Altered Differentiation of Endometrial Mesenchymal Stromal Fibroblasts Is Associated With Endometriosis Susceptibility

Brett McKinnon (brett.mckinnon@dbmr.unibe.ch)
University Hospital of Berne https://orcid.org/0000-0002-9881-1252
Samuel Lukowski
Boehringer Ingelheim https://orcid.org/0000-0002-8598-7902
Sally Mortlock
The University of Queensland
Joanna Crawford
The University of Queensland
Rebecca Johnston
QIMR Berghofer Medical Research Institute https://orcid.org/0000-0001-7503-9514
Kostantinos Nirgianankis
University Hospital of Berne
Michael Mueller
University Hospital of Berne
Grant Montgomery
The University of Queensland https://orcid.org/0000-0002-4140-8139

Article

Keywords: Cellular development, endometrial mesenchymal stromal fibroblasts, endometriosis susceptibility, endometrium

Posted Date: June 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-596345/v1

Version of Record: A version of this preprint was published at Communications Biology on June 20th, 2022. See the published version at https://doi.org/10.1038/s42003-022-03541-3.

Abstract

Cellular development is tightly regulated as mature cells with aberrant functions may initiate pathogenic processes. The endometrium is a highly regenerative tissue, shedding and regenerating each month. Endometrial stromal fibroblasts are regenerated each cycle from mesenchymal stem cells and play a pivotal role in endometriosis, a disease characterised by endometrial cells that grow outside the uterus. Why the cells of some women are more capable of developing into endometriosis lesions is not clear. Using isolated, purified and cultured endometrial cells of mesenchymal origin from 19 women with (n = 10) and without (n = 9) endometriosis we analysed the transcriptome of 33,758 individual cells and compared these to clinical characteristics and *in vitro* growth profiles. We show purified mesenchymal cell cultures include a mix of mesenchymal stem cells and two endometrial stromal fibroblast subtypes with distinct transcriptomic signatures indicative of varied progression through the differentiation processes. The fibroblast subgroup characterised by incomplete differentiation was predominantly (81%) derived from women with endometriosis and exhibited an altered *in vitro* growth profile. These results uncover an inherent difference in endometrial cells of women with endometriosis and highlight the relevance of cellular differentiation and its potential to contribute to disease susceptibility.

Introduction

Every tissue is a complex biological system consisting of heterogenous cell mixtures that developed through cellular differentiation and maturation. Tight regulation of these processes are required to maintain homeostasis. They can be influenced by both cell autonomous and non-autonomous factors. Differentiation is contingent on stochastic interactions and subject to biological variability (1). Any alterations in the differentiation or maturation process may give rise to subtle biological variations, introduce heterogeneity that leads to functional consequences and influence disease susceptibility.

The endometrium is the reproductive tissue that lines the uterus and plays a critical role in reproduction. It is unique in that it is consistently shed and regrown each month, generating up to 10mm of new mucosa. Over the reproductive life of a women it undergoes > 400 cycles of growth, differentiation and shedding (2). The endometrium is made up of luminal and glandular epithelial cells supported by a vascularised stroma with immune infiltration. Endometrial stromal fibroblasts are regenerated from endometrial mesenchymal stem cells (eMSC) located at perivascular locations in basalis, not shed during menstruation (3). In the endometrium, local niche effects largely restrict eMSC differentiation into mesodermal stromal fibroblasts (4) that subsequently differentiate into secretory decidual cells under hormonal stimulation. As a highly regenerative tissue, the endometrium has the potential each month for aberrant differentiation to occur.

Endometriosis is a reproductive disorder characterised by the growth of endometrial tissue outside the uterus. Endometrial cells are thought to enter the peritoneal cavity through retrograde menstruation (5). Up to 80% of women experience retrograde menstruation, however only a proportion develop endometriosis (6). Inherent factors in the cells of some women must underlie an increased disease

susceptibility. A number of studies have reported differences between the endometrium of women with and without endometriosis, although differentially expressed genes have been difficult to consistently replicate (7), potentially due to the dynamic nature of the tissue. Recent large scale genome-wide gene expression studies on endometrial tissue also reported no significant difference in gene expression between women with and without endometriosis once menstrual stage and multiple testing correction was applied (8, 9). Although increasing in power, these studies are limited by the complex milieu of cells and cellular states that may mask subtle differences.

Cells of mesenchymal lineage are some of the most abundant in the endometrium and are strongly implicated in endometriosis pathogenesis. SUSD2 + eMSC have been identified in both peritoneal and menstrual fluid and may have a key role in the establishment and proliferation of ectopic endometrial tissue (10) through their clonogenic and multipotent differentiation capacity (5). eMSC from women with endometriosis had impaired *in vitro* decidualisation (11), as well as altered activation of signalling pathways during decidualisation (12, 13).

To identify inherent variations in the endometrium that could underlie endometrial susceptibility we therefore assessed gene expression of individual mesenchymal-derived cells from the endometrium and their association with clinical parameters and *in vitro* growth. We identified eMSCs and two distinct endometrial stromal fibroblasts populations generated by divergent differentiation from MSC to their mature cell state, one of which was characterised by gene expression profile indicative of an altered immune state and was found significantly more frequently in women with endometriosis. This study links single cell transcriptome data with both functional and clinical characteristics and uncovers a potential role for divergent mesenchymal-derived stromal fibroblast maturation to contribute to endometriosis susceptibility.

Results

Endometrial stromal cell viability and purity from patient samples

Endometrial stromal cells isolated from endometrial biopsies were grown in culture and stored frozen (Fig. 1A). We selected 22 frozen samples for analysis and to ensure pure, viable cells of mesenchymal lineage from thawed preparations conducted two channel FACS sorting with forward and side scatter (Fig. 1B), propidium iodine (PI) exclusion (Fig. 1C) and platelet derived growth factor β+ (PDGFB+) expression (Fig. 1D). The mean cell concentration after initial thaw was 2.19×10^6; (Range 0.172×10^6 – 3.46×10^6) with 90.0% remaining viable after the thawing process. One sample did not yield a sufficient cell concentration, and two samples had a final viability < 80% and were not carried forward to single cell analysis, resulting in a final 19 samples.

Single cell RNA-sequencing

In order to increase the size of the patient cohort, we used a multiplexing approach to combine samples from different patients into four microfluidic runs of the 10X Genomics Chromium Platform (P1: Patients 1–5, P2: 6–10, P3: 11–14 and P4: 15–19) (Fig. 1A**)**. We obtained sequencing data from the four scRNA-seq libraries constructed from our four pools of 19 endometrial stromal cells (ESC).

For each pool, the number of cells obtained were 10,411 (P1), 9,869 (P2), 10,894 (P3) and 9,808 (P4) making a total of 40,982 cells. Demuxlet identified 3,982 doublets that were randomly distributed across each pool (**Supplementary Fig. 1A-D)**. We also excluded 358 ambient cells, 2,884 with > 10% mitochondrial DNA that were considered stressed or dying cells, and cells with either very high (> 6,500), or very low (< 200) numbers of expressed genes. From the initial 40,982 cell dataset we retained 33,758 high quality singlets for analysis with an average read depth of 58,541 (P1), 58,277 (P2), 53,825 (P3) and 55,836 (P4) for each pool. We detected a median of 19,803 (P1), 20,095 (P2), 20,651 (P3) and 22,472 (P4) unique molecular identifiers (UMI) per cell with the total number of genes with measurable expression of 21,885 (P1), 21,736 (P2), 22,178 (P3) and 21,996 (P4). In total, 20,590 unique genes were identified across all four pools with the median number of genes expressed per cell as 3,780 (P1), 3,761 (P2), 3,941 (P3) and 4,050 (P4). After doublet-filtering and quality control, between-pool batch effects were corrected using Harmony (Fig. 1E-F).

We next investigated transcriptome similarity and assessed whether potential cellular subtypes were present through unsupervised Louvain clustering using Seurat v3.0.2 (14). A critical step in deriving relevant data from single cell datasets is selecting the appropriate resolution for clustering. Increasing resolution increases cell clusters, although potentially at the expense of biological relevance (**Supplementary Fig. 2**). Using the clustree (15) package, we produced a cluster tree with 13 levels of resolutions ranging from 0.01-1.0 (Fig. 1G) to visualise the similarity between cells at multiple resolutions, and track how cells move between clusters as resolution is varied. At a coarse resolution (0.1), three distinct nodes were identified that established stable clusters with minimal movement across nodes at increasingly finer resolutions. At this course resolution, there was also a distinct spatial separation for each cluster (Fig. 1H) and the number of cells from each experimental pool within each cluster was consistent (Fig. 1I). In contrast, clustering at a finer resolution (0.6) generated 20 clusters which lacked distinct spatial resolution and revealed mixing between clusters and possible over clustering resulting from technical artifacts (**Supplementary Fig. 2**).

Cell cycle scoring

Genes can be periodically regulated during the cell cycle (16), influencing their transcriptome and affecting the ability to accurately cluster cells based on phenotype. To characterise cell cycle for each cell, we calculated G1, G2M and S scores using the CellCycleScoring function in Seurat and human cell cycle phase gene expression profiles (17). The majority of cells analysed (69.65%) showed a G1 phenotype, whilst 15.77% of cells were classified as G2M and 14.58% were classified as S phase (Fig. 1J). We observed an enrichment of the proliferating cells (G2M) in cluster 1 (54.94%). Cluster 1 also had an increased proportion of cells in S phase (31.94%) compared to clusters 0 (7.88%) and 2 (12.64%). The

majority of cells in cluster 0 and 2 were classified as the quiescent G1 phase (90.80% and 84.11% respectively).

Differential gene expression between clusters reveals discrete signatures

We next examined the differentially expressed genes (DEGs) that underlie these cluster differences (Fig. 2A). Setting a log fold change (logFC) > 0.25 and adjusted p value < $1x10^{-4}$ we found 152 significant DEGs between cells in cluster 0 and all other cells (**Supplementary table 1**). A comparison between cluster 1 and all remaining cells found 707 DEGs (**Supplementary table 2**), and cluster 2 (**Supplementary table 3**) and all other cells found 113 DEGs. Spatial representation of three of the top DEGs in cluster 0 (IGFBP5; logFC = 0.984; adj. p value < $1.0 x10^{-305}$; MMP11; logFC = 0.887; adj. p value < $1.0 x10^{-305}$ and ACTA2; logFC = 0.728; adj. p value < $1.0 x10^{-305}$) (Fig. 2B) revealed strong variation and non-synonymous distribution within the cluster, accompanied by low but consistent expression in the two other clusters. A similar spatial resolution was observed for UBE2S (logFC = 1.67; adj. p value < $1.0 x10^{-305}$) with high expression in cluster 1 but low, consistent expression in the remaining clusters, although both PTTG1 (logFC = 1.833; adj. p value < $1.0 x10^{-305}$) and UBE2C in particular (logFC = 2.02,; adj. p value < $1.0 x10^{-305}$) showed a limited expression confined mostly to cluster 1. In cluster 2, MMP3 (logFC = 2.31; adj. p value < $1.0 x10^{-305}$), CST1 (logFC = 2.19; adj. p value < $1.0 x10^{-305}$ and MMP10; (logFC = 1.56; adj. p value < $1.0 x10^{-305}$) showed significant differential expression, and both CST1 and MMP10 expression were limited mostly to cluster 2. We also performed differential expression analysis between specific cluster pairs (cluster 0 vs 1, 1 vs 2, and 0 vs 2) and detected 242, 246 and 17 significant DEGs, respectively (logFC > 0.5, adj. p value < $1x10^{-4}$; **Supplementary Tables 4, 5 and 6**).

To gain further insight into the biological differences underlying the three clusters, we performed pathway analysis using the top 200 significant DEGs for each cluster using Reactome, KEGG and gene ontology databases (Fig. 2C). This revealed significantly enriched processes involved in extracellular matrix organisation (adj. p value < $1.33x10^{-12}$) and focal adhesion (adj. p value = $1.1x10^{-3}$) for cluster 0. For cluster 1 we observed significant enrichment of cell cycle (adj. p value = $6.97x10^{-12}$), progesterone-mediated oocyte maturation (adj. p value = $4.28x10^{-8}$) and oocyte meiosis (adj. p value = $2.53x10^{-7}$), whereas cluster 2 DEGs were enriched for extracellular matrix organisation (adj. p value = $1.89x10^{-14}$) but also antigen processing and presentation (adj. p value = 0.0049), and allograft rejection (adj. p value = 0.0389), indicating a potentially immune-reactive cell population (Fig. 2C).

Cell- type annotation identification

Mesenchymal maturation can take multiple pathways leading to divergent progeny such as fibroblasts, adipocytes and smooth muscle cells (18). To annotate our transcriptomically defined cell clusters we applied SingleR (19). This utilises the transcriptomic signatures from the Human Primary Cell Atlas, a database curated from publicly available microarray datasets of human primary cells (20). The analysis confirmed a close alignment with cells of mesenchymal lineage, albeit with variations in mesenchymal

progeny distributed across the clusters (Fig. 3A). The five most prevalent cell types identified were fibroblasts (84.78%), mesenchymal stem cells (MSCs, 13.66%), smooth muscle cells (1.21%), induced pluripotent stem (IPS) cells (0.24%) and tissue stem cells (0.11%) (Fig. 3B).

Overlay of the different cell types with the clustering analysis revealed fibroblasts were the predominant cell type of cluster 0, representing (98.34%) of the cells in this cluster, with 0.93% identified as MSCs. As this was the largest cluster of fibroblast cells we designated this cluster '*fibroblast major*'. Similarly, cluster 2 while distinct from cluster 0 was predominately fibroblasts (97.34%) with the inclusion of some MSCs (2.50%) and was subsequently termed '*fibroblast minor*'. The clustering differences observed between the fibroblast clusters (clusters 0 and 2) could not be attributed to cell cycle differences (Fig. 2E). Cluster 1 identified predominantly as MSCs (57.42%), and as such was named the '*MSC cluster*', although 42.53% of cells within this cluster also identified as fibroblasts (Fig. 3C).

Cell fate trajectory between MSCs and fibroblasts

To investigate dynamic biological processes within our dataset we applied pseudotime and RNA velocity analysis using Monocle 2 and scVelo (21, 22) (23, 24). This allowed the opportunity to study cellular differentiation or lineage progression by ordering individual cells along a trajectory, which can then be used to infer the state of individual cells in processes such as cell maturation (Fig. 3D). Overlay of the clustering data on the pseudotime trajectory predictions suggested a directional progression from cluster 1 (*MSC cluster*) as the root cell directing a cell fate lineage towards cluster 0 (*fibroblast major*) (Fig. 3E). This directional progression was also observed in the RNA velocity analysis (Fig. 3F), supporting the hypothesis that the cell differentiation pathway extends from cluster 1 (*MSC cluster*) to cluster 0 (*fibroblast major*). Cluster 2 (*fibroblast minor*) in contrast, was spread uniformly across the differentiation trajectory. Finally, we overlaid cell cycle information onto the trajectory plot and observed that most G2M phase cells, as well as the S phase cells corresponding to MSCs and the less differentiated fibroblasts consistent with the initial cell cycle analysis of each cluster and indicative of higher proliferative ability of the MSCs. (Fig. 3G).

Deconvolution of individual patient cells

To ascertain whether cell types, or cell clusters were associated with clinical phenotype we assigned each cell from the 33,758 cell dataset to the source patient. To demultiplex the individual patient samples in the 4 microfluidic pools we collected SNP genotyping information and used demuxlet (25) to assign each cell to the patient from which it was derived. Demuxlet uses statistical modelling to identify RNA-seq reads that overlap single nucleotide polymorphisms (SNPs). Using SNP data and imputation generated from genotyping the most likely donor for each cell can be identified. Demuxlet identified an average of 1,229 SNPs (range = 13 – 3,970) per cell across pools; P1 (mean = 1,264 SNPs/cell), P2 (mean = 1,270 SNPs/cell), P3 (mean = 1,213 SNPs/cell) and P4 (mean = 1,176 SNPs/cell) (**Supplementary Fig. 3A-D**), allowing the confident assignment of 100% of the cells. Using this method the number of cells analysed for each patient in each pool was identified (Fig. 4A).

Clinical relationship with identified cell types

Using the clinical data we compared the number of MSCs, fibroblasts and smooth muscle cells derived from women with and without endometriosis. In total, 16,650 cells (49.3%) were sourced from 9 women without endometriosis and 17,108 (50.7%) cells from 10 women with endometriosis. The results indicate no significant variation in the percentage of fibroblasts from the endometrium of cases (84.93%) compared to controls (81.74%) **(**Fig. 4B**)**. There was also no difference between the percentage of MSCs derived from the cases (14.48%) versus the controls (17.83%) (Fig. 4C). The percentage of smooth muscle cells derived from cases (0.24%) was larger compared to controls (0.08%) (Fig. 4D) although the difference did not reach significance (p = 0.0738).

Our data indicates that fibroblasts could be split into two distinct groups. We therefore also compared the number of *fibroblast major (cluster 0)*, *fibroblast minor* (cluster 2) and *MSC cluster* (cluster 1) cells that were from women with and without endometriosis (Fig. 4E). The *MSC cluster* (cluster 1) had a total of 9,033 cells of which 4,199 (46.5%) were from women with endometriosis and 4,834 (53.5%) from women without endometriosis. The *fibroblast major cluster* (cluster 0) contained 22,881 cells with 11,418 (49.9%) from women with endometriosis and 11,463 (50.1%) from women without endometriosis. The *fibroblast minor cluster* (cluster 2) which contained 1,844 cells consisted of 1,491 (80.9%) cells from women with endometriosis and only 353 (19.1%) cells from women without endometriosis (Fig. 4F). A Chi-squared test indicated significantly more *fibroblast minor* cells (cluster 2) were from women with endometriosis (p < 0.0001).

Relationship between cell type, cell clusters and *in vitro* growth

Finally, we performed an analysis of cell growth rates for a continuous 100 hour period using a subset of 11 cell preparations and the xCELLigence assay. We compared growth rates of individual cell preparations, endometriosis status, and the percentage of cell type (MSCs and fibroblasts) or cell subset (*MSC cluster, fibroblast major, fibroblast minor*). For individual preparations, the growth rates varied for both endometriosis cases and controls (Fig. 5A). Grouping cell preparation by endometriosis status showed variable rates of proliferation between cases and controls at different time points. Cells from controls had an initial (0–15 hours) increased rate of proliferation with the growth rate eventually plateauing after 35 hours. In contrast the growth rate of cells from endometriosis cases continued to increase until the end of the incubation period (100 hours) resulting in an increased number of cells from endometriosis cases, although the difference was not significant (p = 0.4725)(Fig. 5B).

We also investigated the association between the contents of each cell preparation and growth rates by plotting the correlation between the percentage of each cell type, as determined by scRNA-seq and SingIR ananlysis and cell growth rates against time. The analysis revealed a positive association between MSC content and cell index that reached the strongest correlation between 8.25–9.75 hours (Pearson's r = 0.6364, p = 0.0402) (Fig. 5C). Similarly there was an opposite negative correlation with the percentage of fibroblasts and cell index between 8.5–9.5 hours (Pearson's r = -0.618, p = 0.0478) (Fig. 5D).

As the scRNA-seq analysis identified two subsets of fibroblasts *(fibroblast major* and *fibroblast minor*) we further assessed the association between the percent content of these cells in each cell preparation and growth rates. We found that each subtype displayed contrasting growth profiles, with the *fibroblast major cluster* (cluster 0) showing a non-significant positive association with growth rate and the *fibroblast minor cluster* (cluster 2) showing a significant negative correlation between 17.25–26 hours (Pearson's r = -0.681, p = 0.025) (Fig. 5E). confirming the relative presence of each fibroblasts influenced *in vitro* growth profiles at a later time points after seeding compared to MSCs.

Discussion

Cellular heterogeneity both within tissue and within cell types is a key driver of tissue variation and disease susceptibility. To better understand endometrium and endometrial pathologies such as endometriosis, we assessed cell heterogeneity within the endometrial mesenchymal cell lineage and association with clinical variables and *in vitro* cellular function. By profiling their gene expression at the single-cell level we identified three mesenchymal cell populations; a MSCs and two distinct stromal fibroblasts groups and charted their dynamic changes in gene expression. These data revealed the abundance of MSCs isolated was not related to endometriosis, but was associated with increased short-term *in vitro* growth. In addition, one fibroblast subpopulation displayed a gene expression profile indicative of dysregulated differentiation, altered immune reactivity, and their percentage within each cell preparation inhibited in vitro growth rates. Importantly, this subpopulation was more likely to be derived from the endometrium of women with endometriosis compared to women without endometriosis. These results support a divergence in mesenchymal differentiation that alters fibroblast function and may predispose some women to endometriosis susceptibility.

Single-cell transcriptome analysis has previously revealed insights into endometrial cells, although it has not yet provided insight into clinical observations or endometrial pathologies. Previous analysis of whole endometrium confirmed 6 distinct cell types; endothelial, epithelial (ciliated and unciliated), stromal and immune cells (26), although variations within cell types were not explored, potentially because it was not investigated and underpowered to do so with only 2,149 cells and one biological replicate at each day of the menstrual cycle. In our study we focused on cultured endometrial cells selected via the mesenchymal marker PDGFRβ. Previous single-cell investigations of transcriptomic profiles of endometrial stromal cells showed 64.9% of gene displayed consistent expression between both fresh and cultured cells (27). While subtle variation mediated by niche environment will be lost during *ex vivo* processing, the use of primary cultured cells provides the opportunity to examine cell lineage differentiation in the absence of exogenous cues. It also provides the opportunity to perform experiments at scale, integrate clinical data and importantly assess the transcriptomic relationship to *in vitro* growth characteristics.

Using these cultured samples we identified a significant proportion of MSCs remaining in all 19 culture preparations. MSCs reside in both the basalis and functionalis of the endometrium and are shed during menstruation potentially initiating endometriotic lesion growth (3). In this dataset we did not identify any differences in the MSC populations from women with and without endometriosis. Pseudotime trajectory

and RNA velocity analysis indicated a variable differentiation from the *MSC cluster* to the *fibroblast minor* cluster that may have derived from inherent variability within a subset of MSC cells, or lack of appropriate niche signals in the culture environment. It has previously been shown transcriptomic variations in MSCs are inherited by daughter cells creating variation in the gene expression profile and biological function, potentially leading to increased disease susceptibility (11).

The fibroblast minor cluster was characterised by a transcriptome with potential for extracellular matrix organisation. Some of the most differentially regulated genes included matrix metalloproteinases (MMPs), *MMP3* and *MMP10*, both of which are within the stromelysin subclass of MMPs that have significant roles in extracellular remodelling of laminin fibronectin and gelatin (I-V) and collagens (28). A genetic polymorphism in the promoter of *MMP3* is reported to be associated endometriosis (29) and in the normal menstrual cycle there is no *MMP3* expression in the proliferative phase, with an upregulation during the secretory stage that is significantly higher in women with endometriosis compared to women without (30). Previous evidence reports the focal expression of *MMP3* in developing endometrium (31, 32), data that would be consistent with our cluster analysis. Studies in skin have suggested *MMP10* expression is predominantly limited to epithelial cells (33), but has also been reported in endometriosis (34) and shown to control immune response in macrophages (35). MMP10 expression is increased in the bladder (36), oesophagus (37) and skin cancer (38) and has been shown to be instrumental in bladder tumour cell migration and invasion (39), and wound healing and matrix remodelling in skin cancer (38).

The fibroblast minor cluster was also characterised by strong expression of *CST1*. *CST1* has limited expression in most tissues of the body (40, 41), although deep proteome and transcriptome sequencing confirmed endometrial expression (42). Upregulation of *CST1* has been observed in malignant tumours and is associated with cancer cell proliferation, invasion and tumour recurrence (43–45). The combination of extracellular matrix with upregulation of these genes may provide the fibroblast minor cluster, through its enhanced adhesion and infiltration capabilities, the capacity to establish lesions and thus increase disease susceptibility.

CST1 has also been proposed as a fibroblast senescence marker (46). The presence of senescent decidual endometrial stromal cells has been observed both *in vitro* and *in vivo* (47). Decidual endometrial stromal fibroblasts differentiate from stromal fibroblasts approximately at mid cycle and in response to rising progesterone concentrations. Single-cell sequencing of cultured endometrial stromal cells undergoing induced decidualisation identified the emergence of this subset of senescent decidualised stromal cells and found they were linked to aberrant endometrial biology, increasing susceptibility to recurrent pregnancy loss (48). Cellular senescence is a state of permeant cell-cycle arrest and is accompanied by the secretion of extracellular matrix proteins, proinflammatory cytokines and growth factors (49).

The fibroblast minor cluster we observe with altered immune reactivity may have a corollary to these subset of decidualised endometrial cells produced from divergent differentiation pathways. Variation in the transcriptome of decidualised cells was observed that were acquired during maturation and were

dependent on gene expression profiles of the starting cell. Isolated SUSD2^{+} and SUSD- endometrial stromal cells that underwent differentiation to decidualising stromal cells retained distinct transcriptomic profiles that were characterised by differences in the secretion of inflammatory mediators with the decidualised SUSD2^{+} cells producing significantly more leukaemia inhibitory factor (LIF) and chemokine ligand 7 (CCL7) compared to the decidualised SUSD- stromal fibroblast (4).

Finally, information from this study may also contribute to the understanding of endometriosis progression and not just pathogenesis. There is an increasing understanding of the importance of fibrosis in endometriotic lesions, influencing both disease progression and treatment. Gli1 + marks perivascular MSC-like cells that contribute to organ fibrosis (50) and in endometriosis the immune environment of the peritoneal cavity can stimulate fibrosis through smooth muscle metaplasia (SMM) of endometrial stromal cells (51, 52). The identification of the smooth muscle cells in this dataset supports this hypothesis and may represent a further progression of the differentiation pathway that can be induced when exposed to variations in the extracellular environment.

In summary, previous work on endometriosis has suggested significant differences in the endometrium of women with and without endometriosis, although the mechanisms behind these variations and their contribution to endometrial pathologies is yet to be fully elucidated. Cell heterogeneity derived from variations in cell states or altered maturation pathways is common and may be embedded during cell fate lineage determination and can be leveraged by disease processes. By analysing endometrial stromal cells at a single-cell level with sufficient cell numbers, depth of sequencing and appropriate resolution we have uncovered a divergent mesenchymal differentiation of stromal fibroblasts that is significantly more likely to occur in cells from women with endometriosis. This could increase the susceptibility of the cells of some women to initiate endometriosis lesions at any particular cycle and may represent a potential biomarker. Divergent differentiation of stromal fibroblasts may provide and provide novel targets for future treatment paradigms and warrants further investigation.

Methods

Sample collection

Prior to surgery the relevant institutional review board granted ethical approval for the collection of samples and informed consent was obtained from all patients. Exclusion criteria for the study included abnormal ovulatory menstrual cycles, the use of either hormonal medication in the past 3 months. Patients with prior or current infections and liver dysfunction were also excluded. During laparoscopic surgery endometriotic biopsies were collected via soft curette (Pipelle de Cornier, Laboratorie CCD, France) and stored in Complete IMDM media (10% fetal calf serum (FCS), 1% antibiotics/antimycotics (Invitrogen Life Technologies)) supplemented with 10% dimethyl sulfoxide (DMSO) (Thermo Fischer Scientific, Waltham, MA, USA) using the slow freezing method in a Bicell vessel to -80°C. The pelvic cavity of each patient was subsequently examined, any endometriotic lesions removed and the patient staged according to the revised American Fertility Society staging system (rAFS) (53).

Endometrial stromal cell preparation

The endometrial stromal cells were prepared as described previously (54). Briefly, the tissue was thawed at 37°C, washed with serum-free medium to remove DMSO and dissected into smaller pieces. The tissue was washed in phosphate buffered saline (PBS) and incubated for 90 min at 37°C in the presence of collagenase (10mg/ml, Sigma) and subsequently filtered through 100µm mesh (Falcon) to remove debris and undigested material. This was followed by a second filtration through 40µm mesh which will retain intact epithelial glands and allow individual stromal cells to pass through. Two volumes of IMDM was immediately added to the filtrate containing single stromal cells. The cells were centrifuged 5 x 500g, the supernatant discarded and the pellet resuspended in 1ml fresh complete IMDM.

Cells were maintained in culture using complete media (IMDM, 10% FCS, 1% antibiotic/antimycotic). Growth curves and cell viability were monitored via recording of population doubling and cells maintained in a proliferative state by passaging using a standardised 1:3 split with trypsin/EDTA when cells were approximately 80% confluent. Once sufficient cell stocks were grown for subsequent experiments cells were trypsinsed and counted using the automated Countess Cell Counters (Thermo Fisher Scientific). Cells were stored frozen at $4x10^6$ cells per vial in complete media with 10% DMSO (v/v).

FACS analysis, cell selection and sample pooling

All samples were removed from liquid nitrogen, thawed and washed twice in IMDM complete media and cells counted as described. A final concentration of $2x10^6$ cells were diluted into 50µl of PBS. Cells were incubated in blocking buffer (PBS, 40% FCS, 1% BSA) for 30 min and subsequently incubated with the mouse monoclonal anti-human PDGRFβ + antibody conjugated to BV786-A (Becton Dickinson Cat No; 743038) with a 1:37.5 dilution in PBS, 10% FCS and 1% BSA for 1 hour. Prior to cell sorting, 2µl of propidium iodide (PI) was added to each sample. FACS was performed using the Aria II FACS machine (Becton Dickinson) with a dual-colour setting to select PDGFRβ + positive cells via dedicated excitation and emission settings for Brilliant Violet 421 and cells that excluded PI. Sorted cells were collected in PBS containing 10% FCS, and were then counted and their viability determined by haemocytometer and Trypan Blue staining. Cells with viability < 80% were excluded from further analysis.

Pools of cells from multiple patient samples were generated prior to loading the 10x Genomics Chromium microfluidic chip channels. To obtain a final concentration of 20,000 cells per pool with an equimolar concentration of cells from each sample, we aimed to pool samples as follows: for pools of 5 samples, we added approximately 8,000 cells from each sample, and for pools of 4 samples we added approximately 10,000 cells per sample. This yielded a final count of approximately 20,000 cells per pool due to the expected loss of cells during microfluidic processing.

Genotyping and imputation

DNA samples were isolated from the cell cultures and were genotyped using the Infinium Global Screening Array (Illumina Inc, San Diego). Quality control of genotypes was performed using PLINK(55)

and SNPs with a missing rate of > 5%, minor allele frequency (MAF) < 1×10^{-4} and with a Hardy-Weinberg Equilibrium (HWE) $p < 1 \times 10^{-6}$ were removed, leaving 645,726 SNPs for imputation. Imputation was performed using the 1000 Genomes Phase 3 reference panel. Genotyping data were used to identify the ancestry of each patient using 1000 Genome genotype data and principal component analysis.

Single cell RNA-sequencing and analysis

The FACS single-cell suspensions were used to generate barcoded single-cell 3' cDNA libraries for each of the pools with the Chromium Single cell 3' Gel Bead and library kit v2 (10x Genomics). Library quality control was performed with the Agilent Bioanalyzer High sensitivity DNA chip (Agilent). Denatured libraries were loaded onto an Illumina NovaSeq6000 and sequenced with a 2 x 100 base pair output for an average depth of 54,321 reads/cell.

The cellranger pipeline (v3.0.2) was used to process the sequencing data that included the mkfastq, count and aggr functions. The raw Illumina base call files were demultiplexed into sample-specific FASTQ files using cellranger mkfastq. Quality control (QC) was performed on the sample-specific files and subsequently aligned to the hg38 human reference using STAR (56) within the cellranger count algorithm. Aligned reads were filtered for valid cell barcodes and unique molecular identifiers, and resulting count matrices were combined into a single dataset using the cellranger aggr function. SNP genotyping data was used to identify doublets, multiplets and ambient cells using the Demuxlet software (25). Remaining cells were taken forward for further analysis using the Seurat package (v3.0.2) in R (v3.4.1). We applied the following QC and filtering steps to the raw data: exclude (i) cells with > 10% mitochondrial gene expression, (ii) cells with very low (< 200) or very high (> 6,500) numbers of expressed genes and (iii) genes expressed in very small numbers of cells (≤ 3). Between-cell gene expression was normalised using scTransform (57). Between-pool variation due to technical and biological differences was corrected using filtered and normalised data with Harmony (58).

Bioinformatic analysis

Seurat was subsequently used to perform Louvain clustering of cells with the first 50 principal components and using a parameter sweep across multiple resolutions between 0.01 and 1.0. Cluster stability was assessed using clustree (15) and, based on clustering stability, clustering information from resolution 0.1 was retained for analysis. Differentially expressed genes (DEGs) between each cluster were determined with the Wilcoxon rank sum test in Seurat with minimum percent expressing cells > = 0.25 and minimum absolute log2 fold-change threshold > = 0.25. Gene expression differences were considered significant if the adjusted p-value was < 1×10^{-4} (Benjamini-Hochberg correction for multiple testing) and the absolute log fold expression changes ≥ 0.5. Pathway enrichment analysis was performed with the top 200 DEGs in each cluster using the EnrichR package (59). The enrichment ranking for pathways, ontologies, transcription factor networks and protein network analysis was calculated from the multiplication of a log p-value from Fisher exact test by the Z-score of the deviation of the expected rank.

To identify the potential cell types within the dataset, a transcriptome-based cell-type classification was performed with SingleR (19) interrogating the Human Primary Cell atlas (HPCA) and the Blueprint + Encode reference datasets. Cell fate trajectory was predicted using the Monocle 2 package's pseudotime analysis (21) using the 500 genes with the highest variation in expression across all cells. For the Monocle 2 analysis, variation in gene expression was determined using a log2(counts + 1)-transformed dataset and genes ranked from the most to least variable. Cellular trajectory was further analysed via RNA velocity using dynamic modelling with velocyto (23)and scVelo (60)

Demultiplexing patient samples from single cell pools

Sample demultiplexing was performed using the Demuxlet software (25) to assign cells to genotyped individuals and identify doublets. The position-sorted BAM file produced by the cellranger count function and a VCF file containing the genotype information for each sample were used as input into Demuxlet, where each cell barcode was assigned to a specific sample (or a pair of samples) in the VCF file using the genetic variation sequenced in each cell.

Real-time analysis of cell adhesion and proliferation

Selected ESCs isolated from endometrial biopsies that remained in log phase of growth as determined by population doubling calculations were grown to approximately 80% confluence. Cells were trypsinized and counting using the methods described above. Sixteen-well E-plates (ACEA Bioscience) were pre-incubated with 50µl of prewarmed media and allowed to equilibrate in the incubator at 37°C with 5% CO_2 for 60 min to equilibrate. Each well per plate was inoculated with 10,000 cells in a total volume of 100µl, as this was previously determined as the optimal seeding density for cellular proliferation. The xCELLigence RTCA was set to perform a complete sweep across the plate to record cell growth, as Cell Index, every 15 min. Growth was profiled over a 48 hour period. Cell index data were normalised at the first time point post seeding to account for variations in cell concentrations, and exported for statistical analysis in Graphpad Prism v8.

Data and statistical analysis

All data have been presented unless otherwise stated as mean ± standard error of mean (SEM). Samples sizes for each experiment are indicated in the relevant results section or figure legends and represent biological replicates. Statistical analysis was performed in R (v3.4.1) and GraphPad Prism v8 software.

Declarations

Ethics approval and consent to participate

Tissue sample collection was approved by the Cantonal ethics commission Bern (149/03). Experimental procedures were approved by the Cantonal ethics commission Bern (2019-01146) and the University of Queensland Human Research ethics committee (2016001723).

Acknowledgements

We thank the staff of the Genome Innovation Hub, University of Queensland for constructive discussions on the deconvolution of patient samples and use of the Demuxlet software. This study was supported by a grant from the National Health and Medical Research Council (NHMRC) project grant GNT1147846. GWM was supported by NHMRC Fellowships GNT1078399 and GNT1177194.

References

1. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 2008;135(2):216-26.
2. Jabbour HN, Kelly RW, Fraser HM, Critchley HOD. Endocrine Regulation of Menstruation. Endocrine Reviews. 2006;27(1):17-46.
3. Gargett CE, Masuda H. Adult stem cells in the endometrium. Molecular Human Reproduction. 2010;16(11):818-34.
4. Murakami K, Lee YH, Lucas ES, Chan Y-W, Durairaj RP, Takeda S, et al. Decidualization Induces a Secretome Switch in Perivascular Niche Cells of the Human Endometrium. Endocrinology. 2014;155(11):4542-53.
5. Gargett CE, Schwab KE, Deane JA. Endometrial stem/progenitor cells: the first 10 years. Hum Reprod Update. 2016;22(2):137-63.
6. Halme J, Hammond MG, Hulka JF, Raj SG, Talbert LM. Retrograde menstruation in healthy women and in patients with endometriosis. Obstet Gynecol. 1984;64(2):151-4.
7. McKinnon B, Mueller M, Montgomery G. Progesterone Resistance in Endometriosis: an Acquired Property? Trends Endocrinol Metab. 2018;29(8):535-48.
8. Fung JN, Mortlock S, Girling JE, Holdsworth-Carson SJ, Teh WT, Zhu Z, et al. Genetic regulation of disease risk and endometrial gene expression highlights potential target genes for endometriosis and polycystic ovarian syndrome. Scientific Reports. 2018;8(1):11424.
9. Mortlock S, Kendarsari RI, Fung JN, Gibson G, Yang F, Restuadi R, et al. Tissue specific regulation of transcription in endometrium and association with disease. Human Reproduction. 2020;35(2):377-93.
10. Masuda H, Schwab KE, Filby CE, Tan CSC, Tsaltas J, Weston GC, et al. Endometrial stem/progenitor cells in menstrual blood and peritoneal fluid of women with and without endometriosis. Reproductive BioMedicine Online. 2021.
11. Barragan F, Irwin JC, Balayan S, Erikson DW, Chen JC, Houshdaran S, et al. Human Endometrial Fibroblasts Derived from Mesenchymal Progenitors Inherit Progesterone Resistance and Acquire an Inflammatory Phenotype in the Endometrial Niche in Endometriosis. Biol Reprod. 2016;94(5):118.
12. Yin X, Pavone ME, Lu Z, Wei J, Kim JJ. Increased activation of the PI3K/AKT pathway compromises decidualization of stromal cells from endometriosis. J Clin Endocrinol Metab. 2012;97(1):E35-43.

13. Aghajanova L, Hamilton A, Kwintkiewicz J, Vo KC, Giudice LC. Steroidogenic enzyme and key decidualization marker dysregulation in endometrial stromal cells from women with versus without endometriosis. Biol Reprod. 2009;80(1):105-14.
14. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature Biotechnology. 2018;36(5):411-20.
15. Zappia L, Oshlack A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. GigaScience. 2018;7(7).
16. Liu Y, Chen S, Wang S, Soares F, Fischer M, Meng F, et al. Transcriptional landscape of the human cell cycle. Proceedings of the National Academy of Sciences. 2017;114(13):3473.
17. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016;352(6282):189-96.
18. Dominici M, Le Blanc K, Mueller I, Slaper-Cortenbach I, Marini F, Krause D, et al. Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. Cytotherapy. 2006;8(4):315-7.
19. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nature Immunology. 2019;20(2):163-72.
20. Mabbott NA, Baillie JK, Brown H, Freeman TC, Hume DA. An expression atlas of human primary cells: inference of gene function from coexpression networks. BMC Genomics. 2013;14(1):632.
21. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nature Methods. 2017;14(10):979-82.
22. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature Biotechnology. 2014;32(4):381-6.
23. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;560(7719):494-8.
24. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. bioRxiv. 2019:820936.
25. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat Biotechnol. 2018;36(1):89-94.
26. Wang W, Vilella F, Alama P, Moreno I, Mignardi M, Isakova A, et al. Single-cell transcriptomic atlas of the human endometrium during the menstrual cycle. Nature Medicine. 2020;26(10):1644-53.
27. Krjutškov K, Katayama S, Saare M, Vera-Rodriguez M, Lubenets D, Samuel K, et al. Single-cell transcriptome analysis of endometrial tissue. Human Reproduction. 2016;31(4):844-53.
28. Osteen KG, Yeaman GR, Bruner-Tran KL. Matrix metalloproteinases and endometriosis. Semin Reprod Med. 2003;21(2):155-64.

29. Cardoso JV, Machado DE, da Silva MC, Berardo PT, Ferrari R, Abrao MS, et al. Matrix metalloproteinases 3 polymorphism increases the risk of developing advanced endometriosis and infertility: A case-control study. Eur J Obstet Gynecol Reprod Biol X. 2019;3:100041.
30. Gilabert-Estellés J, Ramón LA, España F, Gilabert J, Vila V, Réganon E, et al. Expression of angiogenic factors in endometriosis: relationship to fibrinolytic and metalloproteinase systems. Hum Reprod. 2007;22(8):2120-7.
31. Rodgers WH, Osteen KG, Matrisian LM, Navre M, Giudice LC, Gorstein F. Expression and localization of matrilysin, a matrix metalloproteinase, in human endometrium during the reproductive cycle. Am J Obstet Gynecol. 1993;168(1 Pt 1):253-60.
32. Rodgers WH, Matrisian LM, Giudice LC, Dsupin B, Cannon P, Svitek C, et al. Patterns of matrix metalloproteinase expression in cycling endometrium imply differential functions and regulation by steroid hormones. J Clin Invest. 1994;94(3):946-53.
33. Madlener M, Mauch C, Conca W, Brauchle M, Parks WC, Werner S. Regulation of the expression of stromelysin-2 by growth factors in keratinocytes: implications for normal and impaired wound healing. Biochem J. 1996;320 (Pt 2)(Pt 2):659-64.
34. Nap AW, Dunselman GA, de Goeij AF, Evers JL, Groothuis PG. Inhibiting MMP activity prevents the development of endometriosis in the chicken chorioallantoic membrane model. Hum Reprod. 2004;19(10):2180-7.
35. McMahon M, Ye S, Izzard L, Dlugolenski D, Tripp RA, Bean AG, et al. ADAMTS5 Is a Critical Regulator of Virus-Specific T Cell Immunity. PLoS Biol. 2016;14(11):e1002580.
36. Seargent JM, Loadman PM, Martin SW, Naylor B, Bibby MC, Gill JH. Expression of matrix metalloproteinase-10 in human bladder transitional cell carcinoma. Urology. 2005;65(4):815-20.
37. Mathew R, Khanna R, Kumar R, Mathur M, Shukla NK, Ralhan R. Stromelysin-2 overexpression in human esophageal squamous cell carcinoma: potential clinical implications. Cancer Detection and Prevention. 2002;26(3):222-8.
38. Kerkelä E, Ala-aho R, Jeskanen L, Lohi J, Grénman R, M-Kähäri V, et al. Differential patterns of stromelysin-2 (MMP-10) and MT1-MMP (MMP-14) expression in epithelial skin cancers. British Journal of Cancer. 2001;84(5):659-69.
39. Zhang G, Miyake M, Lawton A, Goodison S, Rosser CJ. Matrix metalloproteinase-10 promotes tumor progression through regulation of angiogenic and apoptotic pathways in cervical tumors. BMC Cancer. 2014;14(1):310.
40. Lizio M, Abugessaisa I, Noguchi S, Kondo A, Hasegawa A, Hon CC, et al. Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. Nucleic Acids Res. 2018;47(D1):D752-D8.
41. Lin S, Lin Y, Nery JR, Urich MA, Breschi A, Davis CA, et al. Comparison of the transcriptional landscapes between human and mouse tissues. Proc Natl Acad Sci U S A. 2014;111(48):17224-9.
42. Wang D, Eraslan B, Wieland T, Hallström B, Hopf T, Zolg DP, et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. Mol Syst Biol. 2019;15(2):e8503.

43. Cao X, Li Y, Luo RZ, Zhang L, Zhang SL, Zeng J, et al. Expression of Cystatin SN significantly correlates with recurrence, metastasis, and survival duration in surgically resected non-small cell lung cancer patients. Sci Rep. 2015;5:8230.
44. Choi EH, Kim JT, Kim JH, Kim SY, Song EY, Kim JW, et al. Upregulation of the cysteine protease inhibitor, cystatin SN, contributes to cell proliferation and cathepsin inhibition in gastric cancer. Clin Chim Acta. 2009;406(1-2):45-51.
45. Yoneda K, Iida H, Endo H, Hosono K, Akiyama T, Takahashi H, et al. Identification of Cystatin SN as a novel tumor marker for colorectal cancer. Int J Oncol. 2009;35(1):33-40.
46. Keppler D, Zhang J, Bihani T, Lin AW. Novel Expression of CST1 as Candidate Senescence Marker. The Journals of Gerontology: Series A. 2011;66A(7):723-31.
47. Brighton PJ, Maruyama Y, Fishwick K, Vrljicak P, Tewary S, Fujihara R, et al. Clearance of senescent decidual cells by uterine natural killer cells in cycling human endometrium. Elife. 2017;6.
48. Lucas ES, Vrljicak P, Muter J, Diniz-da-Costa MM, Brighton PJ, Kong C-S, et al. Recurrent pregnancy loss is associated with a pro-senescent decidual response during the peri-implantation window. Communications Biology. 2020;3(1):37.
49. Hernandez-Segura A, Nehme J, Demaria M. Hallmarks of Cellular Senescence. Trends Cell Biol. 2018;28(6):436-53.
50. Kramann R, Schneider RK, DiRocco DP, Machado F, Fleig S, Bondzie PA, et al. Perivascular Gli1+ progenitors are key contributors to injury-induced organ fibrosis. Cell Stem Cell. 2015;16(1):51-66.
51. Yan D, Liu X, Guo S-W. Neuropeptides Substance P and Calcitonin Gene Related Peptide Accelerate the Development and Fibrogenesis of Endometriosis. Scientific Reports. 2019;9(1):2698.
52. Liu X, Yan D, Guo S-W. Sensory nerve-derived neuropeptides accelerate the development and fibrogenesis of endometriosis. Human Reproduction. 2019;34(3):452-68.
53. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. Fertil Steril. 1997;67(5):817-21.
54. McKinnon BD, Evers J, Bersinger NA, Mueller MD. Induction of the Neurokinin 1 Receptor by TNFα in Endometriotic Tissue Provides the Potential for Neurogenic Control Over Endometriotic Lesion Growth. The Journal of Clinical Endocrinology & Metabolism. 2013;98(6):2469-77.
55. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4(1).
56. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2012;29(1):15-21.
57. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biology. 2019;20(1):296.
58. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nature Methods. 2019;16(12):1289-96.

59. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):W90-7.
60. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nature Biotechnology. 2020;38(12):1408-14.

Figures

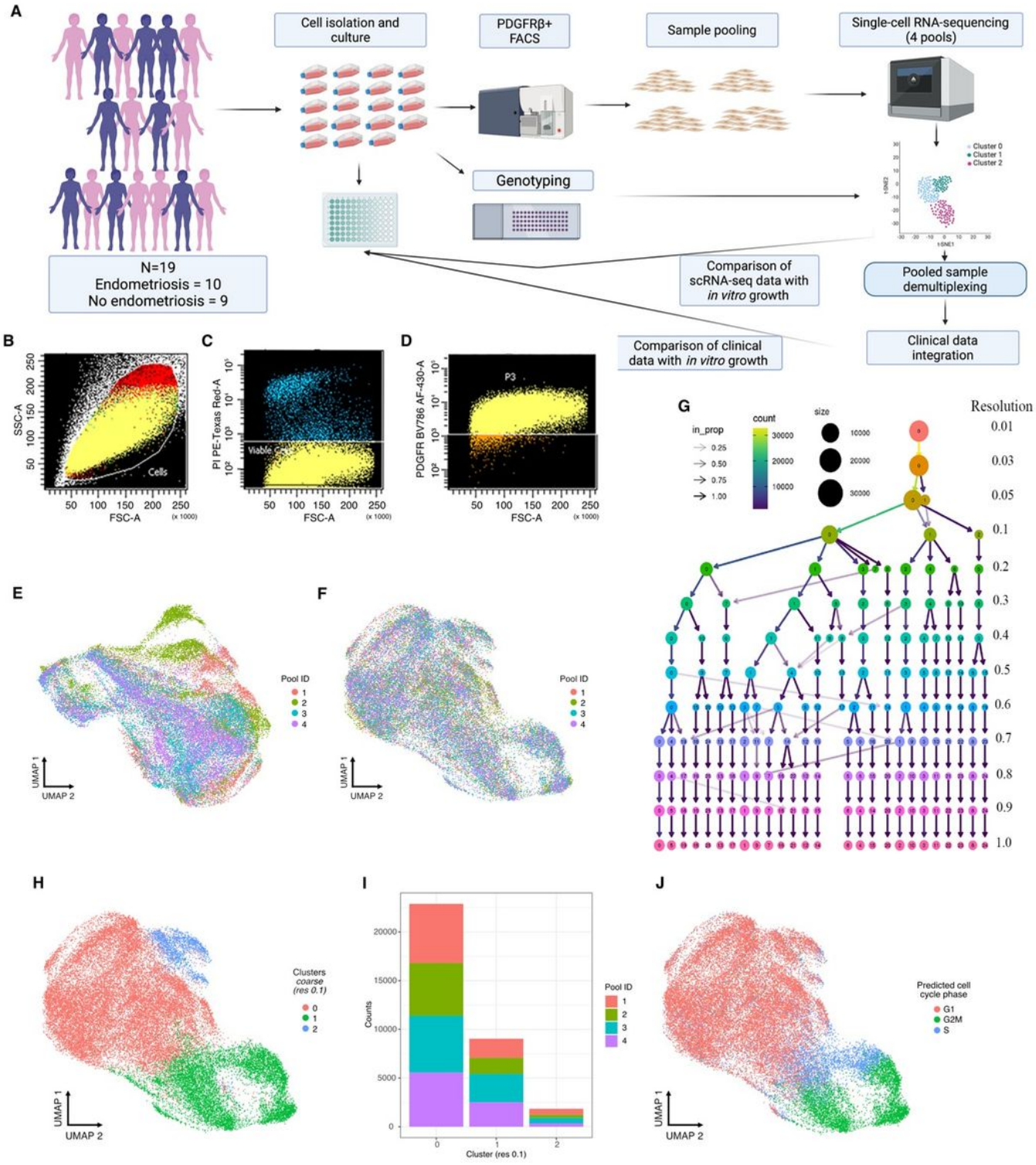

Figure 1

Experimental design and quality control for high throughput single-cell RNA sequencing of purified endometrial mesenchymal cells. (A) Endometrial biopsies were isolated, from 19 different women and cultured. In vitro growth assays, DNA isolation for genotyping and PDGFRβ+ FACS purification was performed. Purified cells from nineteen samples were pooled into 4 lanes and run on the 10X Chromium controller and scRNA-seq data analysed and cell clustering performed. Samples were assigned to their source individual and clusters compared to clinical data. Both scRNA-seq data and clinical data was compared to in vitro growth. Mesenchymal stromal cells were purified via FACS (B) forward and side scatter, (C) viability and (D) PDGFRβ+ expression. UMAP plot distribution of scRNA-seq data was determined both (E) pre- and (F) post harmony correction of between-pool variations introduced through technical variations. (G) Clustree analysis was used to determine the most stable level for cellular clustering. (H) UMAP plot of integrated scRNA-seq data at a clustering resolution of 0.1 identified three distinct clusters with minimal subsequent mixing at finer resolution. (I) At a cluster resolution of 0.1, cell numbers from each pool in each cluster remained stable suggesting clusters were formed from biological differences rather than technical effects. (J) UMAP plot of scRNA-seq data labelled by cell cycle phase inferred by CellcycleSorting. (A) Created with BioRender.com

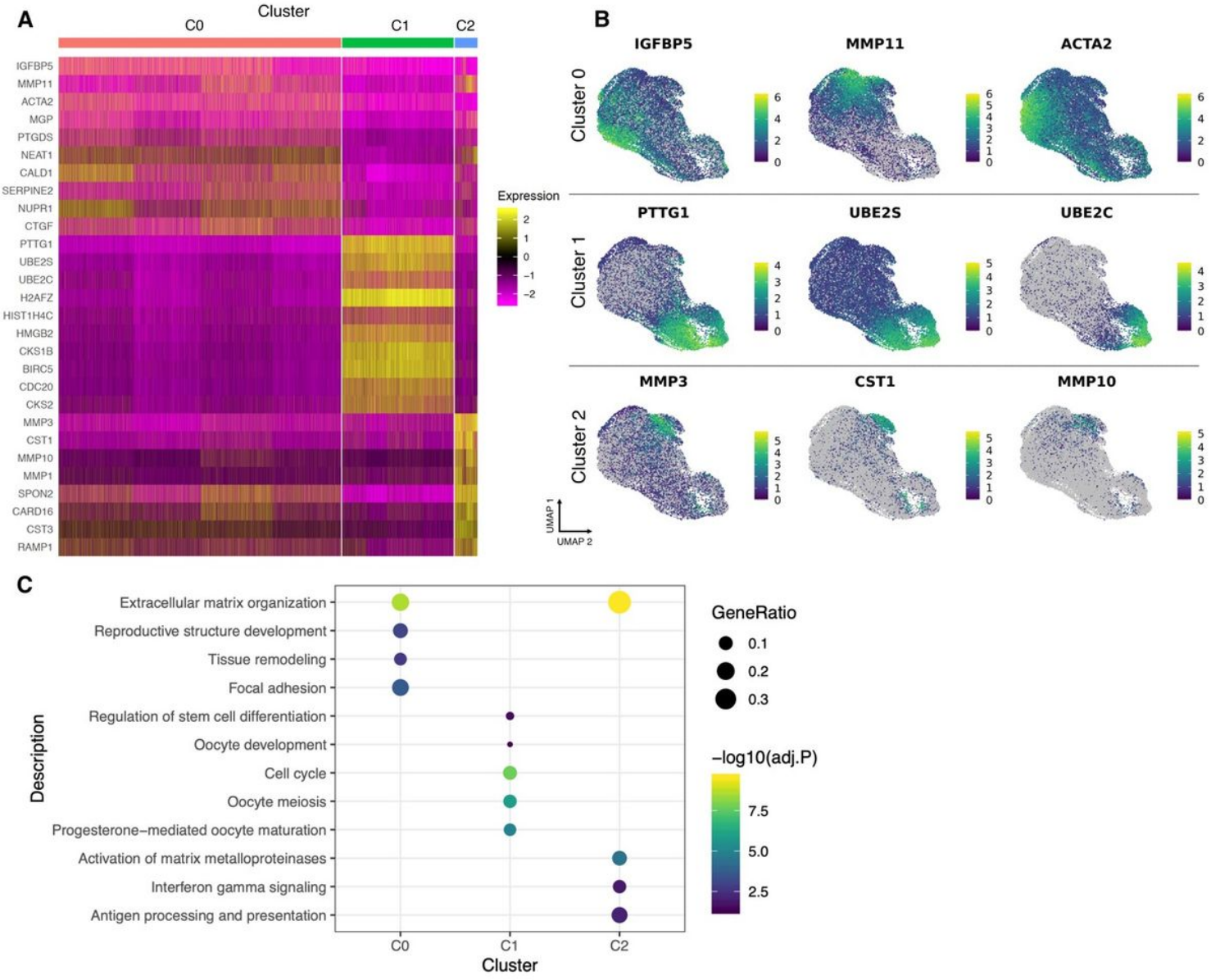

Figure 2

Differential gene expression and cell cycle characteristics of cell clusters. (A) A heatmap representation of the differentially expressed genes (DEGs) across the different clusters. (B) UMAP plots coloured by log normalized expression of the top DEGs per cluster show strong gene expression in cells consistent with the location of the respective cluster, although for each gene the individual cellular expression is variable. IGFBP5, MMP11 and ACTA2 showed strong expression in cluster 0, but in distinct cells. (C) Pathway analysis indicated a role for extracellular matrix organisation in cluster 0, oocyte maturation and meiosis and cluster 1 and activation of matrix metalloproteinases, interferon gamma signalling and antigen processing and presentation in cluster 2.

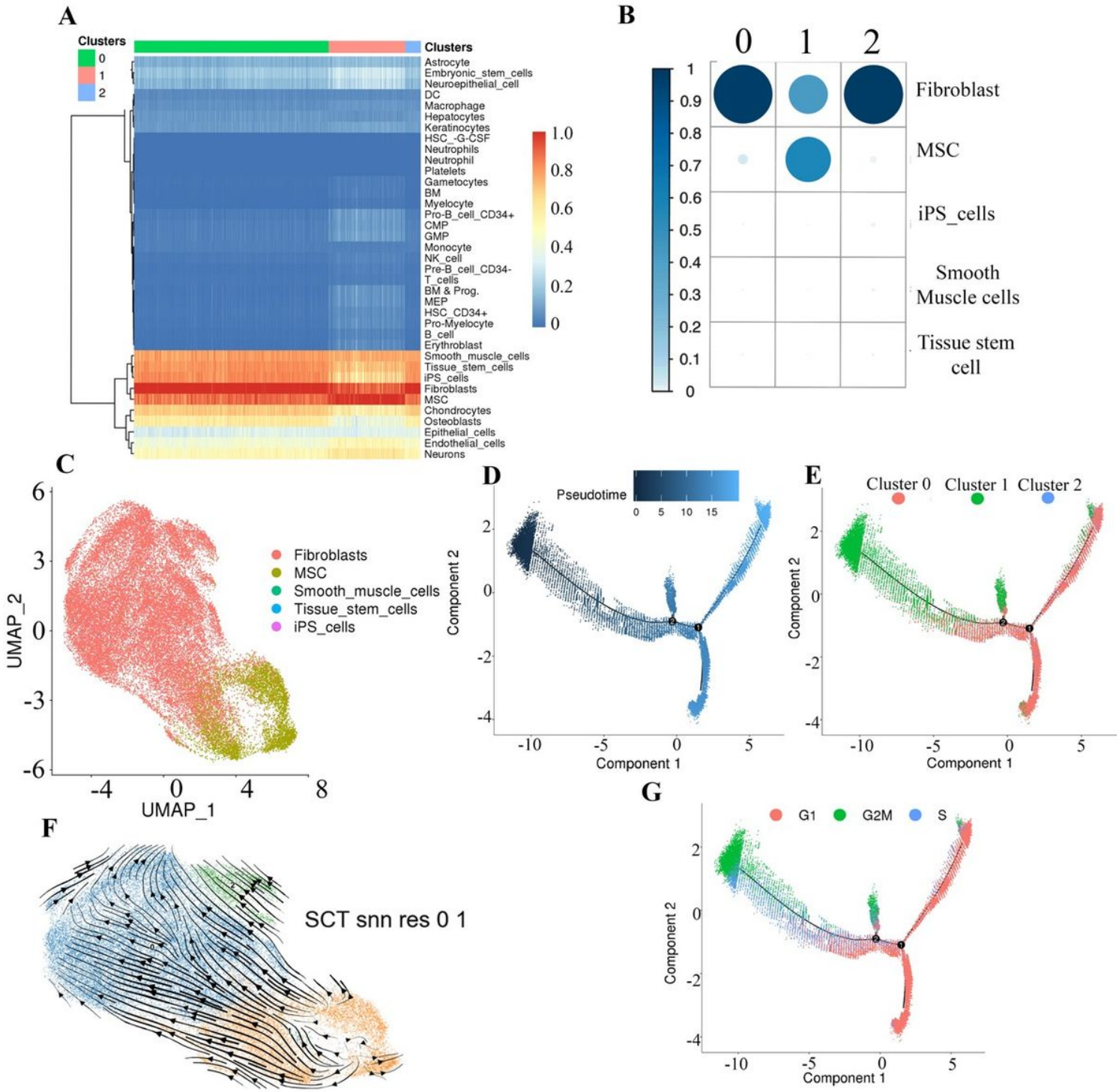

Figure 3

Cell type identification. SingleR was used to assign cell type based on transcriptomic signatures. Strongest correlations were observed in red with weakest in blue. (A) Heatmap visualisation of the transcriptomic signatures aligned strongly with cells of mesenchymal origin. This was consistent across all three clusters, although the actual identity varied across the clusters. (B) Distribution of the top cell type assignments across the three clusters. Circle size represents the proportion of each cell type identified in each cluster. Analysis indicated that fibroblasts were the predominant cells in cluster 0 and

cluster 2. The majority of cluster 1 were MSCs, although 42.53% were considered fibroblast. (C) UMAP plot of scRNA-seq data labelled by cell type. Overlay of cell type on spatially resolved distribution depicts fibroblasts as the predominant cell in cluster 0 and cluster 2 and an association between MSCs with cluster 1. (D) Pseudotime cell fate trajectory analysis using Monocle 2 placed each cell on a continuum based on the similarity of the transcriptome. (E) An overlay of the clusters identified the majority of the cluster 1 MSCs as the root source directing cell fate lineage towards the fibroblast major cluster (cluster 0). The additional fibroblast minor cluster (cluster 2) was uniformly scattered across the continuum. (F) ScVelo analysis supports the developmental trajectory direction of cluster 1 to cluster 0. A similar direction is also taken by cluster 2. (G) An overlay of the cell cycle analysis identifies the majority of G2M phase cells aligned with earlier trajectory and MSCs, transitioning to S phase followed by movement towards the majority G1 stage cells.

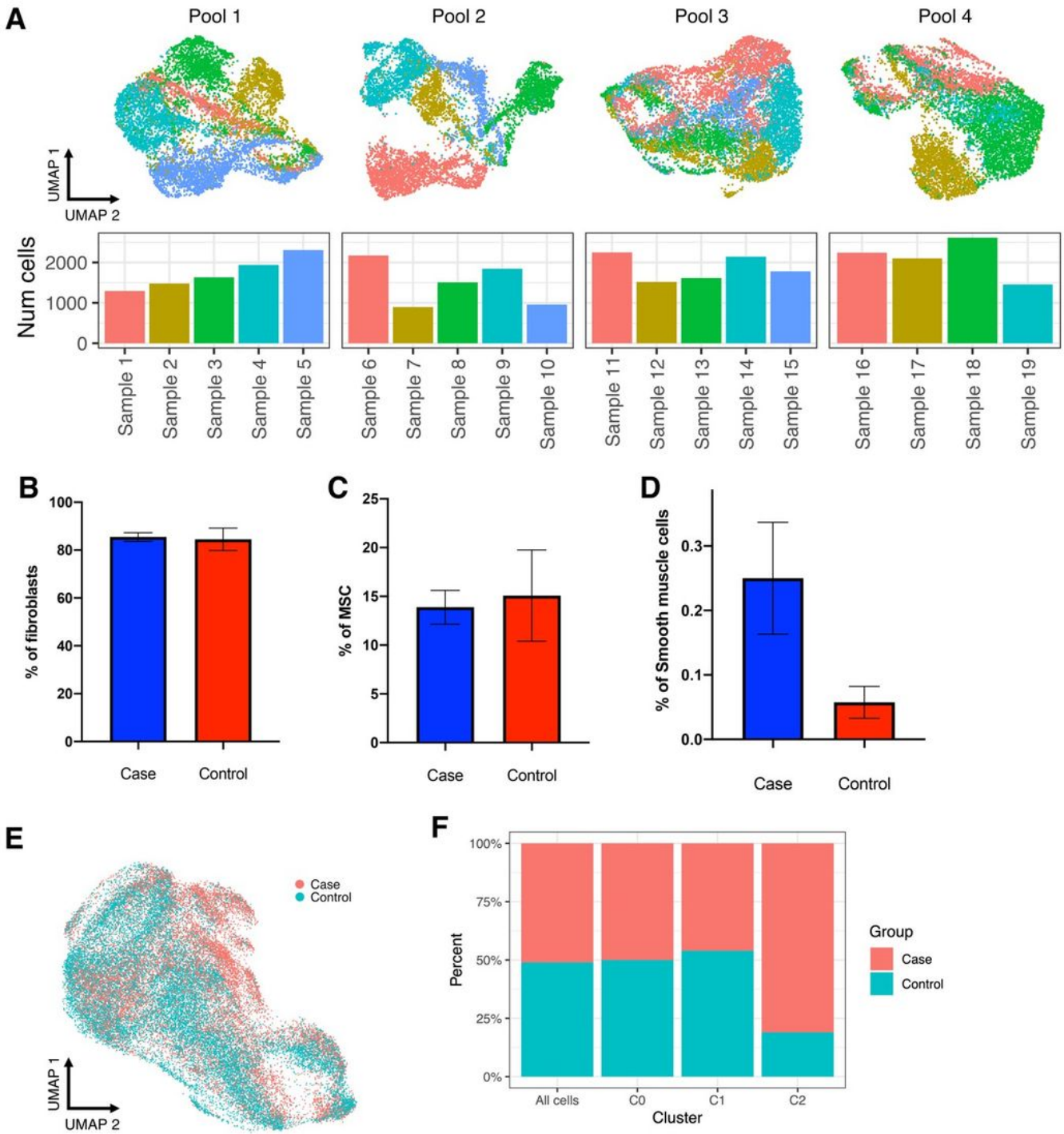

Figure 4

Cell type and cell cluster relationship to endometriosis status. (A) Genotype data and the demuxlet program was used to deconvolute individual samples from the mixed pool of cells; 100% of cells were correctly assigned to the patient donor. Overlaying clinical data and comparing the number of cells derived from women with endometriosis (cases) and without endometriosis (controls) found no significant difference between (B) the number of fibroblasts, (C) the number of MSCs or (D) the number

of smooth muscle cells. (E) UMAP plot of scRNA-seq data labelled by endometriosis status. (F) Percentage of cells assigned case or control status across all cells and per cluster (cluster 0, 1 and 2). A Chi-squared comparison between each cluster confirmed a significant association ($p < 0.0001$) between endometriosis cases and cluster 2.

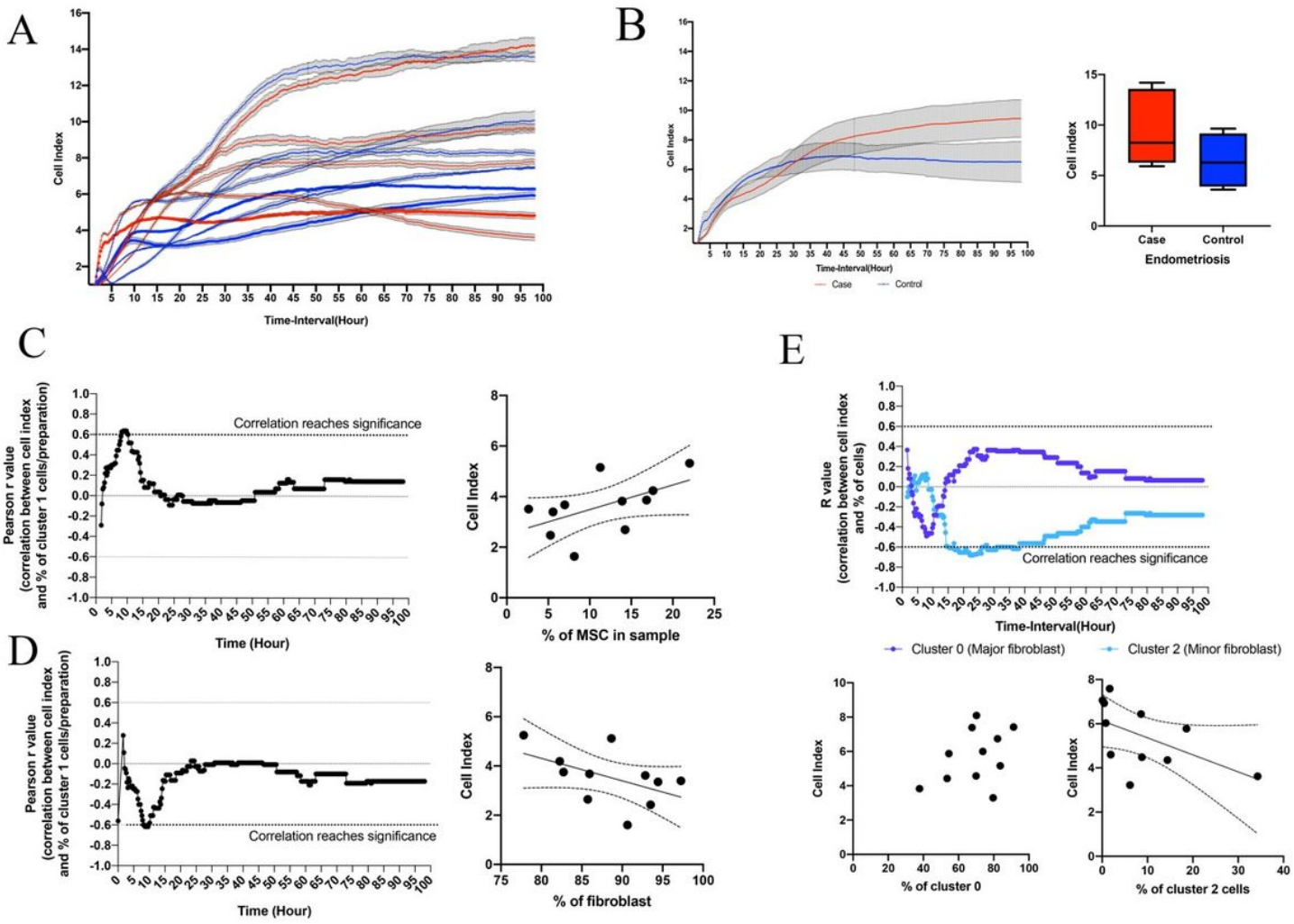

Figure 5

Relationship between cell types and cell clusters and in vitro behaviour. We assessed cell growth rates over a 100 hour period with the xCELLigence assay and compared the cell index to the percentage of cell types, or cell clusters identified in each cell preparation. (A) A comparison of the cell index for each individual preparation showed large variations in growth rates. (B) Categorisation of cell preparations based on endometriosis status showed varying rates of increase for both cases and controls across the 100 hour growth assay. At the 100 hour end point the number of cells was higher for endometriosis cases compared to controls, although not significantly ($p = 0.4725$). We subsequently compared the percentage of cell types in each preparation with the cell index across all time points. This identified an increasingly positive correlation with cell index that reached significance (dotted line) with (C) MSC between 8.25–9.75 hours after cell seeding, peaking at 9.25 (inset correlation graph). Conversely, for fibroblasts (D), a significant negative correlation with cell index was observed at the same time point (inset correlation

graph). (E) Lastly, comparison with the fibroblast cell clusters revealed the association with cell growth was strongest between 17-26 hours after initial seeding, with the slower growth rate of cluster 2 reaching significance (inset correlation graph) during this time period.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.docx
- Supplementarytable1.xlsx
- Supplementarytable2.xlsx
- Supplementarytable3.xlsx
- Supplementarytable4.xlsx
- Supplementarytable5.xlsx
- Supplementarytable6.xlsx