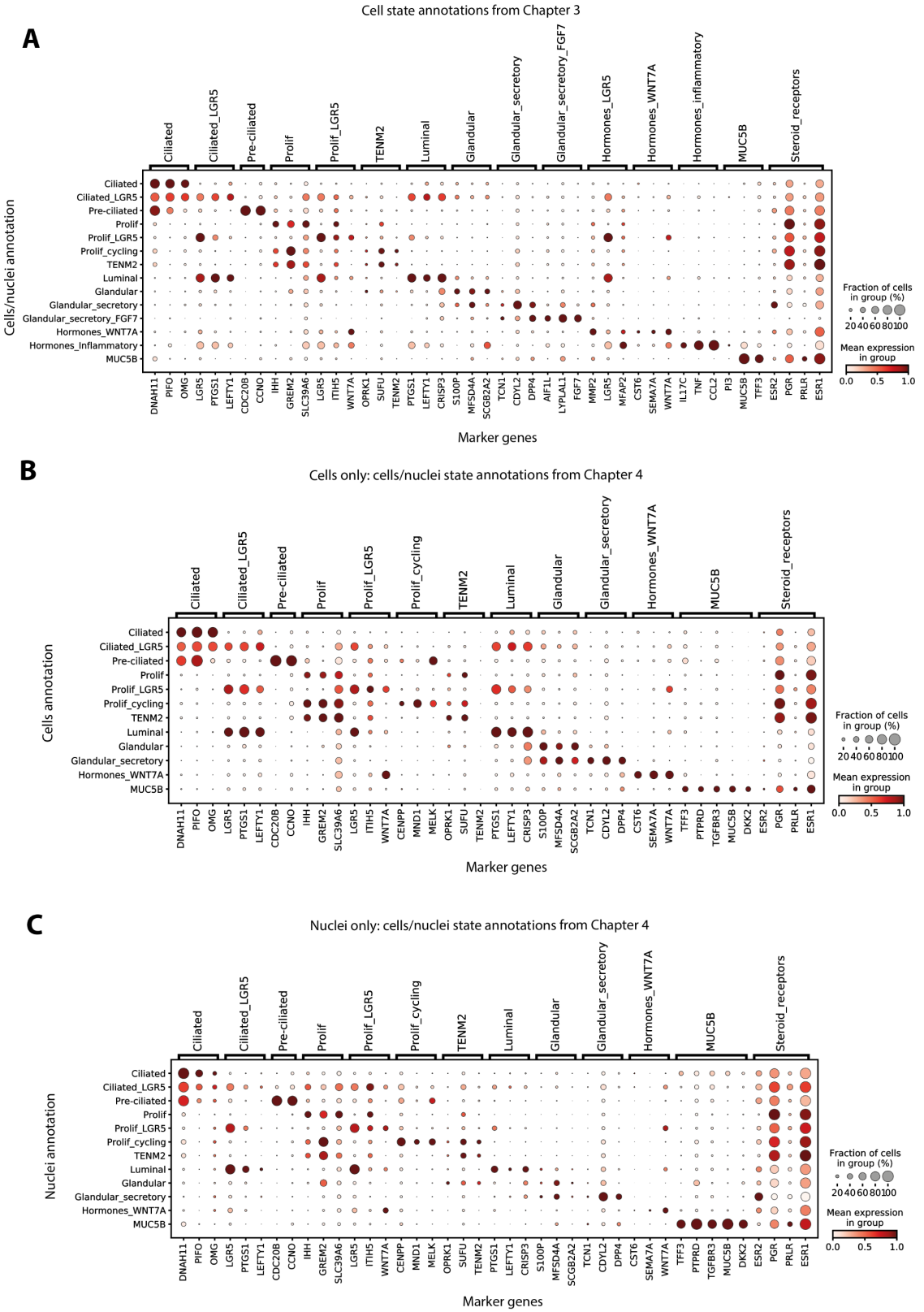


Figure 4.7: Epithelial lineage cells/nuclei states marker gene expression. The dotplots show the variance-scaled, log-transformed expression of marker genes (x-axis) characteristic for each of the cells/nuclei states (y-axis). Expression of steroid hormones is also shown. In (A) the x-axes show the annotations and markers based on the cells' analysis in Chapter 3 for the integrated cells and nuclei manifold, while the expression of the same marker genes and further newly identified marker genes is shown in (B) for cells only and in (C) for nuclei only. *Abbreviations:* Prolif, proliferative.

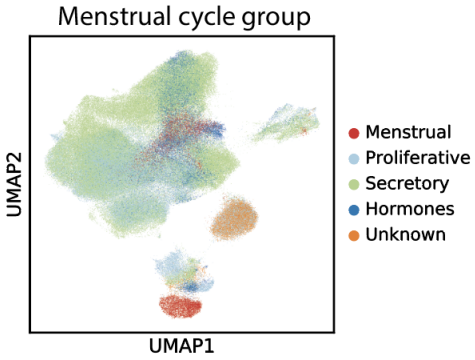
The hormones group

In contrast to Chapter 3, the hormones_LGR5 state could not be identified in the integrated dataset based on its marker gene expression (Figure 4.7A), only the existence of the hormones_WNT7A state was confirmed in the nuclei data (Figure 4.6B). In addition, the hormones_inflammatory state could not be identified in the nuclei data. Based on the marker gene expression, I could only identify this cell state in the Mareckova_Cells dataset and it remained a one donor-specific cluster. This cell cluster was thus excluded from further analyses.

4.2.2.2.2 Epithelial cells/nuclei states across the menstrual cycle

Next, I explored how the above identified cells/nuclei states vary across the menstrual cycle. Five main groups are considered: menstrual, proliferative, secretory, and unknown phases of natural menstrual cycles plus samples from donors taking exogenous hormones (Figure 4.8). The number of donors assigned to each group and the number of cells/nuclei analysed, and the percentual contribution of each group to the total number of cells/nuclei analysed is summarised in Figure 4.8B. To describe the general trends, all samples are considered, irrespective of the controls vs endometriosis cases status.

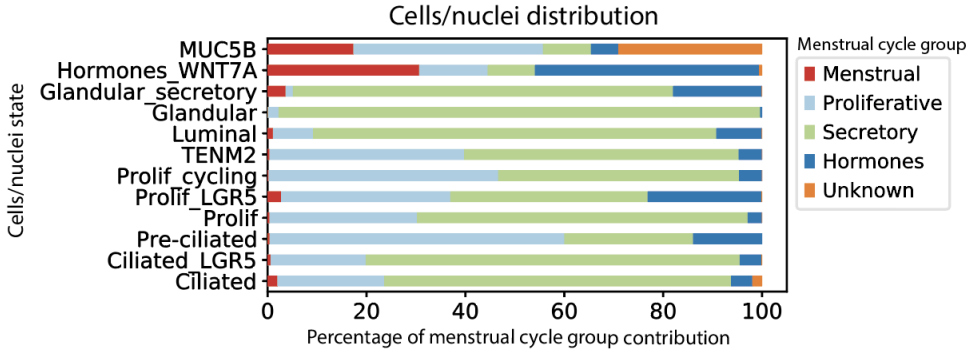
A



B

Menstrual cycle group	Donor number	Cells/nuclei number	% of total cells/nuclei
Menstrual	5	9,467	4%
Proliferative	25	60,040	26%
Secretory	41	130,644	57%
Hormones	17	22,995	10%
Unknown	2	7,921	3%
Total	90	231,067	100%

C



D

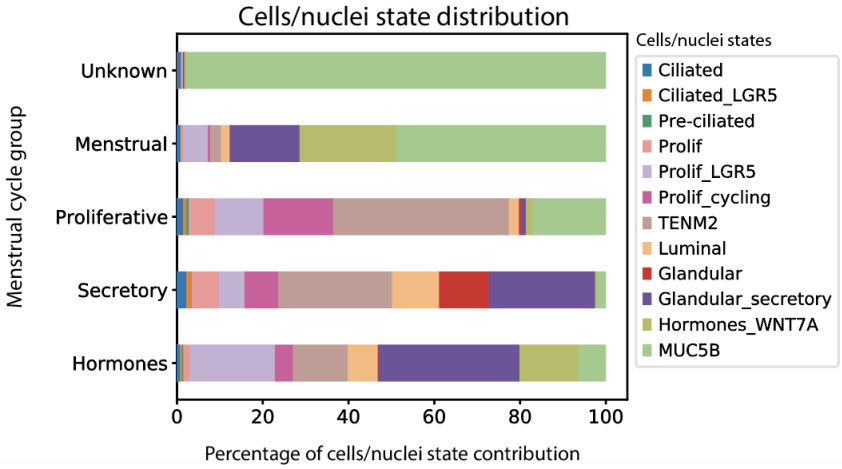


Figure 4.8: Epithelial cells/nuclei across the menstrual cycle. Shown is an UMAP projection of the scRNAseq and snucRNAseq datasets integrated, coloured by menstrual cycle group assignment (A). A summary table of the number of donors and cells/nuclei profiled per group is shown in (B). In (C) the contribution of each menstrual cycle group to the cells/nuclei states is shown, while in (D) the distribution of cells/nuclei states per menstrual cycle group is shown. *Abbreviations*: Prolif, proliferative; scRNAseq, single-cell RNA-sequencing; snucRNAseq, single-nucleus RNA-sequencing; UMAP, uniform manifold approximation and projection.

Menstrual phase

Looking at the menstrual phase, 5 samples were profiled, contributing 4% to the total number of cells/nuclei analysed (Figure 4.8B). Cells and nuclei from the menstrual phase predominantly clustered with those assigned the hormones_WNT7A identity (Figure 4.8 A & C) but one donor sample was mostly composed of MUC5B nuclei and one of proliferative_LGR5 nuclei skewing the proportions observed. The MUC5B nuclei from the menstrual phase formed their own cluster within the MUC5B assigned cells/nuclei state (Figure 4.8A). A smaller number of cells/nuclei belonged to the rest of the identified cells/nuclei states with only a limited number belonging to the glandular state while the glandular_secretory state was more abundant (Figure 4.8C & D). Variation in the data is not surprising given the endometrium sheds and heals simultaneously in a piecemeal fashion, with dramatic histological changes between day 1 of bleeding to the end of the menstrual phase when it has been repaired and bleeding stops.

Proliferative and secretory phases

Unlike in Chapter 3, the distinction between cells/nuclei states belonging to either the proliferative or secretory phases (26% and 57% of total cells/nuclei analysed, respectively) is not as clear-cut. There appears to be more of a continuum (Figure 4.8C), but some phase-specific features remain. E.g. the proliferative phase is characterised by the presence of the proliferative, proliferative_LGR5, proliferative_cycling, TENM2 and MUC5B states in larger proportions when compared to the secretory phase (Figure 4.8C & D). On the other hand, the secretory phase is defined by a larger number of luminal, glandular and glandular_secretory cells/nuclei. With regards to the ciliated group of cells, the pre-

ciliated cells are a feature of the proliferative phase while the ciliated and ciliated_LGR5 states were more abundant in the secretory phase samples. NB, the Wang_Cells dataset skewed these numbers and proportions as discussed above.

The TENM2 population briefly described in the previous section was confided to the secretory phase in Chapter 3, but it appeared that when a larger number of samples is profiled, it is the proliferative phase in which the total number of TENM2 cells/nuclei peak. Unfortunately, for most of the samples I did not have a more fine annotation of the proliferative phase (i.e. early/late) and also the secretory phase (i.e. early/mid/late) which would allow more detailed characterisation. However, based on the samples that such finer annotation was available for, the data suggest that the TENM2 population increases in its numbers towards the end of the proliferative phase and early secretory phase with numbers decreasing towards later stages of the secretory phase (data not shown).

Unknown phase

It was not possible to assign the menstrual phase for two donor samples due to poor histology of the samples upon sectioning and no data available for self-reported days of menstrual cycle. These two samples were mostly made of the MUC5B cells/nuclei state (Figure 4.8) and thus it was also not possible to assign the menstrual stage based on the transcriptomic profiles observed. These samples therefore remain assigned to the unknown group until the clinical metadata is further reviewed by our research nurses for more details.

Hormones

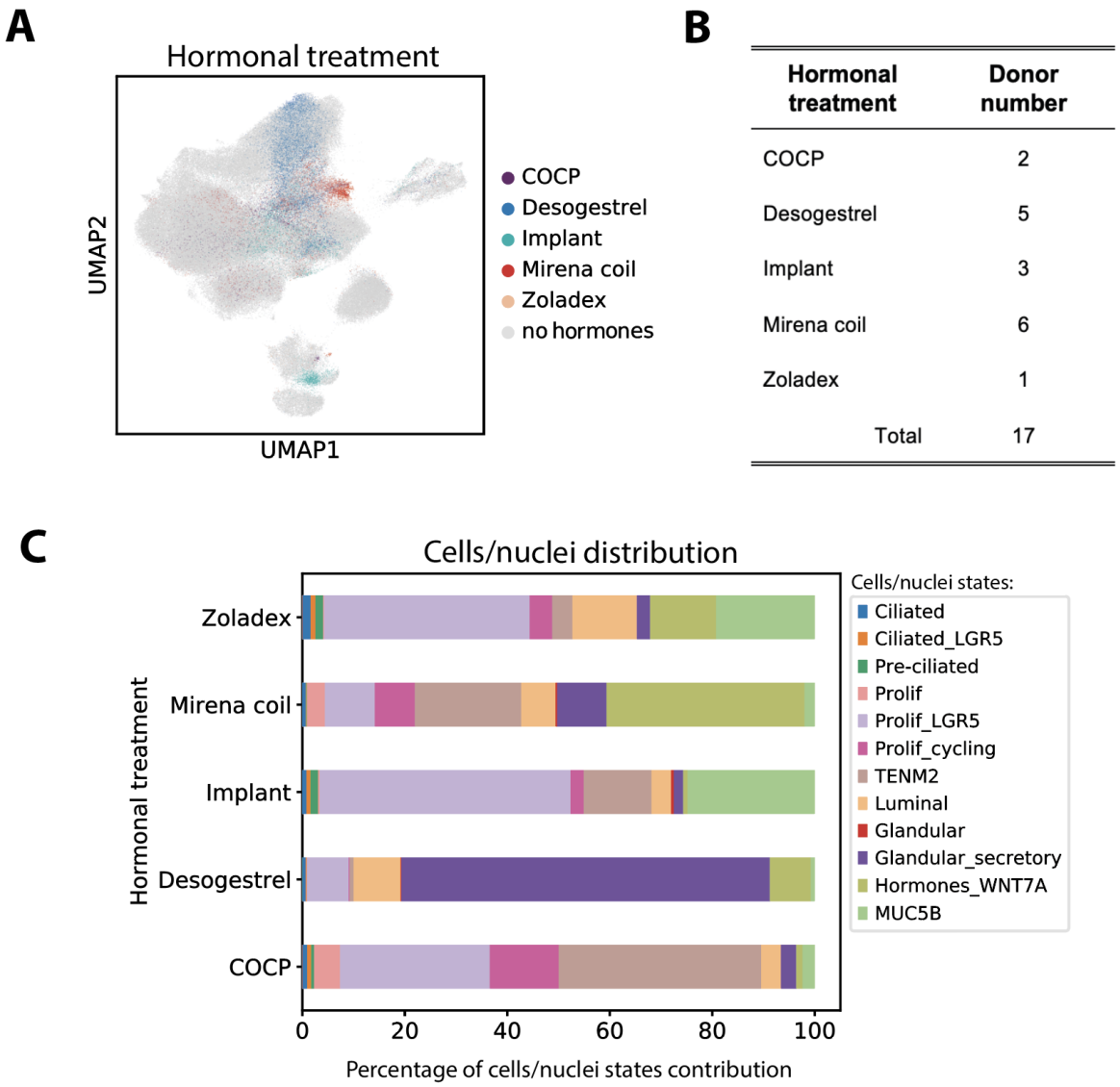
Altogether, 17 samples were collected from donors taking exogenous hormones, contributing ~10% of cells/nuclei to the total analysed. Yet, these donors were taking different forms of exogenous hormones and are thus discussed in more detail in a separate section below.

4.2.2.2.3 Epithelial cells/nuclei states under exogenous hormones

Overall, the 17 donors taking exogenous hormones at sample collection were taking 5 different types of treatment. These included: (i) 2 donors taking the combined oral contraceptive pill (COCP) containing both oestrogen and progestin, (ii) 5 donors taking progestins only orally, brand name Desogestrel, (iii) 3 donors with a progestin implant with a continuous release of the hormone, (iv) 6 donors with a fitted progestin-releasing IUD, brand name Mirena coil, and (v) 1 donor taking GnRH agonist, brand name Zoladex (Figure 4.9A & B). The differences in the cellular composition of the samples between the different treatments (Figure 4.9A & C) is discussed below.

Starting with the COCP treatment, the sample composition resembled proliferative phase endometrium with a high proportion of proliferative, proliferative_LGR5 and TENM2 cells/nuclei states detected (Figure 4.9C). The Desogestrel treatment (progestin only), unlike in Chapter 3 was not characterised by the hormones_inflammatory cell state. As described above, this cell state was one donor-specific and thus excluded from further analyses. Instead, the endometrium of donors taking Desogestrel mostly resembled the secretory phase endometrium with a high proportion of glandular_secretory cells/nuclei (70% of total). Presence

of this secretory population was also observed in Chapter 3, together with the hormones_WNT7A cells/nuclei which were also seen here (Figure 4.9C).



Hormonal treatment	Donor number
COCP	2
Desogestrel	5
Implant	3
Mirena coil	6
Zoladex	1
Total	17

Figure 4.9: Epithelial cells/nuclei in donors taking exogenous hormones. Shown is an UMAP projection of the scRNAseq and snucRNAseq datasets integrated, coloured by type of hormonal treatment (A). A summary table of the number of donors and cells/nuclei profiled per treatment is shown in (B). In (C) the distribution of cells/nuclei states per hormonal treatment is shown. *Abbreviations*: COCP, combined oral contraceptive pill; Prolif, proliferative; scRNAseq, single-cell RNA-sequencing; snucRNAseq, single-nucleus RNA-sequencing; UMAP, uniform manifold approximation and projection.

The progestin-releasing implant was characterised by a high proportion of proliferative_LGR5 cells/nuclei (40%) and a high proportion of MUC5B cells/nuclei (~25%). However, this was due to one donor where the sample consisted of 95% MUC5B cells/nuclei. In contrast, the progestin-releasing IUD Mirena coil was defined by a high proportion of the hormones_WNT7A cells/nuclei state (~37%) which is in agreement with what was observed for the sample described in Chapter 3, making this a specific feature of this treatment. In addition, a higher proportion of cells/nuclei states present in the proliferative phase endometrium were observed with some luminal and glandular_secretory states more characteristic of the secretory phase endometrium also detected (Figure 4.9C).

With regards to the Zoladex treatment, only one donor sample was profiled (Figure 4.9B). The data suggest that under this treatment the endometrium is made of a high proportion of the proliferative_LGR5 cells/nuclei state as well as the luminal, hormones_WNT7A and MUC5B states. This is in line with the treatment causing endometrial atrophy, manifested as a thin *functionalis* layer lined with luminal cells[77].

Lastly, of note is that in most cases where the donors were taking hormonal treatment, the MUC5B population was detected on a per donor basis (data not shown). With regards to the ciliated group of cells/nuclei, lower proportion of all types of ciliated cells/nuclei was seen in samples from donors taking Desogestrel or having the IUD Mirena coil fitted while higher proportions were observed under the Zoladex, Implant and COCP treatments.

4.2.2.2.4 Epithelial cells/nuclei states in endometriosis and controls

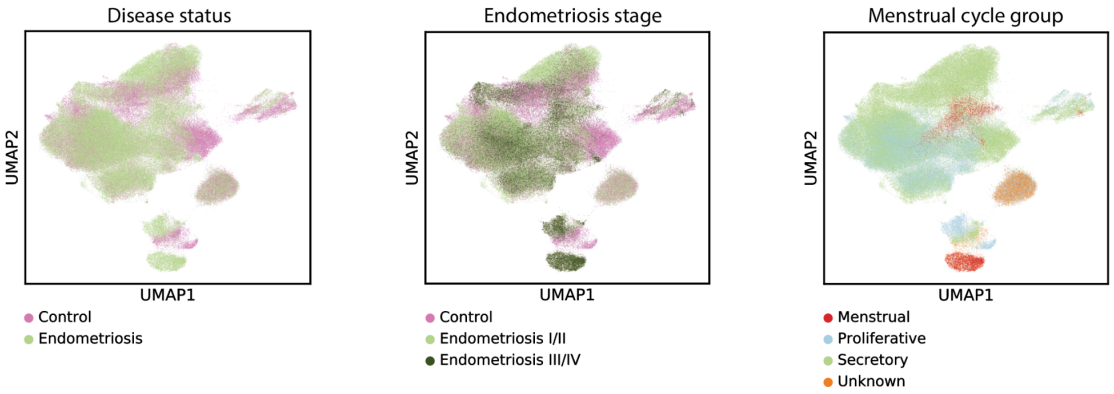
In this section, only samples from donors not taking exogenous hormones and from the proliferative and secretory phases (*n* = 66) are considered. This is due to the low number of samples and cells/nuclei from the menstrual phase and unknown group within the control and endometriosis groups (Figure 4.10B). Cells/nuclei from controls (i.e. women without endometriosis) (*n* = 32) and endometriosis cases (*n* = 41) contributed 38% and 62% of the total cells/nuclei analysed from donors not on hormones, respectively. In addition, the different endometriosis stages I/II and III/IV were all merged under one 'endometriosis' group due to a higher proportion of endometriosis I/II stage samples being from the secretory phase, while endometriosis III/IV stage donors had a higher proportion of donors in the proliferative phase. For this reason, only an UMAP visualisation of the different stages is provided in Figure 4.10A. Taking these steps ensured the number of proliferative and secretory phase samples was balanced between the control and endometriosis groups (Figure 4.10B, green rectangle).

In Chapter 3, the preliminary comparison of controls and cases suggested no differences in the proliferative phase and highlighted some differences during the secretory phase, such as higher proportion of luminal, ciliated_LGR5 and glandular_secretory_FGF7 cell states in controls when compared to endometriosis cases. Looking at a larger number of samples in this chapter, I observed that in the proliferative phase, there is a higher proportion of proliferative_cycling and TENM2 cells/nuclei in endometriosis cases when compared to controls (Figure 4.10C). Moreover, the proportion of MUC5B and all types of ciliated cells/nuclei was higher in controls than in cases. These findings were not observed in Chapter 3.

Considering the secretory phase, controls had a higher proportion of ciliated, ciliated_LGR5, luminal and glandular cells/nuclei states when compared to endometriosis cases (as observed in Chapter 3). In addition, I also observed that in endometriosis cases there is a higher proportion of proliferative_cycling, TENM2 and glandular_secretory cells/nuclei states when compared to controls. With regards to the MUC5B population, a higher proportion of these cells/nuclei was observed in controls (Figure 4.10C).

Next, given the observations and biases noted for the Wang_Cells dataset profiling samples from around the time of the window of implantation, I compared only the data generated as part of the Mareckova_Nuclei dataset to verify the above findings (Figure 4.10D). In the proliferative phase, the findings described above are recapitulated - there is a higher proportion of proliferative_cycling, and TENM2 nuclei in endometriosis cases and a reduced proportion of MUC5B nuclei when compared to controls. The differences in proportions of ciliated cells are less obvious. Considering the secretory phase, the differences noted when both cells and nuclei were analysed appear to be absent/less prominent with only a slight increase in the proportion of luminal and glandular_secretory nuclei in controls, while the endometriosis cases are described by a higher proportion of glandular_secretory nuclei and lower proportion of the MUC5B nuclei (Figure 4.10D).

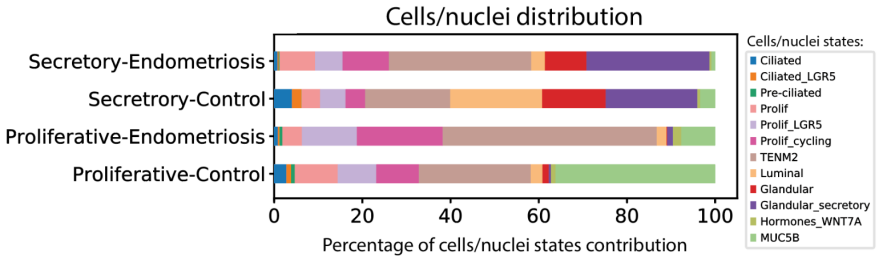
A



B

	Controls		Endometriosis	
	Donor number	% cells/nuclei	Donor number	% cells/nuclei
	32	38%	41	62%
Menstrual phase				
Menstrual	1	1%	4	99%
Proliferative	10	44%	15	56%
Secretory	20	33%	21	67%
Unknown	1	11%	1	89%
Endometriosis stage (rASRM)				
I/II	-	-	24	59%
III/IV	-	-	17	41%

C



D

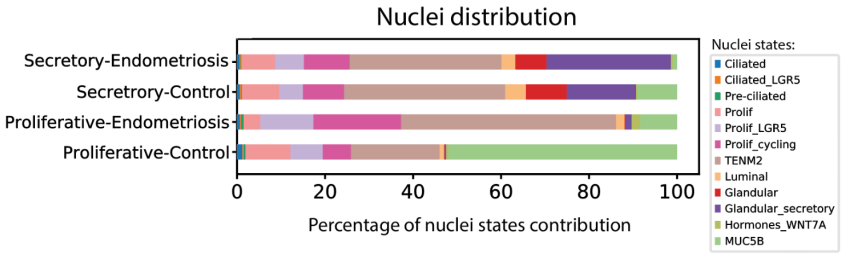


Figure 4.10: Epithelial cells/nuclei in endometriosis and controls. Shown is an UMAP projection of the scRNAseq and snucRNAseq datasets integrated, coloured by disease status, endometriosis stage and menstrual cycle group (A). A summary table of the number of donors and cells/nuclei profiled for controls and endometriosis cases is shown in (B). In (C) the distribution of cells/nuclei states per proliferative and secretory phase for controls and endometriosis cases is shown while in (D) the same is shown for nuclei only. *Abbreviations*: Prolif, proliferative; scRNAseq, single-cell RNA-sequencing; snucRNAseq, single-nucleus RNA-sequencing; UMAP, uniform manifold approximation and projection.

4.4.2.3 The mesenchymal cell lineage

Altogether, 174,970 high QC cells/nuclei (52% cells, 48% nuclei) belonging to the mesenchymal lineage are shown in Figure 4.11 with the main cells/nuclei states annotated and their marker gene expression shown in Figure 4.12. uSMCs were removed from the data upon identification as these came from the organ donor samples only as described in Chapter 3. The next sections cover the identified cells/nuclei states in more detail and how these compared to those annotated in Chapter 3. Moreover, the trends observed across the menstrual cycle, under exogenous hormones and in endometriosis are described.

4.4.2.3.1 Mesenchymal cells/nuclei states identified

The fibroblasts group

As described in Chapter 3, the Fib_C7 population was predominantly detected in the Garcia-Alonso_Cells dataset with organ donor samples, but in this integrated cells/nuclei map also in one sample from the Mareckova_Nuclei dataset (Figure 4.11B & C). This donor sample contributed ~18% of all cells/nuclei assigned the Fib_C7 identity (Figure 4.11B). This population is most likely only detected in samples where deeper layers of the endometrium are sampled, which as the data shows was not the case in the majority of the samples analysed here. Due to this, the Fib_C7 population was excluded from further analyses.

The perivascular group

Similarly, the myometrial mPV_MYH11 cell state was only present in the organ donor samples from the Garcia-Alonso_Cells dataset and was thus not analysed further (Figure 4.11B & C). With regards to the endometrial PV cells/nuclei, low recovery of these was observed in the nuclei data, especially for the ePV_STEAP4

and ePV_AOC3 states with < 5% of all cells/nuclei of these identities being from the Mareckova_Nuclei dataset (Figure 4.11B & C). Moreover, the ePV_STEAP4 population came mostly from the Garcia-Alonso_Cells dataset (> 95%). The only ePV population recapitulated in the nuclei data was the ePV_MMP11 subtype with nuclei making ~30% of all cells/nuclei assigned this identity (Figure 4.11B). The marker genes defining this population were consistent across cells and nuclei and the same as described in Chapter 3 (Figure 4.12).

Stromal cells/nuclei: the non-decidualised group

The eS, eS_cycling, eS_BMP7 and hormones_eS_MMPs populations were described in Chapter 3, but here the hormones_eS_MMPs population could not be detected based on the marker genes defined previously (Figure 4.12A). The eS and eS_cycling states were abundant in the nuclei data, with nuclei making 58% and 55% of all cells/nuclei assigned this annotation, respectively. Same as in Chapter 3, 87% of all cells/nuclei in the eS_BMP7 cluster were from one donor in the Garcia-Alonso_Cells dataset. Only a small number of other donors contributed up to a maximum of 12 cells/nuclei per donor to this cluster. For this reason, this eS_BMP7 state was not considered robust and was not included in further analyses. Apart from the populations defined in Chapter 3, an additional cells/nuclei state was observed and named eS_LAMC3 based on its high expression of *LAMC3.* Further markers identified included *PRSS12* and *CRABP2*, expression of which was shared across the eS and eS_cycling populations (Figure 4.12C).

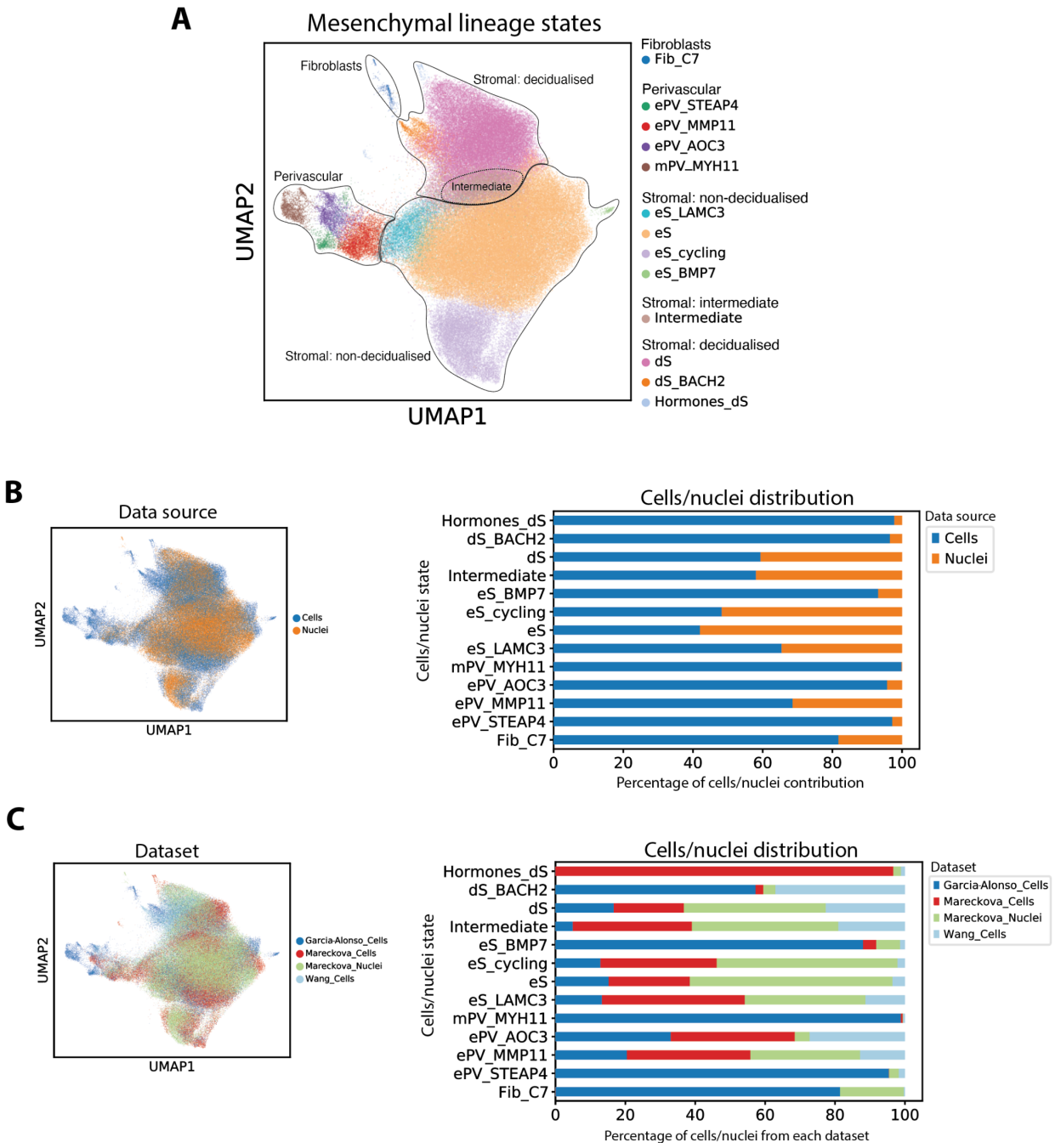


Figure 4.11: Mesenchymal lineage cells/nuclei states identified. Shown are UMAP projections of the scRNAseq and snucRNAseq datasets integrated, coloured by mesenchymal cells/nuclei states identified (A), the different data sources (B), and the datasets integrated (C). In (B, right) the contribution of cells and nuclei to each mesenchymal cells/nuclei state identified is shown and in (C, right) the contribution of each dataset is quantified. *Abbreviations*: dS, decidualised stromal; ePV, endometrial perivascular; eS, endometrial stromal; mPV, myometrial perivascular; scRNAseq, single-cell RNA-sequencing; snucRNAseq, single-nucleus RNA-sequencing; UMAP, uniform manifold approximation and projection.

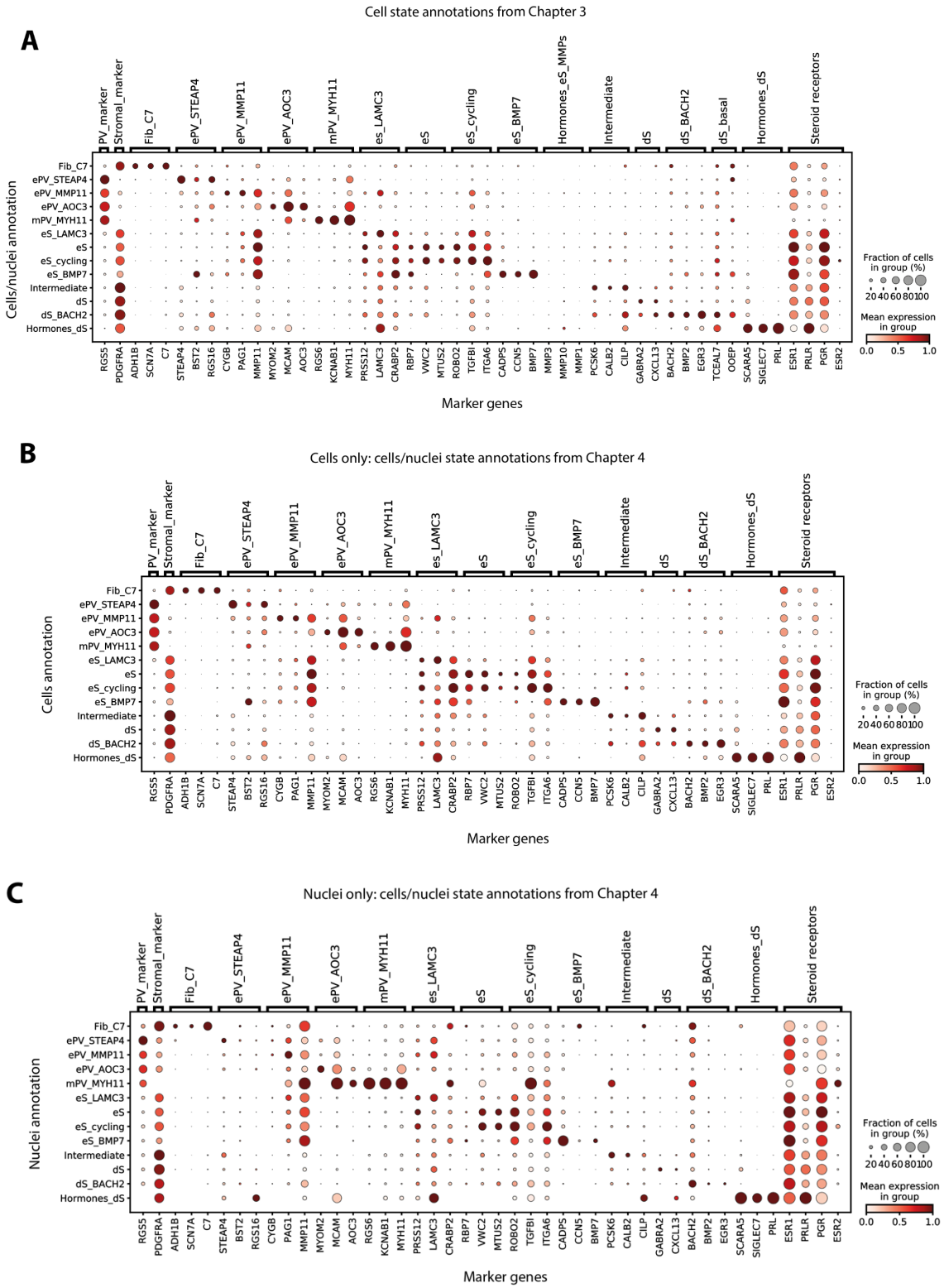


Figure 4.12: Mesenchymal lineage cells/nuclei marker gene expression. The dotplots show the variance-scaled, log-transformed expression of marker genes (x-axis) characteristic for each of the cells/nuclei states (y-axis). Expression of steroid hormones and general PV and stromal markers is also shown. In (A) the x-axes show the annotations and markers based on the cells' analysis in Chapter 3 for the integrated cells and nuclei manifold, while the expression of the same marker genes and further newly identified marker genes are shown in (B) for cells and in (C) for nuclei only. *Abbreviations*: dS, decidualised stromal; ePV, endometrial perivascular; eS, endometrial stromal; mPV, myometrial perivascular.

Stromal cells/nuclei: the intermediate group

The intermediate cell state could also be identified upon scRNAseq and snucRNAseq data integration using the markers described in Chapter 3 (Figure 4.12). The expression of *CILP* (one of the identified markers) was lower in the nuclei data when compared to cells (Figure 4.12B & C).

Stromal cells/nuclei: the decidualised group

In the decidualised group, the dS population was also detected in the nuclei data (contributing ~40%); however, the hormones_dS and dS_BACH2 populations were not recapitulated to the same extent (Figure 4.11). The hormones_dS cell state was observed to be predominantly made of cells from one donor in the Mareckova_Cells dataset, with only 5 cells/nuclei in total contributed by other donors and datasets. This cluster was thus not considered robust and was excluded from further analyses. The dS_BACH2 state was detected in the Garcia-Alonso_Cells and Wang_Cells datasets with nuclei contributing only ~4% to the total number of this population (Figure 4.11B & C). While multiple donors from the Mareckova_Nuclei dataset contributed to this cluster, the highest number of nuclei per donor was 15, suggesting this cells/nuclei state is not fully recapitulated/captured in the nuclei data. In addition, the dS_basal state defined in Chapter 3 could not be identified in the integrated UMAP based on its marker gene expression (Figure 4.12A).

4.4.2.3.2 Mesenchymal cells/nuclei states across the menstrual cycle

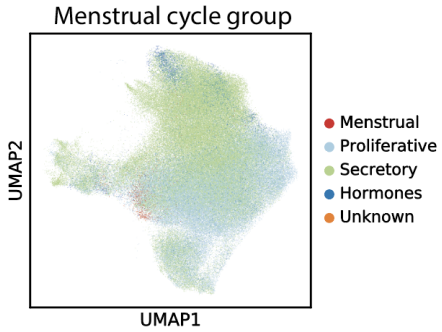
As described above, the recovery of nuclei for the mesenchymal lineage was poor when compared to the epithelial lineage, with low numbers detected especially for the menstrual (1,013 cells/nuclei) and unknown phase (6 cells/nuclei) samples

(Figure 4.13B). These phases are therefore not described in detail, nonetheless their cellular composition is shown in Figure 4.13 after removing the Fib_C7, eS_BMP7 and hormones_dS cells/nuclei. Below, the proliferative and secretory phases as well as samples from donors taking exogenous hormones are discussed in more detail.

Proliferative and secretory phases

As described in Chapter 3, the proliferative phase samples were defined by a high proportion of the eS and eS_cycling states and this finding was recapitulated in this chapter's integrated dataset (Figure 4.13C & D). The proliferative phase samples consisted of ~80% eS cells/nuclei. Moreover, the newly defined eS_LAMC3 state, ePV_AOC3 and ePV_MMP11 populations were also present during the proliferative phase but detected in smaller proportions than in the secretory phase (Figure 4.13D). In the secretory phase samples there was a shift towards a lower proportion of eS and eS_cycling states (making ~35% of all secretory phase cells/nuclei), while the intermediate cell state increased in its numbers as well as the dS cell state, contributing ~12% and 34%, respectively. In Chapter 3, both the intermediate and dS cell populations were reported to be characteristic of the secretory phase samples (Figure 4.13C & D). With regards to the dS_BACH2 and ePV_STEAP4 populations, these were also characteristic of the secretory phase samples; however, as described above, these populations came from the cells' datasets only and thus require further investigation.

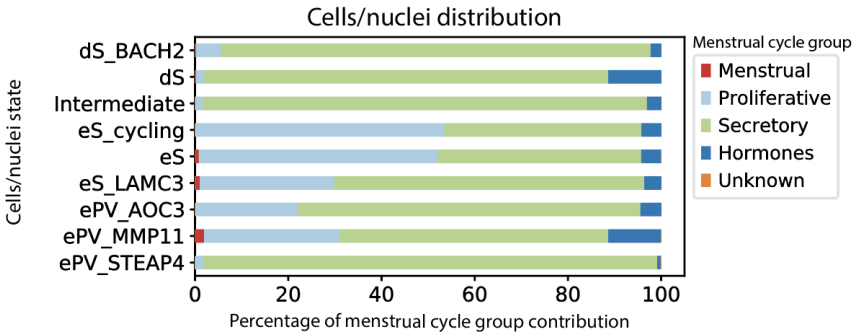
A



B

Menstrual cycle group	Donor number	Cells/nuclei number	% of total cells/nuclei
Menstrual	5	1,013	0.6%
Proliferative	25	57,843	33.5%
Secretory	41	103,558	59.9%
Hormones	17	10,232	5.9%
Unknown	2	6	< 0.0%
Total	90	172,652	100%

C



D

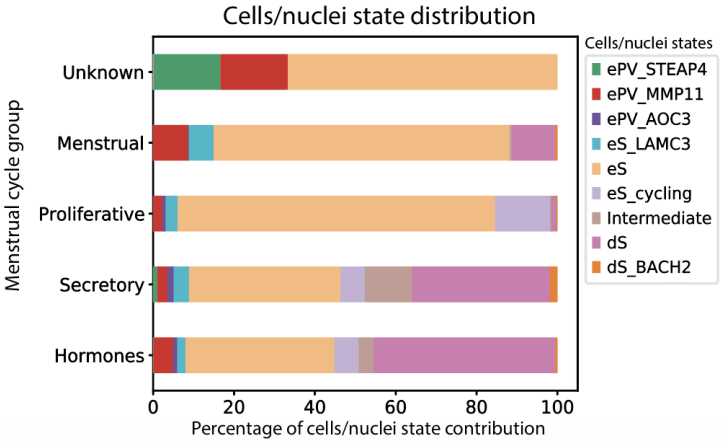


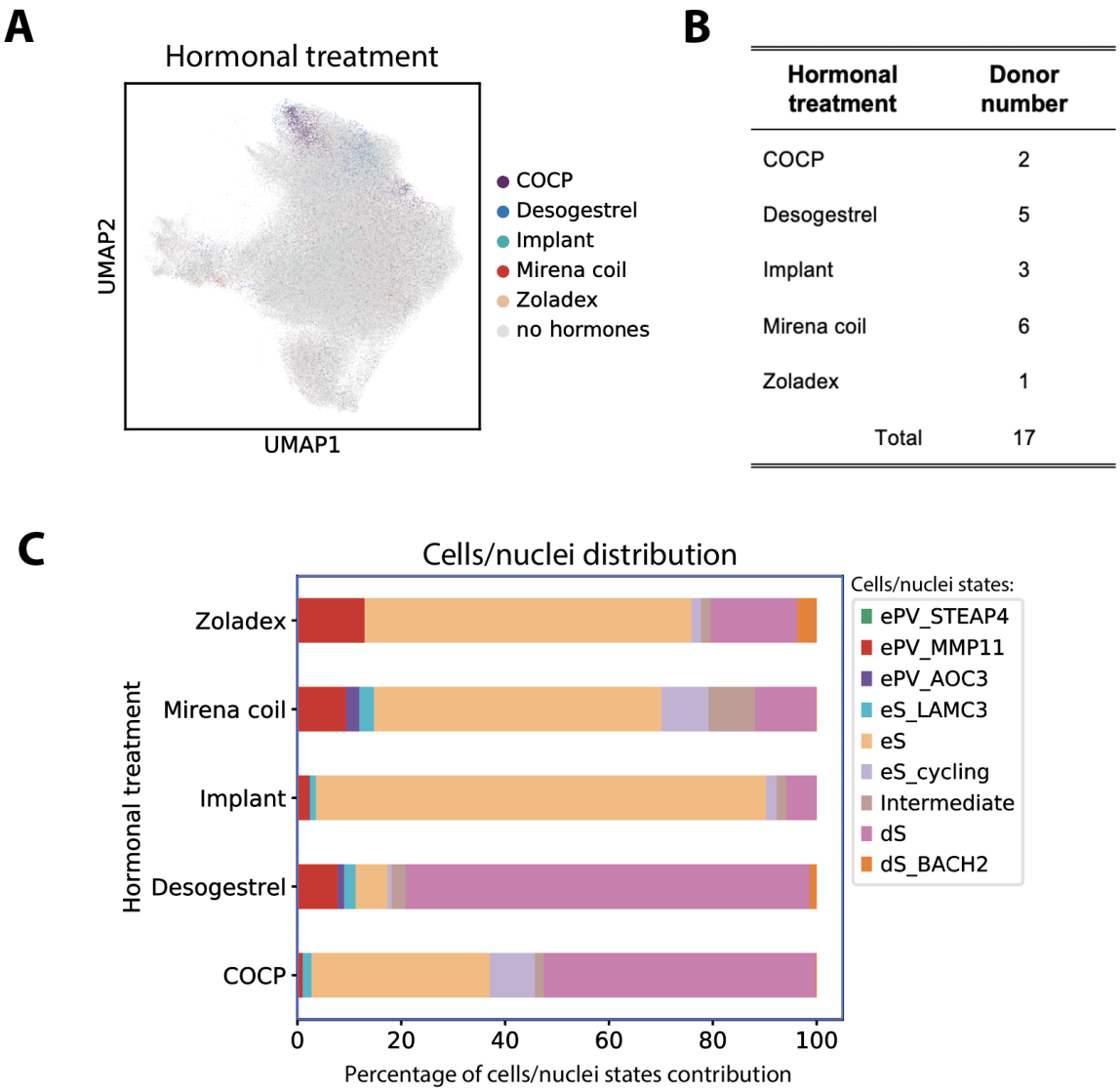
Figure 4.13: Mesenchymal cells/nuclei across the menstrual cycle. Shown is an UMAP projection of the scRNAseq and snucRNAseq datasets integrated, coloured by menstrual cycle group assignment (A). A summary table of the number of donors and cells/nuclei profiled per group is shown in (B). In (C) the contribution of each menstrual cycle group to the cells/nuclei states is shown, while in (D) the distribution of cells/nuclei states per menstrual cycle group is shown. *Abbreviations*: dS, decidualised stromal; ePV, endometrial perivascular; eS, endometrial stromal; scRNAseq, single-cell RNA-sequencing; snucRNAseq, single-nucleus RNA-sequencing; UMAP, uniform manifold approximation and projection.

Hormones

Overall, only ~10,000 cells/nuclei from 17 donors taking exogenous hormones were recovered, with a few donors contributing only a limited number of cells/nuclei. As described for the epithelial lineage section above, the donors analysed were taking 5 different hormonal treatments (Figure 4.14B) and the UMAP visualisation is shown in Figure 4.14A.

The COCP treatment appeared to induce a phenotype similar to the secretory phase endometrium with a reduced proportion of the eS and eS_cycling state, some intermediate cells/nuclei and a higher proportion (~50%) of dS cells/nuclei (Figure 4.14C). The hormones_dS population described in Chapter 3 to be specific to the progestin pill Desogestrel was confirmed to be one donor-specific and thus was not analysed further, but the overall secretory phase phenotype was observed also in this chapter for this hormonal treatment. In this group, the proportion of dS cells/nuclei was highest (~80%) when compared to the rest of the samples taking other hormonal therapy.

In contrast, the progestin-releasing implant induced more of a proliferative phase phenotype with ~85% of cells/nuclei assigned the eS identity (Figure 4.14C). Similarly, the progestin-releasing IUD Mirena was defined by a higher percentage of eS cells/nuclei (~55%) and a smaller proportion of dS cells/nuclei (~12%) which was also observed in Chapter 3. Taking the GnRH agonist, Zoladex, likely induced a similar phenotype to the Mirena coil one but with a reduced number of eS_cycling cells/nuclei and more ePV_MMP11 cells/nuclei; however, due to analysing only one donor sample in this group the results remain inconclusive.



Hormonal treatment	Donor number
COCP	2
Desogestrel	5
Implant	3
Mirena coil	6
Zoladex	1
Total	17

Figure 4.14: Mesenchymal cells/nuclei in donors taking exogenous hormones. Shown is an UMAP projection of the scRNAseq and snucRNAseq datasets integrated, coloured by type of hormonal treatment (A). A summary table of the number of donors and cells/nuclei profiled per treatment is shown in (B). In (C) the distribution of cells/nuclei states per hormonal treatment is shown. *Abbreviations*: COCP, combined oral contraceptive pill; dS, decidualised stromal; ePV, endometrial perivascular; eS, endometrial stromal; scRNAseq, single-cell RNA-sequencing; snucRNAseq, single-nucleus RNA-sequencing; UMAP, uniform manifold approximation and projection.

4.2.2.3.3 Mesenchymal cells/nuclei states in endometriosis and controls

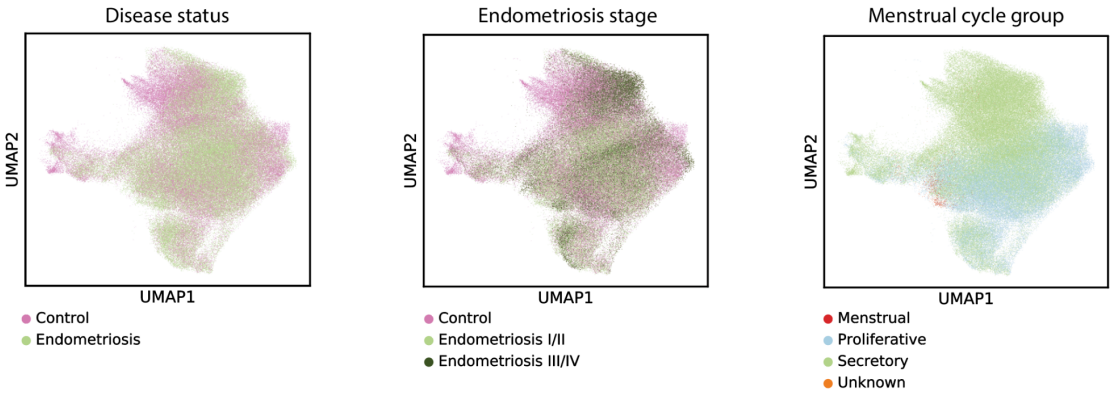
As described for the epithelial lineage, to compare the endometrium in endometriosis and controls, only samples from donors not taking exogenous hormones and from the proliferative and secretory phases (n = 66, green rectangle in Figure 4.15B) are considered here. Moreover, the endometriosis stages I/II and

III/IV are also merged under one ‘endometriosis’ group for the same reasons as described for the epithelial lineage. Only UMAP visualisations of the different disease stages and menstrual cycle groups are shown in Figure 4.15A and percentual contributions to the different groups across controls and endometriosis cases are summarised in Figure 4.15B.

Comparing a smaller number of samples, no differences between cases and controls was reported during the proliferative phase in Chapter 3, while a trend was observed for a smaller proportion of dS cells in endometriosis cases during the secretory phase when compared to controls. Analysing a larger set of samples confirmed that indeed no differences in the cellular composition of the proliferative phase samples could be observed between endometriosis cases and controls (Figure 4.15C) and this was also true when only the nuclei samples were considered (Figure 4.15D).

Similar to Chapter 3, differences were observed between cases and controls when considering the secretory phase samples - a higher proportion of dS cells/nuclei (42% vs 28%) was observed in control samples. However, this result was likely due to the bias of the Wang_Cells dataset profiling 10 secretory phase samples around the time of the window of implantation and late secretory phase samples as described above. To further validate these findings, I thus looked at comparing only the nuclei data generated for my study. In this scenario, secretory phase samples from endometriosis cases have a slightly lower proportion of dS nuclei (28% vs 23%) and slightly higher proportion of the intermediate nuclei (11% vs 6%) when compared to control samples (Figure 4.15D).

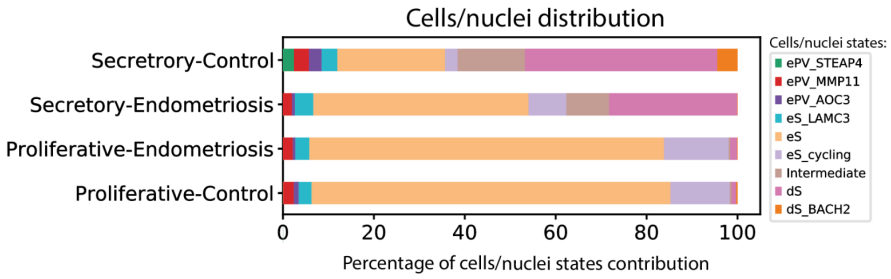
A



B

	Controls		Endometriosis	
	Donor number	% cells/nuclei	Donor number	% cells/nuclei
	32	45%	41	55%
Menstrual phase				
Menstrual	1	3%	4	97%
Proliferative	10	52%	15	48%
Secretory	20	42%	21	58%
Unknown	1	100%	1	0%
Endometriosis stage (rASRM)				
I/II	-	-	24	56%
III/IV	-	-	17	44%

C



D

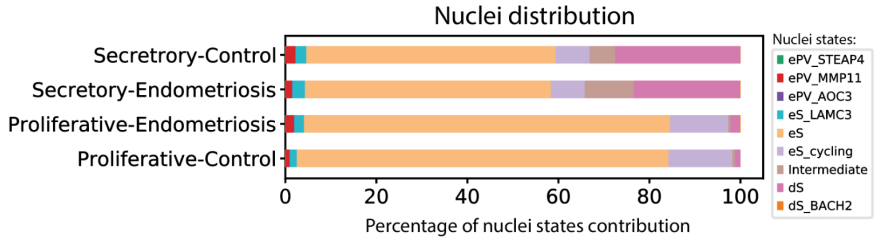


Figure 4.15: Mesenchymal cells/nuclei in endometriosis and controls. Shown is an UMAP projection of the scRNAseq and snucRNAseq datasets integrated, coloured by disease status, endometriosis stage and menstrual cycle group (A). A summary table of the number of donors and cells/nuclei profiled for controls and endometriosis cases is shown in (B). In (C) the distribution of cells/nuclei states per proliferative and secretory phase for controls and endometriosis cases is shown, while in (D) the same is shown for nuclei only. *Abbreviations*: dS, decidualised stromal; ePV, endometrial perivascular; eS, endometrial stromal; scRNAseq, single-cell RNA-sequencing; snucRNAseq, single-nucleus RNA-sequencing; UMAP, uniform manifold approximation and projection.

4.5 Discussion

In this chapter I have shown that it is possible to generate high quality snucRNAseq data for a large number of endometrial samples across the menstrual cycle in health and endometriosis, as well as when donors take exogenous hormones. Moreover, I achieved successful integration of the scRNAseq datasets presented in Chapter 3 and the newly generated snucRNAseq data. Using the marker genes defined in Chapter 3, I could identify and annotate most of the cells/nuclei states in the mesenchymal and epithelial lineages and could examine their distribution across the cycle in controls, endometriosis and under the influence of exogenous hormones. Some of the findings are discussed below in more detail.

4.5.1 Integration of scRNAseq and snucRNAseq data

In order to integrate the cells and nuclei transcriptomic data, the machine learning tool *scVI* (described in more detail in Chapter 3) was used as it is suitable for analysing complex datasets. Using neural networks, the probabilistic model is trained on the data provided and generates a latent representation of each cell in a batch corrected space. To learn the cells' latent representation and remove batch effects, multiple parameters can be adjusted to achieve optimal data integration, including the number of layers of the neural networks the model passes the data through. Here, integrating scRNAseq and snucRNAseq data required a higher number of layers of the neural networks than when integrating across the three different scRNAseq datasets in Chapter 3. This suggests higher complexity of the data and differences between the cells and nuclei datasets that needed to be considered by *scVI* and is in line with the multiple different variables present across the datasets and the number of samples analysed/integrated in this chapter.

Overall, the integration worked well and highlighted some differences that require further investigation, especially, the lower recovery of PV, endothelial and stromal nuclei in the snucRNAseq data and lower number of epithelial cells detected in the scRNAseq data. Multiple factors might be responsible for this observation, including the way in which the endometrial samples are dissociated to obtain single cells. A two-step enzymatic digestion protocol is used to dissociate the tissue, suggesting that in the initial step the collagenase enzyme used readily releases the stromal, PV and endothelial cells from the extracellular matrix while the epithelial glands remain to some extent undigested and require further treatment with trypsin. Due to this, the proportion of stromal, PV and endothelial cells may be higher in scRNAseq data. It may also be that the nuclei extraction protocol is not fully optimised for capturing PV and endothelial nuclei.

Considering the epithelial cells, by separately loading the fraction of epithelial cells after trypsin digestion, I should have obtained higher cell numbers. However, the prolonged dissociation and trypsin treatment resulted in the observed increased number of mitochondrial genes expressed (because of dissociation stress) leading to a higher exclusion rate of these cells due to low QC metrics. Such ‘dissociation signature’ of increased mitochondrial gene expression was described previously in the gut after collagenase digestion[171] and was in contrast to the very low expression of mitochondrial genes detected in the snucRNAseq data. The low mitochondrial gene expression may be due to less dissociation stress, but is also a feature of not profiling genes from the cytoplasm (rich in mitochondrial genes) with snucRNAseq. Lastly, it is also possible that sampling and processing of samples for scRNAseq

(cryopreservation) and snucRNAseq (snap-freezing) have different effects on the viability and recovery of these cells/nuclei.

4.5.2 The epithelial and mesenchymal cells/nuclei states identified

Most of the cells/nuclei populations described in this chapter recapitulated the ones described in Chapter 3, including their marker gene expression. This supports the robustness of these populations across the different types of data (cells and nuclei) when it comes to their annotation to the level described in this thesis. However, some cell states defined in Chapter 3 were not recapitulated in this chapter, confirming that these populations were donor-specific, such as the hormones_inflammatory, hormones_LGR5, eS_BMP7, hormones_dS and dS_basal populations. The dS_BACH2 and glandular_secretory_FGF7 populations require further investigation as only a very small number of these cells/nuclei came from the nuclei dataset, but were contributed by multiple donors. As the endometrium is highly dynamic, it may be that these cells/nuclei were not captured due to not profiling the 'correct' stages of the menstrual cycle.

Chapter 4 new populations

Apart from the populations defined and described in Chapter 3, I also defined the proliferative_cycling and eS_LAMC3 populations in this chapter. The genes characteristic of the proliferative_cycling population are associated with the cell cycle and meiotic division: centromere protein P (*CENPP)*, meiotic nuclear divisions 1 (*MND1*), and maternal embryonic leucine zipper kinase (*MELK).* In Chapter 3 this population was likely not detected due to the small number of samples profiled during the proliferative phase. The eS_LAMC3 population shared some marker gene expression with the eS/eS_cycling populations (e.g. serine protease 12

(*PRSS12*) and cellular retinoic acid binding protein 2 (*CRABP2*)) while the laminin subunit gamma 3 (*LAMC3*) gene was to a lesser extent expressed by the ePV_MMP11 population. This could suggest eS_LAMC3 as a transitioning population between the ePV and eS states. In addition, all of the identified genes are thought to be involved in cell growth and differentiation, but their role in endometrial function is not known, only abnormal CRABP2 expression has been reported in endometrial cancer[204].

While the MUC5B population was described in Chapter 3 in a few samples, I detected a much higher number of donors and nuclei within this population when analysing the nuclei data. The MUC5B population showed great heterogeneity across the donors analysed and new markers were defined for its identification such as protein tyrosine phosphatase receptor type D (*PTPRD*), transforming growth factor beta receptor 3 (*TGFBR3*) and dickkopf WNT signalling pathway inhibitor 2 (*DKK2*). These new markers were likely identified as a result of analysing a larger number of samples in this chapter, but one should note that the nuclei transcriptomic data was aligned also to the intronic regions of the genome and thus the expression of these genes now needs to be validated and the exact location of the MUC5B population determined using e.g. smFISH staining of full-thickness endometrial sections *in situ*.

Datasets' features and cell populations

Throughout this chapter I described certain features of each of the datasets analysed that seemed to skew the data to some extent. For example, the Wang *et al.* dataset had the highest proportion of ciliated, ciliated_LGR5 and luminal cells

characteristic of the window of implantation samples. The Garcia-Alonso *et al*. dataset had unique populations (e.g. Fib_C7, mPV_MYH11) in the whole uterus samples. In addition, samples from this dataset had a high number of ePV_STEAP4 and ePV_AOC3 cells and only a very small number of these was detected in the nuclei data. This may be for the reasons described above, such as differences in the nuclei vs cells isolation protocols, but could also be due to the varied sampling of the endometrium. Next steps will be to stain whole uterus samples with the markers identified in order to validate the presence and location of these populations across the full thickness endometrium. With regards to the ciliated, ciliated_LGR5 and luminal cells presence, focus should be shifted towards samples from around the time of the window of implantation.

Naming of the proliferative and luminal cell populations

In this thesis, I have presented a slightly different naming strategy for epithelial cells in the proliferative phase samples when compared to those reported previously in the Garcia-Alonso *et al*. study[98]. After the datasets integration and cluster identification, I was not able to confidently distinguish between the Luminal 1 and Luminal 2 populations described by the authors and opted for calling cells expressing the markers of the luminal 1 population as luminal only. Moreover, the previously described *SOX9*$^{+}$ proliferative, *SOX9*$^{+}$*LGR5*$^{-}$, *SOX9*$^{+}$*LGR5*$^{+}$ populations are in this thesis referred to as proliferative_cycling, proliferative and proliferative_LGR5 populations. This is due to the low expression of *SOX9* detected in my integrated datasets, likely due to the different genome reference used here to align the data. To avoid confusion, I therefore chose to rename these populations in this thesis; however, importing the cell annotations from the Garcia-Alonso *et al*.

study showed that the identity and naming of these cell populations match (data not shown).

4.5.3 The endometrium across the menstrual cycle

The rationale behind integrating all of the scRNAseq and snucRNAseq data was to obtain a comprehensive cellular map of the endometrium across the cycle as so far such analyses relied on a limited number of cells and samples[97,98]. I was able to characterise the proliferative and secretory phases in a large number of donors, but could only analyse 5 samples from the menstrual phase. Samples from this phase of the cycle are very difficult to obtain but are of great interest to study as re-epithelisation and regeneration occur simultaneously as the endometrium sheds in a piecemeal fashion[20]. A recent study has shown that it is possible to generate scRNAseq data from the menstrual effluent which could help understand the menstrual phase further[145]. In the future, I will be exploring the integration of this dataset together with the data I generated to compare the properties of the shed and *in situ* endometrium.

With regards to the changes across the menstrual cycle, at the time of my thesis writing, only the broad classification of the cycle (proliferative, secretory, menstrual phase) was available for all donors analysed and more detailed annotations were available for a limited number of donors. However, as the endometrium is highly dynamic and its morphology and function changes within days, this staging is not accurate enough for a thorough analysis of the cellular composition and changes across the cycle. In order to pinpoint when certain cell populations appear and disappear, such as e.g. the newly defined TENM2 and intermediate populations, a more fine annotation of the cycle phases is needed and will be provided by the

pathologists I worked with. Moreover, for the two donors with the unknown phase, their clinical data will be reviewed for more information. A more detailed annotation of the menstrual cycle phases will be key in understanding the transition of the endometrium across the cycle, its regeneration and response to hormones. Once this information is available, I will perform trajectory and cell-cell communication analyses to corroborate the cell interactions mediating these processes.

4.5.4 The endometrium under the influence of exogenous hormones

As described in Chapter 3, understanding what effects exogenous hormones have on endometrial function is needed as globally millions of women take hormones for both contraceptive and therapeutic reasons[190,205,206]. In this chapter I have shown that it is possible to generate scRNAseq and snucRNAseq data for samples from donors taking different hormonal therapies, and these therapies induced various phenotypes. For the first time I described the hormones_WNT7A population in donors having the progestin-releasing IUD Mirena coil fitted and that overall, having this treatment at the cellular level resembles the proliferative phase endometrium. A limited number of samples from donors taking the GnRH antagonist Zoladex and COCP (oestrogen and progestin pill) also induced a proliferative phase phenotype, with COCP resembling the late proliferative phase endometrium transitioning into the secretory phase. Given the small number of samples analysed, the effects of Zoladex and COCP need to be studied in a larger set of samples. In contrast, the progestin pill Desogestrel induced a secretory phase phenotype while the progestin-releasing implant resembled the proliferative phase endometrium. Taken together, these findings highlight the heterogeneity of responses observed when taking exogenous hormones, even when the same hormone (i.e. progestin) is used[71]. The cellular and tissue response observed will depend on many factors, including the

route of administration (pills orally vs IUD vs implant), the dose and formulation taken, each individual's metabolism, etc. (reviewed in [71]).

Of note is that a low number of mesenchymal nuclei was recovered from samples of donors on hormones, suggesting that this reduction could be a feature of taking exogenous hormones. Morphologically, taking exogenous hormones has been linked to endometrial thinning[207] supporting such an observation.

4.5.5 Cellular composition of the endometrium in endometriosis and controls

In this chapter I explored whether the differences observed between endometriosis cases and controls in a limited number of samples in Chapter 3 would also be recapitulated in a larger dataset. In order to have a balanced number of samples and cells between controls and endometriosis cases, I only analysed the proliferative and secretory phases. Interestingly, the observed differences were more prominent in the epithelial lineage than the mesenchymal lineage. During the proliferative phase, higher proportions of the epithelial lineage's proliferative_cycling and TENM2 cells/nuclei and reduced proportion of MUC5B cells/nuclei in cases when compared to controls were noted. The presence of the proliferative_cycling population to a bigger extent suggests aberrant proliferation of the endometrium in endometriosis, and has been reported previously[208]. The differences in the TENM2 and MUC5B populations are difficult to comment on as the exact roles of the TENM2 and MUC5B populations remain to be investigated. No differences were found between endometriosis cases and controls when comparing the mesenchymal cell populations in the proliferative phase.

In contrast, differences were observed in the secretory phase samples, but to a much lesser extent when only the nuclei samples were analysed. The unique feature of the window of implantation samples from the Wang *et al.* dataset meant that when all datasets were considered, there was a higher proportion of ciliated, ciliated_LGR5 and luminal epithelial cell states in controls when compared to endometriosis cases. Endometriosis is associated with infertility[209–211], but the exact reasons for this remain unknown. Altered endometrial composition and response to hormones have been suggested[212–215], which also the data here could suggest, but needs to be interpreted with caution due to the inherent biases and limitations of the datasets analysed here. In order to better understand the differences, especially around the time of window of implantation between cases and controls, timed samples from this phase of the cycle should be collected from both endometriosis cases and controls and compared to those of the Wang *et al.* dataset.

Furthermore, in the mesenchymal lineage, higher proportion of PV and dS cells/nuclei in controls can be attributed to the higher number of these cells in the Garcia-Alonso and Wang datasets, respectively. These large differences between cases and controls also disappeared when analysing the nuclei samples only, and just a slight reduction in the proportion of dS and increase in the intermediate population was observed in endometriosis cases. Once again, this warrants further investigation and finer annotation of the menstrual cycle phases to be able to unpick the real differences between cases and controls instead of signatures specific to certain phases of the cycle.

4.5.6 Summary of future directions

Due to time constraints, I was only able to provide a descriptive account of the transcriptomic data generated during my studies in this chapter. However, further analyses and data validation steps are needed to evaluate the descriptive findings reported. Below is a summary of 10 future steps I would like to explore further:

1. *Perform more detailed benchmarking of the scRNAseq and snucRNAseq data using quantitative measures, including machine learning based methods such as logistic regression or correlation methods such as cosine distances to quantify the cells/nuclei transcriptomic similarities.*
2. *Re-analyse the snucRNAseq dataset on its own and use label transfer tools to identify and annotate the nuclei dataset to confirm if the results are robust, i.e. can the same results be obtained using various approaches?*
3. *Annotate and analyse the rest of the cells/nuclei lineages (i.e. immune and endothelial). How do they vary across the cycle and different conditions?*
4. *Obtain more fine annotations of the menstrual cycle phases, i.e. early/late proliferative, early/mid/late secretory assignment and repeat the analyses based on this stratification.*
5. *Stratify patients based on the different endometriosis stages (I/II and III/IV) as well as other pathologies and describe their endometrial transcriptomic signatures.*
6. *Perform differentially expressed genes (DEGs) analyses to identify menstrual cycle phase signatures and differences between controls and endometriosis cases at the cell state level.*

7. *Perform cell-cell communication analyses (i.e. using CellPhoneDB[216]) to understand the cellular interactions driving endometrial regeneration and maturation across the menstrual cycle.*
8. *Define the specific transcriptomic signatures for each of the menstrual cycle phases in health and explore whether these signatures can be regressed out, and thus allow for the identification of a, for example, disease signature that is menstrual cycle phase independent.*
9. *Obtain further samples from donors taking different types of exogenous hormone therapies and describe their characteristics.*
10. *Pinpoint the in-situ location and timing of appearance of the newly defined populations through e.g. spatial transcriptomics/smFISH imaging.*

4.6 Conclusions

Taken together, in this chapter the newly generated snucRNAseq data integrated with the three datasets analysed in Chapter 3 were presented. My aim was to describe the commonalities and differences between the scRNAseq and snucRNAseq while building a comprehensive cellular map of the endometrium in health, endometriosis and under exogenous hormones. I focused on annotating the epithelial and mesenchymal lineages, evaluating which of the cell populations and findings described in Chapter 3 can be recapitulated in the larger dataset shown here. In doing so, I have built an updated map of the endometrium across the menstrual cycle, consisting of the largest number of donor samples, cells/nuclei and conditions analysed. Due to the time constraints faced, I only presented a descriptive summary of the observations made while putting this integrated map

together and provided a list of future analyses I will undertake. In the next chapter I describe my work on *in vitro* models of the endometrium.

Chapter 5: Human endometrium *in vitro*

5.1 Introduction

In vitro models of the endometrium can provide a powerful platform to study endometrial biology and function in both health and disease (please see Chapter 1 for further information). Endometrial epithelial organoids (EEOs) can be used as a model for disorders such as endometriosis, hyperplasia, endometrial cancer, but also embryo implantation as reviewed and postulated extensively[80,91,217,218]. However, an in depth characterisation and benchmarking of this model system across a large number of samples is still lacking to be able to harness its full potential. Single-cell genomics can serve as an effective tool to quantitatively evaluate and assess the EEO system's quality and guide its development so that it successfully recapitulates the physiological or pathological processes of interest *in vitro*[150]. In this chapter, I briefly summarise the work done in this respect and describe the cellular heterogeneity of EEOs and EEOs co-cultured with stromal cells (EEO-ES). Moreover, I also explore the resemblance of cells from both of these model systems to the *in vivo* primary endometrial cells.

5.1.1 Single-cell profiling of 3D endometrial cell *in vitro* model systems

To date a number of studies have profiled EEOs using scRNAseq (Table 5.1); however, they differed greatly in their choice of media and protocols for organoid culture as well as the number of donor samples, disease status and cells profiled. In addition, the number of cell states defined and nomenclature used to assign their identities varied greatly (Table 5.1). Multiple epithelial cell states in EEOs and their response to hormonal stimulation was first reported by Fitzgerald *et al.*[148] Analysing organoids from one donor line they reported an increase in the number of ciliated

cells in response to oestrogen (E2) and increase in secretory cells in response to medroxyprogesterone acetate [(MPA) a progesterone derivative more resistant to metabolism[219]] alone, or in combination with cyclic adenosine monophosphate (cAMP). The Cochrane *et al*. study on the other hand reported only the presence of ciliated and secretory cells and changes in their proportions in response to treatment with a NOTCH pathway inhibitor[149]. Tan *et al*.[99] profiled EEOs from healthy donors (*n* = 2) and endometriosis cases (*n* = 4) without any treatment, but did not provide a detailed annotation of the EEO cellular heterogeneity and whether this varied between controls and endometriosis cases.

Garcia-Alonso *et al*.[98] performed the first and only detailed benchmarking analysis of the EEO model system to date, comparing it to *in vivo* endometrial epithelial cells using 3 EEO lines. Studying the response of EEOs to hormonal stimulation, similarly to the Fitzgerald *et al.* study, they observed an increase in secretory cells after first treating EEOs with E2 followed by treatment with E2, progesterone (P4), cAMP and prolactin (PRL) (for details see Table 5.1). Using computational tools and their endometrial *in vivo* scRNAseq reference, they found that the transcriptomic profiles of EEO ciliated cells almost perfectly matched those of the *in vivo* ciliated cells, and about a quarter of cells assigned the secretory identity matched glandular epithelium *in vivo*. Furthermore, treating EEOs with E2 only led to the emergence of pre-ciliated and E2-induced cell states, which closely matched the pre-ciliated and *SOX9*$^{+}$ cell populations they described *in vivo*, respectively. The results of treating EEOs with WNT/NOTCH pathway inhibitors is not within the scope of this thesis.

Table 5.1: Publications profiling 3D endometrial *in vitro* systems at single-cell level.

	Publication[a]				
	Fitzgerald et al.	**Cochrane et al.**	**Rawlings et al.**	**Garcia-Alonso et al.**	**Tan et al.**
Model system[b]	Organoids	Organoids	Assembloids	Organoids	Organoids
Endometrial samples	*n* = 1 Healthy	*n* = 3 Endoms[c]	*n* = 3 Healthy	*n* = 3 Healthy	*n* = 6 Healthy (*n* = 2) Endoms (*n* = 4)
Media[d]	T	Own media	T with E2 & own media[e]	T	B
Hormonal stimulation					
First phase:	*2 days with:* 10nM E2	N/A	*4 days with:* 10nM E2 (in T media)	*2 days with:* 10nM E2	N/A
Second phase:	*6 days with:* 10nM E2 1µM MPA or 10nM E2 1µM MPA 1µM cAMP	N/A	*4 days with:* 1µM E2 1µM MPA 0.5mM cAMP (in own media)	*4 days with:* 10nM E2 1µM P4 1µM cAMP 20ng/ml PRL	N/A
Cells total	~22,000	~9,000[f,g]	~3,000	~37,000[g]	~6,000
Cell states	Ciliated Epithelial Prolif Secretory Stem Unciliated	Ciliated Secretory	5 epithelial 1 transitional 5 stromal	NH Day 0 NH Day 2 NH Day 6 NH Prolif Estrogen induced Pre-ciliated Ciliated Secretory Secretory cycling KRT17+	Ciliated Glandular prolif MUC5B+

Abbreviations: cAMP, cyclic adenosine monophosphate; E2, oestrogen; Endoms, endometriosis; MPA, Medroxyprogesterone acetate; P4, progesterone; Prolif, proliferative.

Superscripts:

[a] *Listed according to publication date.*

[b] *Organoids are epithelial endometrial organoids grown in Matrigel. Assembloids are epithelial endometrial organoids and stromal cells co-cultured in a synthetic hydrogel.*

[c] *Collected during hysterectomies, all patients had endometriosis, one also leiomyomas.*

[d] *Media composition as described in Turco et al. 2017 (T) and Boretto et al. 2017 (B).*

[e] *Turco et al. 2017 composition for organoid establishment. To culture assembloids the media was supplemented with E2 for 4 days followed by their own media composition with P4, cAMP and MPA.*

[f] *Numbers not provided - estimated from a bar plot in Figure 1E.*

[g] *Excludes cells that were profiled when treated with NOTCH inhibitors (Cochrane et al. & Garcia-Alonso et al.) and WNT inhibitors (Garcia-Alonso et al.).*

Lastly, an assembloid model of the endometrium (i.e. EEO-ES co-cultured in a synthetic hydrogel) was profiled using scRNAseq by Rawlings *et al.*[87] before and after hormonal stimulation to determine the decidualisation response of the cells, and suitability of the model system to study embryo implantation. The authors reported cellular heterogeneity within both the epithelial and stromal compartments, showing differentiation of cells into more decidualised stromal and secretory epithelial cells expressing 'receptivity genes' upon stimulation. Marker genes and likely function for each of the cell subpopulations were provided; however, the benchmarking was performed by comparing the scRNAseq assembloid data with a previously published bulk RNAseq dataset.

5.1.2 Media composition and hormonal stimulation of endometrial cells *in vitro*

As highlighted in Table 5.1 and in the above section, various media compositions and protocols for hormonal stimulation have been used across these studies, with some using their own media, while the rest relied on either the exact media composition, or alterations of these, published by the two seminal papers reporting EEOs development: the Turco *et al.*[84] and Boretto *et al.*[85] studies. The exact differences between the Turco *et al.* and Boretto *et al.* media can be viewed in Appendix 4, but for the purposes of this chapter, it is important to highlight the presence of E2 in the Boretto *et al*. media throughout EEO culture, while E2 is absent in the Turco *et al*. media. Given EEOs are hormone-responsive and E2 is present in the body throughout the menstrual cycle, I hypothesised that the E2-containing media would generate cells more similar to the *in vivo* endometrial cells. I tested this hypothesis in a pilot experiment (Appendix 4) and observed that EEOs cultured in E2-containing media most closely resembled *in vivo* endometrial cells

and thus used this media type for both the EEO and EEO-ES systems presented in this chapter.

In addition, I adapted the hormonal stimulation protocol described by Boretto *et al.*[85] (see section 5.3.6 for more details) using E2 stimulation followed by reduced E2 and P4 treatment to mimic the proliferative and secretory phases of the menstrual cycle. I hypothesised that such treatment would generate a more physiologically-relevant response than the previously reported use of MPA and cAMP for the following reasons: (i) MPA has been synthesised to be metabolically more stable than P4, i.e. making it more potent and thus exhibiting a stronger response, and (ii) cAMP has similarly been shown to induce stromal cell decidualization more rapidly and efficiently than P4 *in vitro*, but also to induce upregulation of different genes to P4 only treatment[220–222]. Here, I characterised the response of EEOs and EEOs co-cultured with stromal cells from donors with/without endometriosis and taking exogenous hormones using a combination of culture media and hormonal stimulation regime not studied before at the single-cell level.

5.2 Aims

1. Establish organoid and stromal cell lines from donors with and without endometriosis (i.e. controls) and those taking exogenous hormones.
2. Establish co-cultures of organoids & stromal cells.
3. Profile both systems (organoids only and co-cultures) at the single cell level after hormonal stimulation mimicking the menstrual cycle.
4. Describe the cellular heterogeneity of these systems and their resemblance to the *in vivo* primary endometrial tissue.

5.3 Materials & Methods

5.3.1 Organoid and stromal cell lines derivation

To derive epithelial organoid and stromal cell lines, endometrial pipelle biopsies were collected and processed as described throughout this thesis for scRNAseq. Briefly, collected biopsies were cut-up, cryopreserved and stored at -80°C until the day of enzymatic digestion (described in Chapter 3, sections 3.3.1 and 3.3.2). After the first step of digestion with collagenase V, termed the collagenase fraction, cells in this fraction were filtered and placed in either plates or culture flasks, the size of which depended on cellular yields and viability. I followed the seeding densities and media volumes recommended by ThermoFisher[223] (Appendix 5) and placed the counted cells in stromal culture media (see Table 5.2 for media composition) for 30 - 45 min inside a 37°C incubator supplied with 5% CO_2. This step was performed to separate the epithelial and stromal cells from within the collagenase fraction. Stromal cells are plastic-adherent and readily attach to the bottom of the flask, while epithelial cells float in the medium. At the end of the incubation period, stromal cell attachment was visually confirmed using the EVOS M5000 Imaging System (Invivogen, AMF5000) and the medium removed and centrifuged (500 x g, 5 min). The supernatant was discarded and the cell pellet used for epithelial organoid derivation as described in section 5.3.3 below. The attached stromal cells in the flask were provided fresh pre-warmed stromal culture media and maintained as described in the next section. Figure 5.1 provides a visual overview of how the organoid and stromal cell lines were established.

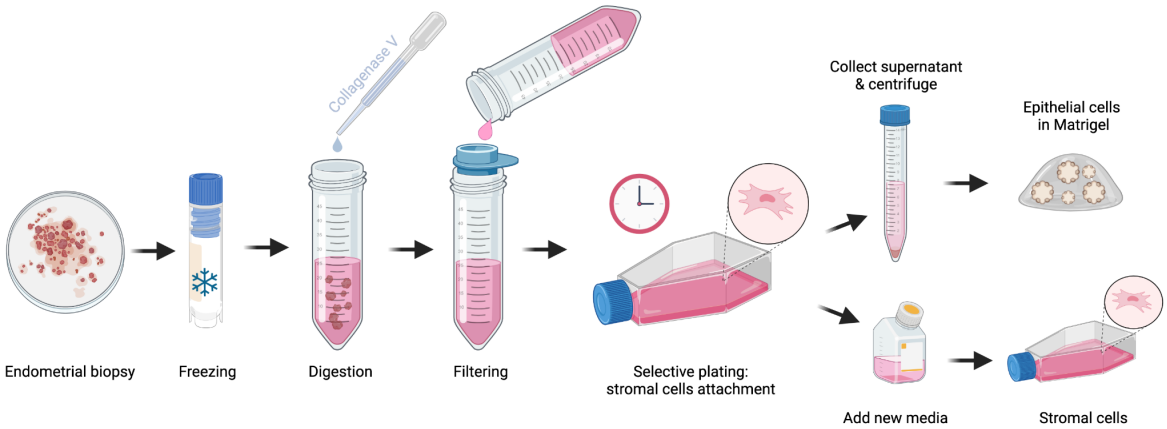


Figure 5.1: Organoid and stromal cell line derivation overview. Prior to digestion, endometrial biopsies were cryopreserved and thawed on the day of derivation. Upon digestion with collagenase V, the digest was filtered, and the released cells placed in a flask to allow for stromal cell attachment. Floating epithelial cells were obtained by collecting the supernatant, followed by centrifugation and embedding the cell pellet in Matrigel domes. Upon culture, these cells developed into 3D organoid structures. The attached stromal cells received fresh media and were cultured as a 2D monolayer until line establishment.

5.3.2 Culturing stromal cells

Endometrial stromal cells were cultured at 37°C in an incubator (5% CO_2) in the stromal culture medium (composition summarised in Table 5.2). The media was changed every 2-3 days. Once cells reached 90-100% confluency, they were lifted from the plates/flasks using pre-warmed TrypLE™ (Gibco, 12604013) for 5 min at 37°C. The dissociation was stopped by adding the stromal culture medium, cells were centrifuged (500 x g, 3 min), counted and placed in new flasks for further expansion (typically passaged at a 1:3 ratio) or cryopreserved and banked. To bank the stromal cell lines, after counting, cell suspensions were centrifuged again and resuspended in 500-1000 µl of ice-cold CryoStor®cell cryopreservation medium and frozen at -80°C as per the manufacturer's instructions using Corning® CoolCell®. The cryovials were then moved and stored in liquid nitrogen. All donor lines were frozen at the earliest passage (P) possible, most typically at P1 or P2.

Table 5.2: Stromal cell culture medium composition.

Reagent	Supplier	Product number	Final concentration
DMEM	Life Technologies	41965039	1 X
FBS	Sigma-Aldrich	F7524	10%
Primocin	Invivogen	ant-pm-1	100 µg / ml

Abbreviations: DMEM, Dulbecco′s Modified Eagle′s Medium; FBS, foetal bovine serum.

5.3.3 Culturing epithelial organoids

The cell pellet obtained after letting the stromal cells of the collagenase fraction attach was resuspended in ice-cold Matrigel (Corning, 356231) to derive epithelial organoids. Volume of Matrigel was adjusted based on the pellet size and cell viability. This was at around 1:3 ratio (pellet volume : Matrigel volume) for samples with poor viability and 1:5 for samples with > 70% viable cells. Organoids were cultured as described previously, combining the protocols of two seminal papers reporting their establishment[84,85]. Briefly, organoids formed inside Matrigel domes, created by dispersing 25 µl drops of the Matrigel-cell suspension into the centre of each well of 48-well plates (Corning, 3548). For the Matrigel to set, the plates were placed inside a 37°C incubator supplied with 5% CO_2 for 15 min before adding 250 µl of the organoids expansion medium (ExM). The ExM (see Table 5.3 for its composition) was supplemented with Y-27632 (a inhibitor of Rho-associated kinase (ROCK)) for the first 2-3 days after culture establishment or passaging, and changed every 2-3 days. Organoid growth was monitored and imaged every 2-3 days using the EVOS M5000 system and organoids passaged every 5-7 days at a ratio most appropriate for their current growth density (typically 1:5). The detailed protocol for passaging of organoids is described in[224]. In brief, organoids were disrupted mechanically, using an electronic pipette with a ‘mix’ function (Rainin, 17014493).

Fragmenting the organoids was achieved by pipetting the suspension up/down 300 times using small bore low-retention pipette tips (Rainin, 30389187), followed by a second round of manual pipetting of ~80-100 times. All washing and centrifugation steps (600 x g, 6 min, 4°C) were performed in ice-cold DMEM/F12 media. All steps requiring direct handling of the organoids were carried out using low-retention pipette tips to prevent organoids from sticking to the plastic.

Table 5.3: Organoid expansion medium composition.

Reagent	Supplier	Product number	Final concentration
DMEM/F12	Life Technologies	12634010	1 X
Noggin	Peprotech	120-10c	100 ng / ml
B27	Life Technologies	12587010	1 X
N2	Life Technologies	17502048	1 X
GlutaMAX	Life Technologies	35050061	1 %
ITS-G	Life Technologies	41400045	1 %
N-Acetyl-L-cysteine	Sigma-Aldrich	A9165-5G	1.25 mM
17-β Oestradiol	Sigma-Aldrich	E4389	1 nM
Primocin	Invivogen	ant-pm-1	100 µg / ml
Nicotinamide	Sigma-Aldrich	N0636	2 mM
bFGF	R&D Systems	233-FB-025	2 ng / ml
R-Spondin 1	BioTechne	4645-RS-01M/CF	500 ng / ml
SB 202190	Sigma-Aldrich	S7067-5MG	10 µM
EGF	Peprotech	AF-100-15	50 ng / ml
FGF-10	Peprotech	100-26	10 ng / ml
A 83-01	Tocris	2939	500 nM
Y-27632*	Merck	688000	10 µM

Abbreviations: DMEM/F12, Advanced Dulbecco's Modified Eagle Medium/Ham's F-12; ITS-G, insulin-transferrin-selenium; bFGF, fibroblast growth factor-basic; EGF, epidermal growth factor; FGF-10, fibroblast growth factor-10.
Superscripts:
* Y-27632 is only added to the media for 2-3 days after thawing or passaging organoids.

5.3.4 Freezing and thawing epithelial organoids

Protocols for freezing and thawing endometrial organoids are published. However, I found following these led to organoid loss and the failure to re-establish organoid growth post-thaw. Therefore, I worked on optimising organoid freezing and thawing using multiple approaches with the most successful one described here in detail. Firstly, the media was removed from the wells and replaced with 250 µl of ice-cold Cell Recovery Solution (Corning, CLS354253). After 1 h incubation at 4°C, contents of the wells were transferred into microcentrifuge tubes using wide-bore low-retention tips (up to 4 wells per tube) and centrifuged (600 x g, 6 min, 4°C). To prevent organoid loss during the centrifugation steps, I replaced using LoBind Eppendorf tubes with normal microcentrifuge tubes. Once the supernatant was discarded, the organoid pellet was gently resuspended in ice-cold CryoStor®cell cryopreservation medium using wide-bore low-retention tips, transferred to pre-cooled cryovials and frozen at -80°C as per the manufacturer's instructions using Corning® CoolCell®. I did not mechanically disrupt the organoids prior to freezing (as per published protocols), but froze the organoids as whole structures. I found that the fragmented organoids did not recover after freezing and thawing. The most satisfactory recovery after freezing/thawing was observed with organoids of ~100-200µm in diameter. Organoids of smaller sizes were difficult to pellet, while larger organoids exhibited a necrotic appearance in the centre upon thawing, and took a longer time to fully recover as they required earlier passaging. For banking and stimulation experiments, all organoid lines were frozen at early passages, most typically at P2 and the vials stored in liquid nitrogen upon freezing.

To thaw frozen organoid lines, I consulted Dr Iva Kelava, Senior Staff Scientist at Wellcome Sanger Institute, and optimised the final protocol as follows: after taking vials from liquid nitrogen, they were placed on dry-ice and thawed in a 37°C water bath for 2-3 min, until a small piece of ice remained. Using wide-bore low-retention tips, the cryovial contents were moved into 5 ml tubes (up to 2 cryovials were pooled per tube). Adding 1 ml at a time, ice-cold DMEM/F12 media was added in a drop-like fashion, while slowly swirling the tube to wash off the freezing media. Up to 5 ml of DMEM/F12 was added this way and the tube centrifuged (600 x g, 6 min, 4°C) to pellet organoids. The supernatant was discarded and organoids resuspended in Matrigel and cultured as described above.

5.3.5 Co-culturing epithelial organoids with stromal cells

To co-culture epithelial organoids with stromal cells, I established a collaboration with Professor Linda Griffith and Dr Juan Gnecco (both at Massachusetts Institute of Technology at the time) to trial their synthetic hydrogel supporting such co-cultures[86] in our lab. All of the reagents (summarised in Table 5.4) were prepared and gifted to me by Professor Griffith's group and Dr Gnecco provided guidance and advice on setting the system up. The protocol and peptide design is described in detail in their manuscript[86]. In brief, the hydrogel was made by functionalising the 8-arm 20 kDA polyethylene glycol vinylsulfone (PEG-VS) with a set of peptides designed to mimic the biophysical properties of the endometrium: integrins (PHSRN-K-RGDS and GFOGER) and matrix-binders (basement membrane binder and fibronectin binder). All were mixed in water and HEPES buffer to final concentrations summarised in Table 5.4, and left to react for 30 min at room temperature.

Table 5.4: Synthetic hydrogel composition.

Reagent	Supplier	Description	Final concentration
1 M HEPES, pH 8.2	Griffith Lab	Buffer in 10x PBS	10 %
10% (w/v) PEG	Griffith Lab	8-arm 20 kDa PEG-VS	13.5 mM
PHSRN-K-RGDS	Griffith Lab	Fibronectin-derived peptide	1.5 mM
GFOGER	Griffith Lab	Collagen I-derived peptide	1.5 mM
BM-binder	Griffith Lab	Basement membrane binder	0.5 mM
FN-binder	Griffith Lab	Fibronectin binder	0.5 mM
XL-LW	Griffith Lab	Dithiol crosslinking peptide	3.04 mM
NF-water	Ambion	Ambion, AM9937	NA

Abbreviations: HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; PBS, phosphate buffered saline; PEG, polyethylene glycol; PEG-VS, polyethylene glycol vinylsulfone; BM, basement membrane; FN, fibronectin; XL-LW, crosslinker; NF, nuclease-free.

During the functionalisation period, both the epithelial organoids and stromal cells were prepared for being mixed with the PEG-VS. Shortly, epithelial organoids were grown from single cells up to ~100 µm in diameter prior to harvesting. In order to do so, organoid cultures ready to be passaged were dissociated into single cells using TrypLE for 20-25 min at 37°C, filtered through 40 µm Flowmi® cell strainers and counted. 10,000 cells were seeded per Matrigel dome and grown in ExM supplemented with Y-27632 for the first 2-3 days and media changes every 2-3 days. Once they reached the appropriate size, the media was removed, organoids incubated with Cell Recovery Solution for 45 min at 4°C to remove Matrigel, pelleted and visually inspected and counted under the microscope (see Figure 5.2). Single-cell stromal cell suspensions were prepared as for passaging or freezing described in section 5.3.2. Organoids and stromal cells were pooled in the same tube, at concentrations of 10 whole intact organoids and 10,000 stromal cells per 1 µl of PEG-VS solution. The organoid and cell mixture was centrifuged (350 x g, 5 min,

4°C), media aspirated and the pellet carefully resuspended in the functionalised PEG-VS solution without creating bubbles. The crosslinker was then added and mixed well by pipetting. 3 µl droplets of the PEG-VS/cell suspension were dispensed into the centre of each well of non-tissue culture treated 96-well plates (Corning, 351172). The plates were inverted upside down and incubated for 30-45 min in a 37°C incubator supplied with 5% CO_2. Once the gelation was complete, 100 µl of ExM was added per well and the media changed every 2-3 days. Only matched (i.e. from the same donor) organoids and stromal cell lines were co-cultured. Passage numbers for both organoid and organoid-stromal co-cultures are summarised in Tables 5.5 & 5.6.

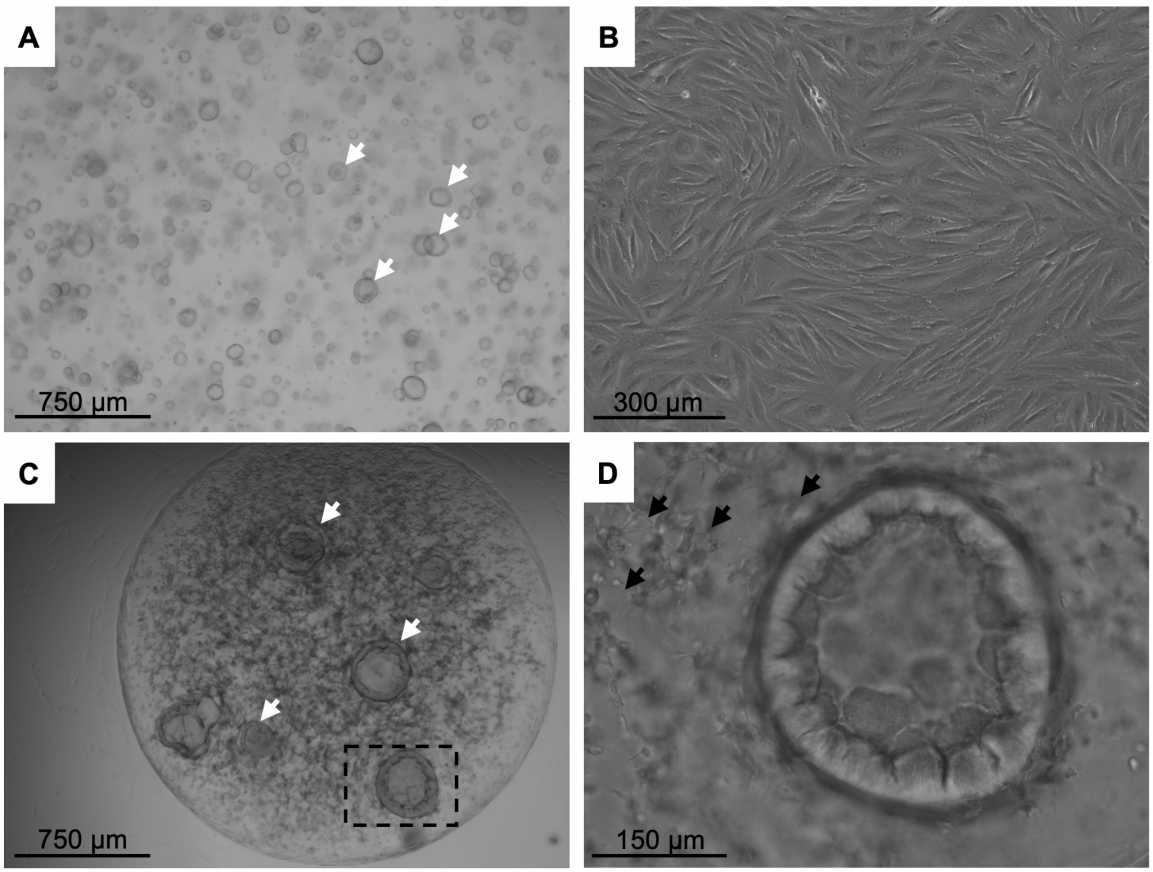


Figure 5.2: Epithelial organoid-stromal cells co-culture. (A) shows whole intact organoids grown from single cells ready for harvesting for setting-up co-cultures. (B) shows stromal cells ready for harvesting. (C) is an example image of organoid-stroma co-culture embedded in synthetic hydrogels. (D) shows a zoom-in of the rectangle area shown in (C). White arrows point out epithelial organoids, black arrows stromal cells.

5.3.6 Hormonal stimulation of epithelial organoids & organoid-stroma co-cultures

Both epithelial organoids and organoid-stroma co-cultures were stimulated with hormones, mimicking the menstrual cycle. Once set-up after passaging (organoids) or assembling in the hydrogel (co-cultures), the wells to be stimulated received ExM, which contained 17-β oestradiol (E2) at 1 nM final concentration for 4 days. Afterwards, for the following 6 days, the cultures received ExM with E2 reduced to 0.1 nM and progesterone (P4) (Sigma-Aldrich, P7756) at 50 ng/ml. This was to mimic the proliferative and secretory phases of the menstrual cycle, respectively. In case of cultures with epithelial organoids only, ExM was supplemented with 10 µM Y-27632 for the initial 2 days (see Figure 5.3) after set-up. In parallel, unstimulated wells were also cultured, receiving ExM containing E2 at 1 nM for 10 days. Media was changed every 2 days during the stimulation experiments. At the end of day 10, cultures were harvested and dissociated into single cells for scRNAseq analysis or fixed and stored for later analyses as described in the next section.

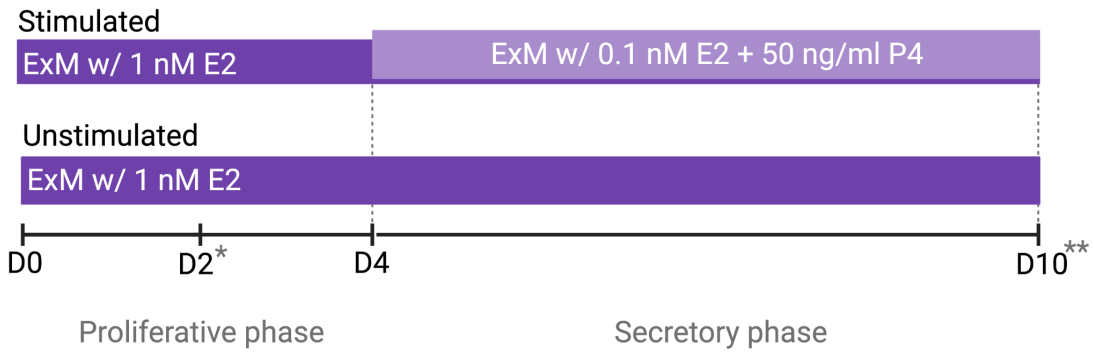


Figure 5.3: Schematic representation of *in vitro* hormonal stimulation. Both organoid cultures and organoid-stromal co-cultures were established to grow under stimulated and unstimulated conditions in parallel for 10 days. Stimulated cultures received ExM with 1 nM E2 for 4 days, followed by 6 days of ExM supplemented with 0.1 nM E2 and 50 ng/ml P4 to mimic the proliferative and secretory phases of the menstrual cycle. Unstimulated cultures received ExM with 1 nM E2 for 10 days. D0 denotes day zero when cultures were set-up. D10** = cell collection for scRNAseq. D2* = 10 µM Y-27632 added for the initial 2 days for epithelial organoid cultures. *Abbreviations*: ExM, expansion medium; w/, with; E2, 17-β oestradiol; P4, progesterone, D, day.

5.3.5 Single-cell RNA-sequencing of organoids and co-cultures

5.3.5.1 Single cell suspension preparation

To obtain single cell suspensions of epithelial organoids and organoid-stromal cell co-cultures, the structures were enzymatically dissociated. In case of organoids, the culture media was removed and 500 µl of ice-cold DMEM/F12 added per well to dislodge the Matrigel domes. Pooling the contents of up to 3 wells into 5 ml centrifuge tubes, these were centrifuged (600 x g, 6 min, 4°C), supernatant discarded and resulting cell pellet resuspended in up to 2 ml of pre-warmed TrypLE and placed on a MACSMix Rotator for 25 min at 37°C. Tubes were checked and vigorously shook after 15 min to help dissociate the 3D structures. At 20 min, any remaining structures were broken down by manually pipetting the solution up/down 50-70 times using 200 µl narrow-bore low-retention tips. Tubes were incubated for a further 5 min at 37°C while rotating and the dissociation stopped by adding 2-3 ml of ice-cold DMEM/F12. Tubes were centrifuged (600 x g, 6 min, 4°C), supernatant discarded and cells resuspended in ~200 µl of 0.04% BSA/PBS and filtered through 40 µm Flowmi® cell strainers and counted. For the hydrogel co-cultures, the same protocol was used with the following modifications: (i) wells were scraped back and forth to dislodge the hydrogels prior to transferring them to a tube, (ii) up to 4 wells were pooled per tube, (iii) centrifuge setting was 300 x g, 5 min, 4°C, (iv) Trypsin and DNase I (final concentration 0.1 mg/ml) were used for dissociation and (v) the dissociation was stopped after 18 minutes followed by pipetting the suspension up/down 20-30 times.

5.3.5.2 Cell loading and target recovery

Organoids were processed in 3 sequencing batches, pooling 3 donors per 10x reaction in all batches, plus 2 donors for one further 10x reaction in batch 3, when 5 donor lines were processed (Table 5.5). Pooling was done on a per experimental condition basis - stimulated and unstimulated cells separately. 10x chips were loaded with targeted viable cell recovery of 12,000 per reaction (i.e. 4,000 cells per donor per condition) when 3 donors were pooled, and 10,000 per reaction in case of pooling 2 donors (i.e. 5,000 cells per donor per condition). For the organoid-stromal co-cultures, cells were processed in 2 batches (Table 5.6). In batch 1, only one donor line was processed and targeted cell recovery was 10,000 viable cells. In batch 2, 3 donors were pooled across four 10x reactions, each with targeted cell recovery of 12,000 viable cells, thus 8,000 cells per donor per condition. To demultiplex the different donors, genotype information was used, but in case this information was not available, cells were labelled prior to loading the 10x chip using TotalSeq™ antibodies as described in Chapter 3, section 3.3.3.3. The antibody concentrations were adjusted to the cellular yields on the day. With regards to the 10x Genomics dual index v3.1 libraries, these were prepared and sequenced as described in Chapter 3, section 3.3.3.4.

Table 5.5: Epithelial organoid lines profiled using scRNAseq.

Batch	Line ID	Metadata	Hormones	Passage	*In vivo* data
1	FX1119	Control	No	P2	Yes
1	FX1254	Endometriosis II	Yes	P2	Yes
1	FX1259	Endometriosis II	No	P2	Yes
2	FX1268	Endometriosis II	No	P4	No
2	FX1260	Control	Yes	P4	No
2	FX1146	Control	No	P4	Yes
3	FX1294	Control	Yes	P3	No
3	FX1125	Endometriosis III	No	P3	Yes
3	SE03	Control	No	P3	Yes
3	SE02	Control	No	P3	Yes
3	FX1289	Endometriosis I	No	P3	Yes

Metadata refers to being controls (i.e. donors without endometriosis) or patients with endometriosis and respective rASRM endometriosis stages. Hormones refers to patients taking hormonal therapy at sample collection. Column *In vivo data* highlights for which donor lines the primary tissue was also profiled at the single-cell level. **Abbreviations:** P, passage; rASRM, revised American Society for Reproductive Medicine staging system; scRNAseq, single-cell RNA-sequencing.

Table 5.6: Organoid-stromal cell co-cultures profiled using scRNAseq.

Batch	Line ID	Metadata	Passage		*In vivo* data
			Organoids	Stromal	
1	FX1146	Control	P6	P6	Yes
2	FX1294	Control - hormones	P3	P5	No
2	FX1125	Endometriosis III	P3	P4	Yes
2	SE03	Control	P3	P4	Yes

Metadata refers to being controls (i.e. donors without endometriosis) or patients with endometriosis and respective rASRM endometriosis stages. Hormones refers to a patient taking hormonal therapy at sample collection. Column *In vivo data* highlights for which donor lines the primary tissue was also profiled at the single-cell level. **Abbreviations:** P, passage; rASRM, revised American Society for Reproductive Medicine staging; scRNAseq, single-cell RNA-sequencing.

5.3.6 Data analysis - *in vitro* data specifics

Both the upstream and downstream data analyses were performed following the same computational pipelines as described in Chapter 3, section 3.3.3 but with a few modifications. First, samples were integrated using *Harmony*[225], correcting for passage number in the case of the epithelial organoids dataset and for genotype when analysing the organoid and stromal co-culture dataset. Second, differential gene expression analyses were performed using the Wilcoxon rank-sum method, provided as a parameter for the *scanpy.tl.rank_genes_groups* function from the *Scanpy* package. The top 3 differentially expressed genes with *p*-value < 0.001 were reported. Lastly, to annotate the organoids cells using the scRNAseq *in vivo* endometrial reference established in Chapter 3 and 4, a regularised logistic regression approach was used as described in Garcia-Alonso *et al.*[98]. Briefly, only a limited number of cell states (considered most robust) for the epithelial organoids analyses were considered and included the ciliated, glandular, glandular_secretory, luminal, MUC5B, proliferative, proliferative_LGR5 and TENM2 populations. For the organoids and stromal cell co-cultures the eS, dS and intermediate cell states were also included. To limit the influence of cell cycle, the cycling cell states in the *in vivo* reference dataset and the G2/M and S genes from *Seurat* were excluded. Subsequently, both the *in vivo* and *in vitro* datasets were subsetted to their shared highly variable genes. Gene expression was log-transformed and normalised by the maximum RNA expression for both datasets. The model was trained with the *in vivo* epithelial and stromal cell identities, with 10,000 iterations, and used to classify the EEO and EEO-ES cells.

5.4 Results

I processed 27 samples for organoid derivation with a line establishment success rate of 93%. Overall, I established 25 organoid and 16 stromal cell lines during my studies (Appendix 6) and adapted the existing methods for freezing and thawing of the organoid lines as described in the Methods section of this chapter. Moreover, I successfully established the co-culturing of matched organoid and stromal cell lines in synthetic hydrogels (n = 4). The next sections describe single-cell transcriptomic profiling of both the organoid and co-culture systems.

5.4.1 Endometrial epithelial organoids (EEOs)

In total, 11 EEO lines underwent scRNAseq analyses upon culturing with and without hormonal stimulation mimicking the menstrual cycle (Figure 5.4B). The endometrial samples came from donors with/without endometriosis (all from proliferative phase) and those taking exogenous hormones at sample collection. Here, I present the transcriptomic data for 8 of these EEO lines (Table 5.7).

Table 5.7: Overview of organoid lines and cell numbers analysed in Chapter 5.

Line ID	Control Endometriosis Hormones	Unstimul	Stimul	Total cell number
FX1146	C	2,846	1,125	3,971
SE02	C	6,971	6,827	13,798
SE03	C	4,815	2,757	7,572
FX1125	E	3,345	3,436	6,781
FX1268	E	2,777	1,823	4,600
FX1289	E	10,091	8,988	19,079
FX1260	C-H	4,360	2,701	7,061
FX1294	C-H	4,243	4,307	8,550
Total	*N/A*	*39,448*	*31,964*	*71,412*

Abbreviations: C, control; E, endometriosis; C-H, control taking exogenous hormones at sample collection; Stimul, stimulated; Unstimul, unstimulated.

5.4.1.1 scRNAseq data generation and QC

Dissociating EEOs into single cells using my adapted protocol yielded highly viable cells (75-90%) and in total, 71,412 high QC cells were generated from the 8 EEO lines: 39,448 cells (55%) were from the unstimulated group and 31,964 cells (45%) from the stimulated group (Table 5.7).

Morphologically, at day 10 the EEOs in the stimulated group exhibited a different phenotype to those in the unstimulated group (Figure 5.4A). The stimulated organoids appeared darker, had thicker lumen and their shape was distorted, i.e. less round and regular than that of the EEOs in the unstimulated group. This did not indicate unhealthy/dying organoids but rather the morphological response to the hormonal stimulation, as upon dissociation into single cells the viability was comparable between the groups (data not shown). In addition, the stimulation protocol (Figure 5.4B) did not alter the QC metrics of these two groups with respect to the number of genes, mRNA molecules and mitochondrial gene expression per cell detected for the unstimulated and stimulated organoids (Figure 5.4C). A slight increase in the percentage of mitochondrial genes per cell was observed in the stimulated group (Figure 5.4C).

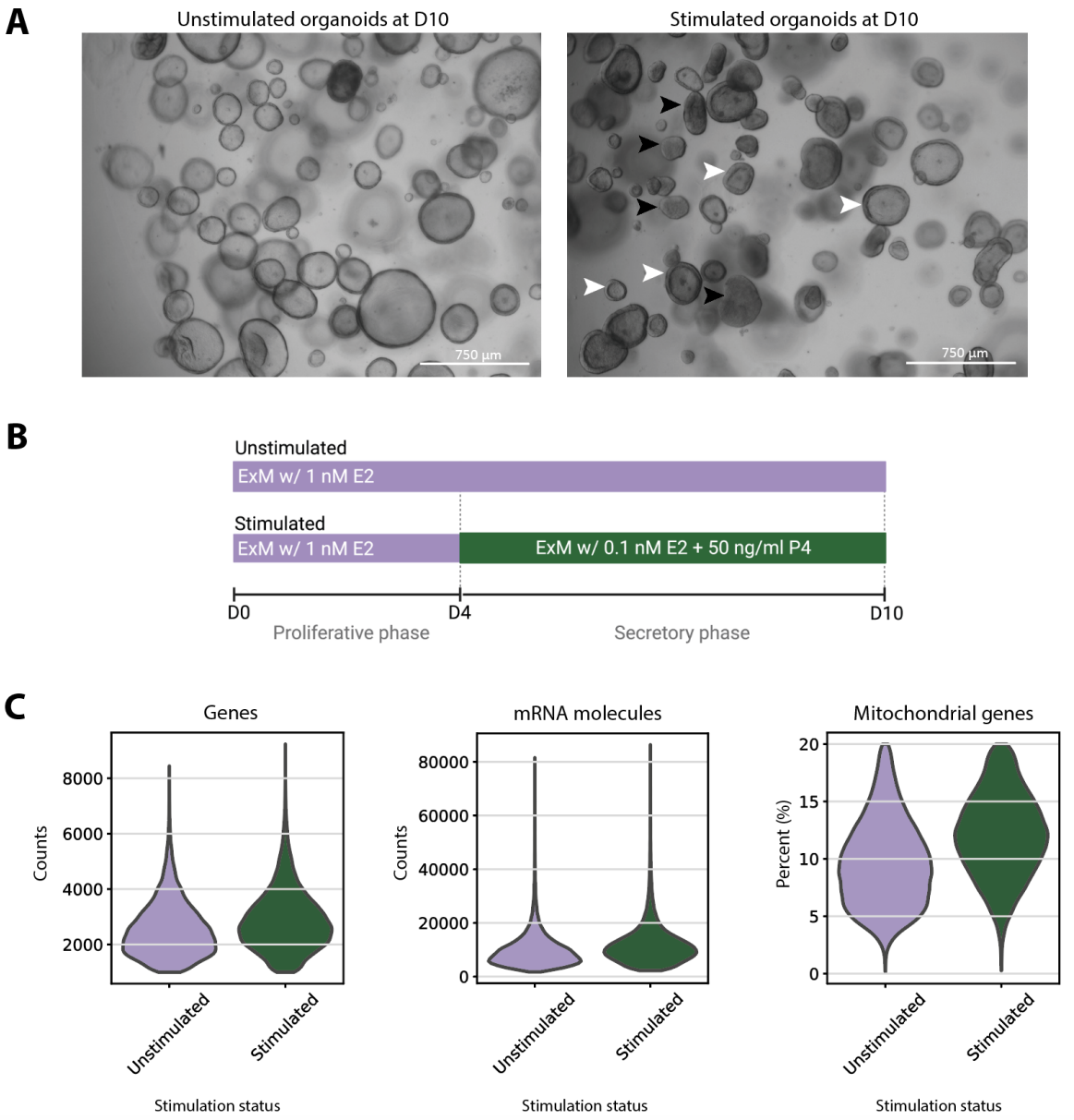


Figure 5.4: Hormonal stimulation of endometrial epithelial organoids. Representative images of organoids at day 10 for unstimulated and stimulated conditions are shown in (A). White arrows highlight example organoids with a thick lumen, black arrows highlight dark organoids. (B) the culture timelines and conditions for the unstimulated and stimulated groups. Violin plots in (C) represent the data distribution for number of genes, number of mRNA molecules and percentage of mitochondrial gene expression on a per cell basis for each of the culture conditions. *Abbreviations*: D, day; E2, 17-β oestradiol; P4, progesterone; w/, with.

5.4.1.2 Endometrial organoids at the single-cell level

Cell state annotation

In total, I defined 14 clusters and cell states when analysing the EEO data and grouped them under 3 main categories considering whether they were predominantly detected in the unstimulated, stimulated or both unstimulated and stimulated organoids (Figure 5.5A). The next sections describe the cell states detected in these three groups in more detail.

Unstimulated & Stimulated group

The ciliated, pre-ciliated, cycling, KRT17/inflammatory and proliferative_LGR5-like cells states were detected in both unstimulated and stimulated groups in similar numbers (Figure 5.5C). The marker genes for the ciliated and pre-ciliated states were the same as those described for *in vivo* endometrium in Chapters 3 & 4 (Figure 5.5D). The cycling cell population was identified through visualisation of the cell cycle phase scores (Figure 5.5B) and expression of genes such as *H2AFZ*, *HMGN2* and *TUBA1B*. The KRT17/inflammatory cell state was assigned its identity based on the expression of markers such as *IL32*, *PLAU* and *KRT17* described previously in EEOs[98]. The proliferative_LGR5-like cell state expressed marker genes characteristic of the proliferative_LGR5 population described in Chapters 3 and 4 (i.e. *WNT7A*, *LGR5* and *ITIH5*) and was thus assigned this name and identity (Figure 5.5D).

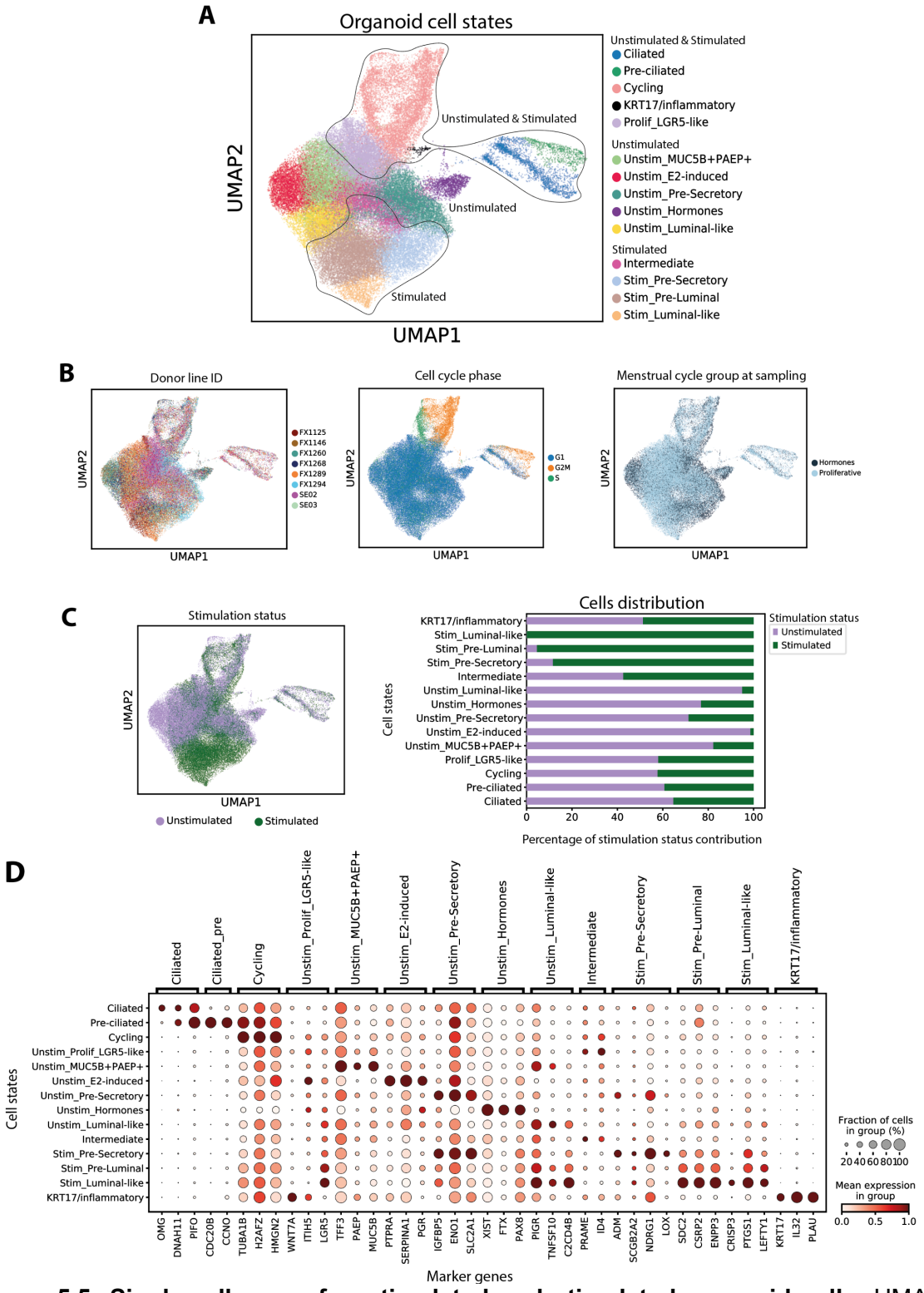


Figure 5.5: Single-cell map of unstimulated and stimulated organoid cells. UMAP projections of the data coloured by the main cells states identified (A), in (B) by donor line ID, cell cycle phase and menstrual cycle phase at sampling. In (C) a UMAP projection coloured by the stimulation status and the percentual contribution of the unstimulated and stimulated organoid cells to the main cell states identified is shown. The dotplot in (D) shows the variance-scaled, log-transformed expression of marker genes (x-axis) characteristic for each of the cell states (y-axis). *Abbreviations*: Stim, stimulated; UMAP, uniform manifold approximation and projection; Unstim, unstimulated.

Unstimulated group

Five cell states were detected at higher proportions (between ~75-98%) in the unstimulated organoids than in the stimulated ones (Figure 5.5C). Firstly, the unstimulated_MUC5B+PAEP+ population expressing *MUC5B*, *PAEP* and *TFF3* (Figure 5.5D) found in all donor lines, but with the FX1289 line contributing ~60% of all cells in this population. Secondly, the unstimulated_E2-induced population expressed genes typical of response to oestrogen (*PGR*, *SERPINA1*, *PTPRA*) previously reported in EEOs[98]. Thirdly, the expression of *SLC2A1*, *ENO1* and *IGFBP5* was characteristic of the unstimulated_Pre-secretory population with similar expression of these markers seen in the stimulated_Pre-secretory cell state (Figure 5.5D). The fourth population was the unstimulated_Hormones group assigned its name as a result of ~50% of this population originating from the two donors that were taking exogenous hormones at sample collection (Figure 5.5B). Marker genes for this population included *PAX8*, *XIST* and *FTX*. Lastly, the unstimulated_Luminal-like population expressed *PIGR*, *CSCD48*, and *TNFSF10* but also *CRISP3*, *PTGS1* and *LEFTY1* to a much lower degree than the stimulated_Luminal-like population (Figure 5.5D).

Stimulated group

Four cell states were identified as specific to the stimulated organoids, including the intermediate population (expressing *PRAME* and *ID4*) with ~60% of this population found in the stimulated organoids (Figure 5.5C). ~90% of the stimulated_Pre-secretory population (expressing *LOX*, *ADM*, *NDRG1* and *SCGB2A2*) was contributed by the stimulated organoids. The stimulated_Pre-luminal and stimulated_Luminal-like populations were made of 95% and 100% of cells from the

stimulated organoids, respectively. Both populations expressed luminal cell markers of *in vivo* endometrium (e.g. *CRISP3*, *LEFTY1*, *PTGS1*) with high expression observed in the stimulated_Luminal-like cells and a lower expression seen in the stimulated_Pre-Luminal cells (Figure 5.5D). Other markers included *CSRP2*, *ENPP3* and *SDC2* also showing the same pattern of expression.

Endometrial epithelial cells in vitro and in vivo

In order to assign the *in vitro* cell state identities I relied on marker gene expression that I established for the different cell states in Chapters 3 and 4, but also logistic regression predictions (Figure 5.6). Using the transcriptomic signatures of the main *in vivo* endometrial cell states (ciliated, glandular, glandular_secretory, luminal, MUC5B, proliferative, proliferative_LGR5 and TENM2), the logistic regression model assigned the ciliated, proliferative, proliferative_LGR5, luminal and glandular cell identities with high probabilities (Figure 5.6B). However, the predicted glandular and luminal population overlapped to a large extent. In addition, the MUC5B population was identified as a separate cluster, but the probability of this assignment was low (Figure 5.6B). Taken together, the data shows that the ciliated, proliferative, proliferative_LGR5, and luminal populations *in vitro* are transcriptomically highly similar to those observed *in vivo*. The glandular_secretory, MUC5B and TENM2 populations could not be detected/were not predicted and thus are likely not recapitulated in the EEO *in vitro* model system presented in this chapter.

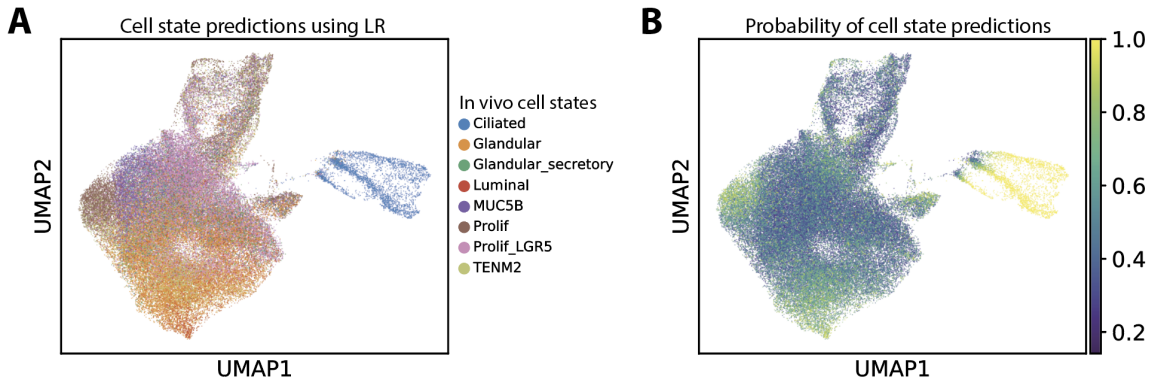


Figure 5.6: *In vivo* to *in vitro* endometrial cell state predictions. UMAP projections of the organoid data with cells coloured by the predicted *in vivo* reference cell states using a logistic regression model (A). In (B) cells are coloured by the probabilities with which the cell states were assigned. *Abbreviations*: LR, logistic regression; Prolif, proliferative; UMAP, uniform manifold approximation and projection.

5.4.1.3 Organoids from donors with/without endometriosis and taking hormones

Even though good mixing of the cells from the different donor lines was observed, some heterogeneity and clusters specific to certain groups of donors could be seen (Figure 5.5B). Next, I therefore explored whether there were any differences between the EEOs from donors with/without endometriosis after the 10-day culture period (Figure 5.4B), in both the unstimulated and stimulated conditions. Secondly, seeing the slight differences in cell state presentation in EEOs from donors taking exogenous hormones at endometrial biopsy collection, I also looked at better characterising the cellular composition of EEOs from these donors. The next sections describe the results obtained.

Organoids from donors with/without endometriosis

Initially, I considered all of the 8 EEO lines when looking at the cellular composition of organoids from endometriosis cases and controls (Figure 5.7A) and observed a higher proportion of the proliferative_LGR5-like, unstimulated_E2-induced and unstimulated_Pre-secretory populations in EEOs from controls (i.e. donors without

endometriosis). The EEOs from endometriosis cases on the other hand had a higher proportion of the unstimulated_MUC5B+PAEP+, unstimulated_Luminal-like, stimulated_Pre-luminal and stimulated_Luminal-like cells (Figure 5.7A). In order to remove any potential heterogeneity introduced by including EEOs from the two donors taking exogenous hormones at sample collection, I next removed these two donor lines from the analyses. In addition, I also divided the dataset into unstimulated EEOs and stimulated EEOs (Figure 5.7B & C) and analysed these separately to understand the differences between control (*n* = 3) and endometriosis (*n* = 3) EEOs when (i) cultured for 10 days in media with low amount of E2 (unstimulated) and (ii) cultured for 10 days in media with low amount of E2 and P4 (stimulated).

Unstimulated EEOs

In the unstimulated conditions, 25% of the EEO cells from endometriosis cases were the unstimulated_MUC5B+PAEP+ state which was higher than in control EEOs (~10% of all cells). Moreover, the proportion of the unstimulated_Luminal-like population was higher in endometriosis EEOs (~20%) when compared to controls (~7%). In contrast, the control EEOs consisted of ~20% proliferative_LGR5-like cells while this cell state contributed ~10% in endometriosis cases. Moreover, a higher proportion of the unstimulated_E2-induced and unstimulated_Pre-secretory populations were characteristic of control EEOs (Figure 5.7B).

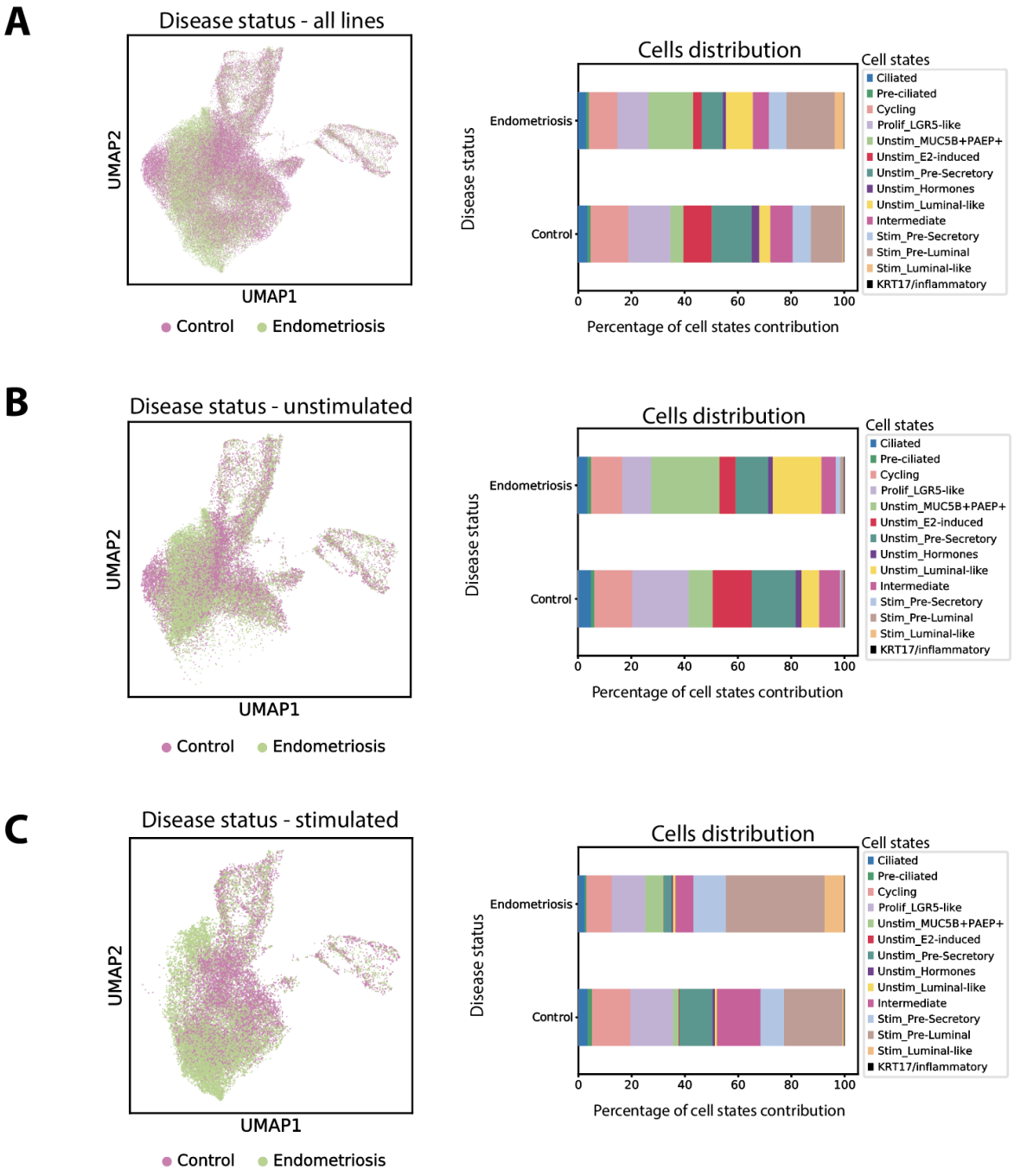


Figure 5.7: EEOs in endometriosis cases and controls. Shown are UMAP projections of the organoid data coloured by disease status. In (A) data for all 8 EEO lines is shown together with the percentual contribution of each cell state identified to EEOs from controls and endometriosis cases. In (B) and (C) data for 6 EEOs is shown - 2 lines from donors taking exogenous hormones have been removed. The cellular composition of EEOs in controls and endometriosis cases is shown for the unstimulated condition (B) and stimulated condition (C). *Abbreviations*: EEO, endometrial epithelial organoids; Prolif, proliferative; Stim, stimulated; UMAP, uniform manifold approximation and projection; Unstim, unstimulated.

Stimulated EEOs

Exposing the EEOs to progesterone led to changes in their cellular composition (Figure 5.7C). In the endometriosis EEOs, the unstimulated_MUC5B+PAEP+ population decreased in its numbers with only ~7% of the total cells assigned this identity. In controls this number was reduced to ~2%. Moreover, in both endometriosis and control EEOs the unstimulated_Luminal-population was almost absent but instead an increase in the stimulated_Pre-luminal population was observed, making up ~37% of endometriosis EEOs and only ~20% of control EEOs. The control EEOs were also characterised by a higher proportion of unstimulated_Pre-secretory (~15%) and the intermediate population (~18%) when compared to endometriosis cases (~3% and ~5%, respectively). Lastly, the stimulated_Luminal-like population was observed to a higher extent in endometriosis EEOs at ~7% as compared to ~1% in controls.

Gene expression differences

Next, I considered the differences between cases and controls at the gene expression level. I performed differentially expressed gene (DEG) analyses for both the unstimulated and stimulated conditions separately. In both the unstimulated and stimulated EEOs, higher expression of e.g. *MMP7* and *UGT2B7* was observed in controls when compared to cases. In EEOs from endometriosis cases higher expression of *TFF3*, *SELENOP* and *CP* was detected (Figure 5.8). In addition, in the stimulated condition, EEOs from controls had higher expression of *SERPINA3*.

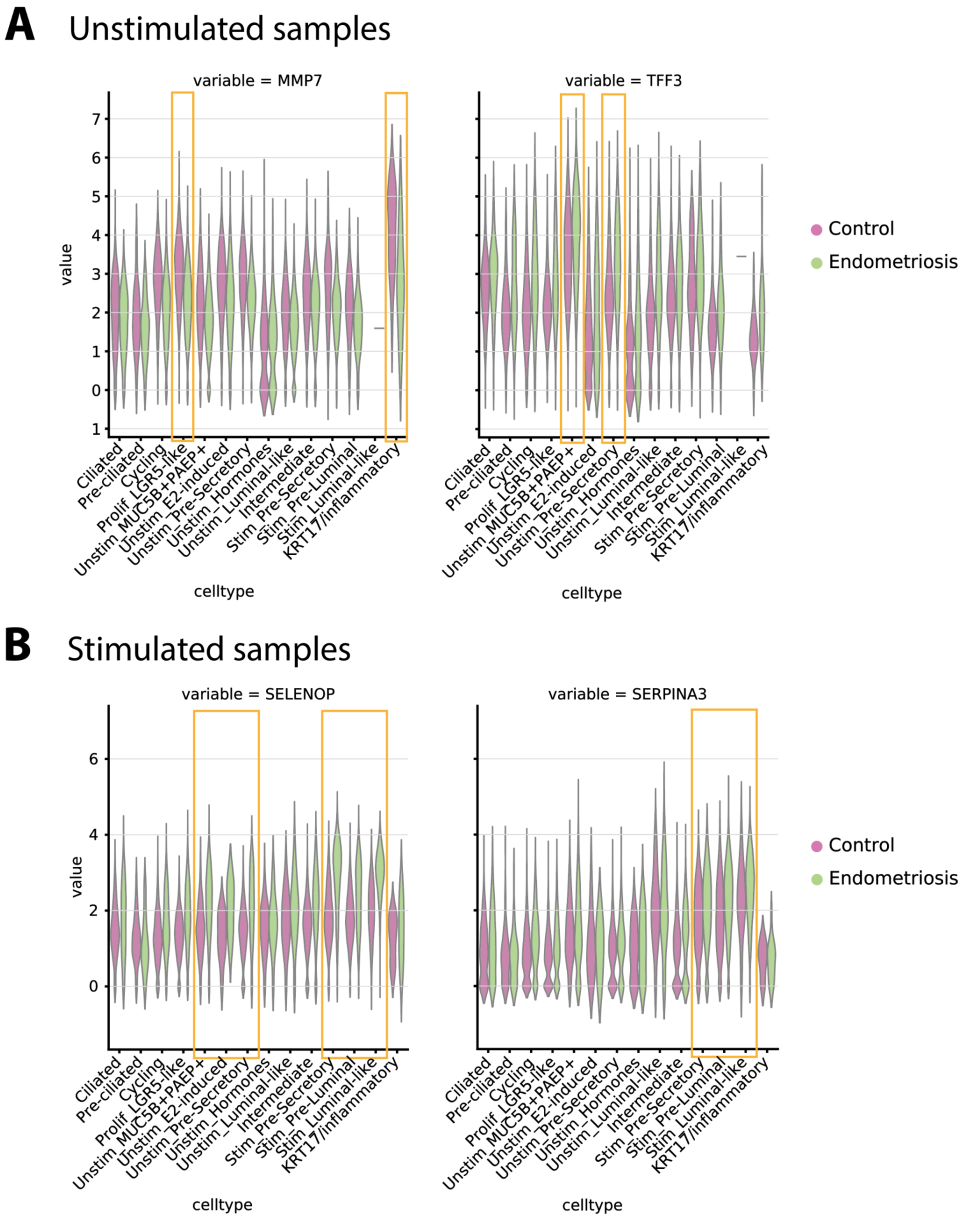


Figure 5.8: Gene expression profiles of unstimulated and stimulated EEOs. Expression levels of example genes that were found to be differentially expressed between cases and controls in the unstimulated condition (A) or stimulated condition (B) are shown as violins for each cell state. Violin colours represent the disease status. Orange rectangles highlight cell state-specific differences in gene expression between control and endometriosis EEOs. *Abbreviations*: EEO, endometrial epithelial organoids; Prolif, proliferative; Stim, stimulated; Unstim, unstimulated.

Looking at the gene expression difference on a cell state level showed that e.g. expression of *MMP7* is indeed higher in control EEOs overall, with biggest differences in expression between cases and controls observed in the proliferative_LGR5-like and KRT17/inflammatory cell states (Figure 5.8A). With

regards to the higher expression of *TFF3* in endometriosis cases, more specifically the biggest differences were detected in the unstimulated_MUC5B+PAEP+ and unstimulated_Pre-secretory populations (Figure 5.8A). In the stimulated condition, higher expression of *SELENOP* and *SERPINA3* in EEOs from endometriosis cases was seen in the stimulated_Pre-Secretory, stimulated_Pre-Luminal and stimulated_Luminal-like cell populations (Figure 5.8B).

Organoids from donors taking exogenous hormones

To explore whether taking exogenous hormones at sample collection affects the cellular composition of the derived EEOs, I included three such samples in my experiments. Due to time constraints, here I only present data for two of these samples: one from a donor taking the progestin-releasing pill Desogestrel and one from a donor taking the progestin-releasing pill Norethisterone (Figure 5.9A). While cells from these donors clustered with those of not taking any exogenous hormones at sample collection, it can be observed that a predominant feature of the EEOs from the donor taking Norethisterone is a high proportion (~28% of all cells) of the unstimulated_E2-induced population and the unstimulated_Hormones population (~10%). This is in contrast to the EEOs from a donor taking Desogestrel, characterised by the presence of the unstimulated_Pre-Secretory population (~21%) and the stimulated_Pre-Secretory state (18%). These differences may be due to the fact that these two hormonal treatments have varied effects on the endometrium (Desogestrel highly suppresses ovulation, leading to amenorrhea, while Norethisterone does not exhibit such profound effects[207], suggesting that these effects may have persisted in the *in vitro* culture and affected the response of EEOs to hormonal stimulation.

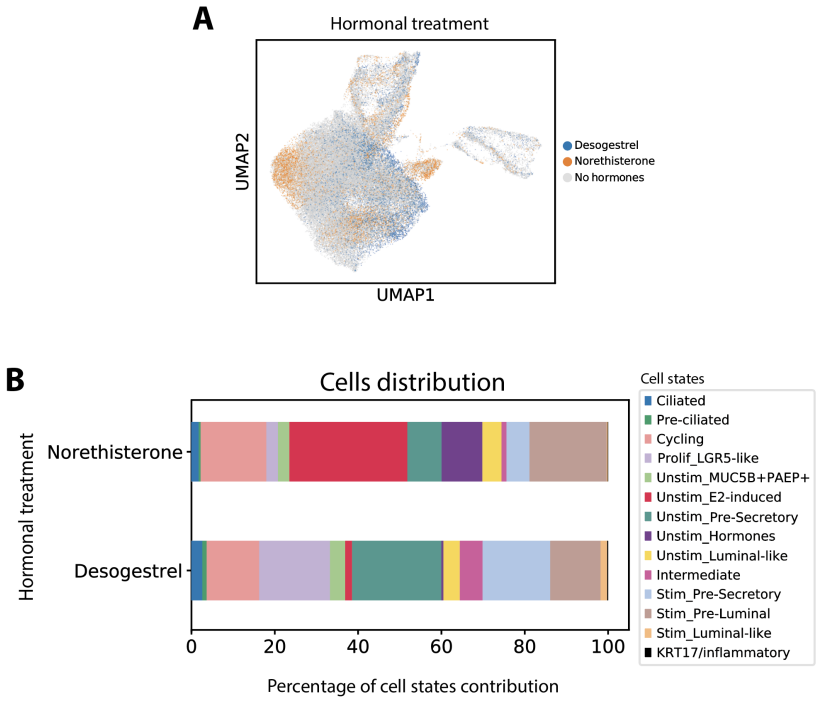


Figure 5.9: EEOs from donors taking exogenous hormones. Shown is an UMAP projection of the data coloured by hormonal treatment taken at sample collection (A). In (B) the cellular composition of samples is shown for the two hormonal treatment groups. *Abbreviations*: EEO, endometrial epithelial organoid; Prolif, proliferative; Stim, stimulated UMAP, uniform manifold approximation and projection; Unstim, unstimulated.

5.4.2 Endometrial epithelial organoids and stromal cell co-cultures

I established four endometrial epithelial organoid and stromal cell co-cultures (EEO-ES) and optimised a protocol for their dissociation into single cells. Upon culturing with and without hormonal stimulation mimicking the menstrual cycle as described above for EEOs, the four EEO-ES lines underwent scRNAseq analyses. The matched EEO and ES cell lines came from donors with/without endometriosis and those taking exogenous hormones at sample collection (Table 5.8).

5.4.2.1 scRNAseq data generation and QC

Cell viability upon tissue dissociation was high (~85-90%) and in total, the data for 46,217 high QC cells are presented in this chapter (Table 5.8). 55% of all cells

analysed came from the unstimulated cultures and 45% from the stimulated ones. Cell recovery for line FX1146 processed in the first batch was ~10x lower than for the rest of the lines processed in batch two, indicating an issue with cell encapsulation and capture inside the 10x Chromium platform (Table 5.8).

Table 5.8: Overview of EEO-ES lines data is presented for in this chapter.

Line ID	Control Endometriosis Hormones	Unstimul	Stimul	Total cell number
FX1146	C	731	919	1,650
SE03	C	9,536	5,596	15,132
FX1125	E	7,015	7,133	14,148
FX1294	C-H	7,931	7,356	15,287
Total	*N/A*	*25,213*	*21,004*	*46,217*

Abbreviations: C, control; C-H, control taking exogenous hormones at sample collection; E, endometriosis; EEO-ES, endometrial epithelial organoid and stromal co-culture; Stimul, stimulated; Unstimul, unstimulated.

Morphologically, at day 10 the EEOs in the EEO-ES co-cultures exhibited a similar phenotype in both the unstimulated and stimulated groups (Figure 5.10A). The EEOs had a thick lumen, and exhibited more folding and tortuosity than EEOs cultured on their own in Matrigel only, as described in the previous sections. Interestingly, the EEOs formed gland-like structures resembling those observed *in vivo* and even lumen-like structures when the EEOs expanded and grew at the periphery of the hydrogel (Figure 5.10Aa). The high viability upon dissociation confirmed the structures formed were healthy and not necrotic even though their shape and darker appearance differed from the EEOs cultured on their own.

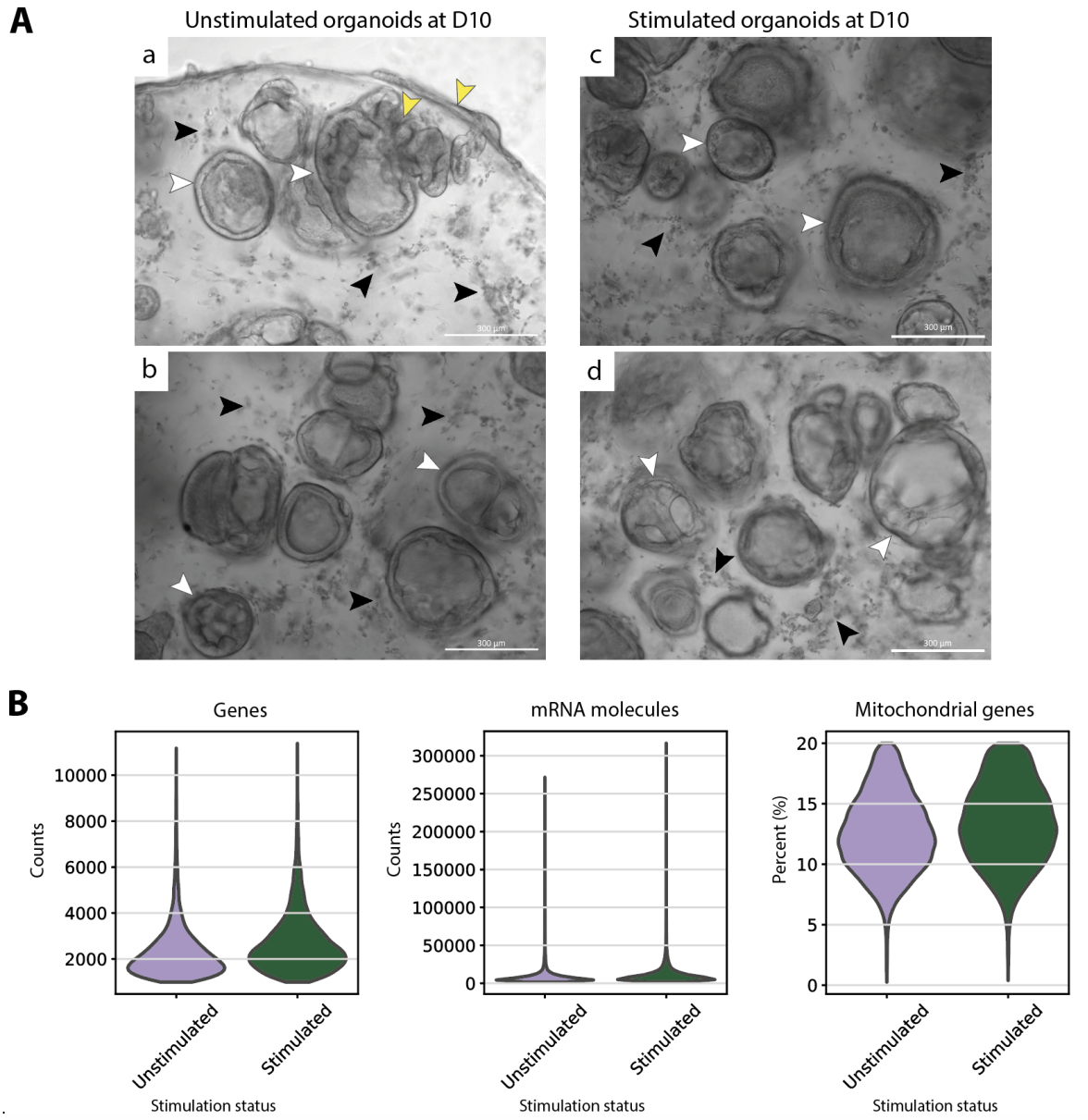


Figure 5.10: Hormonal stimulation of EEO-ES lines. Representative images of organoids at day 10 for unstimulated and stimulated conditions are shown in (A). White arrows highlight the EEOs, black arrows ES cells and yellow arrows in Aa show a folded and tortuous EEO, creating a lumen-like structure on the periphery of the hydrogel. In Ab and Ad organoids forming invaginations and gland-like structures can be seen. Violin plots in (B) represent the data distribution for number of genes, number of mRNA molecules and percentage of mitochondrial gene expression on a per cell basis for each of the culture conditions. *Abbreviations*: D, day; EEO, endometrial epithelial organoids; ES, endometrial stromal.

With regards to scRNAseq data QC, the number of genes, mRNA molecules and mitochondrial gene expression per cell detected for the unstimulated and stimulated EEO-ES was comparable (Figure 5.10B), with only a slight increase in the number

of genes per cell observed in the stimulated group (Figure 5.10B). Of note is that while the mitochondrial gene expression per cell detected for the majority of the cells was within the acceptable range (up to 20%), there was an increase in this metric in the EEO-ES cells when compared to EEOs only (Figure 5.4C). This was likely caused by the dissociation of the EEO-ES cultures and hydrogels with trypsin.

5.4.2.2 Cell state annotation of the co-culture system

Overall, good mixing of the donor lines was achieved through data integration (Figure 5.11B). I defined 15 cell states when analysing the EEO-ES data, grouping these under two main categories: epithelial and stromal cells (Figure 5.11A). The next sections describe the cell states detected in each of the groups, their marker gene expression and resemblance to *in vivo* endometrial cells.

Epithelial cells

The ciliated cells were identified based on canonical marker gene expression described throughout this thesis. Due to the low number of ciliated cells in the EEO-ES system, more granular annotation of this population was not performed. The cycling_epithelial cells were identified based on the cell cycle phase score (Figure 5.11B) and expression of markers described in Chapter 4 (Figure 5.11D). Utilising marker gene expression reported previously and in this thesis, the proliferative-like_SOX9+LGR5- population was assigned its name and identity. It expressed *SOX9*, lacked *LGR5* and in this system highly expressed *FGF18*. The E2-response, unstimulated_MUC5B+PAEP+ and KRT17/inflammatory populations were identified based on the marker genes I defined in the EEO system (Figure 5.11D). This suggests that these populations and markers are robust and present in both

the EEO and EEO-ES culture systems. The unstimulated_Luminal-like cell state shared some marker gene expression with the Luminal/Secretory-like population, but at much lower levels and uniquely expressed *STC2*, *IGFBP5* and *SLC2A1.* The Luminal/Secretory-like population expressed markers of both the luminal (*LEFTY1*, *PTGS1, LGR5*) and secretory cells (*SCGB1D2*) and was thus assigned a name reflecting its luminal and secretory phenotype.

Stromal cells

As described for the cycling_epithelial cells, cycling_stromal cells were identified based on their cell cycle phase score and expression of markers related to the cell cycle (Figure 5.11B & D). The rest of the stromal cells appeared along a continuum with shared marker gene expression and expression of some distinct markers that were used for naming the clusters identified. The eS_THY1+ cell state expressed mesenchymal stem-cell markers such as *THY1*, *PDGFRB*, *NOTCH3* but also markers of perivascular cells (*RGS5*, *MYO1B* and *PAG1*) described in Chapters 3 and 4. The eS_CPE+ population highly expressed *CPE*, *MMP11*, *TDO2* and to a lesser extent *RGS5*. The eS_CXCL12+ cells were characterised by the expression of *CXCL12*, *CCDC80* and *OSR2* (Figure 5.11D). The eS_IGBP6+ population expressed *IGFBP6*, *WLS* and *SFRP1* while the eS_ISG15+ cells uniquely expressed *ISG15*, *MX1*, *IFIT1*, and *IFIT3*. The population expressing markers expressed by decidualised stromal cells (e.g. *MMP2*, *PTGDS, FGF7*) was named dS-like but also highly expressed *CRABP2*, a marker of proliferative phase eS cells (Figure 5.11D).

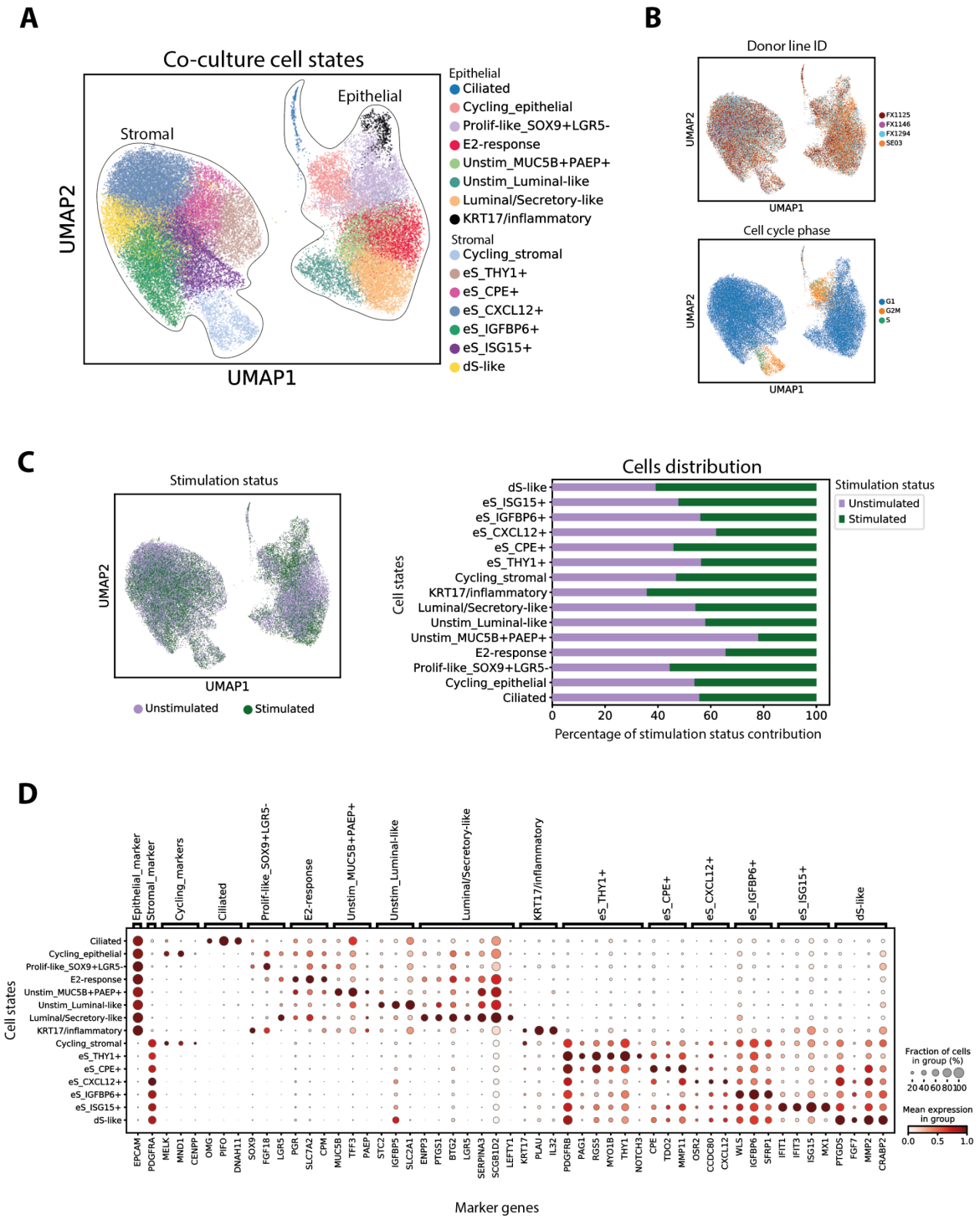


Figure 5.11: Single-cell map of unstimulated and stimulated EEO-ES cultures. Shown are UMAP projections of the data coloured by the main cell states identified (A), in (B) by donor line ID and cell cycle phase. In (C) an UMAP projection coloured by the stimulation status and the percentual contribution of the unstimulated and stimulated EEO-ES cells to the main cell states identified is shown. The dotplot in (D) shows the variance-scaled, log-transformed expression of marker genes (x-axis) characteristic for each of the cell states (y-axis). *Abbreviations*: dS, decidualised endometrial stromal; eS, non-decidualised endometrial stromal; Prolif, proliferative; Stim, stimulated; UMAP, uniform manifold approximation and projection; Unstim, unstimulated.

Unstimulated vs stimulated EEO-ES cells

Comparing the cellular composition of EEO-ES in the unstimulated and stimulated conditions did not show notable differences with most cell states present in both the stimulated and unstimulated groups at proportions expected based on the total number of cells profiled (Figure 5.11C). Only the unstimulated_MUC5B+PAEP+ population consisted of a larger proportion of cells (~78%) from the unstimulated group. Similarly, the E2-response population with ~62% of all its cells coming from the unstimulated group decreased in its numbers upon treatment with progesterone. In contrast, treatment with progesterone led to an increase in the dS-like cells, with ~60% of this population coming from the stimulated group.

Endometrial epithelial and stromal cells in vitro and in vivo

To assign the *in vitro* cell state identities, I considered marker gene expression established for the different cell states in Chapters 3 and 4, literature search, but also logistic regression predictions (Figure 5.12). Using the transcriptomic signatures of the main *in vivo* endometrial epithelial cell states (ciliated, glandular, glandular_secretory, luminal, MUC5B, proliferative, proliferative_LGR5, TENM2), and endometrial stromal cells (dS, eS and intermediate) the logistic regression model assigned the ciliated, proliferative_LGR5, luminal and eS cell identities with high probabilities (Figure 5.12B). Similarly, in the EEO system, the MUC5B population was identified as a separate cluster, but the probability of the assignment/prediction was low (Figure 5.12B). Taken together, the data showed that the ciliated, proliferative_LGR5, luminal and eS populations *in vitro* were transcriptomically highly similar to those observed *in vivo*. The remaining cell populations queried were not predicted to be present in the EEO-ES system, either

at all or with very low probabilities, highlighting differences between the transcriptomic profiles of *in vivo* and *in vitro* endometrial cell states.

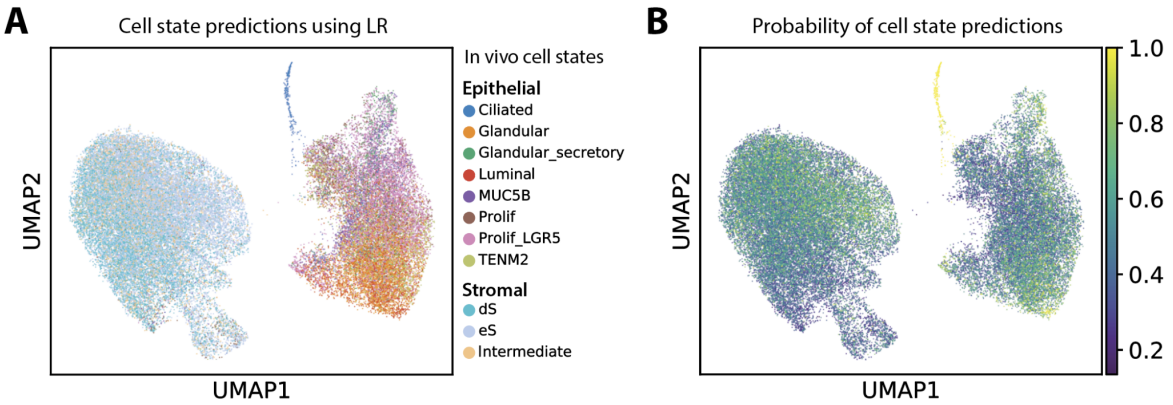


Figure 5.12: *In vivo* to *in vitro* endometrial cell state predictions. Shown are UMAP projections of the EEO-ES system data with cells coloured by the predicted *in vivo* reference cell states using a logistic regression model (A). In (B) cells are coloured by the probabilities with which the cell states were assigned. *Abbreviations*: dS, decidualised endometrial stromal; eS, non-decidualised endometrial stromal; LR, logistic regression; Prolif, proliferative; UMAP, uniform manifold approximation and projection.

5.5 Discussion

In summary, in this chapter I presented the scRNAseq data generated for EEOs (*n* = 8) and EEOs co-cultured with stromal cells (*n* = 4), both in the unstimulated and stimulated conditions. The cultured and analysed cells came from endometrial biopsies of donors with/without endometriosis and those taking exogenous hormones, confirming EEOs can be generated successfully from a variety of samples. I described the cellular heterogeneity of both the epithelial and stromal cell lineages of the two model systems. Even though the data from the various donors integrated well, I observed differences between EEOs from donors taking exogenous hormones at sample collection and those that did not. Moreover, differences in the cellular composition of EEOs from donors with/without endometriosis were identified. Some of the findings are discussed in more detail in the below sections.

5.5.1 Cellular composition of EEOs

Cell state annotation

Annotating and assigning names to the different cell populations was challenging due to the lack of a well-defined cellular reference and marker genes for EEOs, as well as the inconsistencies across media composition and stimulation protocols used by the previous studies (as described in this chapter's introduction). These factors made the extrapolating of previous findings difficult. However, I did find that using the logistic regression model to project the *in vivo* reference cellular identities in combination with the cluster-specific marker genes helped me navigate the cellular heterogeneity observed. The names assigned therefore represent where certain cell populations were like some of those *in vivo* (e.g. proliferative_LGR5-like state) or expressed markers typical of a well-defined population (e.g. luminal), but at a lower level, suggesting that they were transitioning towards such a fate (e.g. pre-luminal, pre-secretory states).

The names used here; however, include a substantial amount of inference and speculation based on one final time point of the culture period and require further refinement through analyses of the culture system at different timepoints, ideally utilising approaches that allow direct measurement of how cell fates and lineage are established. Gene editing tools allowing such studies have been described for e.g. lung organoids[226], offering the opportunity to perturb gene expression in order to model development, function and disease phenotypes in tissue-derived organoids. In the future, such approaches could also be applied to EEOs and help better understand and define the cell fate decisions in both the unstimulated conditions and in response to hormonal stimulation.

In comparison to other scRNAseq EEO studies, the cell states that were recapitulated here (based on marker gene expression) included: the pre-ciliated, ciliated, KRT17/inflammatory and E2-induced states[98]. The exact presence of the MUC5B$^+$ cell state reported by Tan *et al.*[99] cannot be determined, as the cellular heterogeneity and marker gene expression of this cell state was not further characterised in the study. The unstimulated_MUC5B+PAEP+ cell state identified in this chapter should now be investigated in my *in vivo* endometrial cells/nuclei reference reported in Chapters 3 and 4, where heterogeneity of the MUC5B population was noted, but not explored further. It may be that a subset of MUC5B+PAEP+ cells also exists *in vivo*, but was not defined. If as suggested by the Tan *et al*. study, MUC5B$^+$ cells are involved in endometrial regeneration and are found in endometriotic lesions, better understanding of this population both *in vivo* and *in vitro* is required, especially, identifying what signals are needed for its maintenance and proliferation. The Tan *et al*. study reported that growing organoids from sorted MUC5B$^+$ cells led to quicker growth and larger organoid size when compared to MUC5B$^-$ cells. However, EEOs also grew well from MUC5B$^-$ cells and later in culture started to express *MUC5B*. This suggests some sort of plasticity of endometrial cells *in vitro*, warranting further investigation. Once again, studies utilising gene editing tools (e.g. through overexpression or knockout of *MUC5B*) could help dissect the role of *MUC5B* expression in EEO formation and endometrial regeneration.

EEOs grown in unstimulated and stimulated conditions

In this chapter, I reported for the first time the single-cell transcriptomic profiles of cells grown in the media defined by Boretto *et al*.[85] and an adapted version of their hormonal stimulation protocol. Unsurprisingly, the response and transcriptomic profiles of EEOs reported in this chapter differed from those reported in the Garcia-Alonso *et al.* study[98], where the media composition and hormonal stimulation was based on the Turco *et al.* protocols[84]. For example, growing EEOs in media without E2 followed by stimulation with E2 for 2 days succeeded by 4 days of E2, P4, cAMP and PRL led to higher expression of *PAEP*[98], a marker of secretory glandular cells, which I did not observe in my dataset. In my data, the expression of *PAEP* was highest in the unstimulated_MUC5B+PAEP+ population which decreased in its numbers upon stimulation with P4. No distinct MUC5B+ population was reported by Garcia-Alonso *et al.* suggesting that culturing EEOs in media with low amounts of E2 (as reported here) may induce such a cell state/phenotype.

Moreover, given that the most differentiated glandular secretory cell state could not be observed in either the Turco *et al*. nor Boretto *et al.* media and hormonal stimulation protocols, it is likely these protocols model earlier stages of the secretory phase of the menstrual cycle. This would be supported by the morphological and gene expression changes observed. Conventionally though, the *in vivo* menstrual cycle takes 28 days while *in vitro* this has been reduced to 6 days (Garcia-Alonso *et al.*) and 10 days (my study). It is not surprising that the shorter *in vitro* stimulation period and lack of other endometrial cell types in EEOs are not sufficient to induce the more differentiated glandular secretory phenotype. In addition, to achieve the desired hormonal stimulation effect, various concentrations and formulation of the

hormones used need to be tested and benchmarked. Previous studies examining cultured stromal and epithelial cells indicated differential effects of P4 vs MPA vs cAMP stimulation[227,228]. Therefore, reviewing their rate of metabolism *in vitro*, gene expression changes they induce and what concentrations and combinations best mimic the changes associated with the progression of the menstrual cycle are now required for EEOs.

EEOs from donors with/without endometriosis: unstimulated condition

There appeared to be differences between EEOs from endometriosis cases and controls in both the unstimulated and stimulated conditions. In the case of the unstimulated EEOs, the data suggested that the control EEOs exhibited more of a oestrogen-responsive proliferative phase endometrial signature, with a higher proportion of the proliferative_LGR5-like and E2-induced populations when compared to endometriosis EEOs. The significantly higher expression of matrix metalloproteinase 7 (*MMP7*) and UDP Glucuronosyltransferase Family 2 Member B7 (*UGT2B7*) in control EEOs further suggests a differential response of control and endometriosis EEOs to the culture media with low amounts of E2 (i.e. the unstimulated condition) for a number of reasons. Firstly, *MMP7* has been reported to be highly expressed in epithelial cells during the proliferative phase of the menstrual cycle and at the single-cell level specifically expressed by the different proliferative *SOX9*$^{+}$ cell states[98,229]. Secondly, *UGT2B7* has been reported to be important for the final inactivation and elimination of oestrogens[230], reducing their biological activity and preventing their metabolism into hydroxy-oestrogens, which can have a mutagenic potential[231]. In the context of endometriosis, decreased expression of *UGT2B7* has been reported in ovarian endometriomas when

compared to healthy endometrium[232]. The overall disturbed oxidative oestrogen metabolism was linked to an increase of hydroxy-oestrogens and reactive oxygen species in the ectopic location, causing ectopic endometrium proliferation[232]. If such disbalance already exists at the level of the eutopic endometrium, as suggested by the EEO data, it could explain how endometrial epithelial cells misplaced during retrograde menstruation may exhibit an advantage of more extensive proliferation in ectopic locations in women with endometriosis.

With regards to genes upregulated by endometriosis EEOs, high expression of trefoil factor 3 (*TFF3*), ceruloplasmin (*CP*) and selenoprotein P (*SELENOP*) was observed. However, only *TFF3* has been studied in more detail at the level of the endometrium, with its expression localised to endometrial glands (reported to be rare) and the endocervix[233,234]. In endometriosis, *TFF3* has been detected in endometriotic lesions and at increased levels in the peritoneal fluid of women with endometriosis[235]. Moreover, expression of *TFF3* has been reported in endometrial cancers[236,237] and more widely, in a range of chronic inflammatory diseases[238,239]. *TFF3* has been reported to be involved in epithelial cell proliferation, differentiation, as well as migration and antiapoptosis[240–243], all properties that would provide a selective advantage once these cells reach the ectopic locations. The high expression of *TFF3* in endometriosis EEOs, especially in the unstimulated_MUC5B+PAEP+ and pre-secretory populations now require further investigation.

EEOs from donors with/without endometriosis: stimulated condition

With regards to the EEO response to stimulation with P4, the data suggested that EEOs from controls when compared to endometriosis EEOs, responded in a more 'mild' way, i.e. the proliferative phase phenotype and cellular composition was maintained (e.g. observing high *MMP7* and *UGT2B7* expression, presence of the unstimulated_LGR5-like state), but there was a shift towards the intermediate and also luminal-like states. In the case of the endometriosis EEOs, treatment with P4 resulted in a phenotype where a higher number of cells were assigned the pre-luminal and luminal-like identities when compared to control EEOs. Higher expression of serpin family a member 3 (*SERPINA3*) was associated with these cellular states in endometriosis EEOs and requires further investigation, especially, as *SERPINA3* has been reported to be involved in promoting endometrial cancer cell growth and inhibiting apoptosis[244] as well as an important factor in a multitude of other pathologies[245].

Taken together, the results highlighted a differential response to both E2 and P4 treatment in control and endometriosis EEOs, highlighting the utility of the model system to study endometrial cell response to hormones once the system has been fully benchmarked. Such studies are of particular relevance in endometriosis where disturbed E2 and P4 signalling has been reported extensively but with conflicting findings[246,247]. EEOs could serve as a useful tool for functional validation and further study of E2/P4 signalling in the context of the endometrium.

EEOs from donors taking exogenous hormones

The observed differences between donors taking and not taking hormones at sample collection could be explained by a variety of factors. First, it is possible that the effect of exogenous hormones on endometrial cells persists for an extended period of time in the *in vitro* conditions, and the length of the effect can be influenced by the type, formulation and metabolism of hormones taken. Second, any differences seen could have simply been due to the heterogeneity of endometrial cells from different donors, suggesting a donor-specific effect irrespective of taking exogenous hormones. Both scenarios are plausible and will need to be addressed in the future to determine if and how long the effects of exogenous hormones persist *in vitro* and whether this may influence the EEO's response to further perturbations. As described in Chapters 3 and 4, globally millions of women take exogenous hormones on a daily basis, making it a worthwhile endeavour to better understand if and to what point the effects of exogenous hormones are reversed *in vitro*, and thus allow for an unbiased comparison between cultured cells from donors taking and not taking hormonal treatment.

5.5.2 Cellular composition of the EEO-ES system

For the first time in this chapter, I reported the generation of scRNAseq data for the EEO-ES model system that uses a functionalised synthetic hydrogel developed by Professor Linda Griffith's team. In this pilot study, I generated data from four donors only, utilising samples from a variety of conditions (donors with/without endometriosis, taking hormones). The effect of these conditions could not be compared or evaluated due to the low number of lines analysed. However, the general trends with regards to the cellular heterogeneity observed and in comparison to EEOs are discussed in more detail below.

Epithelial cell lineage heterogeneity

One striking difference between the EEO and EEO-ES systems, was the low number of ciliated cells detected in EEO-ES cultures. In the EEO only system, a reduced number of ciliated cells and an increase in the number of secretory cells was reported when EEOs were treated with WNT pathway inhibitors[98]. This suggests that in the EEO-ES system, the stromal cells may be able to induce such a WNT inhibitory environment through cell-cell signalling, leading to the reduction in the number of ciliated cells observed. Stromal cells undergoing decidualisation are known to secrete WNT pathway antagonists[248] and this WNT inhibitory phenomenon was, to some extent, likely recapitulated in this EEO-ES model system. However, such a WNT inhibitory environment is important for the secretory phase glandular/glandular secretory epithelial identity[98], but this cell state was not detected in either the EEO-ES or EEO only model systems. Considering the marker gene expression observed, in both model systems some secretory gene expression was detected (e.g. *SCGB1D2* and *SCGB2A2*), but *SCGB2A2* (a glandular secretory marker) was only lowly expressed in both systems. This further highlights that achieving the glandular secretory phenotype *in vitro* requires additional signalling. Such signalling could, for example, come in the form of endogenous WNT inhibition from a more decidualised stromal cell state than the one achieved in this thesis in the EEO-ES system. This would be of particular relevance, as *in vivo* the cellular crosstalk between epithelial and stromal cells determines their response to endogenous hormones, specifically progesterone, and thus regulates the epithelial cells' secretory phenotype.

Interestingly, a population detected in the EEO-ES system, but not EEOs, was the proliferative-like_SOX9+LGR5- population. The marker gene expression was very similar to that of the proliferative cell state described in Chapters 3 & 4 *in vivo* (i.e. expressing *SOX9* but lacking *LGR5* expression). This was in contrast with the proliferative_LGR5-like population detected in EEOs that expressed *SOX9*, *WNT7A* and *LGR5*, resembling the proliferative_LGR5 population described *in vivo*. Given the spatial location of these different cell states across the endometrium described by the Garcia-Alonso *et al*. study, the results suggest that the EEO-ES system more closely resembles the proliferative glandular cell population found in the deeper regions of the endometrium (*SOX9*$^{+}$*LGR5*$^{-}$), while the EEO system captures the proliferative_LGR5 population (*SOX9*$^{+}$*LGR5*$^{+}$) located in the more superficial part of the endometrium, specifically the lumen.

When comparing marker gene expression, epithelial cell populations found in both the EEO-ES and EEO model systems included the cycling, KRT17/inflammatory and unstimulated_MUC5B+PAEP+ cell states. The E2-response population in EEO-ES did not express the same marker genes as the E2-induced population in EEOs and was thus given a different name. The E2-response population still expressed markers such as *PGR* (indiciating E2 response) but also genes such as solute carrier family 7 member 2 (*SLC7A2*) and carboxypeptidase M (*CPM*) thought to be involved in amino acid transmembrane transporter activity and control of peptide hormone growth factor activity at the cell surface, respectively[249,250]. The cellular and transcriptomic response to E2 and P4 was therefore likely different in the EEO-ES and EEO systems.

Taken together, to fully understand the transcriptomic resemblance and differences of the epithelial cells from both the EEO and EEO-ES systems, in the future, these cells should be integrated and analysed in a joint manifold. Such an approach would allow for a more comprehensive analysis and understanding of how the transcriptomic profiles of epithelial cells change when they are cultured alone (EEOs) or in the presence of stromal cells (EEO-ES).

Stromal cell lineage heterogeneity

Across the stromal cell states identified, more of a continuous marker gene expression was observed when compared to the epithelial cells states, suggesting a more fluid transition between the different cell states. Of interest is the identification of the eS_THY1+ population which expressed markers previously described for endometrial mesenchymal stem cells (e.g. *THY1*, *PDGFRB*, *NOTCH3*)[167,170,251,252] and the expression of perivascular cell markers (e.g. *RGS5*, *MYO1B*, *PAG1*) reported in Chapters 3 and 4 for PV cells. Higher expression of *PDGFRB* and *RGS5* was also observed in the eS_CPE+ cell state which highly expressed carboxypeptidase E (*CPE*), tryptophan 2,3-dioxygenase (*TOD2*) and *MMP11. CPE* expression has not been studied in the endometrium, but is reported to play a role in processing peptide prohormones and in the mouse brain was found to promote differentiation of neural stem cells by negatively regulating the WNT/β-catenin signalling in these cells[253]. Since the WNT/β-catenin signalling pathway has been reported to play a crucial role in endometrial mesenchymal stem cell maintenance and regulation[254], one could speculate that in the EEO-ES system, the eS_THY1+ population is the most undifferentiated stem-cell population uniquely expressing *NOTCH3*, while the eS_CPE+ is a more differentiated/transitioning

population. As discussed for the EEO findings earlier, analysing more time points during the culture period and using lineage-tracing approaches could help better characterise the transitions between the stromal cell states identified, as well as understand what role perivascular cells might play in the establishment and differentiation of what we consider to be/call endometrial stromal cell cultures.

Another population expressing marker genes involved in WNT pathway regulation was the eS_IGFBP6+ cell state, expressing WNT ligand secretion mediator (*WLS*), secreted frizzled related protein 1 (*SFRP1*) and insulin growth factor binding protein 6 (*IGFBP6*). *WLS* is thought to ensure intracellular trafficking of WNT to the cell surface[255,256], while *SFRP1* is a soluble modulator of WNT signalling and has been reported to be elevated in the proliferative phase endometrium and also endometriotic lesions[257]. In addition, expression of *SFRP1* was reported to be most likely directly regulated by E2[257], which would be in line with an increased *IGFBP6* expression in this cell state, also found to increase in expression upon treatment with E2 in a number of different cells[258,259]. Taken together, it can be hypothesised that in response to E2, this cell state plays a role in modulating WNT pathway signalling in the EEO-ES system.

Interestingly, in the EEO-ES system, I also identified the eS_ISG15+ population characterised by high expression of markers involved in interferon signalling and innate immunity functions, such as interferon induced protein with tetratricopeptide repeats 1 and 3 (*IFIT1* and *IFIT3*), MX dynamin like GTPase 1 (*MX1*) and the interferon-induced ISG15 ubiquitin-like modifier (*ISG15*). While I did not detect such a population in my *in vivo* data presented in Chapters 3 and 4, an ISG15+ stromal

population has been reported previously in a scRNAseq *in vivo* study profiling human endometrial stromal cells from proliferative phase endometrium[170]. Both the *in vivo* ISG15+ population and the *in vitro* one defined in this chapter shared the expression of many markers. The *in vivo* study suggested this ISG15+ population could represent an activated stromal cell population, playing a role in immunomodulation and tissue regeneration[170]. However, interferon signalling also plays an important role in implantation and early pregnancy leading to stromal decidualisation[260,261] and abnormal *ISG15* expression has been reported in endometrial cancer[262]. The exact function of this stromal population *in vitro* should now be investigated further.

Lastly, the dS-like population identified in the EEO-ES system expressed markers suggestive of emerging decidualised stromal cells, e.g. prostaglandin D2 synthase (*PTGDS*), *MMP2*, and fibroblast growth factor 7 (*FGF7*), with *PTGDS* and *MMP2* previously reported as markers for decidualised stromal cell[98]. *PTGDS* expression has been found to peak in the mid-secretory phase endometrium[263] and drop when progesterone withdrawal is caused by administering the antiprogestin mifepristone[264]. Therefore, the *PTGDS* expression seen here may have been observed as a result of the EEO-ES system's response to P4. However, the dS-like population reported here also highly expressed cellular retinoic acid binding protein 2 (*CRABP2*), a marker of non-decidualised stroma[98], highlighting that further stimulation/different signalling is needed to induce a true decidual stromal cell differentiation *in vitro*.

Effects of hormonal stimulation

In the EEO-ES system, the effect of the stimulated conditions (i.e. reducing E2 levels and adding P4) did not appear to have a significant effect on the cellular composition of the co-cultures. However, some indication of a difference between the unstimulated and stimulated conditions included the increase in dS-like cells and reduction in unstimulated_MUC5B+PAEP+ and E2-response epithelial cells after treatment with P4. Multiple factors can be responsible for the lack of a more prominent response, including fast metabolism of P4 by having both stromal and epithelial cells growing together as well as unknown permeability of the synthetic hydrogel to both E2 and P4 in this system. Assessing the amount of E2 and P4 before and after media changes could help address the question of how much is likely reaching the cells/to what extent E2 and P4 are being metabolised. Nonetheless, the results show that in this culture system, the stimulation protocol needs to be adjusted to reflect the more complex cellular nature and hydrogel properties of the system.

5.5.3 Summary of future directions

In this chapter, I have presented an overview of the scRNAseq data showing the cellular heterogeneity of 8 EEO lines. Whilst a further 3 samples were analysed, I was not able to demultiplex and analyse these in time for this thesis. The next steps will include analysing the remaining lines and repeating the analyses presented here, so I can validate and better understand the effects of both taking exogenous hormones at sample collection and deriving organoids from donors with/without endometriosis have on the cellular make-up of the EEO lines established.

Secondly, this chapter was of a descriptive nature, but with a larger number of EEO lines included, I will focus on performing further more quantitative analyses, such as cell-cell communication, trajectory inference and DEG analyses to compare the response of the various cell states identified to the unstimulated and stimulated conditions. Moreover, deciphering cell-cell signalling between the epithelial and stromal cells in the EEO-ES co-cultures will be of great interest to better understand the signalling pathways and cellular events leading to the phenotypes observed. Trajectory inference analyses will on the other hand help me evaluate my hypotheses about the stem-cell nature of the eS_THY1+ population and its likely transition/differentiation into the other cell states described.

Thirdly, as mentioned throughout the discussion, in the future, additional time points should be included and analysed to help with defining which cell states appear at what stage of the culture protocol. Moreover, given the lack of a significant response upon treatment with P4 in the EEO-ES system and weak response (e.g. lack of *PAEP* expression) in the EEO system, the hormonal stimulation protocols need to be optimised and benchmarked against the existing data and *in vivo* reference cells. Experimenting with and testing different concentrations, exposure periods and combinations of hormones (E2/P4/MPA/PRL) but also other signalling molecules (cAMP, WNT pathway inhibitors) could help us to achieve the differentiation of the epithelial cells into the glandular/glandular secretory states and the stromal cells into decidualised stroma that have not been recapitulated in the current culture conditions and systems described.

Lastly, the future steps would also include validation of the findings described throughout this chapter and upon further data analysis using e.g. smFISH imaging and immunofluorescence stainings to confirm the presence and location of the cell states described *in situ*. To this end I have fixed and stored EEOs and EEO-ES co-cultures for each of the conditions described. In collaboration with Miss Cecilia Icoresi-Mazzeo (Staff Scientist, Wellcome Sanger Institute) we have conducted preliminary smFISH experiments, showing that it is possible to stain and detect smFISH signals from cryosectioned EEOs and EEO-ES co-cultures.

5.6 Conclusions

In summary, in this chapter I have presented scRNAseq data for EEO lines (n = 8) and EEO-ES lines (n = 4), both in the unstimulated and stimulated conditions. Firstly, my aim was to establish these culture systems in our laboratory and derive endometrial EEO and ES cell lines from a variety of samples (i.e. from donors with/without endometriosis and taking exogenous hormones at sample collection). Succeeding at this aim and generating a small bank of the organoid and stromal cell lines, I next set out to explore the cellular heterogeneity of the two culture systems (EEOs and EEO-ES). Moreover, I looked into characterising the effects endometriosis and exogenous hormones have on the cellular composition of the EEO model system. To help annotate the cell states identified and understand the transcriptomic similarities between endometrial cells *in vivo* and *in vitro*, I utilised my annotated endometrial cell *in vivo* reference dataset from Chapter 3. Moreover, I utilised the marker genes identified in both Chapters 3 and 4 to guide the cell state assignment.

To conclude, I have been able to address the aims set out for this chapter, generating further hypotheses and research questions to test in the future. Both the *in vivo* and *in vitro* cellular maps of the endometrium that I have generated and described throughout Chapters 3-5 have the potential to serve as a great resource for many other scientists working in the field of endometrial biology. The utility and limitations of the work presented in this thesis are further discussed in the next, final discussion and conclusions chapter.

Chapter 6: Final discussion and conclusions

The title of this thesis promised to take the reader on a journey of exploring endometrial cellular heterogeneity and its role in endometriosis using single-cell transcriptomics. By looking at ~700,000 endometrial cells, both *in vivo* and *in vitro*, across the menstrual cycle, under the influence of exogenous hormones and in those with/without endometriosis, new cell populations were identified and many new research questions and hypotheses formed. The cell populations and genes identified are likely to be important in mediating endometrial function, and thus require further investigation and validation as discussed throughout the thesis. For now though, this is where the initial data analysis achievable during my DPhil studies ends, but the journey and exploration undoubtedly continue as I move into my next research role. This final chapter is therefore an opportunity to provide a brief overview of the work presented in this thesis, and to discuss its major strengths and limitations as well as its importance and where it can be taken next.

6.1 Thesis summary

The overall aim of this thesis was to characterise the transcriptomic signatures of endometrial cell states that drive the dynamic changes of the human endometrium, and pinpoint any differences specific to endometriosis. To do so, I generated scRNAseq and snucRNAseq data from 75 individuals (53 endometriosis cases and 22 controls) and integrated the new data with two previously published scRNAseq datasets profiling healthy endometrial samples. To build this large cellular atlas of the endometrium, and to understand the unique features of each of the datasets, in Chapter 3 I first focused on integrating the scRNAseq data only. The data served as a starting point for defining the cell states present, including their marker gene

expression, and generating hypotheses about the differences between the endometrium of women with/without endometriosis. The number of samples analysed was not well-balanced between the control and case groups across the menstrual cycle, but hinted at dysregulated decidualisation in endometriosis cases.

To further validate the findings reported in Chapter 3, I turned to the larger set of samples profiled using snucRNAseq and integrated it with the scRNAseq datasets described in Chapter 3. By doing so in Chapter 4, I confirmed the presence of some of the newly defined cell states, including the epithelial TENM2 and hormones_WNT7A populations, plus the intermediate cell state identified in the mesenchymal cell lineage. The TENM2 and intermediate cell populations appeared around the time of ovulation, while the hormones_WNT7A population was a feature of exogenous hormonal treatment, specifically the Mirena coil IUD. Additionally, the presence of a number of cell states detected in Chapter 3, but not recapitulated in Chapter 4, highlighted their donor-specific properties. In contrast, a cell state detected in larger numbers in the nuclei dataset included the MUC5B population. Lastly, when describing the differences between the endometrium of women with/without endometriosis stratified based on the menstrual cycle phase, findings consistent across the scRNAseq and snucRNAseq data included: (i) higher proportion of cycling and proliferative epithelial cell populations in endometriosis during the proliferative phase and (ii) a slight reduction in the number of dS cells in the secretory phase.

Finally, in my last experimental chapter, Chapter 5, I investigated the cellular heterogeneity of two *in vitro* model systems of the endometrium - EEOs and EEO-ES co-cultures - including their response to hormonal stimulation. Interestingly, the two systems likely modelled two different epithelial cell states found in different spatial locations of the *in vivo* endometrium. In the EEO-ES co-cultures I detected a cell population transcriptomically similar to the one previously described in the deeper regions of the endometrium (*SOX9*$^{+}$*LGR5*$^{-}$) while in the EEO system it was a population described in the luminal region (*SOX9*$^{+}$*LGR5*$^{+}$)[98,154]. Furthermore, the EEO-ES co-cultures did not respond to the stimulation protocol used, while the EEO cells did and exhibited unique features and cellular composition when derived from donors with and without endometriosis. The presence of the unstimulated_MUC5B+PAEP+ population in endometriosis EEOs and differential response to E2 and P4 now warrant further investigation.

6.2 Strengths and limitations

A major strength of this thesis is the inclusion of a large set of endometrial biopsies with detailed clinical metadata from donors with/without endometriosis during their natural cycles, but also when taking hormonal therapy. Larger sets of samples are needed to fully account for the dynamic processes the endometrium undergoes, and the changes that occur when exogenous hormones are administered if we are to unpick what constitutes, for example, a menstrual phase signature from a disease-specific one. Integration of multiple datasets and data sources can be a very powerful tool to achieve such a sample size, but also when evaluating what one's findings mean in the context of previously published work. I see the demonstration of such an approach - integrating multiple datasets and data sources

(cells and nuclei) - as another strength of this thesis. It allowed me to contextualise and harmonise my findings with respect to the previously published studies, helped me evaluate the identification and naming of the different cell states both *in vivo* and *in vitro,* and allowed me to use a large collection of banked samples. The updated cellular map of the endometrium presented in this thesis can now be improved even further by including more available datasets.

On the other hand, a major limitation of this thesis is its descriptive character. Many of the more advanced computational and statistical analyses were not performed due to the time-constraints and challenges faced. A major set-back included the closure of our laboratory for an extended period of time during the Covid-19 pandemic, as well as prolonged cancellations of endometriosis surgeries. For this reason, most of the samples used for scRNAseq and organoid derivation were only collected during the final year of my studies and I thus did not have sufficient time to perform more quantitative analyses. Nonetheless, this initial exploration of the data allowed me to form new questions and hypotheses that I will be able to address following the thesis submission.

Further limitations of the work presented here include the low number of samples from menstrual and premenstrual phases of the cycle as well as a lack of more fine annotation of the menstrual cycle phases. Such annotation and sample inclusion would be key to fully describe the dynamic cellular and molecular changes occurring across the menstrual cycle and characterising how both endometrial shedding and regeneration take place. Moreover, including further samples of full thickness endometrium would have also been of great importance in this endeavour as they

would have provided cues into the events taking place in the deeper *basalis* layer of the endometrium, an area that has been underrepresented and understudied.

Lastly, with regards to the *in vitro* work presented, one of the limitations includes the lack of data for multiple time points during the culture protocol and hormonal stimulation, and not analysing data for three EEO lines in time. A larger number of lines would have been helpful to strengthen the validity of the differences observed when comparing EEOs from donors with/without endometriosis and those taking exogenous hormones at sample collection. On the other hand, one of the strengths of the data generated is the fact that scRNAseq data are available for both the *in vivo* primary endometrial samples as well as the EEO lines from the same donor biopsy. In addition, data are also available for the matched EEO-ES lines and EEOs cultured on their own from the same donors. In the future, this unique data availability shall be harnessed to its full potential and utilised to characterise how endometrial cells *in vivo* and various *in vitro* models differ. It will help by removing an extra layer of complexity that is introduced by donor line heterogeneity when making such comparisons using EEO and EEO-ES lines from different donors.

6.3 Implications of the work presented & future directions

Even when considering the limitations described above, the cellular map of the endometrium presented in this thesis is the most comprehensive one built so far and holds a great potential to serve as a reference for studying the endometrium in both health and disease. Moreover, it can be used as a guide for improving *in vitro* models of the endometrium, making them the most accurate representation of the *in vivo* endometrium they can be. There is no doubt, however, that further work is

needed before we can make full use of this resource. Below, I briefly discuss the potential implications once this work has been finalised.

Endometrial function in health and disease

When it comes to endometrial function, there is still a lot we do not fully understand, including how it regenerates on a monthly basis. Which cells and signals orchestrate the shedding, growth, and differentiation in response to steroid hormones? How and why are these signals disturbed in various pathologies? The map generated in this thesis can help start answering these questions by reconstructing trajectories leading to certain cell fate decisions and what cell-cell communication drives these changes. The Garcia-Alonso *et al.*[98] study has already nicely demonstrated how such analyses can be incredibly informative; however, some cell states reported in this thesis were not detected in the cellular maps reported previously[97,98] and could now help better define the dynamic transitions endometrial cells undergo across the menstrual cycle and importantly, in response to exogenous hormonal therapy.

Another attractive application of the comprehensive map generated as part of my studies is its use in interpreting how genetic variants associated with disease, such as endometriosis, manifest at the cellular level. Linking endometriosis associated single nucleotide polymorphisms (SNPs) with cell type-specific gene expression obtained through scRNAseq could help evaluate the effects of genetic variation on gene expression in individual cells. Performing single-cell expression quantitative trait loci (sc-eQTL) analyses[265–267] on the dataset presented here would be feasible as we have both the genotyping and scRNAseq data for the donors analysed. Yet, due to the dynamic nature of the endometrium, further samples will need to be

profiled for such analyses to be meaningful. However, the linking of disease associated SNPs and cell state-specific gene expression is also possible through other approaches[268,269] that do not require the genotype and scRNAseq data from the same donors, but can still provide insights into whether GWAS-implied genes are concordantly expressed in a particular cell state. As such, the cellular map generated here could be queried for how genetic differences identified for other uterine/endometrial diseases (e.g. fibroids[270], endometrial cancer[271], adenomyosis) contribute to gene expression differences at the single-cell level.

Therapeutic interventions and modelling endometrial cells in vitro

Another important utility of the cellular map of the endometrium *in vivo* and *in vitro*, is the discovery of targets for therapeutic interventions. By distinguishing how cells behave in health and how in disease, we can design more efficient and targeted approaches for treatment and diagnostics. For example, in the case of endometriosis, it is of importance to maintain normal endometrial function and only target the endometrial-like cells in ectopic locations. Current strategies using contraceptive hormonal therapy are not suitable for those individuals who wish to conceive and the severe side-effects that can come with such treatment are also not acceptable for many. A deeper characterisation and comparison of the eutopic and ectopic endometrial cells could aid the identification of therapeutic targets that would specifically target the disease cells only. The recently published Tan *et al*. and Fonseca *et al*. studies[99,144] profiling endometriotic tissues could now be integrated with the large dataset of eutopic endometrial cells I have generated here, in order to investigate whether such targets could be identified.

If successful, the *in vitro* model systems of both the endometrium and endometriotic lesions[92,99] could be used to test and evaluate the effects of any new drugs, assessing their response in a high-throughput manner using single-cell genomics. However, in this regard, these model systems first need to be further benchmarked and validated in order to understand their strengths and weaknesses in modelling the cellular states and processes of interest. A well-defined cellular map of the endometrium presented here will undoubtedly prove to be an invaluable tool in benchmarking these *in vitro* systems.

6.4 Concluding remarks

I still remember the first time I searched PubMed® with the following key words *'endometrium AND single-cell RNA-sequencing.'* There was a single hit - the 2016 Krjutškov *et al.* study profiling ~70 cultured stromal cells[96]. It was autumn 2018, I had just started my DPhil and could not believe no one had ever looked at the endometrium at the single-cell level. Fast-forward to 2023 and the same PubMed® search yields a very different output. A number of studies have analysed the endometrium in health and across various pathologies, providing initial insights into the cellular and molecular functioning of this fascinating tissue. However, despite the key role of the endometrium in perpetuating life and uterine/endometrial disorders having a huge impact on millions of individuals globally[27,272,273], generation of these cellular maps of the uterus/endometrium has lagged behind the generation of cellular maps of other organs of the human body, both in time and number of studies conducted. This is not surprising given the reproductive system and its associated pathologies have been historically understudied[274] and as such the existing knowledge gap requires ever more research attention.

What becomes apparent, however, when sifting through the so far published single-cell endometrial studies is that in most cases they are underpowered to account for the inter- and intra-individual variability inherent to the ever-changing endometrium. This obstacle can be overcome by e.g. integrating various datasets and profiling a larger number of cells and samples - endeavours made possible through technological advancements allowing simultaneous profiling of thousands of cells in parallel[113] and development of new bioinformatics tools[202]. Another challenge though is that the nomenclature, marker gene expression and bioinformatic analyses are far from harmonised across the published datasets, making it difficult to compare the findings and cell states reported in different studies. Detailed annotation of both the transcriptomic and clinical data, data sharing, standardisation of nomenclature, data analysis and results reporting is absolutely essential if the scientific community is to fully benefit from these rich datasets. Submitting, annotating, and making data widely available through the Human Cell Atlas data portal (initiative aiming to map all cells in the human body) is an example of how certain data standards and harmonisation can be met[275,276]. Adding the data generated during my studies will be a major contribution towards the aims of the Human Cell Atlas, and additionally will help shift the spotlight to the underrepresented endometrial cells, including endometrial cells in endometriosis.

To conclude, all research studies generate new insights that allow us to improve our knowledge, but the complexity of the human body and its tissues means that any one study only really scratches the surface of what the data can uncover. My hope is that upon further work on the dataset generated for the endometrium as part of my studies, its integration and harmonisation with other published datasets, it will

prove to be a helpful resource for many scientists working in the field. A resource they can turn to test their hypotheses and ask their research questions that have fascinated and puzzled them about the remarkable power and functioning of the endometrium, that for now, remains a mystery.

6.5 There was more than just the lab work and studying

Finally, I would like to end this thesis by briefly discussing the things I did outside of the realm of working in the lab and staring at the screen trying to make sense out of the results. There are three things very important to me: (i) being able to do 'fun' science while learning new things, (ii) passing this new knowledge on, and (iii) ensuring science does not just stay behind closed doors, but instead is shared in an accessible way and reaches wide audiences. I still find it surprising that hardly anyone has heard of endometriosis, even though it has such a huge impact on the lives of millions of individuals globally. I must admit, the first time I heard about endometriosis was during my undergraduate degree in a specialist lecture about endometrium and endometrial pathologies. It left me wondering, can this be improved? Is there a way in which we can start having conversations about periods, wombs, and pain - topics that have been surrounded by stigma and taboos for so long? And could I possibly also include talking about the research we do to understand endometriosis better? The answer is: ***yes***. In Appendix 7 I have summarised the projects I worked on and awards I received for these public engagement/outreach activities. They truly formed a cherished part of my studies.

Lastly, without doubt I learnt a lot of new things during my studies and did plenty of ‘fun’ science as presented throughout this thesis. All that remains is to address point (ii) passing new knowledge on. Of course, one can think of presenting in lab meetings, talking at conferences, writing articles - all of these have been a great way to discuss my work and get invaluable feedback (see Appendix 8). However, the biggest joy came from getting students (i.e. scientists to be) excited about a topic they might have never even considered to be interesting. Seeing them think about problems differently, ask questions, discover a new topic to read about - that is what has been very rewarding for me. I really enjoyed teaching the MSc Clinical Embryology students, postgraduate entry medical students and students doing the same undergraduate course I once did at the University of Edinburgh about scRNAseq and its use in reproductive biology. It was there where my interest in reproductive biology was first sparked and continues to this day. So I would like to end with a *thank you* to all the wonderful people that happily shared their knowledge and wisdom with me throughout my studies and have in this way, without even knowing, continued fuelling the once sparked interest and curiosity!

References

1. McLennan, C. E. & Rydell, A. H. Extent of endometrial shedding during normal menstruation. *Obstet. Gynecol.* **26**, 605–621 (1965).
2. Jabbour, H. N., Kelly, R. W., Fraser, H. M. & Critchley, H. O. D. Endocrine Regulation of Menstruation. *Endocr. Rev.* **27**, 17–46 (2006).
3. Prianishnikov, V. A. A functional model of the structure of the epithelium of normal, hyperplastic and malignant human endometrium: a review. *Gynecol. Oncol.* **6**, 420–428 (1978).
4. Padykula, H. A. Regeneration in the primate uterus: the role of stem cells. *Ann. N. Y. Acad. Sci.* **622**, 47–56 (1991).
5. Gray, C. A. *et al.* Developmental Biology of Uterine Glands. *Biol. Reprod.* **65**, 1311–1323 (2001).
6. Valentijn, A. J. *et al.* SSEA-1 isolates human endometrial basal glandular epithelial cells: phenotypic and functional characterization and implications in the pathogenesis of endometriosis. *Hum. Reprod.* **28**, 2695–2708 (2013).
7. Tempest, N., Maclean, A. & Hapangama, D. K. Endometrial Stem Cell Markers: Current Concepts and Unresolved Questions. *Int. J. Mol. Sci.* **19**, (2018).
8. Tempest, N. *et al.* Histological 3D reconstruction and in vivo lineage tracing of the human endometrium. *J. Pathol.* **251**, 440–451 (2020).
9. Yamaguchi, M. *et al.* Three-dimensional understanding of the morphological complexity of the human uterine endometrium. *iScience* **24**, 102258 (2021).
10. Tempest, N. *et al.* Novel microarchitecture of human endometrial glands: implications in endometrial regeneration and pathologies. *Hum. Reprod. Update* **28**, 153–171 (2022).
11. Masterton, R., Armstrong, E. M. & More, I. A. The cyclical variation in the percentage of ciliated cells in the normal human endometrium. *J. Reprod. Fertil.* **42**, 537–540 (1975).
12. Berbic, M. & Fraser, I. S. Immunology of normal and abnormal menstruation. *Womens. Health* **9**, 387–395 (2013).
13. Abu-Raya, B., Michalski, C., Sadarangani, M. & Lavoie, P. M. Maternal Immunological Adaptation During Normal Pregnancy. *Front. Immunol.* **11**, (2020).
14. Oreshkova, T., Dimitrov, R. & Mourdjeva, M. A cross-talk of decidual stromal cells, trophoblast, and immune cells: a prerequisite for the success of pregnancy. *Am. J. Reprod. Immunol.* **68**, 366–373 (2012).
15. Gargett, C. E., Schwab, K. E. & Deane, J. A. Endometrial stem/progenitor cells: the first 10 years. *Hum. Reprod. Update* **22**, 137–163 (2016).
16. Reed, B. G. & Carr, B. R. The Normal Menstrual Cycle and the Control of Ovulation. in *Endotext* (South

Dartmouth, 2000).

17. Fehring, R. J., Schneider, M. & Raviele, K. Variability in the Phases of the Menstrual Cycle. *Journal of Obstetric, Gynecologic & Neonatal Nursing* **35**, 376–384 (2006).
18. Mihm, M., Gangooly, S. & Muttukrishna, S. The normal menstrual cycle in women. *Anim. Reprod. Sci.* **124**, 229–236 (2011).
19. Bull, J. R. *et al.* Real-world menstrual cycle characteristics of more than 600,000 menstrual cycles. *NPJ Digit Med* **2**, 1–8 (2019).
20. Garry, R., Hart, R., Karthigasu, K. A. & Burke, C. A re-appraisal of the morphological changes within the endometrium during menstruation: a hysteroscopic, histological and scanning electron microscopic study. *Hum. Reprod.* **24**, (2009).
21. Ferenczy, A. Studies on the cytodynamics of human endometrial regeneration: I. Scanning electron microscopy. *Am. J. Obstet. Gynecol.* **124**, 64–74 (1976).
22. Ferenczy, A., Bertrand, G. & Gelfand, M. M. Proliferation kinetics of human endometrium during the normal menstrual cycle. *Am. J. Obstet. Gynecol.* **133**, 859–867 (1979).
23. Novak, E. & Te Linde, R. W. The Endometrium of the Menstruating Uterus. *J.A.M.A* **83**, 900–906 (1924).
24. Burton, G. J., Watson, A. L., Hempstock, J., Skepper, J. N. & Jauniaux, E. Uterine Glands Provide Histiotrophic Nutrition for the Human Fetus during the First Trimester of Pregnancy. *J. Clin. Endocrinol. Metab.* **87**, 2954–2959 (2002).
25. Salamonsen, L. A. Tissue injury and repair in the female human reproductive tract. *Reproduction* **125**, 301–311 (2003).
26. Shafrir, A. L. *et al.* Risk for and consequences of endometriosis: A critical epidemiologic review. *Best Pract. Res. Clin. Obstet. Gynaecol.* **51**, 1–15 (2018).
27. Zondervan, K. T., Becker, C. M. & Missmer, S. A. Endometriosis. *N. Engl. J. Med.* **382**, 1244–1256 (2020).
28. Giudice, L. C. Clinical practice. Endometriosis. *N. Engl. J. Med.* **362**, 2389–2398 (2010).
29. Simoens, S. *et al.* The burden of endometriosis: costs and quality of life of women with endometriosis and treated in referral centres. *Hum. Reprod.* **27**, 1292–1299 (2012).
30. Ghai, V., Jan, H., Shakir, F., Haines, P. & Kent, A. Diagnostic delay for superficial and deep endometriosis in the United Kingdom. *J. Obstet. Gynaecol.* **40**, 83–89 (2020).
31. Nnoaham, K. E. *et al.* Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. *Fertil. Steril.* **96**, 366–373.e8 (2011).
32. Hudelist, G. *et al.* Diagnostic delay for endometriosis in Austria and Germany: causes and possible consequences. *Hum. Reprod.* **27**, 3412–3416 (2012).
33. Sourial, S., Tempest, N. & Hapangama, D. K. Theories on the pathogenesis of endometriosis. *Int J*

Reprod Med **2014**, 179515 (2014).

34. Sampson, J. A. Peritoneal Endometriosis Due to the. *American Journal of Obstetrics and Gynecology* **14**, 422–469 (1927).
35. Halme, J., Hammond, M. G., Hulka, J. F., Raj, S. G. & Talbert, L. M. Retrograde menstruation in healthy women and in patients with endometriosis. *Obstet. Gynecol.* **64**, 151–154 (1984).
36. Suginami, H. A reappraisal of the coelomic metaplasia theory by reviewing endometriosis occurring in unusual sites and instances. *Am. J. Obstet. Gynecol.* **165**, 214–218 (1991).
37. Jerman, L. F. & Hey-Cunningham, A. J. The role of the lymphatic system in endometriosis: a comprehensive review of the literature. *Biol. Reprod.* **92**, 64 (2015).
38. Uimari, O. *et al.* Genome-wide genetic analyses highlight mitogen-activated protein kinase (MAPK) signaling in the pathogenesis of endometriosis. *Hum. Reprod.* **32**, 780–793 (2017).
39. Kennedy, S. *et al.* ESHRE guideline for the diagnosis and treatment of endometriosis. *Hum. Reprod.* **20**, 2698–2704 (2005).
40. Dunselman, G. A. J. *et al.* ESHRE guideline: management of women with endometriosis. *Hum. Reprod.* **29**, 400–412 (2014).
41. Leonardi, M., Martin, E., Reid, S., Blanchette, G. & Condous, G. Deep endometriosis transvaginal ultrasound in the workup of patients with signs and symptoms of endometriosis: a cost analysis. *BJOG* **126**, 1499–1506 (2019).
42. Leonardi, M., Robledo, K. P., Espada, M., Vanza, K. & Condous, G. SonoPODography: A new diagnostic technique for visualizing superficial endometriosis. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **254**, 124–131 (2020).
43. Leonardi, M., Avery, J. & Hull, M. L. Chapter 19 - Novel diagnostic strategies for endometriosis. in *Immunology of Endometriosis* 297–317 (Academic Press, 2022).
44. Baumgarten & Raine-Fenning. The social and psychological impact of endometriosis on women's lives: a critical narrative review. *Hum. Reprod.*
45. Becker, C. M. *et al.* ESHRE guideline: endometriosis. *Hum Reprod Open* **2022**, hoac009 (2022).
46. Chapron, C., Marcellin, L., Borghese, B. & Santulli, P. Rethinking mechanisms, diagnosis and management of endometriosis. *Nat. Rev. Endocrinol.* **15**, 666–682 (2019).
47. Vercellini, P. *et al.* Medical treatment of endometriosis-related pain. *Best Pract. Res. Clin. Obstet. Gynaecol.* **51**, 68–91 (2018).
48. Streuli, I. *et al.* An update on the pharmacological management of endometriosis. *Expert Opin. Pharmacother.* **14**, 291–305 (2013).
49. Brown, J., Pan, A. & Hart, R. J. Gonadotrophin-releasing hormone analogues for pain associated with endometriosis. *Cochrane Database Syst. Rev.* CD008475 (2010).

50. Ferrero, S., Evangelisti, G. & Barra, F. Current and emerging treatment options for endometriosis. *Expert Opin. Pharmacother.* **19**, 1109–1125 (2018).
51. Olive, D. L. Optimizing gonadotropin-releasing hormone agonist therapy in women with endometriosis. *Treat. Endocrinol.* **3**, 83–89 (2004).
52. Maharajaa, S. P. K., Asally, R., Markham, R. & Manconi, F. Endometriotic lesions. *Journal of Endometriosis and Pelvic Pain Disorders* **11**, 62–76 (2019).
53. Clement, P. B. The pathology of endometriosis: a survey of the many faces of a common disease emphasizing diagnostic pitfalls and unusual and newly appreciated aspects. *Adv. Anat. Pathol.* **14**, 241–260 (2007).
54. Johnson, N. P. *et al.* World Endometriosis Society consensus on the classification of endometriosis. *Hum. Reprod.* **32**, 315–324 (2017).
55. American Society for Reproductive Medicine. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. *Fertil. Steril.* **67**, 817–821 (1997).
56. Montanari, E. *et al.* Comparison of #Enzian classification and revised American Society for Reproductive Medicine stages for the description of disease extent in women with deep endometriosis. *Hum. Reprod.* **37**, 2359–2365 (2022).
57. Zondervan, KT, Becker, CM, Koga, K, Missmer SA, Taylor RN, Viganò, P. Endometriosis. *Nat Rev Dis Primers* **4**, (2018).
58. Naftalin, J. *et al.* How common is adenomyosis? A prospective study of prevalence using transvaginal ultrasound in a gynaecology clinic. *Hum. Reprod.* **27**, 3432–3439 (2012).
59. Yu, O, Schulze-Rath, R, Grafton,J, Hansen, K, Scholes, D, Reed, SD. Adenomyosis incidence, prevalence and treatment: United States population-based study 2006–2015. *Am. J. Obstet. Gynecol.* **223**, 94.e1–94.e10 (2020).
60. Is, P. H. F. The symptomatology of adenomyosis. *Best Pract. Res. Clin. Obstet. Gynaecol.* **20**, 547–555 (2006).
61. Giuliani, E., As-Sanie, S. & Marsh, E. E. Epidemiology and management of uterine fibroids. *Int. J. Gynaecol. Obstet.* **149**, (2020).
62. Singh, G. & Puckett, Y. Endometrial Hyperplasia. in *StatPearls [Internet]* (StatPearls Publishing, 2022).
63. Nees, L. K. *et al.* Endometrial hyperplasia as a risk factor of endometrial cancer. *Arch. Gynecol. Obstet.* **306**, 407–421 (2022).
64. Bakour, S. H., Khan, K. S. & Gupta, J. K. The risk of premalignant and malignant pathology in endometrial polyps. *Acta Obstet. Gynecol. Scand.* **79**, (2000).
65. Uglietti, A. *et al.* Endometrial polyps detected at ultrasound and rate of malignancy. *Arch. Gynecol. Obstet.* **289**, (2014).

66. Chaudhry, S., Reinhold, C., Guermazi, A., Khalili, I. & Maheshwari, S. Benign and Malignant Diseases of the Endometrium. *Top. Magn. Reson. Imaging* **14**, 339 (2003).
67. Uimari, O., Nazri, H. & Tapmeier, T. Endometriosis and Uterine Fibroids (Leiomyomata): Comorbidity, Risks and Implications. *Front. Reprod. Health* **3**, (2021).
68. Choi, E. J. *et al.* Comorbidity of gynecological and non-gynecological diseases with adenomyosis and endometriosis. *Obstet Gynecol Sci* **60**, 579–586 (2017).
69. Tan, G. C. & Yee Khong, T. Cyclic Endometrium and Exogenous Hormone Effect. *Gynecologic and Obstetric Pathology, Volume 1* 383–408 (2019).
70. Deligdisch, L. Hormonal pathology of the endometrium. *Mod. Pathol.* **13**, 285–294 (2000).
71. Bastianelli, C. *et al.* Effects of progestin-only contraceptives on the endometrium. *Expert Rev. Clin. Pharmacol.* **13**, 1103–1123 (2020).
72. Ludwig, H. The morphologic response of the human endometrium to long-term treatment with progestational agents. *Am. J. Obstet. Gynecol.* **142**, 796–808 (1982).
73. Song, J. Y. & Fraser, I. S. Effects of progestogens on human endometrium. *Obstet. Gynecol. Surv.* **50**, 385–394 (1995).
74. Dallenbach-Hellweg, G. Endometrial morphology during long-term use of levonorgestrel-releasing intrauterine devices. *Int. J. Gynecol. Pathol.* **7**, 95–96 (1988).
75. Critchley, H. O. *et al.* Morphological and functional features of endometrial decidualization following long-term intrauterine levonorgestrel delivery. *Hum. Reprod.* **13**, 1218–1224 (1998).
76. Phillips, V., Graham, C. T., Manek, S. & McCluggage, W. G. The effects of the levonorgestrel intrauterine system (Mirena coil) on endometrial morphology. *J. Clin. Pathol.* **56**, 305 (2003).
77. West, C. P. & Baird, D. T. Suppression of ovarian activity by Zoladex depot (ICI 118630), a long-acting luteinizing hormone releasing hormone agonist analogue. *Clin. Endocrinol.* **26**, 213–220 (1987).
78. Bellofiore, N. *et al.* First evidence of a menstruating rodent: the spiny mouse (Acomys cahirinus). *Am. J. Obstet. Gynecol.* **216**, 40.e1–40.e11 (2017).
79. Critchley, H. O. D. *et al.* Menstruation: science and society. *Am. J. Obstet. Gynecol.* **223**, 624 (2020).
80. Maenhoudt, N., De Moor, A. & Vankelecom, H. Modeling Endometrium Biology and Disease. *J Pers Med* **12**, (2022).
81. Fitzgerald, H. C., Schust, D. J. & Spencer, T. E. In vitro models of the human endometrium: evolution and application for women's health. *Biol. Reprod.* **104**, 282–293 (2021).
82. Murphy, A. R., Campo, H. & Kim, J. J. Strategies for modelling endometrial diseases. *Nat. Rev. Endocrinol.* **18**, 727–743 (2022).
83. Nishida M., Kasahara K., Kaneko M., Iwasaki H. & Hayashi K. [Establishment of a new human endometrial adenocarcinoma cell line, Ishikawa cells, containing estrogen and progesterone receptors].

Nihon Sanka Fujinka Gakkai Zasshi **37**, 1103–1111 (1985).

84. Turco, M. Y. *et al.* Long-term, hormone-responsive organoid cultures of human endometrium in a chemically defined medium. *Nat. Cell Biol.* **19**, 568–577 (2017).
85. Boretto, M. *et al.* Development of organoids from mouse and human endometrium showing endometrial epithelium physiology and long-term expandability. *Development* **144**, 1775–1786 (2017).
86. Gnecco, J. S. *et al.* Organoid co-culture model of the cycling human endometrium in a fully-defined synthetic extracellular matrix reveals epithelial-stromal crosstalk. *bioRxiv* 2021.09.30.462577 (2022) doi:10.1101/2021.09.30.462577.
87. Rawlings, T. M. *et al.* Modelling the impact of decidual senescence on embryo implantation in human endometrial assembloids. *Elife* **10**, e69603 (2021).
88. Ahn, J. *et al.* Three-dimensional microengineered vascularised endometrium-on-a-chip. *Hum. Reprod.* **36**, 2720–2731 (2021).
89. Lancaster, M. A. & Huch, M. Disease modelling in human organoids. *Dis. Model. Mech.* **12**, (2019).
90. Schutgens, F. & Clevers, H. Human Organoids: Tools for Understanding Biology and Treating Diseases. *Annu. Rev. Pathol.* **15**, 211–234 (2020).
91. Chumduri, C. & Turco, M. Y. Organoids of the female reproductive tract. *J. Mol. Med.* **99**, 531–553 (2021).
92. Boretto, M. *et al.* Patient-derived organoids from endometrial disease capture clinical heterogeneity and are amenable to drug screening. *Nat. Cell Biol.* **21**, 1041–1051 (2019).
93. Cindrova-Davies, T. *et al.* Menstrual flow as a non-invasive source of endometrial organoids. *Commun Biol* **4**, 651 (2021).
94. Wiwatpanit, T. *et al.* Scaffold-Free Endometrial Organoids Respond to Excess Androgens Associated With Polycystic Ovarian Syndrome. *J. Clin. Endocrinol. Metab.* **105**, (2020).
95. Abbas, Y. *et al.* Generation of a three-dimensional collagen scaffold-based model of the human endometrium. *Interface Focus* **10**, 20190079 (2020).
96. Krjutškov, K. *et al.* Single-cell transcriptome analysis of endometrial tissue. *Hum. Reprod.* **31**, (2016).
97. Wang, W. *et al.* Single-cell transcriptomic atlas of the human endometrium during the menstrual cycle. *Nat. Med.* **26**, 1644–1653 (2020).
98. Garcia-Alonso, L. *et al.* Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nat. Genet.* **53**, 1698–1711 (2021).
99. Tan, Y. *et al.* Single-cell analysis of endometriosis reveals a coordinated transcriptional programme driving immunotolerance and angiogenesis across eutopic and ectopic tissues. *Nat. Cell Biol.* **24**, 1306–1318 (2022).
100. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).

101. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
102. Wang, J, Fan, HC, Behr, B, Quake, SR. Genome-wide Single-Cell Analysis of Recombination Activity and De Novo Mutation Rates in Human Sperm. *Cell* **150**, 402–412 (2012).
103. Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626 (2012).
104. Sanders, A. D. *et al.* Single-cell analysis of structural variations and complex rearrangements with tri-channel processing. *Nat. Biotechnol.* **38**, (2020).
105. Fan, X. *et al.* SMOOTH-seq: single-cell genome sequencing of human cells on a third-generation sequencing platform. *Genome Biol.* **22**, 1–19 (2021).
106. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
107. Mulqueen, R. M. *et al.* Highly scalable generation of DNA methylation profiles in single cells. *Nat. Biotechnol.* **36**, 428–431 (2018).
108. Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11**, 817–820 (2014).
109. Cusanovich, D. A. *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
110. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
111. Peterson, V. M. *et al.* Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936–939 (2017).
112. Shahi, P., Kim, S. C., Haliburton, J. R., Gartner, Z. J. & Abate, A. R. Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci. Rep.* **7**, 1–12 (2017).
113. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604 (2018).
114. Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018).
115. Parikh, K. *et al.* Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature* vol. 567 49–55 Preprint at https://doi.org/10.1038/s41586-019-0992-y (2019).
116. Proserpio, V. & Lönnberg, T. Single-cell technologies are revolutionizing the approach to rare cells. *Immunol. Cell Biol.* **94**, 225–229 (2016).
117. Braga, F. A. V. *et al.* A cellular census of human lungs identifies novel cell states in health and in asthma. *Nature Medicine* vol. 25 1153–1163 Preprint at https://doi.org/10.1038/s41591-019-0468-5 (2019).
118. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter

droplets. *Cell* **161**, 1202–1214 (2015).

119. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).

120. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.* **38**, 708–714 (2020).

121. Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* **9**, 1–12 (2017).

122. Baran-Gale, J., Chandra, T. & Kirschner, K. Experimental design for single-cell RNA sequencing. *Brief. Funct. Genomics* **17**, 233–239 (2018).

123. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).

124. Barthelson, R. A., Lambert, G. M., Vanier, C., Lynch, R. M. & Galbraith, D. W. Comparison of the contributions of the nuclear and cytoplasmic compartments to global gene expression in human cells. *BMC Genomics* **8**, 340 (2007).

125. Bakken, T. E. *et al.* Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS One* **13**, e0209648 (2018).

126. Lake, B. B. *et al.* A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci. Rep.* **7**, 6031 (2017).

127. Chen, A, *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).

128. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).

129. Lewis, S. M. *et al.* Spatial omics and multiplexed imaging to explore cancer biology. *Nat. Methods* **188**, 997–1012 (2021).

130. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).

131. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* **37**, 1452–1457 (2019).

132. Zhang, B. *et al.* Characterizing cellular heterogeneity in chromatin state with scCUT&Tag-pro. *Nat Biotechnol* **40**, 1220–1230 (2022).

133. McKinnon, B. D. *et al.* Altered differentiation of endometrial mesenchymal stromal fibroblasts is associated with endometriosis susceptibility. *Communications Biology* **5**, 1–14 (2022).

134. Cao, D., Chan, R. W. S., Ng, E. H. Y., Gemzell-Danielsson, K. & Yeung, W. S. B. Single-cell RNA sequencing of cultured human endometrial CD140b+CD146+ perivascular cells highlights the importance

of in vivo microenvironment. *Stem Cell Research & Therapy* **12**, 206 (2021).

135. Zhang, X. *et al.* Single-cell transcriptome analysis uncovers the molecular and cellular characteristics of thin endometrium. *FASEB J.* **36**, e22193 (2022).

136. Lv, H. *et al.* Deciphering the endometrial niche of human thin endometrium at single-cell resolution. *Proc Natl Acad Sci U S A* **119**, e2115912119 (2022).

137. Liu, Z. *et al.* Single-cell transcriptomic analysis of eutopic endometrium and ectopic lesions of adenomyosis. *Cell Biosci.* **11**, 51 (2021).

138. Chen, P. *et al.* The Immune Atlas of Human Deciduas With Unexplained Recurrent Pregnancy Loss. *Frontiers in Immunology* **12**, 689019 (2021).

139. Lai, Z.-Z. *et al.* Single-cell transcriptome profiling of the human endometrium of patients with recurrent implantation failure. *Theranostics* **12**, 6527–6547 (2022).

140. Lucas, E. S. *et al.* Recurrent pregnancy loss is associated with a pro-senescent decidual response during the peri-implantation window. *Commun Biol* **3**, 37 (2020).

141. Hashimoto, S. *et al.* Comprehensive single-cell transcriptome analysis reveals heterogeneity in endometrioid adenocarcinoma tissues. *Sci. Rep.* **7**, 14225 (2017).

142. Guo, Y.-E. *et al.* Phenotyping of immune and endometrial epithelial cells in endometrial carcinomas revealed by single-cell RNA sequencing. *Aging* **13**, 6565–6591 (2021).

143. Ma, J. *et al.* Single-cell transcriptomic analysis of endometriosis provides insights into fibroblast fates and immune cell heterogeneity. *Cell & Bioscience* **11**, 125 (2021).

144. Fonseca, M. A. S. *et al.* Single-cell transcriptomic analysis of endometriosis. *Nat. Genet.* **55**, 255–267 (2023).

145. Shih, A. J. *et al.* Single-cell analysis of menstrual endometrial tissues defines phenotypes associated with endometriosis. *BMC Med.* **20**, 315 (2022).

146. Bunis, D. G. *et al.* Whole-Tissue Deconvolution and scRNAseq Analysis Identify Altered Endometrial Cellular Compositions and Functionality Associated With Endometriosis. *Front. Immunol.* **12**, 788315 (2021).

147. Baugh, L. *et al.* Integrating endometrial proteomic and single cell transcriptomic pipelines reveals distinct menstrual cycle and endometriosis-associated molecular profiles. *bioRxiv* (2022) doi:10.1101/2022.01.29.22269829.

148. Fitzgerald, H. C., Dhakal, P., Behura, S. K., Schust, D. J. & Spencer, T. E. Self-renewing endometrial epithelial organoids of the human uterus. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23132–23142 (2019).

149. Cochrane, D. R. *et al.* Single cell transcriptomes of normal endometrial derived organoids uncover novel cell type markers and cryptic differentiation of primary tumours. *J. Pathol.* **252**, 201–214 (2020).

150. Camp, J. G., Wollny, D. & Treutlein, B. Single-cell genomics to guide human stem cell and tissue

engineering. *Nat. Methods* **15**, 661–667 (2018).

151. Noyes, R. W., Hertig, A. T. & Rock, J. Dating the endometrial biopsy. *Am. J. Obstet. Gynecol.* **122**, 262–263 (1975).

152. Tran, H. T. N. *et al.* A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* **21**, 1–32 (2020).

153. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).

154. Tempest, N., Baker, A. M., Wright, N. A. & Hapangama, D. K. Does human endometrial LGR5 gene expression suggest the existence of another hormonally regulated epithelial stem cell niche? *Hum. Reprod.* **33**, 1052–1062 (2018).

155. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 1–12 (2018).

156. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 1–12 (2017).

157. Heaton, H. *et al.* Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620 (2020).

158. Bernstein, NJ *et al.* Solo: Doublet Identification in Single-Cell RNA-Seq via Semi-Supervised Deep Learning. *Cell Systems* **11**, 95–101.e5 (2020).

159. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 1–5 (2018).

160. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

161. Garcia-Alonso, L. *et al.* Single-cell roadmap of human gonadal development. *Nature* **607**, 540–547 (2022).

162. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281–291.e9 (2019).

163. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

164. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **3**, 861 (2018).

165. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, (2020).

166. Young, M. Single cell differential expression testing. http://constantamateur.github.io/2020-04-10-scDE/.

167. Schwab, K. E. & Gargett, C. E. Co-expression of two perivascular cell markers isolates mesenchymal

stem-like cells from human endometrium. *Hum. Reprod.* **22**, 2903–2911 (2007).

168. Spitzer, T. L. *et al.* Perivascular human endometrial mesenchymal stem cells express pathways relevant to self-renewal, lineage specification, and functional phenotype. *Biol. Reprod.* **86**, 58 (2012).

169. Salamonsen, L. A., Hutchison, J. C. & Gargett, C. E. Cyclical endometrial repair and regeneration. *Development* **148**, dev199577 (2021).

170. Queckbörner, S. *et al.* Stromal Heterogeneity in the Human Proliferative Endometrium-A Single-Cell RNA Sequencing Study. *J Pers Med* **11**, 448 (2021).

171. Uniken Venema, W. T. *et al.* Gut mucosa dissociation protocols influence cell type proportions and single-cell gene expression levels. *Sci. Rep.* **12**, 9897 (2022).

172. Diaz-Gimeno, P. A *et al.* A genomic diagnostic tool for human endometrial receptivity based on the transcriptomic signature. *Fertil. Steril.* **95**, 50–60.e15 (2011).

173. Suhorutshenko, M. *et al.* Endometrial receptivity revisited: endometrial transcriptome adjusted for tissue cellular heterogeneity. *Hum. Reprod.* **33**, 2074–2086 (2018).

174. GeneCards Human Gene Database. TENM2 Gene - GeneCards. https://www.genecards.org/cgi-bin/carddisp.pl?gene=TENM2&keywords=TENM2.

175. Tucker, R. P., Kenzelmann, D., Trzebiatowska, A. & Chiquet-Ehrismann, R. Teneurins: transmembrane proteins with fundamental roles in development. *Int. J. Biochem. Cell Biol.* **39**, 292–297 (2007).

176. Liu, L. & Hu, D. SUFU reduced pancreatic cancer cell growth by Wnt/β-catenin signaling pathway. *Food Sci. Technol.* **42**, (2022).

177. SUFU negative regulator of hedgehog signaling [Homo sapiens (human)] - Gene - NCBI. https://www.ncbi.nlm.nih.gov/gene/51684.

178. Wang, Y., van der Zee, M., Fodde, R. & Blok, L. J. Wnt/B-Catenin and Sex Hormone Signaling In Endometrial Homeostasis and Cancer. *Oncotarget* **1**, 674–684 (2010).

179. Kiewisz, J., Wasniewski, T. & Kmiec, Z. Participation of WNT and β-Catenin in Physiological and Pathological Endometrial Changes: Association with Angiogenesis. *Biomed Res. Int.* **2015**, (2015).

180. Ma, A.-Y., Xie, S.-W., Zhou, J.-Y. & Zhu, Y. Nomegestrol Acetate Suppresses Human Endometrial Cancer RL95-2 Cells Proliferation In Vitro and In Vivo Possibly Related to Upregulating Expression of SUFU and Wnt7a. *Int. J. Mol. Sci.* **18**, 1337 (2017).

181. Ka, H. *et al.* Regulation of Expression of Fibroblast Growth Factor 7 in the Pig Uterus by Progesterone and Estradiol1. *Biol. Reprod.* **77**, 172–180 (2007).

182. FGF7 fibroblast growth factor 7 [Homo sapiens (human)] - Gene - NCBI. https://www.ncbi.nlm.nih.gov/gene/2252.

183. Werner, S. & Grose, R. Regulation of wound healing by growth factors and cytokines. *Physiol. Rev.* **83**, 835–870 (2003).

184. Nie, G. *et al.* Inhibiting uterine PC6 blocks embryo implantation: an obligatory role for a proprotein convertase in fertility. *Biol. Reprod.* **72**, 1029–1036 (2005).

185. Nie, G. Y., Li, Y., Minoura, H., Findlay, J. K. & Salamonsen, L. A. Specific and transient up-regulation of proprotein convertase 6 at the site of embryo implantation and identification of a unique transcript in mouse uterus during early pregnancy. *Biol Reprod* **68**, 439–447 (2003).

186. Okada, H., Nie, G. & Salamonsen, L. A. Requirement for proprotein convertase 5/6 during decidualization of human endometrial stromal cells in vitro. *J. Clin. Endocrinol. Metab.* **90**, 1028–1034 (2005).

187. Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018).

188. Burney, R. O. *et al.* Gene Expression Analysis of Endometrium Reveals Progesterone Resistance and Candidate Susceptibility Genes in Women with Endometriosis. *Endocrinology* **148**, 3814–3826 (2007).

189. Bulun, S.E. *et.al*. Progesterone resistance in endometriosis: Link to failure to metabolize estradiol. *Mol. Cell. Endocrinol.* **248**, 94–103 (2006).

190. United Nations. *Contraceptive Use by Method 2019: Data Booklet*. (United Nations, 2019).

191. Smith-McCune, K. *et al.* Differential Effects of the Hormonal and Copper Intrauterine Device on the Endometrial Transcriptome. *Sci. Rep.* **10**, 6888 (2020).

192. Rao, D. A. *et al.* A protocol for single-cell transcriptomics from cryopreserved renal tissue and urine for the Accelerating Medicine Partnership (AMP) RA/SLE network. *bioRxiv* 275859 (2018) doi:10.1101/275859.

193. Donlin, L. T. *et al.* Methods for high-dimensional analysis of cells dissociated from cryopreserved synovial tissue. *Arthritis Res. Ther.* **20**, 139 (2018).

194. Guillaumet-Adkins, A. *et al.* Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol.* **18**, 1–15 (2017).

195. Aldridge, S. & Teichmann, S. A. Single cell transcriptomics comes of age. *Nat. Commun.* **11**, 1–4 (2020).

196. van den Brink, S. C. *et al.* Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017).

197. O'Flanagan, C. H. *et al.* Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. *Genome Biol.* **20**, 1–13 (2019).

198. Lacar, B. *et al.* Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat. Commun.* **7**, 11022 (2016).

199. Adam, M., Potter, A. S. & Potter, S. S. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. *Development* **144**, 3625–3632 (2017).

200. Wu, H., Kirita, Y., Donnelly, E. L. & Humphreys, B. D. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *J. Am.*

Soc. Nephrol. **30**, 23–32 (2019).

201. Ding, J. *et al.* Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **38**, 737–746 (2020).
202. Zappia, L. & Theis, F. J. Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape. *Genome Biol.* **22**, 1–18 (2021).
203. Yang, S. *et al.* Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol.* **21**, 1–15 (2020).
204. Egan, D. *et al.* CRABP2 – A novel biomarker for high-risk endometrial cancer. *Gynecol. Oncol.* **167**, 314–322 (2022).
205. Hoffman, S.R. *et al*. Reasons for hormonal contraceptive use in a cohort of African-American women living in the Detroit area. *Contraception* **102**, 346–348 (2020).
206. Jones, R. K. *Beyond Birth Control: The Overlooked Benefits Of Oral Contraceptive Pills*. (2011).
207. Dinh, A. *et al*. A review of the endometrial histologic effects of progestins and progesterone receptor modulators in reproductive age women. *Contraception* **91**, 360–367 (2015).
208. Park, J. S. *et al.* Endometrium from women with endometriosis shows increased proliferation activity. *Fertil. Steril.* **92**, 1246–1249 (2009).
209. Tomassetti, C. & D'Hooghe, T. Endometriosis and infertility: Insights into the causal link and management strategies. *Best Pract. Res. Clin. Obstet. Gynaecol.* **51**, 25–33 (2018).
210. Tanbo, T. & Fedorcsak, P. Endometriosis-associated infertility: aspects of pathophysiological mechanisms and treatment options. *Acta Obstet. Gynecol. Scand.* **96**, 659–667 (2017).
211. Filip, L. *et al.* Endometriosis Associated Infertility: A Critical Review and Analysis on Etiopathogenesis and Therapeutic Approaches. *Medicina* **56**, 460 (2020).
212. Young, S.L. *et al.* Oestrogen and progesterone action on endometrium: a translational approach to understanding endometrial receptivity. *Reprod. Biomed. Online* **27**, 497–505 (2013).
213. Wei, Q., St Clair, J. B., Fu, T., Stratton, P. & Nieman, L. K. Reduced expression of biomarkers associated with the implantation window in women with endometriosis. *Fertil. Steril.* **91**, 1686–1691 (2009).
214. Da Broi, M. G., Ferriani, R. A. & Navarro, P. A. Ethiopathogenic mechanisms of endometriosis-related infertility. *JBRA Assisted Reproduction* **23**, 273 (2019).
215. Lessey, B.A. *et al*. Endometrial receptivity in the eutopic endometrium of women with endometriosis: it is affected, and let me show you why. *Fertil. Steril.* **108**, 19–27 (2017).
216. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nature Protocols* **15**, 1484–1506 (2020).
217. Nikolakopoulou, K. & Turco, M. Y. Investigation of infertility using endometrial organoids. *Reproduction*

161, R113–R127 (2021).

218. Alzamil, L., Nikolakopoulou, K. & Turco, M. Y. Organoid systems to study the human female reproductive tract and pregnancy. *Cell Death Differ.* **28**, 35–51 (2021).

219. Zhang, J.-W. *et al.* Metabolic profiling and cytochrome P450 reaction phenotyping of medroxyprogesterone acetate. *Drug Metab. Dispos.* **36**, 2292–2298 (2008).

220. Okada, H., Tsuzuki, T. & Murata, H. Decidualization of the human endometrium. *Reprod. Med. Biol.* **17**, 220–227 (2018).

221. Gellersen, B. & Brosens, J. Cyclic AMP and progesterone receptor cross-talk in human endometrium: a decidualizing affair. *J. Endocrinol.* **178**, 357–372 (2003).

222. Yoshie, M., Kusama, K. & Tamura, K. Molecular Mechanisms of Human Endometrial Decidualization Activated by Cyclic Adenosine Monophosphate Signaling Pathways. *Journal of Mammalian Ova Research* **32**, 95–102 (2015).

223. Useful numbers for cell culture - UK. https://www.thermofisher.com/uk/en/home/references/gibco-cell-culture-basics/cell-culture-protocols/cell-culture-useful-numbers.html.

224. Turco, M., Burton, G., Gardner, L., Koo, B.-K. & Moffett, A. Derivation and long-term expansion of human endometrial and decidual organoids. (2017) doi:10.1038/protex.2017.030.

225. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

226. Sun, D. *et al.* A functional genetic toolbox for human tissue-derived organoids. *Elife* **10**, e67886 (2021).

227. Tang, B., Guller, S. & Gurpide, E. Cyclic adenosine 3',5'-monophosphate induces prolactin expression in stromal cells isolated from human proliferative endometrium. *Endocrinology* **133**, 2197–2203 (1993).

228. Brosens, J. J., Hayashi, N. & White, J. O. Progesterone receptor regulates decidual prolactin expression in differentiating human endometrial stromal cells. *Endocrinology* **140**, 4809–4820 (1999).

229. Graesslin, O. *et al*. Metalloproteinase-2, -7 and -9 and tissue inhibitor of metalloproteinase-1 and -2 expression in normal, hyperplastic and neoplastic endometrium: a clinical-pathological correlation study. *Ann. Oncol.* **17**, 637–645 (2006).

230. Guillemette, C. Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. *The Pharmacogenomics Journal* **3**, 136–158 (2003).

231. Zhu, B. T. & Conney, A. H. Functional role of estrogen metabolism in target cells: review and perspectives. *Carcinogenesis* **19**, 1–27 (1998).

232. Hevir, N. *et al.* Disturbed balance between phase I and II metabolizing enzymes in ovarian endometriosis: A source of excessive hydroxy-estrogens and ROS? *Mol. Cell. Endocrinol.* **367**, 74–84 (2013).

233. Madsen, J., Nielsen, O., Tornøe, I., Thim, L. & Holmskov, U. Tissue localization of human trefoil factors 1, 2, and 3. *J. Histochem. Cytochem.* **55**, 505–513 (2007).

234. Wiede, A. *et al.* Synthesis and localization of the mucin-associated TFF-peptides in the human uterus. *Cell Tissue Res.* **303**, 109–115 (2001).

235. Henze, D. *et al.* Endometriosis Leads to an Increased Trefoil Factor 3 Concentration in the Peritoneal Cavity but Does Not Alter Systemic Levels. *Reprod. Sci.* **24**, 258–267 (2017).

236. Bignotti, E. *et al.* Trefoil factor 3: a novel serum marker identified by gene expression profiling in high-grade endometrial carcinomas. *Br. J. Cancer* **99**, 768–773 (2008).

237. Mhawech-Fauceglia, P. *et al.* Trefoil factor family 3 (TFF3) expression and its interaction with estrogen receptor (ER) in endometrial adenocarcinoma. *Gynecologic Oncology* **130**, 174–180 (2013).

238. Du, T.-Y. *et al.* Circulating serum trefoil factor 3 (TFF3) is dramatically increased in chronic kidney disease. *PLoS One* **8**, e80271 (2013).

239. Srivastava, S. *et al.* Serum human trefoil factor 3 is a biomarker for mucosal healing in ulcerative colitis patients with minimal disease activity. *J. Crohns. Colitis* **9**, 575–579 (2015).

240. Göke, M. N. *et al.* Trefoil peptides promote restitution of wounded corneal epithelial cells. *Exp. Cell Res.* **264**, 337–344 (2001).

241. Kjellev, S. The trefoil factor family - small peptides with multiple functionalities. *Cell. Mol. Life Sci.* **66**, 1350–1369 (2009).

242. Taupin, D. R., Kinoshita, K. & Podolsky, D. K. Intestinal trefoil factor confers colonic epithelial resistance to apoptosis. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 799–804 (2000).

243. Dignass, A., Lynch-Devaney, K., Kindon, H., Thim, L. & Podolsky, D. K. Trefoil peptides promote epithelial migration through a transforming growth factor beta-independent pathway. *J. Clin. Invest.* **94**, 376–383 (1994).

244. Yang, G.-D. *et al.* SERPINA3 promotes endometrial cancer cells growth by regulating G2/M cell cycle checkpoint and apoptosis. *Int. J. Clin. Exp. Pathol.* **7**, 1348–1358 (2014).

245. de Mezer, M. *et al.* SERPINA3: Stimulator or Inhibitor of Pathological Changes. *Biomedicines* **11**, 156 (2023).

246. Marquardt, R. M., Kim, T. H., Shin, J.-H. & Jeong, J.-W. Progesterone and Estrogen Signaling in the Endometrium: What Goes Wrong in Endometriosis? *Int. J. Mol. Sci.* **20**, (2019).

247. Poorasamy, J., Sengupta, J., Patil, A. & Ghosh, D. Progesterone Resistance in Endometriosis. *Eur. Med. J. Reprod. Health* **8**, 51–63 (2022).

248. Tulac, S. *et al.* Dickkopf-1, an inhibitor of Wnt signaling, is regulated by progesterone in human endometrial stromal cells. *J. Clin. Endocrinol. Metab.* **91**, 1453–1461 (2006).

249. SLC7A2 solute carrier family 7 member 2 [Homo sapiens (human)] - Gene - NCBI. https://www.ncbi.nlm.nih.gov/gene/6542.

250. GeneCards Human Gene Database. CPM gene - GeneCards. https://www.genecards.org/cgi-

bin/carddisp.pl?gene=CPM.

251. Masuda, H., Anwar, S. S., Bühring, H. J., Rao, J. R. & Gargett, C. E. A novel marker of human endometrial mesenchymal stem-like cells. *Cell Transplant.* **21**, 2201–2214 (2012).

252. Dominici, M. *et al.* Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* **8**, 315–317 (2006).

253. Selvaraj, P. *et al.* Neurotrophic Factor-α1: A Key Wnt-β-Catenin Dependent Anti-Proliferation Factor and ERK-Sox9 Activated Inducer of Embryonic Neural Stem Cell Differentiation to Astrocytes in Neurodevelopment. *Stem Cells* **35**, 557–571 (2016).

254. Li, T. *et al.* WNT5A Interacts With FZD5 and LRP5 to Regulate Proliferation and Self-Renewal of Endometrial Mesenchymal Stem-Like Cells. *Frontiers in cell and developmental biology* **10**, 837827 (2022).

255. Bänziger, C. *et al.* Wntless, a Conserved Membrane Protein Dedicated to the Secretion of Wnt Proteins from Signaling Cells. *Cell* **125**, 509–522 (2006).

256. Bartscherer, K., Pelte, N., Ingelfinger, D. & Boutros, M. Secretion of Wnt Ligands Requires Evi, a Conserved Transmembrane Protein. *Cell* **125**, 523–533 (2006).

257. Cheng, C-W. *et al.* Transcript profile and localization of Wnt signaling–related molecules in human endometrium. *Fertil. Steril.* **90**, 201–204 (2008).

258. Zhao, Y.-Y. *et al.* Transcriptional activation of insulin-like growth factor binding protein 6 by 17β-estradiol in SaOS-2 cells. *Exp. Mol. Med.* **41**, 478 (2009).

259. Martin, J. L., Coverley, J. A., Pattison, S. T. & Baxter, R. C. Insulin-like growth factor-binding protein-3 production by MCF-7 breast cancer cells: stimulation by retinoic acid and cyclic adenosine monophosphate and differential effects of estradiol. *Endocrinology* **136**, 1219–1226 (1995).

260. Park, Y. & Han, S. J. Interferon Signaling in the Endometrium and in Endometriosis. *Biomolecules* **12**, 1554 (2022).

261. Li, Q. *et al.* Identification and Implantation Stage-Specific Expression of an Interferon-α-Regulated Gene in Human and Rat Endometrium. *Endocrinology* **142**, 2390–2400 (2001).

262. Zhao, X. *et al.* Interferon-stimulated gene 15 promotes progression of endometrial carcinoma and weakens antitumor immune response. *Oncol. Rep.* **47**, 110 (2022).

263. Catalano, R. D., Wilson, M. R., Boddy, S. C. & Jabbour, H. N. Comprehensive expression analysis of prostanoid enzymes and receptors in the human endometrium across the menstrual cycle. *Mol. Hum. Reprod.* **17**, 182–192 (2011).

264. Catalano, R. D. *et al.* Mifepristone induced progesterone withdrawal reveals novel regulatory pathways in human endometrium. *Mol. Hum. Reprod.* **13**, 641–654 (2007).

265. Neavin, D. *et al.* Single cell eQTL analysis identifies cell type-specific genetic control of gene expression

in fibroblasts and reprogrammed induced pluripotent stem cells. *Genome Biol.* **22**, 76 (2021).

266. Nathan, A. *et al.* Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature* **606**, 120–128 (2022).

267. Yazar, S. *et al.* Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* **376**, eabf3041 (2022).

268. Jia, P., Hu, R., Yan, F., Dai, Y. & Zhao, Z. scGWAS: landscape of trait-cell type associations by integrating single-cell transcriptomics-wide and genome-wide association studies. *Genome Biol.* **23**, 220 (2022).

269. Elmentaite, R. *et al.* Cells of the human intestinal tract mapped across space and time. *Nature* **597**, 250–255 (2021).

270. Sliz, E. *et al.* Evidence of a causal effect of genetic tendency to gain muscle mass on uterine leiomyomata. *Nat. Commun.* **14**, 1–14 (2023).

271. Wang, X., Glubb, D. M. & O'Mara, T. A. 10 Years of GWAS discovery in endometrial cancer: Aetiology, function and translation. *eBioMedicine* **77**, 103895 (2022).

272. Datir, S. G. & Bhake, A. Management of Uterine Fibroids and Its Complications During Pregnancy: A Review of Literature. *Cureus* **14**, e31080 (2022).

273. Liu, Z., Doan, Q. V., Blumenthal, P. & Dubois, R. W. A systematic review evaluating health-related quality of life, work impairment, and health-care costs and utilization in abnormal uterine bleeding. *Value Health* **10**, 183–194 (2007).

274. Mercuri, N. D. & Cox, B. J. Meta-Research: A Poor Research Landscape Hinders the Progression of Knowledge and Treatment of Reproductive Diseases. *bioRxiv* 2021.11.16.468787 (2021) doi:10.1101/2021.11.16.468787.

275. Rozenblatt-Rosen, O., Stubbington, M. J. T., Regev, A. & Teichmann, S. A. The Human Cell Atlas: from vision to reality. *Nature* **550**, 451–453 (2017).

276. Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, e27041 (2017).

Appendices

Appendix 1:
Comparison of fresh vs frozen endometrial biopsies for scRNAseq

(As presented at the World Endometriosis Congress 2021)

BACKGROUND

The human endometrium:
- has heterogeneous architecture, comprised of many different cell types
- is the likely tissue of the origin of endometriosis
- prior published data focuses on bulk RNA-sequencing that characterises 'average' gene expression of whole tissue
- single-cell RNA-sequencing (scRNAseq) allows characterisation at the cellular level

scRNAseq:
- mostly used with freshly collected samples, freezing/thawing of tissues *may* reduce the quality of the cells studied
- cryopreservation and banking of frozen samples is highly desirable in a clinical set-up
- banking allows for more controlled experimental design and thus generation of robust, reproducible data

AIMS

In order to address the challenges of collecting and storing relevant clinical samples, the objectives of this work were to:

1) Compare the transcriptomic profiles of cryopreserved and fresh tissue samples
2) Characterise the cellular composition of endometrial tissue in fresh and frozen samples using scRNAseq

MATERIALS & METHODS

Sample processing:

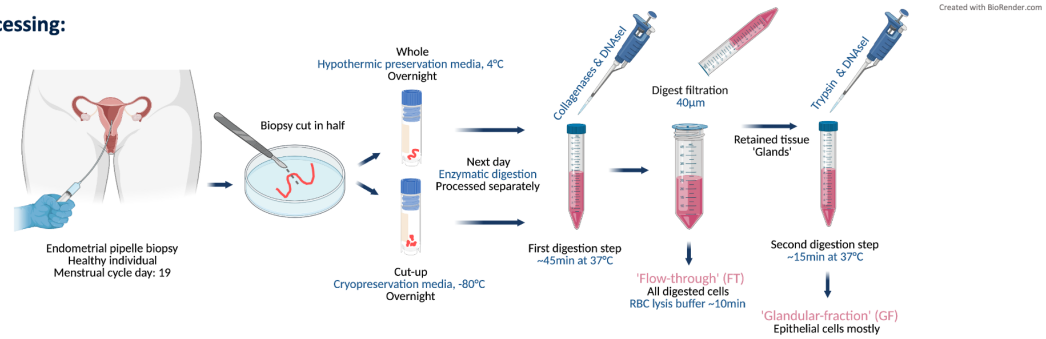


Figure 1: Experimental design. Endometrial pipelle biopsy was divided, half stored at 4°C in preservation solution overnight and the other frozen at -80°C . The next day samples were processed simultaneously. Two-step enzymatic dissociation yielded two cellular fractions per sample: flow-through (FT) containing all cells, and glandular fraction (GF): enriched for epithelial glandular cells.

Data generation & analysis: Cells were loaded onto the 10X Chromium Controller using Chromium Single Cell 3' Kit v3.1. aiming for recovery of ~3,000 cells/reaction. Libraries were prepared according to manufacturer's protocol and sequenced (Illumina Hiseq4000) to ~40,000 reads/cell. CellRanger v3.1.0 and Scanpy were used for alignment and data analysis. Cell types were annotated using regularised logistic approach and a reference dataset generated for freshly processed healthy endometrium[1].

RESULTS

Quality control metrics for fresh & frozen samples

The library size and complexity as well as number of cells recovered and sequenced was comparable between the fresh and frozen samples (*Table 1*). The number of genes, individual mRNA molecules and percentage of mitochondrial genes detected per cell did not vary between conditions (*Figure 2*). Higher mitochondrial gene expression was detected in the epithelial cells enriched fraction termed 'GF.' This was independent of sample freezing (*Figure 2C*).

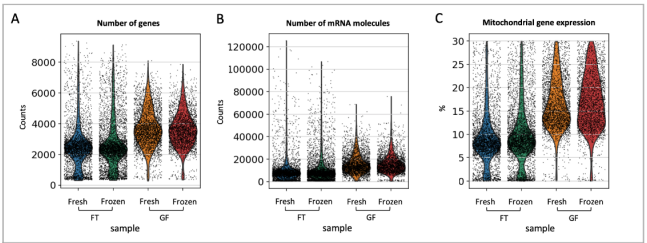


Figure 2: Plots of quality control metrics. Number of genes (A), number of mRNA molecules (B) and percentage of mitochondrial genes (C) detected per cell in the different samples. *(FT= flow-through fraction, GF = glandular fraction)*

Table 1: Data metrics
(FT= flow-through fraction, GF = glandular fraction, CR= CellRanger)

Sample	CR estimated cell numbers	Median genes/cell	Mean number of reads/cell	Reads confidently mapped to genome
Fresh_FT	4,495	2,221	34,994	81.3%
Frozen_FT	5,328	2,255	36,668	85.2%
Fresh_GF	4,428	3,054	39,146	84.2%
Frozen_GF	4,166	2,796	39,317	84.1%

Clustering & cell state identification

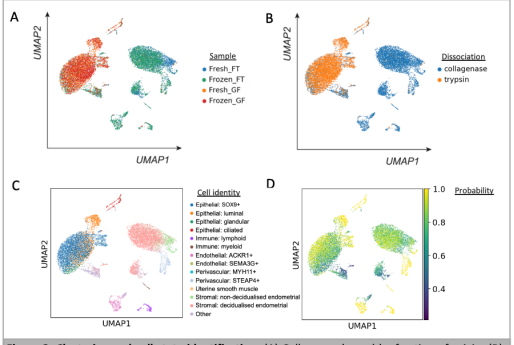


Figure 3: Clustering and cell state identification. (A) Cells are coloured by fraction of origin. (B) Colours correspond to the two tissue dissociation steps. (C) Different cell states identified and (D) the probabilities with which these were assigned. *(FT= flow-through fraction, GF = glandular fraction)*

Freezing has no effect on cell distribution nor assigned identities *(Figure 3)*. The main difference comes from the tissue dissociation process: cells obtained after collagenase incubation cluster closely together, as do cells obtained from trypsin treatment. The logistic regression model assigns cell identities with high probabilities and detects 10 distinct cell states.

RESULTS

Cell identities are the same for fresh & frozen samples

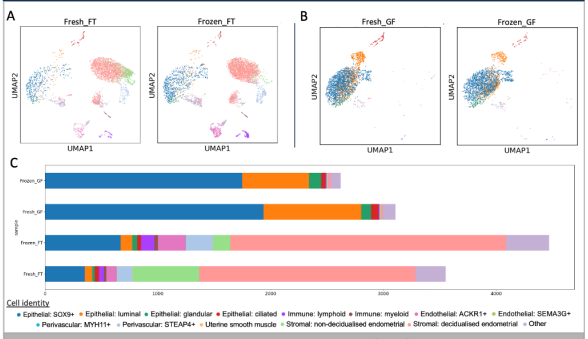


Figure 4: UMAP visualisation of cells identified & their quantification. (A & B) show UMAP distribution of cells identified in both the FT and GF fresh and frozen fractions. The actual numbers of cells detected per cell identity and the different conditions are shown in (C). *(FT= flow-through fraction, GF = glandular fraction)*

The 10 different cell states identified (*epithelial:* SOX9+, luminal, glandular, ciliated; *stromal:* non-decidualised endometrial, decidualised endometrial; *immune:* myeloid, lymphoid; *endothelial:* ACKR1+; *perivascular:* STEAP4+) are detected in both fresh and frozen samples in similar proportions *(Figure 4)*, showing that freezing does not negatively influence a certain cell type/state nor causes its loss. Limited uterine smooth muscle, endothelial SEMA3G+ and perivascular PVMYH11+ cells were identified, due to the menstrual cycle phase timing and the biopsy being superficial. Cells identified as 'other' require further investigation.

Gene expression profile does not vary

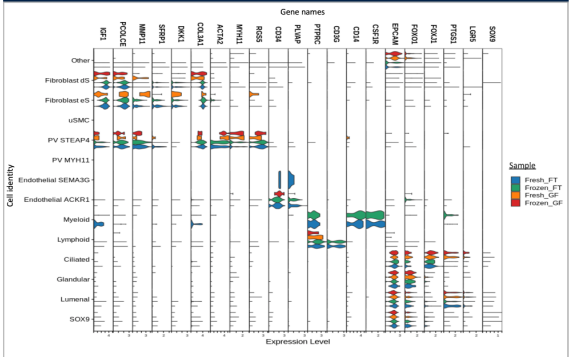


Figure 5: Gene expression profiles of fresh and frozen samples. Expression levels of example genes that distinguish different cell identities are shown as violins. Violin colours represent the fraction of origin. *(FT= flow-through fraction, GF = glandular fraction)*

The cells' expression profiles do not change as a result of freezing. Similar expression values are observed for the fresh and frozen samples *(Figure 5)* for both highly expressed genes, such as EpCAM, and genes with much lower expression, such as SOX9 or LGR5. All genes are expressed at similar levels across all cell identities, irrespective of freezing.

CONCLUSIONS

Cryopreservation of endometrial superficial biopsies does not:

- reduce the cell numbers detected
- impact the quality of libraries generated
- affect cellular transcriptomic profiles
- lead to observable loss of any cell states/types

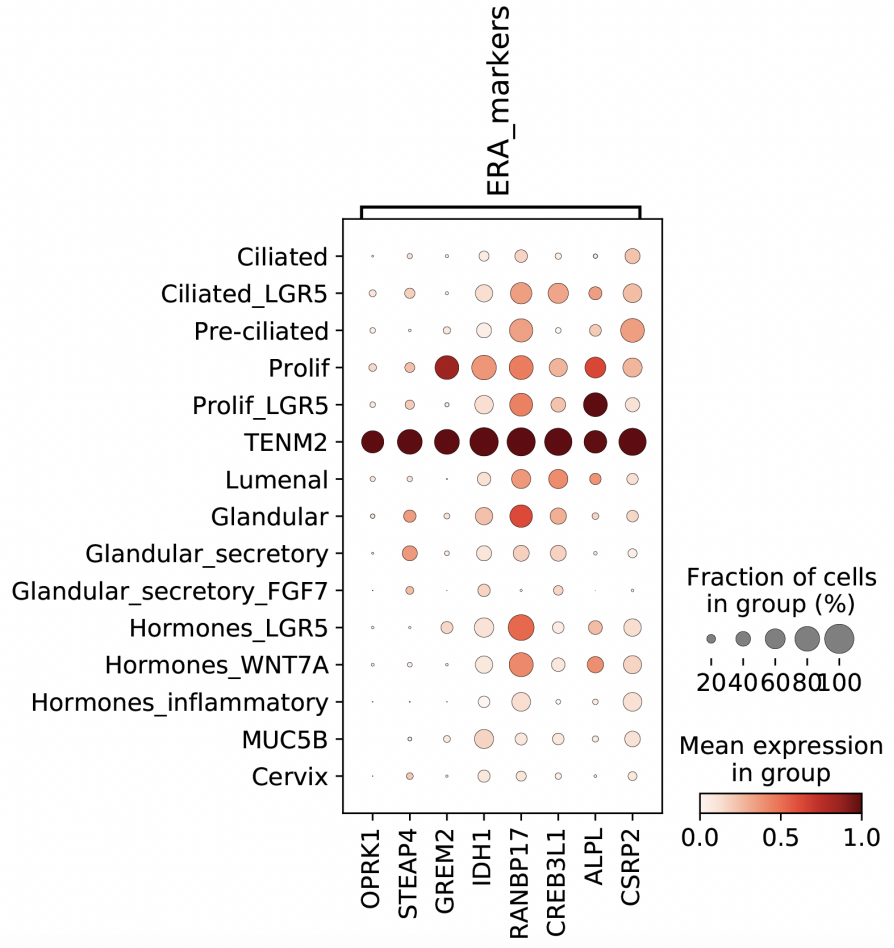
It is therefore feasible to cryopreserve endometrial samples using our protocol for scRNAseq experiments.

REFERENCES

1. Luz Garcia-Alonso, Louis-François Handfield, et al. (2021) Mapping the temporal and spatial dynamics of the human endometrium *in vivo* and *in vitro*. *bioRxiv 2021.01.02.425073*; doi: https://doi.org/10.1101/2021.01.02.425073

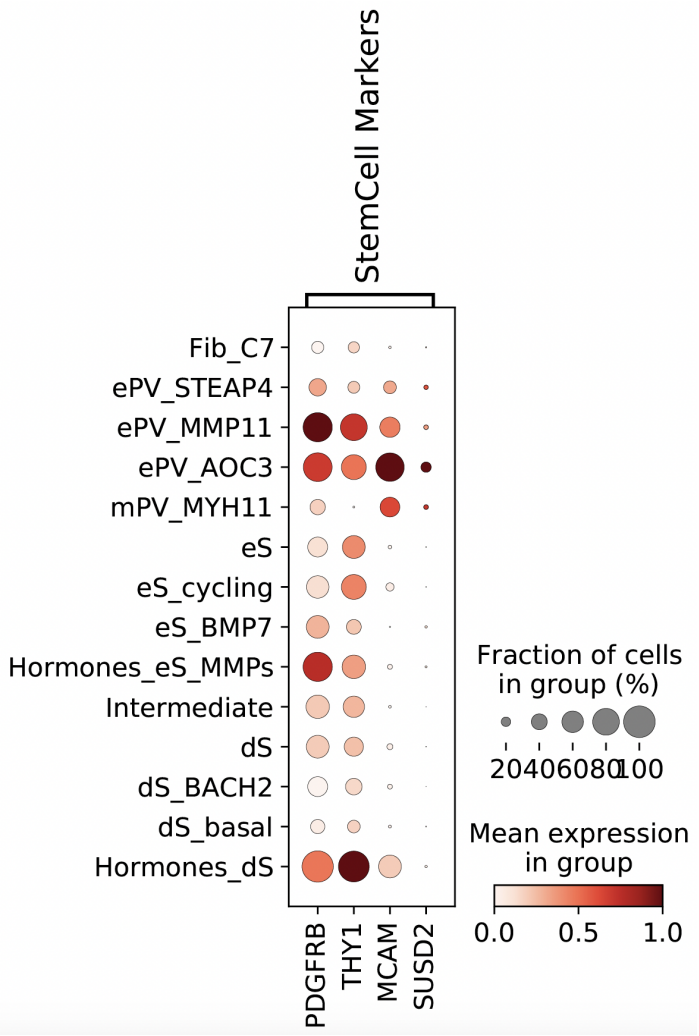
As shown in the above slides, my initial experiments looking at tissue cryopreservation and subsequent scRNAseq analyses showed that there were no differences in the cellular profiles and I thus used this cryopreservation methods throughout this thesis.

Appendix 2:
Gene expression profile associated with endometrial receptivity



The dotplot above shows the variance-scaled, log-transformed expression of 8 marker genes (x-axis) previously reported as differentially expressed in the receptive endometrium, for the cell lineages identified in Chapter 3 (y-axis). These marker genes were specifically expressed by the newly defined TENM2 population. *Abbreviations:* ERA, endometrial receptivity array; Prolif, proliferative.

Appendix 3:
Expression of previously described mesenchymal stem cell markers



The dotplot above shows the variance-scaled, log-transformed expression of 4 marker genes (x-axis) previously reported as characteristic of endometrial mesenchymal stem cell markers, for the mesenchymal cell states identified in Chapter 3 (y-axis). These marker genes were expressed by the newly defined ePV_AOC3 and ePV_MMP11 populations. *Abbreviations:* dS, decidualised stromal cells; eS, endometrial stromal cells; Fib, fibroblast.

Appendix 4:
Comparing media composition defined by Turco *et al*. & Boretto *et al.*

Investigating the response of endometrial organoids to different culture media compositions and hormonal stimulation regimes using single-cell transcriptomics

Magda Marečková[1,2†], Louis-François Handfield[2†], Iva Kelava[2], Carmen Sancho-Serra[2], Rebecca Dragovic[1], Jen Southcombe[1], Karin Hellner[1], Christian Becker[1*], Krina Zondervan[1*], Roser Vento-Tormo[2*]

1 Nuffield Department of Women's and Reproductive Health, University of Oxford, UK
2 Wellcome Sanger Institute, UK
†equal contribution, *corresponding author

Introduction

The human endometrium:
- is the inner mucosal lining of the womb
- comprises epithelial glands surrounded by stroma
- undergoes cycles of shedding and rapid repair on a monthly basis
- is re-epithelised and proliferates under the influence of oestrogen (E2)
- matures and becomes secretory under the influence of progesterone (P4)

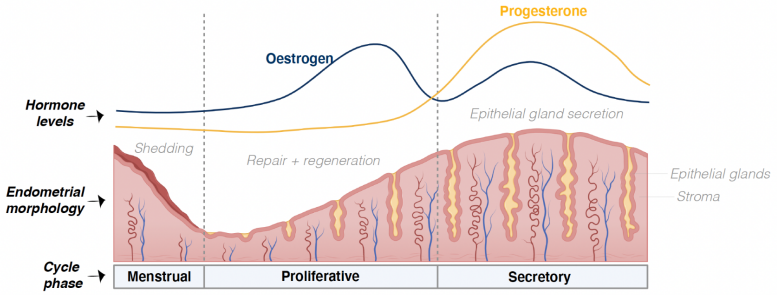


Figure 1: Schematic overview of the endometrium across the menstrual cycle

Epithelial endometrial organoids:
- have been generated using different endometrial organoid media (EOM)[1,2]
- are hormone-responsive
- can be stimulated with hormones to mimic the menstrual cycle
- different stimulation protocols have been reported[1,2]

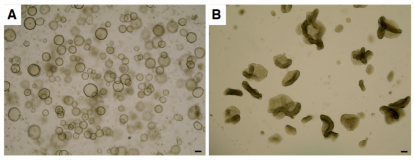


Figure 2: Endometrial organoids morphology when grown in EOM (A), and after hormonal stimulation (B).

Objectives

1) Investigate the effects of (*a*) EOM compositon and (*b*) hormonal stimulation on the cellular profile of endometrial organoids.

2) Benchmark endometrial organoids (unstimulated & stimulated) against our in vivo endometrial reference[3].

Materials & Methods

Organoids derivation

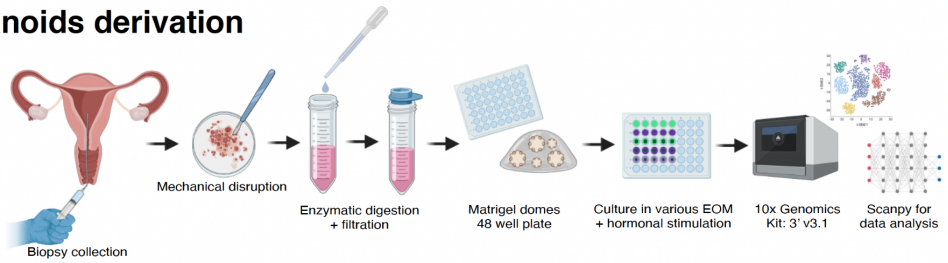


Figure 3: Endometrial organoid derivation, culture set-up, single-cell capture & analysis.

Endometrial organoids media

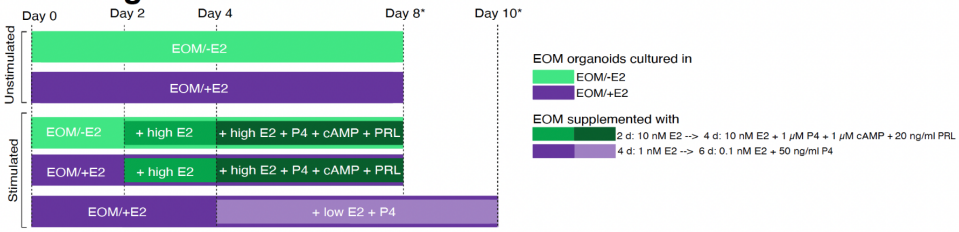


Figure 4: Experimental set-up for culturing organoids in different EOM and hormonal stimulation regimes.
* collection for single-cell RNA-sequencing

Abbreviations: cAMP, cyclic adenosine monophosphate; PRL, prolactin

	Component (final concentration)																	
Medium	DMEM/F12	B27	N2	Noggin	FGF-10	L-glutamine	Glutamax	NAC	Primocin	R-spondin-1	Nicotinamide	EGF	HGF	A-83-01	ITS	bFGF	p38i	E2	Y-27632
EOM/+E2	1X	50X	1X	100 ng/ml	10 ng/ml	--------	1 %	1.25 mM	100 μg/ml	500 ng/ml	2 mM	50 ng/ml	--------	500 nM	1 %	2 ng/ml	10 μM	1 nM	10 μM*
EOM/-E2	1X	50X	1X	100 ng/ml	100 ng/ml	2 mM	--------	1.25 mM	100 μg/ml	500 ng/ml	10 nM	50 ng/ml	50 ng/ml	500 nM	--------	--------	--------	--------	10 μM*

Table 1: Endometrial organoid media (EOM) composition
* used for organoid formation after passaging only

Abbreviations: FGF-10, fibroblast growth factor 10; NAC, N-acetyl cysteine; EGF, epithelial growth factor; HGF, human growth factor; ITS, insulin transferrin selenium; bFGF, basic fibroblast growth factor

Results

EOM composition alters the cells' transcriptomic profiles

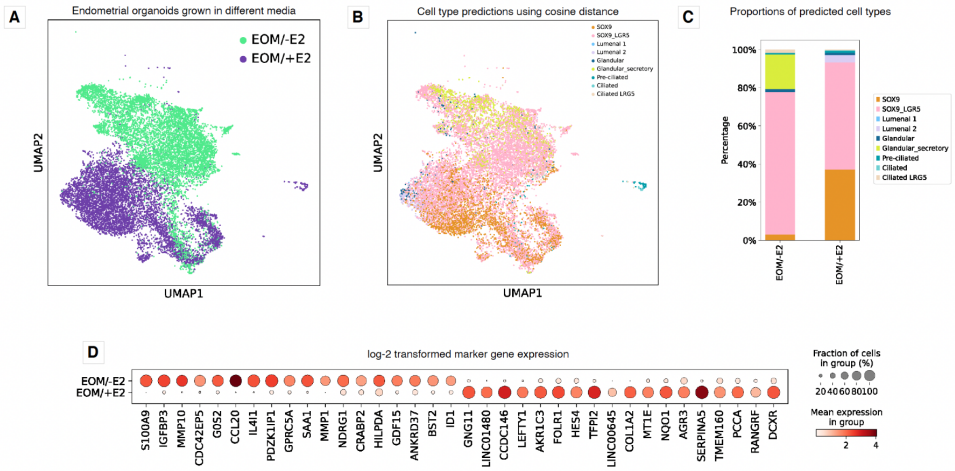


Figure 5: Single-cell transcriptomic data for unstimulated organoids. (A) UMAP visualisation of organoid cells cultured in different media. Predicted cellular identities of organoid cells based on our in vivo endometrial reference using cosine distances as a maesure of similarity (B). Cell-type proportions predicted for each culture medium used (C). Dot plot showing expression of genes characteristic for each media condition (D).

Hormonal stimulation protocols alter the cells' transcriptomes

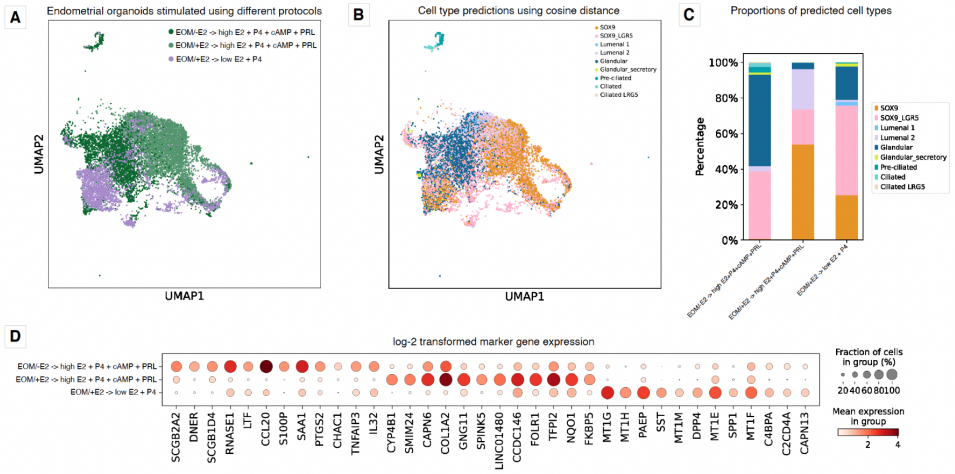


Figure 6: Single-cell transcriptomic data for stimulated organoids. (A) UMAP visualisation of organoid cells hormonally stimulated using diffrent protocols. Predicted cellular identities of organoid cells based on our in vivo endometrial reference using cosine distances as a measure of similarity (B). Cell-type proportions predicted for each stimulation protocol (C). Dot plot showing expression of genes characteristic for each stimulation condition (D).

Endometrial organoids closely resemble in vivo endometrium

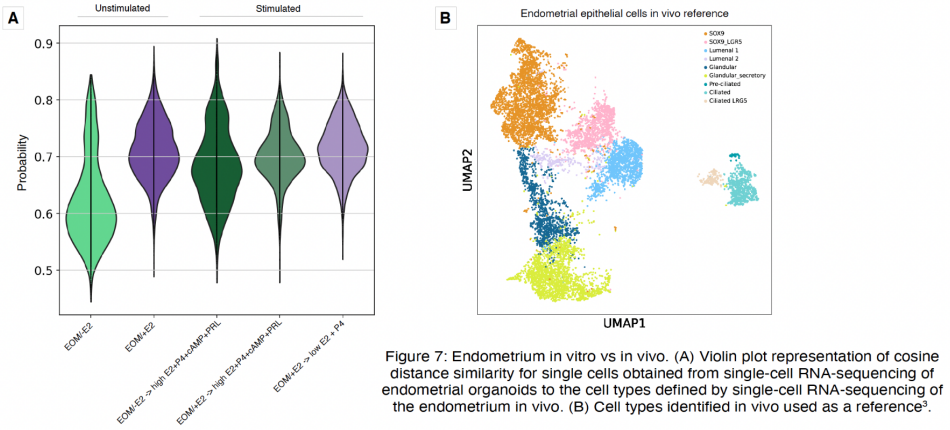


Figure 7: Endometrium in vitro vs in vivo. (A) Violin plot representation of cosine distance similarity for single cells obtained from single-cell RNA-sequencing of endometrial organoids to the cell types defined by single-cell RNA-sequencing of the endometrium in vivo. (B) Cell types identified in vivo used as a reference[3].

Conclusions

Organoids cultured using different EOM and stimulation protocols:

- differ in their transcriptomic profiles and proportions of cell types identified
- stimulation generates secretory phase cellular phenotypes
- culture in EOM/+E2 most closely resembles in vivo endometrial cells

These notable effects of media composition and hormonal stimulation regimes on organoid cell differentiation and gene expression profiles have implications for down-stream applications of the in vitro system to study the human endometrium in health and disease.

References:

1. Turco, M. Y. et al. Long-term, hormone-responsive organoid cultures of human endometrium in a chemically defined medium. Nat. Cell Biol. 19, 568–577 (2017).
2. Boretto, M. et al. Development of organoids from mouse and human endometrium showing endometrial epithelium physiology and long-term expandability. Development 144, 1775–1786 (2017).
3. Garcia-Alonso, L. et al. Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. Nat Genet 53, 1698–1711 (2021).

Initial trial experiments comparing the Turco *et al*. (EOM/-E2) and Boretto *et al.* (EOM/+E2) protocols and combination of their hormonal stimulation regimes.

Appendix 5:
Cell culture - seeding densities and media volumes used

	Surface area (cm^2)	Seeding density*	Cells at confluency^	Trypsin (ml)	Growth medium (ml)
Culture plates					
6-well	9.6	0.3×10^6	1.2×10^6	1	1 to 3
12-well	3.5	0.1×10^6	0.5×10^6	0.4 to 1	1 to 2
24-well	1.9	0.05×10^6	0.24×10^6	0.2 to 0.3	0.5 to 1.0
48-well	1.1	0.03×10^6	0.12×10^6	0.1 to 0.2	0.2 to 0.4
96-well	0.32	$0.01\ 10^6$	0.04×10^6	0.05 to 0.1	0.1 to 0.2
Flasks					
T-25	25	0.7×10^6	2.8×10^6	3	3–5
T-75	75	2.1×10^6	8.4×10^6	5	8–15
T-175	175	4.9×10^6	23.3×10^6	17	35–53
T-225	225	6.3×10^6	30×10^6	22	45–68

*Seeding density is given for each culture vessel type as follows: Flasks: Cells per vessel; Culture plates: Cells per well.

^The number of cells on a confluent plate or flask will vary with cell type. For this table, HeLa cells were used.

The table is based on adapted from the *Useful Numbers for Cell Culture* table provided by ThermoFisher on their website: https://www.thermofisher.com/uk/en/home/references/gibco-cell-culture-basics/cell-culture-protocols/cell-culture-useful-numbers.html

Appendix 6:
List of all endometrial epithelial organoid and stromal cell lines derived

Line ID	EEO	ES	Metadata	Hormones	Data in thesis[$]
FX9006	Yes	Yes	Control	No	No
FX1156	Yes	Yes	Endometriosis IV	No	No
FX1233	Yes	Yes	Endometriosis II	No	No
FX1160	Yes	No	Endometriosis III	No	No
FX1236	Yes	No	Endometriosis II	No[*]	No
FX1168	Yes	No	Endometriosis IV	Yes	No
FX1176	Yes	Yes	Endometriosis IV	No	No
FX1249	Yes	No	Endometriosis I	Yes	No
FX1251	Yes	No	Endometriosis III	No	No
FX1267	Yes	No	Endometriosis III	No	No
FX1258	Yes	No	Endometriosis II	Yes	No
A66[^]	Yes	Yes	Unknown	Unknown	No
FX1291	Yes	Yes	Control	Yes	No
FX1285	Yes	No	Endometriosis II	No	No
FX1119	Yes	Yes	Control	No	Yes
FX1254	Yes	Yes	Endometriosis II	Yes	Yes
FX1259	Yes	Yes	Endometriosis II	No	Yes
FX1268	Yes	Yes	Endometriosis II	No	Yes
FX1260	Yes	Yes	Control	Yes	Yes
FX1146	Yes	Yes	Control	No	Yes - both
FX1294	Yes	Yes	Control	Yes	Yes - both
FX1125	Yes	Yes	Endometriosis III	No	Yes - both
SE03	Yes	Yes	Control	No	Yes - both
SE02	Yes	No	Control	No	Yes
FX1289	Yes	Yes	Endometriosis I	No	Yes

Metadata refers to being controls (i.e. donors without endometriosis) or patients with endometriosis and respective rASRM endometriosis stages. Hormones refers to a patient taking hormonal therapy at sample collection.

Abbreviations: EEOs, endometrial epithelial organoid lines; ES, endometrial stromal lines.

Superscripts:
[$] *single-cell RNA-sequencing data presented in this thesis: Yes – both means data presented for both EEOs and ES lines, otherwise Yes means for EEOs only.*
[*] *donor reported not having periods for 10 months prior to surgery.*
[^] *sample is from the organ donor programme, limited metadata is available.*

Appendix 7:
List of public engagement/outreach projects and awards received

2022 - 2023: The SILBERSALZ Institute - Film School

Europe & online: https://www.documentary-campus.com/training/silbersalz-institute

Selected as one of 12 scientists to receive a year-long experience of working with media professionals on developing one's own research into a production for mainstream media.

2022: Wellcome Centre for Human Genetics Public Engagement Awards 2022

University of Oxford, UK

Winner of the Project category, awarded for Unheard of, the hidden voices of endometriosis podcast series.

2021 - 2022: Unheard of - The Hidden Voices of Endometriosis Podcast series

Online, world-wide: https://unheardof.libsyn.com

Co-host and co-producer of a podcast series about endometriosis

Interviewing doctors, researchers, patients

Scripting & editing episodes

Social media content scripting, editing & dissemination

2020: Wellcome Centre for Human Genetics Public Engagement Awards 2020

University of Oxford, UK

Winner of the Early Career Researcher category, awarded for consistent commitment to public engagement through a variety of methods.

2020: Young Scientist Award 2020

Fast Forward Science, Germany

International competition in science video production & communication

Link to my winning video on endometriosis: https://cutt.ly/yVgtWXx

2019: Digital Media Intern at the Royal Institution

Royal Institution, UK

12-week science communication placement focusing on the use of digital media

Managing social media platforms (Twitter, Facebook, Instagram, Youtube)

Content creation for social media, videos, internal communication

e.g. post on endometriosis: https://twitter.com/Ri_Science/status/1195281446564483072

Writing scripts, filming and editing short videos & audio recordings

Podcast production & editing

e.g. podcast about our research centre: https://soundcloud.com/royal-institution/endometriosis

2019: Max Perutz Science Writing Award 2019 (shortlisted)

Medical Research Council, UK

Annual writing competition receiving over 100 entries from MRC-funded PhD students

Writing about why my research matters for non-scientists: https://cutt.ly/dVgo5Tv

2018 - ongoing: Outreach and Public Engagement with Research

Throughout the UK and beyond

During my DPhil studies I have been involved in many public-facing events, including.:

Developing workshops for primary & secondary school students
(Cheney School Oxford, Summer Teaching Programme Oxford)

Bringing experiments & genetics to schools & libraries
(Oxford Hands on Science, Science at the Library)

Discussing future of genetics research with young people
(Future Forum with British Science Association)

Talking to patient groups about endometriosis research
(in the UK, Slovenia, Czech Republic)

Talking at festivals and events
(British Science Festival, Endometriosis Pain through Art, The Future of Endometriosis)

Seminars & talks about running a podcast
(Oxford University, Mind the Research Gap conference)

Appendix 8:
List of scientific meetings and conferences attended

Conferences/meetings attended and presented at:

2023: World Congress on Endometriosis - talk

2023: Society for Reproductive Investigation - talk

2022: Single-Cell Biology - poster

2022: European Society of Human Reproduction and Embryology meeting - talk

2022: Human Uterus (HUTER) Project Final Steering Group Meeting - talk

2022: Organoids as Tools for Fundamental Discovery and Translation - poster

2021: Edinburgh-Oxford Women's Health Networking Event- talk (runner-up best talk)

2021: World Congress on Endometriosis - poster (runner-up best poster)

2019: ENITEC meeting - talk

Conferences attended, but not presented at:

2020: Emerging Technologies in Single-Cell Research

2020: Single-Cell Biology

2019: NextGen Omics London

2019: Society for Reproductive Investigation meeting

2018: NextGen Omics London

Appendix 9:
List of academic awards received, roles held and publications

Academic awards:

2022: Transition into your next career (6-month postdoc salary)
Oxford MRC-DTP; University of Oxford
· funding to be used after thesis submission to develop new skills needed for next job

2022: Goodger and Schorstein Research Scholarship in Medical Sciences
University of Oxford
· scholarship to support my DPhil studies

2021: Associate Fellow of the Higher Education Academy
Advance HE
· awarded upon completion of the Advancing Teaching and Learning programme in the Sciences stream and assessment of one's teaching portfolio
· programme is run by the Centre for Teaching and Learning, University of Oxford

2019: Funding for Oxford Biomedical Data Science Training Programme
Oxford MRC-DTP; University of Oxford
· awarded £6,000 to undertake this highly competitive training programme

2019: Funding to complete an internship at the Royal Institution, UK
Oxford MRC-DTP; University of Oxford
· competitive 12-week science communication placement

2018 – 2021: Oxford & Medical Research Council Doctoral Training Partnership - Nuffield Department of Women's & Reproductive Health: Medical Sciences Graduate School Studentship
University of Oxford
· scholarship to fund and support my DPhil studies and research

Academic and non-academic roles held:

2020 - 2022: Trained peer supporter for students in Medical Sciences Division
· training included active learning & assertivness
· providing wellbeing support across division & organising welfare events

2020 - 2021: NDWRH student committee president
· representing students in the Department across Departmental and Divisional steering and working groups
· organising events, symposium day, writing newsletters, managing committee members

2019 - 2022: MRC-DTP student representative
· representing all MRC-funded students at the MRC-DTP steering group meetings
· running and organising the annual MRC-DTP student symposium

Publications:

Garcia-Alonso, L., Lorenzi, V., Mazzeo, C.I. et al. Single-cell roadmap of human gonadal development. Nature **607**, 540–547 (2022). https://doi.org/10.1038/s41586-022-04918-4

Marečková, M., Massalha, H., Lorenzi, V. Et et al. Mapping human reproduction with single-cell genomics. Annual Review of Genomics and Human Genetics 2022 23:1, 523-547. https://doi.org/10.1146/annurev-genom-120121-114415

Siemienowicz, K.J., Wang, Y., Marečková, M. et al. Early pregnancy maternal progesterone administration alters pituitary and testis function and steroid profile in male fetuses. Sci Rep **10**, 21920 (2020). https://doi.org/10.1038/s41598-020-78976-x