Single-cell RNAseq uncovers involution mimicry as an aberrant development pathway during breast cancer metastasis

Fatima Valdes-Mora[1,2#*], Robert Salomon[3,4#], Brian Gloss[5#], Andrew MK. Law[5], Kendelle Murphy[6], Daniel L. Roden[2,3], Lesley Castillo[5,7], Yolanda Colino-Sanguino[1], Zoya Kikhtyak[5], Nona Farbehi[3], James RW. Conway[6], Samantha R. Oakes[2,7], Neblina Sikta[8], Seán I. O'Donoghue[8,9,10], Thomas Cox[2,11], Paul Timpson[2,6], Christopher J. Ormandy[2,7^] and David Gallego-Ortega[2,5^*].

1. Histone Variant Laboratory, Genomics and Epigenetics Division, Garvan Institute of Medical Research. Sydney, NSW Australia.
2. St. Vincent's Clinical School, Faculty of Medicine, UNSW Sydney, NSW, Australia.
3. Garvan-Weizmann Centre for Cellular Genomics. Garvan Institute of Medical Research. Sydney, NSW, Australia.
4. *Current:* Institute for Biomedical Materials and Devices, University of Technology Sydney. Sydney, NSW, Australia
5. Tumour Development Laboratory, The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Sydney, NSW Australia.
6. Invasion and Metastasis Laboratory, The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Sydney, NSW Australia.
7. Cancer Biology Laboratory, The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Sydney, NSW Australia.
8. BioVis Centre, Genomics and Epigenetics Division, Garvan Institute of Medical Research. Sydney, NSW Australia.
9. CSIRO Data61, Eveleigh NSW 2015, Australia
10. School of Biotechnology and Biomolecular Sciences, University of New South Wales, Kensington NSW 2052, Australia
11. Matrix and Metastasis Laboratory, The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Sydney, NSW Australia.

joint-first authors
^ joint-senior authors
* Correspondence: d.gallego@garvan.org.au / f.valdesmora@garvan.org.au

Summary

Both luminal and basal breast cancer subtypes originate in the mammary luminal progenitor cell compartment. Basal breast cancer is associated with younger age, early relapse, and high mortality rate. Here we used unbiased droplet-based single-cell RNAseq to elucidate the cellular basis of tumour progression during the specification of the basal breast cancer subtype from the luminal progenitor population. Basal–like cancer cells resembled the alveolar lineage that is specified upon pregnancy and showed molecular features indicative of an interaction with the tumour microenvironment (TME) including EMT, hypoxia, lactation and involution. Involution signatures in luminal breast cancer tumours with alveolar lineage features were associated with worse prognosis and features of basal breast cancer. Our high-resolution molecular characterisation of the tumour ecosystem also revealed a highly interactive cell-cell network reminiscent of an involution process. This involution mimicry involves malignant education of cancer-associated fibroblasts and myeloid cell recruitment to support tissue remodelling and sustained inflammation. Our study shows how luminal breast cancer acquires a post-lactation developmental program to shift molecular subtype and promote tumour progression, with potential to explain the increased risk of cancer during the post-partum period.

Keywords: scRNAseq, basal breast cancer, alveolar lineage, cancer-associated fibroblast, involution, tissue remodelling.

1. Introduction

The mammary gland is a unique organ that undergoes a series of developmental processes that mostly occur postnatally, with profound tissue morphogenesis during puberty and pregancy[1]. During pregnancy, alveolar milk-secretory epithelial cells differentiate from the CD61+ luminal progenitor population (also EpCAM^{high}/Sca1^{+}/CD49b^{+})[2] driven by Prolactin- and Progesterone-induced up-regulation of Elf5, the master transcriptional regulator of the alveolar cell fate[3,4].

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[5,6]. Involution is a highly complex multicellular process, neutrophils and macrophages engulf the apoptotic bodies of the collapsing alveoli and regulate inflammation by suppressing the adaptive immune response; fibroblasts are activated to fill the areas left by the dying epithelium in a pseudo-wound healing process[7] and intense adipogenesis is activated. Thus, the exquisitely regulated morphogenesis of the mammary gland requires a high degree of cell-to-cell communication within the mammary epithelium, but also throughout a vast array of cell types from the tissue microenvironment.

One of the most active areas of research in the last years has been the study of the contribution of normal process during mammary gland development to breast cancer. Some mammary epithelial cell lineages have been proposed as the cell of origin of the different breast cancer subtypes[8,9], including those with specific genetic aberrations[9,10]. The association of developmental cues with breast cancer is not restricted to the epithelial compartment, as tissue morphogenesis encompasses dramatic changes within the epithelial ductal tree as well as its surrounding stroma. Thus, these events of tissue remodelling and often hijacked in breast cancer as drivers of tumour progression[11]. In this context, pregnancy-associated breast cancer (PABC) is correlated with metastatic disease and poor prognosis (reviewed in[12]). Despite the effects of the hormonal load associated with the onset of pregnancy, there is increasing evidence pointing to a direct effect of the TME and the events of tissue remodelling that occur in the mammary gland during pregnancy, and in particular those associated to involution, as the driving force behind the higher mortality associated to PABC[7,12,13].We have previously shown that pregnancy-associated transcriptional networks driven by Elf5 specifies the basal subtype of breast cancer, promoting a lethal phenotype, characterised by endocrine insensitivity and resistance

to therapy[14] and the promotion of metastatic disease through a profound alteration of the TME[15].

Here we use mammary specific induction of Elf5 expression in the context of the luminal MMTV/PyMT mouse mammary cancer, to study the tumour progression mechanisms associated to the specification of the basal breast cancer subtype at single-cell resolution. This study provides an unprecedented characterisation of the intense TME remodelling acquired by the activation of pregnancy-induced developmental pathways during breast carcinogenesis. Our analysis revealed the cellular diversity of breast tumours, characterising cell subclasses that resemble the mammary epithelial lineages and identifying strong plasticity within the luminal progenitor compartment, the origin of luminal and basal cancers. We exposed the functional heterogeneity of the different cancer cell species in their contribution to breast cancer progression: EMT, collagen deposition, inflammation and hypoxia. Using our high-resolution molecular landscape of breast tumours, we mapped cell-to-cell interactions within the TME, revealing a multicellular mechanism of tumour progression that encompasses immune suppression and tissue remodelling, two central processes involved in the promotion of metastatic disease[15]. We demonstrate that this mechanism of tumour progression is originated in the alveolar cancer cell lineage as a consequence of a subverted involution process. Finally, we provide evidence at the single-cell resolution of the molecular pathways that alveolar cells elicit to recruit Myeloid-Derived Suppressor Cells (MDSCs) and activate specific cancer-associated fibroblasts (CAFs) subpopulations; two central players in this process of involution mimicry.

2. Materials and Methods

Animal statement, doxycycline induction

To promote tumour progression, ELF5 expression was specifically induced in mammary epithelial cells since puberty (6 weeks-old) through doxycycline containing food as performed previously [15]. At ethical endpoint (7-13% tumour/body weight) mice were euthanized and tumours harvested and processed for single cell digestions as previously performed [15].

Tumour digestion and FACS

Mammary tumours were collected from MMTV-PyMT mice at 14 weeks of age, when tumour weights were 10±3% of total weight and digested to single cell suspension as performed before[15] (Extended Methods). Briefly, cells were sorted on the FACSAriaIII (Becton Dickinson) following automated compensation (FACSDiVa) using a 100um nozzle. Cell were sorted using a purity mask to ensure purities of the sorted population in excess of 99%. Antibodies with conjugated fluorophores for sorting stromal cells used included CD140b (clone REA634) from MACS (Miltenyi Biotech), CD140a (clone APA5) from Biolegend, and CD34 (clone RAM34) from BD Pharmingen.

Auto macs cleanup

The viability of the cells was assessed by FACS. All tumours that contained less than 80% viable cells were labelled with Annexin specific MACS beads and dead cells were removed by passing the labelled cells through the autoMACS® Pro (Miltenyi). Highly 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 Dropseq [16] following the Online Drop-seq Laboratory Protocol version 3.1 (www.mccarrolllab.com/dropseq). 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 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/ per run were sequenced.

Bioinformatic analysis (see Extended Methods and SuppFig, 1)

Second Harmonic Generation (SHG) collagen imaging and analysis

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 tumour samples, 5 representative regions of interest (512 μm x 512 μm) per tumour 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 tumour 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.

Polarized light microscopy of picrosirius red stained sections and quantification of collagen content and birefringence.

Paraffin-embedded samples were cut into 4 μm sections and stained with 0.1% picrosirius red for fibrillar collagen according to manufacturer's instructions. As previously performed [17,18] polarized light imaging was performed on a Leica DM6000 fitted with a polarizer in combination with a transmitted light analyser. Quantitative intensity measurements of fibrillar collagen content and birefringent signal were carried out using in house scripts in ImageJ. The relative area of red-orange (high birefringent) fibres, yellow (medium birefringent) fibres, and green (low birefringent) fibres (as a % of total fibres) was calculated.

Hypoxia analysis

Tumours and lungs were harvested and fixed in 10% neutral-buffered formalin overnight. FFPE tissues were sectioned to 4μm for IHC staining as described in [15]. Following the manufacturer's instructions, the FITC-conjugated pimonidazole primary antibody (Hypoxyprobe, 1:1000) and the anti-FITC secondary antibody (Hypoxyprobe 1:100) was used to stain for the pimonidazole adducts. Quantification of hypoxia in IHC was performed using Andy's DAB Algorithm [19].

Mammary gland whole mounting and immunohistochemistry

Mouse mammary glands from FVB/N mice with established lactation and mammary tumours from PyMT/WT and PyMT/ELF5 were fixed, defatted, dehydrated and carmine stained and paraffin embedded as previously described [20]. Sections were stained with haematoxylin and eosin and with a mouse milk specific protein antibody

(Accurate Chemical & Scientific CO, Cat #YNRMMSP) under conditions previously described [20].

Western Blot and Immunohistochemistry

Western Blot analyses of protein were performed as previously detailed in [14]. The primary antibodies not described in Kalyuga et al were used as follows: E-Cadherin (BD bioscience #610181, 1:10,000) and Vimentin (Leica Biosystems # VIM-572-L-CE).

Mouse mammary glands and tumours from PyMT mice were fixed, defatted, dehydrated, stained in Carmine alum and Paraffin embedded as previously described [20]. Sections were stained with the rabbit anti-mouse milk proteins antibody (Accurate Chemical & Scientific CO, # YNRMTM) diluted 1:12,000 as per [20].

3. Results

Unbiased massively-parallel scRNAseq captures cell heterogeneity of MMTV-PyMT mammary tumours.

A total of fifteen MMTV-PyMT/WT and /ELF5 mouse mammary tumours were harvested and analysed using Drop-seq [16]. A total of 15,702 cells passing the QC filters (see material and methods) were visualised using tSNE. Importantly cells from the biological replicates contributed to all cell clusters, indicating consistent cell sampling between all replicates. We then identified the presence of the main cellular compartments in the tumours using cell signatures defined by Aran and colleagues[21] (Fig. 1A). The majority of cells were defined as epithelial cells (63.16%); tumour associated stroma represented 20.41% of the cells identified; and the tumour-infiltrated immune cell compartment accounted for 16.44%. The composition of PyMT tumours found using Drop-seq was consistent with cell surface marker analysis (FACS) using EpCAM to define epithelial cells, CD45 to define leukocyte and non-stained cells to define the stroma (Fig. 1B). A tSNE visualisation of the three cellular compartments and their associated top expressed markers is shown in Figure 1C; and we used the RNA expression of canonical markers for each compartment to verify their identity (SuppFig. 2A). Consistent with the role of Elf5 as a master regulator of mammary cell specification [3,4], we found a striking polarisation in the main epithelial cluster associated to the sample genotype (Fig. 1D). This result clearly shows that

induction of ELF5 expression in PyMT tumours generate a profound transcriptional redefinition of the epithelial cell compartment, this effect was not observed in the other cellular compartments

We then used differential gene expression analysis and k-means clustering to identify cell diversity within the major cell compartments, classifying the tumour cells into 16 different cell classes (Fig. 1E). A cluster tree representation[22] of the relationship among the different clusters revealed two main families in the immune compartment, three in the stromal compartment and two in the cancer compartment; all of those defined very early in the phylogeny (Fig. 1F). The top differential marker genes that define stromal and immune cell compartments is shown in SuppFig. 2B. Such cell diversity within the major compartments was also validated using xCell algorithm [21] (Fig. 1G). The xCell signatures clearly identified endothelial cells and various classes of fibroblasts in the stromal compartment; a clear segregation between T-cells and all the other immune species such as B-cells and the diverse array of myeloid cells. In the cancer epithelial cell compartment a separate cluster of cancer cells (Fig. 1E cluster 10) consistent with an epithelial basal population (characterised by Krt5 expression[1], SuppFig. 2A) and a cluster rich of cell cycle related genes (Fig. 1E cluster 7) were the only two epithelial clusters not related to the previously shown cell polarisation driven by induced Elf5 expression.

PyMT cancer cells are organised in a structure that resembles the mammary gland epithelial hierarchy.

In our Drop-seq analysis, the major polarisation by ELF5 genotype was confined to the epithelial cell compartment (Fig 1D). Thus, we performed an independent k-means clustering on the epithelial subset cohort of cells (6475 PyMT/WT and 3233 PyMT/Elf5 cells) at an increased resolution. Epithelial tumour cells were distributed in 11 clusters visualised using tSNE representation (Fig. 2A). Figure 2B shows the strong genotype-driven cell polarisation in the epithelial compartment, which was not due to batch effects (SuppFig. 3). The tSNE plot from Figure 2B shows that clusters 1 and 6 were formed mostly in response to forced ELF5 expression (>2 Fold of cells belong to PyMT/ELF5 mice, while clusters 0, 7, 9, 4 and 3 were formed mostly by WT cells (>2 Fold of cells belong to PyMT/WT mice) (Fig. 2C). Clusters 2, 5, 8 and 10 were present in both WT and forced ELF5 expressing tumours (<2 Fold of cells from any of the genotypes). This distribution is consistent

with the expression of *Elf5* in each of these clusters (SuppFig. 4A). Concomitantly with the scRNAseq data, FACS analysis of the cancer epithelial cells revealed that Elf5 forced differentiation of the luminal progenitors, defined by a decrease in the proportion of Sca1^{low}/CD49b^{+} progenitor cells [2] within the luminal (EpCam+/CD49f+) population (Fig 2D).

We plotted each of our epithelial cells against the gene expression signatures for the epithelial cell lineages of the normal mammary gland [23] to reveal the phenotypic similarity of our MMTV-PyMT epithelial cell clusters to those lineages currently recognised. Ternary plots shown in Fig. 2E revealed that most cells and their clusters fell closest to the luminal progenitor (LP) signature, but clusters 7 and 8 were located towards the basal (B) region while cluster 10 and a portion of the cells that belonged to cluster 9 were located towards the mature luminal (ML) signature. Importantly, Cluster 10 showed very low levels of the PyMT oncogene (SuppFig. 4A), suggesting that this cluster could be formed of untransformed normal cells that were trapped within the mass of the tumours, this hypothesis is consistent with the small number of cells that belonged to this cluster. To increase the definition provided by these signatures we also annotated PyMT clusters using the mammary epithelial lineage markers defined by Bach and colleagues [24] (Fig. 2F). Cluster 0 and 7 expressed genes consistent with Procr+ multipotent/stem cells. Cluster 9 cells expressed hormone sensing (Hs) lineage markers, cluster 8 expressed markers for Basal (B)/ Myoepithelial (Myo) cells, clusters 1, 2, 3, 4, and 6 presented markers consistent with luminal cells, with clusters 2, 3, and 4 presenting progenitor markers while cluster 1 and 6 presented markers consistent with alveolar differentiated cells. By far, Clusters 5 and 10 (enriched for normal cells) were the most undefined clusters, showing mixed luminal and myoepithelial markers. Figure 2G shows the top differential markers associated to these malignant states of the main lineages of the mammary gland. Interestingly cluster 5 markers were strongly consistent with cycling cells, indicating that dividing cells are clustered together irrespective of their origin and could explain the reason why this cluster shows mixed gene markers.

Dynamic relationship and states of the malignant lineages of PyMT tumours: implications for the cell of origin of cancer.

To further investigate the differentiation states present within the luminal lineage in the PyMT tumours, we performed pseudotime analysis using Monocle (Fig.

3). We used the mammary gland hierarchy gene signature [23] to build the pseudotime structure of the PyMT cancer epithelial cells identifying 7 discrete states (Fig. 3A). Consistent with previous results, Elf5 had a profound impact upon the structure of the epithelial cell diversity biasing the composition of the cells towards one of the end fate axis (right hand side) of the pesudotime plot (dominated by state 7) at the expense of a depletion of states 2, 4, 5 and 6. We then visualised the Monocle defined cell states using the coordinates from the tSNE (Fig. 3B) which allowed us to overlap the cell cluster identity defined in Figure 2 with the pseudotiming states from Fig. 3B (Fig. 3C and SuppFig. 4B). A deconvoluted visualisation is shown in SuppFig. 4C and 4D.

The alveolar lineage was clearly defined by 2 distinct cell identities (cell clusters 1 and 6); the sensor lineage was composed of cluster 9, and the myoepithelial/basal lineage was formed by clusters 7 and 8. An undifferentiated cluster with multipotent/stem characteristics (cluster 0, apart from the cluster 5 of the cycling cells) sat in a position equidistant from the myoepithelial and the luminal lineages (alveolar and sensor). Interestingly, and consistently with the literature where stem cell properties are confounded by myoepithelial properties [25], cluster 7 was classified by Monocle as part of the basal/myoepithelial lineage, however, this cluster presents strong multipotent characteristics similar to those present in cluster 0. The most difficult compartment to define was the luminal progenitor (represented by cluster 2, 3 and 4) presumably due to their strong plasticity and subtle differential features.

To gain insights in this definition, we used GSVA [26] and metasignature visualisation to score each of the clusters according to the mammary gland epithelial signatures defined by Pal and colleagues. (Fig. 3D). Hierarchical clustering based on GSVA scores associated Cluster 2 with the alveolar lineage, suggesting its identity as a pre-alveolar committed luminal progenitor, or a cell type that sits in the hinge between the luminal progenitor and the alveolar lineages. Similarly, Cluster 4 was associated to the hormone-sensing lineage represented by Cluster 9, indicating a luminal progenitor committed towards this lineage. In this scenario, Cluster 3 remained undefined so we classified it as uncommitted luminal progenitor. tSNE visualisations of these metasignatures are shown in Fig 3D which further confirmed the cell distribution of the three main mammary epithelial lineages. The distribution of the different designated lineages in the two genotypes is shown in Fig. 3E, showing

less heterogeneity in PyMT/ELF5 tumours induced by a strong enrichment of the alveolar lineage. Finally, Figure 3F shows the relationship between clusters at different resolutions. These cluster tree groups are consistent with the cell states defined in the trajectory analysis. The phylogeny revealed an early separation between the two major lineages, the myoepithelial/undifferentiated (Cluster 0 from resolution 0.1) and luminal lineages (Cluster 1 from resolution 0.1).

Altogether, the combination analysis of gene signatures, gene markers and pseudotiming enabled the precise annotation of PyMT cell clusters within the mammary hierarchy proposed by Pal et al. and Bach et al, identifying a large luminal lineage that retains most of the cell diversity and strong plasticity, a basal/myoepithelial compartment and a hormone-sensing lineage (Fig. 3G).

Molecular mechanisms of cancer progression associated to cancer cells of Alveolar origin.

We next sought to validate some of the functional findings found in cancer cells of alveolar origin (Clusters 1 and 6) characteristic of PyMT/ELF5 tumours, including the decrease of cell proliferation, lack of EMT and increase of hypoxia [14,15,27] and SuppTable 1. Figure 4A shows the distribution of cell cycle marker signatures [28] in the epithelial cell compartment. These signatures identified a specific group exclusively populated by cells undergoing G2M/S. PyMT/ELF5 tumours showed an enrichment in the G1 phase, this is consistent with our previous reports that ELF5 increased cell cycle time by delaying entry into S phase [14], and that ELF5 reduced cell proliferation in PyMT tumours [15].

We then performed GSVA for specific signatures of known functions of ELF5, and functions suggested by the Hallmark analysis from SupplFig. 5, which included lactation, involution, EMT and hypoxia. This analysis revealed strong inductions of lactation, involution, MET and hypoxia by ELF5 (Fig. 4B). A tSNE visualisation of the EMT gene signature showed a strong association with myoepithelial and undifferentiated cells and a negatively enrichment in cells of other lineages, including alveolar cells (Fig. 4C) which it is a known characteristic of ELF5 [14,15,27]. Western Blot on total cell lysates of PyMT/WT and /ELF5 tumours for canonical markers of EMT that have been previously identified to be regulated by ELF5 via repression of SNAI2 [27] confirmed that ELF5 tumours were depleted of such EMT markers at the protein level. Although E-Cadherin levels were not significantly

different, ELF5 tumours with alveolar features show a clear reduction of Vimentin (Vim) indicating MET (Fig. 4C).

We have previously shown that PyMT/ELF5 tumours presented extensive vascular leakiness, a fact that can lead to increased hypoxia [15]. Cells enriched for a hypoxic gene signature concentrated in myoepithelial and alveolar cells (Fig. 4D). We confirmed these findings using a hypoxyprobe to identify hypoxic regions in the tumours. Immunohistochemistry showed extensive hypoxic regions in PyMT/ELF5 tumours compared to their WT counterparts, indicating that cancer alveolar cells are strong drivers of events leading to hypoxia or these cells are able to survive better in hypoxic regions. This phenotype was strong enough to induce hypoxia in the metastases of a highly oxygenated tissue such as the lung (Fig. 4D).

As expected, the lactation signature was strongly enriched in cells of alveolar origin (Fig. 4E, top tSNE plot). Immunohistochemistry using anti-milk antibodies revealed a strong production of milk in PyMT/ELF5 tumours compared to PyMT/WT tumours (Fig. 4E, IHC pictures). In the normal mammary gland, milk stasis generates the signals that trigger mammary involution, and in this scenario, where milk cannot leave the tumour, we identified a late involution gene signature [29] associated with the luminal lineage in PyMT tumours, specifically enriched in ELF5 alveolar cells (Fig. 4E, bottom tSNE plot).

Taken together, single-cell transcriptomics enables the identification of biological properties of mammary tumours and, in particular, the role of involution mimicry as a driver of sustained inflammation associated with the Elf5-alveolar subtype in luminal breast cancer.

Characterisation of cancer-associated fibroblasts in PyMT tumours

The involution process involves cell apoptosis and tissue remodelling, the latter of which is orchestrated by stromal and immune cells [31]. Among these cell types, fibroblasts have a critical role in the extracellular matrix remodelling and immune suppression during involution [32]. Thus, we further explored the role of cancer-associated fibroblasts (CAFs) in aberrant involution in our model.

Unsupervised clustering on the fibroblast subset (2,255 pooled cells from PyMT/WT and PyMT/ELF5 tumours, revealed three major fibroblast clusters (res 0.1) (Fig. 5A, SuppFig. 6A-B). Overlaying of the cell cycle signature in the tSNE

showed that a cell subgroup of Cluster 2 corresponded to dividing cells (SuppFig. 6C, upper panel).

CAFs are known to be highly diverse, which it was confirmed in our system as classical fibroblast markers[33] failed to be specific to any particular transcriptomically-defined clusters (SuppFig. 6D). Previous unbiased scRNAseq analyses in human tumours has classified two major fibroblast subtypes [34,35]: activated myofibroblasts, involved in tissue remodelling architecture by physical forces; and secretory CAFs related to ECM synthesis and cyto- and chemokine production [36]. Fibroblast clusters 0 (FC0) and 1 (FC1) were enriched for a secretory CAF signature while the activated myofibroblast signature was concentrated in fibroblast cluster 2 (FC2) (Fig. 5B). Functional annotation using GSVA analysis (SuppFig. 7A) revealed that secretory FC0 were governed by hallmarks of EMT and were enriched for a desmoplastic signature[37], characterised by genes involved in extra cellular matrix (ECM) interactions, suggesting a tissue remodelling function (Fig. 5C and D). Secretory FC1 clearly showed immune-related functions (top enriched hallmarks “Interferon alpha response”, “IL6/JAK/STAT3 signalling” and “Interferon gamma response”), with genes involved in cytokine-cytokine receptor interactions, monocyte recruitment and the complement cascade[37]. FC2 were enriched for a contractile signature, characterised by the expression of genes involved in actin cytoskeleton organisation.

Taken together, we have functionally characterised three different classes of fibroblasts in PyMT tumours: 2 secretory-CAFs classes, FC0 defined as ECM-CAFs and FC1 defined as inflammatory-CAFs (iCAFs); and a contractile myofibroblast class (FC2).

CAFs from PyMT/ELF5 mammary tumours show features of mimicry involution.

Unsupervised clustering of the fibroblasts found in PyMT tumours further identified several subclasses of secretory and contractile fibroblasts (Fig. 5E). Mammary involuting fibroblasts have unique properties that include a high production of fibrillar collagen, ECM remodelling and immune suppression through monocyte recruitment[32]. This specialised class of fibroblast involved in involution are characterised by the expression of Col1a1, Cxcl12, Tgfb1 and Mmp3. We identified a group of ECM-CAFs and iCAFs with high expression of these involution markers in PyMT tumours (Fig. 5E middle and bottom panel, clusters F2 and F3). We thus

referred these two clusters as involution-iCAFs and involution-ECM-CAFs, respectively. The complete classification of PyMT fibroblasts is shown in Fig. 5F. Involution-CAFs were enriched in PyMT/ELF5 tumours where >2 fold of involution-CAFs, Cluster F2 and F3, belonged to PyMT/ELF5 mice (Fig. 5G). Supplementary Figure 7B shows that such genotype polarisation was not due to batch effects. Consistently, GSVA of the CAF-involution signature and involution marker genes were significantly enriched in the CAFs from ELF5 tumours compared to the WT (Fig. 5H).

The functional annotation of the involution fibroblasts revealed specific differences within the ECM-CAFs or iCAF subpopulations (SuppFig. 7C). For example, the top pathways of the involution ECM-CAFs that were unique of this ECM subgroup, fatty acid metabolism and peroxisome hallmarks, suggested that these cells may correspond to adipocyte-like fibroblasts arisen through a de-differentiation process of mature adipocytes [38] or transdifferentiation process of activated fibroblasts [39]. Adipocyte-derived fibroblasts secrete fibronectin and collagen I and have invasive properties [38,40]. Moreover, during the stromal remodelling stage in involution, there is a re-differentiation of the adipocytes [31,41]. Involution iCAFs, on the other hand, showed pathways linked to a wound-healing process, including inflammation through the complement cascade and coagulation [29,42,43]. The late stages of involution share many wound-healing properties such as high immune cell infiltration and immune suppression [44], thus these involution iCAFs might be responsible for these two immune-related processes.

Our finding on CAFs activation mediated by Elf5-alveolar epithelial cells has never been reported before, thus we performed orthogonal validation for the role of CAFs in ECM remodelling by the production of collagen fibres in both PyMT/WT and PyMT/Elf5 tumours. Second Harmonic Generation (SHG) imaging and picrosirius red staining of tumour sections showed an increase in fibrillar collagen coverage in PyMT/ELF5 tumours compared with their PyMT/WT counterparts (Fig. 6A and B). Polarised light microscopy showed that PyMT/ELF5 tumours had a significantly higher proportion of thicker and mature collagen fibres than PyMT/WT tumours (Fig. 6C). Further analysis of matrix ultrastructure revealed a more complex spatial arrangement (peak alignment (+/-10 degrees from peak)) of collagen fibres in PyMT/ELF5 tumours than in PyMT/WT tumours (Fig 6D). Taken together these

results show higher fibroblast activity both in collagen deposition and in their rearrangement in PyMT/ELF5 tumours.

Characterisation of the cell-to-cell interactions involved in the cancer-associated involution mimicry.

Involution is a multicellular process, thus we used CellphoneDB to identify networks of cell-to-cell communication within PyMT tumours[45]. CellphoneDB is a manually curated repository of ligand-receptor interacting pairs combined with a statistical framework to predict cell-to-cell connections. CellphoneDB analysis in PyMT tumours revealed an intricate network of cell interactions. Figure 7A shows the interactome of all cell clusters defined in this study (Figs. 1, 2 and 5, also SuppFig. 7 D). These interactions were ranked from strongest to weakest (Fig 7A, count above mean 0.3) which allowed us to categorise them in three 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://odonoghuelab.org:8000), cell types are grouped by unsupervised clustering of common signalling pathways.

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 [31]. We interestingly found that when the most restrictive cut offs were applied (Fig. 7B, 1st and 2nd Tier, blue lines), the ECM-CAFs (involution and fibroblast cluster 2) and involution iCAFs showed the largest number of interactions between themselves and also with the two Elf5 enriched-alveolar epithelial cells clusters (epithelial Clusters 1 and 6), the myoepithelial cells and the luminal progenitor hormone sensing cells (LP Hs). In the immune system interactive network (Fig. 7B, 3rd and 4rd Tiers, green lines), the myeloid compartment showed the highest number of connections and interestingly, myeloid cells communicate with the two alveolar epithelial clusters, with ECM CAFs (involution and fib cluster 2), involution iCAFs, myofibroblasts and with the

endothelium. Involution is associated with the M2-like innate response, thus these results are in line with an activated and interactive innate myeloid response [6,29,46].

We then investigated what particular ligand/receptor interactions were found among CAFs, alveolar epithelial and myeloid cells in order to find common pathways linked to involution (Fig. 7C). Overall, we found pathways associated with the multi-functional aspects of involution (TGFβ axis) [47,48]; networks associated with an immune suppressive ecosystem (Cxcl12 and Dpp4 axis) [32,49] and ECM remodelling (IGF axis) [50] (Fig. 7C). Transforming growth factor β2 and 3 (TGFβ2 and TGFβ3) are ligands produced by CAFs and the two alveolar epithelial clusters, while their receptor TGFβR3 is exclusively expressed in involution ECM and iCAFs and in iCAFs-1 (Fig. 7C and D).

The chemokine Cxcl12 is exclusively expressed in the involution iCAFs and two of its known receptors, CXC chemokine receptor 4 (CXCR4) and atypical chemokine receptor 3 (ACKR3 or CXCR7), are expressed in iCAFs-1 and B and T cells, respectively. Dipeptidyl peptidase-4 (Dpp4) expression was restricted to iCAFs-1 and showed a highly interactive axis; Dpp4 connected with the involution iCAFs through Cxcl12 and with myeloid cells through Cxcl2 and Cxcl10. The IGF1/IGF1R signalling pathway was found to be highly specific to the two involution CAF subtypes, ECM and immune, in which a predicted positive feedback loop among these cells was observed. Of note is that IGF signalling is classically associated with collagen and ECM remodelling in fibroblasts [50].

The myeloid cell compartment is also intimately associated with the process of involution. We have previously demonstrated that tumour infiltrated myeloid-derived suppressor cells (MDSCs) are involved in the formation of lung metastasis in PyMT/ELF5 alveolar tumours[15]. As tumour infiltrated myeloid cells constitute a low abundance population[51] we kept it as a unique cell cluster, however, our scRNAseq analysis was able to map the cellular mechanism for MDSC expansion, recruitment and malignant activation previously described in these alveolar tumours[52], identifying the cell of origin of well-known signalling pathways involved in this process (Fig. 7D). G-CSF (Csf3), a potent growth factor that attracts and recruits MDSCs, was expressed in cancer cells associated with the alveolar clusters, which would explain the increased infiltration of MDSCs in PyMT/ELF5 tumours[15]. Malignant activation of MDSCs is also driven by M-CSF (Cfs1), mainly expressed by CAFs. In addition,

VEGF (Vegfa), was highly secreted by alveolar, basal cancer cells and CAFs, and IL6 is mainly secreted by CAFs and endothelial cells, these are factors that contribute to MDSC expansion. Potent MDSCs chemo-attractants were expressed by a diverse range of cell types within the tumour; Cxcl1 is ubiquitously expressed in most cell types, while Cxcl2 was highly expressed in the myeloid cell cluster, suggesting this is a mechanism that tumours may use to maintain a feedback loop enabling continuous myeloid cell recruitment and sustained inflammation. Other recruitment factors such as Cxcl12 and Ccl2 are specifically secreted by CAFs. Altogether, our results show that MDSCs recruitment and malignant activation is not just driven by cancer cells but by a complex molecular network orchestrated by alveolar cancer cells and involving multiple cell compartments in the TME, mainly CAFs, endothelial cells and tumour infiltrated myeloid species.

In summary, we have found that cells in the TME of PyMT tumours are heavily involved in the promotion of tumour progression. We have characterised the intercellular molecular pathways originated in alveolar cancer cells that resemble a deregulated established involution. This involution mimicry results in an inflammatory process characterised by malignant CAF education and recruitment of MDSCs, which ultimately results in the acquisition of functional traits of tumour progression.

Involution signatures in the context of Elf5 predicts for poor prognosis in breast cancer patients.

High *ELF5* expression is associated with basal breast cancers[14]. We have previously shown that in luminal tumours with higher expression of ELF5 show basal-like characteristics and poor prognosis[15], thus these can be considered as the early steps towards the differentiation into a basal tumour, underscoring ELF5 as a driver of the basal-like phenotype.

We used the METABRIC dataset [30] to investigate whether the involution signature correlated with patient outcome. In line with previous reports[14], we found a significant association between high *ELF5* expression and poorer patient prognosis. As expected, higher proportion of basal molecular subtype fell into the ELF5-high group, but also a proportion of luminal cancers (Fig. 8A). Importantly, *ELF5* expression or the involution signature alone did not have

predictive value both in the basal or the luminal context (Fig. 8B, 8C top panels); however, the presence of the involution gene signature in *ELF5*-high patients was associated with poor prognosis (Fig. 8B, 8C, bottom panels). It is important to note that some genes of the involution signature are also drivers of general inflammation, our results highlight that the acquisition of hallmarks of tumour progression associated to involution are only meaningful when ELF5 is expressed and thus in the context of a process of involution mimicry.

4. Discussion

The MMTV-PyMT model gives rise to mammary tumours of luminal progenitor origin as classified by conventional transcriptomics [9,53]. Our findings however underscore the expansive plasticity of the luminal progenitor compartment identifying "malignant states" associated with the mammary epithelium cell lineage specification. Our results unite tumour heterogeneity with developmental cues and highlights the intrinsic capacity of tumours to develop distinct molecular subtypes, supporting the current understanding of progenitor cells as the cell of origin of basal and luminal cancers (reviewed in [8])

Here we study the functional consequences of the specification of the ER-basal subtype of breast cancer through the differentiation of the luminal progenitor cells towards the pregnancy-derived alveolar lineage [3,63]. Our data uncovers that alveolar cancer cells activate cell-extrinsic features reminiscent of mammary morphogenesis, orchestrating cell-to-cell interactions within the TME that result in the promotion of metastatic cues (Fig. 7). Our results support the function of the transcription factor ELF5 in the specification of basal breast cancer and subsequent acquisition of tumour progression cues based on extracellular signals and the alteration of the tumour microenvironment.

We have previously shown that Elf5-driven tumours presented enhanced cancer immunotolerance, vascular leakiness and angiogenesis, which in turn produced a dramatic increase of metastasis, similar to that induced by pregnancy [15]. This has profound therapeutic implications in the clinical setting, as patients with luminal tumours receive standard-of-care anti-estrogen therapy. In this context, adoption of the alveolar fate provides a route of escape from endocrine therapy, as these cells are insensitive to estrogen [14] and subsequently give rise to a basal-like tumour [14] with

much poorer survival rate [15]. 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 that show higher Elf5 expression activate involution mimicry pathways that are associated with poor prognosis. In addition, the pro-metastatic effects of Elf5 are concomitant with mimicry of mammary involution. In this line of evidence, involution has been previously shown to be involved in post-partum associated breast cancer (PABC) [13] however the molecular drivers of this process had not been not fully identified.

Our Elf5-restricted MMTV-PyMT model reproduces the specific 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 which includes tissue remodelling with fibroblast activation to reorganise the ECM and accommodate the epithelial reduction, and myeloid infiltration to remove all apoptotic bodies [5,6]. Our data presented here supports the hypothesis that the induced expression of *Elf5* stimulates the production of milk by alveolar cells, which in turn accumulates and triggers an inflammation-mediated TME remodelling associated to aberrant involution. This mimicry involution process results in alveolar apoptosis that is rapidly restored by ELF5-driven differentiation of luminal progenitors and a chronic induction of inflammation by involution-iCAFs and myeloid cells encompassing ECM remodelling driven by involuting ECM-CAFs (Fig 7E).

Our scRNAseq analyses have also allowed us to identify putative molecular candidates responsible for this epithelium-fibroblast-myeloid crosstalk during mimicry of involution (Fig. 7). We find that only alveolar epithelial cells express TGFβ2 and 3 ligands and their receptor TFGβR3 is expressed in involution CAFs. TGFβ is linked to tumour malignancy by affecting both cancer cells and the TME, including CAF activation and immune suppression [47] suggesting that this could be a specific pathway driving mimicry involution. In fact, TGFβ3 is induced in response to milk stasis during the first stage of involution for the induction of alveolar apoptosis and TGFβ signalling is also key in the second stages of normal involution for the induction of ECM deposition and immune suppression [48]. Altogether, this suggests that TGFβ signalling might be one of the pivotal pathways triggered by Elf5-alveolar cells and thus a potent pathway to be therapeutically targeted.

We also find specific pathways of crosstalk among involution CAFs, other CAF-types and the tumour-infiltrating immune system, which might be activated following this ELF5-mediated TGFβ activation. These include the expression of Cxcl12 by the involution-iCAFs and IGF1 from both ECM and inflammatory involution CAFs. Cxcl12 is a stromal chemokine that attracts leukocytes through CXCR4[67] and enhances cell adhesion and survival through ACKR3[68], these receptors are expressed in T/B cells and iCAFs, respectively. Interestingly, involuting mammary fibroblasts that highly express Cxcl12 induce monocyte recruitment and are associated with blockade of CD8+ T cell tumor infiltration [69]. Cxcl12 is also a known recruitment factor for MDSCs [52]. In this context, CAFs also show high expression of other classical markers for MDSC recruitment and activation, including M-CSF, IL-6 and Ccl2 [52] underpinning the high infiltration and activation of MDSCs previously described in Elf5-high tumours [15]. 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 the inhibition the recruitment effector T cells [70] and by the promotion of chronic inflammation [71]. This suggests the M1-innate response from myeloid cells and fibroblasts is inhibited by Dpp4-expressing iCAFs. In addition, Costa and colleagues recently demonstrated that a subtype of CAFs from human breast tumours was responsible for the increase of regulatory T cells through the expression of Dpp4 [49], so the involvement of Tregs in Elf5-mediated immunosuppression might be another area to further explore. Thus, our data suggests that most of the immunosuppressive signals come from the presence of activated MDSCs in the tumour, mainly recruited by involution CAFs signaling as a response to the involution mimicry process initiated by alveolar cancer cells (Fig. 7E).

Altogether our data illustrates a highly interactive TME, particularly with CAFs and myeloid cells, with cancer-differentiated cells that are pivotal initiating the events that result in tumour immunetolerance and metastasis as a consequence of mimicry involution. Our data highlights the relevance of targeting cancer-associated cell species as a strategy for anticancer therapy, an approach particularly important in the context of PABC.

5. Acknowledgments

We thank Dr Marina Pajic and Dr Christine Chaffer from the Garvan Institute for their review and feedback on the manuscript. This work is supported by Cancer Institute NSW Career Development Fellowship (DG00625) and Cancer Council NSW project grant (RG18-03) to DGO; National Breast Cancer Foundation postdoctoral fellowship (2013-2018, PF-13-11) and Cancer Institute NSW Career Development Fellowship (2019-2021, CDF181218) to FVM; and the National Health and Medical Research Council project grant (NHMRC 1068753) to CJO; Australian Government Research Training Program (RTP) Scholarship to AMKL and UNSW Sydney University Tuition Fee Scholarship (TFS) to YCS.

6. Legends to Figures

Figure 1. High-resolution cell composition of MMTV-PyMT tumours.
A) Heatmap showing differential expression of the top expressed genes contributing to the epithelial (red), stromal (blue) and immune (green) clusters and their percentage. The bottom panel shows the enrichment value for each of the signatures of the xCell algorithm for each cluster is shown. **B)** Representative contour plot of the cell composition of a MMTV-PyMT tumor analysed by FACS defined by EpCAM antibodies (epithelial cells), CD45 (leukocytes) and double negative cells (stroma). **C)** tSNE visualisation shows the coordinates of each analysed cell after dimensional reduction coloured by its main cell lineage. The dot plot shows the top differential markers form each of the main cell lineages and their level of expression. **D)** tSNE plot showing the distribution of cells per genotype. **E**) tSNE plot showing cell clusters defined by a k-means based clustering algorithm, and their relative frequency (bottom). **F)** Cluster tree modelling the phylogenic relationship of the different clusters in each of the main cell types compartments at different clustering resolutions. Dashed red line shows the resolution chosen. Coloured circles in the cluster tree represent the origin of the clusters represented in the tSNE plot shown in panel E. **G)** Cell identification using score values for each of the metasignatures of the xCell algorithm in the immune and stromal compartment divided by cluster.

Figure 2. Annotation of the cancer epithelial cell diversity in PyMT tumours.
A) tSNE visualisation of the cell groups defined by k-means clustering analysis. **B)** Distribution of cells by genotype in the defined tSNE dimensions. **C)** Sankey diagram showing the contribution of each of the genotypes to each cell clusters. Cluster numbers are coloured by the dominant genotype (>2-fold cell content of one genotype), Elf5 (red), WT (green). **D)** Scatter plot showing FACS data to define the % alveolar versus luminal progenitors using canonical antibodies that define the epithelial mammary gland hierarchy (EpCAM, CD49f, Sca1 and CD49b), in PyMT tumours (WT n = 6 and Elf5 n = 5). Bottom panels are representative FACS plots of one of the replicates for each genotype. **E)** Ternary plots using the gene signatures of the mammary epithelial hierarchy as in Pal et al. 2017. The colour palette corresponds with the defined cluster identity in Figure 2A. **F)** 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 cluster. These marker genes were grouped according to each mammary epithelial cell type as defined by Bach, et al.: Hormone sensing differentiated (Hs-d, dark pink), Hormone-sensing progenitor (Hs-p, light pink), Luminal progenitor (LP, orange), Alveolar differentiated (Alv-d, dark red), Alveolar progenitor (Alv-p, light red), Basal (B, light purple), Myoepithelial (Myo, dark purple), Undifferentiated (Multi, light blue). **G)** Dot plot of the expression level 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 panel C.

Figure 3. Lineage trajectory of PyMT cancer cells.
A) Pseudotiming alignment of the PyMT cancer epithelial cells along the gene signatures that define the main lineages of the mammary gland epithelial hierarchy using the DDRTree method in Monocle2. Right panels show the distribution of the cell states by genotype. **B)** Projection of the states defined by pseudotime analysis into tSNE clustering coordinates overall and per pseudotime state (miniaturised tSNE plots). Right panels show the projection by genotype. **C)** Overlay representation of the cell identities (k-means clustering as per Fig.3) and cell lineage identification (pseudotime analysis), the proportion of cells in each cluster that belong to each defined state is shown in the bar chart (right hand side). **D)** Enrichment analysis (GSVA score) for the gene signatures that define the main mammary gland lineages: Basal, Luminal Progenitor (LP) and Mature Luminal (ML) for each of the clusters. Bottom panels show the expression of each of the gene signatures at single cell resolution. The top bar shows the assigned mammary epithelial cell type as per section C. **E)** Frequency of the different cell lineages in each genotype. **F)** Cluster tree showing the phylogeny relationship of the different clusters. Red arrow shows the resolution used (0.7). **G)** Illustration of cell diversity of PyMT tumours based on the canonical structure of the mammary gland epithelial lineages.

Figure 4. Functional validation of cancer related features associated to PyMT/Elf5 tumours.
A) Cell cycle stages of the PyMT cancer cells as defined by gene expression signatures using tSNE coordinates and their deconvolution. Circled area shows the cycling cluster (C5 in Fig. 3) characterised by a total absence of G1 cells. The quantification of the proportion of cells in each stage grouped by genotype is shown in the bar chart. **B)** Enrichment GSVA analysis of gene expression metasignatures of cancer-related and Elf5-related hallmarks associated to PyMT/WT (green) and /Elf5 (red) tumours. **C)** tSNE representation of the EMT gene expression metasignature. Right panel shows a western blot of canonical EMT markers (E-Cadh, E-Cadherin and Vim, vimentin) on PyMT/WT or ELF5 full tumour lysates. Note: the two images correspond to the same western blot gel cropped to show the relevant samples. **D)** Hypoxia metasignature is shown in the tSNE plot, bottom panel shows a bar plot of the extension of the hypoxic areas in PyMT/WT (green) and /Elf5 (red) tissue sections (tumours and lung metastasis) stained using IHC based on hypoxyprobe binding, representative images are shown in the right panels. **E)** Lactation and Late Involution (stage 4, S4) metasignatures is shown in the tSNE plots. Pictures show IHC with an anti-milk antibody in tissue sections 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 tumours.

Figure 5. Annotation of the cancer-associated fibroblast diversity in PyMT tumours.

A) tSNE plot groups defined by k-means clustering analysis showing a total of three cell clusters defined within the fibroblast subtype. **B)** Metasignatures of Cancer-associated fibroblast (CAF) signature (upper plot) and myfibroblasts (bottom panel) plotted in the fibroblast tSNE. **C)** Desmoplastic (upper plot), Inflammatory (middle plot) and Contractile (bottom plot) metasignatures plotted in the fibroblast tSNEs from public data [37]. **D)** Violin plots displaying marker genes for each of the three fibroblast clusters defined in section 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. Middle plot: tSNE illustration of the involution metasignature from [32]. Bottom section: Violin plots on the nine cell clusters for individual genes from the involution signature. **F)** Cell tree classification of the main lineages of CAFs in MMTV-PyMT/WT and ELF5 tumours. **G)** Upper plot: Distribution of fibroblasts by genotype (Elf5: red), WT: green) in the defined tSNE dimensions. Bottom plot: Sankey diagram showing the contribution percentage of each of the genotypes to the cell clusters. Cluster numbers are coloured by the dominant genotype (>2-fold cell content of one genotype), Elf5 (red), WT (green). **H)** GSVA enrichment analysis of involuting mammary fibroblast metasignatures associated to PyMT/WT and /Elf5 tumours. Violin plots of Cxcl12, Mmp3 and Col1a1 genes in all fibroblasts of each genotype. PyMT/WT (green) PyMT/Elf5 (red).

Figure 6. Functional validation of ECM remodelling associated with CAFs in PyMT/WT and /Elf5 tumours.

A) Representative bright field images and quantification of total coverage of picrosirius red-stained PyMT/WT and PyMT/ELF5 tumours sections n=4 mice per genotype with 10 regions of interest (ROI) per tumour. **B)** Representative maximum intensity projections of SHG signal and quantification of SHG signal intensity at depth (µm) and at peak in PyMT/WT and PyMT/ELF5 tumour sections, n=6 mice per genotype with 6 ROI per tumour. **C)** Polarised light imaging of picrosirius red stained PyMT/WT and PyMT/ELF5 tumour sections, and quantification of total signal intensity acquired via polarised light. Thick remodelled fibres/high birefringence (red-orange), medium birefringence (yellow) and less remodelled fibres/low birefringence (green) n=4 mice per genotype with 10 ROI. **D)** 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.

Figure 7. Interactome of PyMT tumours.

A) Heatmap of the cell-cell interactions of all cell types from PYMT tumours based on Cellphone DB. Cell classification is based on the global annotation shown in

Supp.Fig. 7C. For simplicity, F6 and F8 were removed and F1 and F7 (myofibroblasts) were pooled together as a single cluster. The scale at the right-hand side shows the interaction strength based on the statistical framework included in CellphoneDB (count of statistically significant ($p<0.01$) interactions above mean= 0.3, see methods). **B)** Graphical representation of all significant cell-cell interactions identified by CellphoneDB 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. The red circles correspond to the cell types from the epithelial compartment; the blue triangles represent the cells from the stromal compartment and the green squares are the cells from the immune compartment. The size of geometric figures is relative to the number of cells involved in the interactions (display as count). Different number splits were applied to establish the most significant interactions for the fibroblast and immune cell types. Fibroblast showed the strongest interactions (highlighted as blue lines) when a number split of 67 (1st Tier) and 50 (2nd Tier) were used. The immune system showed weaker interactions (highlighted as green lines) at a number split of 15 (3rd Tier) and 11 (4th Tier). **C)** Representative dot plots of ligand (no background)-receptor (red background) pairs. The size of the circles 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. **E)** Proposed molecular model of involution mimicry driven by Elf5 where CAFs and MDSCs are the major cell types involved.

Figure 8. Involution is a poor prognosis factor in the context of Elf5 expression in luminal and basal breast cancer.

A) Kaplan-Meier survival curves based on Elf5 expression using the METABRIC cohort. Patient were segregated according to Elf5 expression levels based on tertiles. Elf5-high patients (red) were defined as the top-tertile and Elf5-low patients (green) as the bottom-tertile, Log-rank p values <0.05 are shown in red. The bar chart (bottom panel) corresponds to the distribution of the PAM50 classified breast cancer subtypes in the top and bottom Elf5 expressing tertile of patients. **B)** Upper panels: Survival analysis (Kaplan-Meier curves) for the expression of Elf5 (left hand side) and the Involution signature (right hand side) in basal breast cancer patients. Bottom panel: Kaplan-Meier survival curves for the late involution signature segregated by *ELF5* expression according to tertiles. Green, inv. low, bottom third; Blue, inv. mid, middle third and Red, inv. high, top third. Log-rank p values <0.05 are shown in red. **C)** As per B) but in luminal patients.

7. References

1. Oakes, S.R., Gallego-Ortega, D. & Ormandy, C.J. The mammary cellular hierarchy and breast cancer. *Cell Mol Life Sci* **71**, 4301-24 (2014).
2. Shehata, M. *et al.* Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res* **14**, R134 (2012).
3. Oakes, S.R. *et al.* The Ets transcription factor Elf5 specifies mammary alveolar cell fate. *Genes Dev* **22**, 581-6 (2008).
4. Gallego-Ortega, D., Samantha R Oakes, Heather J Lee, Catherine L Piggin, Christopher J Ormandy. ELF5, normal mammary development and the heterogeneous phenotypes of breast cancer. *Breast Cancer Management* **2**, 489-498 (2013).
5. Inman, J.L., Robertson, C., Mott, J.D. & Bissell, M.J. Mammary gland development: cell fate specification, stem cells and the microenvironment. *Development* **142**, 1028-42 (2015).
6. Stein, T., Salomonis, N. & Gusterson, B.A. Mammary gland involution as a multi-step process. *J Mammary Gland Biol Neoplasia* **12**, 25-35 (2007).
7. Schedin, P., O'Brien, J., Rudolph, M., Stein, T. & Borges, V. Microenvironment of the involuting mammary gland mediates mammary cancer progression. *J Mammary Gland Biol Neoplasia* **12**, 71-82 (2007).
8. Visvader, J.E. Cells of origin in cancer. *Nature* **469**, 314-22 (2011).
9. Lim, E. *et al.* Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res* **12**, R21 (2010).
10. Molyneux, G. & Smalley, M.J. The cell of origin of BRCA1 mutation-associated breast cancer: a cautionary tale of gene expression profiling. *J Mammary Gland Biol Neoplasia* **16**, 51-5 (2011).
11. Fantozzi, A. & Christofori, G. Mouse models of breast cancer metastasis. *Breast Cancer Res* **8**, 212 (2006).
12. Schedin, P. Pregnancy-associated breast cancer and metastasis. *Nat Rev Cancer* **6**, 281-91 (2006).
13. Lyons, T.R. *et al.* Postpartum mammary gland involution drives progression of ductal carcinoma in situ through collagen and COX-2. *Nat Med* **17**, 1109-15 (2011).
14. Kalyuga, M. *et al.* ELF5 suppresses estrogen sensitivity and underpins the acquisition of antiestrogen resistance in luminal breast cancer. *PLoS Biol* **10**, e1001461 (2012).
15. Gallego-Ortega, D. *et al.* ELF5 Drives Lung Metastasis in Luminal Breast Cancer through Recruitment of Gr1+ CD11b+ Myeloid-Derived Suppressor Cells. *PLoS Biol* **13**, e1002330 (2015).
16. Macosko, E.Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).
17. Cazet, A.S. *et al.* Targeting stromal remodeling and cancer stem cell plasticity overcomes chemoresistance in triple negative breast cancer. *Nat Commun* **9**, 2897 (2018).
18. Vennin, C. *et al.* Transient tissue priming via ROCK inhibition uncouples pancreatic cancer progression, sensitivity to chemotherapy, and metastasis. *Sci Transl Med* **9**(2017).
19. Law, A.M.K. *et al.* Andy's Algorithms: new automated digital image analysis pipelines for FIJI. *Sci Rep* **7**, 15717 (2017).

20. Oakes, S.R. *et al.* A mutation in the viral sensor 2'-5'-oligoadenylate synthetase 2 causes failure of lactation. *PLoS Genet* **13**, e1007072 (2017).
21. Aran, D., Hu, Z. & Butte, A.J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* **18**, 220 (2017).
22. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7**(2018).
23. Pal, B. *et al.* Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. *Nat Commun* **8**, 1627 (2017).
24. Bach, K. *et al.* Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. *Nat Commun* **8**, 2128 (2017).
25. Prater, M.D. *et al.* Mammary stem cells have myoepithelial cell properties. *Nat Cell Biol* **16**, 942-50, 1-7 (2014).
26. Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
27. Chakrabarti, R. *et al.* Elf5 inhibits the epithelial-mesenchymal transition in mammary gland development and breast cancer metastasis by transcriptionally repressing Snail2. *Nat Cell Biol* **14**, 1212-22 (2012).
28. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-96 (2016).
29. Stein, T. *et al.* 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**, R75-91 (2004).
30. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346-52 (2012).
31. Watson, C.J. Involution: apoptosis and tissue remodelling that convert the mammary gland from milk factory to a quiescent organ. *Breast Cancer Res* **8**, 203 (2006).
32. Guo, Q. *et al.* Physiologically activated mammary fibroblasts promote postpartum mammary cancer. *JCI Insight* **2**, e89206 (2017).
33. Cortez, E., Roswall, P. & Pietras, K. Functional subsets of mesenchymal cell types in the tumor microenvironment. *Semin Cancer Biol* **25**, 3-9 (2014).
34. Lambrechts, D. *et al.* Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med* **24**, 1277-1289 (2018).
35. Puram, S.V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624 e24 (2017).
36. Kalluri, R. The biology and function of fibroblasts in cancer. *Nat Rev Cancer* **16**, 582-98 (2016).
37. Davidson, S. *et al.* Single-cell RNA sequencing reveals a dynamic stromal niche within the evolving tumour microenvironment. *bioRxiv* (2018).
38. Bochet, L. *et al.* Adipocyte-derived fibroblasts promote tumor progression and contribute to the desmoplastic reaction in breast cancer. *Cancer Res* **73**, 5657-68 (2013).
39. Lorenz, K. *et al.* Multilineage differentiation potential of human dermal skin-derived fibroblasts. *Exp Dermatol* **17**, 925-32 (2008).
40. Hillers, L.E., D'Amato, J.V., Chamberlin, T., Paderta, G. & Arendt, L.M. Obesity-Activated Adipose-Derived Stromal Cells Promote Breast Cancer Growth and Invasion. *Neoplasia* **20**, 1161-1174 (2018).
41. Zwick, R.K. *et al.* Adipocyte hypertrophy and lipid dynamics underlie mammary gland remodeling after lactation. *Nat Commun* **9**, 3592 (2018).

42. Amara, U. *et al.* Interaction between the coagulation and complement system. *Adv Exp Med Biol* **632**, 71-9 (2008).
43. Lilla, J.N., Joshi, R.V., Craik, C.S. & Werb, Z. 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**, 13792-803 (2009).
44. Martinson, H.A., Jindal, S., Durand-Rougely, C., Borges, V.F. & Schedin, P. Wound healing-like immune program facilitates postpartum mammary gland involution and tumor progression. *Int J Cancer* **136**, 1803-13 (2015).
45. Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347-353 (2018).
46. O'Brien, J. *et al.* Alternatively activated macrophages and collagen remodeling characterize the postpartum involuting mammary gland across species. *Am J Pathol* **176**, 1241-55 (2010).
47. Bellomo, C., Caja, L. & Moustakas, A. Transforming growth factor beta as regulator of cancer stemness and metastasis. *Br J Cancer* **115**, 761-9 (2016).
48. Flanders, K.C. & Wakefield, L.M. Transforming growth factor-(beta)s and mammary gland involution; functional roles and implications for cancer progression. *J Mammary Gland Biol Neoplasia* **14**, 131-44 (2009).
49. Costa, A. *et al.* Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. *Cancer Cell* **33**, 463-479 e10 (2018).
50. Blackstock, C.D. *et al.* 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**, 7264-74 (2014).
51. Valdes-Mora, F. *et al.* Single-Cell Transcriptomics in Cancer Immunobiology: The Future of Precision Oncology. *Front Immunol* **9**, 2582 (2018).
52. Sevko, A. & Umansky, V. Myeloid-derived suppressor cells interact with tumors in terms of myelopoiesis, tumorigenesis and immunosuppression: thick as thieves. *J Cancer* **4**, 3-11 (2013).
53. Lim, E. *et al.* Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* **15**, 907-13 (2009).
54. Fu, N.Y. *et al.* Identification of quiescent and spatially restricted mammary stem cells that are hormone responsive. *Nat Cell Biol* **19**, 164-176 (2017).
55. Van Keymeulen, A. *et al.* Lineage-Restricted Mammary Stem Cells Sustain the Development, Homeostasis, and Regeneration of the Estrogen Receptor Positive Lineage. *Cell Rep* **20**, 1525-1532 (2017).
56. Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479**, 189-93 (2011).
57. Wang, D. *et al.* Identification of multipotent mammary stem cells by protein C receptor expression. *Nature* **517**, 81-4 (2015).
58. Wuidart, A. *et al.* Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nat Cell Biol* **20**, 666-676 (2018).
59. Nguyen, Q.H. *et al.* Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nat Commun* **9**, 2028 (2018).
60. Bland, P. & Howard, B.A. Mammary lineage restriction in development. *Nat Cell Biol* **20**, 637-639 (2018).
61. Rios, A.C., Fu, N.Y., Lindeman, G.J. & Visvader, J.E. In situ identification of bipotent stem cells in the mammary gland. *Nature* **506**, 322-7 (2014).

62. Lilja, A.M. *et al.* Clonal analysis of Notch1-expressing cells reveals the existence of unipotent stem cells that retain long-term plasticity in the embryonic mammary gland. *Nat Cell Biol* **20**, 677-687 (2018).
63. Lee, H.J. *et al.* Progesterone drives mammary secretory differentiation via RankL-mediated induction of Elf5 in luminal progenitor cells. *Development* **140**, 1397-401 (2013).
64. Janerich, D.T. & Hoff, M.B. Evidence for a crossover in breast cancer risk factors. *Am J Epidemiol* **116**, 737-42 (1982).
65. MacMahon, B., Cole, P. & Brown, J. Etiology of human breast cancer: a review. *J Natl Cancer Inst* **50**, 21-42 (1973).
66. Nichols, H.B. *et al.* Breast Cancer Risk After Recent Childbirth: A Pooled Analysis of 15 Prospective Studies. *Ann Intern Med* (2018).
67. Mohle, R. *et al.* The chemokine receptor CXCR-4 is expressed on CD34+ hematopoietic progenitors and leukemic cells and mediates transendothelial migration induced by stromal cell-derived factor-1. *Blood* **91**, 4523-30 (1998).
68. Burns, J.M. *et al.* A novel chemokine receptor for SDF-1 and I-TAC involved in cell survival, cell adhesion, and tumor development. *J Exp Med* **203**, 2201-13 (2006).
69. Guo, Q., Betts, C., Pennock, N., Mitchell, E. & Schedin, P. Mammary Gland Involution Provides a Unique Model to Study the TGF-beta Cancer Paradox. *J Clin Med* **6**(2017).
70. Barreira da Silva, R. *et al.* Dipeptidylpeptidase 4 inhibition enhances lymphocyte trafficking, improving both naturally occurring tumor immunity and immunotherapy. *Nat Immunol* **16**, 850-8 (2015).
71. Schurmann, C. *et al.* 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 (2012).

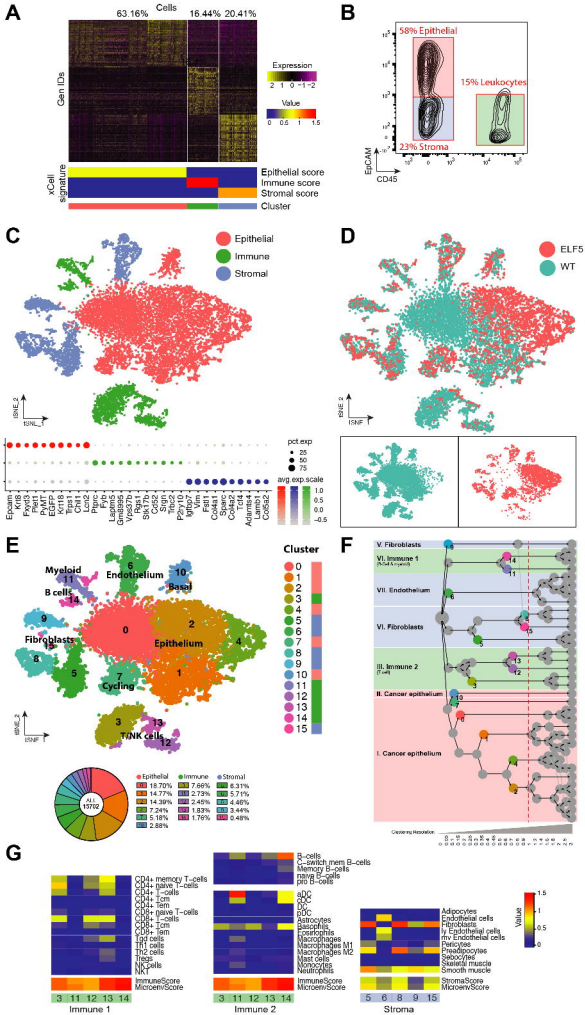

Figure 1. Valdes-Mora et al. 2019

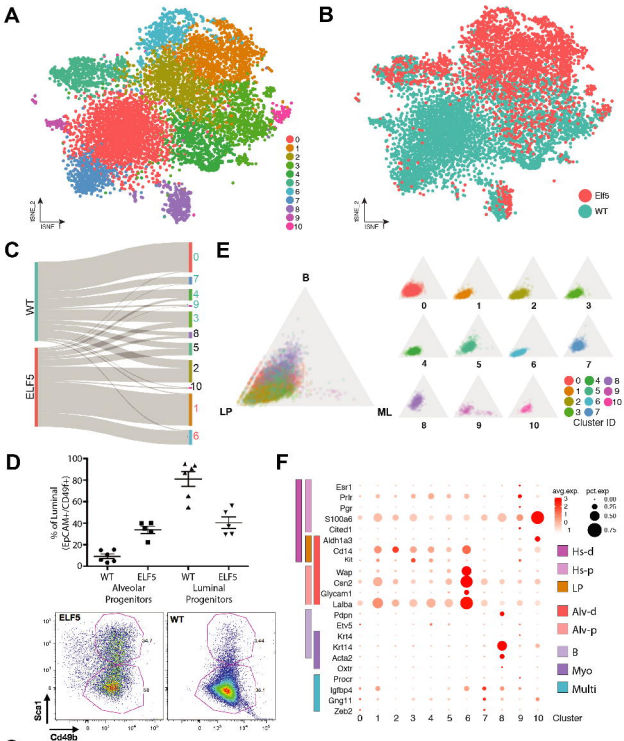

Figure 2 Valdes-Mora et al. 2019

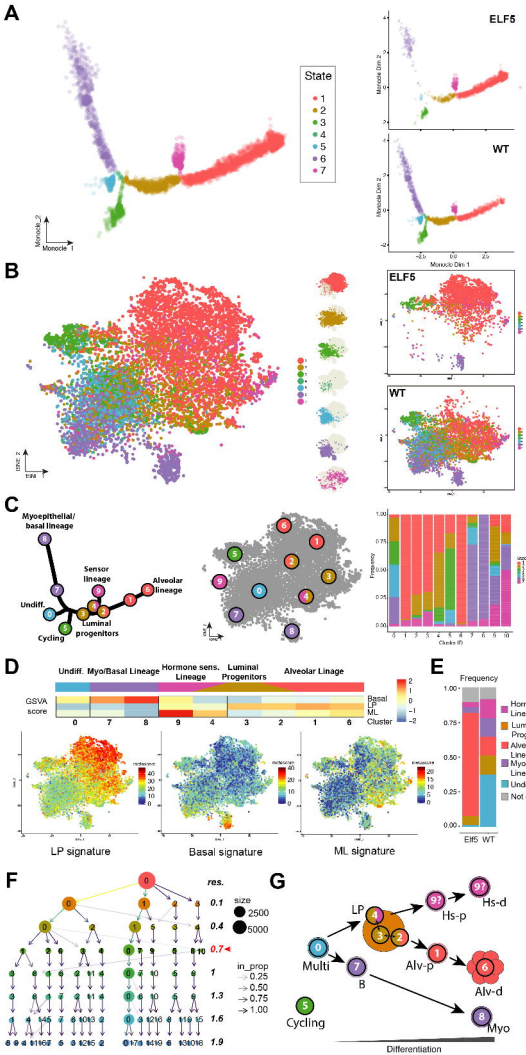

Figure 3 Valdes-Mora et al. 2019

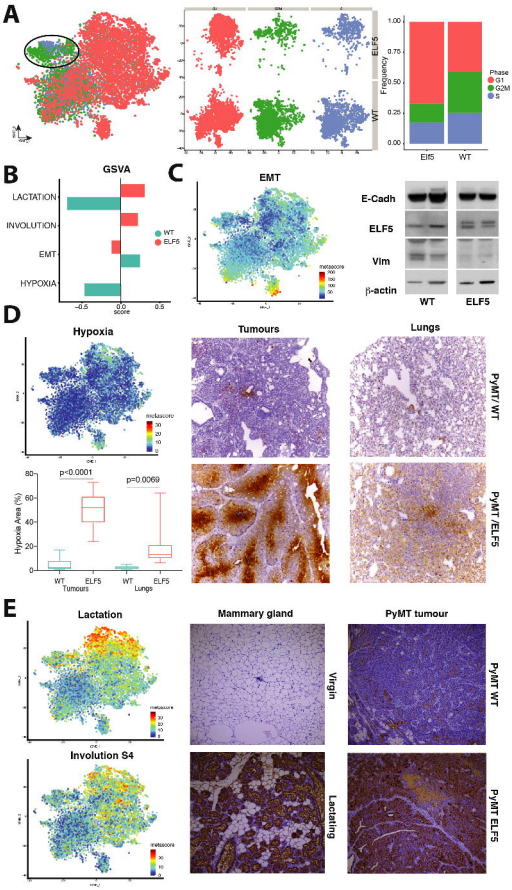

Figure 4 Valdes-Mora et al. 2019

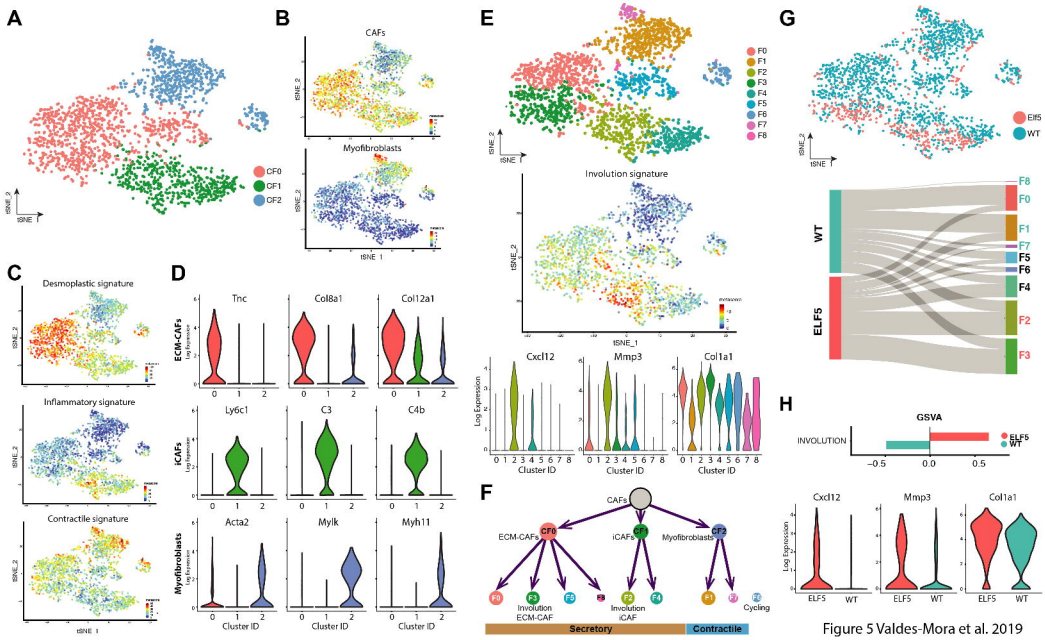

Figure 5 Valdes-Mora et al. 2019

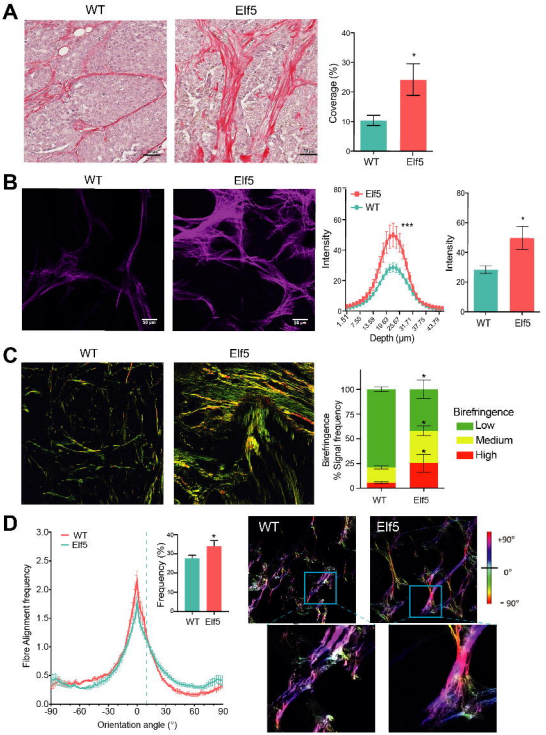

Figure 6 Valdes-Mora et al. 2019

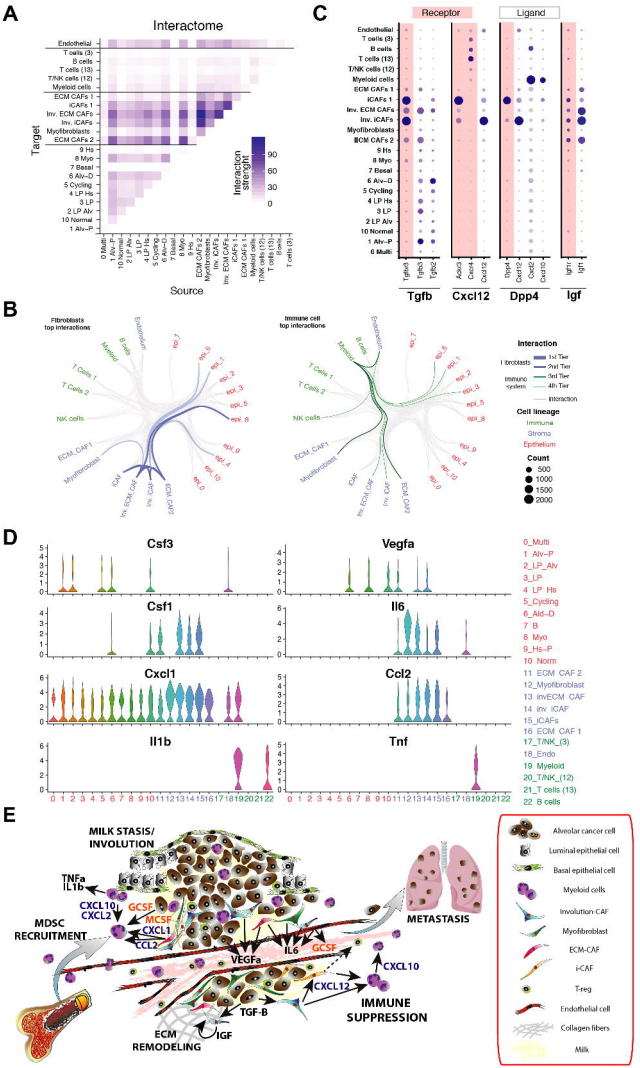

Figure 7. Valdes Mora et al. 2019

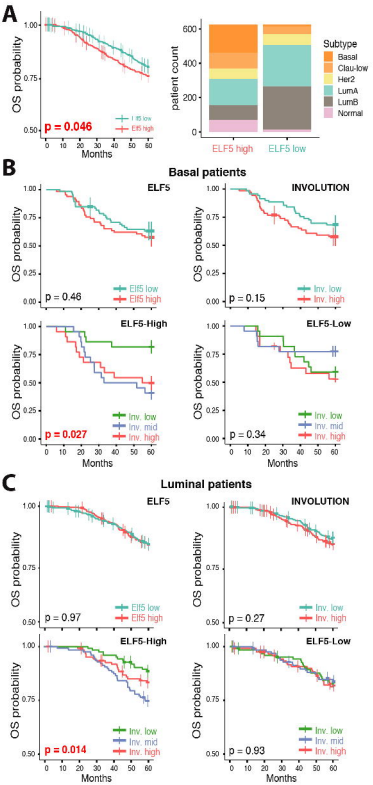

Figure 8. Valdes Mora et al. 2019