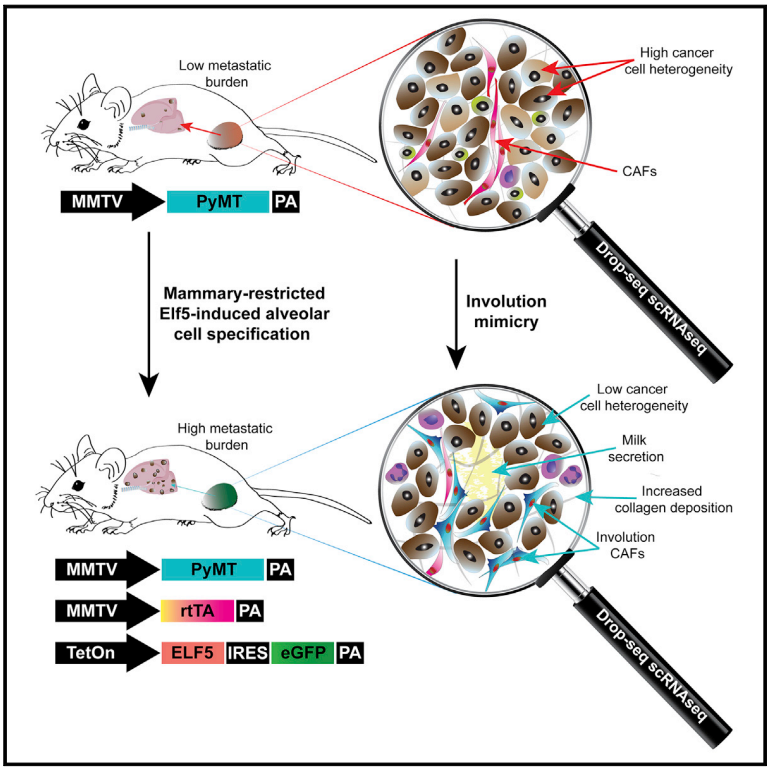
Resource

Cell Reports

Single-cell transcriptomics reveals involution mimicry during the specification of the basal breast cancer subtype

Graphical abstract



Highlights

- Drop-seq reveals a strong cancer cell heterogeneity in PyMT mammary tumors
- Elf5 redefines cancer cell states toward basal-like with pro-metastatic features
- PyMT/Elf5 tumors show cellular and molecular characteristics of involution mimicry
- Involution CAFs in PyMT/Elf5 tumors interconnect the TME and cancer cells

Authors

Fátima Valdés-Mora, Robert Salomon, Brian Stewart Gloss, ..., Paul Timpson, Christopher John Ormandy, David Gallego-Ortega

Correspondence

fvaldesmora@ccia.org.au (F.V.-M.), d.gallego@garvan.org.au (D.G.-O.)

In brief

Using single-cell transcriptomics, Valdés-Mora et al. show that Elf5 drives alveolar lineage differentiation in PyMT tumors, resulting in cancer cell states with basal-like characteristics and a multicellular process of unresolved lactation, with concomitant acquisition of metastasis. This study may partially explain the poor prognosis of breast cancer associated with childbirth.

Valdés-Mora et al., 2021, Cell Reports *35*, 108945
April 13, 2021
https://doi.org/10.1016/j.celrep.2021.108945

CellPress

Cell Reports

CellPress
OPEN ACCESS

Resource

Single-cell transcriptomics reveals involution mimicry during the specification of the basal breast cancer subtype

Fátima Valdés-Mora,[1,2,3,4,10,*] Robert Salomon,[2,4,5,10] Brian Stewart Gloss,[3,6,10] Andrew Man Kit Law,[3,6] Jeron Venhuizen,[6] Lesley Castillo,[6] Kendelle Joan Murphy,[3,6] Astrid Magenau,[6] Michael Papanicolaou,[3,6,7] Laura Rodriguez de la Fuente,[2,3,6] Daniel Lee Roden,[3,4,6] Yolanda Colino-Sanguino,[1,2,3] Zoya Kikhtyak,[6] Nona Farbehi,[4] James Ronald William Conway,[6] Neblina Sikta,[1] Samantha Richelle Oakes,[3,6] Thomas Robert Cox,[3,6] Seán Ignatius O'Donoghue,[1,8,9] Paul Timpson,[3,6] Christopher John Ormandy,[3,6,11] and David Gallego-Ortega[3,4,6,11,12,*]

[1]Genomics and Epigenetics Theme, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia
[2]Personalised Medicine, Children's Cancer Institute, Sydney, NSW 2031, Australia
[3]St. Vincent's Clinical School, Faculty of Medicine, UNSW Sydney, NSW 2010, Australia
[4]Garvan-Weizmann Centre for Cellular Genomics. Garvan Institute of Medical Research, Sydney, NSW 2010, Australia
[5]Institute for Biomedical Materials and Devices, University of Technology Sydney, Sydney, NSW 2007, Australia
[6]Cancer Theme, The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia
[7]School of Life Sciences, University of Technology Sydney, Sydney, NSW 2007, Australia
[8]CSIRO Data61, Eveleigh, NSW 2015, Australia
[9]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Kensington, NSW 2052, Australia
[10]These authors contributed equally
[11]Senior author
[12]Lead contact
*Correspondence: fvaldesmora@ccia.org.au (F.V.-M.), d.gallego@garvan.org.au (D.G.-O.)
https://doi.org/10.1016/j.celrep.2021.108945

SUMMARY

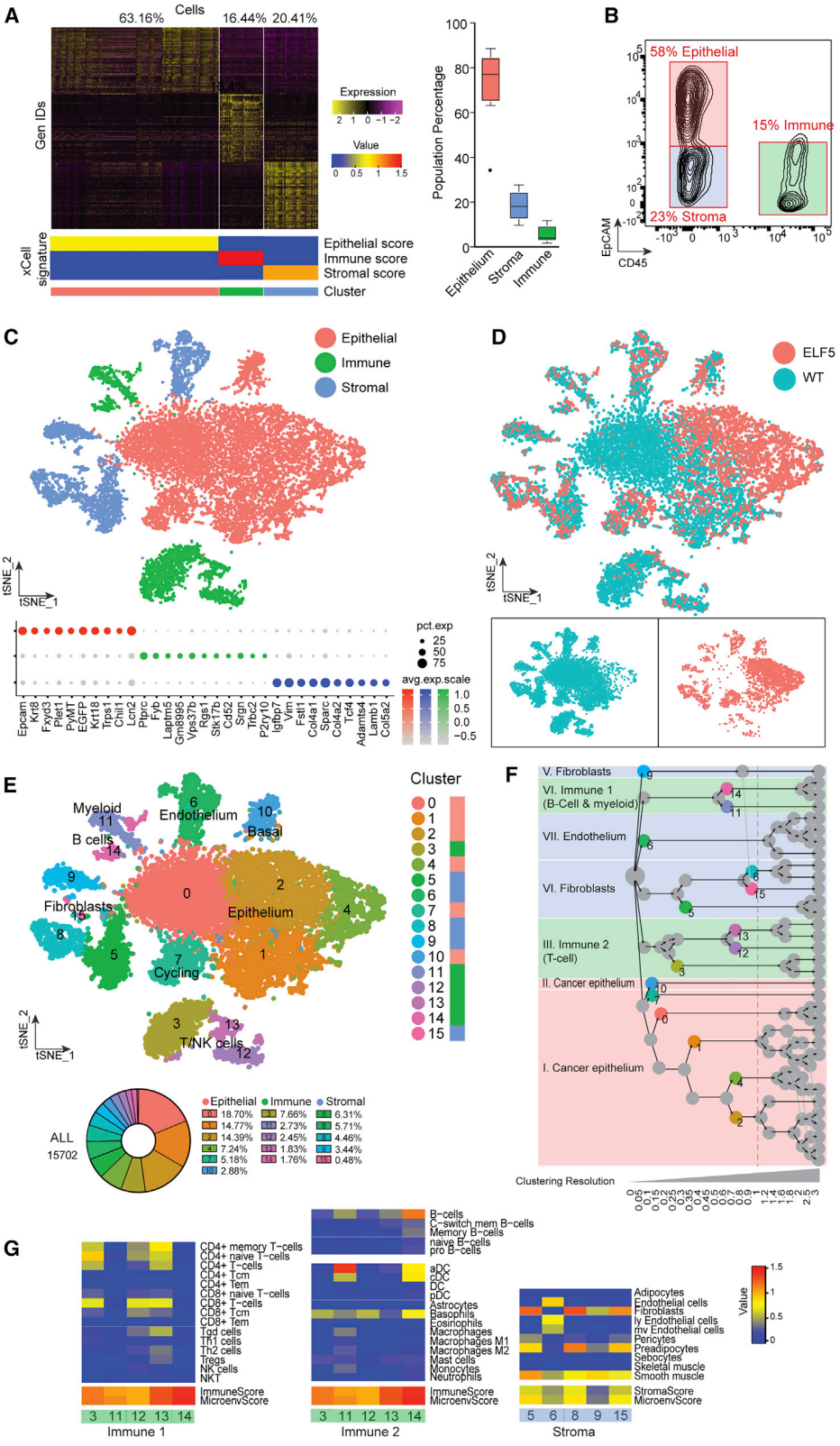
Basal breast cancer is associated with younger age, early relapse, and a high mortality rate. Here, we use unbiased droplet-based single-cell RNA sequencing (RNA-seq) to elucidate the cellular basis of tumor progression during the specification of the basal breast cancer subtype from the luminal progenitor population in the MMTV-PyMT (mouse mammary tumor virus-polyoma middle tumor-antigen) mammary tumor model. We find that basal-like cancer cells resemble the alveolar lineage that is specified upon pregnancy and encompass the acquisition of an aberrant post-lactation developmental program of involution that triggers remodeling of the tumor microenvironment and metastatic dissemination. This involution mimicry is characterized by a highly interactive multicellular network, with involution cancer-associated fibroblasts playing a pivotal role in extracellular matrix remodeling and immunosuppression. Our results may partially explain the increased risk and poor prognosis of breast cancer associated with childbirth.

INTRODUCTION

The mammary gland is a unique organ that mostly develops postnatally, undergoing profound tissue morphogenesis during puberty and pregnancy (Oakes et al., 2014). During pregnancy, the transcription factor ELF5 drives the differentiation of the alveolar milk-secretory epithelium (Gallego-Ortega et al., 2013; Oakes et al., 2006, 2008) from the luminal progenitor cell population (Shehata et al., 2012; Visvader and Stingl, 2014). Transcriptional profiling at single-cell resolution has recently shown, however, that this differentiation process is more complex, with multiple subtypes and states involved in the differentiation of the alveoli during pregnancy (Bach et al., 2017; Nguyen et al., 2018). Different mammary epithelial cell lineages have been proposed as the cell of origin of the breast cancer subtypes (Fu et al., 2020; Lim et al., 2010; Visvader, 2011), including those with specific genetic aberrations (Lim et al., 2010; Molyneux and Smalley, 2011; Visvader and Stingl, 2014).

Tissue morphogenesis associated with pregnancy and lactation concludes with the developmental process of mammary involution, a process whereby the lactating mammary gland returns to a quasi-nulliparous state (Inman et al., 2015; Stein et al., 2007). Involution is an exquisitely regulated multicellular process that requires a high degree of cell-to-cell communication within the mammary epithelium and throughout different cell types from the tissue microenvironment, encompassing innate immune control of inflammation, a pseudo-wound-healing program of activated fibroblasts and intense adipogenesis (Schedin et al., 2007; Watson and Kreuzaler, 2009).

Developmental processes of mammary remodeling are often hijacked in breast cancer as drivers of tumor progression (Fantozzi and Christofori, 2006). Pregnancy-associated breast



(legend on next page)

cancer (PABC) correlates with metastatic disease and poor prognosis and is generally associated with younger women (Bladström et al., 2003; Callihan et al., 2013; Daling et al., 2002; Goddard et al., 2019; Johansson et al., 2011; Schedin, 2006; Stensheim et al., 2009; Whiteman et al., 2004). Despite the effects of the hormonal load associated with pregnancy, there is increasing evidence pointing to a direct effect of the tumour microenvironment (TME) and the events of tissue remodeling that occur in the mammary gland during pregnancy, particularly those associated with involution, as the driving force behind the higher mortality associated with PABC (Lyons et al., 2011; Schedin, 2006; Schedin et al., 2007).

We have previously shown that pregnancy-associated transcriptional networks driven by the transcription factor Elf5 specify the basal subtype of breast cancer, promoting a lethal phenotype, characterized by endocrine insensitivity and resistance to therapy (Kalyuga et al., 2012) and the promotion of metastatic dissemination through a profound alteration of the TME (Gallego-Ortega et al., 2015). Here, we use spatially and time-controlled induction of Elf5 within the epithelial cells of MMTV-PyMT mammary tumors to study tumor progression mechanisms associated with the specification of the basal breast cancer subtype at single-cell resolution. Our analysis revealed intense tumor cell diversity, and despite the majority of cancer cells being of luminal origin, we identified additional cell classes that present markers typically found in myoepithelial and basal lineages. Besides the contribution of cell-intrinsic cues of the alveolar cell differentiation to breast cancer progression (Chakrabarti et al., 2012; Kalyuga et al., 2012), we exposed additional cell-extrinsic effects on the TME (Gallego-Ortega et al., 2015), such as epithelial-mesenchymal transition (EMT), collagen deposition, inflammation, vascular leakiness, and hypoxia. Finally, we uncovered an involution mimicry process associated with tumor progression elicited by alveolar cancer cells. This multicellular mechanism encompasses immune suppression by recruiting of myeloid-derived suppressor cells (MDSCs) and the activation of specific subpopulations of cancer-associated fibroblasts (CAFs).

RESULTS

Unbiased massively parallel single-cell RNA sequencing (scRNA-seq) captures cell heterogeneity of MMTV-PyMT mammary tumors

In this study, we used two mouse models, the MMTV-PyMT mouse mammary tumor model (PyMT/WT) and a mammary restricted (MMTV) doxycycline-inducible (rtTA) Elf5 PyMT model (PyMT/ELF5) (Gallego-Ortega et al., 2015; Guy et al., 1992; Lin et al., 2003; Maglione et al., 2001; Oakes et al., 2008) (Figure S1A). Consistent with previous reports, specific induction of Elf5 in the mammary gland epithelium of PyMT tumors increased lung metastases and induced leaky vasculature (Figures S1B–S1D; Videos S1 and S2) (Gallego-Ortega et al., 2015). Eleven mouse mammary tumors at endpoint (14-week-old FBVn background) from the MMTV-PyMT/WT and PyMT/ELF5 models were harvested and analyzed using Drop-seq (Macosko et al., 2015). Cells passing quality control (QC) filter (15,702) (Figures S2A and S2B) contributed to all cell clusters, indicating consistent sampling between replicates (Figure S2C). As expected, Elf5 expression was higher in PyMT/ELF5 tumors (Figure S2D). The presence of major cell lineages was annotated using gene expression metasignatures (Aran et al., 2017), defining epithelial cells (63.16%), stroma (20.41%), and tumor-infiltrated immune cells (16.44%). These lineages showed consistency among tumor replicates (Figure 1A; Table S1) and were confirmed with cell surface classification by flow cytometry (Figure 1B). A tSNE visualization of the three main cellular compartments and their associated top-expressed and canonical markers is shown in Figure 1C and Figure S3A, respectively. This dataset is accessible in an interactive Shiny application (see Additional resources in the STAR Methods).

Consistent with the role of Elf5 as a master regulator of the alveolar cell specification (Gallego-Ortega et al., 2013; Oakes et al., 2008), induction of Elf5 expression in PyMT tumors generated a profound transcriptional redefinition of the epithelial cell compartment, which was not observed in the other lineages (Figure 1D).

Characterization of the cell diversity of each main lineage based on differential gene expression analysis and k-means clustering (Figure 1E; Table S2) revealed two families in the immune compartment (T/natural killer [NK] cells and B/myeloid cells), three in the stromal compartment (endothelium and several fibroblast classes) and two in the cancer compartment; all of them were defined early in the phylogeny (Figure 1F). The top differential marker genes that define stromal and immune cell compartments re shown in Figure S3B. Such cell diversity was also confirmed using the xCell algorithm (Aran et al., 2017) (Figure 1G).

PyMT cancer cell diversity resembles aspects of different mammary epithelial lineages

To further explore Elf5 action in PyMT tumors, we subset and re-clustered the epithelial cell compartment (6,475 PyMT/WT and

Figure 1. High-resolution cell composition of MMTV-PyMT tumors

(A) Top-expressed genes contributing to the epithelial (red), stromal (blue), and immune (green) clusters and their percentage. Bottom panel: heatmaps for the signatures of the xCell algorithm. Right: boxplot for the distributions of all cells from all replicates in the three main compartments. Error bars show SEM.

(B) Representative contour plot of the cell composition of a PyMT tumor analyzed by flow cytometry, $EpCAM^{+}$ (epithelial cells), $CD45^{+}$ (leukocytes), and $EpCAM^{-}/CD45^{-}$ (stroma).

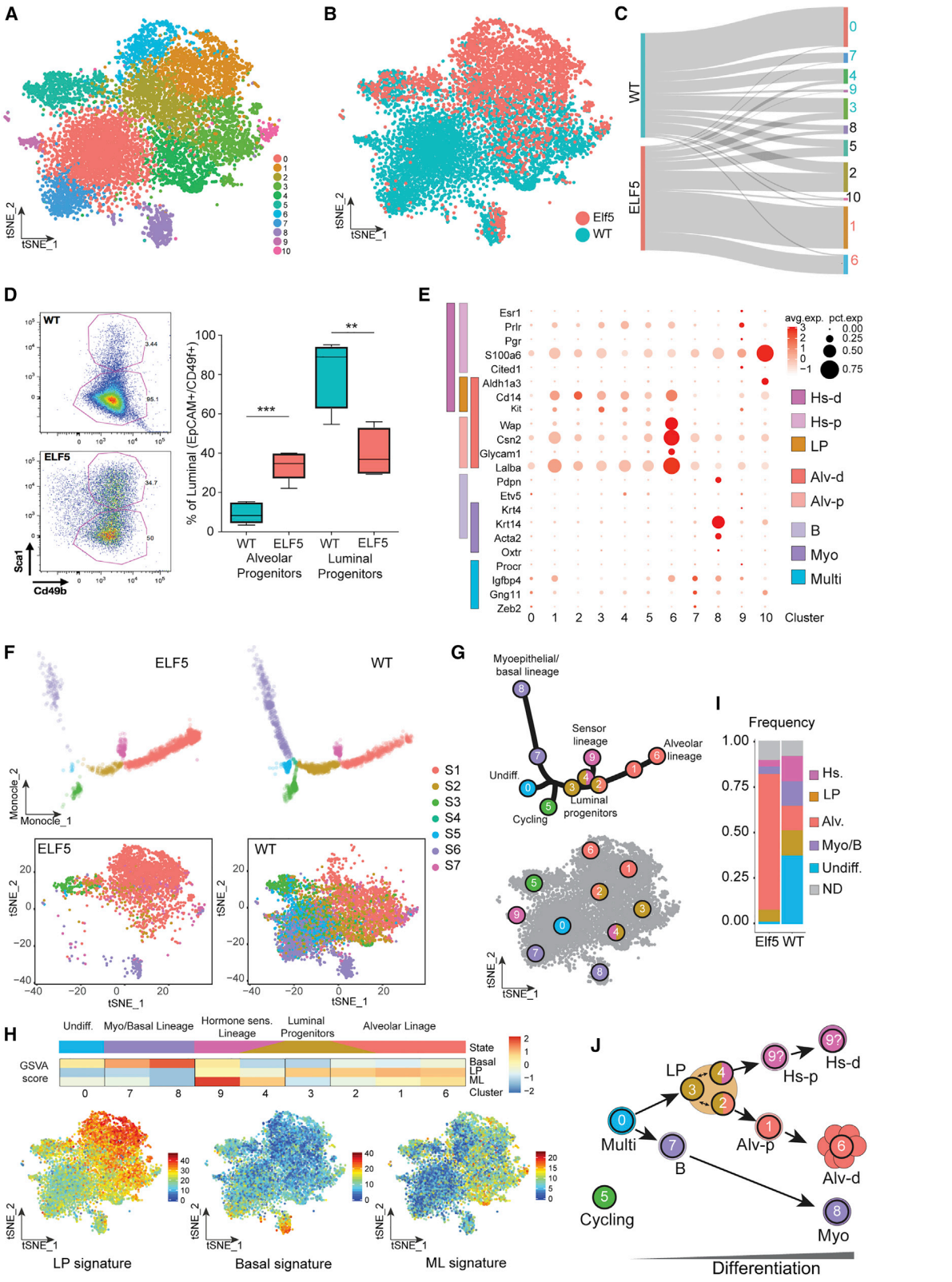
(C) t-distributed stochastic neighbor embedding (tSNE) plot of PyMT tumors colored by main cell lineage. The dot plot shows the top differential markers from each of the main cell lineages.

(D and E) tSNE plot of the distribution of cells per genotype (D) and cell clusters defined by a k-means-based clustering algorithm (E) and their relative frequency (bottom).

(F) Cluster tree modeling the phylogenic relationship of clusters in each cell lineage at increasing resolutions. Dashed red line shows the resolution chosen.

(G) Cell identification score values (xCell algorithm) of the cell clusters in the immune and stromal compartment.

See also Figures S1–S3, Tables S1 and S2, and Video S1.



(legend on next page)

3,233 PyMT/ELF5 cells), revealing intense heterogeneity with 11 epithelial clusters (Figure 2A; Table S2; see also the Shiny application). The strong genotype-driven cell enrichment in the epithelial compartment (Figure 2B) was not due to batch effects (Figure S3C). Epithelial cluster 1 (C1) and C6 were formed mostly in response to forced Elf5 expression, while C0, C7, C9, C4, and C3 were most abundant in wild-type (WT) cells (Figure 2C), a distribution consistent with Elf5 expression in these clusters (Figure S3D). Concomitantly, flow cytometry analysis of cancer epithelial cells revealed that ELF5 forced differentiation of the luminal progenitors (Shehata et al., 2012) within the luminal population (Figure 2D).

We then annotated the cells from our MMTV-PyMT epithelial clusters using previously defined canonical markers of the mouse mammary epithelium (Bach et al., 2017) (Figure 2E). C0 and C7 expressed genes consistent with Procr^{+} multipotent/stem cells. C9 cells expressed hormone-sensing (Hs) lineage markers, C8 expressed markers for basal (B)/myoepithelial (Myo) cells; C1, C2, C3, C4, and C6 presented markers consistent with luminal cells, with C2, C3, and C4 presenting progenitor markers, while C1 and C6 presented markers consistent with differentiated alveolar cells. C5 corresponded with cycling cells (Figure S3E), and C10 showed low PyMT oncogene expression levels (Figure S3D), suggesting an early transformed origin of normal cells trapped within the tumor mass. A complete functional annotation of each epithelial cluster using gene-set variation analysis (GSVA) (Hänzelmann et al., 2013) can be found in Table S3 and Data S1 and S2.

Trajectory analysis reveals that Elf5 drives an alveolar basal-like cell fate on PyMT tumors

To further investigate the dynamic states present the PyMT tumors, we performed pseudotime trajectory analysis along the mammary gland hierarchy gene signature (Pal et al., 2017), identifying seven discrete states (Figure 2F, upper panels). Elf5 had a profound impact upon the structure of epithelial cell diversity, biasing the composition of the cell lineages toward one of the end-fate axes (right-hand side) of the pseudotime plot dominated by state 1 (S1) at the expense of a depletion of S2, S4, S5, and S6. Overlay of the cell cluster identity (C) with the pseudotiming states (S) (Figure 2F bottom panels, and Figure 2G) confirmed that the alveolar lineage state (red, S1) corresponded to two distinct cell identities (C1 and C6); the sensor lineage state (pink, S7) was composed of C9, and the myoepithelial/basal lineage state (purple, S6) was formed by C7 and C8. An undifferentiated state with multipotent/stem characteristics (blue, S5), which correspond to C0, sat in a position equidistant from the myoepithelial and the luminal lineages (alveolar and sensor). Interestingly, and consistent with the literature where stem cell properties are confounded by myoepithelial properties (Prater et al., 2014), C7 was classified as part of the basal/myoepithelial lineage; however, this cluster presents strong multipotent characteristics similar to those present in C0. The luminal progenitor state (brown, S2, represented by C2, C3, and C4) presented subtle differential features within its clusters, presumably due to their strong plasticity.

Trajectory analysis was complemented by GSVA and metasignature visualization (Figure 2H). The luminal progenitor population was derived from CD29^{lo}CD24^{+}CD61^{+} mammary epithelial cells; thus, it contains progenitors of both the Hs and alveolar lineages (Sheridan et al., 2015). These analyses associated C2 with the alveolar lineage, suggesting its identity as a pre-alveolar committed luminal progenitor, or a cell type that sits in the transition zone between the luminal progenitor and the alveolar lineages. Similarly, C4 was associated with the Hs lineage represented by C9, indicating a luminal progenitor committed toward this lineage. In this scenario, C3 remained undefined, so we classified it as an uncommitted luminal progenitor. The distribution of the different designated lineages in the two genotypes (Figure 2I), which showed less heterogeneity in PyMT/ELF5 tumors induced by a strong enrichment of the alveolar lineage, suggesting a skewing toward alveolar lineage differentiation.

Altogether, this analysis identified a large luminal lineage as previously demonstrated in bulk RNA sequencing (RNA-seq) (Lim et al., 2010). However, our single-cell-resolution data revealed further cancer cell diversity and underscored strong cell plasticity (Figure 2J). PyMT/ELF5 tumors, on the other hand, were less heterogeneous and dominated by the alveolar lineage.

Molecular mechanisms of cancer progression associated with cancer cells of alveolar origin

We next sought to validate some of the functional findings found in cancer cells of alveolar origin (C1 and C6), which were characteristic of PyMT/ELF5 tumors.

Figure 2. Cancer epithelial cell diversity of PyMT tumors

(A) tSNE plot of the epithelial cell groups defined by k-means clustering analysis.

(B) Distribution of cells by genotype.

(C) Contribution the genotypes to each cell clusters. Cluster numbers are colored by the dominant genotype (>2-fold cell content of one genotype).

(D) Representative flow cytometry plots of one of the replicates defining the percentage of alveolar and luminal progenitors for each genotype (left) and their quantification (right) (WT, aqua, n = 6; ELF5, red, n = 5; unpaired t test, ***p < 0.001, **p < 0.01); error bars show SEM.

(E) Dot plot representing the expression level (red jet) and the number of expressing cells (dot size) of the transcriptional mammary gland epithelium markers in each PyMT epithelial cluster as defined by Bach et al. (2017), including hormone-sensing differentiated (Hs-d), hormone-sensing progenitor (Hs-p), luminal progenitor (LP), alveolar differentiated (Alv-d), alveolar progenitor (Alv-p), basal (B), myoepithelial (Myo), and undifferentiated (Multi).

(F) Top: pseudotime analysis of the PyMT/WT and /ELF5 cancer epithelial cells along the trajectory of the mammary gland epithelial hierarchy. Bottom: projection of the pseudotime states into tSNE clustering coordinates.

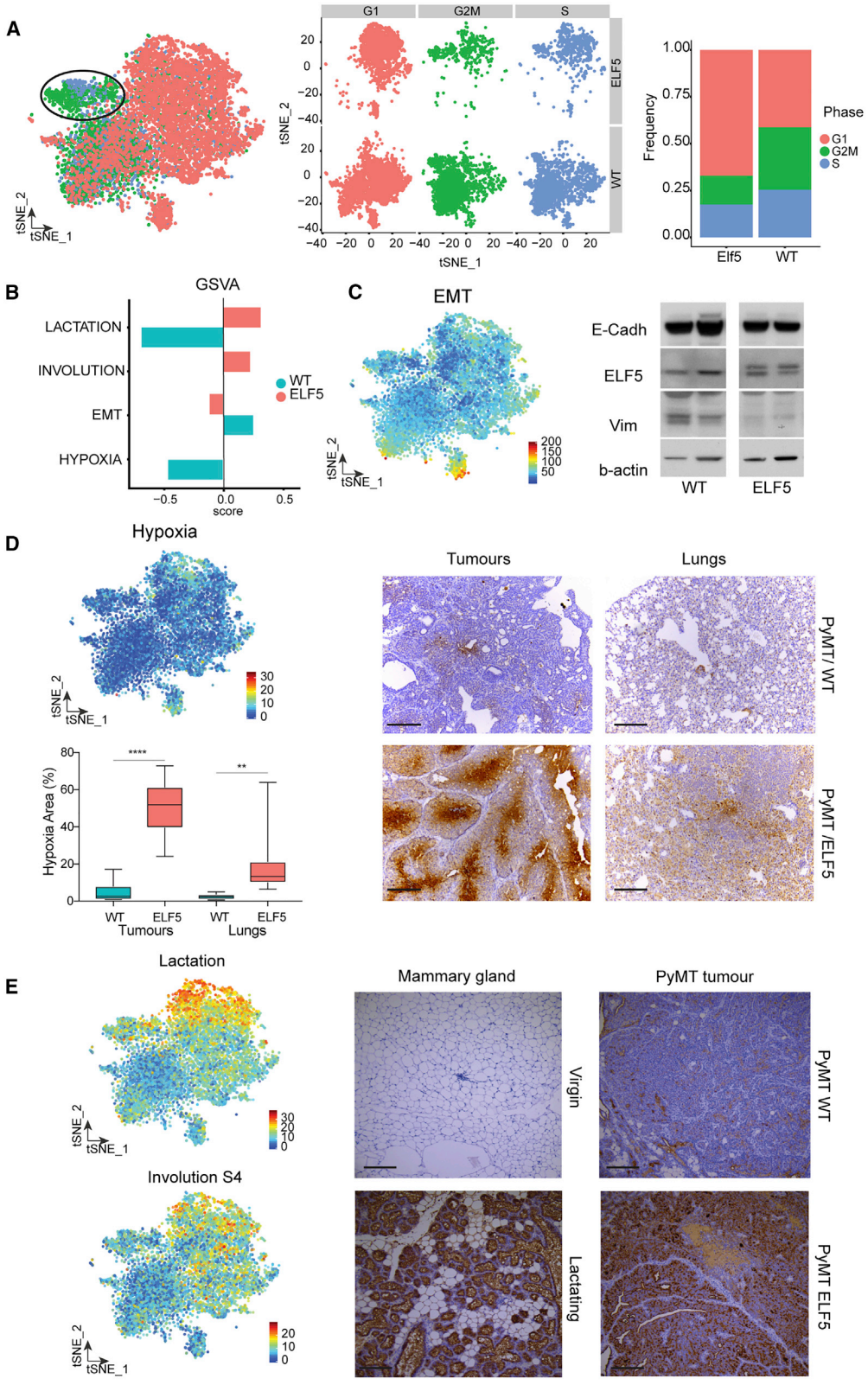
(G) Schematic representation of the overlayed cell states (S) and cluster cell identities (C).

(H) GSVA score for the gene signatures that define the main mammary gland lineages (basal, LP, and mature luminal [ML]) for each of the clusters. Bottom: tSNE representations of each metasignautre for LP, basal, and ML. The top bar shows the assigned mammary epithelial cell type as per (G).

(I) Frequency of the different cell lineages in each genotype.

(J) Illustration of cell diversity of PyMT tumors based on the canonical structure of the mammary gland epithelial lineages.

See also Figure S3, Tables S2 and S3, and Data S2.



(legend on next page)

Cell-cycle marker signatures (Tirosh et al., 2016) identified a specific group exclusively populated by cells in G2M or S phase of the cell cycle (Figure 3A). PyMT/ELF5 tumors showed enrichment in the G1 phase, and this is consistent with our previous reports of cell-intrinsic effects of Elf5 increasing cell-cycle time by delaying entry into S phase (Kalyuga et al., 2012), as well as a report that Elf5 reduced cell proliferation in PyMT tumors (Gallego-Ortega et al., 2015).

GSVA for specific signatures of known functions of Elf5 and functions suggested by the hallmark analysis from Data S2 and Table S3 revealed strong enrichment of lactation, involution, and hypoxia pathways and a decrease in EMT in PyMT/ELF5 tumors (Figure 3B). This was confirmed by an association of the EMT signature restricted to myoepithelial and undifferentiated cells and is consistent with previous reports showing cell-intrinsic effects of Elf5 (Chakrabarti et al., 2012) and a decrease in the mesenchymal marker vimentin in whole lysates of PyMT/ELF5 tumors at the protein level (Figure 3C).

Cells enriched for a hypoxic gene signature concentrated in myoepithelial and alveolar cell clusters (Figure 3D; Data S2; Table S3), and this is consistent the with intense vascular leakiness of PyMT/ELF5 tumors (Figure S1D) (Gallego-Ortega et al., 2015). Immunohistochemistry (IHC) with hypoxyprobe-1 identified extensive hypoxic areas within PyMT/ELF5 tumors, underscoring the extensive tissue remodeling effects of the specification of the basal-like alveolar epithelium (Figure 3D).

As expected, the lactation signature was strongly enriched in cells of alveolar origin (Figure 3E). IHC revealed a strong production of milk in PyMT/ELF5 tumors compared to PyMT/WT tumors and, to a similar extent, a lactating normal mammary gland (Figure 3E). In the normal mammary gland, milk stasis generates the signals that trigger mammary involution. In this scenario, where milk cannot leave the tumor, we identified a late-involution gene signature (Stein et al., 2004) associated with the luminal lineage in PyMT tumors, specifically enriched in Elf5 alveolar cells (Figure 3E).

These results point out to a mechanism of involution mimicry as a plausible driver of the cell-extrinsic effect of Elf5 remodeling of the TME and the acquisition of aggressive traits of tumor progression associated with the transition of the luminal-to-basal subtype of breast cancer.

Characterization of CAFs in PyMT tumors

Involution is a multistep and multicellular process that involves alveolar cell death and tissue remodeling, the latter orchestrated by stromal and immune cells (Watson, 2006). Among these cell types, fibroblasts have a critical role during the extracellular matrix (ECM) remodeling and immune suppression steps (Guo et al., 2017b). Thus, we further explored the role of CAFs in aberrant involution in our model as a potential key event that could fuel Elf5-driven metastasis.

Unsupervised clustering of the fibroblast subset (2,255 cells from PyMT/WT and PyMT/ELF5 tumors) revealed three major fibroblast clusters (Figures 4A, S4A, and S4B; Table S2), including a fibroblast subset from C2 enriched for a G2/M signature (Figure S4C; see also the Shiny application).

CAFs are known to be highly diverse, which was confirmed in our system, as canonical fibroblast markers (Cortez et al., 2014) failed to specify= any particular transcriptomically defined fibroblast cluster (Figure S4D). Previous unbiased scRNA-seq analyses in human tumors have classified two major fibroblast subtypes (Lambrechts et al., 2018; Puram et al., 2017): (1) activated myofibroblasts, involved in tissue remodeling architecture by physical forces; and (2) secretory CAFs, related to ECM synthesis and cyto- and chemokine production (Kalluri, 2016). Fibroblast cluster 0 (CF0) and CF1 were enriched for a secretory CAF signature, while the activated myofibroblast signature was concentrated in CF2 (Figure 4B). Functional annotation using GSVA (Figure S5A) and metasignatures (Davidson et al., 2018) (Figure 4C and D) of the secretory CAFs identified a remodeling function for CF0 (i.e., ECM-CAFs), with a desmoplastic signature of genes involved in ECM interactions, and an inflammatory-related function for CF1 (i.e., iCAFs), featuring genes involved in cytokine-cytokine receptor interactions, monocyte recruitment, and the complement cascade (Davidson et al., 2018). CF2 was enriched for a contractile signature, characterized by the expression of genes involved in actin cytoskeleton organization, and thus identified with myofibroblasts.

This classification is consistent with the recently reported nomenclature from the PyMT model (Bartoschek et al., 2018) (Figures S4E and S4F).

CAFs from PyMT/ELF5 mammary tumors show features of involution mimicry

Unsupervised clustering of the fibroblasts found in PyMT tumors further identified several subclasses of secretory and contractile fibroblasts (Figure 4E; Table S2). Involuting mammary fibroblasts have unique properties that include a high production of fibrillar collagen, ECM remodeling, and immune suppression through monocyte recruitment (Guo et al., 2017b). This specialized class of fibroblasts involved in involution is characterized by the

Figure 3. Orthogonal validation of cancer-related features associated with PyMT/ELF5 tumors

(A) tSNE plot of the cell-cycle stages of the PyMT cancer cells. Circled area shows the cycling cluster. The quantification of the proportion of cells in each stage grouped by genotype is shown in the bar chart.

(B) GSVA analysis of gene expression meta-signatures of cancer-related and Elf5-related hallmarks associated with PyMT/WT (aqua) and /ELF5 (red) tumors.

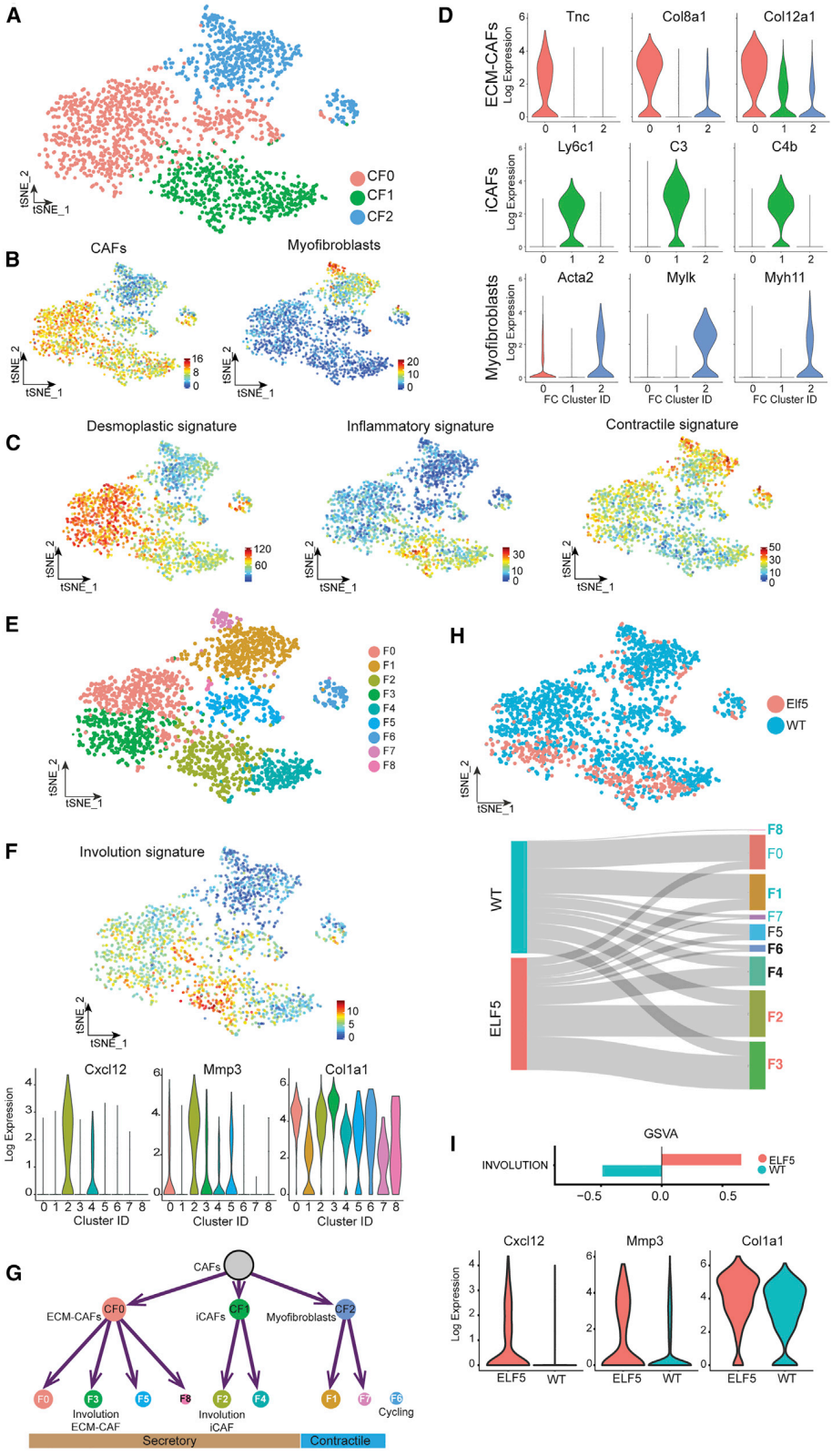
(C) tSNE plot of the EMT metasignature. Right: western blot of canonical EMT markers (E-Cadh, E-cadherin; Vim, vimentin) on PyMT/WT or /ELF5 full tumor lysates.

(D) tSNE plot of the hypoxia metasignature, and boxplot (bottom) of the extension of the hypoxic areas in PyMT/WT (aqua, n = 5) and /ELF5 (red, n = 5) tissue sections (tumors and lung metastasis) stained using IHC based on hypoxyprobe binding. Representative images are shown in the right panels. **$p < 0.01$ and ****$p < 0.0001$. Error bars show SEM.

(E) tSNE plots for lactation and late involution (stage 4 [S4]) metasignatures. Anti-milk IHC pictures from a lactating mammary gland at established lactation compared with a mammary gland from an aged-matched virgin mouse and in PyMT/WT and /ELF5 tumors.

Scale bars (D and E), 200 μm.

See also Figure S1, Table S3, and Data S2.



(legend on next page)

expression of Col1a1, Cxcl12, Tgfb1, and Mmp3. Increased fibrillar collagen and Cxcl12 expression are also considered markers of CAF activation (Sun et al., 2010). We used IHC to confirm the simultaneous upregulation of three of these markers in fibroblasts during mammary gland involution (Figure S6). In PyMT tumors, we identified a group of ECM-CAFs (cluster F3) and iCAFs (cluster F2), with high expression of three of these involution markers (*Col1a1*, *Cxcl12*, and *Mmp3*) in PyMT/ELF5 tumors, (Figure 4F), hereafter referred to as involution iCAFs (F2) and involution-ECM-CAFs (F3). The complete classification of PyMT fibroblasts is shown in Figure 4G. Both involution CAFs were enriched in PyMT/ELF5 tumors (Figure 4H), and this was not due to batch effects (Figure S5B). Consistently, GSVAs of the CAF-involution signature and involution marker genes were significantly enriched in the CAFs from PyMT/ELF5 tumors (Figure 4I).

The functional annotation of the involution fibroblasts revealed specific differences within the ECM-CAF or iCAF subpopulations (Figure S5C). For example, involution ECM-CAFs showed unique fatty acid metabolism and peroxisome hallmarks, suggesting that these cells may correspond to adipocyte-like fibroblasts (Bochet et al., 2013). Involution iCAFs showed pathways linked to a wound-healing process, including inflammation through the complement cascade and coagulation (Amara et al., 2008; Lilla et al., 2009; Stein et al., 2004). All of these processes of involution CAFs commonly occur during late involution (Martinson et al., 2015; Watson, 2006; Zwick et al., 2018).

We validated the role of CAFs in ECM remodeling by measuring the deposition of collagen in tumor sections using microscopy. PyMT/ELF5 tumors showed increased fibrillar collagen coverage compared with PyMT/WT tumors (Figures 5A and 5B), with a significantly higher proportion of thicker and mature collagen fibers than (Figure 5C), and a more complex spatial arrangement (peak alignment ±10 degrees from peak) (Figure S7A).

Subsequently, we histologically mapped the location of the involution CAFs in PyMT tumors by IHC (Figure 5D) using the proposed involution-CAF markers transcriptionally identified (*Col1A1*, *Mmp3*, and *Cxcl12*) (Figure S7B). We used an antibody only recognizing monomer alpha-1 type 1 collagen prior fiber assembly, thus labeling the collagen-I-producing cells. Both COL1A1- and CXCL12-expressing fibroblasts were infiltrated throughout the cancer epithelial foci and organized in clusters, especially in PyMT/ELF5 tumors, where these cells were also more abundant (Figures 5D and S7C). MMP3 IHC staining shows positive expression in both intra-foci stromal cells and adjacent epithelium but negative expression in epithelial areas away from the stroma. This gradient is consistent with a paracrine effect of MMP3 secretion from involution CAFs (Figures 5D and S7C), and it was more noticeable in the PyMT/ELF5 tumors compared to the PyMT/WT counterparts.

These results functionally and histologically characterize the increased presence of involution CAFs in PyMT/ELF5 tumors, confirming the correlation between *Col1a1*, *Mmp3*, and *Cxc12* mRNA and their protein levels and their utility to identify an involution process.

Involution signature predicts for poor prognosis in luminal breast cancer patients in the context of Elf5 expression

We have previously shown that luminal breast tumors with higher expression of *ELF5* show basal-like characteristics and poor prognosis (Gallego-Ortega et al., 2015); thus, these can be considered the early steps toward the differentiation into a basal tumor, underscoring ELF5 as a driver of the basal-like phenotype. In fact, high *ELF5* expression is associated with basal breast cancers (Kalyuga et al., 2012).

In the METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) dataset (Curtis et al., 2012), we confirmed a significant association between high *ELF5* expression and poorer patient prognosis (Figure 6A), where a higher proportion of the basal molecular subtype fell into the *ELF5*-high group. Similarly, an involution signature predicted for poor prognosis independent of other pathological characteristics such as Ki67, basal subtype, and age (<40 years) (Figures 6A and 6B). Consistent with our hypothesis, patients with higher expression of *ELF5* showed significantly higher levels of expression of the involution signature, and this was also confirmed in The Cancer Genome Atlas (TCGA) dataset (Berger et al., 2018; Cancer Genome Atlas, 2012; Ciriello et al., 2015) (Figure 6C). Interestingly, in luminal patients, involution mimicry predicted poor prognosis only in the patient subset with high *ELF5* expression and not in the low-expressing subset (Figure 6D). Taken together, these results highlight the relationship between *ELF5* expression and the involution process in human breast cancer.

Characterization of the cell-to-cell communication involved in involution mimicry

Involution is a multicellular process; thus, we used CellphoneDB to identify cell-to-cell communication networks within PyMT tumors

Figure 4. Cancer-associated fibroblast (CAF) diversity in PyMT tumors

(A) tSNE plot defined by k-means clustering analysis of the fibroblast subpopulation as defined in Figure 1E (prefix CF).

(B) tSNE plots of the metasignatures of the secretory CAFs and myofibroblasts.

(C) Desmoplastic, inflammatory, and contractile metasignatures.

(D) Violin plots displaying marker genes for each of the three fibroblast clusters defined in (A): ECM-CAFs (CF0), immune-CAFs (iCAFs, CF1), and myofibroblasts (CF2).

(E) tSNE plot defined by unsupervised clustering analysis showing heterogeneity of each of the identified fibroblast species.

(F) tSNE plot of the involution metasignature. Bottom: violin plots on the nine fibroblast clusters for individual genes from the involution signature.

(G) Cell tree classification of the main lineages of CAFs.

(H) tSNE plot for the distribution of fibroblasts by genotype (ELF5, red; WT, aqua). Bottom: contribution of each of the genotypes to the cell clusters. Cluster numbers are colored by the dominant genotype (>2-fold cell content of one genotype).

(I) GSVA enrichment analysis of involuting mammary fibroblast metasignatures associated with PyMT/WT and /ELF5 tumors. Bottom: violin plots of *Cxcl12*, *Mmp3*, and *Col1a1* genes in all fibroblasts of each genotype.

See also Figures S4–S6 and Table S2.

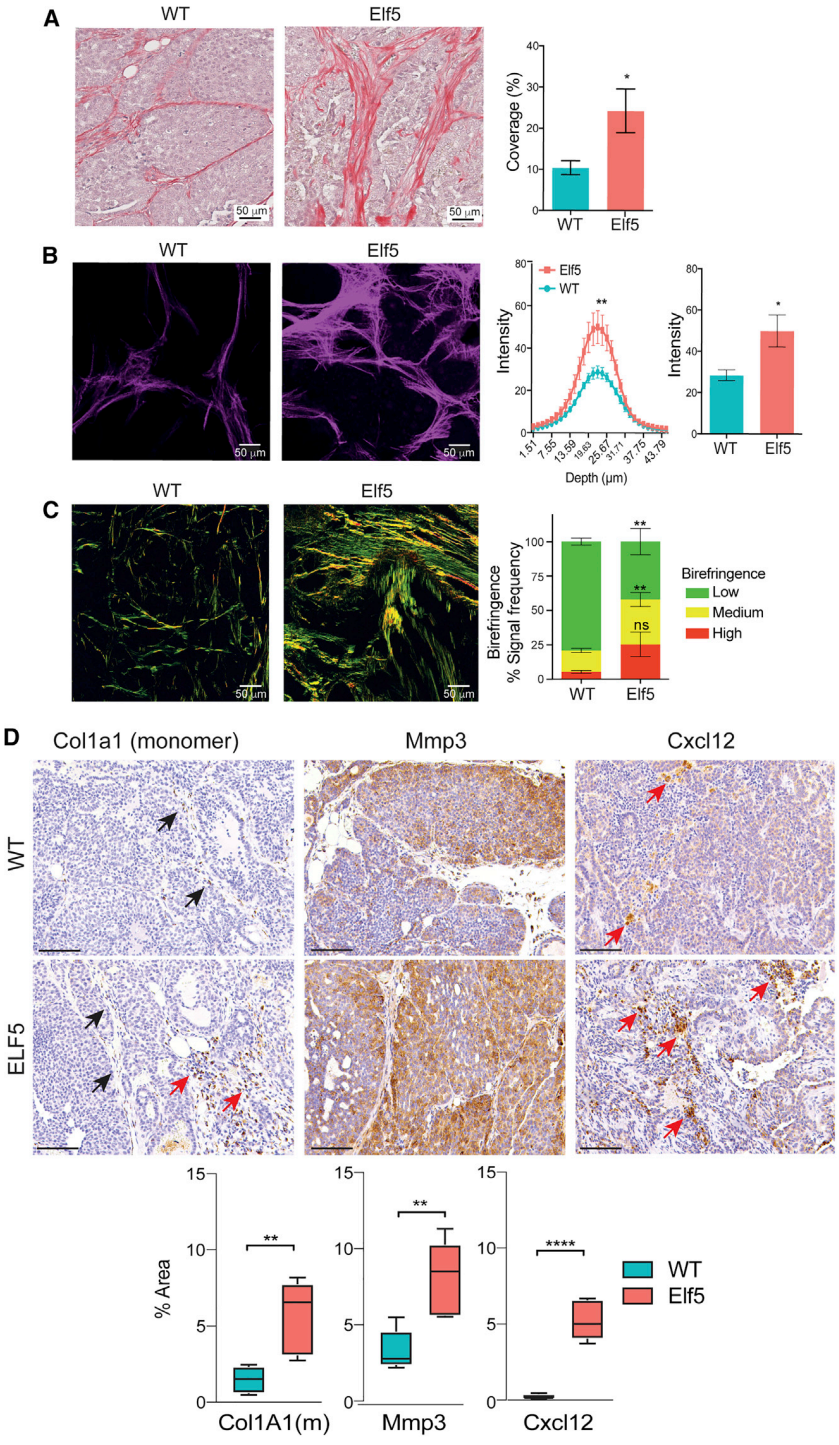


Figure 5. Orthogonal validation of involution CAFs in PyMT/WT and PyMT/ELF5 tumors

(A) Representative bright-field images and bar graphs of quantification of total coverage of picrosirius-red-stained PyMT/WT and PyMT/ELF5 tumor sections. n = 4 mice per genotype with 10 regions of interest (ROIs) per tumor.

(B) Representative maximum intensity projections of second harmonic generation (SHG) signal and quantification of SHG signal intensity at depth (μm) (left graph; paired Wilcoxon test) and at peak (right graph) in PyMT/WT and PyMT/ELF5 tumor sections. n = 6 mice per genotype with 6 ROIs per tumor.

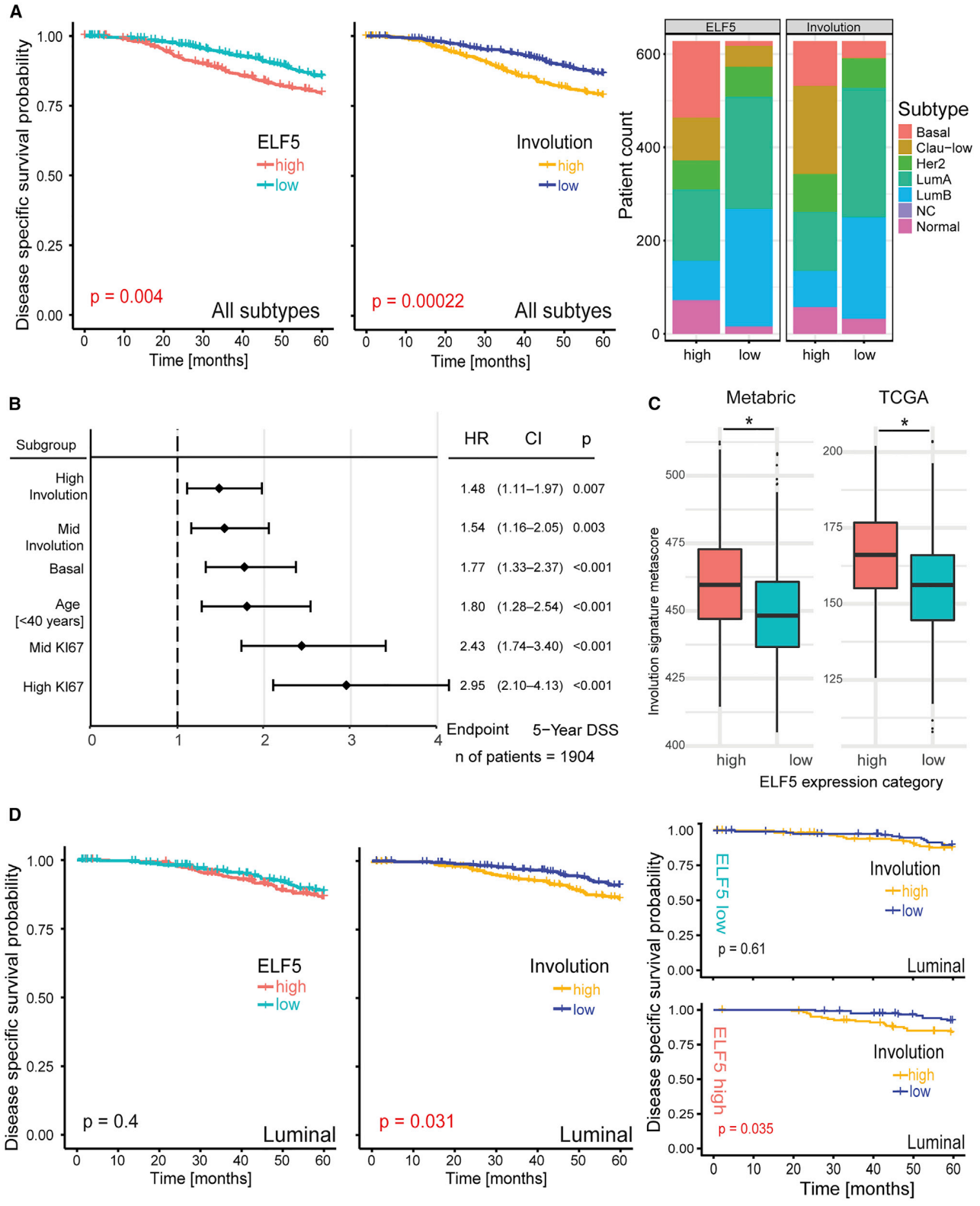
(C) Polarized light imaging of picrosirius-red-stained PyMT/WT and PyMT/ELF5 tumor sections and its quantification. Thick remodeled fibers/high birefringence (red-orange), medium birefringence (yellow), and less remodeled fibers/low birefringence (green) are shown. n = 4 mice per genotype with 10 ROIs.

(D) Representative images (top panels) of the IHC analysis of COL1A1, MMP3, and CXCL12 in PyMT/WT and PyMT/ELF5 tumors and their quantification as percentage of area stained (bottom boxplots). Black arrows show positive staining on elongated cells while red arrows show positive staining in rounded cells. Data correspond to n = 5 tumors with at least five images per tumor; Scale bars, 100 μm. *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001. Scale bars (A–C), 50 μm. Error bars show SEM.
See also Figure S7.

(Vento-Tormo et al., 2018). CellphoneDB analysis using all cell clusters defined in this study (Figures 1, 2, and 4; summarized in Figure S5D) in PyMT tumors revealed a complex interactome network of cell interactions (Figure 7A). Cellular interconnections were ranked in three strength levels: (1) fibroblast-fibroblast interactions were the strongest, followed by fibroblast-epithelium connections; (2) endothelium-fibroblast or -epithelium and epithelium-epithelium interactions were ranked as mid interactions; and (3) the weakest connections were found in the immune cell compartment. A complete classification of the predicted cell-to-cell interactions can be explored using our HTML-based interactive tool (https://galdeslab.github.io/CellReportsManuscript/). These results show a pivotal role of CAFs as hubs of communication within the TME.

Involution accounts for an interactive cell network where epithelial cells, fibroblasts, and immune cells are the key players (Watson, 2006). In our model, ECM-CAFs (both involution and ECM-CAFs 2) and involution iCAFs showed the strongest connections (Figure 7B, 1st and 2nd tier, blue lines) with the two Elf5-enriched-alveolar epithelial cells clusters, myoepithelial cells, and luminal progenitor Hs cells. In the immune interactome (Figure 7B, 3rd and 4th tiers, green lines), the myeloid compartment showed the highest number of connections, communicating with the two Elf5-enriched-alveolar epithelial clusters, ECM-CAFs (both involution and ECM-CAFs 2), involution iCAFs, myofibroblasts, and the endothelium. Involution is associated with the M2-like innate response; thus, these results are in line

CellPress
OPEN ACCESS



(legend on next page)

with an activated and interactive innate myeloid response (O'Brien et al., 2010; Stein et al., 2004, 2007).

Further investigation of particular ligand-receptor connections among CAFs, alveolar epithelial cells, and myeloid cells (Figure 7C) revealed pathways associated with the multi-functional aspects of involution (transforming growth factor β [TGF-β] axis) (Bellomo et al., 2016; Flanders and Wakefield, 2009), networks associated with an immune suppressive ecosystem (Cxcl12 and Dpp4 axis) (Costa et al., 2018; Guo et al., 2017b), and ECM remodeling (insuline growth factor [IGF] axis) (Blackstock et al., 2014).

The myeloid cell compartment is also intimately associated with the process of involution. We have previously demonstrated that despite their low abundance (Valdés-Mora et al., 2018), tumor-infiltrated MDSCs promote lung metastasis in PyMT/ELF5 tumors (Gallego-Ortega et al., 2015). Our scRNA-seq analysis mapped the cellular mechanism for MDSC expansion, recruitment, and malignant activation previously described in these tumors (Sevko and Umansky, 2013), identifying the cell of origin of well-known signaling pathways involved in this process (Figure 7D). Our results show that MDSC recruitment and malignant activation in PyMT/ELF5 tumors (Gallego-Ortega et al., 2015) are driven not only by alveolar cancer cells but also by a complex molecular intercellular network. Tumor-infiltrated myeloid species maintain a feedback loop that enables continuous myeloid cell recruitment, via *Csf3* (G-CSF) and *Vegfa*, and sustained inflammation with molecules, including *Cxcl1* (also expressed by other cell compartments) and *Cxcl2*. CAFs also support MDSC expansion through *Csf1* (M-CSF), *Vegfa*, *Il6*, *Cxcl12*, and *Ccl2* expression and the endothelium with additional *Il6* expression.

In summary, the alveolar lineage supports tumor progression, orchestrating a molecular network of cell-to-cell communication associated with an inflammatory process of involution mimicry with malignant CAF activation and immune suppression via MDSC recruitment (Figure 7E).

DISCUSSION

The MMTV-PyMT model gives rise to mammary tumors of luminal progenitor origin as classified by conventional transcriptomics (Lim et al., 2009, 2010). Our findings, however, underscore the expansive plasticity of the luminal progenitor compartment, identifying "malignant states" associated with the mammary epithelial cell lineage specification. Our results unite tumor heterogeneity with developmental cues and highlight the intrinsic capacity of tumors to develop distinct molecular subtypes, supporting the current understanding of progenitor cells as the cell of origin of basal and luminal cancers (Visvader, 2011).

Here, we study the functional consequences of the specification of the estrogen receptor negative (ER-) basal subtype of breast cancer through the differentiation from the luminal progenitor cells toward the pregnancy-derived alveolar lineage (Lee et al., 2013; Oakes et al., 2008). Elf5-specified alveolar cancer cells subsequently activate cell-extrinsic mechanisms within the TME reminiscent of mammary morphogenesis, which were ultimately associated with the acquisition of tumor progression cues linked to cancer dissemination, similar to pregnancy-induced metastasis in the PyMT/WT model (Gallego-Ortega et al., 2015).

Clinically, patients with luminal tumors receive endocrine therapy as standard-of-care; thus, adoption of the alveolar fate provides a route of escape from endocrine therapy, as these cells are insensitive to estrogen and subsequently give rise to a basal-like tumor (Kalyuga et al., 2012), with much lower survival rates (Gallego-Ortega et al., 2015). This is consistent with the identification of Elf5 as the specific marker for a secretory mature, differentiated luminal cell type, and their associated gene signatures were highly represented in the basal-like (triple-negative breast cancer) TNBC patients (Nguyen et al., 2018). We have found that expression of *ELF5* is characteristic of luminal cancers that are in transit to a basal-like subtype; thus, breast cancer patients who show higher *ELF5* expression activate involution mimicry pathways that are associated with poor prognosis. Also, the pro-metastatic effects of Elf5 are concomitant with mimicry of mammary involution. In this line of evidence, involution has been previously implicated in the reduced survival of PABC (Lyons et al., 2011); however, the molecular drivers of this process had not been fully identified.

Our Elf5-restricted MMTV-PyMT model reproduces the particular effect of pregnancy on the specification of the alveolar cell lineage in the context of cancer, shedding light upon the molecular mechanisms involved in these effects of pregnancy in breast cancer. Mammary involution is a wound-healing-like process that includes tissue remodeling with fibroblast activation to reorganize the ECM and accommodate the epithelial reduction and myeloid infiltration to remove all cell fragments (Inman et al., 2015; Stein et al., 2007). Our data suggest that induced Elf5 expression in cancerous cells stimulates milk production, which in turn accumulates and triggers an inflammation-mediated TME remodeling associated with an aberrant involution. This mimicry involution process results in the loss of alveolar cells that are rapidly replaced by Elf5-driven differentiation of luminal progenitors and a chronic induction of inflammation by involution iCAFs and myeloid cells encompassing ECM remodeling driven by

Figure 6. Involution signature is a poor prognosis factor and is associated with *ELF5* expression in breast cancer patients

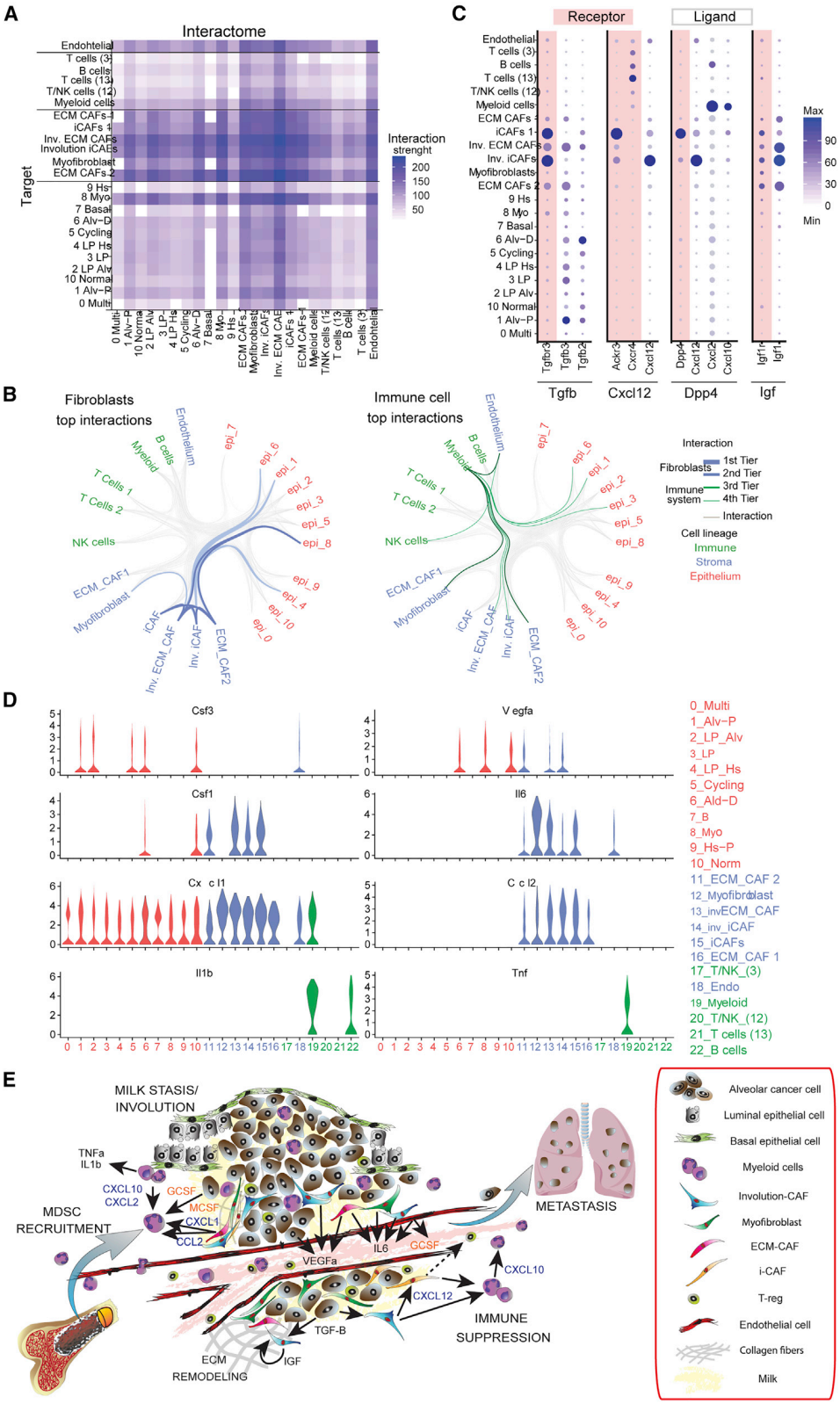
(A) Kaplan-Meier survival curves from the METABRIC patient cohort in relation to *ELF5* and involution signature expression. High and low classifications are based on top and bottom tertiles. Right panel shows the distribution of the PAM50 classification in the top and bottom tertile of *ELF5*- or involution-expressing patients.

(B) Cox multivariate analysis of METABRIC patient cohort showing hazard ratios of the involution signature independent of subtype, age, and Ki67 status.

(C) Expression of the involution signature in patients categorized by high and low *ELF5* expression (tertiles) in the METABRIC dataset (microarray units) and the TCGA dataset (RNA-seq log10 units). *ELF5*-high patient tertile (red) and *ELF5*-low patients (green) are shown.

(D) Same analysis as (A), but only in luminal patients. Right panels show the influence of the involution signature in *ELF5*-low (top panel) and *ELF5*-high (bottom panel) subsets of luminal patients. Log-rank p values are shown.

CellPress
OPEN ACCESS



(legend on next page)

involuting ECM-CAFs (Figure 7E). As Elf5-driven involution signals are present in human breast cancers, further work is necessary to assess the potential benefit of targeting involution mimicry as a plausible strategy for anti-cancer therapy.

Our scRNA-seq analyses have also allowed us to identify putative molecular candidates responsible for this epithelium-fibroblast-myeloid crosstalk during mimicry of involution (Figure 7C). We find that only alveolar epithelial cells express *Tgfb2* and *Tgfb3* ligands, while their receptor, *Tgfbr3*, is expressed in involution CAFs. TGF-β is linked to tumor malignancy by affecting both cancer cells and the TME, including CAF activation and immune suppression (Bellomo et al., 2016), suggesting that this could be a specific pathway driving mimicry involution. In fact, TGF-β3 is induced in response to milk stasis during the first stage of involution for the induction of alveolar apoptosis, and TGF-β signaling is also key in the second stage of normal involution for the induction of ECM deposition and immune suppression (Flanders and Wakefield, 2009). Altogether, this suggests that TGF-β signaling might be one of the pivotal pathways triggered by Elf5-alveolar cells and thus a potent pathway for therapeutic intervention.

We also find specific crosstalk pathways among involution CAFs, other CAF-types, and the tumor-infiltrating immune system, which might be activated following this Elf5-mediated TGF-β activation. Interestingly, involuting mammary fibroblasts that highly express *Cxcl12* induce monocyte recruitment and are associated with blockade of $CD8^+$ T cell tumor infiltration (Guo et al., 2017a). Cxcl12 is also a known recruitment factor for MDSCs (Sevko and Umansky, 2013). Dpp4 is a ligand from one of the iCAFs groups that could also build the communication bridge between *Cxcl12*-expressing involution iCAFs and *Cxcl2*- and *Cxcl10*-expressing myeloid cells. Dpp4 cleaves and inactivates these chemokines, promoting an immunosuppressive environment by inhibiting the recruitment of effector T cells (Barreira da Silva et al., 2015) and promoting chronic inflammation (Schürmann et al., 2012). This also suggests that Dpp4-expressing iCAFs inhibit the M1-innate response from myeloid cells and fibroblasts.

Thus, our data suggest that during the transition from the luminal-to-basal breast cancer subtype, alveolar cells orchestrate a TME remodeling characterized by a mimicry of involution, resulting in a multicellular process, with the involution CAFs playing a pivotal role in the acquisition of immune suppression and traits of tumor progression (Figure 7E). Altogether, our data highlight the relevance of targeting cancer-associated cell species as a strategy for anti-cancer therapy, an approach particularly important in the context of PABC.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animals
  - Mouse models
  - Animals for mammary gland differentiation series
- METHOD DETAILS
  - Tumor digestion
  - Auto macs clean up
  - Drop-seq
  - Flow cytometry analysis
  - Drop-seq data processing
  - Clustering tree
  - Monocle
  - Gene set variation analysis
  - Metasignatures
  - Tumor protein isolation and western blot
  - Immunohistochemistry
  - Polarized light microscopy
  - Hypoxia analysis
  - Cell to cell communication prediction
  - Multiphoton microscopy
  - Immunofluorescence
  - Clinical samples and survival analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.108945.

ACKNOWLEDGMENTS

We thank the Histopathology Facility from the Garvan Institute. This work is supported by Cancer Institute NSW (fellowship DG00625 and project grant RG18-03) and a donation from the Suttons family (to D.G.-O. and C.J.O.); the National Breast Cancer Foundation (fellowship 2013-2018, PF-13-11) and Cancer Institute NSW (fellowship 2019-2021, CDF181218) (to F.V.-M.); and the National Health and Medical Research Council (project grant NHMRC

Figure 7. Interactome of PyMT tumors

(A) Heatmap of the cell-cell connections of all cell types from PyMT tumors based on CellphoneDB. For simplicity, F6 and F8 were removed and F1 and F7 (myofibroblasts) were pooled together as a single cluster. The scale jet shows the interaction strength based on the statistical framework included in CellphoneDB.

(B) Graphical representation of all significant cell-cell connections identified by CellphoneDB; all identified connections are shown in gray. Fibroblast showed the strongest interactions, with a split of 67 (1^{st} tier) and 50 (2^{nd} tier). The immune system showed weaker interactions, with a split of 15 (3^{rd} tier) and 11 (4^{th} tier).

(C) Representative dot plots of ligand-receptor pairs. Circle size is relative to the number of cells within each annotated cluster that showed a positive expression of each gene, and the blue gradient represents the average scaled expression.

(D) Violin plots of genes from canonical pathways known to recruit and expand MDSCs in all cell clusters. Red, epithelial compartment; blue, stroma; and green, immune cells.

(E) Proposed molecular model of involution mimicry driven by Elf5 expression.

See also Figure S5.

1068753 and fellowship NHMRC 1137061) (to C.J.O.). T.R.C. and M.P. are supported by the NHMRC (1158590 and 1140125), Cancer Institute NSW (CDF171105), Cancer Council NSW (RG19-09), and the Susan G. Komen for the Cure Foundation (CCR17483294).

AUTHOR CONTRIBUTIONS

Conceptualization, F.V.-M., C.J.O.and D.G.-O.; methodology, F.V.-M. and D.G.-O.; investigation and validation, F.V.-M., R.S., A.M.K.L., L.C., M.P., L.R.d.l.F., Z.K., and D.G.-O.; formal analysis, F.V.-M., B.S.G., J.V., and D.G.-O.; resources, J.R.W.C., S.R.O., T.R.C., and C.J.O.; data curation, F.V.-M., B.S.G., J.V., D.L.R., Y.C.-S., N.S., S.I.O., and D.G.-O.; visualization, F.V.-M., B.S.G., J.V., N.S., S.I.O., and D.G.-O.; funding acquisition, F.V.-M., D.G.-O., and C.J.O.; writing – original draft, F.V.-M. and D.G.-O.; writing – review & editing, all authors; supervision, F.V.-M., P.T., C.J.O and D.G.-O..

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: June 8, 2020
Revised: October 29, 2020
Accepted: March 14, 2021
Published: April 13, 2021

REFERENCES

Amara, U., Rittirsch, D., Flierl, M., Bruckner, U., Klos, A., Gebhard, F., Lambris, J.D., and Huber-Lang, M. (2008). Interaction between the coagulation and complement system. Adv. Exp. Med. Biol. *632*, 71–79.

Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220.

Bach, K., Pensa, S., Grzelak, M., Hadfield, J., Adams, D.J., Marioni, J.C., and Khaled, W.T. (2017). Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. Nat. Commun. *8*, 2128.

Barreira da Silva, R., Laird, M.E., Yatim, N., Fiette, L., Ingersoll, M.A., and Albert, M.L. (2015). Dipeptidylpeptidase 4 inhibition enhances lymphocyte trafficking, improving both naturally occurring tumor immunity and immunotherapy. Nat. Immunol. *16*, 850–858.

Bartoschek, M., Oskolkov, N., Bocci, M., Lövrot, J., Larsson, C., Sommarin, M., Madsen, C.D., Lindgren, D., Pekar, G., Karlsson, G., et al. (2018). Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. Nat. Commun. *9*, 5150.

Bellomo, C., Caja, L., and Moustakas, A. (2016). Transforming growth factor β as regulator of cancer stemness and metastasis. Br. J. Cancer *115*, 761–769.

Berger, A.C., Korkut, A., Kanchi, R.S., Hegde, A.M., Lenoir, W., Liu, W., Liu, Y., Fan, H., Shen, H., Ravikumar, V., et al.; Cancer Genome Atlas Research Network (2018). A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. Cancer Cell *33*, 690–705.e9.

Blackstock, C.D., Higashi, Y., Sukhanov, S., Shai, S.Y., Stefanovic, B., Tabony, A.M., Yoshida, T., and Delafontaine, P. (2014). Insulin-like growth factor-1 increases synthesis of collagen type I via induction of the mRNA-binding protein LARP6 expression and binding to the 5′ stem-loop of COL1a1 and COL1a2 mRNA. J. Biol. Chem. *289*, 7264–7274.

Bladström, A., Anderson, H., and Olsson, H. (2003). Worse survival in breast cancer among women with recent childbirth: results from a Swedish population-based register study. Clin. Breast Cancer *4*, 280–285.

Bochet, L., Lehuédé, C., Dauvillier, S., Wang, Y.Y., Dirat, B., Laurent, V., Dray, C., Guiet, R., Maridonneau-Parini, I., Le Gonidec, S., et al. (2013). Adipocyte-derived fibroblasts promote tumor progression and contribute to the desmoplastic reaction in breast cancer. Cancer Res. *73*, 5657–5668.

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

Callihan, E.B., Gao, D., Jindal, S., Lyons, T.R., Manthey, E., Edgerton, S., Urquhart, A., Schedin, P., and Borges, V.F. (2013). Postpartum diagnosis demonstrates a high risk for metastasis and merits an expanded definition of pregnancy-associated breast cancer. Breast Cancer Res. Treat. *138*, 549–559.

Cancer Genome Atlas, N.; Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

Cazet, A.S., Hui, M.N., Elsworth, B.L., Wu, S.Z., Roden, D., Chan, C.L., Skhinas, J.N., Collot, R., Yang, J., Harvey, K., et al. (2018). Targeting stromal remodeling and cancer stem cell plasticity overcomes chemoresistance in triple negative breast cancer. Nat. Commun. *9*, 2897.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401–404.

Chakrabarti, R., Hwang, J., Andres Blanco, M., Wei, Y., Lukačišin, M., Romano, R.A., Smalley, K., Liu, S., Yang, Q., Ibrahim, T., et al. (2012). Elf5 inhibits the epithelial-mesenchymal transition in mammary gland development and breast cancer metastasis by transcriptionally repressing Snail2. Nat. Cell Biol. *14*, 1212–1222.

Ciriello, G., Gatza, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., et al.; TCGA Research Network (2015). Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell *163*, 506–519.

Cortez, E., Roswall, P., and Pietras, K. (2014). Functional subsets of mesenchymal cell types in the tumor microenvironment. Semin. Cancer Biol. *25*, 3–9.

Costa, A., Kieffer, Y., Scholer-Dahirel, A., Pelon, F., Bourachot, B., Cardon, M., Sirven, P., Magagna, I., Fuhrmann, L., Bernard, C., et al. (2018). Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. Cancer Cell *33*, 463–479.e10.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al.; METABRIC Group (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.

Daling, J.R., Malone, K.E., Doody, D.R., Anderson, B.O., and Porter, P.L. (2002). The relation of reproductive factors to mortality from breast cancer. Cancer Epidemiol. Biomarkers Prev. *11*, 235–241.

Davidson, S., Efremova, M., Riedel, A., Mahata, B., Pramanik, J., Huuhtanen, J., Kar, G., Vento-Tormo, R., Hagai, T., Chen, X., et al. (2018). Single-cell RNA sequencing reveals a dynamic stromal niche within the evolving tumour microenvironment. bioRxiv.

Fantozzi, A., and Christofori, G. (2006). Mouse models of breast cancer metastasis. Breast Cancer Res. *8*, 212.

Flanders, K.C., and Wakefield, L.M. (2009). Transforming growth factor-(beta)s and mammary gland involution; functional roles and implications for cancer progression. J. Mammary Gland Biol. Neoplasia *14*, 131–144.

Fu, N.Y., Nolan, E., Lindeman, G.J., and Visvader, J.E. (2020). Stem Cells and the Differentiation Hierarchy in Mammary Gland Development. Physiol. Rev. *100*, 489–523.

Gallego-Ortega, D., Oakes, S.R., Lee, H.J., Piggin, C.L., and Ormandy, C.J. (2013). ELF5, normal mammary development and the heterogeneous phenotypes of breast cancer. Breast Cancer Manag. *2*, 489–498.

Gallego-Ortega, D., Ledger, A., Roden, D.L., Law, A.M., Magenau, A., Kikhtyak, Z., Cho, C., Allerdice, S.L., Lee, H.J., Valdes-Mora, F., et al. (2015). ELF5 Drives Lung Metastasis in Luminal Breast Cancer through Recruitment of Gr1+ CD11b+ Myeloid-Derived Suppressor Cells. PLoS Biol. *13*, e1002330.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of

complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, pl1.

Goddard, E.T., Bassale, S., Schedin, T., Jindal, S., Johnston, J., Cabral, E., Latour, E., Lyons, T.R., Mori, M., Schedin, P.J., and Borges, V.F. (2019). Association Between Postpartum Breast Cancer Diagnosis and Metastasis and the Clinical Features Underlying Risk. JAMA Netw. Open *2*, e186997.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

Guo, Q., Betts, C., Pennock, N., Mitchell, E., and Schedin, P. (2017a). Mammary Gland Involution Provides a Unique Model to Study the TGF-β Cancer Paradox. J. Clin. Med. *6*, 10.

Guo, Q., Minnier, J., Burchard, J., Chiotti, K., Spellman, P., and Schedin, P. (2017b). Physiologically activated mammary fibroblasts promote postpartum mammary cancer. JCI Insight *2*, e89206.

Guy, C.T., Cardiff, R.D., and Muller, W.J. (1992). Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. Mol. Cell. Biol. *12*, 954–961.

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics *14*, 7.

Inman, J.L., Robertson, C., Mott, J.D., and Bissell, M.J. (2015). Mammary gland development: cell fate specification, stem cells and the microenvironment. Development *142*, 1028–1042.

Johansson, A.L., Andersson, T.M., Hsieh, C.C., Cnattingius, S., and Lambe, M. (2011). Increased mortality in women with breast cancer detected during pregnancy and different periods postpartum. Cancer Epidemiol. Biomarkers Prev. *20*, 1865–1872.

Kalluri, R. (2016). The biology and function of fibroblasts in cancer. Nat. Rev. Cancer *16*, 582–598.

Kalyuga, M., Gallego-Ortega, D., Lee, H.J., Roden, D.L., Cowley, M.J., Caldon, C.E., Stone, A., Allerdice, S.L., Valdes-Mora, F., Launchbury, R., et al. (2012). ELF5 suppresses estrogen sensitivity and underpins the acquisition of antiestrogen resistance in luminal breast cancer. PLoS Biol. *10*, e1001461.

Lambrechts, D., Wauters, E., Boeckx, B., Aibar, S., Nittner, D., Burton, O., Bassez, A., Decaluwé, H., Pircher, A., Van den Eynde, K., et al. (2018). Phenotype molding of stromal cells in the lung tumor microenvironment. Nat. Med. *24*, 1277–1289.

Law, A.M.K., Yin, J.X.M., Castillo, L., Young, A.I.J., Piggin, C., Rogers, S., Caldon, C.E., Burgess, A., Millar, E.K.A., O'Toole, S.A., et al. (2017). Andy's Algorithms: new automated digital image analysis pipelines for FIJI. Sci. Rep. *7*, 15717.

Lee, H.J., Gallego-Ortega, D., Ledger, A., Schramek, D., Joshi, P., Szwarc, M.M., Cho, C., Lydon, J.P., Khokha, R., Penninger, J.M., and Ormandy, C.J. (2013). Progesterone drives mammary secretory differentiation via RankL-mediated induction of Elf5 in luminal progenitor cells. Development *140*, 1397–1401.

Lilla, J.N., Joshi, R.V., Craik, C.S., and Werb, Z. (2009). Active plasma kallikrein localizes to mast cells and regulates epithelial cell apoptosis, adipocyte differentiation, and stromal remodeling during mammary gland involution. J. Biol. Chem. *284*, 13792–13803.

Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.L., Gyorki, D.E., Ward, T., Partanen, A., et al.; kConFab (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat. Med. *15*, 907–913.

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.

Lin, E.Y., Jones, J.G., Li, P., Zhu, L., Whitney, K.D., Muller, W.J., and Pollard, J.W. (2003). Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. Am. J. Pathol. *163*, 2113–2126.

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.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Maglione, J.E., Moghanaki, D., Young, L.J., Manner, C.K., Ellies, L.G., Joseph, S.O., Nicholson, B., Cardiff, R.D., and MacLeod, C.L. (2001). Transgenic Polyoma middle-T mice model premalignant mammary disease. Cancer Res. *61*, 8298–8305.

Martinson, H.A., Jindal, S., Durand-Rougely, C., Borges, V.F., and Schedin, P. (2015). Wound healing-like immune program facilitates postpartum mammary gland involution and tumor progression. Int. J. Cancer *136*, 1803–1813.

Mayorca-Guiliani, A.E., Madsen, C.D., Cox, T.R., Horton, E.R., Venning, F.A., and Erler, J.T. (2017). ISDoT: in situ decellularization of tissues for high-resolution imaging and proteomic analysis of native extracellular matrix. Nat Med *23*, 890–898.

Molyneux, G., and Smalley, M.J. (2011). The cell of origin of BRCA1 mutation-associated breast cancer: a cautionary tale of gene expression profiling. J. Mammary Gland Biol. Neoplasia *16*, 51–55.

Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273.

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. Nat. Commun. *9*, 2028.

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.

Oakes, S.R., Hilton, H.N., and Ormandy, C.J. (2006). The alveolar switch: coordinating the proliferative cues and cell fate decisions that drive the formation of lobuloalveoli from ductal epithelium. Breast Cancer Res. *8*, 207.

Oakes, S.R., Naylor, M.J., Asselin-Labat, M.L., Blazek, K.D., Gardiner-Garden, M., Hilton, H.N., Kazlauskas, M., Pritchard, M.A., Chodosh, L.A., Pfeffer, P.L., et al. (2008). The Ets transcription factor Elf5 specifies mammary alveolar cell fate. Genes Dev. *22*, 581–586.

Oakes, S.R., Gallego-Ortega, D., and Ormandy, C.J. (2014). The mammary cellular hierarchy and breast cancer. Cell. Mol. Life Sci. *71*, 4301–4324.

Pal, B., Chen, Y., Vaillant, F., Jamieson, P., Gordon, L., Rios, A.C., Wilcox, S., Fu, N., Liu, K.H., Jackling, F.C., et al. (2017). Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. Nat. Commun. *8*, 1627.

Pereira, B., Chin, S.F., Rueda, O.M., Vollan, H.K., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.J., et al. (2016). The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat. Commun. *7*, 11479.

Prater, M.D., Petit, V., Alasdair Russell, I., Giraddi, R.R., Shehata, M., Menon, S., Schulte, R., Kalajzic, I., Rath, N., Olson, M.F., et al. (2014). Mammary stem cells have myoepithelial cell properties. Nat. Cell Biol. *16*, 942–950.

Puram, S.V., Tirosh, I., Parikh, A.S., Patel, A.P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C.L., Mroz, E.A., Emerick, K.S., et al. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell *171*, 1611–1624.e1624.

Salomon, R., Kaczorowski, D., Valdes-Mora, F., Nordon, R.E., Neild, A., Farbehi, N., Bartonicek, N., and Gallego-Ortega, D. (2019). Droplet-based single cell RNAseq tools: a practical guide. Lab Chip *19*, 1706–1727.

Schedin, P. (2006). Pregnancy-associated breast cancer and metastasis. Nat. Rev. Cancer *6*, 281–291.

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.

Schürmann, C., Linke, A., Engelmann-Pilger, K., Steinmetz, C., Mark, M., Pfeilschifter, J., Klein, T., and Frank, S. (2012). The dipeptidyl peptidase-4 inhibitor linagliptin attenuates inflammation and accelerates epithelialization in wounds of diabetic ob/ob mice. J. Pharmacol. Exp. Ther. *342*, 71–80.

Sevko, A., and Umansky, V. (2013). Myeloid-derived suppressor cells interact with tumors in terms of myelopoiesis, tumorigenesis and immunosuppression: thick as thieves. J. Cancer *4*, 3–11.

Shehata, M., Teschendorff, A., Sharp, G., Novcic, N., Russell, I.A., Avril, S., Prater, M., Eirew, P., Caldas, C., Watson, C.J., and Stingl, J. (2012). Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. Breast Cancer Res. *14*, R134.

Sheridan, J.M., Ritchie, M.E., Best, S.A., Jiang, K., Beck, T.J., Vaillant, F., Liu, K., Dickins, R.A., Smyth, G.K., Lindeman, G.J., and Visvader, J.E. (2015). A pooled shRNA screen for regulators of primary mammary stem and progenitor cells identifies roles for Asap1 and Prox1. BMC Cancer *15*, 221.

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., and Gusterson, B.A. (2007). Mammary gland involution as a multi-step process. J. Mammary Gland Biol. Neoplasia *12*, 25–35.

Stensheim, H., Møller, B., van Dijk, T., and Fosså, S.D. (2009). Cause-specific survival for women diagnosed with cancer during pregnancy or lactation: a registry-based cohort study. J. Clin. Oncol. *27*, 45–51.

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., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Sun, X., Cheng, G., Hao, M., Zheng, J., Zhou, X., Zhang, J., Taichman, R.S., Pienta, K.J., and Wang, J. (2010). CXCL12 / CXCR4 / CXCR7 chemokine axis and cancer progression. Cancer Metastasis Rev. *29*, 709–722.

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196.

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.

Valdes-Mora, F., Handler, K., Law, A.M.K., Salomon, R., Oakes, S.R., Ormandy, C.J., and Gallego-Ortega, D. (2018). Single-Cell Transcriptomics in Cancer Immunobiology: The Future of Precision Oncology. Front. Immunol. *9*, 2582.

Vennin, C., Chin, V.T., Warren, S.C., Lucas, M.C., Herrmann, D., Magenau, A., Melenec, P., Walters, S.N., Del Monte-Nieto, G., Conway, J.R.W., et al. (2017). Transient tissue priming via ROCK inhibition uncouples pancreatic cancer progression, sensitivity to chemotherapy, and metastasis. Sci. Transl. Med. *9*. https://doi.org/10.1126/scitranslmed.aai8504.

Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B., Park, J.E., Stephenson, E., Polański, K., Goncalves, A., et al. (2018). Single-cell reconstruction of the early maternal-fetal interface in humans. Nature *563*, 347–353.

Visvader, J.E. (2011). Cells of origin in cancer. Nature *469*, 314–322.

Visvader, J.E., and Stingl, J. (2014). Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes Dev. *28*, 1143–1158.

Watson, C.J. (2006). Involution: apoptosis and tissue remodelling that convert the mammary gland from milk factory to a quiescent organ. Breast Cancer Res. *8*, 203.

Watson, C.J., and Kreuzaler, P.A. (2009). The role of cathepsins in involution and breast cancer. J. Mammary Gland Biol. Neoplasia *14*, 171–179.

Whiteman, M.K., Hillis, S.D., Curtis, K.M., McDonald, J.A., Wingo, P.A., and Marchbanks, P.A. (2004). Reproductive history and mortality after breast cancer diagnosis. Obstet. Gynecol. *104*, 146–154.

Zappia, L., and Oshlack, A. (2018). Clustering trees: a visualization for evaluating clusterings at multiple resolutions. Gigascience *7*, giy083.

Zwick, R.K., Rudolph, M.C., Shook, B.A., Holtrup, B., Roth, E., Lei, V., Van Keymeulen, A., Seewaldt, V., Kwei, S., Wysolmerski, J., et al. (2018). Adipocyte hypertrophy and lipid dynamics underlie mammary gland remodeling after lactation. Nat. Commun. *9*, 3592.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Purified Rat Anti-Mouse CD16/CD32 (Mouse BD Fc Block)	BD Biosciences	Cat#553141
Rat Gamma Globulin	Jackson ImmunoResearch Laboratories	Cat#012-000-002; RRID:AB_2337135
PerCP/Cyanine5.5 anti-mouse CD49b	Biolegend	Cat#108915; RRID:AB_1595599
APC anti-human/mouse CD49f	Biolegend	Cat#313615; RRID:AB_2734290
APC/Cyanine7 anti-mouse Ly-6A/E (Sca-1)	Biolegend	Cat#108125; RRID:AB_10639725
PE/Cyanine7 anti-mouse CD45	Biolegend	Cat#103114; RRID:AB_312979
PE anti-mouse CD326 (Ep-CAM)	Biolegend	Cat#118205; RRID:AB_1134176
PerCP/Cyanine5.5 anti-Mouse CD31	BD PharMingen	Cat#562861
Anti-Elf5 (N-20)	Santa Cruz Biotechnology	Cat#SC-9645; RRID:AB_640106
Anti-Vimentin	Leica Biosystems	Cat#VIM-572-L-CE
Anti-E-cadherin	BD Biosciences	Cat#610181, RRID:AB_397580
Anti-β-Actin	Sigma-Aldrich	Cat#A1978; RRID:AB_476692
HRP-donkey anti-goat	Santa Cruz Biotechnology	Cat#sc-2020, RRID:AB_631728
HRP-sheep anti-mouse	Amersham Biosciences	Cat#25005363
Anti-COL1A1	Cell Signaling	Cat#39952
Anti- MMP3	Abcam	Cat#ab52915; RRID:AB_881243
Anti-SDF1	Abcam	Cat#ab9797; RRID:AB_296627
Anti-mouse milk proteins	Accurate Chemical & Scientific CO	Cat#YNRMTM
Anti-CD31	BD PharMingen	Cat#550274; RRID:AB_393571
Goat-anti-rat Alexa Fluor 647	Thermo Fisher	Cat#a-21247; RRID:AB_141778
Chemicals, peptides, and recombinant proteins		
Collagenase	Sigma Aldrich	Cat#C9891
Hyaluronidase	Sigma Aldrich	Cat#H3506
Dispase	Roche	Cat#04942078001
Ammonium chloride	Sigma Aldrich	Cat#A9434
DNase I	Roche	Cat#10104159001
Ficoll PM-400	GE healthcare	Cat#17-0300-10
Sarkosyl	Sigma Aldrich	Cat#L7414
Droplet generation Oil	Bio-Rad	Cat#186-4006
Perfluorooctanol (PFO)	Sigma Aldrich	Cat#370533
Barcode Oligo dT primer ON Beads (Drop-seq) Macosko-201	Chemgenes	Cat#113015B
6X SSC	Life Technologies	Cat#15557-036
Maxima H-Reverse transcriptase	ThermoFisher Scientific	Cat#EP0753
Exonuclease I	New England BioLabs	Cat#M0293S
Agencourt AMPure XP beads	Beckman Coulter	Cat#A63881
Qtracker 655 Vascular Labels	Thermo Fisher Scientific	Cat#Q21021MP
Protease cocktail	Roche	Cat#11836170001
Phosphatase cocktail	Roche	Cat#04906845001
Benzonase nuclease	Sigma Aldrich	Cat#9025-65-4
BOND Epitope Retrieval Solution 1 (pH 6.0)	Leica Biosystems	Cat#AR9961
BOND Epitope Retrieval Solution 2 (pH 9.0)	Leica Biosystems	Cat#AR9640

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Target Retrieval Solution	Dako	Cat#S1699
Tissue-Tek O.C.T. Compound	Sakura	Cat#4583
Direct Red 80- picrosirius red	Sigma Aldrich	Cat#365548
Critical commercial assays		
Dead Cell Removal Kit	Miltenyi Biotec	Cat#130-090-101
KAPA HiFi HotStart ReadyMix PCR kit	KAPABIOSYSTEMS	Cat#KR0370
Agilent High Sensitivity DNA Kit	Agilent Technologies	Cat#5067-4626
PhiX Control v3	Illumina	Cat#FC-110-3001
Nextera XT DNA Library kit	Illumina	Cat#FC-131-1024
Nextseq 500 High Output v.2 kit	Illumina	Cat#FC-404-2005
NuPAGE 4 to 12% Bis-Tris acrylamide gels	Thermo Fisher Scientific	Cat#NP0322
BOND Polymer Refine Detection kit	Leica Biosystems	Cat#DS9800
EnVision+ System/HRP, Rabbit (AEC+)	Dako	Cat#K4009
Hypoxyprobe Plus Kit	HypoxyprobeTM	Cat#HP2-200Kit
Deposited data		
Raw data	This paper	GSE158677
Molecular Taxonomy of Breast Cancer International Consortium (METABRIC)	Pereira et al., 2016	https://www.cbioportal.org/
The Cancer Genome Atlas (TCGA)	Berger et al., 2018; Ciriello et al., 2015; Cancer Genome Atlas, 2012	https://www.cbioportal.org/
Gene mouse hallmark gene list	Subramanian et al., 2005; Mootha et al., 2003	http://bioinf.wehi.edu.au/software/MSigDB
Experimental models: Organisms/strains		
Mouse: FVB/N-Tg(MMTV-PyVT)	Guy et al., 1992	N/A
Mouse: FVB/N-Tg(MMTV-PyVT, MMTV-rTA, TetON-Elf5-IRES-eGFP)	Gallego-Ortega et al., 2015	N/A
Mouse: FVB/NJAusb	Inbred strain from Australian BioResources (origin The Jackson Laboratory)	MGI Cat# 6200618, RRID:MGI:6200618
Oligonucleotides		
Oligos for genotyping: Elf5/EGFP_FW: GCACCATCTTCTTCAAGGACGAC	This paper	N/A
Oligos for genotyping: Elf5/EGFP_RV: GGTCTTTGCTCAGGGCGGAC	This paper	N/A
Oligos for genotyping: rtTA/MTB_FW: TGCCGCCATTATTACGACAAGC	This paper	N/A
Oligos for genotyping: rtTA/MTB_RV: ACCGTACTCGTCAATTCCAAGGG	This paper	N/A
Oligos for genotyping: PyMT_FW: CGGCGGAGCGAGGAACTGAGGAGAG	This paper	N/A
Oligos for genotyping: PyMT_RV: TCAGAAGACTCGGCAGTCTTAGGCG	This paper	N/A
Primers for Drop-seq see Table S4		
Software and algorithms		
Seurat (v Seurat_2.3.4)	Butler et al., 2018	https://satijalab.org/seurat/
Clustering tree	Zappia and Oshlack, 2018	https://github.com/lazappi/clustree
Shiny	R studio	https://shiny.rstudio.com
Monocle (v2 ole)	Trapnell et al., 2014	http://monocle-bio.sourceforge.net/
GSVA	Hänzelmann et al., 2013	http://www.bioconductor.org/packages/release/bioc/html/GSVA.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CellphoneDB	Vento-Tormo et al., 2018	https://www.cellphonedb.org/
Metasignatures	This paper	https://github.com/GaldesLab/CellReportsManuscript/blob/master/PYMT_DROPseq_scRNA/R/plotMetaScore.r
Andy's algorithms	Law et al., 2017	https://github.com/andlaw1841/Andy-s-Algorithm
ImageJ	NIH	RRID: SCR_003070
MATLAB	MathWorks	RRID:SCR_001622
GraphPad Prism	GraphPad Software	RRID: SCR_002798
R v3.4.1	The R Foundation for Statistical Computing	https://www.r-project.org/
Picrosirius Red birefringence analyzer	Vennin et al., 2017	https://github.com/TCox-Lab/PicRed_Biref
Orientation analysis script	Mayorca-Guiliani et al., 2017	https://github.com/TCox-Lab/Collagen_Orientation
Other		
Bioinformatics analyses code used in this paper	This paper	https://github.com/GaldesLab/CellReportsManuscript/
Bioinformatics analyses code for CellPhoneDB	This paper	https://galdeslab.github.io/CellReportsManuscript/
Interactive Shiny app for Drop-seq data	This paper	https://galdeslab.github.io/CellReportsManuscript/PyMT.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, David Gallego-Ortega (d.gallego@garvan.org.au).

Materials availability

This study did not generate new unique reagents

Data and code availability

The accession number for the Drop-seq data reported in this paper is Gene Expression Omnibus (GEO):GSE158677.

The code generated for the downstream analysis and interactive tools used in this manuscript can be found at https://github.com/GaldesLab/CellReportsManuscript.

Direct links to interactive visualization tools are:

- CellPhonDB interface https://galdeslab.github.io/CellReportsManuscript/
- Shiny application for Drop-seq data clustering tool https://galdeslab.github.io/CellReportsManuscript/PyMT.html/

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

All mice of this study are in a pure FBV/n background with more than 20 generations of backcrossing. Mice were bred at the Australian BioResources Pty Ltd (ABR) facility in Moss Vale, NSW, and housed during the study at the Biological Testing Facility (BTF) at Garvan Institute's both under specific pathogen-free conditions. All mice used in this work were females due to the nature of this study involving breast cancer and pregnancy. All animal experiments carried out according to guidelines contained within the NSW (Australia) Animal Research Act 1985, the NSW (Australia) Animal Research Regulation 2010 and the Australian code of practice for the care and use of animals for scientific purposes, (8th Edition 2013, National Health and Medical Research Council (Australia)). All experiments involving mice have been approved by the St. Vincent's Campus Animal Research Committee AEC #14/27, #17/03 and #19/02.

Mouse models

Two different genetically engineered mouse models were used in this study: MMTV-Polyoma Middle T antigen (PyMT/WT) and mammary restricted (MMTV) doxycycline-inducible (rtTA) Elf5 expression mouse (PyMT/ELF5) (see Figure S1 for breeding strategy). PyMT/WT mouse model (gift from Dr. William J. Muller, McGill University) has been previously described (Guy et al., 1992; Lin et al., 2003; Maglione et al., 2001). The PyMT/ELF5 model was generated by crossing the MMTV-PyMT with the doxycycline (DOX) inducible Elf5 Knock In mouse line (Oakes et al., 2008) as previously described (Gallego-Ortega et al., 2015). Briefly, Elf5 transgene expression is controlled by a doxycycline inducible promoter (TetON, tetracycline response element) and the expression of the tetracycline transactivator protein (rtTA) is controlled by the MMTV promoter, a mammary-epithelium specific promoter. Elf5 expression was specifically induced in mammary epithelial cells since puberty of female mice (6 weeks-old) through doxycycline containing food (700 mg/Kg of Doxycycline, Gordon's Specialty Stockfeeds) as performed previously (Gallego-Ortega et al., 2015). At ethical endpoint, (10% ± 3% tumor/body weight, which approximately corresponds to 14-week-old animals) PyMT/WT and PyMT/ELF5 were euthanized and size-matched tumors harvested and processed for single cell digestions (see below). Genotyping was performed at the Garvan Molecular Genetics facility (NATA accredited, ISO 17025) by PCR of DNA extracted from the mouse tail tip using three set of primers: Elf5/EGFP (GCACCATCTTCTTCAAGGACGAC and GGTCTTTGCTCAGGGCGGAC), rtTA/MTB (TGCCGCCAT TATTACGACAAGC and ACCGTACTCGTCAATTCCAAGGG) and PyMT (CGGCGGAGCGAGGAACTGAGGAGAG and TCAGAA GACTCGGCAGTCTTAGGCG). The touchdown PCR conditions were 94°C for 10 s of initial denaturation, followed by 10 cycles of 94°C 10 s, 65-55°C for 30 s and 72°C for 1 min and 10 s and then 31 cycles of 94°C 10 s, 55°C for 30 s and 72°C for 1 min and 10 s; the final extension is 72°C for 3 minutes. All animals used in this study are heterozygous for Elf5, MTB, and PyMT (PyMT/Elf5) or heterozygous for PyMT and MTB or heterozygous only for PyMT (PyMT/WT).

Animals for mammary gland differentiation series

Adult female mice (8-10 weeks old) were timed mated and pregnancy confirmed by visualizing a vaginal plug on the morning after mating (0.5 days post coitum (d.p.c)). Inguinal mammary gland wholemounts were prepared from pregnant dams at 18.5 d.p.c or at 4.5 days post-partum (established lactation) or after 1 or 4 days after forced involution. For post-partum and involution time points, litters were normalized to 7 and forced involution induced by removal of pups at 10.5 days post-partum respectively. Inguinal mammary gland wholemounts from 8-10-week-old virgin female mice were used as control.

METHOD DETAILS

Tumor digestion

Mammary tumors were collected from aged-matched MMTV-PyMT mice at 14 weeks of age, when tumor weights were 10 ± 3% of total weight, no noticeable differences in tumor sizes were observed between genotypes according to previous results (Gallego-Ortega et al., 2015). Areas of necrosis were excluded, and only regions of epithelial mammary carcinomas were used for digestion and analysis. Tumors were manually dissected into 3-5mm pieces using a surgical scalpel blade before being chopped to 100 μm with maximum blade force on a McIlwain Tissue chopper. Tumor samples were incubated with 15,000 U of collagenase (Sigma Aldrich Cat# C9891) and 5,000 U of hyaluronidase (Sigma Aldrich Cat# H3506) dissolved in DMEM high glucose with 5% FBS at 37°C shaking in 220 rpm for 1 hour. Samples were briefly disrupted with a pipette every 15 min during the incubation time to ensure the tissues were sufficiently resuspended. 0.25% trypsin (GIBCO Cat# 15090-046) in 1mM EGTA was used to digest the samples for 1min in 37°C waterbath before proceeding with a 5-minute incubation with 5mg/mL of dispase (Roche Cat# 04942078001) dissolved in PBS in 37°C waterbath. Red blood cells were then lysed with 0.8% ammonium chloride (Sigma Aldrich Cat# A9434) dissolved in water for 5 min in 37°C waterbath. Samples were washed with PBS containing 2% FBS and spun at 1200 rpm for 5min at 4°C between each step. The supernatant was aspirated and 1mg/mL DNase I (Roche Cat# 10104159001) was mixed with the sample before incubation with each step. Finally, cells were filtered through a 40 μM nylon mesh (Corning) and resuspended in PBS with 2% FBS. Alive cells were selected using the Auto macs clean up procedure explained below.

An overview of the experimental design for the Drop-seq study on PyMT/WT and /ELF5 tumors is shown in Figure S2A, a total of 5 mammary carcinomas for PyMT/WT and 6 for PyMT/ELF5 genotype were used for the Drop-seq study. Figure S2B shows QC cut offs and Figure S2C shows the cell distribution in a tSNE plot and reveals no batch effect due to individual tumors within each genotype with the exemption of WT#5 that contains a higher proportion of T cells, that sole form cluster 3 in Figure 1E. This noticeable increase of T cells is likely caused by the incorporation of a proximal lymph node during cell preparation of this particular tumor. Importantly this cluster did not drive any of the conclusions of this study.

Auto macs clean up

The viability of the cells was assessed by FACS. All tumors that contained less than 80% alive cells were labeled with Annexin specific MACS beads using the Dead Cell Removal Kit (Miltenyi Biotec Cat# 130-090-101) following the manufacturers' instructions and dead cells were removed by passing the labeled cells through the autoMACS® Pro (Miltenyi) (see more details at Salomon et al., 2019). A preparation of cells containing a high proportion of viable cells (> 85% viability assessed by DAPI in FACS) were loaded into the microfluidic Drop-seq pipeline.

Drop-seq

Cells were captured using the microfluidic devices described for Drop-seq (Macosko et al., 2015) following the Online Drop-seq Laboratory Protocol version 3.1 (http://mccarrolllab.com/dropseq/) with some modifications explained here. Briefly, to build the in-house Drop-seq device, we purchased the PDMS co-flow microfluidic droplet generation chip from FlowJem using the CAD file design from Macosko et al.; three syringe pumps (Legato® 100 syringe pump, KD Scientific Cat# KDS-788100) and a magnetic bead mixer (VP Cat# 772DP-N42-5-2). Tumor-dissociated single cells at 75 cells/μl in 0.01% BSA in PBS were loaded into one of the syringes; the Drop-seq barcoded beads in lysis buffer (6% Ficoll PM-400, 0.2% Sarkosyl, 20mM EDTA, 200 mM Tris pH 7.5 and 50 mM DTT) at 280,000 beads/ml, under constant suspension with the magnetic bead mixer, were loaded in the second syringe, and the third syringe contained the droplet generation oil (Bio-Rad Cat# 186-4006). Flow rates were set for 4 ml/hour for the bead and cell suspensions, and 15 ml/hour for the oil, resulting in ~105 μm diameter droplets. We performed cell-bead droplet captures of 7 min to obtain 0.5ml of beads that will capture between 3000 to 3,500 cells or STAMPs (~2.5% of beads containing a cell). The generated droplets were broken by adding perfluorooctanol (Sigma Cat# 370533) in 6X SCC and washed in 6X SSC. The beads with barcoded RNAs were retrotranscribed (Maxima H-Reverse transcriptase, ThermoFisher Scientific Cat# EP0753 using TSO primer, see Table S4) and treated with exonuclease I (New England BioLabs Cat# M0293S). The beads containing cDNA were then washed, counted and split to perform PCR amplification, we put 4000 beads (to generate approximately 100 STAMPs) per PCR (KAPA HiFi HotStart ReadyMix PCR kit, KAPABIOSYSTEMS Cat# KR0370 and primer "TSO_PCR" from Table S4), the PCR program was the same than Drop-seq lab protocol (v 3.1). The PCRs were purified using AMPure XP beads (Beckman Coulter Cat# A63882) at a 0.6X ratio following the manufactures' instructions. Pooled PCRs per sample were checked for quality, fragment size and concentration using the Agilent High Sensitivity DNA Kit (Agilent Cat# 5067-4626) and run in the 2100 Bioanalyzer Instrument (Agilent). A total of 600 pg of cDNA were then tagmented using the reagents from Nextera XT DNA Library kit (Illumina Cat# FC-131-1024) but following Macosko et al. specifications and custom primers (Table S4, P5-TSO_Hybrid and N70X, the later primer sets for multiplexing). The tagmented libraries were purified doing two rounds of clean ups with AMPure XP beads, the first one at a ratio of 0.6X and a second one at a ratio of 1X and eluted in 10 μl H2O. The tagmented and multiplexed cDNA libraries were sequenced in Nextseq 500 using Nextseq 500 High Output v.2 kit (75 cycles, Illumina Cat# FC-404-2005) following Macosko et al. recommendations (see Table S4 for the custom sequencing primer," Read1CustomSeqB") with the following modifications. Drop-seq denatured libraries were loaded at 1.3pM final concentration and were spiked in with 10% of 1.8pM PhiX Control v3 (Illumina Cat# FC-110-3001). Sequencing specifications were as follows: 26bp Read1, 51bp Read 2 and single Nextera indexing (8bp). A total of ~3,000 cells or STAMPs/ per run were sequenced.

Flow cytometry analysis

Flow cytometry analysis were performed on BD FACSymphony™ High-Speed Cell Analyzer or the FACSAriaIII (Becton Dickinson) following automated compensation (FACSDiVa) using a 100um nozzle. Unspecific binding was prevented using 1:80 of Mouse BD Fc Block (BD Biosciences, CA, USA, Cat# 553141) and 1:500 of Rat Gamma Globulin (Jackson ImmunoResearch Laboratories Inc., PA, USA, Cat# 012-000-002) in a buffer containing DNase I (100 μg/mL, Roche Cat# 10104159001). Alive cells were assessed by negative DAPI staining (0.5 μg/ml; Thermo Fisher Scientific Cat# D1306). Subsequent immunostaining was performed using the following fluorescently tagged antibodies and dilutions: anti-mouse CD49b (1:100, Biolegend Cat# 108915), anti-human/mouse CD49f (1:200, Biolegend Cat# 313615), anti-mouse Ly-6A/E (1:100, Biolegend Cat# 108125), anti-mouse CD45 (1:400, Biolegend Cat# 103114), anti-mouse Ep-CAM (1:600, Biolegend Cat# 118205) and anti-Mouse CD31 (1:40, BD PharMingen Cat# 562861). Data was analyzed using FlowJo software (version 10.4.2).

Drop-seq data processing

The sequencing output was analyzed using the McCaroll lab core computational protocol with a custom genome (mm10 plus Trinity assemblies of transgene sequences [Grabherr et al., 2011]) and gene annotation (gencode vM14 plus). Seurat (v Seurat_2.3.4 [Butler et al., 2018]) was the main platform for downstream analysis.

A total of 26,613 cells were sequenced (18,828 from WT tumors and 7,785 from ELF5 tumors) (Figure S2A). First, we removed low-quality cells by modeling mitochondrial to nuclear gene content to < 15%(Macosko et al., 2015) and considering differences between homeostatic tissues and tumors. We subsequently removed outlier cells that contained more than 4,000 genes as they could potentially constitute cell doublets (Figure S2B). Thus, DGE matrices were trimmed for quality metrics (> 200 genes, < 15% mitochondrial genes, and identified genes expressed in at least 3 cells), as a result 15,702 high quality cells (11,490 PyMT/WT and 4,212 PyMT/ELF5) with a total of 28,945 genes proceeded with downstream analysis. A total of 6,176 informative genes were identified based on expression and variance and organized into principal components. Two thirds of the total variation of the system was defined by the first 20 PCs. Downstream analysis was performed according to Butler et al. with UMI number regression and 20 principal components of variable genes being used for dimensional reduction (tSNE) and cluster calling (Butler et al., 2018) (Figure S2C). Elf5 expression was restricted to the PyMT/ELF5 tumors (Figure S2D). PyMT tumors are derived from the signal of the Polyoma middle-T oncogene in congenic FBVn animals of pure background, thus the effects of the environment are minimized as the individuals are exposed to identical controlled standard laboratory conditions.

Clustering tree

Clustering tree (Zappia and Oshlack, 2018) was employed in conjunction with cluster validation, and split error estimation across a range of clustering resolutions to identify optimal resolution values. CCA was performed on experiments split by genotype using the union of the top 2000 variable genes for each genotype. No cells were discarded from downstream analysis to retain unique subpopulations between the experiments and downstream analysis was performed as above using CCA in place of PCA.

Monocle

Monocle (v2 ole (Trapnell et al., 2014)) was used to assemble cells assigned to epithelial clusters along a pseudotime vector generated from single cell expression of "Gorsmy" (Pal et al., 2017). States were assigned using DDRTree according to the manual.

Gene set variation analysis

GSVA (Hänzelmann et al., 2013) was calculated for averaged expression values for clusters or Poisson distributed counts data using gene mouse hallmark gene lists downloaded from http://bioinf.wehi.edu.au/software/MSigDB (Subramanian et al., 2005; Mootha et al., 2003).

Metasignatures

Metasignatures were generated by calculating the sum of scaled expression scores for all genes nominated for the signature in each cell. The relative contribution of each gene to the score was used to rank genes in the signature.

Tumor protein isolation and western blot

Tumor pieces of a maximum of 0.5 cm x 0.5 cm were lysed in 200 ul of tissue lysis buffer (1.5 mM MgCl2, 0.2 mM EDTA, 0.3 mM NaCl, 25 mM HEPES pH 7.5, 20 mM β-glycerophosphate, 0.1% Triton X-100) supplemented with a cocktail of protease/phosphatase inhibitors (100 μg/ml PMSF, 0.5 mM DTT, protease cocktail Roche Cat# 11836170001, phosphatase cocktail Roche Cat# 04906845001) and 400 U/ml of Benzonase® nuclease (Sigma-Aldrich Cat# 9025-65-4). Tumor chunks in the lysis buffer were homogenized using a dounce homogenizer with tight pestle at 4°C until the tissue was completely dissolved. Lysed tumors were centrifugated at 1,400 rpm for 20 min at 4°C and the supernatant was incubated at room temperature for 1 hour under rotation. Then the sample was centrifugated at 8000 rpm for 10 min at 4°C and the aqueous phase was isolated and quantitated for total protein using Bradford protein assay (Bio-Rad Cat# 5000002) following manufacturer's instructions.

40 ug of protein lysates were prepared in Invitrogen LDS sample buffer (Cat# NP0007 (4X)) plus Sample Reducing Agent (Invitrogen Cat# NP0004 10X) and separated on precast NuPAGE 4 to 12% Bis-Tris acrylamide gels (Thermo Fisher Scientific, Cat# NP0322) in MOPS buffer, transferred to a PVDF membrane (Bio-Rad). Membranes were blocked with 5% skim milk in TBS-Tween buffer for 1 hour at room temperature, primary antibody for 2 hours at room temperature ors overnight at 4°C, and a horseradish peroxidase (HRP)-linked secondary antibody for 1 hour at room temperature. The primary antibodies used were: anti-Elf5 (N-20) (1:500, Santa Cruz Cat# sc-9645), anti-vimentin (1:1000, Leica Biosystems Cat# VIM-572-L-CE), anti-E-cadherin (1:10,000, BD Biosciences Cat# 610181), anti-β-Actin (1:2000, Sigma-Aldrich Cat# A1978). Secondary antibodies were HRP-donkey anti-goat (1: 1000, Santa Cruz Cat# sc-2020) and HRP-sheep anti-mouse (1:5000, Amersham Biosciences, Cat# 25005363).

Chemiluminescence detection was done using Western Lightning Plus-ECL, Enhanced Chemiluminescence Substrate kit (Perkin-Elmer Cat# NEL103E001EA) on Fuji Medical X-ray Film (Fujifilm).

Immunohistochemistry

IHC was performed at the Garvan's Histopathology facility. Mouse mammary glands at different stages of development from FVB/N mice and mammary tumors and lungs from PyMT/WT and PyMT/ELF5 mice were fixed, defatted, dehydrated, and paraffin embedded using the Leica Peloris II tissue processor following the manufacturers' recommendations. The paraffin embedded tissues were cut into 4-μm sections using the Leica Microtome RM2235 and placed on glass microscope slides. These slides were stained for hematoxylin and eosin (H&E), or immunostained using the fully automated Leica Bond RX. Briefly, sections were deparaffinized in xylene and rehydrated in ethanol grading; antigen retrieval was performed using the heat-induced epitope retrieval method either using the BOND Epitope Retrieval Solution 1 (pH 6.0) (Leica Cat# AR9961) or BOND Epitope Retrieval Solution 2 (pH 9.0) (Leica Cat# AR9640) for 30 min, peroxidase quenching was performed with 3% H2O2. The primary antibodies were incubated for 1 h: Anti-COL1A1 (1:200, Cell Signaling Cat# 39952), -MMP3 (1:100, Abcam Cat# ab52915) and -CXCL12/SDF1 (1:500, Abcam Cat# ab9797) and the Leica BOND Polymer Refine Detection kit (Cat# DS9800) was used as secondary detection method.

For the immunostaining of milk in mammary tumors and mammary glands, antigen retrieval was done in a water bath for 20 mins in citrate buffer, pH 6.0 (Dako Cat# S1699) and the next steps for the immunostaining were done using the Dako Autostainer. The rabbit anti-mouse milk proteins primary antibody (1:12,000, Accurate Chemical & Scientific CO, Cat# YNRMTM) was incubated for 1h followed by 30-min incubation of EnVision+ System/HRP Rabbit (Dako, Cat# K4009). Antigen visualization was carried out using the Liquid DAB+ Substrate chromogen system (Cat# K3467).

Polarized light microscopy

Paraffin-embedded samples were cut into 4 μm sections and stained with 0.1% picrosirius red for fibrillar collagen (Direct Red 80, Sigma Cat# 365548) according to manufacturer's instructions. As previously performed (Cazet et al., 2018; Vennin et al., 2017) polarized light imaging was performed on a Leica DM6000 fitted with a polarizer in combination with a transmitted light analyzer. The relative area of red-orange (high birefringent) fibers, yellow (medium birefringent) fibers, and green (low birefringent) fibers (as a % of total fibers) was calculated.

Hypoxia analysis

Hypoxia analysis were performed using the Hypoxyprobe Plus Kit (Hypoxyprobe™ Cat# HP2-200Kit) following the manufactures' instructions. Briefly, 60 mg/kg body weight of pimonidazole HCL (Hypoxyprobe™-1) were intraperitoneally injected into PyMT/WT and PyMT/Elf5 animals at ethical endpoint (~at 14 weeks). Tumors and lungs were harvested the next day after pimonidazole HCL injection and were fixed in 10% neutral-buffered formalin overnight. FFPE tissues were sectioned to 4μm for IHC staining as described above. FITC conjugated to anti-pimonidazole mouse IgG1 monoclonal antibody (FITC-MAb1, clone 4.3.11.3) at 1:1000 and a peroxidase conjugated rabbit anti-FITC as a secondary reagent (1:100) were used to stain for the protein adducts of pimonidazole in hypoxic tissue.

Cell to cell communication prediction

CellphoneDB (Vento-Tormo et al., 2018) (https://www.cellphonedb.org/) was performed for 100 cells expressing the highest number of genes in each cluster. A specific interaction was considered as significant if $p < 0.01$ and mean score > 0.3. Expression values of ligand/receptor gene pairs were plotted using Seurat DotPlot function for all cells in each cluster. Interactome of Figure 7B was generated using the parameters of more than 10 significant interactions with a mean score greater than 0.3, number cut as more than 10 connections and number split 10.

Multiphoton microscopy

Second Harmonic Generation (SHG) signal was acquired using a 25x 0.95 NA water objective on an inverted Leica DMI 6000 SP8 confocal microscope. Excitation source was a Ti:Sapphire femtosecond laser cavity (Coherent Chameleon Ultra II), tuned to a wavelength of 880 nm and operating at 80 MHz. Signal intensity was recorded with RLD HyD detectors (420/40 nm)). For tumor samples, 5 representative regions of interest (512 μm x 512 μm) per tumor were imaged, for CDMs 3 representative areas of 3 technical replicates over a 3D z stack (20 μm depth; 20 μm depth for CDM, with a z-step size of 1.26 μm). SHG signal coverage in tumor samples was measured with ImageJ (National Institutes of Health, Bethesda, MD, USA). For CDMs mean SHG intensity was measured using MATLAB (Mathworks). Representative images of maximum projections are shown.

Blood vessel patency analysis was performed injecting 10 μl of quantum dots blood tracers (Qtracker 655 Vascular Labels, Thermo Fisher Scientific Cat# Q21021MP) through the tail vein of the animals (Video S1 and S2). Images were acquired with a 25x NA0.95 water objective. A dichroic filter (560 nm) was used to separate the GFP signal from quantum dot emission, which were further selected with band pass filters (525/50 and 617/73, respectively).

Immunofluorescence

Mammary tumors from PyMT/WT and PyMT/ELF5 animals were harvested at ethical endpoint and frozen in Tissue-Tek® O.C.T. Compound (Sakura Cat# 4583). Cut frozen sections (8 μm) were air-dried for 30 minutes and then fixed in 4% paraformaldehyde (PFA) for 5 minutes. After washing 3 times with PBS, the sections were blocked in serum free protein block (Dako Cat# X0909) for 60 minutes. Sections were stained with anti-CD31 antibody (1:400, BD PharMingen Cat# 550274) overnight at 4°C. After washing, the samples were blocked in serum free protein block for 5 minutes. Then, the samples were incubated for 1 hour in secondary antibody (1:600, Goat-anti-rat Alexa Fluor 647, Invitrogen Cat# a-21247). The samples were washed and incubated for 10 minutes in DAPI (1:1000, Thermo Fisher Scientific Cat# D1306). After washing, the sections were covered with coverslips using mounting medium (Confocal Matrix® Micro Tech Lab). The sections were analyzed and photographed one day after mounting them, using the fluorescent microscope (Leica DM 5500).

Clinical samples and survival analysis

We used the METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) (Curtis et al., 2012; Pereira et al., 2016) dataset and the TCGA (The Cancer Genome Atlas) dataset (Berger et al., 2018; Cancer Genome Atlas, 2012; Ciriello et al., 2015) for patient survival analysis. These datasets contain RNaseq and microarray data as well as detailed clinical information from breast cancer patients. Data was accessed using cBioPortal (https://www.cbioportal.org/) (Gao et al., 2013; Cerami et al., 2012).

Patient survival analysis was performed using the "survminer" package (https://github.com/kassambara/survminer). Survival Curves were drawn using 'ggplot2'. R package version 0.4.3 on METABRIC data accessed from the R-Based API for Accessing the MSKCC Cancer Genomics Data Server, CGDSR (R package cgdsr version 1.2.10). Cohorts were split by ELF5 expression and then by metascores and overall survival was compared by cox proportional hazards analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical details for sample size and the threshold for statistical significance can be found in the figure legends. Unless otherwise stated in the figure legend and or method details, data in bar graphs are expressed as mean ± SEM and the statistical test was unpaired t test. Data represented as boxplot show the minimum and maximum (whiskers); the median of the lower half of the dataset (quartile 1), the median and the median of the upper half of the dataset (quartile 3) (box).

Quantification of the hypoxia assays and IHC was performed using Andy's DAB Algorithm (Law et al., 2017) through FIJI (ImageJ). CDMs mean SHG intensity was measured using MATLAB (Mathworks). Quantitative intensity measurements of fibrillar collagen content and birefringent signal were carried out using in house scripts in ImageJ and can be found in https://github.com/GaldesLab/CellReportsManuscript and https://github.com/TCox-Lab/PicRed_Biref and https://github.com/TCox-Lab/Collagen_Orientation .

The graphs and statistical analyses associated were performed using GraphPad Prism.

ADDITIONAL RESOURCES

An interactive Shiny clustering tool of the Drop-seq datasets was developed to enable the access and analysis of the data in a user-friendly and interactive manner. The website can be found at https://galdeslab.github.io/CellReportsManuscript/PyMT.html.

Cell Reports, Volume *35*

Supplemental information

Single-cell transcriptomics reveals involution mimicry during the specification of the basal breast cancer subtype

Fátima Valdés-Mora, Robert Salomon, Brian Stewart Gloss, Andrew Man Kit Law, Jeron Venhuizen, Lesley Castillo, Kendelle Joan Murphy, Astrid Magenau, Michael Papanicolaou, Laura Rodriguez de la Fuente, Daniel Lee Roden, Yolanda Colino-Sanguino, Zoya Kikhtyak, Nona Farbehi, James Ronald William Conway, Neblina Sikta, Samantha Richelle Oakes, Thomas Robert Cox, Seán Ignatius O'Donoghue, Paul Timpson, Christopher John Ormandy, and David Gallego-Ortega

Figure S1 related to Figure 1 and Figure 3

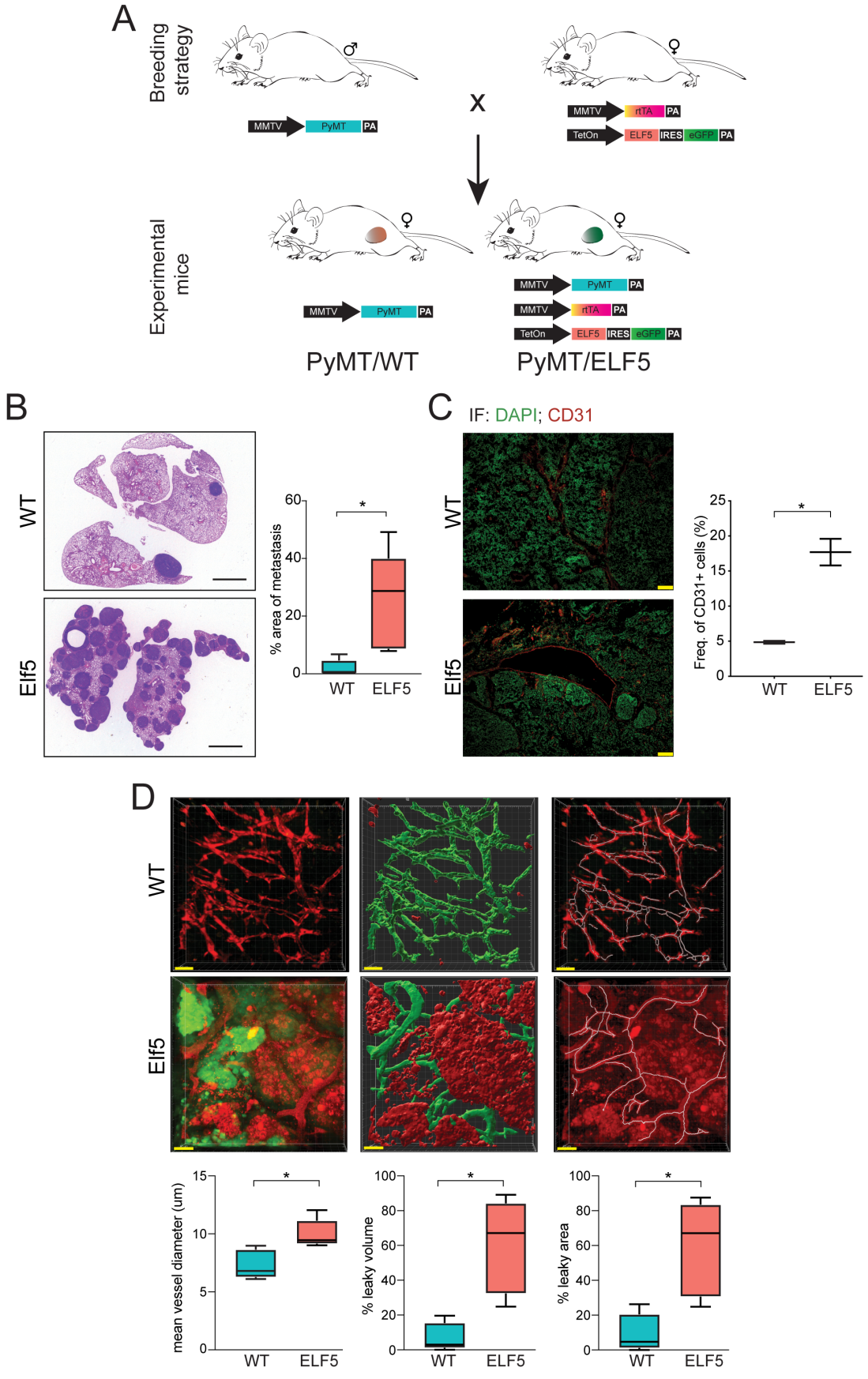
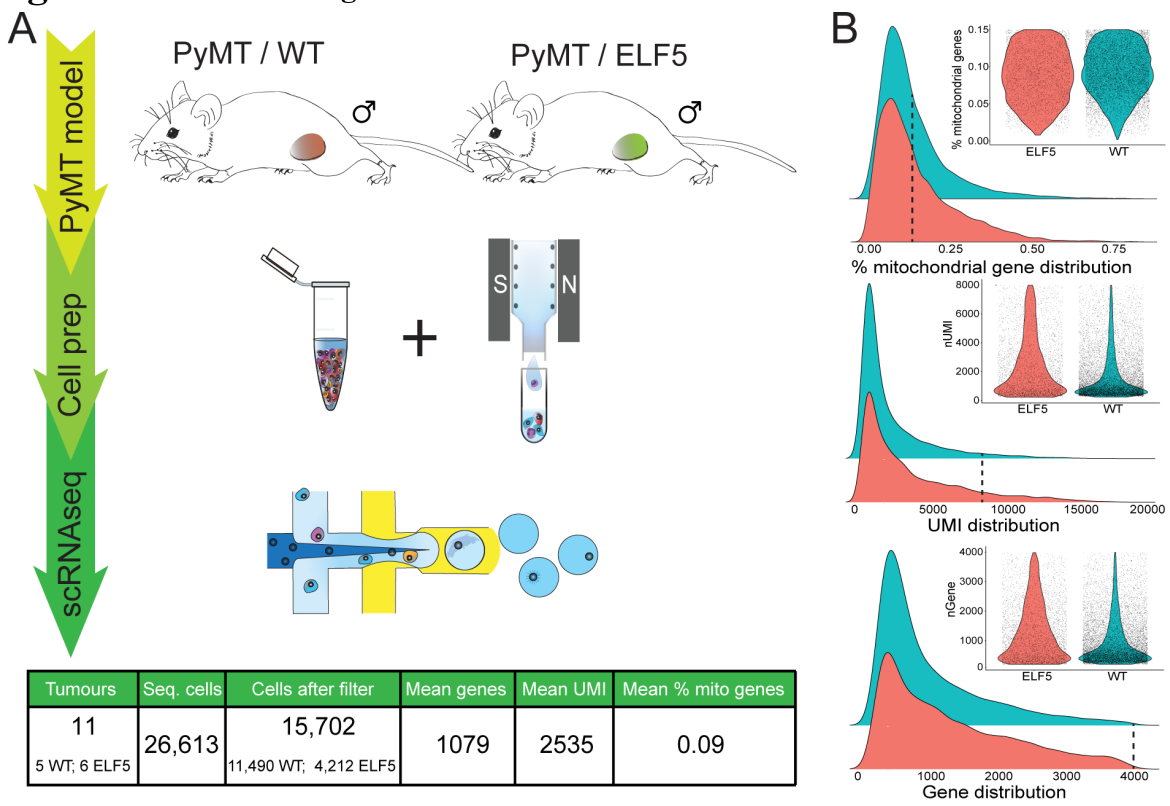
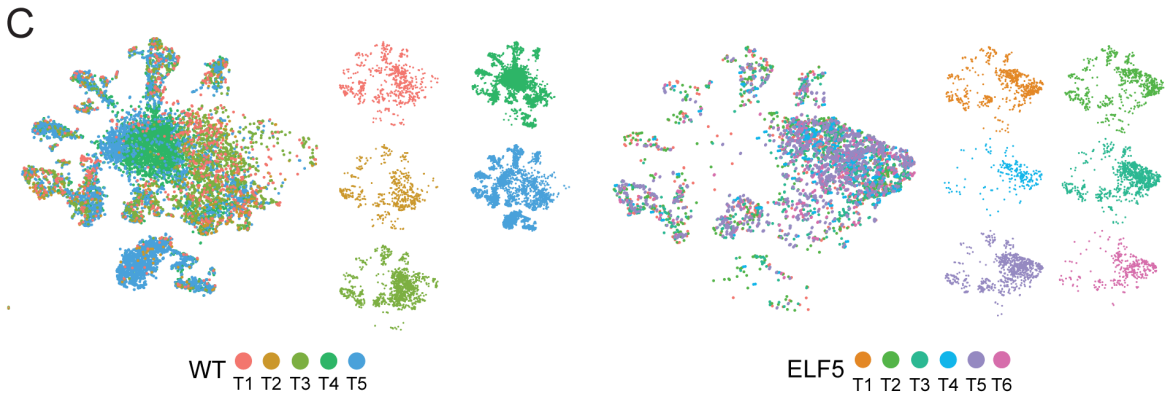


Figure S1. A) Schematic representation of the transgenic models used in this study, MMTV-PyMT and mammary restricted (MMTV) doxycycline-inducible (rtTA) Elf5 expression mouse**. B)** Images of histological sections of lungs from 14 week-old PyMT/WT and /ELF5 mice and the quantification of the area covered by metastases (n=5). Scale bar = 2.5 mm. Box plot representing the % area of lung metastasis. **C)** Representative pictures of immunofluorescence of blood vessels using CD31 antibody (scale bar = 100 μm) and box plot for the FACS quantification of CD31+ cells in dissociated tumours (n=3). **D)** Representative 3D reconstructions of mammary tumours imaged by multi-photon microscopy. Contrast in the tumour vasculature was achieved by tail-vein injection of blood quantum dots in the mice 20 minutes before imaging. Scale bar= 50 μm. Box plots represent the quantification of the vascular patency and vessel thickness (n=4). * denotes p-value <0.05.

Figure S2 related to Figure 1



Tumours	Seq. cells	Cells after filter	Mean genes	Mean UMI	Mean % mito genes
11 5 WT; 6 ELF5	26,613	15,702 11,490 WT; 4,212 ELF5	1079	2535	0.09



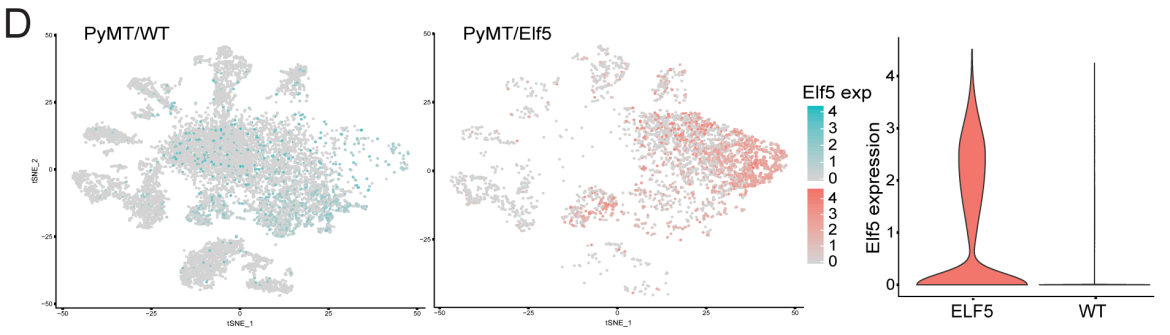


Figure S2. A) Overview of the experimental design for the Drop-seq analysis on PyMT/WT and /ELF5 tumours including a summary of the main experimental characteristics (bottom table). **B)** Quality cut offs (dashed lines) applied for the single cell RNAseq analysis, percentage of mitochondrial genes (15%), number of Unique Molecular Identifiers (UMI) per cell (8,000) and number of genes identifies per cell (4,000). The resultant

distribution after cut offs is shown in the insets**. C)** tSNE plots showing the distribution of cells per tumour replicate both merged and deconvoluted by tumour, on the left, five tumours from PyMT/WT and on the right, six tumours from Elf5. **D**) Feature and violin plots of Elf5 expression in the analysed PyMT/WT and PyMT/Elf5 tumours.

Figure S3 related to Figure 1 and Figure 2

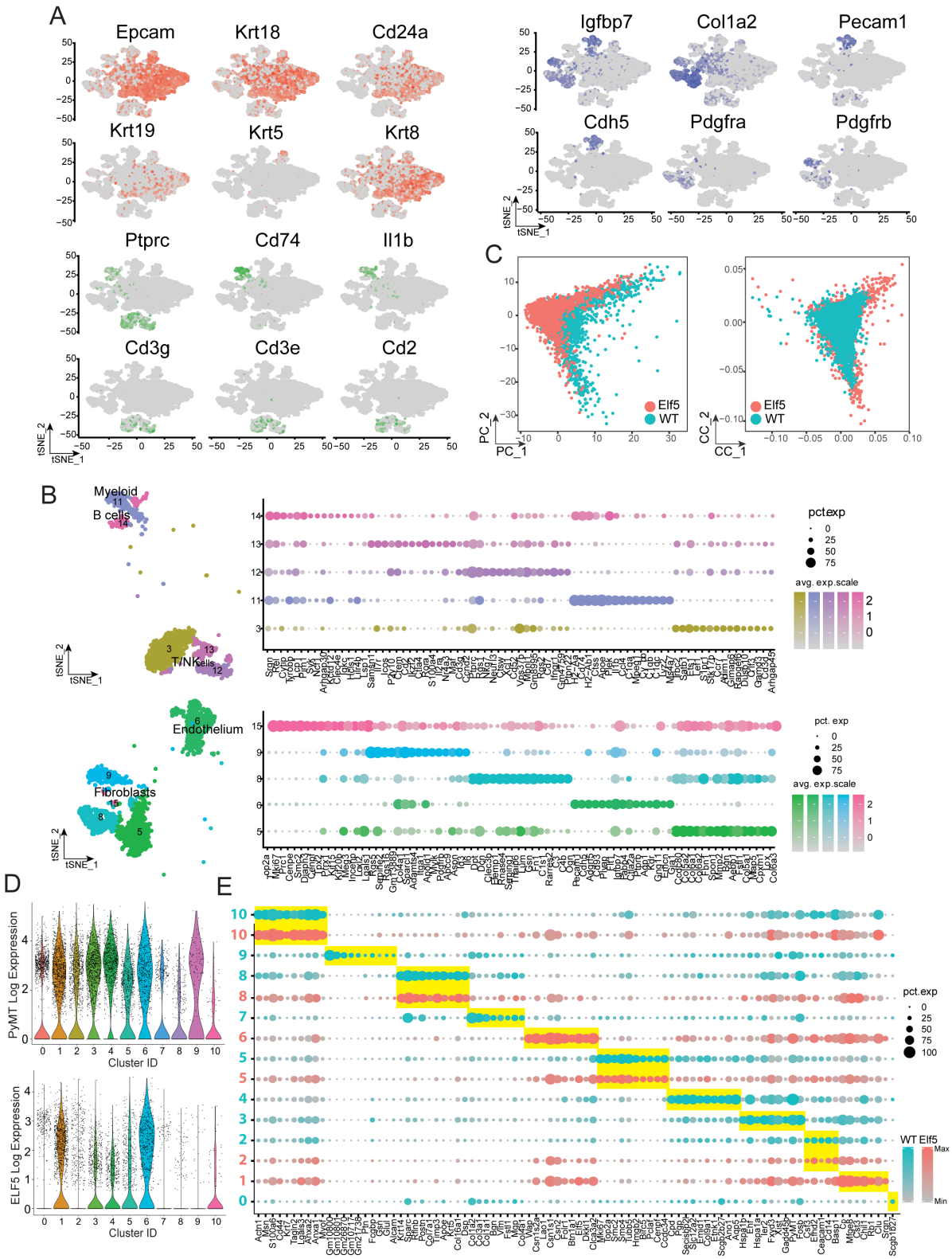


Figure S3. A) Feature tSNE plots showing the expression of canonical markers of each of the main cell lineages, red epithelial, blue stroma and green immune. **B)** Visualisation (tSNE and dot plots) of the top differential genes for each of the defined clusters in the immune lineage (top) and in the stroma (bottom). C) Comparison of the principal component analysis before (left panel) and after (right panel) CCA alignment grouped by genotype **D)** Violin plots showing Elf5 and PyMT gene expression in each epithelial cell cluster. **C)** Dot plot of the expression levels of the top differential marker genes in each of the PyMT clusters coloured by genotype. The yellow rectangles highlight the top genes represented by each cluster. The size of the dots represents the percentage of cells/cluster that express each particular gene (pct. exp) and the colour gradient shows the level of expression for each gene/cluster. Note both colours are shown only when the cluster was populated similarly by both genotypes according to Fig. 2C.

Figure S4 related to Figure 4

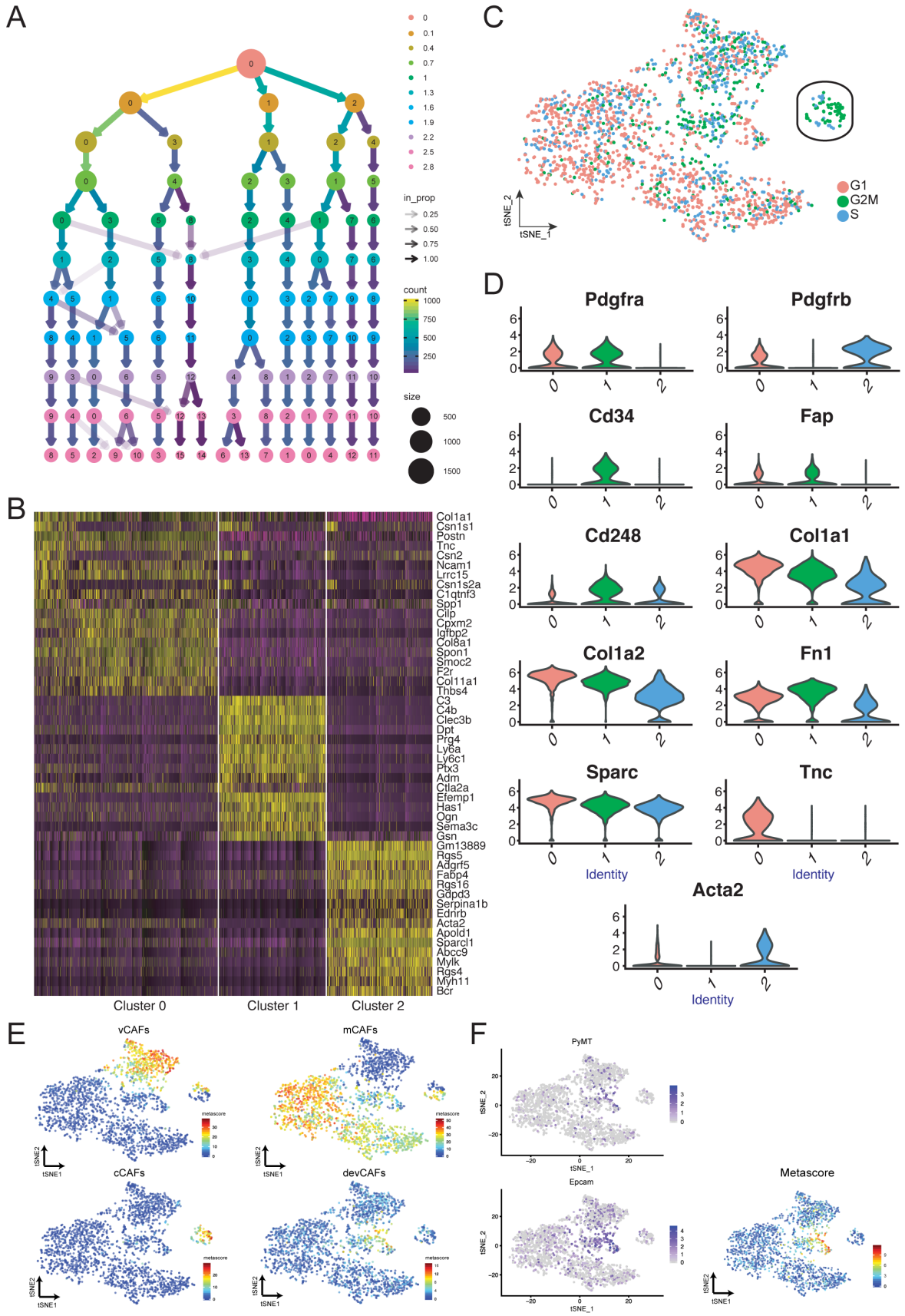


Figure S4. A) Cluster tree modelling the phylogenic relationship of the different clusters within the cancer-associated fibroblasts at different clustering resolutions. **B)** Gene-expression heatmap of the top expressed genes for each fibroblast lineage (FC). **C)** tSNE representation of the cell cycle stages of the cancer-associated fibroblasts as defined by gene expression signatures using tSNE coordinates. Circled area shows the cycling cluster characterised by a total absence of G1 cells. **D)** Violin plots displaying the classical marker genes that define CAFs for each of the three fibroblast clusters defined in resolution 0.1 (FC). **E**) Meta-signatures using the gene profiles from Bartoschek et al for cross-reference of the nomenclature previously used defining fibroblast subsets in the PyMT model: Vascular CAFs (vCAFs) corresponding with myofibroblasts with a contractile signature; matrix CAFs (mCAFs), described as the subpopulation with the strongest ECM signature, correspond with the secretory type of fibroblasts and in more extend to the ECM-CAFs within this group; cycling CAFs (cCAFs) completely overlap with the cell cycle signature, as shown in panel C and development CAFs (devCAFs) with an epithelial origin corresponds to a small population further clustered in resolution 1 (Figure 4E, cluster F5). **F)** Identification of a subset of fibroblasts that expressed the epithelial markers of PyMT and EpCAM, consistent with devCAFs in the Bartoschek et al publication.

Figure S5 related to Figures 4 and 7

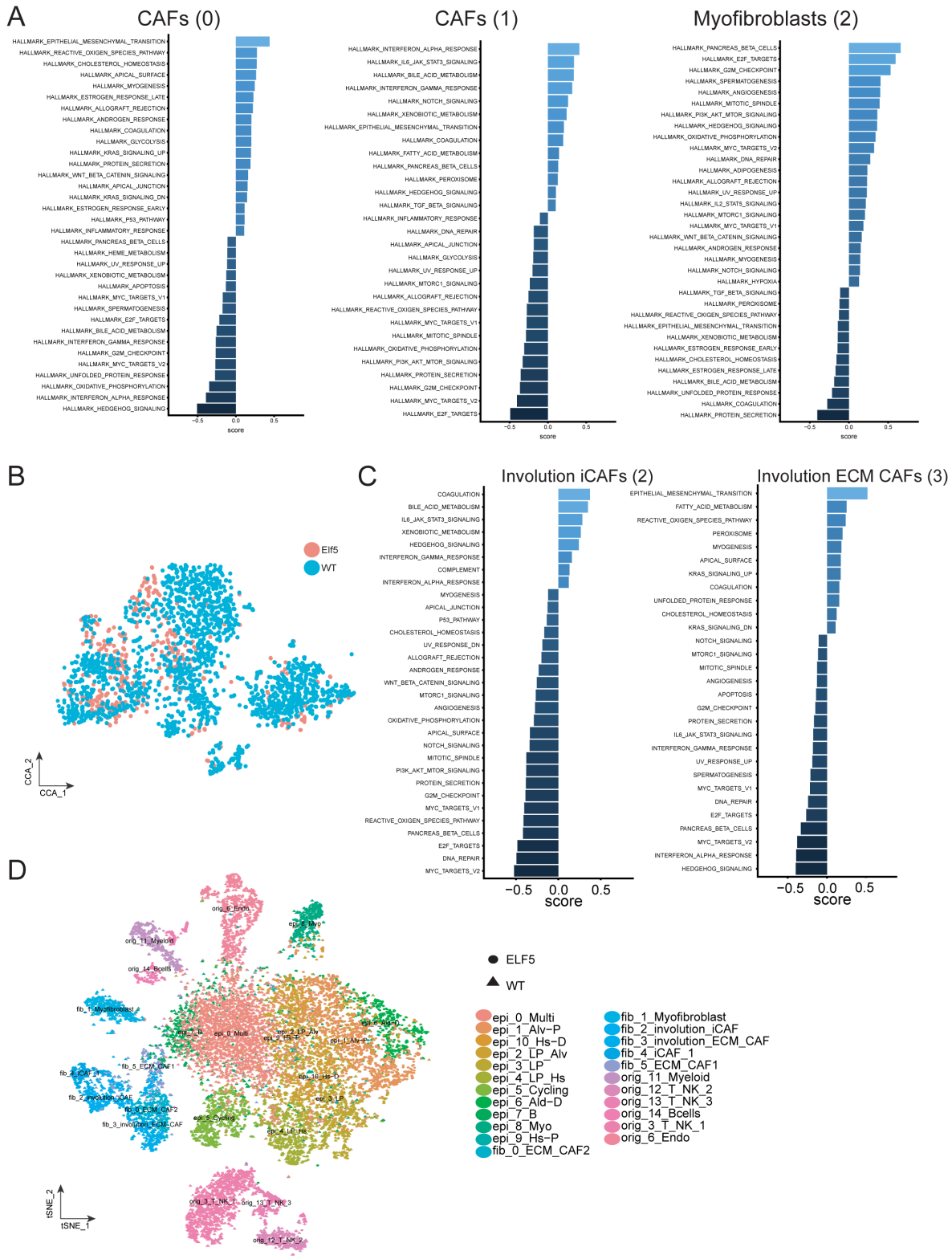


Figure S5. A) Waterfall plot of the GSVA analysis for the H Hallmarks from MSigDB in the CAF cluster defined by resolution 0.1 (CF); Cluster CF0 and CF1 were defined as “CAFs” and CF2 as “Myofibroblasts”. The variation score shows the hallmarks enriched (light blue, positive values) and not enriched (dark blue, negative values) for each cluster. **B)** tSNE representation of the CCA analysis of the CAF cell compartment revealing a similar polarisation between the two genotypes, PyMT/WT (blue) PyMT/ELF5 (red). **C)** Waterfall plot of the GSVA analysis for the H Hallmarks from MSigDB in the involution CAFs clusters defined by

resolution 1 (clusters F2 and F3). The variation score shows the hallmarks enriched (light blue, positive values) and not enriched (dark blue, negative values) for each cluster. **D)** tSNE plot showing the assignment of each cell type for the analysis of cell to cell interactions using CellphoneDB. The circles correspond to the cells from the Elf5 genotype and the triangles are the cells coming from the WT genotype.

Figure S6 related to Figure 4

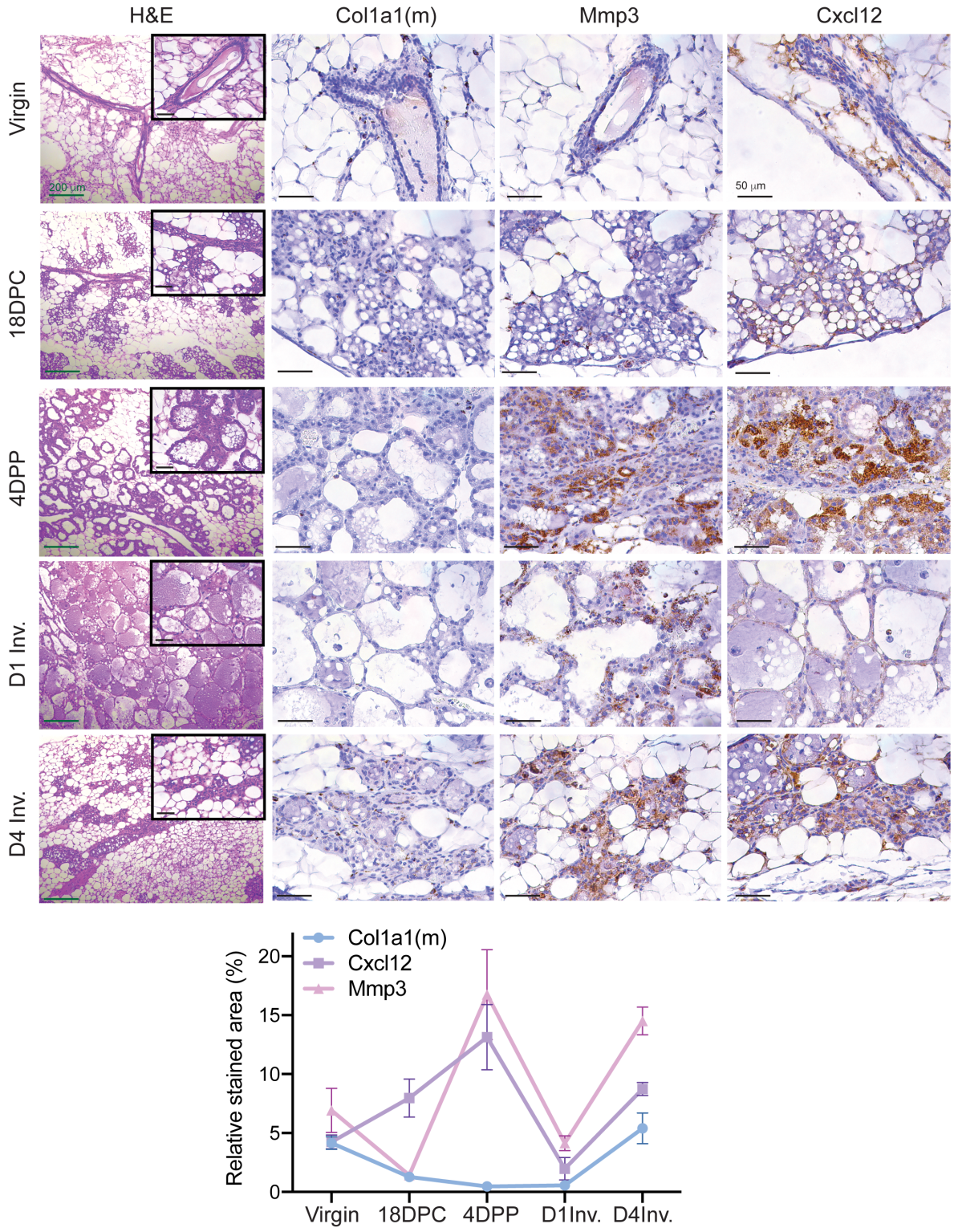


Figure S6. Histological analysis of a mammary gland differentiation series featuring virgin (V), late pregnancy (18 days *post-coitum*, 18DPC), established lactation (4 days *post-partum*, 4DPP), early involution (involution day 1, Inv. D1) and late involution (4 days involution, Inv. D4). First column shows representative images of Haemotoxylin and Eosin (H&E) staining of the mammary glands at low (10X, scale bar 200 µm) and high (40X, scale bar 50 µm) power magnification (inset). Representative images of the immunohistochemistry staining using Col1a1 (monomer), Mmp3 and Cxcl12 antibodies are shown at 40X magnification. Scale bar 50

μm. Bottom panel shows the automatic FIJI-based quantification of the positive staining represented as % positively stained area (n=2 for each stage and 5 sections per animal, SEM ±SD).

Figure S7 related to Figure 5

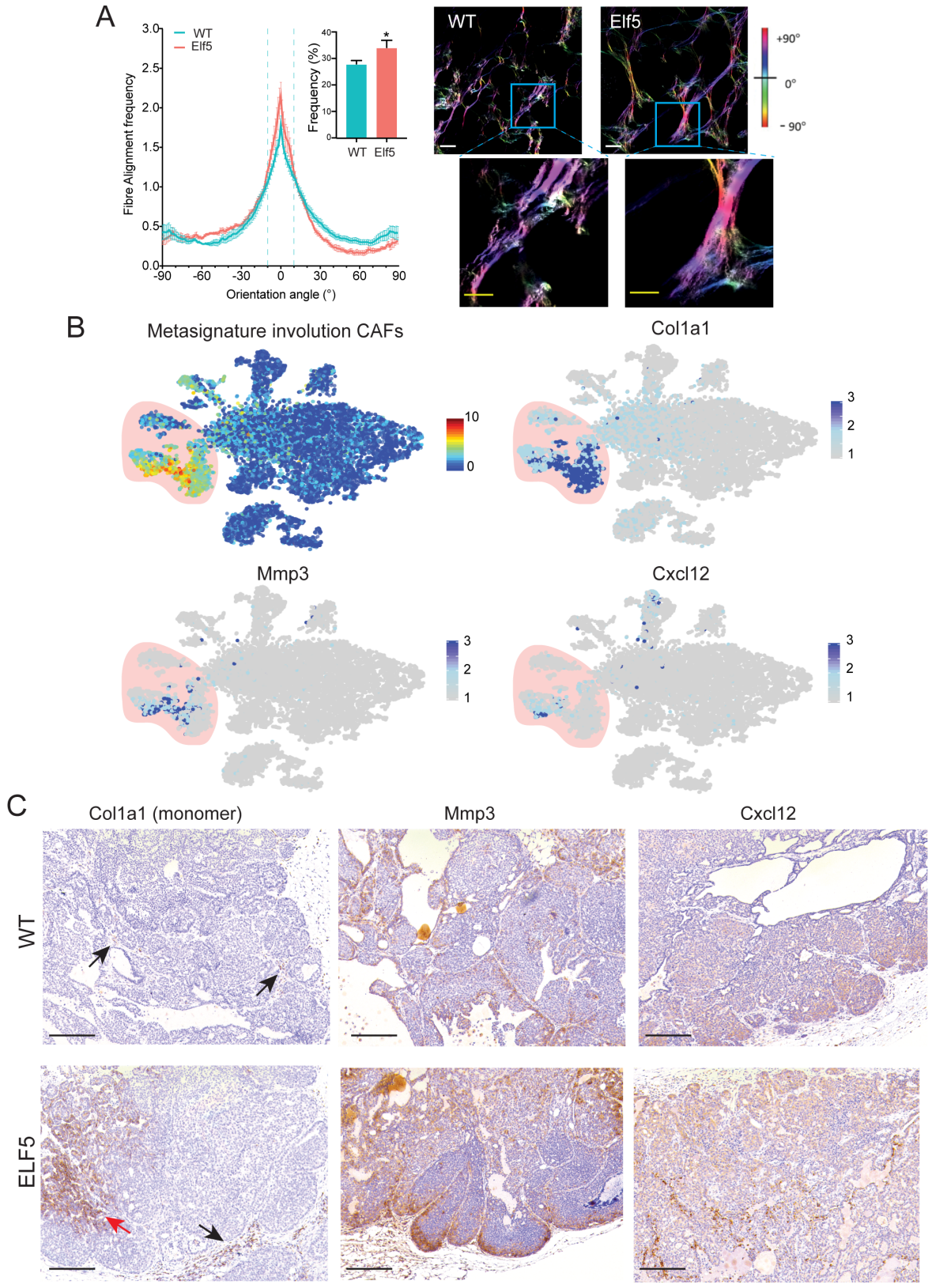


Figure S7. A) SHG images of PyMT/WT and PyMT/ELF5 tumours assessed for differences in fibre orientation angle and quantification of frequency of fibre alignment ranging from the peak alignment. Different colours correspond to specific angles of orientation n=6 PyMT/WT and n=4 PyMT/ELF5. Inset shows the cumulative frequency of fibre alignment +/- 10 degrees from peak. White scale bar= 50 µm, Yellow scale bar= 25 µm

(SEM ± SD, t-test, * denotes p-value <0.05.). **B)** Feature tSNE plots of all cell in tumours and metascore highlighting the expression of the invCAF markers (Col1a1, Mmp3 and Cxcl12) within the fibroblast compartment (red background, as annotated in Fig. 1). **C)** Low power (10X) magnification of representative images on the immunohistochemistry analysis of the Col1a1 (monomers), Mmp3 and Cxcl12 proteins on PyMT/WT and /ELF5 tumours. Black arrows denote cells with elongated structure and red arrows denote cells more infiltrated in the tumours and with a more rounded shape. Scale bar= 200 μm.

Data S1. Functional heterogeneity of PyMT tumours is associated to lineage origin (cont). Related to Figure 2:

Cluster 0 (C0) was consistent with undifferentiated cells or uncommitted progenitors (depleted Notch signalling). C0 was characterised by a high enrichment on proliferative and cell cycle associated hallmarks suggesting that these could be a cell of origin of a highly aggressive cancer. Interestingly C0 was enriched on inflammatory associated hallmarks, such as IL6/Stat3, Interferon α and γ, inflammatory response, allograft rejection and ROS. This cluster is consistent with immunogenic tumours such as those of the TNBC subtype.

The luminal progenitor group was formed by clusters Cluster 2, 3 and 4. Cluster 2 was functionally similar in both genotypes with a slight cell enrichment in Elf5 tumours. C2 was characterised by a profound depletion on differentiation pathways such as Notch and TGFβ as well as proliferative and metabolic pathways, this suggests that C2 has a highly multipotent progenitor profile with high epithelial score (depletion of EMT), consistent with a transitory amplified progenitor. C3 was characterised by the expression of genes signatures associated to IFN α and γ, this suggests that through the differentiation to the specific lineages and, in particular, in the alveolar compartment, IFNα and g signalling gets progressively lost. For example, C0 presented enrichment for IFNα and g, this is conserved in the luminal progenitor cells C3 and only partially conserved in C1 (IFNα enriched and IFNγ depleted) to end up negatively both enriched in C6 and poorly enriched in C9. Interestingly, C3 as a luminal progenitor was depleted for signatures associated to hormone response (androgen, estrogen response early and late, IL2/Stat5) as well as hallmarks associated with TME communication (e.g. hypoxia, coagulation, angiogenesis) and inflammation (inflammatory response, ROS, complement) and a markedly low EMT, proliferative and differentiation pathways. This cluster was consistent with a non-committed luminal progenitor identity, with no clear commitment to either alveolar or the hormone sensing lineages. Albeit a slight enrichment in WT tumours, the functional activity of these cells is very similar between genotypes indicating a true luminal progenitor identity. Amongst the few differences that could be found between C3_WT and C3_Elf5 cells, it stood out an increase in the score for the signatures related to Notch and TGFβ, this could be consistent with a bias towards a transition to a more differentiated state in Elf5 tumours.

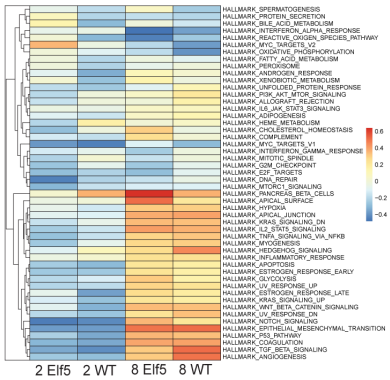
The functional annotation for Cluster 4 is consistent with the previous results of marker genes and trajectory analysis shown in Figs 3 and 4, placing this cluster as a luminal progenitor committed towards the hormone-sensing lineage. Cluster 4 was heavily enriched in WT tumours where Hormone Sensing differentiated cells (C9) are much more frequent. LPs are characteristically depleted in hormone signalling hallmarks, however C4 scores for hormone related hallmarks (ER_early [-0.110]; ER_late [-0.128]; Androgen [-0.017]) were comparatively higher than C2 (ER_early [-0.-0.323]; ER_late [-0.171]; Androgen [-0.164])and C3 (ER_early [-0.209]; ER_late [-0.168]; Androgen [-0.238]), consistent with their transition towards a Hs lineage but still retaining the functional characteristics of a luminal progenitor. Similarly to the other clusters, C4 Elf5 cells presented high Notch signalling enrichment score compared to C4 WT.

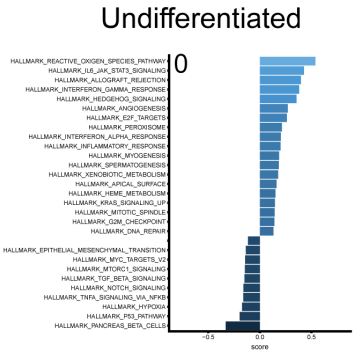
Consistently with previous studies, Elf5-driven cell clusters associated to the alveolar lineage, clusters 1 and 6 (C1 and C6), presented a negative enrichment on estrogen responsive genes, which is presumably associated to the opposing transcriptional force that Elf5 drives on ER signalling (Gallego-Ortega et al., 2015; Kalyuga et al., 2012). Similarly, a downregulation of genes involved in EMT were characteristics of these two clusters (Chakrabarti et al., 2012; Gallego-Ortega et al., 2015). These clusters were also characterised by an enrichment of the IL6/Stat3 signalling pathway, a pathway associated with tumour progression and metastasis but also with induction of cancer immunetolerance (Johnson et al., 2018), and cholesterol homeostasis, also associated with breast cancer progression, inflammation and metastasis. Interestingly aspects of innate immune related pathways (Hallmark Complement) were highly enriched in both alveolar clusters. Reduced cell cycle progression and cell proliferation has been associated to Elf5 expression (Kalyuga et al., 2012) and tumour growth (Gallego-Ortega et al., 2015) this is consistent with the observed decline of Hallmarks related to cell cycle progression (G2M, E2F, mitotic spindle and Myc targets) and cell growth and proliferation (i.e Hedgehog). All these characteristics have been previously associated to Elf5 rich tumours including in the PyMT mouse model (Gallego-Ortega et al., 2015). In addition, C6 was characterised by enriched genes in response to hypoxia, coagulation and involved in angiogenesis, suggesting a strong sensitivity of these cells to signals of the extracellular ecosystem, as Elf5 tumours have been shown to exhibit strong angiogenesis, and vascular leakiness (Gallego-Ortega et al., 2015). Similarly, cells from C6 seemed to be responding to oxidative

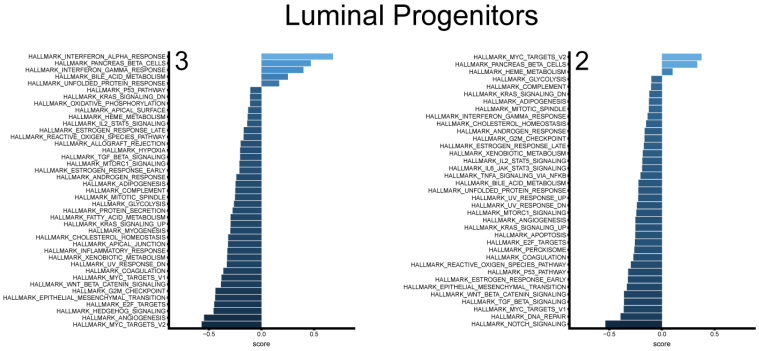
stress (ROS and Xenobiotic signalling pathways), this is associated with the formation of chronic inflammation and recruitment of leukocytes, and the depletion on IFN signalling in this cluster might be associated to increase of immune tolerance and recruitment of immunosuppressive leukocytes, something that has been previously shown in this model associated with Elf5-high tumours (Gallego-Ortega et al., 2015).
The hormone sensing linage was represented by Cluster 9. This cluster was characterised by an enrichment of a number of hallmarks associated to hormonal response, Estrogen, (late and early), androgen and IL2/Stat5 pathways. Interestingly and similarly to what it was observed in C6, these more differentiated cells in both lineages were strongly associated to communication with the TME, including hypoxia, angiogenesis and inflammation associated hallmarks.
The basal/myoepithelial subtype was formed by clusters 7 and 8. Cluster 7 (C7) is consistent with cells from the myoepithelial lineage. These cells were heavily involved in inflammation-associated hallmarks (complement, allograft rejection, inflammatory response, etc.) and hallmarks associated with TME interactions such as hypoxia and angiogenesis. Such characteristics were also found in the undifferentiated cells from C0, and often undifferentiated multipotent cells and myoepithelial cells showed similar gene signatures. However, in direct contrast with cells from C0, C7 cells present a strong mesenchymal phenotype, characterised by EMT and a higher myogenesis score, and an enrichment in Notch signalling, indicating restriction of stem cell activity and commitment towards a more differentiated fate.

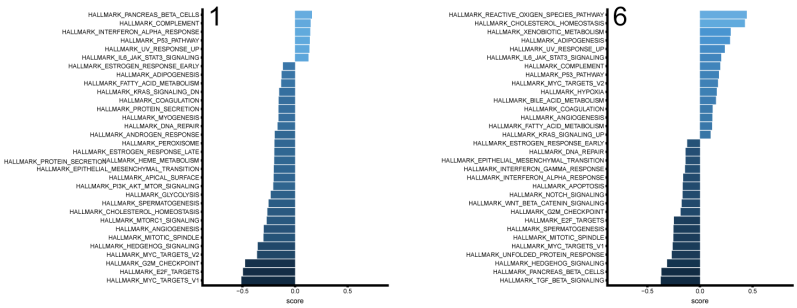
Cluster 8 was consistent with a basal/ myoepithelial lineage and cells of this lineage were more frequent in WT tumours, presumably by the luminal skew produced by the Elf5 differentiation force. Cells that belong to this cluster presented enrichment for Notch and TGFβ pathways, indicating a differentiated stage, however these cells are rich in EMT markers, typical of a myoepithelial lineage. Amongst the differences by genotype, a negative enrichment of innate immune inflammatory properties (complement, IFNγ, and oxidative stress signals) as well as reduced proliferative hallmarks (mitotic spindle, G2M checkpoints and E2F targets) in Elf5 cells compared with WT cells.

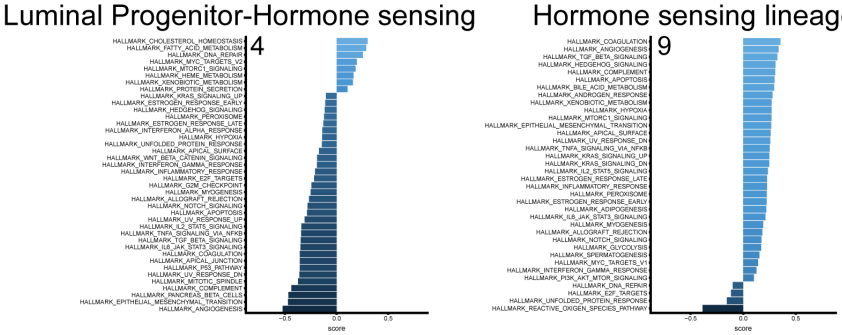
Data S2 related to Figures 2 and 3:

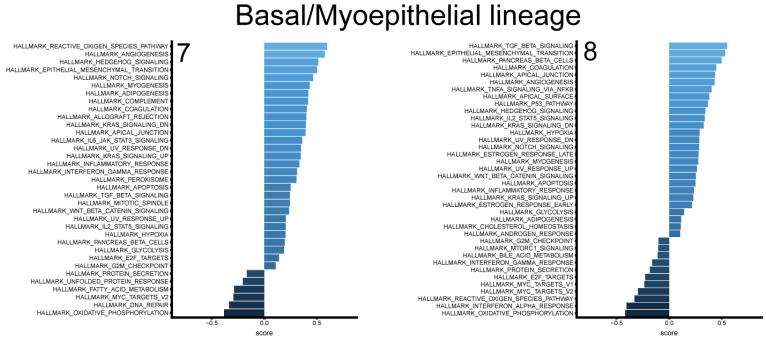












Data S2. GSVA analysis for the H Hallmarks from MSigDB in the cancer epithelial clusters. Waterfall plots are presented for clusters mainly formed by one genotype and a heatmap is presented in the clusters formed by the two genotypes (according to Figure 2C). The variation score shows the hallmarks enriched (light blue, positive values) and decreased (dark blue, negative values) for each cluster.

Table S1. Percentage of cells per tumour replicate for each cluster, Related to Figure 1.

Tumours	Clusters															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ELF5 T1	1.4	6.3	31.1	0.3	25.4	5.0	5.8	10.0	4.7	2.4	2.8	0.9	0.8	0.9	2.0	0.3
ELF5 T2	1.6	4.8	24.8	2.8	27.2	6.0	9.4	5.4	6.3	2.5	3.6	1.2	1.1	0.6	2.0	0.6
ELF5 T3	0.9	9.5	39.6	1.1	25.6	3.9	3.3	6.0	4.7	0.3	2.7	0.6	0.6	0.2	0.7	0.1
ELF5 T4	2.6	8.6	50.9	0.9	19.8	1.3	4.3	3.0	4.3	0.4	1.7	1.3	0.0	0.4	0.4	0.0
ELF5 T5	0.9	11.1	42.8	0.0	20.0	3.1	3.7	7.9	5.9	0.6	2.1	1.0	0.2	0.5	0.0	0.0
ELF5 T6	0.6	14.5	39.6	0.0	25.4	2.8	2.0	6.3	3.4	1.1	2.3	0.6	0.0	0.6	0.6	0.3
WT T1	10.0	23.2	17.3	2.4	3.8	7.3	9.1	5.3	9.5	1.5	3.6	2.0	1.4	0.8	2.4	0.4
WT T2	3.3	39.4	10.5	4.4	3.1	5.6	5.1	5.9	10.3	1.8	2.6	1.6	2.1	0.7	2.9	0.7
WT T3	3.3	32.8	26.6	0.3	3.9	5.5	5.3	8.5	6.1	1.6	3.2	1.3	0.5	0.1	0.5	0.3
WT T4	51.9	14.4	0.9	0.7	0.0	5.2	4.5	3.3	1.7	4.2	2.4	4.0	3.1	2.6	0.9	0.2
WT T5	13.6	10.2	2.2	23.7	0.2	9.5	7.1	4.1	4.1	5.9	3.3	4.0	4.4	3.3	3.4	1.0

Table S3. Functional annotation of each cell cluster from the tumour epithelium, Related to Figures 2 and 3

Mammary lineage	Cluster number	HALLMARK White bg: positive enrichment Gray bg: negative enrichment	Function
Undifferentiated	0	Hedgehog Signalling	Cell proliferation
		E2F Targets	
		Mitotic Spindle	
		G2M Checkpoint	
		Reactive Oxygen Species Pathway	Inflammation/Immune Response
		IL6 JAK STAT3 Signalling	
		Allograft Rejection	
		Interferon Gamma Response	
		Interferon Alpha Response	
		Inflammatory Response	
		Xenobiotic Metabolism	
		Angiogenesis	Interaction with TME
		Notch Signalling	Cell differentiation
		Epithelial Mesenchymal Transition	EMT
Luminal progenitors	2	Notch signalling	Cell Differentiation
		TGF beta Signalling	
		Mitotic Spindle	Cell Proliferation
		G2M Checkpoint	
		E2F Targets	
		Myc Targets v1	
		Cholesterol Homeostasis	Cell Metabolism
		Bile acid metabolism	
		Peroxisome	
	3	Interferon Alpha Response	Immune response
		Interferon Gamma Response	
		Androgen Response	Hormone Response
		Estrogen Response Late	
		Estrogen Response Early	
		IL2 STAT5 Signalling	
		Hypoxia	Interaction with TME
		Angiogenesis	
		Coagulation	Inflammation
		Inflammatory Response	
		Reactive Oxygen Species Pathway	

		Complement	
		Epithelial Mesenchymal Transition	EMT
		Mitotic Spindle	Cell Proliferation
		G2M Checkpoint	
		E2F Targets	
		Myc Targets v1	
		Myc Targets v2	
		Hedgehog Signalling	
		TGF beta Signalling	Cell Differentiation
		Myogenesis	
	4	Cholesterol Homeostasis	Cell Metabolism
		Fatty Acid metabolism	
		Heme metabolism	
		Epithelial Mesenchymal Transition	EMT
		TNFA signalling via NFKB	Inflammation
		IL6 JAK STAT3 Signalling	
		Complement	
		Coagulation	
Alveolar lineage	1	IL6 JAK STAT3 Signalling	Inflammation
		Complement	
		Interferon Alpha Response	
		Estrogen Response Early	Hormone Response
		Estrogen Response Late	
		Epithelial Mesenchymal Transition	EMT
		Mitotic Spindle	Cell proliferation
		G2M Checkpoint	
		E2F Targets	
		Myc Targets v1	
		Myc Targets v2	
		Hedgehog Signalling	
		Interferon Gamma Response	Immune Response
	6	IL6 JAK STAT3 Signalling	Inflammation
		Complement	
		Coagulation	
		Cholesterol Homeostasis	Cell Metabolism
		Hypoxia	Interaction with TME
		Angiogenesis	
		Reactive Oxygen Species Pathway	Oxidative stress
		Estrogen Response Early	Hormone Response
		Epithelial Mesenchymal Transition	EMT
		Mitotic Spindle	Cell proliferation
		G2M Checkpoint	
		E2F Targets	
		Myc Targets v1	
		Hedgehog Signalling	
		Interferon Gamma Response	Immune Response
		Interferon Alpha Response	
Hormone sensing lineage	9	Androgen Response	Hormone Response
		Estrogen Response Late	
		Estrogen Response Early	
		IL2 STAT5 Signalling	
		Hypoxia	Interaction with TME
		Angiogenesis	
		Coagulation	Inflammation
		Complement	
		Inflammatory Response	

		IL6 JAK STAT3 Signalling	
		Allograft Rejection	
		Interferon Gamma Response	
Basal/Myoepithelial lineage	7	Complement	Inflammation
		Coagulation	
		Allograft Rejection	
		Inflammatory Response	
		Reactive Oxygen Species Pathway	
		IL6 JAK STAT3 Signalling	
		Interferon Gamma Response	
		Angiogenesis	Interaction with TME
		Hypoxia	
		Myogenesis	Cell differentiation
		Notch Signalling	
		Epithelial Mesenchymal Transition	EMT
	8	Notch Signalling	Cell differentiation
		TGF beta Signalling	
		Myogenesis	
		Epithelial Mesenchymal Transition	EMT

Table S4. Drop-seq primers from Makosco et al 2015, Related to STAR Methods

Primer Name	3'-5' Sequence	Application
Template_Switching_Oligo	AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG	Drop-seq: RT for sscDNA
TSO_PCR	AAGCAGTGGTATCAACGCAGAGT	Drop-seq: Second strand synthesis and cDNA amplification
P5-TSO_Hybrid	AATGATACGGCGACCACCGAGATCTACACGCCTG TCCGCGGAAGCAGTGGTATCAACGCAGAGT*A*C	Drop-seq: Nextera library prep
Nextera_N701	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTC TCGTGGGCTCGG	Drop-seq: Nextera library prep multiplexing
Nextera_N702	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTC TCGTGGGCTCGG	Drop-seq: Nextera library prep multiplexing
Nextera_N703	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTC TCGTGGGCTCGG	Drop-seq: Nextera library prep multiplexing
Nextera_N704	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTC TCGTGGGCTCGG	Drop-seq: Nextera library prep multiplexing
Nextera_N705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTC TCGTGGGCTCGG	Drop-seq: Nextera library prep multiplexing
Read1CustomSeqB	GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC	Drop-seq: Sequencing