Mapping the complex paracrine response to hormones in the human breast at single-cell resolution

Lyndsay M Murrow[1], Robert J Weber[1,2], Joseph Caruso[3], Christopher S McGinnis[1], Alexander D Borowsky[4], Tejal A Desai[5], Matthew Thomson[6], Thea Tlsty[3], and Zev J Gartner[1,7]†

[1] Department of Pharmaceutical Chemistry and Center for Cellular Construction, University of California San Francisco, San Francisco CA 94158.
[2] Medical Scientist Training Program (MSTP), University of California, San Francisco, San Francisco, California, USA.
[3] Department of Pathology and Helen Diller Cancer Center, University of California San Francisco, San Francisco CA 94143.
[4] Center for Comparative Medicine, University of California Davis, Davis CA 95696.
[5] Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco CA 94158.
[6] Computational Biology, Caltech, Pasadena CA 91125.
[7] Chan Zuckerberg Biohub, University of California San Francisco, San Francisco CA 94158.

† Corresponding author: Email: zev.gartner@ucsf.edu (ZJG)

Abstract

The rise and fall of estrogen and progesterone across menstrual cycles and during pregnancy controls breast development and modifies cancer risk. How these hormones uniquely impact each cell type in the breast is not well understood, because many of their effects are indirect– only a fraction of cells express hormone receptors. Here, we use single-cell transcriptional analysis to reconstruct *in silico* trajectories of the response to cycling hormones in the human breast. We find that during the menstrual cycle, rising estrogen and progesterone levels drive two distinct paracrine signaling states in hormone-responsive cells. These paracrine signals trigger a cascade of secondary responses in other cell types, including an "involution" transcriptional signature, extracellular matrix remodeling, angiogenesis, and a switch between a pro- and anti-inflammatory immune microenvironment. We observed similar cell state changes in women using hormonal contraceptives. We additionally find that history of prior pregnancy alters epithelial composition, increasing the proportion of myoepithelial cells and decreasing the proportion of hormone-responsive cells. These results provide systems-level insight into the links between hormone cycling and breast cancer risk.

Keywords: mammary gland, single-cell RNA sequencing, menstrual cycle, parity, pregnancy history, hormone signaling, breast cancer risk

Introduction

The rise and fall of estrogen and progesterone with each menstrual cycle and during pregnancy controls cell growth, survival, and tissue morphology in the breast. The impact of these changes is profound, and lifetime exposure to cycling hormones is a major modifier of breast cancer risk.

Each additional year of menstrual cycling due to early age of menarche or late menopause leads to a 3-5% increased risk of breast cancer (Collaborative Group on Hormonal Factors in Breast Cancer, 2012). In contrast, pregnancy has two opposing effects on breast cancer risk: it increases the short-term risk by up to 25% (Lambe et al., 1994), but decreases lifetime risk by up to 50% for early pregnancies (Britt et al., 2007). Multiple mechanisms have been proposed to explain the opposing effects of pregnancy on breast cancer risk. The direct link between pregnancy and the long-term reduction in risk remains an open question, but has been attributed to the effects of pregnancy-induced lobuloalveolar differentiation, including a decrease in the hormone-responsiveness of the epithelium and reduced frequency of tumor-susceptible cell populations (Britt et al., 2007; Russo et al., 1992). In contrast, stromal as well as epithelial changes are thought to drive the increased short-term risk following pregnancy. High hormone levels during pregnancy and lactation promote growth in cells with preexisting oncogenic mutations (Haricharan et al., 2013). Following lactation, involution drives a suite of stromal changes that have been proposed to provide a favorable tumor microenvironment, including extracellular matrix (ECM) remodeling, recruitment of phagocytic M2-like macrophages, and increased angiogenesis (Lyons et al., 2011; O'Brien et al., 2010; Schedin et al., 2007).

While the cellular and molecular changes that occur during pregnancy and involution have been well-studied, particularly in mice, far less is known about the underlying mechanisms that link breast cancer risk and menstrual cycling. The menstrual cycle is characterized by alternating periods of epithelial expansion and regression (Anderson et al., 1982; Söderqvist et al., 1997), and histological analyses of paraffin-embedded human tissue sections have identified alterations in epithelial architecture and stromal organization (Ramakrishnan et al., 2002; Vogel et al., 1981). However, a systems-level understanding of how different cell populations respond to cycling hormone levels, and whether these responses parallel those observed during pregnancy and involution, remains unclear.

One major barrier to understanding the mechanistic links between the menstrual cycle and breast cancer risk is that many of the effects of hormone signaling within the breast are indirect. The estrogen and progesterone receptors (ER/PR) are expressed in only 10-15% of luminal cells within the epithelium, known as hormone-responsive or hormone-receptor positive (HR+) luminal cells (Clarke et al., 1997). Thus, most of the effects of hormone receptor activation are mediated by a complex cascade of paracrine signaling events. In addition to HR+ luminal cells, the human breast comprises two other epithelial cell types—hormone-insensitive (HR-) luminal cells (also termed luminal progenitors), which are the secretory cells that produce milk during lactation, and myoepithelial cells, which contract to move milk through the ducts—as well as multiple stromal cell types including fibroblasts, adipocytes, immune cells, and endothelial cells. Each of these populations is likely affected by paracrine signaling downstream of hormone receptor activation, with distinct effects in each cell type. Additional barriers to understanding the effect of hormone signaling include major differences in glandular architecture and stromal composition and complexity between humans and model organisms like the mouse (Dontu and Ince, 2015; Parmar and Cunha, 2004). Notably, ER expression is restricted to the epithelium in humans but is also expressed in the stroma in rodents (Mueller et al., 2002; Palmieri et al., 2004). Thus, understanding the consequences of epithelial-stromal crosstalk downstream of estrogen and progesterone requires studying these processes in humans or human models.

A second challenge for understanding the role of estrogen and progesterone in breast cancer risk is that the dynamic rise and fall of hormones may be as important as absolute levels: although each additional year of menstrual cycling increases the risk of breast cancer, serum hormone levels do not themselves directly correlate with risk (Schernhammer et al., 2013). Despite this, prior studies have not directly investigated the effects of hormone dynamics on cell state across all cell types in the breast. One barrier is the marked differences in hormone levels and kinetics (4-5 versus 28 days) between the rodent estrous cycle and human menstrual cycle (Dontu and Ince, 2015). While a human explant model has been developed to more accurately capture the acute response of the human breast to hormone treatment (Tanos et al., 2013), this models remains limited to relatively short ~24h treatments with estrogen and progesterone and is not fully integrated with the vascular and immune compartments. It is therefore unclear how well both mouse models and human tissue explants recapitulate the full time course and spectrum of cellular interactions that define the human menstrual cycle *in vivo*. As it enables unbiased analysis of the full repertoire of cell types within the human mammary gland, single-cell RNA sequencing (scRNAseq) is particularly well-suited to investigate this problem. Despite this, prior scRNAseq studies in human have primarily focused on the epithelium, and have not examined the response to hormone receptor activation (Nguyen et al., 2018).

Here, we use scRNAseq to trace the transcriptional changes that occur in the human breast in response to cycling hormone levels, using healthy tissue from a cohort of reduction mammoplasty patients. Since hormone receptors are master regulators of breast development, we hypothesized that pregnancy history and menstrual cycle stage would be the major sources of variability between specimens (Figure 1A). Therefore, we first compared the transcriptional profile of nulliparous and parous women and identified major changes in the cellular composition of the breast following pregnancy. We found that prior history of pregnancy was associated with striking changes in epithelial composition, and we propose that these changes are consistent with the protective effect of pregnancy on lifetime breast cancer risk. By decoupling these effects of pregnancy on cell proportions, we then used menstrual cycle staging and pseudotemporal analysis to map cell state changes in response to fluctuating hormone levels across the menstrual cycle, and used "virtual experiments" in an additional cohort of patients with hormonal contraceptive use at the time of surgery to confirm key findings. We found that the changing hormonal microenvironment across the menstrual cycle led to wide-ranging cell state changes in both the epithelium and stroma. In addition to transcriptional signatures consistent with tissue expansion and remodeling, we uncovered changes that mimic those seen during postpartum involution (Lyons et al., 2011; O'Brien et al., 2010; Schedin et al., 2007). Thus, our findings suggest that similar mechanisms contribute to both the short-term increased breast cancer risk following pregnancy and the lifetime increased risk due to total number of menstrual cycles. Overall, these results provide a comprehensive map of the cycling human breast and identify the cellular changes that underlie breast cancer risk.

Results

scRNAseq identifies three major epithelial and four major stromal cell types in the human breast

To determine how cycling estrogen and progesterone levels affect cell composition and cell state in the human breast, we performed scRNAseq on 43,021 cells collected from reduction

mammoplasties from five age-matched, premenopausal donors without hormonal contraceptive use (Table S1, Figure 1B). To obtain an unbiased snapshot of both the epithelium and stroma, we collected live/singlet cells identified on the basis of forward and side scatter and lack of DAPI staining. For four samples, we additionally collected purified luminal and myoepithelial cells to provide additional confirmation of downstream clustering results (Figure S1A). We used the 10X Chromium system to prepare cell-barcoded cDNA libraries and sequenced approximately 3,000 cells per sample and sort condition (Table S2).

To investigate how the proportion and transcriptional state of each cell type changed in response to hormone levels, we first identified the major cell types present within the human breast. Sorted luminal and myoepithelial cell populations were enriched for the epithelial keratins KRT19 and KRT14, respectively (Figure S1B), and were well-resolved by TSNE dimensionality reduction (Figure S1C). Unbiased clustering identified three main epithelial populations—one myoepithelial cell type (C1) and two luminal cell types (C2-C3)—and four stromal populations (C4-C7) (Figure 1C). Hierarchical clustering and marker analysis identified the three epithelial populations as myoepithelial, hormone-responsive (HR+) luminal, and hormone-insensitive (HR-) luminal cells, and the four stromal populations as fibroblast, endothelial, vascular accessory, and immune cells (Figures 1D-E, Figure S1D-E). HR+ luminal cells expressed hormone receptors (ESR1/PGR) and markers such as amphiregulin (AREG) (Figure 1E, Figure S1F) (Ciarloni et al., 2007; Fridriksdottir et al., 2015). The HR+ luminal and HR- luminal clusters described here closely match the two luminal cell populations identified by a previous scRNAseq analysis of the human breast (Nguyen et al., 2018). The authors reported that the transcriptional signatures for these two populations most closely matched microarray expression data for what has been termed $EpCAM^{+}/CD49f^{-}$ "mature luminal cells" and $EpCAM^{+}/CD49f^{+}$ "luminal progenitors" (Lim et al., 2009; 2010). As recent mouse data suggests that the ER+ and ER- cell populations are maintained by independent lineage-restricted progenitors (Van Keymeulen et al., 2017; Wang et al., 2017), we propose the nomenclature "hormone-responsive/HR+ luminal" and "hormone-insensitive/HR- luminal" for these two cell types.

Parity leads to a change in epithelial composition

The breast undergoes numerous changes during pregnancy and involution, and we hypothesized that these changes would be a major driver of sample-to-sample variability in our dataset. To identify a parity signature, we focused our initial analysis on the 14,797 cells in the live/singlet sort gate to get an unbiased view of how the overall composition of the breast changed with history of pregnancy. Based on clustering results, we observed a striking change in epithelial composition in parous women, characterized by an increase in the proportion of myoepithelial cells within the epithelium and a decrease in the proportion of HR+ luminal cells within the luminal compartment (Figure 2A).

We confirmed the increase in myoepithelial proportions by flow cytometry analysis of EpCAM and CD49f expression in 10 additional women. Parity was associated with an increase in the average proportion of myoepithelial cells from 18% to 44% of the epithelium (Figure 2B). The myoepithelial cell fraction correlated with pregnancy history ($R^2 > 0.8$) but not with other discriminating factors such as race, body mass index (BMI), or age (Figure 2B, S2A). As FACS processing steps may affect tissue composition, we performed two additional analyses. First, we

reanalyzed a previously published microarray dataset from total RNA isolated from breast core needle biopsies (Peri et al., 2012), and found a significant increase in the myoepithelial markers TP63, KRT5, and KRT14 relative to luminal keratins (Table S3). Second, we performed immunohistochemistry on matched formalin-fixed, paraffin-embedded tissue. Staining for the myoepithelial marker p63 and luminal KRT7 confirmed a significant increase in the ratio of p63+ myoepithelial cells to KRT7+ luminal cells in intact tissue sections (Figure 2C).

We next analyzed our clustering results from the 11,410 cells in the luminal sort gate to confirm the decreased frequency of HR+ versus HR- luminal cells we observed in the live/singlet gate. While the separation between the hormone-responsive and hormone-insensitive luminal cell populations is not always distinct by FACS (Figure S1A) (Lim et al., 2010), transcriptome analysis clearly distinguished between these two cell types and demonstrated a marked increase in the proportion of HR- luminal cells relative to HR+ luminal cells in parous samples (Figure 2D). To verify these results in intact tissue sections, we performed immunohistochemistry for ER and PR. There was a trend toward decreased expression of both receptors in parous samples, although only the change in double-positive ER+/PR+ cells was statistically significant (Figure S2B). This is consistent with previous studies which consistently found decreased expression of ER and/or PR in parous samples, but with varying degrees of statistical significance (Battersby et al., 1992; Muenst et al., 2017; Taylor et al., 2009). We speculate that this variability is due to changes in ER and PR expression, stability, and nuclear localization across the menstrual cycle based on hormone receptor activation status (Battersby et al., 1992; Métivier et al., 2003; Petz and Nardulli, 2000). Supporting this, we found that although ER transcript and protein levels correlate across tissue sections, they do not correlate on a per-cell basis (Figure S2C). Therefore, we sought to identify another marker to more reliably distinguish between HR+ and HR- cell populations, and identified keratin 23 (KRT23) as highly enriched in HR- cells (Figure S2D), as was also reported by a previous scRNAseq study (Nguyen et al., 2018). Immunohistochemistry for KRT23 and ER confirmed that these two proteins are expressed in mutually exclusive populations (Figure 2E). KRT23 thus represents a discriminatory marker between the two luminal populations that does not fluctuate with hormone signaling as the receptors themselves do. Staining for KRT23 in intact tissue sections confirmed a significant increase in KRT23+ HR- luminal cells from 7% to 21% in parous samples (Figure 2F). Together, these results demonstrate a striking change in epithelial composition with parity (Figure 2G).

Pseudotemporal analysis of individual epithelial cell types orders samples according to menstrual cycle stage

As hormone signaling controls breast development during each menstrual cycle in addition to pregnancy, we predicted that menstrual cycle stage would be a second major source of variability between samples (Figure 2G). We used previously described morphological criteria to stage each sample along the menstrual cycle (Longacre and Bartow, 1986; Ramakrishnan et al., 2002). Blinded analysis placed three samples in the follicular phase (stages I-II) and two in the luteal phase (stages III-IV) (Figure 2H). We confirmed this staging using a previously published bulk RNA sequencing dataset (Pardo et al., 2014) to develop an average menstrual cycle score for each sample (Figure S2E, Table S4). Additionally, since PR is upregulated following estrogen exposure, we predicted that samples in the luteal phase would have a higher proportion of PR+ cells. To normalize the effects of parity on the proportion of HR+ luminal cells, we quantified

the percentage of PR+ cells within the KRT7+/KRT23- HR+ luminal cell population in our five sequenced samples and found a significant increase in the percentage of PR+ cells in the luteal phase of the menstrual cycle (Figure S2F). Finally, as a measure of transcriptional variation between each sequenced sample, we quantified the earth mover's distance (EMD) between each sample in principal component (PC) space, representing the minimum cost of moving one distribution onto another. We chose this approach rather than distance-based similarity metrics since EMD measures the transcriptional variation between entire cell populations—containing multiple cell types—rather than between individual cells. Using hierarchical clustering of the EMD results, we found that samples were grouped according to their predicted menstrual cycle stage (Figure 2I). Thus, the transcriptional differences observed between samples mapped directly onto changes in hormone signaling state that vary with menstrual cycle stage.

To understand how HR+ luminal cells change across the menstrual cycle, we performed pseudotemporal ordering of single cells along a cell state trajectory (Trapnell et al., 2014). Strikingly, this temporal ordering matched the predicted menstrual cycle stage for each sample based on morphological and transcriptional analysis (Figure 2J). To test whether hormone cycling also drives transcriptional changes within a non-hormone-responsive cell population via paracrine signaling, we performed a similar analysis in myoepithelial cells. As in the HR+ population, ordering of single myoepithelial cells placed each sample according to its predicted menstrual cycle stage (Figure S2G). Together, these data demonstrate that hormone levels and their paracrine signaling effectors are the major driver of transcriptional changes—and inter-sample variability—within both hormone-responsive and hormone-insensitive cell populations.

Two transcriptional states emerge in hormone-responsive luminal cells as estrogen and progesterone levels increase

Pseudotime analysis revealed that the HR+ luminal population transitioned from one to two distinct transcriptional states as progesterone increased in the luteal phase (Fig 2J). To investigate these diverging cell states, we used principal component analysis (PCA) and non-negative matrix factorization (NMF) to perform detailed sub-clustering analysis on the HR+ cell population. We chose NMF because this clustering method most accurately distinguishes between groups of cells with high similarity (Zhu et al., 2017), as is the case for classifying cell signaling states rather than distinct cell types. We identified three clusters of HR+ luminal cells (Figure 3A, Figure S3A-B). Each cluster primarily localized to one branch of the pseudotime trajectory (Figure 3B), representing one cell state enriched in the follicular phase (HR+ 1) and two cell states enriched in the luteal phase (HR+ 2 and 3) (Figure 3C, Figure S3C). Notably, all three clusters were present in the early luteal phase, suggesting that both HR+2 and HR+3 emerged simultaneously as estrogen and progesterone levels increased. In support of this, we used the Velocyto tool (La Manno et al., 2018) and found that RNA velocity estimation predicted a similar bifurcating cell state trajectory as subclustering and Monocle analysis (Figure S3D).

We next used pseudotime ordering to identify genes that changed as HR+ cells transitioned from the follicular to the luteal phase (Figure 3D). Upregulated transcripts across both luteal-phase cell states included previously described ER targets such as amphiregulin (AREG) (Ciarloni et al., 2007) and the trefoil factors TFF1 and TFF3 (May and Westley, 2015; Rio et al., 1987),

transcription factors associated with luminal differentiation such as ID2 (Seong et al., 2018) and HES1 (Bouras et al., 2008), and chemokines such as CXCL13 and CXCL3. This analysis also identified genes with branch-specific changes in gene expression levels (Figure 3E). In the HR+ 2 branch, we found upregulation of RANK ligand (TNFSF11) and WNT4, which are well-characterized downstream mediators of progesterone receptor activation that signal to myoepithelial cells and HR- luminal cells (Joshi et al., 2015; Tanos et al., 2013). The HR+ 3 branch was characterized by expression of the PR targets KLF4 (Shimizu et al., 2010) and SOX4 (Graham et al., 1999), as well as upregulation of a hypoxia gene signature and pro-angiogenic factors such as VEGFA and ANGPTL4. Interestingly, a previous study using microdialysis of healthy human breast tissue found that VEGF levels increased in the luteal phase of the menstrual cycle (Dabrosin, 2003). As estrogen response elements have been identified in the 5' and 3' untranslated regions of the VEGFA gene (Hyder et al., 2000), our results suggest that this increased expression is, in part, a direct effect of hormone signaling to a subpopulation of HR+ luminal cells.

To confirm these results *in vivo*, we performed marker analysis to identify genes specific to each cluster that could be used for immunohistochemical staining (Figure S3E). We identified LRRC26 as a marker of the WNT4/RANKL-expressing cluster HR+ 2 and P4HA1 as a marker of the hypoxia/pro-angiogenic cluster HR+ 3 (Figure 3F). To directly test whether the luteal-phase cell states were a result of estrogen and progesterone signaling, we performed immunostaining in additional tissue sections from donors using progestin-based or combined estrogen/progestin-based forms of hormonal contraceptives (Table S5). LRRC26 protein expression was upregulated in both the luteal phase and in women using hormonal contraception (Figure 3G) and marks a distinct set of luminal cells from P4HA1 (Figure 3H, Figure S3F). Moreover, these two subpopulations co-occurred within the same regions of the breast, demonstrating that they are not an artifact of sample processing. Together, these results demonstrate that high estrogen and progesterone in the luteal phase reveal two diverging transcriptional states in HR+ cells, one that signals via RANK ligand and WNT4 to the surrounding epithelium and a second that, in part, signals to the surrounding vasculature via VEGF signaling (Figure 3I).

Paracrine effects of hormone signaling in hormone-insensitive epithelial cells

To investigate cell state changes in hormone-insensitive epithelial populations downstream of paracrine signaling from HR+ luminal cells, we performed sub-clustering analysis on myoepithelial cells and HR- luminal cells as described above. NMF identified two subpopulations of myoepithelial cells (Figure 4A, Figure S4A-B), which represented a switch in cell states between the follicular phase (MEP 1) and the luteal phase (MEP 2) of the menstrual cycle (Figure 4B, Figure S4C). Marker analysis (Figure S4D) and gene set enrichment analysis (Subramanian et al., 2005) of differentially expressed genes identified signaling pathways upregulated in each phase of the menstrual cycle. Similar to the HR+ 3 subcluster of hormone-responsive cells, myoepithelial cells in the luteal phase had enrichment of transcripts involved in hypoxia and angiogenesis such as VEGFA and ANGPTL4 (Figure 4C). Luteal-phase myoepithelial cells were also enriched for signaling pathways involved in epithelial-mesenchymal transition (EMT) (TPM2, MYL9, ACTA2/SMA) and anchoring junction proteins (DST, ACTN1), suggesting that changes in actomyosin contractility and cell-ECM interactions

partly underlie the morphological changes seen in the breast epithelium across the menstrual cycle (Figure 4C). Follicular-phase myoepithelial cells were enriched for pathways involved in ribosome biogenesis and protein synthesis, including Myc target genes (Figure S4E).

A similar analysis in HR- luminal cells identified three subpopulations (Figure 4D, Figures S4F-G). As in myoepithelial cells, there was a switch in cell states between the follicular and luteal phases of the menstrual cycle, with one subpopulation enriched in the early follicular phase (stage I), one in the late follicular phase (stage II), and one in the luteal phase (stages III-IV) (Figure 4E, Figure S4H). Similar to other epithelial cell types, signaling pathways associated with hypoxia were upregulated in luteal-phase hormone-insensitive cells. The luteal phase was also associated with hallmarks of TNF-alpha signaling, including increased expression of TNFAIP6 and CCL4 (Figure 4F, Figure S4I). TNF itself was upregulated in both the luteal phase and the early follicular phase that immediately follows, suggesting that the TNF-alpha signature represented an autocrine signaling response (Figure S4J). Comparison of the early and late follicular phases uncovered a transcriptional signature in the early follicular phase that was similar to that identified during post-lactational involution (Clarkson et al., 2004; Stein et al., 2004). This "involution" gene signature was characterized by upregulation of death receptor ligands such as TNFSF10 (TRAIL) and of genes involved in the defense and immune response including the acute phase genes SAA1/2 and LCN2, complement components CFB and C1R, and the immunoglobulin-related gene LTF (Figure 4G). Moreover, expression of the phagocytic receptors CD14 and MARCO in this cluster suggests that HR- cells play a role as non-professional phagocytes in the clearance of apoptotic cells during the early follicular phase (Figure 4H), similar to what has been described during involution (Monks et al., 2008).

Finally, based on our finding that a subset of HR+ luminal cells upregulated WNT4 in the luteal phase, we examined whether canonical WNT signaling was activated in luteal-phase myoepithelial cells and HR- luminal cells. Additionally, to test whether WNT activation was a result of estrogen and progesterone signaling, we performed immunostaining in tissue sections from donors using progestin-based or combined estrogen/progestin-based forms of hormonal contraception. The WNT effector TCF7 was upregulated in both cell types during the luteal phase (Figure 4I). Staining in matched tissue sections confirmed that TCF7 protein is expressed in a majority of myoepithelial cells (~70%) and a subset of HR- luminal cells (~4%) both in the luteal phase and following hormonal contraceptive use (Figure 4J). Together, these data demonstrate that paracrine signaling following hormone receptor activation leads to a dramatic change in cell state in hormone-insensitive epithelial cell populations (Figure 4K). Moreover, in HR- luminal cells, the early follicular phase is characterized by an "involution" transcriptional signature that closely mimics changes seen during post-lactational mammary gland regression, suggesting that similar mechanisms control both processes.

Paracrine signaling downstream of hormone receptor activation supports a proangiogenic microenvironment in the luteal phase

Expression of pro-angiogenic factors such as VEGFA, TNF, and ANGPTL4 from both hormone-responsive and hormone-insensitive cell types was increased in the luteal phase, suggesting that there was a switch to a proangiogenic microenvironment as estrogen and progesterone levels increased. To test this prediction, we performed sub-clustering analysis to dissect the changes

that occur in endothelial cells across the menstrual cycle and identified two distinct endothelial populations (Figure 5A) that represented vascular or lymphatic endothelial cells, based on expression of the markers PDPN and PLVAP (Figure 5B) (Hirakawa et al., 2003; Niemelä et al., 2005). Within the lymphatic endothelium, sub-clustering identified a change in transcriptional states between the follicular and luteal phases of the menstrual cycle (Figure 5C, Figure 5A-B), characterized by upregulation of pathways involved in TNF-alpha signaling, hypoxia, and EMT in the luteal phase (Figure 5D, Figure S5C). A similar analysis in vascular endothelial cells identified three cell states representing cells in the follicular, early luteal, and late luteal phases (Figure 5E, Figures S5D-E). Vascular endothelial cells in the early luteal phase were enriched for genes associated with angiogenesis and blood vessel remodeling including PGF, EDN1, and VEGF receptors. Similar to the lymphatic endothelium, vascular endothelial cells in the late luteal phase had enrichment of pathways involved in EMT, hypoxia, and TNF-alpha signaling (Figure 5F, Figure S5F). Two VEGF receptors control angiogenesis: previous studies demonstrated that KDR promotes endothelial sprouting in response to VEGF and FLT1 antagonizes this response (Jakobsson et al., 2010). We found that expression of KDR was restricted to the early luteal phase, whereas FLT1 was expressed in both the early and late luteal phases (Figure 5G). Along with gene set enrichment analyses, these results suggest that new blood vessel formation mainly occurs at the beginning of the luteal phase and is followed by blood vessel remodeling or maturation in the late luteal phase.

The proangiogenic microenvironment of the luteal phase was also reflected in the transcriptional state of vascular accessory cells. Clustering resolved four cell populations (Figure 5H, Figure S5G) that represented pericytes or smooth muscle cells based on the expression of smooth muscle actin (ACTA2) (Figure 5I). In both accessory cell types, there was a switch in cell state between the follicular/early luteal phase and late luteal phase of the menstrual cycle (Figure 5J, Figure S5H). Similar to the blood and lymphatic endothelium, the late luteal phase was characterized by upregulation of TNF-alpha signaling in both cell types, as well as a hypoxic gene signature in pericytes and pathways involved in EMT in smooth muscle cells (Figure 5K, Figure S6I-J). Together, these data dissect the transcriptional changes in endothelial cells and accessory cell types that underlie angiogenesis and vascular remodeling during the menstrual cycle (Figure 5L).

Remodeling of the stromal and immune microenvironments in response to estrogen and progesterone

Histological and immunohistochemical analyses of human breast tissue sections have identified alterations in stromal organization and ECM composition across the menstrual cycle (Ferguson et al., 1992; Hallberg et al., 2010). To dissect the transcriptional changes that underlie this stromal remodeling, we performed sub-clustering analysis and identified three subpopulations of fibroblasts (Figure 6A, Figure S6A-B). We identified the FB 3 cluster as a preadipocyte population based on expression of genes involved in adipogenesis such as ADIRF and Adipsin (CFD) (Figure 6B). Compared to fibroblasts, preadipocytes were highly enriched for expression of proteoglycans such as DCN and OGN and non-fibrous ECM proteins such as DPT, whereas fibroblasts had increased expression of genes involved in ECM remodeling such as MMP3 and TIMP1 (Figure 6C, Figure S6C). Within the fibroblast population, we identified two cell states representing a switch between the follicular and luteal phases of the menstrual cycle (Figure 6D,

Figure S6D). The luteal phase was characterized by upregulation of ECM proteins including collagen family members (COL3A1, COL1A2) and fibronectin (FN1), ECM remodeling proteins such as matrix metalloproteinases (MMP10, MMP14) and LOXL2, and cytokines and growth factors such as IL6 and TGFB3 (Figure 6E-F, Figure S6E). Notably, TGFB3 signaling is a major signaling molecule involved in post-lactational involution that enhances phagocytosis by mammary epithelial cells (Fornetti et al., 2016), suggesting that TGFB3 secreted by fibroblasts at the end of the luteal phase activates the subset of HR- luminal cells identified in the early follicular phase that express "involution" markers including phagocyte receptors (Fig 4G).

Finally, we examined changes in the immune microenvironment across the menstrual cycle. Clustering identified 5 immune cell populations (Figure 6G), comprising two CD68+ macrophage populations representing M1- or M2-like polarized macrophages and three lymphocyte populations representing CD20+ (MS4A1) B cells, CD8+ T cells, and IgA-producing IgJ+ plasma cells (Figure 6H, Figure S6F). M1-like macrophages expressed pro-inflammatory growth factors and cytokines such as INHBA and IL6, whereas tissue-remodeling M2-like macrophages expressed the scavenger receptors CD163 and MSR1. Notably, our clustering data suggested that there was a switch from an M2-like to an M1-like macrophage polarization and increased recruitment of B and T lymphocytes during the luteal phase of the menstrual cycle (Figure 6I, Figure S6G), consistent with previous immunohistochemical studies demonstrating a decreased frequency of CD163+ macrophages and increased frequency of CD8 T cells in the luteal phase (Schaadt et al., 2017). Using flow cytometry analysis, we confirmed an increase in the proportion of SSC^{low} lymphocytes within the CD45+ immune cell population, and a decrease in the proportion of CD163+ M2-like macrophages within the CD45+/CD68+ population in luteal-phase samples relative to the follicular phase (Figure 6J). This switch to a proinflammatory immune microenvironment in the luteal phase coincided with increased expression of cytokines such as TNF in HR- luminal cells (Figure S4J) and IL6 in fibroblasts (Figure 6E). Together, these data suggest that a transition from a tissue-remodeling immune microenvironment to a pro-inflammatory microenvironment occurs between the follicular and luteal phases of the menstrual cycle (Figure 6K).

Hormonal contraceptive use provides a "virtual experiment" to confirm key findings

As we observed similar upregulation of key marker genes such as LRRC26 and TCF7 in women using hormonal contraceptives as we did in the luteal phase of the menstrual cycle (Figure 3G, Figure 4J), we performed "virtual experiments" to test the effects of hormone combinations and dynamics on downstream signaling responses by analyzing an additional three donors with hormonal contraceptive use at the time of surgery (Table S6). scRNAseq (Table S7) and clustering identified three major epithelial populations and four major stromal populations (Figure S7A-C) that directly corresponded to the seven cell types identified in our original five sequenced samples (Figure 1D-E).

Using the defined hormonal perturbations represented in this new dataset, we first investigated the HR+ cell population to identify the specific signaling pathways activated by combined estrogen/progesterone signaling (E/P) versus progesterone alone (P) (Figure 7A). To measure the transcriptional variation in HR+ cells between each sample, we quantified the EMD in PC space and identified two discrete groups; samples within the either the follicular or luteal phase of the

menstrual cycle clustered with each other, and samples from donors using hormonal contraception clustered with those in the luteal phase (Figure 7B). Interestingly, combined hormonal contraception was most similar to the late luteal phase (stage IV), whereas progestin-based contraception was most similar to the early luteal phase (stage III), suggesting that both estrogen and progesterone are required for the full luteal-phase transcriptional response. Supporting this, ER target genes such as AREG, TFF1, and TFF3 were specifically upregulated by combined E/P treatment, although other key downstream regulators such as TNFSF11 (RANKL), WNT4, and CXCL13 were upregulated in both hormone treatment groups to varying degrees (Figure 7C, Figure S7D). Surprisingly, we did not observe upregulation of genes such as VEGFA or ANGPTL4 in either hormonal contraceptive group, despite high expression of PR target genes such as SOX4 and KLF4 in the progestin-only contraceptive group similar to levels seen in the luteal-phase (Figure S7D). These data suggest that VEGFA expression and the pro-angiogenic response in the late luteal phase either depend on cycling rather than sustained hormone levels or on specific temporal ordering of ER/PR activation.

Based on these results, we next asked whether the transcriptional states observed in HR+ luminal cells following P-alone or combined E/P treatment led to different downstream paracrine signaling responses in the stroma. EMD and clustering in fibroblasts identified two groups, with samples from donors using hormonal contraception most similar to those in the luteal phase (Figure 7D). We predicted that, as in HR+ luminal cells, fibroblasts would require combined E/P for the full "luteal-phase" transcriptional response, including upregulation of ECM molecules and ECM remodeling proteins. Indeed, although many luteal-phase transcripts such as MMP14, IGFBP2, and DCN were highly upregulated in both hormone-treatment groups relative to the follicular phase, ECM proteins such as COL1A1/2, COL3A1, and FN1 were most highly induced in the combined E/P sample (Figure 7E, Figure S7E).

Finally, we took advantage of the different dynamics of serum hormone levels in donors using hormonal contraception to ask whether the "involution" transcriptional signature observed in HR- luminal cells during the early follicular phase was a consequence of hormone signaling dynamics or represented a homeostatic response. In contrast to HR+ cells and fibroblasts, HR- luminal cells from donors using hormonal contraception were most similar to those in the follicular phase rather than the luteal phase (Figure 7F). Consistent with this, markers of the "involution" cluster during the early follicular phase of the menstrual cycle such as TNFSF10 (TRAIL), CD14, and SAA2 (Figure 4G) were also upregulated in both P and E/P treatment groups relative to similar levels as the early follicular phase (Figure 7G, Figure S7F). Interestingly, we also observed expression of the WNT effector TCF7 in both hormone treatment groups at similar levels to that seen during the luteal phase (Figure S7F). Together, these results suggest that rather than serving as a direct response to changing hormone levels, the "involution" phenotype represents a homeostatic response in HR- luminal cells. Moreover, whereas "involution" and WNT activation are temporally separated during the menstrual cycle, our data suggests that this temporal regulation is lost when normal hormone dynamics are disrupted.

Discussion

In this study, we combined scRNAseq with immunostaining, flow cytometry, and "virtual experiments" to reveal changes in cell composition and cell state associated with the cycling of

hormones in the human breast. Our key insight was that, after accounting for the effect of parity, sample-to-sample variability primarily represented the physiological responses to hormone signaling in the breast, since samples were collected at different points in each woman's menstrual cycle. Indeed, we demonstrated that pregnancy history and menstrual cycle were the two major sources of variation between our samples, with prior pregnancy history leading to a change in epithelial cell proportions and menstrual cycle stage leading to changes in transcriptional state across all epithelial and stromal cell types.

Unbiased clustering uncovered a dramatic change in epithelial composition in women with prior history of pregnancy, characterized by an increased proportion of myoepithelial cells relative to luminal cells and of HR- luminal cells relative to HR+ luminal cells. Previous work has described two tumor-protective features of myoepithelial cells: they are highly resistant to malignant transformation (Lakhani and O'Hare, 2001) and also act as a natural barrier that prevents tumor cell invasion (Sternlicht et al., 1997). Thus, our data suggests that pregnancy protects against breast cancer risk both by decreasing the relative frequency of luminal cells—the tumor cell-of-origin for most breast cancer subtypes (Keller et al., 2012; Melchor et al., 2014; Molyneux et al., 2010)—and by suppressing progression to invasive carcinoma. Moreover, over 80% of all breast cancers express estrogen and/or progesterone receptors (Howlader et al., 2014), and epidemiological studies demonstrate that pregnancy specifically reduces the risk of HR+ breast cancer (Ma et al., 2006). In mice, early pregnancy leads to a lifelong decrease in the overall proportion of HR+ cells (Meier-Abt et al., 2014). Our findings suggest that a similar decrease in the proportion of HR+ cells occurs in the parous human breast. We propose that this decrease is a second mechanism that may contribute to the protective effect of pregnancy against breast cancer.

Interestingly, while it has been suggested that the protective effect of pregnancy is partly due to differentiation of stem and progenitor cells within the mammary epithelium (Choudhury et al., 2013; Meier-Abt et al., 2013), we find that parity led to a change in the proportions of cell types that already existed within the nulliparous breast rather than the emergence of new "differentiated" cell states. However, one outstanding question is whether the cellular transcriptional response to estrogen and progesterone is altered in parous versus nulliparous women. A previous study using bulk RNA sequencing of purified cell types suggested that, in HR- luminal cells, the transcriptional signatures of parous and nulliparous samples were more distinct in the luteal phase of the menstrual cycle than the follicular phase (Choudhury et al., 2013). Our data suggests that this difference may be partly caused by a decrease in the proportion of HR+ cells in the parous breast; if the magnitude of paracrine signaling scales with the proportion of HR+ cells, a reduction in HR+ cells following pregnancy would lead to a corresponding overall reduction in paracrine signaling downstream of hormone activation. However, we cannot rule out the possibility that pregnancy also leads to a change in the differentiation status of HR+ or HR- luminal cells that is only revealed in the luteal phase. Identifying such a scenario would require additional single-cell data from parous samples in the luteal phase of the menstrual cycle.

Second, we used menstrual cycle staging and transcriptional analysis to dissect the changes that occur in cell state across the menstrual cycle in all epithelial and stromal cell populations. Importantly, we found that the transcriptional state of samples clustered with menstrual cycle

stage but not with other factors such as age, BMI, or race. Subclustering and pseudotemporal analysis of HR+ cells across the menstrual cycle identified a bifurcation in cell state, in which a single population in the follicular phase split into two distinct states in the luteal phase. We find that both luteal phase cell states co-occur within the same region of the breast, suggesting that these two states do not reflect gross microenvironmental differences such as epithelial or vascular density over different regions of the breast. However, these bifurcating cell states may reflect more local changes in microenvironment. Two other possibilities are that: 1) bifurcation is driven by stochastic fluctuations in gene expression, or 2) it represents two cell states already present in the follicular phase—such as different estrogen receptor signaling states or cell cycle stages—that we lack the resolution to detect in our data. Additional scRNAseq studies at higher sequencing depth or using tissue explants to achieve finer-grained temporal resolution following hormone treatment will be required to address this question.

Sub-clustering analysis also uncovered changes in cell state across all hormone-insensitive cell types—epithelial and stromal—downstream of the paracrine signaling cascade originating in HR+ cells. Strikingly, many of these changes closely mimic those seen during the pregnancy/lactation/involution cycle that have been linked with a transient increased breast cancer risk following pregnancy (Lyons et al., 2011; O'Brien et al., 2010; Schedin et al., 2007). Similar to pregnancy and lactation, high levels of progesterone in the luteal phase promote epithelial proliferation (Anderson et al., 1982; Söderqvist et al., 1997). We found that HR+ luminal cells in the luteal phase split into two distinct paracrine signaling states: one cell state signals partly via proangiogenic and hypoxia-induced factors, likely contributing to remodeling of the stroma, while a second cell state signals via RANK ligand and WNT to myoepithelial and HR- luminal epithelial cells. The first subpopulation has not been previously characterized, while the second is consistent with previous studies demonstrating that RANK and WNT control progesterone-mediated epithelial proliferation (Joshi et al., 2015). These latter signals may also be permissive of growth in cells with preexisting oncogenic mutations. In contrast, the fraction of apoptotic cells in the epithelium peaks between the late luteal and early follicular phase (Anderson et al., 1982). Consistent with this, we identified a previously undescribed subpopulation of HR- cells in the early follicular phase with a transcriptional signature closely matching that described for post-lactational involution (Clarkson et al., 2004; Stein et al., 2009), including upregulation of immune mediators and phagocytic receptors.

In the stroma, we uncovered tumor promoting microenvironmental changes in all cell types that parallel the cellular changes seen during pregnancy and involution (Lyons et al., 2011; O'Brien et al., 2010; Schedin et al., 2007). As hormone levels increase, we found induction of pathways involved in angiogenesis and blood vessel remodeling in endothelial cells, as well as hallmarks of ECM deposition and remodeling in fibroblasts. Notably, we observed the concurrent upregulation of a pro-angiogenic and hypoxic gene signatures in multiple epithelial and stromal cell types. A previous study identified these same pathways as highly enriched following involution in the mouse mammary gland. More importantly from the perspective of breast cancer risk, this "hypoxia/pro-angiogenic" signature identified breast cancers with increased metastatic activity (Stein et al., 2009), suggesting that these pathways support tumor cell invasion and metastasis. Finally, we described a switch between a pro-inflammatory M1-like macrophage polarization in the luteal phase and a tissue-remodeling M2-like macrophage polarization in the follicular phase. M2-like macrophages have previously been shown to promote cancer

progression (Mantovani et al., 2002). Together, these data suggest that some of the same mechanisms underlie both the increased short-term breast cancer risk following pregnancy and the lifetime increased risk due to menstrual cycle number. Moreover, in samples from donors using hormonal contraception, we observed stromal changes that paralleled those seen during the luteal phase of the menstrual cycle, as well as evidence of an "involution" transcriptional signature in HR- cells similar to the early follicular phase. Thus, we speculate that similar signaling pathways underlie the increased risk of breast cancer that has been observed in women using hormonal contraception (Mørch et al., 2018).

In summary, these results provide a comprehensive, systems-level view of the cellular and transcriptional changes that control normal breast development and breast cancer risk in response to cycling hormones. This single-cell analysis establishes a link between hormone cycling during pregnancy or the menstrual cycle and a variety of well-established pro- and anti-tumorigenic cellular signatures: we identify tumor-protective changes in epithelial cell proportion with pregnancy and tumor-promoting changes in cell state across the menstrual cycle that are similar to changes seen during involution. As the breast is one of the only human organs that undergoes repeated cycles of morphogenesis and involution, this study serves as a roadmap to the cell state changes associated with the dynamic human breast. Moreover, it provides a foundation for similar system-level studies dissecting the how the paracrine communication networks downstream of hormone signaling are altered during HR(+) breast cancer progression. A better understanding of cellular and molecular response to hormone receptor activation will aid in identifying women at higher risk for breast cancer and may inform new strategies for cancer prevention.

Methods

Tissue samples and preparation

Reduction mammoplasty tissue samples were obtained from the Cooperative Human Tissue Network (Nashville, TN) and the Kaiser Foundation Research Institute (Oakland, CA). Tissues were obtained as de-identified samples and all subjects provided written informed consent. When possible, medical reports were obtained with personally identifiable information redacted. Use of the breast tissues to conduct the studies described above were approved by the UCSF Committee on Human Research under Institutional Review Board protocol No. 158396. A portion of each sample was fixed in formalin and paraffin-embedded using standard procedures. The remainder was dissociated mechanically and enzymatically to obtain epithelial-enriched organoids. Tissue was minced, followed by enzymatic dissociation with 200 U/mL collagenase type III (Worthington CLS-3) and 100 U/mL hyaluronidase (Sigma H3506) in RPMI 1640 with HEPES (Cellgro 10-041-CV) plus 10% (v/v) dialyzed FBS, penicillin, streptomycin, amphotericin B (Lonza 17-836E), and gentamicin (Lonza 17-518) at 37 C for 16h. This cell suspension was centrifuged at 400 x g for 10 min and resuspended in RPMI 1640 plus 10% FBS. Organoids enriched for epithelial cells and associated stroma were collected after serial filtration through 150 µm and 40 µm nylon mesh strainers. The final filtrate contained stromal cells consisting primarily of fibroblasts, endothelial cells, and immune cells. Following centrifugation, epithelial organoids and filtrate were frozen and maintained at -180 °C until use.

Dissociation to single cells and sorting for scRNA-seq

The day of sorting, epithelial organoids from the 150 µm fraction were thawed and digested to single cells by trituration in 0.05% trypsin for 2 min, followed by trituration in 5 U/mL dispase (Stem Cell Technologies 07913) plus 1 mg/mL DNase I (Stem Cell Technologies 07900) for 2 min. Single-cell suspensions were resuspended in HBSS supplemented with 2% FBS, filtered through a 40 µm cell strainer, and pelleted at 400 x g for 5 minutes. The pellets were resuspended in 10 mL of complete mammary epithelial growth medium with 2% v/v FBS without GA-1000 (MEGM) (Lonza CC-3150). Cells were incubated in a 37 °C for 2 hours, rotating on a hula mixer, to regenerate surface antigens. Cells were pelleted at 400 x g for 5 minutes and resuspended in phosphate buffered saline supplemented with 1% BSA at a concentration of 1 million cells per 100 µL, and incubated with primary antibodies. Cells were stained with Alexa 488-conjugated anti-CD49f to isolate myoepithelial cells, PE-conjugated anti-EpCAM to isolate luminal epithelial cells, and biotinylated antibodies for lineage markers CD2, CD3, CD16, CD64, CD31, and CD45 to remove hematopoietic (CD16/CD64-positive), endothelial (CD31-positive), and leukocytic (CD2/CD3/CD45-positive) lineage cells by negative selection (Lin-). Sequential incubation with primary antibodies was performed for 15 min at room temperature in PBS with 1% BSA, and cells were washed with PBS with 1% BSA. Biotinylated primary antibodies were detected with a streptavidin-Brilliant Violet 785 conjugate. After incubation, cells were washed once in PBS with 1% BSA and resuspended in PBS with 2% BSA and 1 ug/mL DAPI for live/dead discrimination. Cell sorting was performed on a FACSAria II cell sorter. 5,000-10,000 unsorted (DAPI-), luminal (DAPI-/Lin-/CD49f-/EpCAMhigh), or myoepithelial (DAPI-/Lin-/CD49f+/EpCAMlow) cells were collected for each sample and resuspended in PBS plus 1% BSA at a concentration of 1000 cells/µL.

Antibodies and dilutions used (µL/million cells): FITC-EpCAM (1.5 µL; BD 550257, clone AD2), APC-CD49f (4 µL; Stem Cell Technologies 10109, clone VU1D9), Biotin-CD2 (8 µL; Biolegend 313636, clone GoH3), Biotin-CD3 (8 µL; BD 55325, clone RPA-2.10), Biotin-CD16 (8 µL; BD 55338, clone HIT3a), Biotin-CD64 (8 µL; BD 555526, clone 10.1), Biotin-CD31 (4 µL; Invitrogen MHCD31154, clone MBC78.2), Biotin-CD45 (1 µL; Biolegend 304004, clone HI30), BV785-Streptavidin (1 µL; Biolegend 405249).

scRNAseq library preparation

cDNA libraries were prepared using the 10X Genomics Single Cell V2 (10X Genomics, 2017) standard workflow (CG00052 Single Cell 3' Reagent Kit v2: User Guide Rev B). Library concentrations were quantified using high sensitivity DNA Bioanalyzer chips (Agilent, 5067-4626), the Illumina Library Quantification Kit (Kapa Biosystems KK4824), and Qubit dsDNA HS Assay Kit (Thermo Fisher Q32851). Each library was separately sequenced on a lane of a HiSeq4500 for an average of ~100,000 reads/cell.

scRNAseq data processing with the Cell Ranger package

Cell Ranger software version 1.31 was used to align sequences, filter data and count unique molecular identifiers (UMIs). Data were mapped to the human reference genome GRCh37. The resulting sequencing statistics are summarized in Table S2 and Table S7. Data from multiple samples were aggregated into a single data set using the Cell Ranger Aggr pipeline, which down-samples the read depth of different lanes to normalize across the data set. We performed three sets of data aggregation: the 5 samples listed in Table S2 were aggregated for initial cell type clustering and analyses of changes due to pregnancy and menstrual cycle (Figures 1-6), the 3

samples listed in Table S7 were aggregated for cell type identification in donors using hormonal birth control (Figure S7), and all 8 samples were aggregated for analysis of relative expression of specific marker genes with different hormone treatments relative to different stages of the menstrual cycle (Figure 7).

Quality control and cell type identification using Seurat
Cell type identification was performed using the Seurat package (version 2.3.1) in R. Aggregated data was filtered to remove cells that had fewer than 200 genes and genes that appeared in fewer than 3 cells. Cells with a Z score of 4 or greater for the total number of genes expressed were presumed to be doublets and removed from analysis. Cells with a Z score of 3 or greater for percent of mitochondrial genes were presumed to be low diversity and removed from analysis. This filtering removed 916 cells from analysis of the 5 aggregated samples (Table S2) to give 23,708 genes across 42.103 cells and 170 cells from analysis of the 3 aggregated hormonal contraception samples (Table S7) to give 21,527 genes across 4,419 cells. The remaining cells were log transformed and scaled to a total of 1e4 molecules per cell, and variation due to the number of UMI and percent of mitochondrial genes was regressed out.

Variable genes were defined as genes with an average expression between 0.0125 and 3 and a z-score of at least 0.5. For initial cell type identification, batch effects were corrected by identifying genes with an AUC > 0.6 for an individual sample and removing these from the list of variable genes. We then performed PCA on the resulting list of variable genes. Statistically significant PCs as determined by visual inspection of elbow plots were used as an input for TSNE visualization. Finally, we performed k nearest neighbor (KNN) modularity optimization-based clustering to identify cell types using Seurat's FindClusters function.

Menstrual cycle scoring
Gene sets representing differentially expressed transcripts the follicular or luteal phase of the menstrual cycle were taken from Pardo et al., and the top 20 follicular- or luteal-phase genes were identified after excluding genes not expressed in our scRNAseq dataset. As these gene sets represented differentially expressed transcripts in microdissected epithelium, we randomly selected equal total numbers of luminal epithelial cells (C2 and C3) from the unsorted and luminal sort gates for each sample. We then calculated the averaged log-normalized expression of genes in each set to define follicular- or luteal-phase specific gene scores. These scores were scaled by their root mean square and normalized to their maximum value to give a follicular or luteal score ranging from 0-1 for each cell. Finally, an average menstrual cycle score was calculated for each sample by subtracting the follicular score from the luteal score for each cell and plotting the mean across all cells in a sample.

Measuring transcriptional variability between samples with Earth Mover's Distance
To measure differences in transcriptional state due to menstrual cycle stage without confounding effects of parity on cell proportions, we sampled equal numbers of each of the most abundant cell types (C1-C6) for each sample. We then performed PCA on the subsampled matrix using the most variable genes identified as described above (without correcting for batch effects). We chose this approach rather than using the PCs identified in Seurat, as the PCs in Seurat were calculated on the entire dataset containing variable numbers of each cell type, which varied based on parity and amount of associated stroma. Following PCA, we used the Munkres

assignment algorithm in Matlab (Munkres Assignment Algorithm version 1.0.0.0, Yi Cao) to quantify the minimum cost of moving one sample's distribution across in PC space onto another sample's distribution, in a measure known as Earth Mover's Distance (EMD). We chose this approach rather than distance-based or other similarity metrics since EMD measures the transcriptional variation between entire cell populations (containing multiple cell types) rather than between individual cells. Finally, hierarchical clustering of the EMD between each sample pair was calculated using complete linkage to identify samples that were more similar to each other in PC space. To measure differences in transcriptional state due to hormonal contraceptives, we performed similar analyses with the following modifications. First, we sampled equal numbers of each of the analyzed cell types (e.g. HR+ luminal cells, fibroblasts, etc.) rather than multiple cell types. Second, we performed the Munkres assignment algorithm on the first two PCs identified by PCA of each individual cell type in Seurat.

Pseudotime analysis with Monocle 2

Pseudotime trajectories were constructed for HR+ luminal cells and myoepithelial cells using the Monocle 2 R package (version 2.6.4). For each cell type, we performed PCA on genes expressed in at least 5% of the total cells. Statistically significant PCs as determined by visual inspection of elbow plots were used as an input for TSNE analysis. We next used an unsupervised procedure called dpFeature to select genes that varied between clusters as identified by graph-based density peak clustering in TSNE-space. We selected the 1,500 most significantly differentially expressed genes and used these genes to perform reverse graph embedding dimensionality reduction, followed by manifold learning to fit a trajectory to the reduced dimension data and infer a pseudotime value for each cell. Cells in the follicular phase were chosen as the root of each trajectory. For HR+ luminal cells, we additionally used this ordering to identify genes that varied across pseudotime or along each branch and fit a smooth spline to each gene's expression across pseudotime, using the differentialGeneTest or BEAM functions, respectively. We then clustered genes with similar variation across pseudotime and visualized these changes using the plot_pseudotime_heatmap and plot_genes_branched_heatmap functions.

Cell state identification using non-negative matrix factorization (NMF)

For each cell type, we selected variable genes as described above (without batch correction). We used the NMF (version 0.21.0) package in R to perform sub-clustering by non-negative matrix factorization. NMF attempts to find a mathematical approximation for $A \approx WH$, where, for our dataset, A represents a matrix of n variable genes by m cells, W represents an n x k matrix of gene loadings for each cluster k, and H represents a k x m matrix of cell loadings for each cluster k. To estimate the optimum number of clusters, k, we randomly sampled 1000 cells for each cell type, initialized W and H using a random seed, and performed 30 NMF runs to obtain consensus clustering results. We calculated the cophenetic correlation coefficient and dispersion for each consensus clustering matrix and chose the optimum value for k as proposed in Brunet et al. (Brunet et al., 2004). Using this optimum value of k, we then re-ran NMF on the full set of cells for each cell type, using non-negative double singular value decomposition (nnsvd) to approximate appropriate initial values of W and H. For each cell type, we also performed PCA in Seurat on the most variable genes as described above, and plotted the NMF results in PC-space to visually confirm clustering results.

RNA velocity estimation with Velocyto

RNA velocity estimation was performed using the Velocyto (version 0.5) R package as described in La Manno et al. Spliced and unspliced matrices were prepared using the Velocyto Python command line interface. Genes were filtered to remove spliced transcripts expressed in fewer than 20 cells, with fewer than 30 total counts, or with greater than 0.5 average counts and unspliced transcripts expressed in fewer than 20 cells, with fewer than 20 total counts, or with greater than 0.05 average counts. RNA velocities were estimated based on calculating a cell-cell distance from the correlation of each cell in PC space, nearest-cell pooling (k=25) and a fit quantile of 0.02. These velocities were visualized by projecting velocity vector fields into PC space using Gaussian smoothing on a regular grid.

Gene set enrichment analysis

To identify gene sets upregulated in each phase of the menstrual cycle, we performed marker analysis on the cell states identified by NMF analysis, using the likelihood-ratio test for single gene expression as part of the Seurat package (McDavid et al., 2013). We selected the set of marker genes overexpressed by at least 1.5-fold in each cell state and used the Broad's gene set enrichment analysis tool (http://software.broadinstitute.org/gsea/msigdb/annotate.jsp) to compute overlaps with hallmark, KEGG, and gene ontology (GO) gene sets (Subramanian et al., 2005). A corrected p-value of 0.05 was used as a cutoff to identify significant enrichment of a gene set, and the top 10 most significantly-enriched gene sets were plotted for each cell state.

Fluorescent Immunohistochemistry

For immunofluorescent staining, formalin-fixed paraffin-embedded tissue sections were deparaffinized and rehydrated using standard methods. Endogenous peroxides were blocked using 3% hydrogen peroxide in PBS, and antigen retrieval was performed in 0.1 M citrate buffer pH 6.0. Sections were blocked for 5 minutes at room temperature using Lab Vision Ultra-V block (Thermo TA-125-UB) and rinsed with TNT wash buffer (1X Tris-buffered saline with 5 mM Tris-HCl and 0.5% TWEEN-20). Primary antibody incubations were performed for 1 hour at room temperature or overnight at 4°C. Sections were washed three times for 5 min each with TNT wash buffer, incubated with Lab Vision UltraVision LP Detection System HRP Polymer (Thermo Fisher TL-060-HL) for 15 minutes at room temperature, washed, and incubated with one of three colors of TSA amplification reagent at a 1:50 dilution. After tyramide signal amplification, antibody complexes were removed by boiling in citrate buffer, followed by blocking and incubation with additional primary antibodies as above. Finally, sections were rinsed with deionized water and mounted using Vectashield HardSet Mounting Media with DAPI (Vector H-1400). Immunfluorescence was analyzed by spinning disk confocal microscopy using a Zeiss Cell Observer Z1 equipped with a Yokagawa spinning disk and running Zeiss Zen Software.

Antibodies, TSA reagents, and dilutions used: p63 (1:2000; CST 13109, clone D2K8X), KRT7 (1:4000; Abcam AB68459, clone EPR1619Y), KRT23 (1:2000; Abcam AB156569, clone EPR10943), ER (1:4000; Thermo RMM-9101-S, clone SP1), LRRC26 (1:2000; Thermo PA5-63285), TCF7 (1:2000; CST 2203, clone C63D9), PR (1:3000; CST 8757, clone D8Q2J), P4HA1 (1:9000; Thermo PA5-55353), FITC-TSA (2 min; Perkin Elmer NEL701A001KT), Cy3-TSA (3 min; Perkin Elmer NEL744001KT), Cy5-TSA (7 min; Perkin Elmer NEL745E001KT).

RNA FISH analysis of ESR1 transcripts

Combined RNA FISH and immunofluorescence analysis of estrogen receptor transcript (RNAscope Probe Hs-ESR1; ACD 310301) and protein (anti-ER; Thermo RMM-9101-S, clone SP1) was performed using the RNAscope in situ hybridization kit (RNAscope Multiplex Fluorescent Reagent Kit V2, ACD 323100) according to the manufacturer's instructions and fluorescent immunohistochemistry protocol outlined above with the following modifications. Immunostaining for ER was performed prior to *in situ* hybridization, using the hydrogen peroxide and antigen retrieval solutions supplied with the RNAscope kit and the mildest recommended conditions. After ER immunostaining and tyramide signal amplification, *in situ* hybridization for ESR1 was performed according to the manufacturer's instructions, followed by immunostaining for KRT7 as described above. For all RNA FISH experiments, we used positive (PPIB) and negative controls (DAPB) to verify staining conditions and probe specificity.

Flow cytometry analysis of myoepithelial and immune cell populations
Flow cytometry analysis of myoepithelial cell populations was performed as described above (*Dissociation to single cells and sorting for scRNA-seq).* Flow cytometry analysis of immune cell populations was performed as described above except that the filtrate fraction, enriched for stromal fibroblasts and immune cells, was used rather than the 150 µm epithelial-enriched organoid fraction. Cells were stained with Pacific Blue-conjugated anti-CD45 to identify immune cells, FITC-conjugated anti-CD68 to identify macrophages, and PE-conjugated anti-CD163 to identify M2-like polarized macrophages.

Antibodies and dilutions used (µL/million cells): EpCAM, CD49f, and lineage negative-selection antibodies and dilutions for myoepithelial cell FACS analysis are described above. PB-CD45 (5 µL; BIolegend 304012, clone HI30), FITC-CD68 (5 µL; Biolegend 333805, clone Y1/82A), PE-CD163 (5 µL; Biolegend 333605, clone GHI/61)

Data availability
Raw gene expression and barcode count matrices will be uploaded to the Gene Expression Omnibus.

Acknowledgments
We thank Drs. Tom Norman and Jonathan Weissman for technical support and for generously providing access to equipment and computing resources. Sequencing was performed in the Center for Advanced Technology at UCSF. This research was supported in part by grants from the Department of Defense Breast Cancer Research Program (W81XWH-10-1-1023 and W81XWH-13-1-0221), NIH (U01CA199315 and DP2 HD080351-01), the NSF (MCB-1330864), and the UCSF Center for Cellular Construction (DBI-1548297), an NSF Science and Technology Center. Z.J.G is a Chan-Zuckerberg BioHub Investigator. L.M.M is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2239-15).

Author Contributions
Conceptualization, L.M.M., R.J.W., and Z.J.G.; Methodology, L.M.M., R.J.W., M.T., and Z.J.G.; Software L.M.M., R.J.W., and C.S.M.; Investigation, L.M.M., R.J.W., J.C., and A.D.B.; Resources, T.T. and Z.J.G.; Writing – Original Draft, L.M.M. and Z.J.G.; Writing – Review & Editing, L.M.M., R.J.W., A.D.B., M.T., T.T. and Z.J.G.; Visualization, L.M.M.; Supervision, T.A.D., T.T., and Z.J.G.

References

Anderson, T.J., Ferguson, D.J., and Raab, G.M. (1982). Cell turnover in the “resting” human breast: influence of parity, contraceptive pill, age and laterality. Br. J. Cancer *46*, 376–382.

Battersby, S., Robertson, B.J., Anderson, T.J., King, R.J., and McPherson, K. (1992). Influence of menstrual cycle, parity and oral contraceptive use on steroid hormone receptors in normal breast. Br. J. Cancer *65*, 601–607.

Bouras, T., Pal, B., Vaillant, F., Harburg, G., Asselin-Labat, M.-L., Oakes, S.R., Lindeman, G.J., and Visvader, J.E. (2008). Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. Cell Stem Cell *3*, 429–441.

Britt, K., Ashworth, A., and Smalley, M. (2007). Pregnancy and the risk of breast cancer. Endocr Relat Cancer *14*, 907–933.

Brunet, J.-P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proceedings of the National Academy of Sciences *101*, 4164–4169.

Choudhury, S., Almendro, V., Merino, V.F., Wu, Z., Maruyama, R., Su, Y., Martins, F.C., Fackler, M.J., Bessarabova, M., Kowalczyk, A., et al. (2013). Molecular Profiling of Human Mammary Gland Links Breast Cancer Risk to a p27+ Cell Population with Progenitor Characteristics. Stem Cell *13*, 117–130.

Ciarloni, L., Mallepell, S., and Brisken, C. (2007). Amphiregulin is an essential mediator of estrogen receptor alpha function in mammary gland development. Proceedings of the National Academy of Sciences *104*, 5455–5460.

Clarke, R.B., Howell, A., Potten, C.S., and Anderson, E. (1997). Dissociation between steroid receptor expression and cell proliferation in the human breast. Cancer Research *57*, 4987–4991.

Clarkson, R.W.E., Wayland, M.T., Lee, J., Freeman, T., and Watson, C.J. (2004). Gene expression profiling of mammary gland development reveals putative roles for death receptors and immune mediators in post-lactational regression. Breast Cancer Res *6*, R92–R109.

Collaborative Group on Hormonal Factors in Breast Cancer (2012). Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. The Lancet Oncology *13*, 1141–1151.

Dabrosin, C. (2003). Variability of vascular endothelial growth factor in normal human breast tissue in vivo during the menstrual cycle. J. Clin. Endocrinol. Metab. *88*, 2695–2698.

Dontu, G., and Ince, T.A. (2015). Of Mice and Women: A Comparative Tissue Biology Perspective of Breast Stem Cells and Differentiation. J Mammary Gland Biol Neoplasia *20*, 51–62.

Ferguson, J.E., Schor, A.M., Howell, A., and Ferguson, M.W. (1992). Changes in the

extracellular matrix of the normal human breast during the menstrual cycle. Cell Tissue Res. *268*, 167–177.

Fornetti, J., Flanders, K.C., Henson, P.M., Tan, A.-C., Borges, V.F., and Schedin, P. (2016). Mammary epithelial cell phagocytosis downstream of TGF-β3 is characterized by adherens junction reorganization. Cell Death Differ. *23*, 185–196.

Fridriksdottir, A.J., Kim, J., Villadsen, R., Klitgaard, M.C., Hopkinson, B.M., Petersen, O.W., and Rønnov-Jessen, L. (2015). Propagation of oestrogen receptor-positive and oestrogen-responsive normal human breast cells in culture. Nature Communications *6*, 8786.

Graham, J.D., Hunt, S.M., Tran, N., and Clarke, C.L. (1999). Regulation of the expression and activity by progestins of a member of the SOX gene family of transcriptional modulators. J. Mol. Endocrinol. *22*, 295–304.

Hallberg, G., Andersson, E., Naessén, T., and Ordeberg, G.E. (2010). The expression of syndecan-1, syndecan-4 and decorin in healthy human breast tissue during the menstrual cycle. Reprod. Biol. Endocrinol. *8*, 35.

Haricharan, S., Dong, J., Hein, S., Reddy, J.P., Du, Z., Toneff, M., Holloway, K., Hilsenbeck, S.G., Huang, S., Atkinson, R., et al. (2013). Mechanism and preclinical prevention of increased breast cancer risk caused by pregnancy. eLife *2*, 4984–24.

Hirakawa, S., Hong, Y.-K., Harvey, N., Schacht, V., Matsuda, K., Libermann, T., and Detmar, M. (2003). Identification of Vascular Lineage-Specific Genes by Transcriptional Profiling of Isolated Blood Vascular and Lymphatic Endothelial Cells. Am. J. Pathol. *162*, 575–586.

Howlader, N., Altekruse, S.F., Li, C.I., Chen, V.W., Clarke, C.A., Ries, L.A.G., and Cronin, K.A. (2014). US Incidence of Breast Cancer Subtypes Defined by Joint Hormone Receptor and HER2 Status. JNCI J Natl Cancer Inst *106*, 747.

Hyder, S.M., Nawaz, Z., Chiappetta, C., and Stancel, G.M. (2000). Identification of functional estrogen response elements in the gene coding for the potent angiogenic factor vascular endothelial growth factor. Cancer Research *60*, 3183–3190.

Jakobsson, L., Franco, C.A., Bentley, K., Collins, R.T., Ponsioen, B., Aspalter, I.M., Rosewell, I., Busse, M., Thurston, G., Medvinsky, A., et al. (2010). Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting. Nat Cell Biol *12*, 943–953.

Joshi, P.A., Waterhouse, P.D., Kannan, N., Narala, S., Fang, H., Di Grappa, M.A., Jackson, H.W., Penninger, J.M., Eaves, C., and Khokha, R. (2015). RANK Signaling Amplifies WNT-Responsive Mammary Progenitors through R-SPONDIN1. Stemcr *5*, 31–44.

Keller, P.J., Arendt, L.M., Skibinski, A., Logvinenko, T., Klebba, I., Dong, S., Smith, A.E., Prat, A., Perou, C.M., Gilmore, H., et al. (2012). Defining the cellular precursors to human breast cancer. Proc. Natl. Acad. Sci. U.S.a. *109*, 2772–2777.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber,

K., Kastriti, M.E., Lonnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature Publishing Group *17*, 529.

Lakhani, S.R., and O'Hare, M.J. (2001). The mammary myoepithelial cell--Cinderella or ugly sister? Breast Cancer Res *3*, 1–4.

Lambe, M., Hsieh, C., Trichopoulos, D., Ekbom, A., Pavia, M., and Adami, H.O. (1994). Transient increase in the risk of breast cancer after giving birth. N. Engl. J. Med. *331*, 5–9.

Lim, E., Vaillant, F., Di Wu, Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.-L., Gyorki, D.E., Ward, T., Partanen, A., et al. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nature Publishing Group 1–9.

Lim, E., Wu, D., Pal, B., Bouras, T., Asselin-Labat, M.-L., Vaillant, F., Yagita, H., Lindeman, G.J., Smyth, G.K., and Visvader, J.E. (2010). Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. Breast Cancer Res *12*, R21.

Longacre, T.A., and Bartow, S.A. (1986). A correlative morphologic study of human breast and endometrium in the menstrual cycle. Am. J. Surg. Pathol. *10*, 382–393.

Lyons, T.R., O'Brien, J., Borges, V.F., Conklin, M.W., Keely, P.J., Eliceiri, K.W., Marusyk, A., Tan, A.-C., and Schedin, P. (2011). Postpartum mammary gland involution drives progression of ductal carcinoma in situ through collagen and COX-2. Nat Med *17*, 1109–1115.

Ma, H., Bernstein, L., Pike, M.C., and Ursin, G. (2006). Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. Breast Cancer Res *8*, 3232.

Mantovani, A., Sozzani, S., Locati, M., Allavena, P., and Sica, A. (2002). Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. Trends Immunol. *23*, 549–555.

May, F.E.B., and Westley, B.R. (2015). TFF3 is a valuable predictive biomarker of endocrine response in metastatic breast cancer. Endocr Relat Cancer *22*, 465–479.

McDavid, A., Finak, G., Chattopadyay, P.K., Dominguez, M., Lamoreaux, L., Ma, S.S., Roederer, M., and Gottardo, R. (2013). Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. Bioinformatics *29*, 461–467.

Meier-Abt, F., Brinkhaus, H., and Bentires-Alj, M. (2014). Early but not late pregnancy induces lifelong reductions in the proportion of mammary progesterone sensing cells and epithelial Wnt signaling. Breast Cancer Res *16*, 209.

Meier-Abt, F., Milani, E., Roloff, T., Brinkhaus, H., Duss, S., Meyer, D.S., Klebba, I., Balwierz, P.J., van Nimwegen, E., and Bentires-Alj, M. (2013). Parity induces differentiation and reduces Wnt/Notch signaling ratio and proliferation potential of basal stem/progenitor cells isolated from

mouse mammary epithelium. Breast Cancer Res *15*, R36.

Melchor, L., Molyneux, G., Mackay, A., Magnay, F.-A., Atienza, M., Kendrick, H., Nava-Rodrigues, D., López-García, M.Á., Milanezi, F., Greenow, K., et al. (2014). Identification of cellular and genetic drivers of breast cancer heterogeneity in genetically engineered mouse tumour models. The Journal of Pathology *233*, 124–137.

Métivier, R., Penot, G., Hübner, M.R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. Cell *115*, 751–763.

Molyneux, G., Geyer, F.C., Magnay, F.-A., McCarthy, A., Kendrick, H., Natrajan, R., Mackay, A., Grigoriadis, A., Tutt, A., Ashworth, A., et al. (2010). BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell Stem Cell *7*, 403–417.

Monks, J., Smith-Steinhart, C., Kruk, E.R., Fadok, V.A., and Henson, P.M. (2008). Epithelial cells remove apoptotic epithelial cells during post-lactation involution of the mouse mammary gland. Biol. Reprod. *78*, 586–594.

Mueller, S.O., Clark, J.A., Myers, P.H., and Korach, K.S. (2002). Mammary gland development in adult mice requires epithelial and stromal estrogen receptor alpha. Endocrinology *143*, 2357–2365.

Muenst, S., Mechera, R., Däster, S., Piscuoglio, S., Ng, C.K.Y., Meier-Abt, F., Weber, W.P., and Soysal, S.D. (2017). Pregnancy at early age is associated with a reduction of progesterone-responsive cells and epithelial Wnt signaling in human breast tissue. Oncotarget *8*, 22353–22360.

Mørch, L.S., Hannaford, P.C., and Lidegaard, Ø. (2018). Contemporary Hormonal Contraception and the Risk of Breast Cancer. N. Engl. J. Med. *378*, 1265–1266.

Nguyen, Q.H., Pervolarakis, N., Blake, K., Ma, D., Davis, R.T., James, N., Phung, A.T., Willey, E., Kumar, R., Jabart, E., et al. (2018). Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. Nature Communications *9*, 2028.

Niemelä, H., Elima, K., Henttinen, T., Irjala, H., Salmi, M., and Jalkanen, S. (2005). Molecular identification of PAL-E, a widely used endothelial-cell marker. Blood *106*, 3405–3409.

O'Brien, J., Lyons, T., Monks, J., Lucia, M.S., Wilson, R.S., Hines, L., Man, Y.-G., Borges, V., and Schedin, P. (2010). Alternatively activated macrophages and collagen remodeling characterize the postpartum involuting mammary gland across species. Am. J. Pathol. *176*, 1241–1255.

Palmieri, C., Saji, S., Sakaguchi, H., Cheng, G., Sunters, A., O'Hare, M.J., Warner, M., Gustafsson, J.-A., Coombes, R.C., and Lam, E.W.-F. (2004). The expression of oestrogen receptor (ER)-beta and its variants, but not ERalpha, in adult human mammary fibroblasts. J. Mol. Endocrinol. *33*, 35–50.

Pardo, I., Lillemoe, H.A., Blosser, R.J., Choi, M., Sauder, C.A.M., Doxey, D.K., Mathieson, T., Hancock, B.A., Baptiste, D., Atale, R., et al. (2014). Next-generation transcriptome sequencing of the premenopausal breast epithelium using specimens from a normal human breast tissue bank. Breast Cancer Res *16*, R26.

Parmar, H., and Cunha, G.R. (2004). Epithelial-stromal interactions in the mouse and human mammary gland in vivo. Endocr Relat Cancer *11*, 437–458.

Peri, S., de Cicco, R.L., Santucci-Pereira, J., Slifker, M., Ross, E.A., Russo, I.H., Russo, P.A., Arslan, A.A., Belitskaya-Lévy, I., Zeleniuch-Jacquotte, A., et al. (2012). Defining the genomic signature of the parous breast. BMC Med Genomics *5*, 46.

Petz, L.N., and Nardulli, A.M. (2000). Sp1 binding sites and an estrogen response element half-site are involved in regulation of the human progesterone receptor A promoter. Mol. Endocrinol. *14*, 972–985.

Ramakrishnan, R., Khan, S.A., and Badve, S. (2002). Morphological changes in breast tissue with menstrual cycle. Mod. Pathol. *15*, 1348–1356.

Rio, M.C., Bellocq, J.P., Gairard, B., Rasmussen, U.B., Krust, A., Koehl, C., Calderoli, H., Schiff, V., Renaud, R., and Chambon, P. (1987). Specific expression of the pS2 gene in subclasses of breast cancers in comparison with expression of the estrogen and progesterone receptors and the oncogene ERBB2. Proceedings of the National Academy of Sciences *84*, 9243–9247.

Russo, J., Rivera, R., and Russo, I.H. (1992). Influence of age and parity on the development of the human breast. Breast Cancer Res. Treat. *23*, 211–218.

Schaadt, N.S., Alfonso, J.C.L., Schönmeyer, R., Grote, A., Forestier, G., Wemmert, C., Krönke, N., Stoeckelhuber, M., Kreipe, H.H., Hatzikirou, H., et al. (2017). Image analysis of immune cell patterns in the human mammary gland during the menstrual cycle refines lymphocytic lobulitis. Breast Cancer Res. Treat. *164*, 305–315.

Schedin, P., O'Brien, J., Rudolph, M., Stein, T., and Borges, V. (2007). Microenvironment of the Involuting Mammary Gland Mediates Mammary Cancer Progression. J Mammary Gland Biol Neoplasia *12*, 71–82.

Schernhammer, E.S., Sperati, F., Razavi, P., Agnoli, C., Sieri, S., Berrino, F., Krogh, V., Abbagnato, C., Grioni, S., Blandino, G., et al. (2013). Endogenous sex steroids in premenopausal women and risk of breast cancer: the ORDET cohort. Breast Cancer Res *15*, R46.

Seong, J., Kim, N.-S., Kim, J.-A., Lee, W., Seo, J.-Y., Yum, M.K., Kim, J.-H., Park, I., Kang, J.-S., Bae, S.-H., et al. (2018). Side branching and luminal lineage commitment by ID2 in developing mammary glands. Development dev.165258.

Shimizu, Y., Takeuchi, T., Mita, S., Notsu, T., Mizuguchi, K., and Kyo, S. (2010). Krüppel-like factor 4 mediates anti-proliferative effects of progesterone with G_0/G_1 arrest in human endometrial epithelial cells. J. Endocrinol. Invest. *33*, 745–750.

Söderqvist, G., Isaksson, E., Schoultz, von, B., Carlström, K., Tani, E., and Skoog, L. (1997). Proliferation of breast epithelial cells in healthy women during the menstrual cycle. Am. J. Obstet. Gynecol. *176*, 123–128.

Stein, T., Morris, J.S., Davies, C.R., Weber-Hall, S.J., Duffy, M.-A., Heath, V.J., Bell, A.K., Ferrier, R.K., Sandilands, G.P., and Gusterson, B.A. (2004). Involution of the mouse mammary gland is associated with an immune cascade and an acute-phase response, involving LBP, CD14 and STAT3. Breast Cancer Res *6*, R75–R91.

Stein, T., Salomonis, N., Nuyten, D.S.A., van de Vijver, M.J., and Gusterson, B.A. (2009). A mouse mammary gland involution mRNA signature identifies biological pathways potentially associated with breast cancer metastasis. J Mammary Gland Biol Neoplasia *14*, 99–116.

Sternlicht, M.D., Kedeshian, P., Shao, Z.M., Safarians, S., and Barsky, S.H. (1997). The human myoepithelial cell is a natural tumor suppressor. Clin. Cancer Res. *3*, 1949–1958.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences *102*, 15545–15550.

Tanos, T., Sflomos, G., Echeverria, P.C., Ayyanan, A., Gutierrez, M., Delaloye, J.-F., Raffoul, W., Fiche, M., Dougall, W., Schneider, P., et al. (2013). Progesterone/RANKL is a major regulatory axis in the human breast. Sci Transl Med *5*, 182ra55–182ra55.

Taylor, D., Pearce, C.L., Hovanessian-Larsen, L., Downey, S., Spicer, D.V., Bartow, S., Pike, M.C., Wu, A.H., and Hawes, D. (2009). Progesterone and estrogen receptors in pregnant and premenopausal non-pregnant normal human breast. Breast Cancer Res. Treat. *118*, 161–168.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Van Keymeulen, A., Fioramonti, M., Centonze, A., Bouvencourt, G., Achouri, Y., and Blanpain, C. (2017). Lineage-Restricted Mammary Stem Cells Sustain the Development, Homeostasis, and Regeneration of the Estrogen Receptor Positive Lineage. CellReports *20*, 1525–1532.

Vogel, P.M., Georgiade, N.G., Fetter, B.F., Vogel, F.S., and McCarty, K.S. (1981). The correlation of histologic changes in the human breast with the menstrual cycle. Am. J. Pathol. *104*, 23–34.

Wang, C., Christin, J.R., Oktay, M.H., and Guo, W. (2017). Lineage-Biased Stem Cells Maintain Estrogen-Receptor-Positive and -Negative Mouse Mammary Luminal Lineages. CellReports *18*, 2825–2835.

Zhu, X., Ching, T., Pan, X., Weissman, S.M., and Garmire, L. (2017). Detecting heterogeneity in single-cell RNA-Seq data by non-negative matrix factorization. PeerJ *5*, e2888.

Figure titles and legends

Figure 1. scRNAseq identifies the major epithelial and stromal cell types in the human breast (See also Figure S1, Table S1, and Table S2). (A) Overview of approach: We predicted that pregnancy history and menstrual cycle stage would be the two the major sources of variability between reduction mammoplasty samples, and that scRNA analysis of nulliparous and parous women from different stages of the menstrual cycle could identify pregnancy-related changes in the breast and map transcriptional state to changing hormone levels across the menstrual cycle. (B) Schematic of scRNAseq workflow: Reduction mammoplasty samples were processed to a single cell suspension, followed by FACS purification and scRNAseq. (C) TSNE dimensionality reduction and KNN clustering of the combined data from all five samples. (D) *Left:* Hierarchical clustering of each cell type based on the log-transformed mean expression profile, along with their putative identities (Spearman's correlation, Ward linkage). *Right:* pseudocolored H&E section and simplified diagram depicting the organization of the human mammary gland. (E) Heatmap highlighting marker genes used to identify each cell type. For visualization purposes, we randomly selected 100 cells from each cluster.

Figure 2. Pregnancy history and menstrual cycle stage are two major sources of variation in the human breast. (See also Figure S2 and Table S3). (A) TSNE plot of aggregated live/singlet cells from nulliparous (RM263, RM272, RM273) and parous (RM264 RM284) samples, with the percent of each epithelial cell type highlighted. (B) FACS analysis of the percent of EpCAM^{-}/CD49f^{high} myoepithelial cells within the Lin^{-} epithelial population in nulliparous or parous samples. (C) Samples were immunostained with the myoepithelial marker p63 and pan-luminal marker KRT7, and the ratio of p63+ myoepithelial cells to KRT7+ luminal cells was quantified for the indicated samples. Scale bars 50 µm. (D) TSNE plot highlighting aggregated cells from nulliparous or parous samples in the luminal sort gate, with the percent of each luminal cell type highlighted. (E) Samples were immunostained with KRT23, KRT7 and ER, and the percent of ER+ cells within the KRT23- (HR-) and KRT23+ (HR+) luminal cell populations was quantified. Scale bars 100 µm. (F) Parous or nulliparous samples were immunostained with KRT23 and KRT7, and the percent of KRT23+ cells within the luminal cell population was quantified. Scale bars 50 µm. (G) Prior history of pregnancy led to an increase in the proportion of myoepithelial cells and decrease in the proportion of HR+ luminal cells in the human breast. We predicted that menstrual cycle stage would introduce a second major source of transcriptional variation in the human breast. (H) Representative images of H&E stained sections and predicted menstrual cycle stage based on blinded morphological analysis. Scale bars 100 µm. (I) *Left:* Schematic depicting the earth mover's distance (EMD) between two samples. *Right*: Heatmap representing the EMD for each pair of samples. Samples were ordered using complete linkage hierarchical clustering. (J) Pseudotime ordering of HR+ luminal cells from the indicated samples and density plot depicting the relative frequency of each sample along inferred pseudotime.

Figure 3. Two diverging transcriptional states emerge in hormone-responsive luminal cells as estrogen and progesterone levels increase (See also Figure S3, Table S5). (A) PCA and clustering of aggregated data from all samples identified three cell states in HR+ luminal cells. (B) Pseudotime analysis revealed a bifurcating trajectory in HR+ luminal cells from one to two cell states. (C) Frequency of each cell state across the menstrual cycle. (D) Heatmap of genes

with pseudotime-dependent changes in gene expression. (E) Heatmap of genes with branch-specific, pseudotime-dependent changes in gene expression. (F) Bar chart and feature plot depicting the log-normalized expression of LRRC26 or P4HA1. (G) Immunostaining for LRRC26 and KRT7 in the indicated samples and quantification of the percent of LRRC26+ luminal cells. Scale bars 50 µm. (H) Immunostaining for LRRC26, P4HA1, and KRT7 and quantification of the relative intensity of P4HA1 signal in LRRC26-/KRT7+ and LRRC26+/KRT7+ regions of interest. Scale bars 20 µm. See also Figure S3F. (I) High estrogen and progesterone levels drive the emergence of two cell states in HR+ luminal cells. One cell state expresses high levels of paracrine factors such as RANKL and WNT4 that signal to the surrounding epithelium, and the second expresses high levels of VEGFA and other stromal paracrine factors.

Figure 4. Paracrine effects of hormone signaling in hormone-insensitive epithelial cells (See also Figure S4, Table S5). (A) PCA and clustering identified two cell states in myoepithelial cells. (B) Frequency of each cell state across the menstrual cycle in myoepithelial cells. (C) Volcano plot for genes differentially expressed between the luteal- and follicular-phases in myoepithelial cells and bar chart of the ten most significantly enriched gene sets in the luteal phase. (D) PCA and clustering of aggregated data identified three cell states in HR- luminal cells. (E) Frequency of each cell state across the menstrual cycle in HR- luminal cells (F) Genes differentially expressed between the luteal- and follicular-phases in HR- luminal cells and the ten most significantly enriched gene sets in the luteal phase. (G) Genes differentially expressed between the early- and late-follicular phases in HR- luminal cells and the ten most significantly enriched gene sets in each phase. (H) Log-normalized expression of CD14 and MARCO in the indicated HR- luminal clusters. (I) Log normalized expression of the WNT effector TCF7 in the indicated cell types. (J) Immunostaining for the pan-luminal marker KRT7, TCF7, and the basal marker p63 or HR- luminal cell marker KRT23 in samples from the indicated menstrual cycle phase and quantification of the percent of TCF7+ myoepithelial (p63+/KRT7-) or HR- luminal (KRT23+/KRT7+) cells. Scale bars 50 µm. (K) Paracrine signaling from HR+ luminal cells drives WNT activation and the additional indicated gene expression programs in myoepithelial cells and HR- luminal cells. HR- luminal cells in the early follicular phase upregulate a transcriptional signature that parallels postpartum involution.

Figure 5. Paracrine signaling downstream of hormone receptor activation supports a proangiogenic microenvironment (See also Figure S5). (A) Clustering identified two endothelial cell types. (B) Log-normalized expression of blood or lymphatic endothelial markers PLVAP and PDPN in the indicated endothelial cell clusters. (C) *Left:* PCA and sub-clustering identified two cell states in lymphatic endothelial cells. *Right*: Frequency of each cell state across the menstrual cycle. (D) Bar chart of the ten most significantly enriched gene sets in the luteal phase in lymphatic endothelial cells. (E) *Left: Sub-clustering* identified three cell states in blood endothelial cells. *Right*: Frequency of each cell state across the menstrual cycle. (F) Ten most significantly enriched gene sets in blood endothelial cells in the early (left) or late (right) luteal phase. (G) Log-normalized expression of the VEGF receptors FLT1 and KDR in the indicated blood endothelial cell clusters. (H) Sub-clustering identified four cell states in vascular accessory cells. (I) Log-normalized expression of the smooth muscle marker ACTA2 in the indicated vascular accessory cell clusters. (J) Frequency of each cell state across the menstrual cycle in pericytes (left) or smooth muscle cells (right). (K) Ten most significantly enriched gene sets in

the luteal phase for pericytes (left) or smooth muscle cells (right). (L) Paracrine signaling downstream of hormone activation supports a proangiogenic microenvironment in the early luteal phase, followed by upregulation of TNF-alpha signaling and a hypoxic gene signature in the late luteal phase.

Figure 6. Remodeling of the stromal and immune microenvironments across the menstrual cycle (See also Figure S6). (A) Sub-clustering identified three cell states in fibroblasts. (B) Log-normalized expression of the preadipocyte markers ADIRF and CFD (Adipsin) in the indicated stromal clusters. (C) Volcano plot of genes differentially expressed between fibroblasts and preadipocytes and bar chart of the ten most significantly enriched gene sets in preadipocytes. (D) Frequency of fibroblast cell states across the menstrual cycle. (E) Genes differentially expressed between the luteal- and follicular-phases in fibroblasts and the ten most significantly enriched gene sets in the luteal phase. (F) Paracrine signaling downstream of hormone activation drives ECM deposition and remodeling in luteal-phase fibroblasts. (G) Clustering identified five cell types in the immune cell population. (H) *Left:* Heatmap highlighting marker genes used to identify each cell type. *Right:* hierarchical clustering of each cell type based on its log-transformed mean expression profile, and putative identities (Spearman's correlation, Ward linkage). (I) Frequency of each immune cell type across the menstrual cycle based on TSNE and clustering results. (J) Representative FACS plots of CD45, CD68, and CD163 and quantification of the percent of $CD45^{+}/SSC^{low}$ lymphocytes and $CD163^{+}$ M2-like macrophages ($CD45^{+}/SSC^{high}/CD68^{+}$) in follicular or luteal phase samples. (K) Paracrine signaling during the luteal phase drives a switch from an M2-like to an M1-like macrophage polarization and recruitment of B and T lymphocytes.

Figure 7. Analysis of samples from donors using hormonal contraceptives dissects the effects of altered hormone combinations and dynamics. (See also Figure S7, Table S6, and Table S7). (A) Schematic depicting relative estrogen and progestin levels across the natural menstrual cycle and in donors using progestin-based (P) or combination (E/P) hormonal contraceptives. Samples from donors using hormonal contraceptives were used as a "virtual experiment" to test the effects of hormone treatment on downstream signaling pathways in the indicated cell types. (B) *Left*: PC plot of HR+ luminal cells from the indicated menstrual cycle stage or hormone-treatment group. *Right*: Heatmap representing the EMD for each pair of samples, ordered by hierarchical clustering (complete linkage). (C) Bar chart depicting the log normalized expression of AREG or TNFSF11 (RANKL) in HR+ luminal cells from the indicated menstrual cycle stage or hormone-treatment group. (D) *Left*: PC plot of fibroblasts from the indicated menstrual cycle stage or hormone-treatment group. *Right*: Heatmap representing the EMD for each pair of samples (complete linkage). (E) Log normalized expression of COL3A1 or MMP14 in fibroblasts from the indicated menstrual cycle stage or hormone-treatment group. (F) *Left*: PC plot of HR- luminal cells from the indicated menstrual cycle stage or hormone-treatment group. *Right*: Heatmap representing the EMD for each pair of samples (complete linkage). (G) Log normalized expression of TNFSF10 (TRAIL) or CD14 in HR- luminal cells from the indicated menstrual cycle stage or hormone-treatment group.

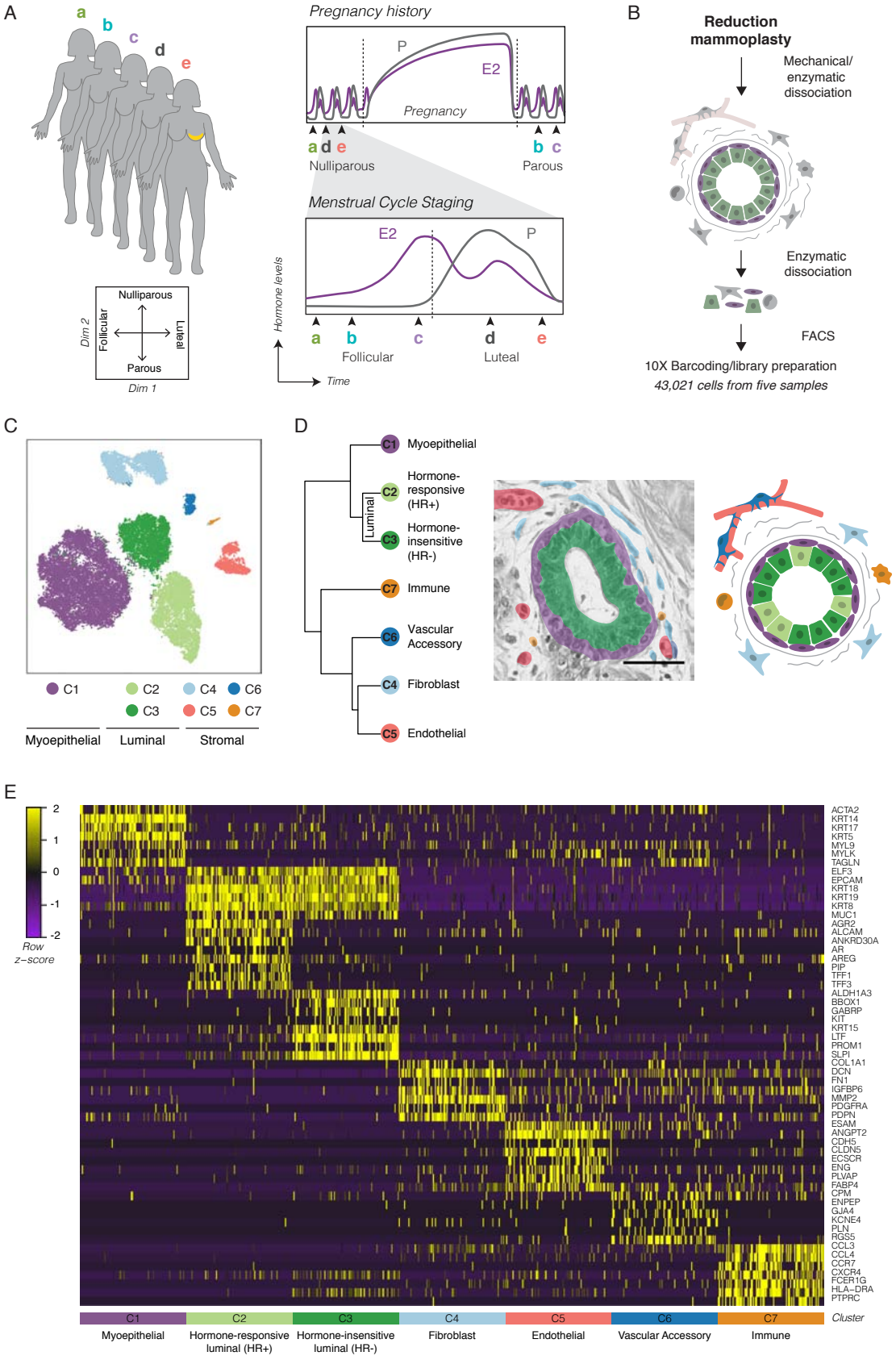

A
a
b
c
d
e
Nulliparous
Follicular
Luteal
Parous
Dim 2
Dim 1
Pregnancy history
P
E2
Pregnancy
a d e
Nulliparous
b c
Parous
Menstrual Cycle Staging
E2
P
Hormone levels
a
b
c
d
e
Follicular
Luteal
Time
B
Reduction mammoplasty
Mechanical/ enzymatic dissociation
Enzymatic dissociation
FACS
10X Barcoding/library preparation
43,021 cells from five samples
C
C1
C2
C3
C4
C5
C6
C7
Myoepithelial
Luminal
Stromal
D
C1 Myoepithelial
C2 Hormone-responsive (HR+)
C3 Hormone-insensitive (HR-)
Luminal
C7 Immune
C6 Vascular Accessory
C4 Fibroblast
C5 Endothelial
E
2
1
0
-1
-2
Row z-score
ACTA2
KRT14
KRT17
KRT5
MYL9
MYLK
TAGLN
ELF3
EPCAM
KRT18
KRT19
KRT8
MUC1
AGR2
ALCAM
ANKRD30A
AR
AREG
PIP
TFF1
TFF3
ALDH1A3
BBOX1
GABRP
KIT
KRT15
LTF
PROM1
SLPI
COL1A1
DCN
FN1
IGFBP6
MMP2
PDGFRA
PDPN
ESAM
ANGPT2
CDH5
CLDN5
ECSCR
ENG
PLVAP
FABP4
CPM
ENPEP
GJA4
KCNE4
PLN
RGS5
CCL3
CCL4
CCR7
CXCR4
FCER1G
HLA-DRA
PTPRC
C1
C2
C3
C4
C5
C6
C7
Cluster
Myoepithelial
Hormone-responsive luminal (HR+)
Hormone-insensitive luminal (HR-)
Fibroblast
Endothelial
Vascular Accessory
Immune

Figure 1

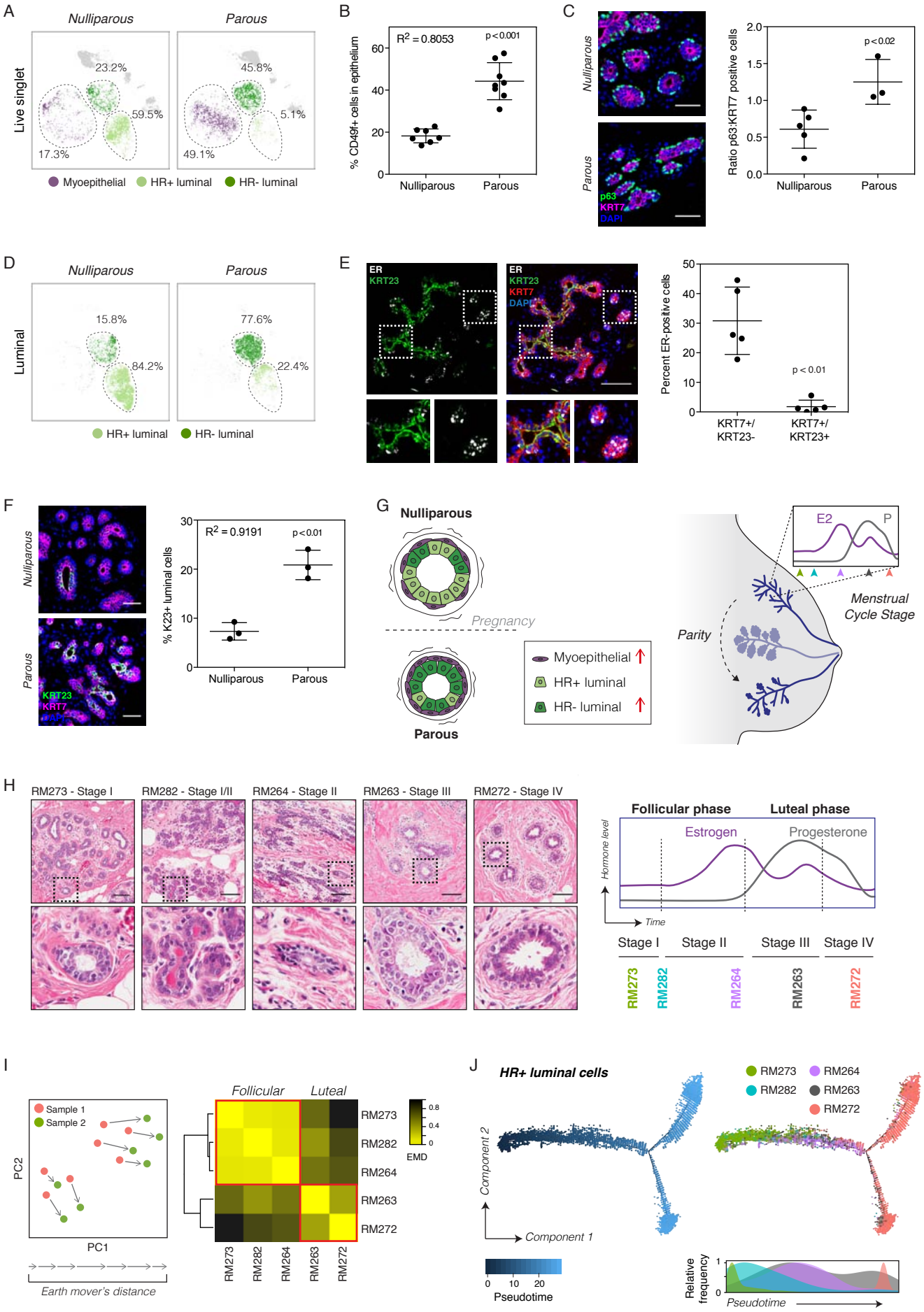

A
Nulliparous
Parous
Live singlet
23.2%
59.5%
17.3%
45.8%
5.1%
49.1%
Myoepithelial
HR+ luminal
HR- luminal
B
R2 = 0.8053
p < 0.001
% CD49f+ cells in epithelium
Nulliparous
Parous
C
Nulliparous
Parous
p63
KRT7
DAPI
p < 0.02
Ratio p63:KRT7 positive cells
D
Luminal
15.8%
84.2%
77.6%
22.4%
HR+ luminal
HR- luminal
E
ER
KRT23
KRT7
DAPI
p < 0.01
Percent ER-positive cells
KRT7+/ KRT23-
KRT7+/ KRT23+
F
R2 = 0.9191
p < 0.01
% K23+ luminal cells
G
Nulliparous
Pregnancy
Parous
Myoepithelial
HR+ luminal
HR- luminal
E2
P
Menstrual Cycle Stage
Parity
H
RM273 - Stage I
RM282 - Stage I/II
RM264 - Stage II
RM263 - Stage III
RM272 - Stage IV
Follicular phase
Luteal phase
Estrogen
Progesterone
Hormone level
Time
Stage I
Stage II
Stage III
Stage IV
RM273
RM282
RM264
RM263
RM272
I
Sample 1
Sample 2
PC2
PC1
Earth mover's distance
Follicular
Luteal
EMD
J
HR+ luminal cells
Component 2
Component 1
Pseudotime
Relative frequency

Figure 2

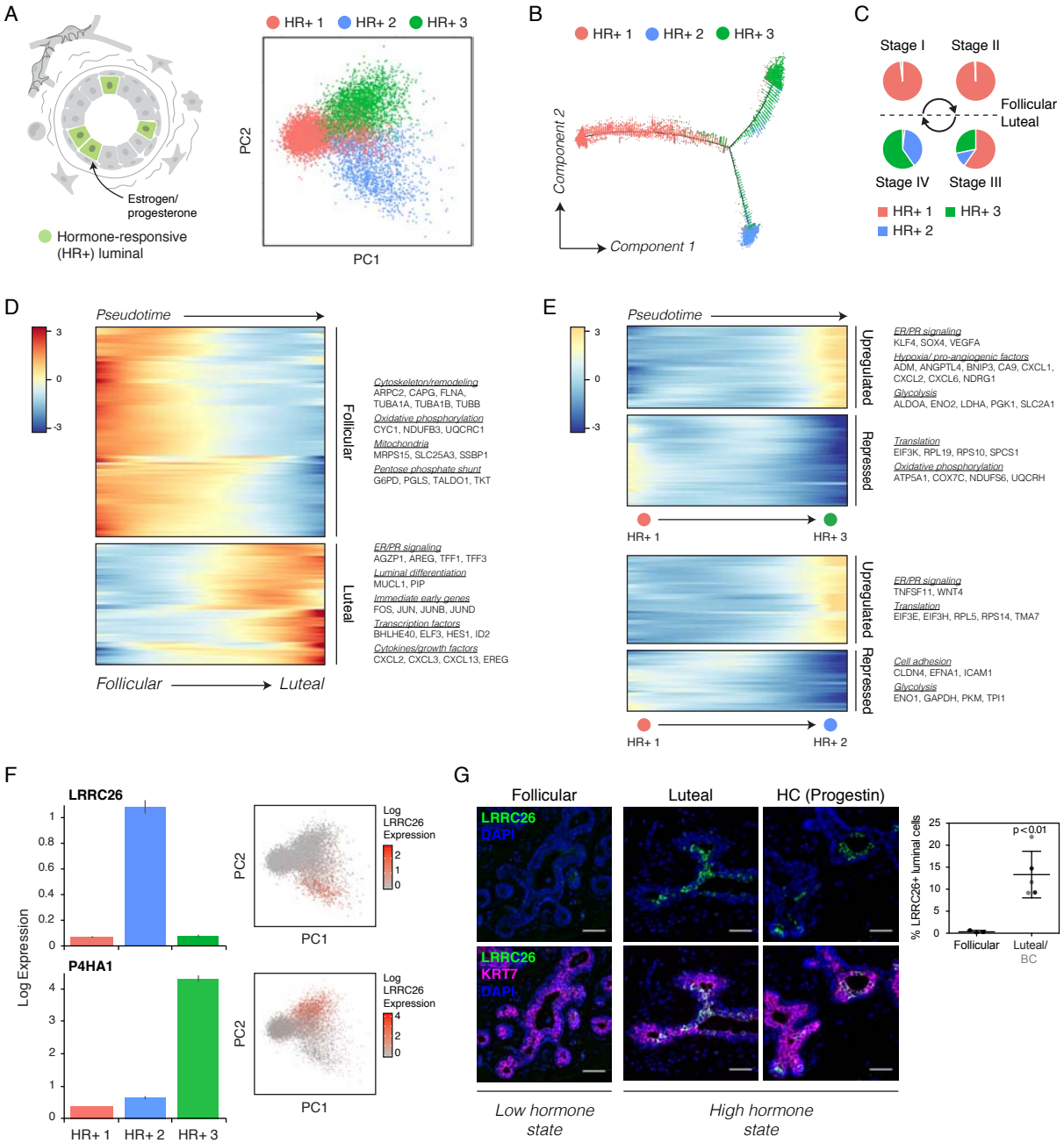

A
Estrogen/ progesterone
Hormone-responsive (HR+) luminal
HR+ 1
HR+ 2
HR+ 3
PC2
PC1
B
HR+ 1
HR+ 2
HR+ 3
Component 2
Component 1
C
Stage I
Stage II
Follicular
Luteal
Stage IV
Stage III
HR+ 1
HR+ 3
HR+ 2
D
Pseudotime
3
0
-3
Cytoskeleton/remodeling
ARPC2, CAPG, FLNA, TUBA1A, TUBA1B, TUBB
Oxidative phosphorylation
CYC1, NDUFB3, UQCRC1
Mitochondria
MRPS15, SLC25A3, SSBP1
Pentose phosphate shunt
G6PD, PGLS, TALDO1, TKT
Follicular
ER/PR signaling
AGZP1, AREG, TFF1, TFF3
Luminal differentiation
MUCL1, PIP
Immediate early genes
FOS, JUN, JUNB, JUND
Transcription factors
BHLHE40, ELF3, HES1, ID2
Cytokines/growth factors
CXCL2, CXCL3, CXCL13, EREG
Luteal
Follicular
Luteal
E
Pseudotime
3
0
-3
Upregulated
ER/PR signaling
KLF4, SOX4, VEGFA
Hypoxia/ pro-angiogenic factors
ADM, ANGPTL4, BNIP3, CA9, CXCL1, CXCL2, CXCL6, NDRG1
Glycolysis
ALDOA, ENO2, LDHA, PGK1, SLC2A1
Repressed
Translation
EIF3K, RPL19, RPS10, SPCS1
Oxidative phosphorylation
ATP5A1, COX7C, NDUFS6, UQCRH
HR+ 1
HR+ 3
Upregulated
ER/PR signaling
TNFSF11, WNT4
Translation
EIF3E, EIF3H, RPL5, RPS14, TMA7
Repressed
Cell adhesion
CLDN4, EFNA1, ICAM1
Glycolysis
ENO1, GAPDH, PKM, TPI1
HR+ 1
HR+ 2
F
LRRC26
P4HA1
Log Expression
HR+ 1
HR+ 2
HR+ 3
PC2
PC1
Log LRRC26 Expression
G
Follicular
Luteal
HC (Progestin)
LRRC26
DAPI
LRRC26
KRT7
DAPI
Low hormone state
High hormone state
% LRRC26+ luminal cells
p<0.01
Follicular
Luteal/ BC
H
LRRC26
KRT7
DAPI
LRRC26
P4HA1
P4HA1
KRT7 + ROI
LRRC26+ ROI
Rlative intensity P4HA1
p<0.001
LRRC26- KRT7+ ROI
LRRC26+ KRT7+ ROI
I
Follicular
Luteal
VEGF
RANKL/ WNT4
E/P

Figure 3

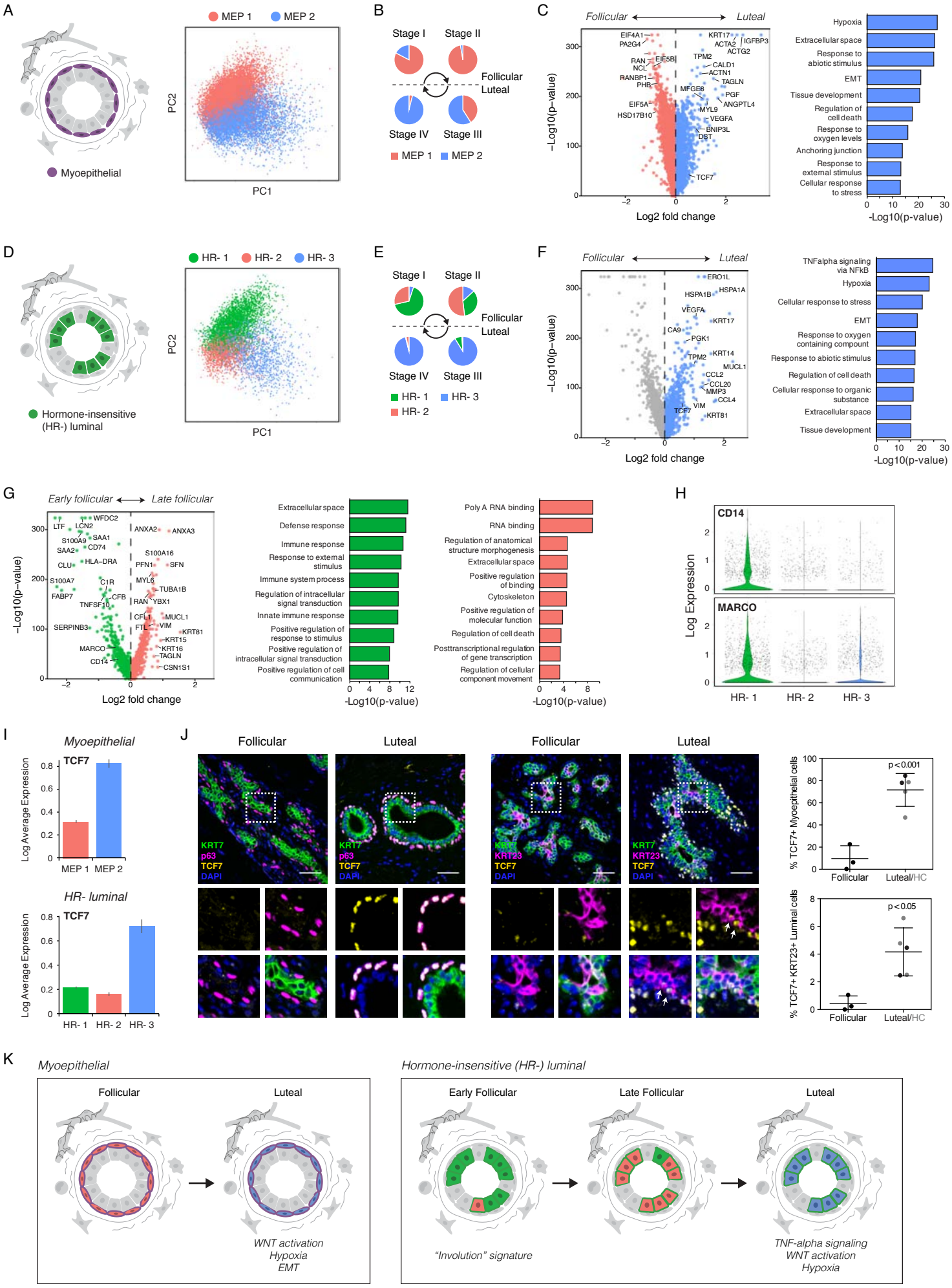

A
MEP 1
MEP 2
PC2
PC1
Myoepithelial
B
Stage I
Stage II
Follicular
Luteal
Stage IV
Stage III
MEP 1
MEP 2
C
Follicular
Luteal
−Log10(p−value)
Log2 fold change
Hypoxia
Extracellular space
Response to abiotic stimulus
EMT
Tissue development
Regulation of cell death
Response to oxygen levels
Anchoring junction
Response to external stimulus
Cellular response to stress
-Log10(p-value)
D
HR- 1
HR- 2
HR- 3
PC2
PC1
Hormone-insensitive (HR-) luminal
E
Stage I
Stage II
Follicular
Luteal
Stage IV
Stage III
HR- 1
HR- 3
HR- 2
F
Follicular
Luteal
−Log10(p−value)
Log2 fold change
TNFalpha signaling via NFkB
Hypoxia
Cellular response to stress
EMT
Response to oxygen containing compount
Response to abiotic stimulus
Regulation of cell death
Cellular response to organic substance
Extracellular space
Tissue development
-Log10(p-value)
G
Early follicular
Late follicular
−Log10(p−value)
Log2 fold change
Extracellular space
Defense response
Immune response
Response to external stimulus
Immune system process
Regulation of intracellular signal transduction
Innate immune response
Positive regulation of response to stimulus
Positive regulation of intracellular signal transduction
Positive regulation of cell communication
-Log10(p-value)
Poly A RNA binding
RNA binding
Regulation of anatomical structure morphogenesis
Extracellular space
Positive regulation of binding
Cytoskeleton
Positive regulation of molecular function
Regulation of cell death
Posttranscriptional regulation of gene transcription
Regulation of cellular component movement
-Log10(p-value)
H
CD14
MARCO
Log Expression
HR- 1
HR- 2
HR- 3
I
Myoepithelial
TCF7
Log Average Expression
MEP 1
MEP 2
HR- luminal
TCF7
Log Average Expression
HR- 1
HR- 2
HR- 3
J
Follicular
Luteal
Follicular
Luteal
KRT7
p63
TCF7
DAPI
KRT7
p63
TCF7
DAPI
KRT7
KRT23
TCF7
DAPI
KRT7
KRT23
TCF7
DAPI
% TCF7+ Myoepithelial cells
p<0.001
Follicular
Luteal/HC
% TCF7+ KRT23+ Luminal cells
p<0.05
Follicular
Luteal/HC
K
Myoepithelial
Follicular
Luteal
WNT activation
Hypoxia
EMT
Hormone-insensitive (HR-) luminal
Early Follicular
Late Follicular
Luteal
"Involution" signature
TNF-alpha signaling
WNT activation
Hypoxia

Figure 4

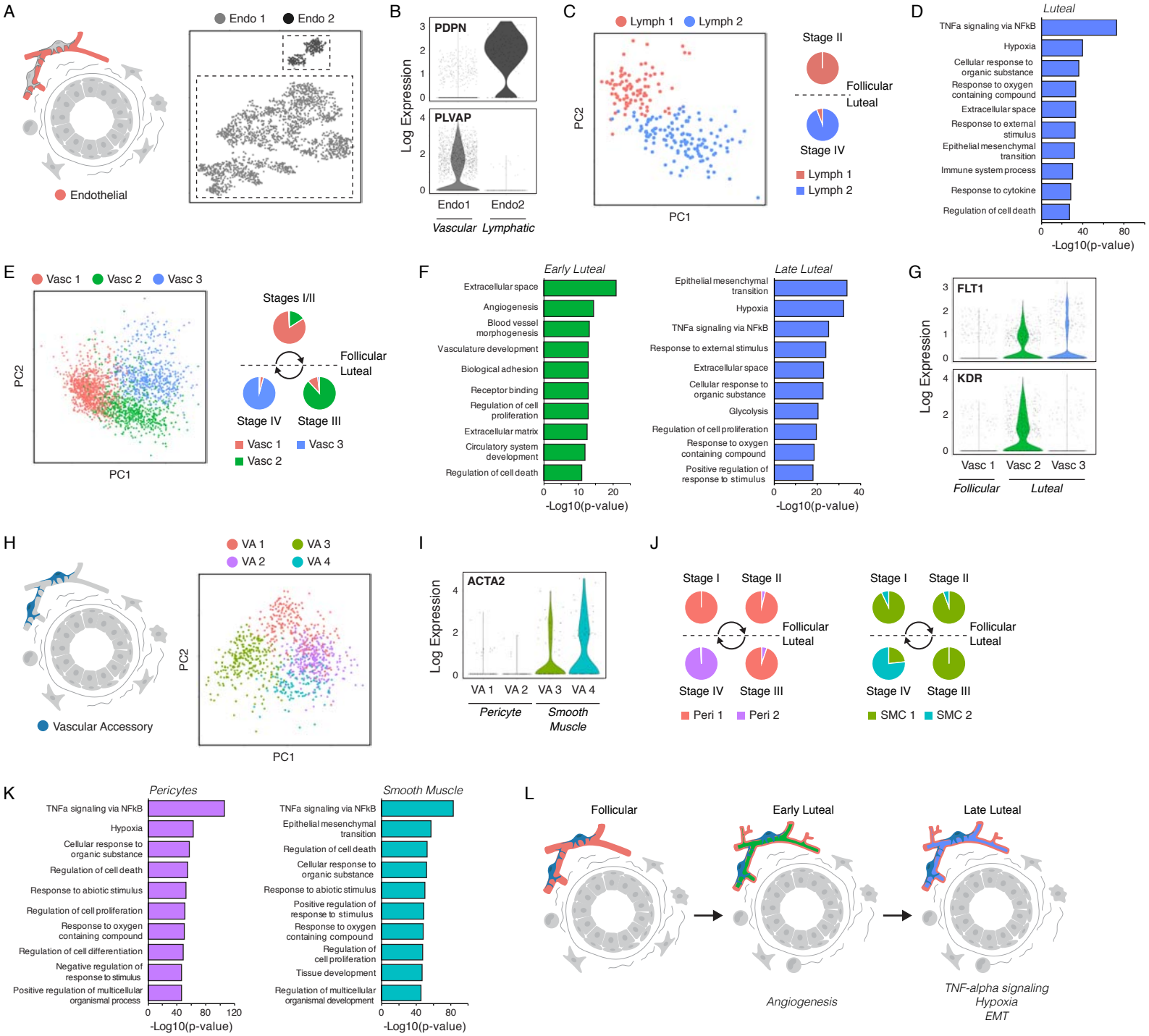

A
Endo 1
Endo 2
Endothelial
B
PDPN
PLVAP
Log Expression
Endo1
Endo2
Vascular
Lymphatic
C
Lymph 1
Lymph 2
PC2
PC1
Stage II
Follicular
Luteal
Stage IV
Lymph 1
Lymph 2
D
Luteal
TNFa signaling via NFkB
Hypoxia
Cellular response to organic substance
Response to oxygen containing compound
Extracellular space
Response to external stimulus
Epithelial mesenchymal transition
Immune system process
Response to cytokine
Regulation of cell death
-Log10(p-value)
E
Vasc 1
Vasc 2
Vasc 3
PC2
PC1
Stages I/II
Follicular
Luteal
Stage IV
Stage III
Vasc 1
Vasc 3
Vasc 2
F
Early Luteal
Extracellular space
Angiogenesis
Blood vessel morphogenesis
Vasculature development
Biological adhesion
Receptor binding
Regulation of cell proliferation
Extracellular matrix
Circulatory system development
Regulation of cell death
-Log10(p-value)
Late Luteal
Epithelial mesenchymal transition
Hypoxia
TNFa signaling via NFkB
Response to external stimulus
Extracellular space
Cellular response to organic substance
Glycolysis
Regulation of cell proliferation
Response to oxygen containing compound
Positive regulation of response to stimulus
-Log10(p-value)
G
FLT1
KDR
Log Expression
Vasc 1
Vasc 2
Vasc 3
Follicular
Luteal
H
VA 1
VA 2
VA 3
VA 4
Vascular Accessory
PC2
PC1
I
ACTA2
Log Expression
VA 1
VA 2
VA 3
VA 4
Pericyte
Smooth Muscle
J
Stage I
Stage II
Follicular
Luteal
Stage IV
Stage III
Peri 1
Peri 2
Stage I
Stage II
Follicular
Luteal
Stage IV
Stage III
SMC 1
SMC 2
K
Pericytes
TNFa signaling via NFkB
Hypoxia
Cellular response to organic substance
Regulation of cell death
Response to abiotic stimulus
Regulation of cell proliferation
Response to oxygen containing compound
Regulation of cell differentiation
Negative regulation of response to stimulus
Positive regulation of multicellular organismal process
-Log10(p-value)
Smooth Muscle
TNFa signaling via NFkB
Epithelial mesenchymal transition
Regulation of cell death
Cellular response to organic substance
Response to abiotic stimulus
Positive regulation of response to stimulus
Response to oxygen containing compound
Regulation of cell proliferation
Tissue development
Regulation of multicellular organismal development
-Log10(p-value)
L
Follicular
Early Luteal
Late Luteal
Angiogenesis
TNF-alpha signaling
Hypoxia
EMT

Figure 5

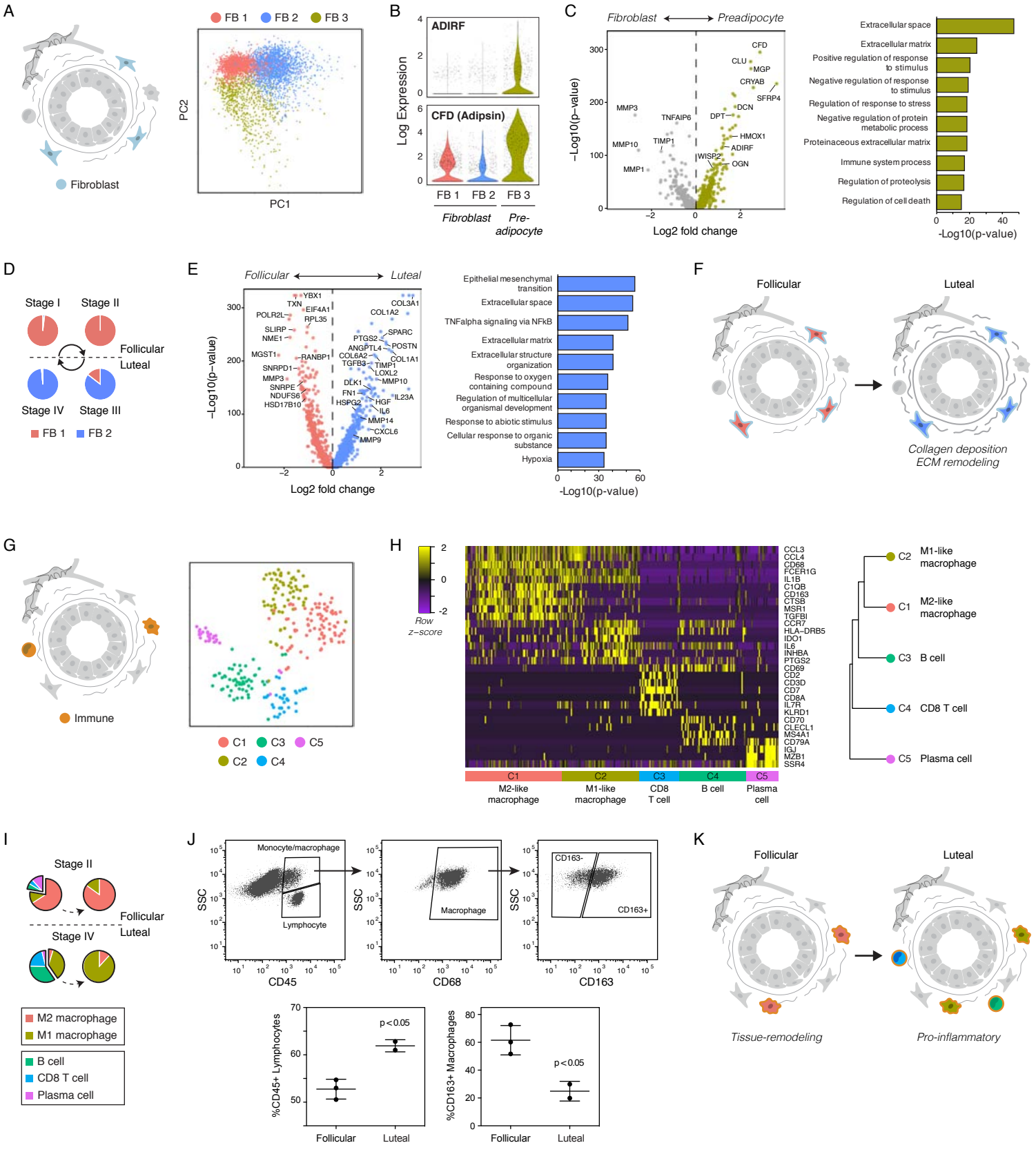

A
Fibroblast
FB 1
FB 2
FB 3
PC2
PC1
B
ADIRF
CFD (Adipsin)
Log Expression
FB 1
FB 2
FB 3
Fibroblast
Pre-adipocyte
C
Fibroblast
Preadipocyte
−Log10(p−value)
Log2 fold change
CFD
CLU
MGP
CRYAB
SFRP4
DCN
DPT
HMOX1
ADIRF
OGN
WISP2
MMP3
TNFAIP6
MMP10
TIMP1
MMP1
Extracellular space
Extracellular matrix
Positive regulation of response to stimulus
Negative regulation of response to stimulus
Regulation of response to stress
Negative regulation of protein metabolic process
Proteinaceous extracellular matrix
Immune system process
Regulation of proteolysis
Regulation of cell death
-Log10(p-value)
D
Stage I
Stage II
Follicular
Luteal
Stage IV
Stage III
FB 1
FB 2
E
Follicular
Luteal
−Log10(p−value)
Log2 fold change
YBX1
TXN
EIF4A1
POLR2L
RPL35
SLIRP
NME1
MGST1
RANBP1
SNRPD1
MMP3
SNRPE
NDUFS6
HSD17B10
COL3A1
COL1A2
SPARC
PTGS2
POSTN
ANGPTL4
COL6A2
COL1A1
TGFB3
TIMP1
LOXL2
MMP10
DKK1
FN1
IL23A
HSPG2
HGF
IL6
MMP14
CXCL6
MMP9
Epithelial mesenchymal transition
Extracellular space
TNFalpha signaling via NFkB
Extracellular matrix
Extracellular structure organization
Response to oxygen containing compound
Regulation of multicellular organismal development
Response to abiotic stimulus
Cellular response to organic substance
Hypoxia
-Log10(p-value)
F
Follicular
Luteal
Collagen deposition
ECM remodeling
G
Immune
C1
C3
C5
C2
C4
H
Row z−score
CCL3
CCL4
CD68
FCER1G
IL1B
C1QB
CD163
CTSB
MSR1
TGFBI
CCR7
HLA−DRB5
IDO1
IL6
INHBA
PTGS2
CD69
CD2
CD3D
CD7
CD8A
IL7R
KLRD1
CD70
CLECL1
MS4A1
CD79A
IGJ
MZB1
SSR4
C1 M2-like macrophage
C2 M1-like macrophage
C3 CD8 T cell
C4 B cell
C5 Plasma cell
I
Stage II
Follicular
Luteal
Stage IV
M2 macrophage
M1 macrophage
B cell
CD8 T cell
Plasma cell
J
Monocyte/macrophage
Lymphocyte
Macrophage
CD163-
CD163+
SSC
CD45
CD68
CD163
%CD45+ Lymphocytes
%CD163+ Macrophages
p < 0.05
Follicular
Luteal
K
Follicular
Luteal
Tissue-remodeling
Pro-inflammatory

Figure 6

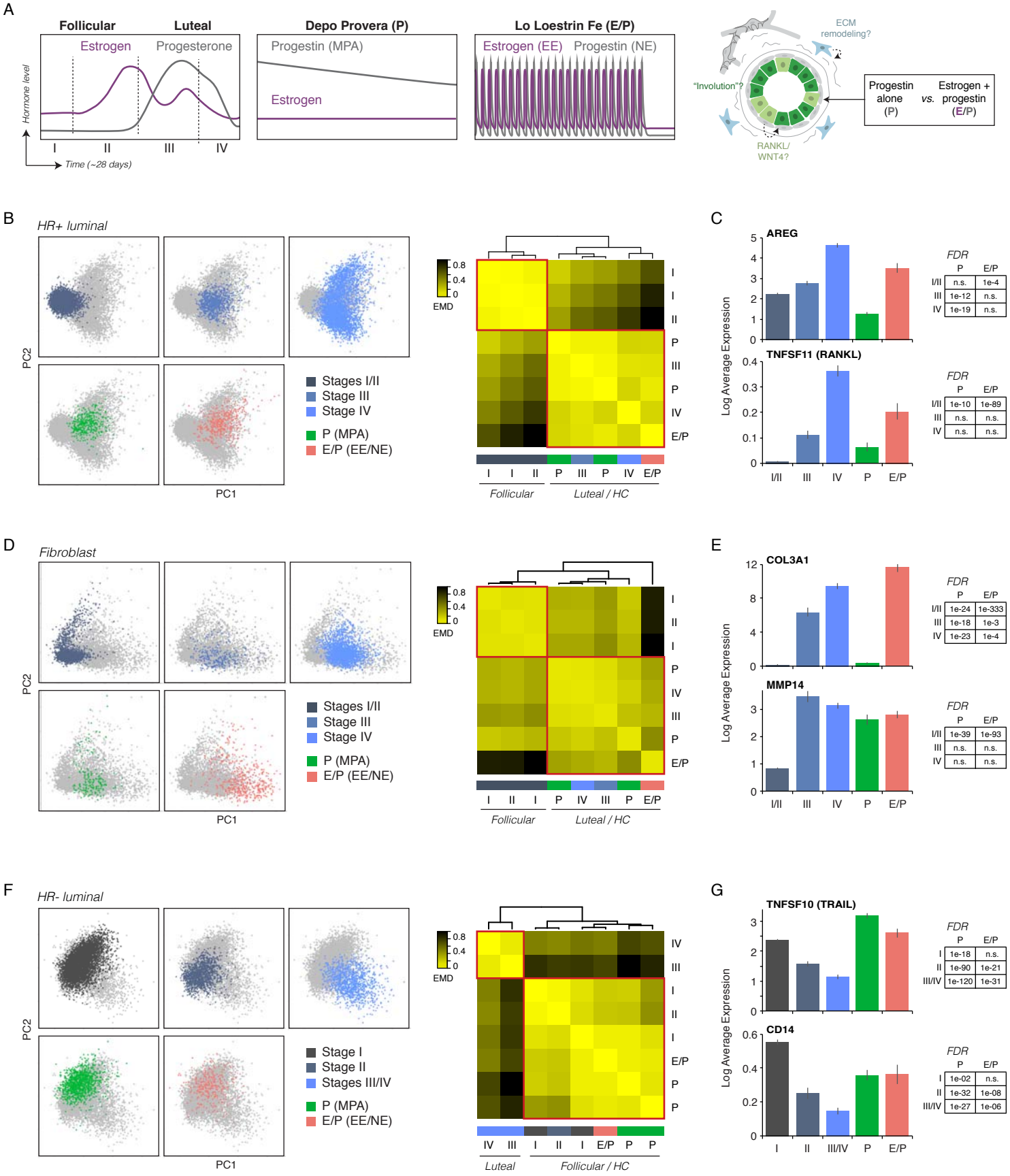

A
Follicular
Luteal
Estrogen
Progesterone
Hormone level
I
II
III
IV
Time (~28 days)
Depo Provera (P)
Progestin (MPA)
Estrogen
Lo Loestrin Fe (E/P)
Estrogen (EE)
Progestin (NE)
ECM remodeling?
"Involution"?
RANKL/ WNT4?
Progestin alone (P)
vs.
Estrogen + progestin (E/P)
B
HR+ luminal
PC2
PC1
Stages I/II
Stage III
Stage IV
P (MPA)
E/P (EE/NE)
EMD
0.8
0.4
0
I I II P III P IV E/P
Follicular
Luteal / HC
C
AREG
TNFSF11 (RANKL)
Log Average Expression
FDR
P E/P
I/II n.s. 1e-4
III 1e-12 n.s.
IV 1e-19 n.s.
FDR
P E/P
I/II 1e-10 1e-89
III n.s. n.s.
IV n.s. n.s.
I/II III IV P E/P
D
Fibroblast
PC2
PC1
Stages I/II
Stage III
Stage IV
P (MPA)
E/P (EE/NE)
EMD
0.8
0.4
0
I II I P IV III P E/P
Follicular
Luteal / HC
E
COL3A1
MMP14
Log Average Expression
FDR
P E/P
I/II 1e-24 1e-333
III 1e-18 1e-3
IV 1e-23 1e-4
FDR
P E/P
I/II 1e-39 1e-93
III n.s. n.s.
IV n.s. n.s.
I/II III IV P E/P
F
HR- luminal
PC2
PC1
Stage I
Stage II
Stages III/IV
P (MPA)
E/P (EE/NE)
EMD
0.8
0.4
0
IV III I II I E/P P P
Luteal
Follicular / HC
G
TNFSF10 (TRAIL)
CD14
Log Average Expression
FDR
P E/P
I 1e-18 n.s.
II 1e-90 1e-21
III/IV 1e-120 1e-31
FDR
P E/P
I 1e-02 n.s.
II 1e-32 1e-08
III/IV 1e-27 1e-06
I II III/IV P E/P

Figure 7

Supplemental Information Titles and Legends

Figure S1. Sorting strategy for scRNAseq experiments and marker analysis of cell type clusters. Related to Figure 1. (A) FACS plots depicting sort gates used for sequencing. (B) Violin plot depicting the log-normalized expression of the myoepithelial marker KRT14, the luminal marker KRT19, and the stromal marker VIM in cells from the indicated sort gates. (C) TSNE dimensionality reduction of the combined data from all five samples resolved sorted myoepithelial and luminal cell populations. (D) TSNE dimensionality reduction and KNN clustering of each sample. (E) Bar chart depicting the Log2 fold change in expression of the top 15 marker genes for each cluster relative to all other clusters. Gene names in grey represent markers that are enriched in more than one cluster. (F) *Left:* Bar chart depicting the log normalized expression of ESR1 and PGR in the indicated clusters. *Right*: TSNE dimensionality reduction plots depicting cells with expression of ESR1 or PGR in red.

Figure S2. Identification of changes in cell proportion and transcriptional state with pregnancy history and menstrual cycle stage. Related to Figure 2. (A) Scatter plots depicting the correlation between the percent of EpCAM^{-}/CD49f^{high} myoepithelial cells within the Lin^{-} epithelial population as measured by flow cytometry analysis with the indicated parameters. (B) Parous or nulliparous were immunostained with ER, PR, and the pan-luminal marker KRT7, and the percentage of ER+, PR+, or double-positive ER+/PR+ luminal (KRT7+) cells was quantified. Scale bars 100 µm. (C) Multiplexed *in situ* hybridization of estrogen receptor transcript (ESR1) and immunostaining for estrogen receptor protein (ER) and KRT7. *Right:* Plots depicting the correlation between ESR1 and ER across multiple tissue sections or within individual cells. Scale bars 25 µm. (D) Violin plot depicting the log-normalized expression of KRT23 in the indicated clusters. (E) *Top:* Scatter plot depicting the follicular or luteal score for each cell from the indicated samples, representing the scaled average expression of differentially expressed follicular- or luteal-phase transcripts (see also Table S4). *Bottom:* Bar chart depicting the average menstrual cycle score for the indicated samples, representing the difference between the average follicular and luteal scores for each sample. (F) Follicular or luteal phase samples were immunostained with PR, KRT23, and KRT7, and the percentage of PR+ hormone-responsive luminal (KRT7+/KRT23-) cells was quantified. Scale bars 50 µm. (G) Pseudotime ordering of myoepithelial cells from the indicated samples and density plot depicting the relative frequency of each sample along inferred pseudotime.

Figure S3. Cell state changes in hormone-receptor positive luminal cells with menstrual cycle stage. Related to Figure 3. (A) *Left*: Consensus matrices of NMF results averaging 30 runs for rank k=2-5 for 1000 randomly sampled HR+ luminal cells. *Right*: Cophenetic correlation coefficients and dispersion for hierarchically clustered consensus matrices run for k=2-7. (B) Heatmap of the top 40 genes contributing to each principal component (PC) in HR+ luminal cells. For visualization purposes, the top 100 positive and negative cells ordered by PC scores are shown. (C) Principal component plot of HR+ luminal cells from each sample colored by NMF sub-clustering results. (D) Observed and predicted future cell states using RNA velocity estimation are shown on the first two PCs for HR+ luminal cells. Velocities were based on nearest-cell pooling (k = 25) and visualized using Gaussian smoothing on a regular grid. (E) Bar chart depicting the Log2 fold change in expression of the top 10 marker genes for each cluster

relative to all other clusters. (F) Representative image of immunostaining for LRRC26, P4HA1, and KRT7. Scale bars 50 µm.

Figure S4. Cell state changes in hormone-insensitive epithelial cells in response to menstrual cycle stage. Related to Figure 4. (A) *Left*: Consensus matrices of NMF results averaging 30 runs for rank k=2-4 for 1000 randomly sampled myoepithelial cells. *Right*: Cophenetic correlation coefficients and dispersion for hierarchically clustered consensus matrices run for k=2-7. (B) Heatmap of the top 40 genes contributing to each principal component (PC) in myoepithelial cells. For visualization purposes, the top 100 positive and negative cells ordered by PC scores are shown. (C) Principal component plot of myoepithelial cells from each sample colored by NMF sub-clustering results. (D) Bar chart depicting the Log2 fold change in expression of the top 10 marker genes for each myoepithelial cluster relative to all other clusters. (E) Bar chart of the ten most significantly enriched gene sets in myoepithelial cells in follicular phase. (F) *Left*: Consensus matrices of NMF results averaging 30 runs for rank k=2-4 for 1000 randomly sampled HR- luminal cells. *Right*: Cophenetic correlation coefficients and dispersion for hierarchically clustered consensus matrices run for k=2-7. (G) Heatmap of the top 40 genes contributing to each principal component (PC) in HR- luminal cells. For visualization purposes, the top 100 positive and negative cells ordered by PC scores are shown. (H) PC plot of HR- luminal cells from each sample colored by NMF sub-clustering results. (I) Bar chart depicting the Log2 fold change in expression of the top 10 marker genes for each HR- luminal cell cluster relative to all other clusters. (J) Violin plot depicting the log-normalized expression of TNF in the indicated HR- luminal cell clusters.

Figure S5. Cell state changes in endothelial and vascular accessory cells in response to menstrual cycle stage. Related to Figure 5. (A) *Top*: Consensus matrices of NMF results averaging 30 runs for rank k=2-3 for lymphatic endothelial cells. *Bottom*: Cophenetic correlation coefficients and dispersion for hierarchically clustered consensus matrices run for k=2-7. (B) Principal component plot of lymphatic endothelial cells colored by sample. (C) Volcano plot of genes differentially expressed between the luteal-phase and follicular-phase clusters in lymphatic endothelial cells and bar chart of the ten most significantly enriched gene sets in the follicular phase. (D) *Top*: Consensus matrices of NMF results averaging 30 runs for rank k=2-3 for 1000 randomly sampled vascular endothelial cells. *Bottom*: Cophenetic correlation coefficients and dispersion for hierarchically clustered consensus matrices run for k=2-7. (E) Principal component plot of vascular endothelial cells colored by sample. (F) Volcano plots of genes differentially expressed between the indicated menstrual cycle phases in vascular endothelial cells and bar chart of the ten most significantly enriched gene sets in the follicular phase. (G) *Top*: Consensus matrices of NMF results averaging 30 runs for rank k=2-4 for vascular accessory cells. *Bottom*: Cophenetic correlation coefficients and dispersion for hierarchically clustered consensus matrices run for k=2-7. (H) Principal component plot of vascular accessory cells colored by sample. (I) Volcano plot of genes differentially expressed between the follicular-phase and late luteal-phase clusters in pericytes and bar chart of the ten most significantly enriched gene sets in the follicular phase. (J) Volcano plots of gene differentially expressed between the follicular-phase and late luteal-phase clusters in smooth muscle cells and bar chart of the ten most significantly enriched gene sets in the follicular phase.

Figure S6. Changes in the stromal and immune microenvironments in response to menstrual cycle stage. Related to Figure 6. (A) *Left*: Consensus matrices of NMF results averaging 30 runs for rank k=2-5 for 1000 randomly sampled fibroblasts. *Right*: Cophenetic correlation coefficients and dispersion for hierarchically clustered consensus matrices run for k=2-7. (B) Heatmap of the top 40 genes contributing to each principal component (PC) in fibroblasts. For visualization purposes, the top 100 positive and negative cells ordered by PC scores are shown. (C) Bar chart depicting the Log2 fold change in expression of the top 10 marker genes for each fibroblast/preadipocyte cluster relative to all other clusters. (D) Principal component plot of fibroblasts/preadipocytes from each sample colored by NMF sub-clustering results. (E) Bar chart of the ten most significantly enriched gene sets in fibroblasts in the follicular phase. (F) Violin plot depicting the log-normalized expression of the immune cell marker genes CD68, CD163, INHBA, CD79A, MS4A1 (CD20), IGJ, and CD8A in the indicated clusters. (G) Principal component plot of immune cells colored by sample.

Figure S7. Identification of cell type clusters and marker analysis in samples from donors using hormonal contraceptives. Related to Figure 7. (A) TSNE dimensionality reduction of the combined data from three donors colored by hormonal contraceptive type. (B) *Left:* TSNE dimensionality reduction and KNN clustering of the combined data from three samples identified seven cell types. *Right:* Hierarchical clustering of each cell type based on the log-transformed mean expression profile, along with their putative identities (Spearman's correlation, Ward linkage). (C) Heatmap highlighting marker genes used to identify each cell type. For visualization purposes, we randomly selected 100 cells from each cluster 1-6 and 50 cells from cluster 7 (Immune). (D) *Top:* Heatmap depicting the relative expression of the specified genes in HR+ luminal cells from the indicated menstrual cycle stage or hormone-treatment group. *Bottom:* Bar chart depicting the log normalized expression of WNT4 or SOX4 in HR+ luminal cells from the indicated menstrual cycle stage or hormone-treatment group. (E) *Top:* Heatmap depicting the relative expression of the specified genes in fibroblasts from the indicated menstrual cycle stage or hormone-treatment group. *Bottom:* Bar chart depicting the log normalized expression of DCN or IGFBP2 in fibroblasts from the indicated menstrual cycle stage or hormone-treatment group. (F) *Top:* Heatmap depicting the relative expression of the specified genes in HR- luminal cells from the indicated menstrual cycle stage or hormone-treatment group. *Bottom:* Bar chart depicting the log normalized expression of SAA1 or TCF7 in HR- luminal cells from the indicated menstrual cycle stage or hormone-treatment group.

Table S1. Donor information for reduction mammoplasty samples. Related to Figure 1.

Table S2. Summary statistics for sequencing of five reduction mammoplasty samples. Related to Figure 1.

Table S3. Microarray data for selected genes from Peri *et al.* Related to Figure 2 and Figure S2.

Table S4. Top 20 differentially expressed genes in the luteal or follicular phases from Pardo *et al*. used to develop a Menstrual Cycle Score for each sample. Related to Figure 2 and Figure S2.

Table S5. Donor information for additional reduction mammoplasty samples used in immunohistochemical analyses. Related to Figures 3-4.

Table S6. Donor information for sequenced reduction mammoplasty samples with hormonal contraceptive use. Related to Figure 7.

Table S7. Summary statistics for sequencing of three reduction mammoplasty samples from donor using hormonal contraception. Related to Figure 7.

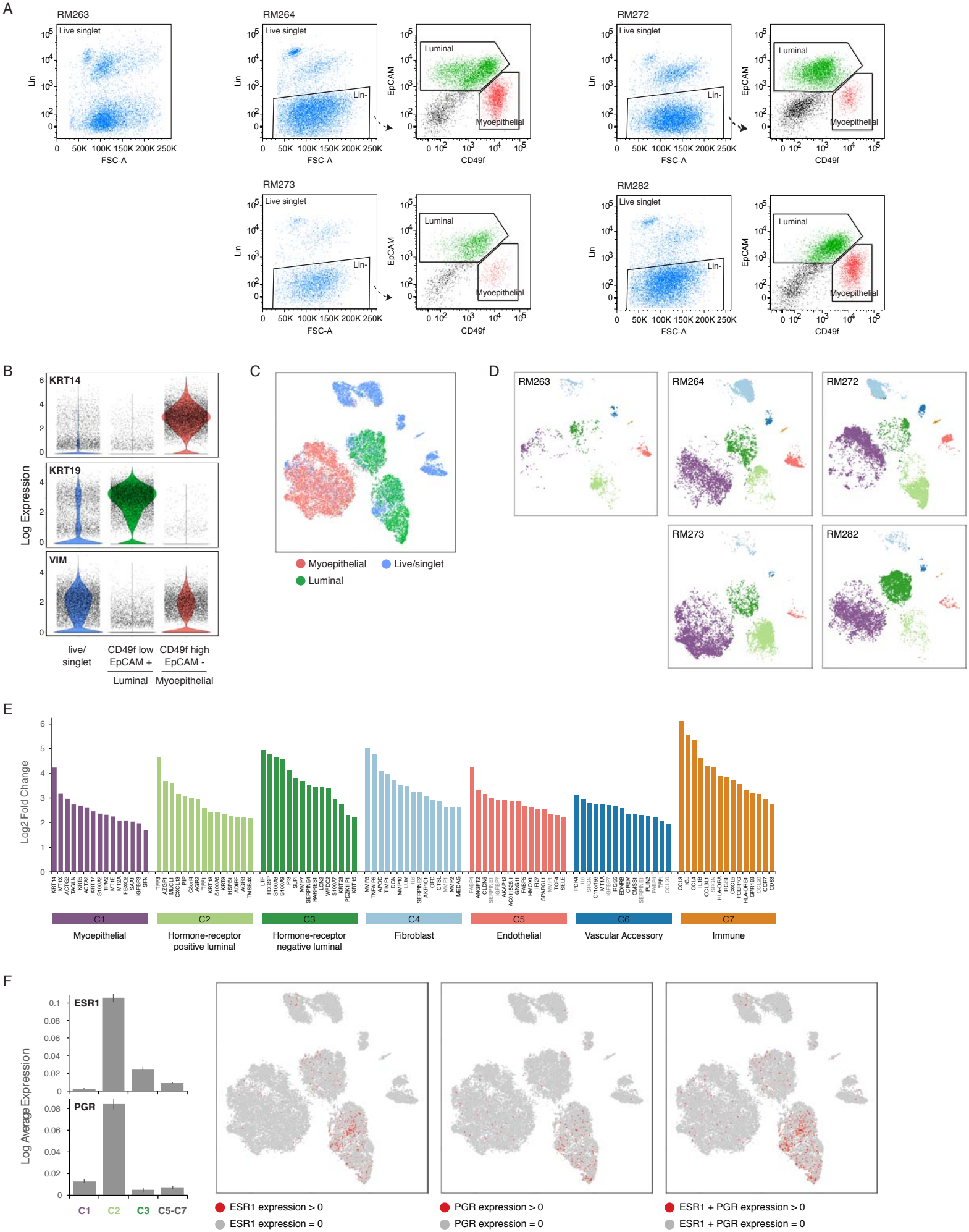

A
RM263
RM264
RM272
RM273
RM282
Live singlet
Lin
FSC-A
Lin-
EpCAM
CD49f
Luminal
Myoepithelial
B
KRT14
KRT19
VIM
Log Expression
live/ singlet
CD49f low EpCAM +
CD49f high EpCAM -
Luminal
Myoepithelial
C
Myoepithelial
Live/singlet
Luminal
D
E
Log2 Fold Change
C1
Myoepithelial
C2
Hormone-receptor positive luminal
C3
Hormone-receptor negative luminal
C4
Fibroblast
C5
Endothelial
C6
Vascular Accessory
C7
Immune
F
ESR1
PGR
Log Average Expression
C1
C2
C3
C5-C7
ESR1 expression > 0
ESR1 expression = 0
PGR expression > 0
PGR expression = 0
ESR1 + PGR expression > 0
ESR1 + PGR expression = 0

Figure S1

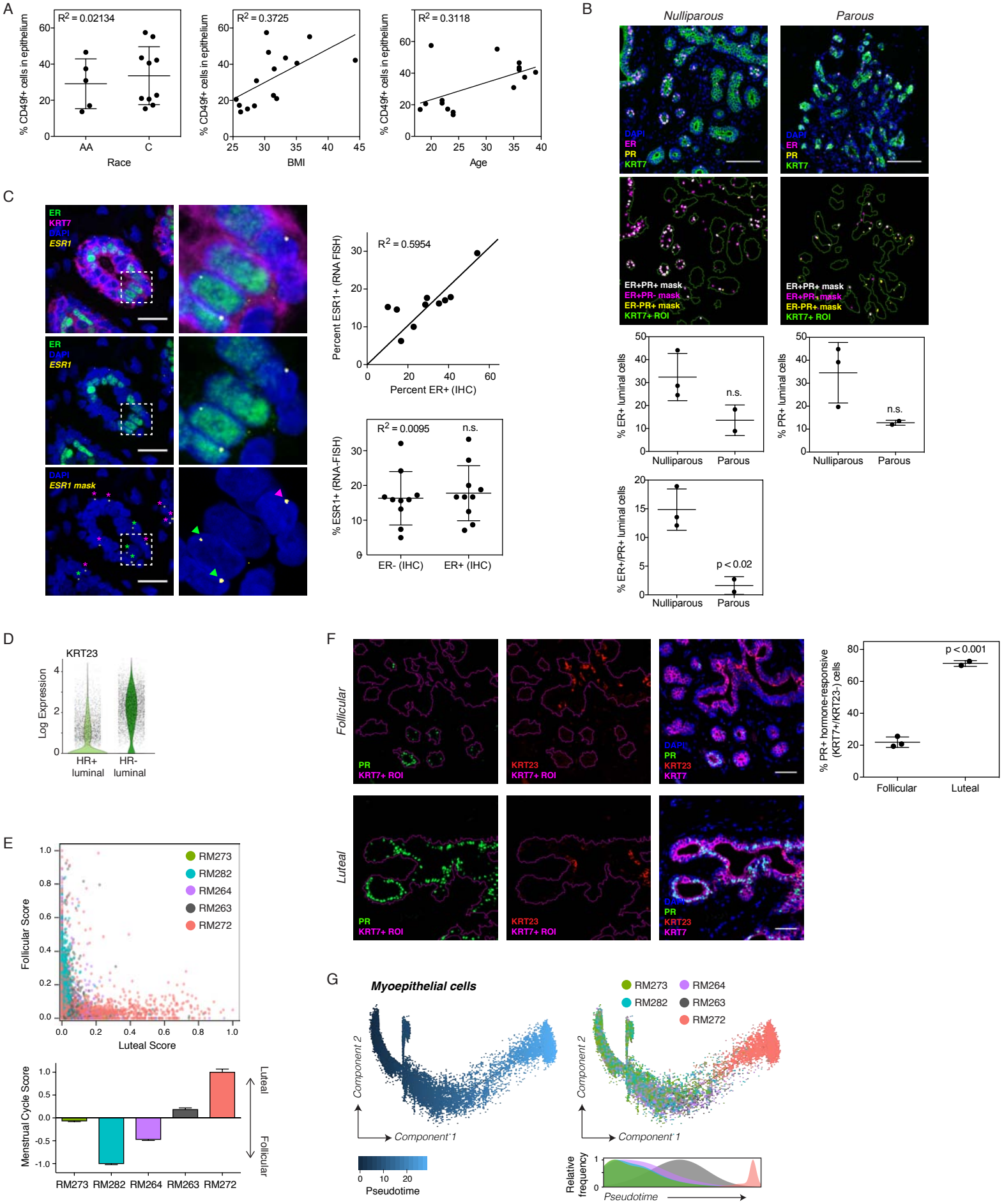

A
R2 = 0.02134
% CD49f+ cells in epithelium
AA
C
Race
R2 = 0.3725
BMI
R2 = 0.3118
Age
B
Nulliparous
Parous
DAPI
ER
PR
KRT7
ER+PR+ mask
ER+PR- mask
ER-PR+ mask
KRT7+ ROI
% ER+ luminal cells
% PR+ luminal cells
% ER+/PR+ luminal cells
n.s.
p < 0.02
C
ER
KRT7
DAPI
ESR1
ESR1 mask
R2 = 0.5954
Percent ESR1+ (RNA FISH)
Percent ER+ (IHC)
R2 = 0.0095
% ESR1+ (RNA-FISH)
ER- (IHC)
ER+ (IHC)
D
KRT23
Log Expression
HR+ luminal
HR- luminal
E
RM273
RM282
RM264
RM263
RM272
Follicular Score
Luteal Score
Menstrual Cycle Score
Luteal
Follicular
F
Follicular
Luteal
PR
KRT7+ ROI
KRT23
KRT7
p < 0.001
% PR+ hormone-responsive (KRT7+/KRT23-) cells
G
Myoepithelial cells
Component 1
Component 2
Pseudotime
Relative frequency

Figure S2

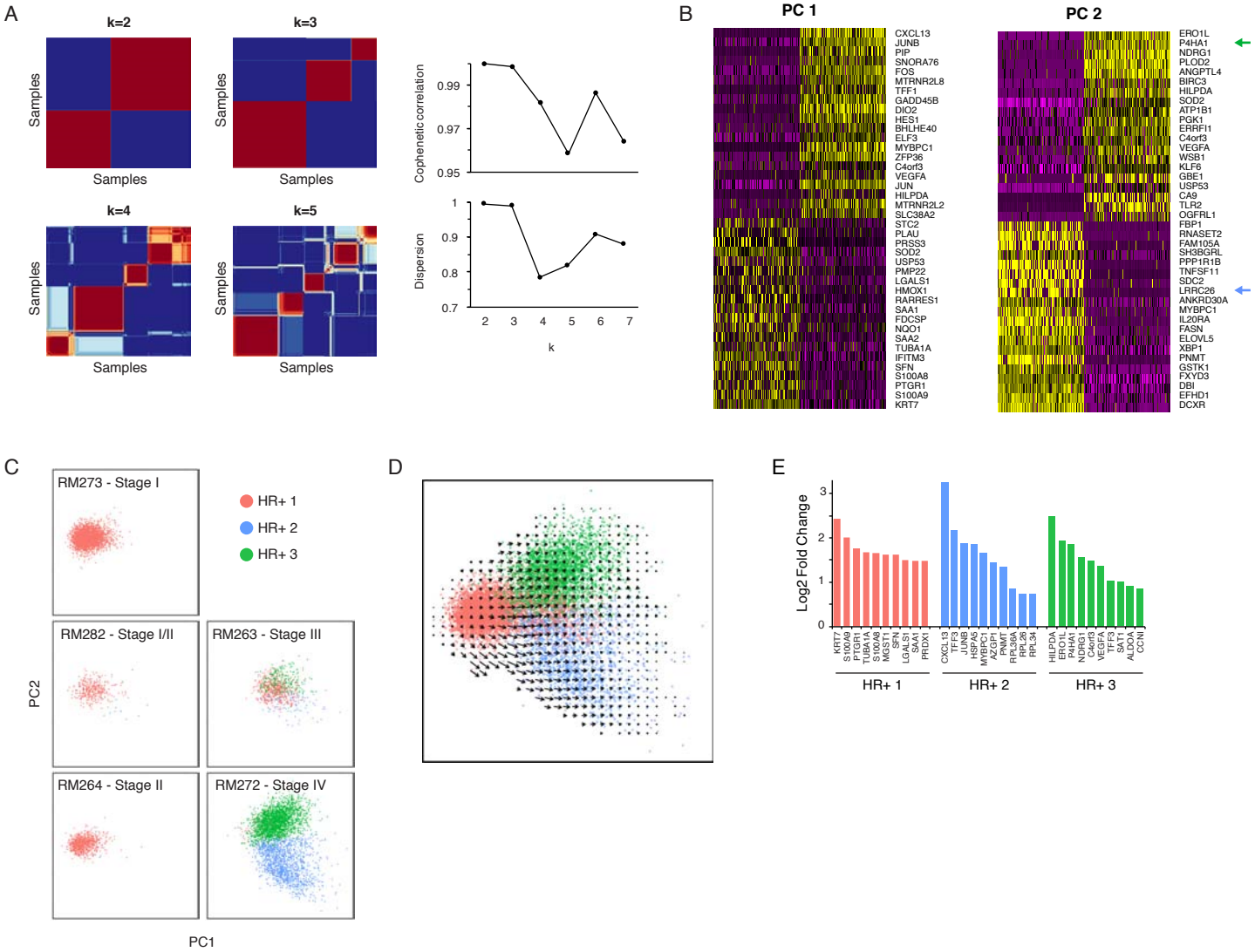

A
k=2
k=3
k=4
k=5
Samples
Cophenetic correlation
Dispersion
k
B
PC 1
PC 2
CXCL13
JUNB
PIP
SNORA76
FOS
MTRNR2L8
TFF1
GADD45B
DIO2
HES1
BHLHE40
ELF3
MYBPC1
ZFP36
C4orf3
VEGFA
JUN
HILPDA
MTRNR2L2
SLC38A2
STC2
PLAU
PRSS3
SOD2
USP53
PMP22
LGALS1
HMOX1
RARRES1
SAA1
FDCSP
NQO1
SAA2
TUBA1A
IFITM3
SFN
S100A8
PTGR1
S100A9
KRT7
ERO1L
P4HA1
NDRG1
PLOD2
ANGPTL4
BIRC3
HILPDA
SOD2
ATP1B1
PGK1
ERRFI1
C4orf3
VEGFA
WSB1
KLF6
GBE1
USP53
CA9
TLR2
OGFRL1
FBP1
RNASET2
FAM105A
SH3BGRL
PPP1R1B
TNFSF11
SDC2
LRRC26
ANKRD30A
MYBPC1
IL20RA
FASN
ELOVL5
XBP1
PNMT
GSTK1
FXYD3
DBI
EFHD1
DCXR
C
RM273 - Stage I
HR+ 1
HR+ 2
HR+ 3
RM282 - Stage I/II
RM263 - Stage III
RM264 - Stage II
RM272 - Stage IV
PC2
PC1
D
E
Log2 Fold Change
HR+ 1
HR+ 2
HR+ 3
F
LRRC26
KRT7
DAPI
LRRC26
P4HA1

Figure S3

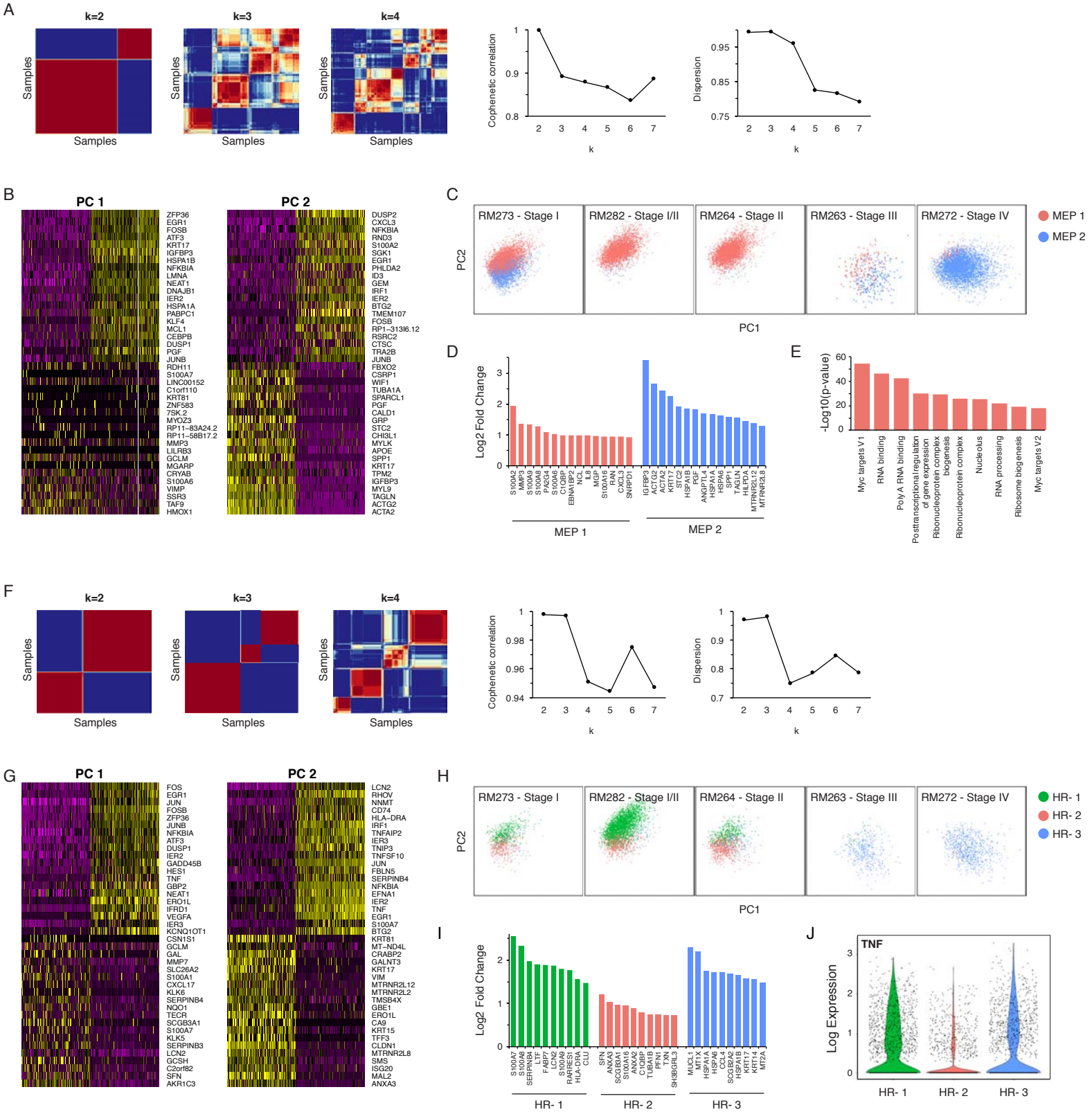

A
k=2
k=3
k=4
Samples
Cophenetic correlation
Dispersion
k
B
PC 1
PC 2
C
RM273 - Stage I
RM282 - Stage I/II
RM264 - Stage II
RM263 - Stage III
RM272 - Stage IV
MEP 1
MEP 2
PC2
PC1
D
Log2 Fold Change
MEP 1
MEP 2
E
-Log10(p-value)
Myc targets V1
RNA binding
Poly A RNA binding
Posttranscriptional regulation of gene expression
Ribonucleoprotein complex biogenesis
Ribonucleoprotein complex
Nucleolus
RNA processing
Ribosome biogenesis
Myc targets V2
F
G
H
HR- 1
HR- 2
HR- 3
I
J
TNF
Log Expression

Figure S4

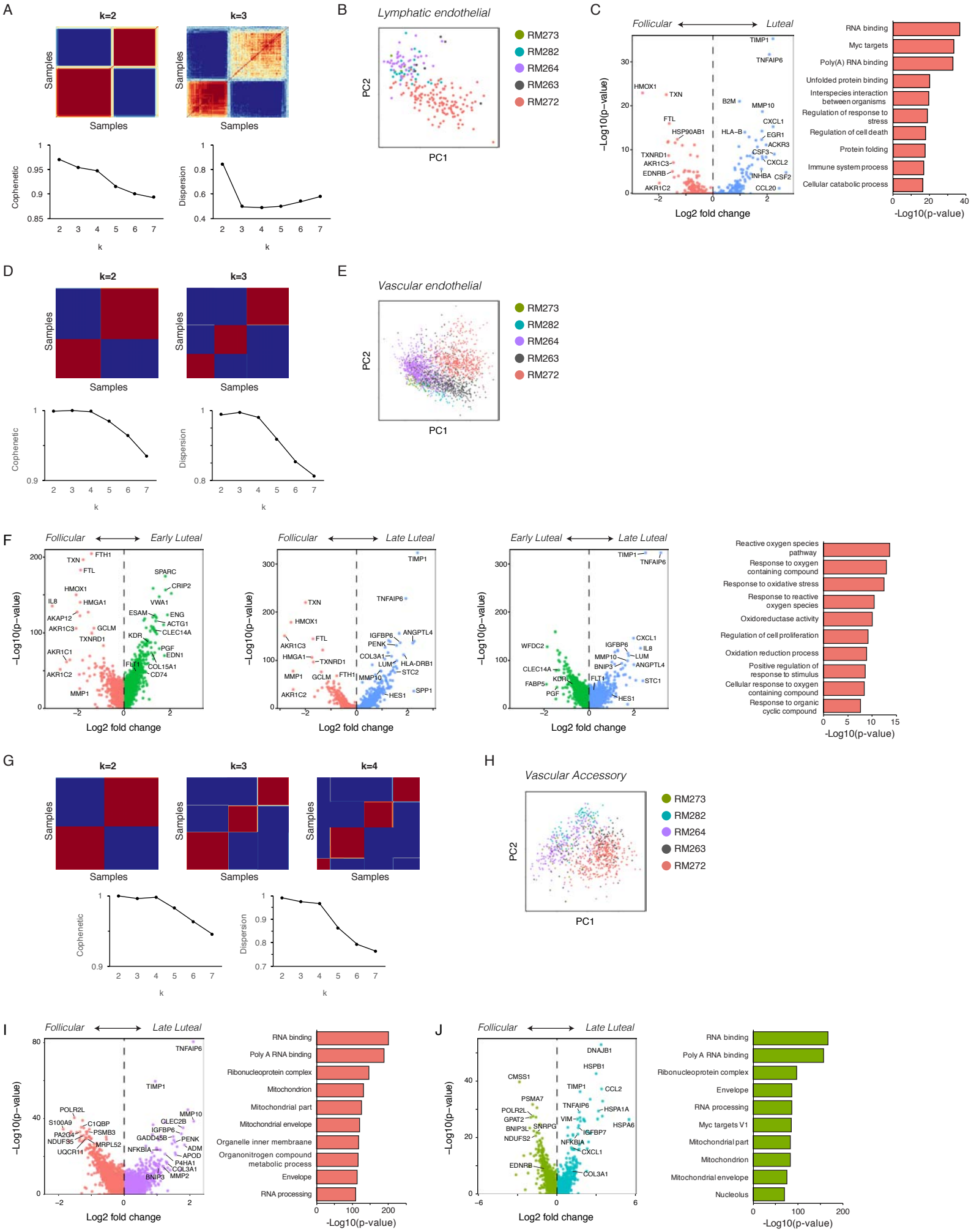

A
k=2
k=3
Samples
Cophenetic
Dispersion
k
B
Lymphatic endothelial
RM273
RM282
RM264
RM263
RM272
PC1
PC2
C
Follicular
Luteal
−Log10(p−value)
Log2 fold change
RNA binding
Myc targets
Poly(A) RNA binding
Unfolded protein binding
Interspecies interaction between organisms
Regulation of response to stress
Regulation of cell death
Protein folding
Immune system process
Cellular catabolic process
-Log10(p-value)
D
E
Vascular endothelial
F
Early Luteal
Late Luteal
Reactive oxygen species pathway
Response to oxygen containing compound
Response to oxidative stress
Response to reactive oxygen species
Oxidoreductase activity
Regulation of cell proliferation
Oxidation reduction process
Positive regulation of response to stimulus
Cellular response to oxygen containing compound
Response to organic cyclic compound
G
k=4
H
Vascular Accessory
I
RNA binding
Poly A RNA binding
Ribonucleoprotein complex
Mitochondrion
Mitochondrial part
Mitochondrial envelope
Organelle inner membraane
Organonitrogen compound metabolic process
Envelope
RNA processing
J
Envelope
RNA processing
Myc targets V1
Mitochondrial part
Mitochondrion
Mitochondrial envelope
Nucleolus

Figure S5

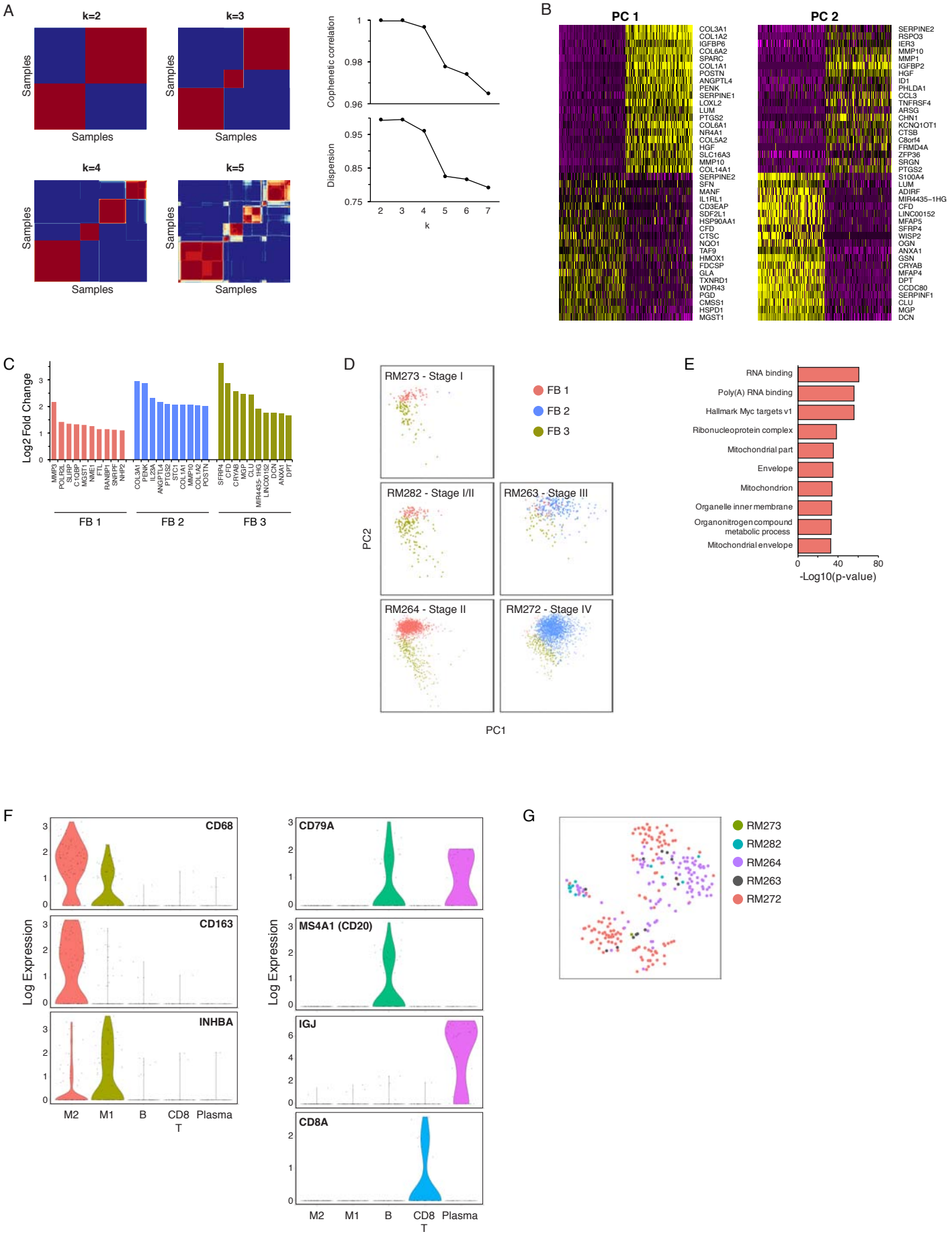
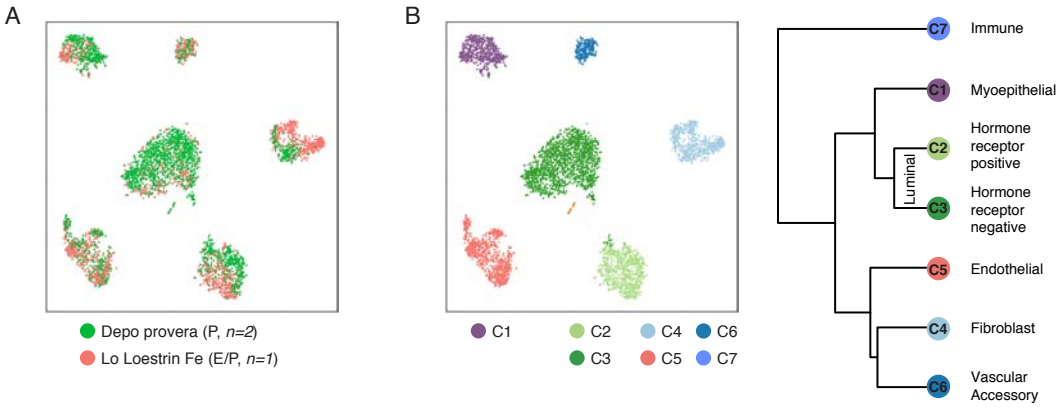
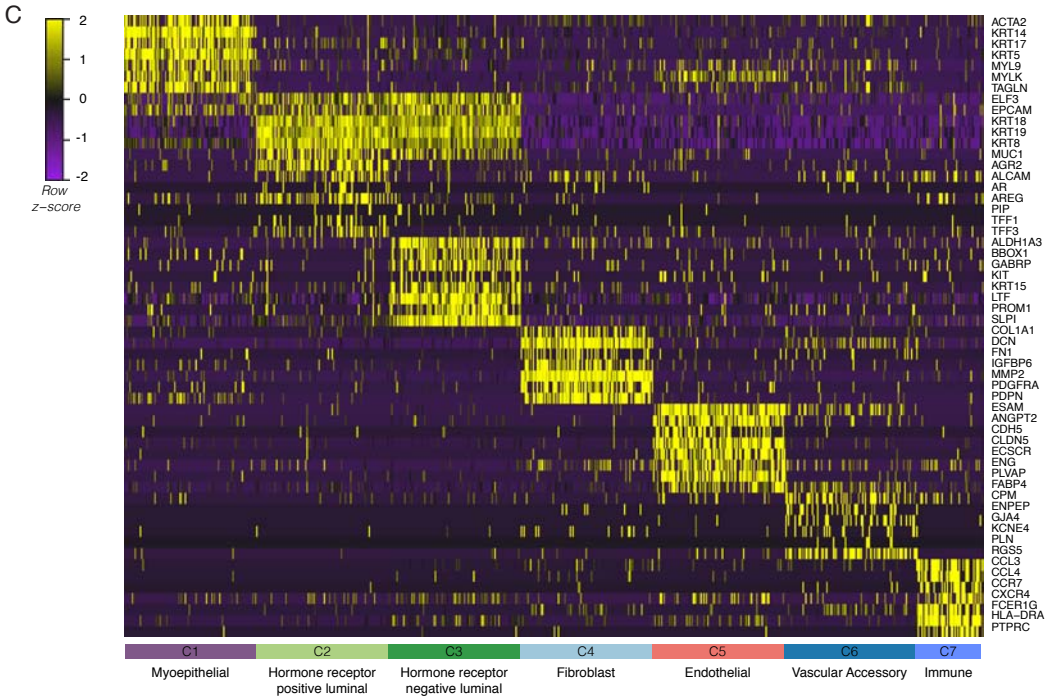

A
k=2
k=3
k=4
k=5
Samples
Cophenetic correlation
Dispersion
k
B
PC 1
PC 2
C
Log2 Fold Change
FB 1
FB 2
FB 3
D
RM273 - Stage I
RM282 - Stage I/II
RM263 - Stage III
RM264 - Stage II
RM272 - Stage IV
PC1
PC2
E
RNA binding
Poly(A) RNA binding
Hallmark Myc targets v1
Ribonucleoprotein complex
Mitochondrial part
Envelope
Mitochondrion
Organelle inner membrane
Organonitrogen compound metabolic process
Mitochondrial envelope
-Log10(p-value)
F
CD68
CD163
INHBA
CD79A
MS4A1 (CD20)
IGJ
CD8A
Log Expression
M2
M1
B
CD8 T
Plasma
G
RM273
RM282
RM264
RM263
RM272

Figure S6


A
Depo provera (P, n=2)
Lo Loestrin Fe (E/P, n=1)
B
C1
C2
C3
C4
C5
C6
C7
Immune
Myoepithelial
Hormone receptor positive
Hormone receptor negative
Luminal
Endothelial
Fibroblast
Vascular Accessory


C
Row z-score
ACTA2
KRT14
KRT17
KRT5
MYL9
MYLK
TAGLN
ELF3
EPCAM
KRT18
KRT19
KRT8
MUC1
AGR2
ALCAM
AR
AREG
PIP
TFF1
TFF3
ALDH1A3
BBOX1
GABRP
KIT
KRT15
LTF
PROM1
SLPI
COL1A1
DCN
FN1
IGFBP6
MMP2
PDGFRA
PDPN
ESAM
ANGPT2
CDH5
CLDN5
ECSCR
ENG
PLVAP
FABP4
CPM
ENPEP
GJA4
KCNE4
PLN
RGS5
CCL3
CCL4
CCR7
CXCR4
FCER1G
HLA-DRA
PTPRC
C1 Myoepithelial
C2 Hormone receptor positive luminal
C3 Hormone receptor negative luminal
C4 Fibroblast
C5 Endothelial
C6 Vascular Accessory
C7 Immune

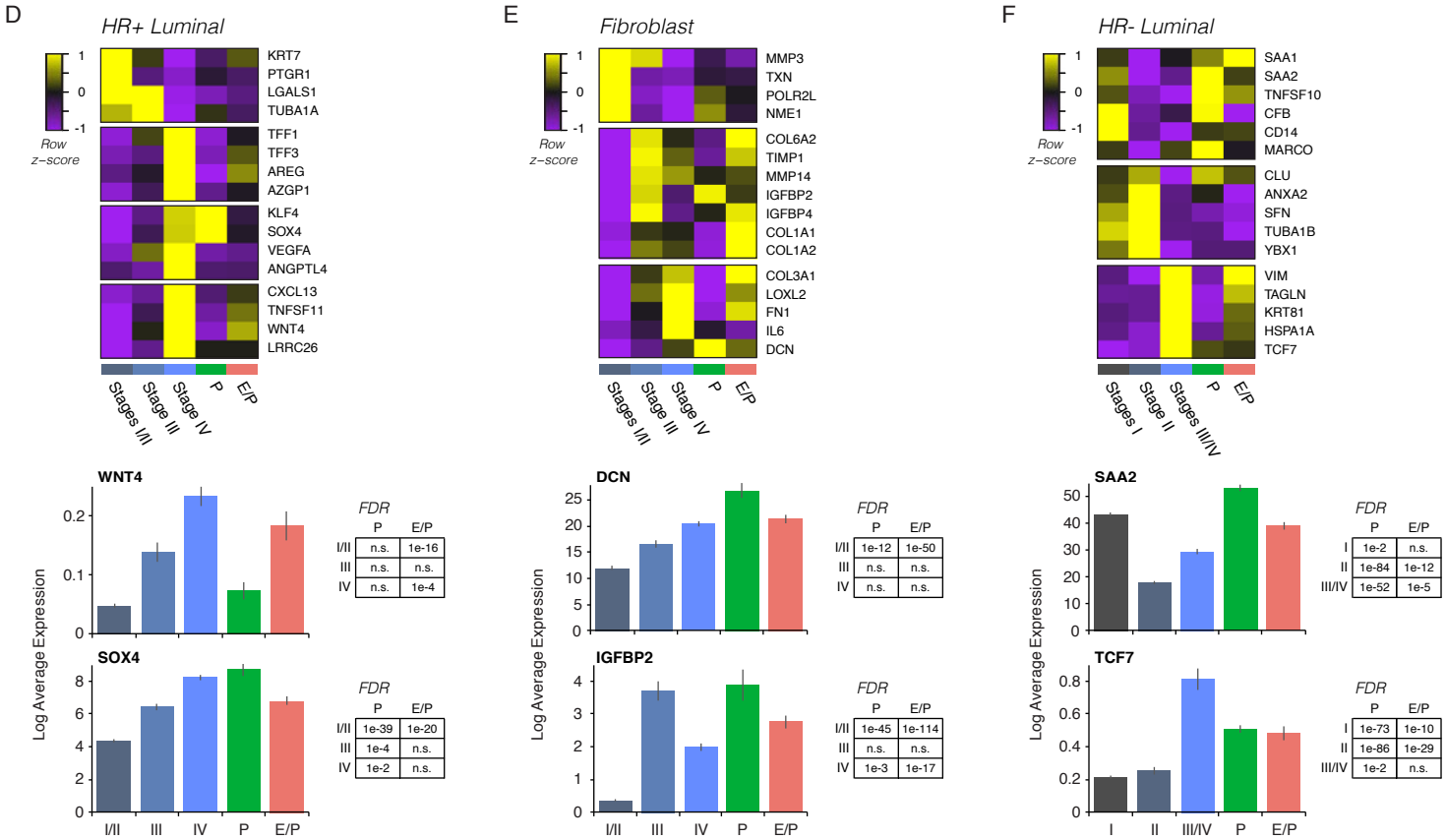

D
HR+ Luminal
Row z-score
KRT7
PTGR1
LGALS1
TUBA1A
TFF1
TFF3
AREG
AZGP1
KLF4
SOX4
VEGFA
ANGPTL4
CXCL13
TNFSF11
WNT4
LRRC26
Stages I/II
Stage III
Stage IV
P
E/P
E
Fibroblast
MMP3
TXN
POLR2L
NME1
COL6A2
TIMP1
MMP14
IGFBP2
IGFBP4
COL1A1
COL1A2
COL3A1
LOXL2
FN1
IL6
DCN
F
HR- Luminal
SAA1
SAA2
TNFSF10
CFB
CD14
MARCO
CLU
ANXA2
SFN
TUBA1B
YBX1
VIM
TAGLN
KRT81
HSPA1A
TCF7
Stages I
Stage II
Stages III/IV
WNT4
SOX4
DCN
IGFBP2
SAA2
TCF7
Log Average Expression
FDR
P E/P
I/II n.s. 1e-16
III n.s. n.s.
IV n.s. 1e-4
I/II 1e-39 1e-20
III 1e-4 n.s.
IV 1e-2 n.s.
I/II 1e-12 1e-50
III n.s. n.s.
IV n.s. n.s.
I/II 1e-45 1e-114
III n.s. n.s.
IV 1e-3 1e-17
I 1e-2 n.s.
II 1e-84 1e-12
III/IV 1e-52 1e-5
I 1e-73 1e-10
II 1e-86 1e-29
III/IV 1e-2 n.s.

Figure S7

Table S1. Donor information for reduction mammoplasty samples. Related to Figure 1.

Sample ID	Age	Race	BMI	Gravidity	Parity	HC
RM263	24	C	27.3	0	0	none
RM264	37	AA	31.5	6	3	none
RM272	23	C	26	0	0	none
RM273	24	AA	26.3	0	0	none
RM282	36	C	44.3	2	2	none

Table S2. Summary statistics for sequencing of five reduction mammoplasty samples. Related to Figure 1.

Sample ID	Sort gate	Number of reads	Percent reads mapped confidently to transcriptome	Estimated cell number	Median genes per cell
RM263	Unsorted	331,865,524	60.7%	2,784	2,765
RM264	Unsorted	665,013,964	58.1%	4,198	1,552
	Luminal	343,078,479	63.2%	2,271	1,207
	Basal	332,878,058	62.0%	2,754	1,469
RM272	Unsorted	656,991,798	54.6%	3,503	1,319
	Luminal	337,334,049	55.2%	4,213	1,351
	Basal	337,758,232	54.8%	5,007	601
RM273	Unsorted	338,792,766	65.8%	1,805	2,590
	Luminal	353,798,722	66.2%	2,501	2,772
	Basal	335,833,078	62.6%	5,142	1,891
RM282	Unsorted	314,274,078	53.6%	2,965	2,359
	Luminal	330,627,133	54.7%	2,772	2,888
	Basal	317,563,455	51.0%	3,106	2,095

Table S3. Microarray data for selected genes from Peri *et al.* Related to Figure 2 and Figure S2.

Gene Symbol	Probe ID	Log_2 FC†	P-value	FDR
KRT5	201820_at	0.41	0.000	0.01
KRT14	209351_at	0.34	0.001	0.02
TP63	209863_s_at	0.32	0.003	0.04
KRT8	209008_x_at	0.20	0.083	0.22
KRT18	201596_x_at	0.16	0.183	0.36
KRT19	201650_at	0.05	0.775	0.87

† Log_2 fold-change parous versus nulliparous samples

Data from: Peri et al. *BMC Medical Genomics* 2012, 5:46.

Table S4. Top 20 differentially expressed genes in the luteal or follicular phases from Pardo *et al*. used to develop a Menstrual Cycle Score for each sample. Related to Figure 2 and Figure S2.

Gene Symbol	Log FC†	P-value	FDR
DIO2	5.05	2.9E-13	6.4E-10
TNFSF11	4.89	1.4E-07	4.1E-05
CXCL13	4.52	2.2E-12	3.9E-09
PTHLH	3.48	1.6E-10	1.4E-07
PNMT	3.31	4.1E-04	3.3E-02
SLC26A3	3.28	1.2E-06	2.5E-04
PTPRN	2.73	5.9E-05	6.9E-03
HIST1H1B	2.72	6.5E-14	2.2E-10
MMP3	2.72	1.2E-08	5.5E-06
TDRD12	2.72	2.4E-04	2.1E-02
ECEL1	2.68	1.5E-06	3.0E-04
MKI67	2.54	7.2E-14	2.2E-10
CYP24A1	2.53	1.2E-07	3.5E-05
TOP2A	2.53	2.5E-16	2.3E-12
HIST1H3C	2.52	1.7E-08	7.5E-06
CDC25C	2.50	4.1E-11	5.7E-08
HMMR	2.45	1.9E-10	1.6E-07
RP1	2.43	5.4E-04	4.0E-02
HIST1H3F	2.41	3.4E-14	1.5E-10
HIST1H2BO	2.40	2.9E-07	7.0E-05

Gene Symbol	Log FC†	P-value	FDR
CRTAC1	-1.20	5.2E-05	6.4E-03
KBTBD11	-1.28	3.8E-05	5.0E-03
HOXB6	-1.41	2.3E-04	2.0E-02
TFCP2L1	-1.54	5.9E-05	6.9E-03
KRT16	-1.64	4.8E-04	3.7E-02
SPAG6	-1.69	1.5E-04	1.5E-02
PI3	-1.70	1.3E-04	1.3E-02
CYP4X1	-1.70	5.3E-05	6.5E-03
SCGB1D2	-1.88	6.4E-04	4.6E-02
CP	-2.04	8.3E-05	9.3E-03
TDRD1	-2.31	8.5E-05	9.5E-03
GLRA3	-2.31	9.5E-05	1.0E-02
PCK1	-2.32	6.2E-05	7.2E-03
HMGCS2	-2.44	2.0E-04	1.8E-02
MUCL1	-2.57	1.2E-04	1.2E-02
ALB	-2.70	8.1E-06	1.3E-03
CYP4Z1	-2.91	9.1E-06	1.5E-03
CSN2	-3.08	3.7E-04	3.0E-02
CPB1	-3.79	1.7E-07	4.7E-05
SCGB2A1	-4.84	1.0E-07	3.3E-05

† Log fold-change luteal-phase versus follicular-phase samples

Data from: Pardo et al. *Breast Cancer Research* 2014, 16:R26.

Table S5. Donor information for additional reduction mammoplasty samples used in immunohistochemical analyses. Related to Figures 3-4.

Sample ID	Age	Race	BMI	Gravidity	Parity	HC	HC format
RM222	19	AA	unknown	0	0	Depo Provera	progestin (medroxy-progesterone acetate)
RM247	22	C	31.9	0	0	TriNessa	combined (ethinyl estradiol/norgestimate)
RM256	32	C	37.1	2	2	Mirena	progestin (levonorgestrel)

Table S6. Donor information for sequenced reduction mammoplasty samples with hormonal contraceptive use. Related to Figure 7.

Sample ID	Age	Race	BMI	Gravidity	Parity	HC	HC format
RM222	19	AA	unknown	0	0	Depo Provera	progestin (medroxy-progesterone acetate)
RM248	19	C	25.5	0	0	Lo Loestrin FE	combined (ethinyl estradiol/norethindrone acetate)
RM249	23	C	41	3	0/1 (unclear)	Depo Provera	progestin (medroxyprogesterone acetate)

Table S7. Summary statistics for sequencing of three reduction mammoplasty samples from donor using hormonal contraception. Related to Figure 7.

Sample ID	Sort gate	Number of reads	Percent reads mapped confidently to transcriptome	Estimated cell number	Median genes per cell
RM222	Unsorted	346,749,732	61.5%	833	2,914
RM248	Unsorted	333,160,029	59.8%	2,075	2,602
RM249	Unsorted	329,434,014	63.3%	1,647	3,036