Single-cell genomic analysis of triple-negative breast cancer fibroblasts uncovers evolutionarily conserved features and potential therapeutic targets

Ana Paula Delgado[1,2], Alice Nemajerova[1], Jinyu Li[1], Natalia Marchenko[1], Jonathan Preall[3], Ute M. Moll[1], Mikala Egeblad[2,3], Scott Powers[1,2,3,*]

[1]Department of Pathology and Cancer Center, Renaissance School of Medicine, Stony Brook, New York; [2]Graduate Program in Genetics, Stony Brook University, Stony Brook, New York; [3]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

*corresponding author, email: scott.powers@stonybrook.edu

Summary

To explore cancer associated fibroblasts (CAFs) in triple-negative breast cancers (TNBC), we performed scRNA-seq analysis of fibroblasts from murine and human TNBCs. We observed three distinct CAF subtypes in mouse TNBC: two that are intermingled and adjacent to tumor cells, and one that is more distal. We present evidence that progression of CAFs from normal resident fibroblasts/pericytes involves upregulation of their Pdgf and Tgfb receptors along with reciprocal ligand upregulation in other cells within the tumor microenvironment. Additionally, extracellular matrix, glycolytic, and mitochondrial respiratory genes are strongly upregulated in all CAFs. Activation of extracellular matrix genes specifically in CAFs and not in normal fibroblasts provides numerous targets for CAF-based therapeutics, many of which are conserved in CAFs from human TNBC. In contrast, the subtype structure of CAFs was less conserved, which along with their transcriptional heterogeneity suggests that molecular targeting of CAFs is more practical than targeting CAF subtypes.

Introduction

There is renewed interest in exploring cancer-associated fibroblasts (CAFs) and their extracellular matrix components as targets for anti-cancer therapy (Chen et al., 2021; Sahai et al., 2020). One of the main motivations for this revival has been the discovery of discrete CAF subtypes, engendering the hope that targeting specific CAF subtypes will be more effective than targeting all CAFs indiscriminately. In support of this notion, several groups have discovered CAF subtypes that are determinants of response to specific cancer treatments. In breast cancer, CD146-positive (*MCAM*-positive) CAFs were found to maintain estrogen-receptor (ER) expression and sensitivity to tamoxifen, whereas CD146-negative CAFs had opposing effects (Brechbuhl et al., 2017). In non-small cell lung cancer, Hu et al. discovered three CAF subtypes that differed markedly in their capacity to secrete growth factors that counteract the effectiveness of targeted therapies (Hu et al., 2021). Finally, in several cancer types, the presence of LRRC15-positive CAFs was found to be strongly correlated with intrinsic resistance to immune checkpoint therapy (Dominguez et al., 2020).

Most CAF subtypes have been characterized by selective expression of cell-surface or intracellular marker proteins, but a less biased genome-wide method that is currently on hand is single-cell transcriptomics and cluster analysis (Kanzaki and Pietras, 2020). The advantage of the latter method is that the expression of thousands of genes can be analyzed, providing an opportunity to gain deeper insight into the different CAF subtypes, as opposed to making inferences based on the expression of a limited set of markers. This genomic approach to CAF subtype discovery has been previously applied to triple-negative breast cancer, including the transgenic MMTV-PyMT and transplantable 4T1 mouse models as well as human TNBCs (Bartoschek et al., 2018; Friedman et al., 2020; Wu et al., 2020). However, these three reports differ considerably in respect to the number and characteristics of TNBC CAF subtypes. One goal of our work here was to seek to understand the discrepancies of these prior reports. Additionally, we sought to extend the previous findings by performing comprehensive analysis of

the changes that occur in CAFs during cancer development, which we did by performing integrated analysis of single-cell transcriptomes of mammary fibroblasts from normal mice compared to three different mouse models of TNBC. This was not done in prior reports, and importantly this allowed us to distinguish CAF markers and biological processes that are upregulated during tumorigenesis as opposed to CAF markers and biological processes that are intrinsic to the cell-of-origin. Additionally, to determine if the discrepancies between the prior report of human TNBC CAFs and the reports of murine CAFs was due to evolutionary changes, we examined evolutionary conservation of the analysis of CAFs in murine TNBC by performing comparative single-cell transcriptome and cluster analysis of CAFs from human TNBC. The latter analysis enabled us to designate a set of potential therapeutic targets -- CAF markers encoding extracellular components that are evolutionarily conserved and significantly upregulated in cancer relative to normal tissue.

Results

Three transcriptional subtypes of CAFs in murine PyMT breast tumors

Our first step was to perform single-cell RNA sequencing (scRNA-seq) of mammary tumors from four different MMTV-PyMT mice. We used Seurat to perform cluster analysis of the combined single-cell transcriptomes and identified ten distinct clusters. By differential gene expression analysis we were able to assign cell types to the ten clusters, including four clusters comprising the tumor epithelial cells, all of which expressed epithelial markers such as *Epcam*, but differed with respect to expression of differentiation or proliferative markers. The proliferative cluster is labeled *Birc5*$^+$ in the UMAP projection shown in Figure 1A. Similarly the luminal cluster is labeled *Krt8*$^+$, the more differentiated cluster in labeled *Ltf*$^+$, and the invasive cluster is labeled *Krt14*$^+$ (Figure 1A). We detected two *Ptprc*$^+$ immune cell clusters, one comprised of lymphocytes (*Nkg7*$^+$) and the other containing myeloid cells (*Cd74*$^+$) (Figure 1A). There were four stromal cell clusters, including *Pecam1*$^+$ endothelial cells and three different

types of *Col1a1*$^+$ expressing fibroblasts (Figure 1A). We performed differential gene expression analysis and provisionally labeled the three *Col1a1*$^+$ clusters based on their most selectively expressed gene: *Spon1*$^+$, *Clec3b*$^+$, and *Rgs5*$^+$ (Figure 1B). Although *Rgs5* is sometimes considered as a definitive marker for pericytes and used to exclude cells from downstream analysis of CAFs (Elyada et al., 2019; Friedman et al., 2020), *Rgs5*$^+$ cells, along with cells from the other two *Col1a1*$^+$ clusters, expressed a multitude of fibroblast markers. Additionally, this class of fibroblastic cells has been previously designated as CAFs and shown to be localized away from the vasculature in both breast and melanoma tumors (Bartoschek et al., 2018; Davidson et al., 2020; Wu et al., 2020).

The expression of the ten most differentially expressed genes within the three *Col1a1*$^+$ clusters is displayed in a heatmap (Figure 1C). Included in the ten genes of the *Spon1*$^+$ cluster are four that encode extracellular matrix components (*Col8a1, Col12a1, Spon1, Tnc*); for the *Clec3b*$^+$ cluster there are three genes encoding proteins involved in immunity (*Cxcl12, C3, C4b*); and for the *Rgs5*$^+$ cluster there are four genes encoding components of actomyosin contractile filaments (*Acta2, Mylk, Myl9, Tpm2*). These markers for the three CAF subtypes are also found in the three CAF subtypes revealed by scRNA-seq in melanoma described by Teichmann's group (desmoplastic, immune, and contractile) (Davidson et al., 2020) as well as the three CAF subtypes identified by scRNA-seq in MMTV-PyMT mammary tumors by Gallego-Ortega's group (extracellular matrix, immune, and myofibroblast) (Valdes-Mora et al., 2021).

We noted in Figure 1C that *Clec3b*$^+$ CAFs expressed three of the stem cell markers (*Ly6c1*, *Dpt*, *Ly6a*) that are prominently expressed in the least differentiated fibroblast lineage discovered in a large scale scRNA-seq study of over twenty normal and diseased tissues .(Buechler et al., 2021). Another marker from that study, *Lrrc15*, was not found in normal fibroblasts but found in activated myofibroblasts from cancer, arthritis, fibrosis, and wounds, and we noted that it was expressed in *Spon1*$^+$ CAFs (Figure 1C). Even though the expression of the CAF marker smooth muscle actin (*Acta2*) is higher in *Rgs5*$^+$ CAFs, its expression in *Spon1*$^+$

CAFs confirmed its identity as activated myofibroblasts (Figure 1D). Other canonical CAF markers were differentially expressed amongst the three CAF subtypes. Expression of *Pdgfra* and *Pdpn* were restricted to *Spon1*$^+$ and *Clec3b*$^+$ CAFs, whereas *Pdgfrb* expression is detected only in *Rgs5*$^+$ and *Spon1*$^+$ CAFs (Figure 1D). Fibroblast specific protein (*S100a4*) is selectively expressed in *Clec3b*$^+$ CAFs whereas the *Mcam* (*Cd146*) is selectively expressed in *Rgs5*$^+$ CAFs (Figure 1D).

Additionally, we determined that the three CAFs differentially expressed collagens that serve different functions. *Col1a1* and *Col1a2*, encoding collagen type I proteins, form thick collagen fibrils and is expressed most strongly in *Spon1*$^+$ CAFs, followed by *Clec3b*$^+$ CAFs, and then *Rgs5*$^+$ CAFs which express the lowest amount (Figure 2A) (Mayne and Burgeson, 1987). *Col3a1*, encoding a collagen type III protein, plays a key role in early stages of wound healing and forms thinner collagen fibrils; is expressed most strongly in *Clec3b*$^+$ CAFs, followed by *Spon1*$^+$ CAFs, then *Rgs5*$^+$ CAFs (Figure 2A). *Col4a1* and *Col4a2*, encoding collagen type IV proteins, does not form fibrils but instead is a key component of basement membranes; it is expressed most strongly in *Rgs5*$^+$ CAFs (Figure 2A). Additionally, *Clec3b*$^+$ CAFs selectively expressed the hyaluronan synthase gene *Has1*, which produces a major component of the extracellular matrix (Figure 2A).

Spon1$^+$ *CAFs are tumor-adjacent whereas Clec3b*$^+$ *CAFs are confined to the stroma*

Having identified markers that are selectively expressed in the three CAF subtypes, our next step was to use these selective markers to determine their location within tumor tissue. We used RNA in situ hybridization (ISH) to visualize *Spon1* and *Clec3b* RNA molecules in hematoxylin-stained, formalin-fixed tissue samples of mammary tumors from MMTV-PyMT mice. The RNAscope probes to *Spon1* and *Clec3b* were differentially labeled with a cyan dye and magenta dye, respectively. Shown in Figure 2B are images from eight different tumors. In all eight images, *Spon1*$^+$ and *Clec3b*$^+$ CAFs display non-overlapping, discrete locations within

the tumor. *Spon1*$^{+}$ CAFs are located adjacent to the epithelial tumor nests, whereas *Clec3b*$^{+}$ CAFs are located within the surrounding stroma. In areas containing stromal adipocytes, *Clec3b* RNA staining is observed within some adipocytes, consistent with previous observations of its expression in murine adipose tissue (Plasko, 2021). We note that due to their large size and other properties adipocytes are not robustly profiled by standard scRNA-seq methodology.

We quantified the localization of *Spon1*$^{+}$ and *Clec3b*$^{+}$ CAFs to areas containing tumor epithelial cells as opposed to stromal areas that were devoid of tumor cells. The results of analyzing a total of ten different slides (five each from two different tumors) are shown in Figure 2C. *Spon1* staining comprised up to 6% of the total area containing tumor epithelial cells, and less than 1% of the total stromal area; whereas *Clec3b* staining comprised up to 1.5% of the stromal area, but less than 0.2% of the tumor epithelial area (Figure 2C).

We detected *Clec3b* and *Spon1* signals in normal mouse tissue, with the *Spon1* signal coming from cells in between the two ducts and *Clec3b* signals being more distal (Figure 2D). We also detected tumor-adjacent *Spon1*$^{+}$ CAFs and distal *Clec3b*$^{+}$ CAFs in both C3-SV40 T-antigen (Figure 2E) and MMTV-Erbb2 mammary tumors (Figure 2F).

Rgs5^{+} *CAFs are tumor-adjacent and intermingle with Spon1+ CAFs*

The first indication that *Rgs5*$^{+}$ CAFs were intermingled with *Spon1*$^{+}$ CAFs came from indirect evidence. As shown in Figure 1D, both of these CAFs express *Pdgfrb* and *Acta2*, and the presence of *Pdgfrb*$^{+}$ *Spon1*$^{-}$ (Figure 2G) and *Acta2*$^{+}$ *Spon1*$^{-}$ (Figure 3A) cells intermingled with doubly positive cells suggested that *Rgs5*$^{+}$ CAFs were localized to the same areas as *Spon1*$^{+}$ CAFs and 3A). This was confirmed by multiplexed fluorescent RNA-ISH, using differentially labeled probes for all three CAF markers. As shown in Figures 3B and 3C, both *Rgs5* and *Spon1* probes were detected in between tumor cell nests, which appear dark blue from 4′,6-diamidino-2-phenylindole (DAPI) staining. There was more *Spon1* signal than *Rgs5*, consistent

with the relative cell numbers of the two CAF subtypes estimated by scRNA-seq (Figure 1A). *Clec3b* signal is distally located away from the tumor nests (Figure 3B and 3C).

Comparing CAFs to their normal counterparts

An important goal of our study was to understand what expression changes occurred in CAFs as part of cancer progression. Although the cell-of-origin for most CAFs have not been rigorously pinpointed, several studies support the concept that CAFs derive from tissue-resident cells (Arina et al., 2016; Kobayashi et al., 2022; Sahai et al., 2020). To enable comparison of TNBC CAFs to normal breast tissue-resident cells we performed scRNA-seq analysis of mammary glands from mice matched for age and genetic background with the MMTV-PyMT mammary tumors. By differential gene expression analysis we assigned cell types to seventeen clusters, including four clusters comprising the epithelial compartment: a luminal cluster (*Krt8*$^+$), two myoepithelial clusters (*Krt14*$^+$ and *Krt17*$^+$), and a developmental cluster (*Areg*$^+$) (Figure 3D). Tumor stromal clusters included six immune cell clusters, three endothelial cell clusters, pericytes (*Rgs5*$^+$) and three *Col1a1*$^+$ fibroblastic clusters: *Spon1*$^+$, *Clec3b*$^+$, and *C3*$^+$ (Figure 3D). The Rgs5+ cluster expressed other canonical pericyte markers including desmin (*Des*) and smooth-muscle actin (*Acta2*), as well as twenty pericyte markers from comparison to single-cell RNA data using Enrichr (Xie et al., 2021).

To gain insight into the three *Col1a1*$^+$ fibroblastic clusters, we compared their one-hundred top differentially expressed genes (ranked by FDR) to three fibroblast lineages discovered in the scRNA-seq study of over twenty normal tissues. Two of these lineages were found in all tissues (*Pi16*$^+$, *Col15a1*$^+$) and another lineage was found in the spleen and lymph nodes (*Ccl19*$^+$) (Buechler et al., 2021). Notably, over 80% of the top one-hundred differentially expressed genes of the mammary gland *Clec3b*$^+$ cluster matched differentially expressed genes of the *Pi16*$^+$ lineage, considered to be the most stem-like of all fibroblast lineages (Figure 3E). The other two mammary fibroblast clusters also had over 80% identical matches with markers

for the three fibroblast lineages: *Spon1*+ fibroblasts matched equally to both the *Pi16*$^{+}$ and *Col15a1*$^{+}$ lineages, whereas *C3*$^{+}$ fibroblasts largely matched to the *Ccl19*$^{+}$ lineage (Figure 2E).

Armed with this analysis of tissue-resident fibroblasts and pericytes from normal mammary glands, we next performed side-by-side scRNA-seq profiling of normal mammary glands, MMTV-PyMT mammary tumors, and C3-promoter driven SV40 T-antigen mammary tumors, and compared CAFs with their corresponding normal tissue-resident cells. We also included publicly available data from *Brca1*$^{-}$/*Brca1*$^{-}$ *Trp53*$^{-}$/*Trp53*$^{-}$ mammary tumors (Wang et al., 2019). We again used Seurat to perform comparative scRNA-seq analysis across these four conditions. This software first identifies cross-dataset anchors to perform batch effect corrections. The resultant nineteen clusters identified are shown in Figure 3F. There were eight epithelial cells clusters, including a *Snai1*$^{+}$ cluster that represented tumor cells from *Brca1*$^{-}$ /*Brca1*$^{-}$ *Trp53*$^{-}$/*Trp53*$^{-}$ and C3-promoter driven SV40 T-antigen mammary tumors that had undergone epithelial-to-mesenchymal transition (EMT) (Figure 3F). Both of these mammary tumor genotypes resemble the basal subtype of human TNBC, whereas the PyMT model resembles luminal TNBC.

Again there were three *Col1a1*$^{+}$ clusters labeled *Spon1*$^{+}$, *Clec3b*$^{+}$, and *Rgs5*$^{+}$ (Figure 3G). Overall, MMTV-PyMT tumors contained a greater percentage of CAFs (9%*)* than either C3-SV40 T-antigen tumors (5% CAFs) or *Brca1*$^{-}$/*Brca1*$^{-}$ *Trp53*$^{-}$/*Trp53*$^{-}$ tumors (4%CAFs). Additionally, the three different mammary tumor types showed a different proportion of the three CAFs: MMTV-PyMT tumors had less *Clec3b*$^{+}$ cells and more *Rgs5*$^{+}$ cells compared to normal mammary tissue, similar to the findings of Teichmann's group in melanoma (Davidson et al., 2020); C3 SV40 T-antigen tumors had expanded the proportion of *Clec3b*$^{+}$ cells; and *Brca1*$^{-}$ /*Brca1*$^{-}$ *Trp53*$^{-}$/*Trp53*$^{-}$ tumors had a large expansion of *Rgs5*$^{+}$ cells (Figure 3G).

Next, to gain insight into CAF progression we set out to determine which of the ten most differentially expressed genes within the three MMTV-PyMT CAF subtypes (Figure 1C) were expressed due to cancer progression *versus* as a result of their cell-of-origin. As shown in

Figure 3H, all ten of the selective *Spon1*$^{+}$ CAF markers are expressed due to cancer progression. In sharp contrast, all ten *Clec3b*$^{+}$ CAF markers are expressed at similar levels in both normal and PyMT cells, and apart from *Higd1b*, the same holds true for *Rgs5*$^{+}$ CAF markers (Figures 4A-4B). As CAF markers are sometimes assumed to be cancer-specific, these results indicate that several of the CAF markers reported in the literature may have nothing to do with cancer per se, but rather reflect the transcriptional state of the cell-of-origin. This prompted us to examine canonical CAF markers. For this analysis, we compared normal tissue-resident cells to all three different TNBC models. *Pdgrb* stood out as the canonical CAF marker that is the most dependent on carcinogenesis, and is induced by cancer in both *Spon1*$^{+}$ and *Rgs5*$^{+}$ CAFs (Figure 4C). Although *Pdgfra* is also induced by cancer, its expression in normal *Spon1* fibroblasts is similar to its level in PyMT *Spon1*$^{+}$ CAFs. Similarly, *Acta2* expression in *Rgs5* CAFs is at levels similar to those found in normal tissue-resident *Rgs5*+ cells (Figure 4C). *Pdpn* expression is present in normal *Spon1*$^{+}$ and *Clec3b*$^{+}$ fibroblasts although induced to higher levels in CAFs, whereas *S100a4* expression is cancer-dependent and largely restricted to *Clec3b*$^{+}$ CAFs (Figure 4D). Finally, *Fap* expression, which is lower in our mammary tissues than others have reported for other tissue types, is nearly identical in normal and cancer samples (Figure 4E).

Gene sets overrepresented in the transition from normal tissue resident cells to CAFs

To gain a genome-wide perspective on the gene expression changes accompanying CAF cancer progression, we performed gene set enrichment analysis using the differentially (over)expressed genes in MMTV-PyMT CAFs along with GO term enrichment analysis for both biological processes and cellular components (http://geneontology.org). The most significant in all three CAF subtypes was the enrichment in cellular components of genes encoding extracellular matrix proteins (Figure 4F). Related to this category is the broader group of genes encoding all components of the extracellular region, which includes secreted proteins as well as

transmembrane receptors. This group was also enriched in all three CAFs (Figure 4F). The second most significant enrichment in all three subtypes was enrichment of genes encoding proteins found in the mitochondrial membrane (Figure 4F) which was mirrored in enrichment of genes involved in ATP biosynthesis (Figure 4G).

We observed enrichment for transforming growth factor beta, Wnt signaling and the glycolytic process specifically in the *Spon1*$^+$ and *Clec3b*$^+$ CAFs (Figure 4G). However, a larger group of gene sets were enriched selectively in *Spon1*$^+$ and *Rgs5*$^+$ CAFs, including focal adhesion, cell substrate junction, wound healing, myofibril assembly, extracellular matrix organization, cell-matrix adhesion, and cell migration (Figures 4F,4G). Thus ,a high degree of functional overlap accompanies their co-localization inside tumors.

To determine whether the genes overexpressed in MMTV-PyMT CAFs were also overexpressed in the two other mouse TNBC models, we examined the expression of genes representative of extracellular matrix, mitochondrial, and glycolytic gene sets. *Col1a1*, *Col3a1*, and *Col4a1* were found to be significantly overexpressed in all three TNBC models, displaying the same CAF subtype selectively across all models (Figure 5A). Similarly, the mitochondrial respiratory genes *Cox5b*, *Uqcrc*, and *Atp5j2*, along with the glycolytic genes *Pkm*, *Gapdh*, and *Aldoa* are induced in all CAF subtypes in each TNBC model (Figure 5B, 5C).

Upregulation of TGFβ and PDGF receptors in CAFs in parallel with upregulation of TGFβ and PDGF ligands in other cell types

One of the gene sets enriched in *Spon1*$^+$ and *Clec3b*$^+$ CAFs compared to normal tissue-resident cells is TGFBR signaling (Figure 4G). We noted that this included upregulation of *Tgfbr2* expression in both of these CAFs and upregulation of *Tgfbr3* in *Clec3b*$^+$ CAFs, across all three TNBC models (Figure 5D). By applying NicheNet, a computational method that predicts ligand–target links between interacting cells by searching for co-upregulation of ligands and receptors along with testing whether gene sets associated with the response to a given

ligand are induced in the target cell (Browaeys et al., 2020), we found that tumor cells in the PyMT model indeed upregulated the ligands encoded by *Tgfb2* and *Tgfb3*, and that the predicted targets of these ligands were equally upregulated in *Spon1*$^+$ CAFs (Figure 5E). Along the same notion, we previously noted that the CAF markers *Pdgfra* and *Pdgfrb* were selectively upregulated in *Spon1*$^+$ and *Clec3b*$^+$ CAFs, and *Spon1*$^+$ and *Rgs5*$^+$ CAFs, respectively (Figure 4C). Accordingly, we examined expression of Pdgf receptor ligands (along with Tgf receptor ligands) in different cell types of both normal and tumor samples. In both the MMTV-PyMT and C3-T-antigen tumor models, their lymphocytes expressed higher levels of *Tgfb1* compared to normal breast tissue, whereas myeloid cells expressed considerably higher levels of *Pdgfa* and somewhat higher levels of *Pdgfb* (Figure 5F). In the MMTV-PyMT model but not the C3 T-antigen model, endothelial cells expressed higher levels of *Tgfb1*, *Pdgfa*, and *Pdgfb*, and tumor epithelial cells expressed higher levels of *Tgfb2* and *Tgfb3* (Figure 5F).

We used RNA in situ hybridization (ISH) to visualize *Tgfbr2* and *Tgfb3* RNA molecules in mammary tumors from MMTV-PyMT mice. The RNAscope probes to *Tgfbr2* and *Tgfb3* were differentially labeled with a cyan dye and magenta dye, respectively. Shown in Figure 5G is a representative image of the magenta *Tgfb3* probe brightly staining the tumor epithelial cells, whereas the cyan *Tgfbr2* probe is staining the fibroblasts lying in between tumor cell nests, establishing this reciprocal signaling system. Based on these findings, a model for how secretion of these two classes of growth factors in MMTV-PyMT mammary tissues stimulates the development of CAFs through corresponding receptors is shown in Figure 5H.

Human TNBCs exhibit three CAF subtypes with varying similarity to murine CAF subtypes

To compare our results from the 3 mouse models to human TNBC, we performed scRNA-seq analysis of three human TNBC samples and integrated the three datasets using Seurat. Differential expression analysis revealed that there were seven epithelial cells clusters, five immune cell clusters, and one endothelial cell cluster (Figure 6A). There were three collagen

expressing clusters labeled *COL1A1*$^+$, *COL3A1*$^+$, and *RGS5*$^+$ (Figure 6A). The only lead marker of murine CAFs that is conserved in human CAFs is *RGS5*. *CLEC3B* was not expressed at appreciable levels in any of these clusters, and as shown below, *SPON1* expression is not conserved either.

The expression profiles of *COL1A1*, *COL3A1* and *COL4A1* and other key genes are shown in Figure 6B. Analogous to the murine *Clec3b*$^+$ cluster, the *COL3A1*$^+$ cluster preferentially expresses *COL3A1*, and analogous to the murine *Spon1*$^+$ cluster, the *COL1A1*$^+$ cluster preferentially expresses *COL1A1*. Consonant with the murine *Rgs5*$^+$ cluster, the human *RGS5*$^+$ cluster selectively overexpresses *COL4A1* and *ACTA2*. These results largely coincide with the analysis of human TNBC CAFs by Swarbick's group (Wu et al., 2020). Based on the selective expression of *TAGLN* and *FAP* in our *COL1A1*$^+$ and *COL3A1*$^+$ clusters (Figure 6B), our *COL1A1*$^+$ cluster corresponds to their myCAF cluster, and our *COL3A1*$^+$ corresponds to their iCAF cluster. Additionally, corresponding to the murine *Clec3b*$^+$ cluster, the *COL3A1*$^+$ cluster preferentially expresses genes encoding hyaluronic acid synthase (Figure 6C).

Systematic analysis of the overlap in the differentially expressed genes (markers) within the three CAF clusters for human and murine TNBC revealed >80% conservation of markers for the *RGS5*$^+$ clusters, but no such corresponding conservation for either the *COL1A1*$^+$ or *COL3A1*$^+$ clusters, which based on overall markers are equally related to murine *Spon1*$^+$ or *Clec3b*$^+$ clusters (Figure 6D). Examples of markers that have 'switched sides' include *C3*, *TNXB*, and even *SPON1* itself (Figures 6E and 6F).

Heterogeneity within CAF subtypes

The lack of strong evolutionary conservation of the two of the CAF subgroups prompted us to reconsider the biological importance of CAF subtypes. What if CAF subtypes themselves were heterogenous, and what we assumed were common functions of one group of CAFs were being carried out by different subgroups? To examine the heterogeneity of CAF subtypes, we

performed subclustering analysis. We found three distinct subgroups of *Clec3b*$^+$ fibroblasts that separated several key *Clec3b*$^+$ markers. One subgroup selectively expressed Clec3b along with extracellular matrix genes *Mfap5* and *Tnxb* and immune gene *Ly6c1* ; another subgroup selectively expresses the immune genes *C3* and *Cxcl12* that underlie in part other groups designation of the *Clec3b*$^+$ subtype as 'immune' or 'inflammatory' along with extracellular matrix genes *Mgp* and *Bgn*; and the third subgroup selectively expresses *Col3a1* along with immune markers *Ccl7* and *Cxcl1* (Figure 6H). These results show that all three subgroups express both immune genes and extracellular matrix genes. Although there may be functional distinction between the groups, there isn't one aligned with these two groups of genes. We next wanted to see if any of the subgroups were associated with genes induced during cancer formation. With the first subgroup, *C3* was expressed equally in both normal and cancer-associated fibroblasts, but *Bgn* and *Igf1* were both cancer-induced; in the second subgroup, *Clec3b* and *Dpt* were expressed equally in normal and cancer fibroblasts, but *Tnxb* and *Mfap5* were cancer-induced; and in the third group *Ccl7* was expressed equally in normal and cancer fibroblasts, but *Col3a1* and *Plac8* were cancer-induced (Figure 6I).

The most evolutionarily conserved *RGS5*$^+$ CAF subtype showed divergent evolution upon subcluster analysis . Both murine and human *RGS5*$^+$ CAFs could be divided into two subgroups, with one subgroup selectively expressing *COL4A1*, *FN1*, *MGP*, and *PDGFRB*; and the other subgroup selectively expressing *SPARCL1*, *ACTA2*, and *MYLK* However, both *RGS5* and *BGN* switched subgroups (Figure S1A-S1D). Each of the other TNBC CAF clusters showed evidence of heterogeneity, including the murine *Spon1*$^+$ cluster and the human *COL1A1*$^+$ and *COL3A1*$^+$ clusters (Figure S1E-S1J).

Spatial heterogeneity within CAF subtypes

To test whether the transcriptional heterogeneity we observed within murine *Clec3b*$^+$ CAFs was reflected in spatial organization, we compared the location of the two subgroups of *Clec3b*$^+$

CAFs defined by selective expression of *Ccl7* and *Clec3b* itself. First, we plotted the expression of *Ccl7* and *Clec3b* for each of the single cells in the *Clec3b*$^+$ cluster. We observed distinct heterogeneity, with some cells expressing relatively more *Ccl7*, some relatively more *Clec3b*, some expressing equal amounts, and some expressing low amounts of both genes (Figure 7A), consistent with the subclustering results. Next, we used dual-labelled ISH probes to visualize the presence of CAFs selectively expressing *Ccl7* compared to the presence of CAFs selectively expressing *Clec3b*. As shown in Figure 7B, the selectively expressing *Ccl7* CAFs were located adjacent to the tumor cells, while the selectively expressing *Clec3b* CAFs being even more tumor-distal.

Conserved CAF Therapeutic Targets

Recently, monoclonal antibodies directed against two extracellular matrix proteins produced and secreted by CAFs, i.e. LRCC5 and MFAP5, have achieved impressive preclinical anti-tumor efficacy results (Purcell et al., 2018; Yeung et al., 2019). We examined the expression of the corresponding target genes *Lrrc15* and *Mfap5* in mouse and human TNBC CAFs. *Lrrc15* expression in normal fibroblasts is minimal and induced in *Spon1*$^+$ CAFs in all three mouse models of TNBC (Figure 7C). Unfortunately, although it is also selectively expressed in the corresponding CAF subtype (*COL1A1*$^+$) in human TNBC, it is expressed in less than 10% of the cells (Figure 7D). *Mfap5* expression in normal fibroblasts is higher than *Lrrc15* and although it is induced in *Spon1*$^+$ and Clec3b+ CAFs in all three mouse models of TNBC (Figure 7C), it is expressed in less than 40% of the cells in human *COL1A1*$^+$ CAFs (Figure 7D).

We decided to look in more detail at other possible targets within the large number of extracellular matrix genes that we found were minimally expressed in normal fibroblasts but significantly upregulated in MMTV-PyMT CAFs. By filtering for those that were expressed in less than 15% of normal fibroblasts, and with a corresponding > 40% increase in CAFs, we found a total of eighteen extracellular matrix genes that were mostly expressed in *Spon1*$^+$ CAFs,

although one was selectively expressed in *Clec3b*$^{+}$ CAFs (*Fbn1*) and two were selectively expressed in *Rgs5*$^{+}$ CAFs (*Tagln* and *Serpine2*) (Figure 7E). We removed one of these genes, *Igfbp3*, from further consideration since it was more prominently expressed in endothelial cells rather than in CAFs (Figure S2A). Of the remaining seventeen genes, sixteen were expressed in human TNBC CAFs (Figure S2B) but only seven were found to be expressed at sufficiently high proportion (in > 40% of CAFs) (Figure 7F). Of these seven ECM genes, two were expressed at lower levels in most cell types (*CD59* and *TIMP1*), leaving five genes that were robustly and selectively expressed in TNBC CAFs (Figure 7G).

Discussion

We had four major goals for this study, including (1) reconciling discrepancies between previous reports that also had utilized scRNA-seq to characterize TNBC CAF subtypes; (2) determining which markers and other aspects of CAFs were a consequence of cancer development as opposed to their cell-of-origin; (3) exploration of the evolutionary conservation and divergence between murine and human TNBC CAFs; and (4) pinpointing potential therapeutic strategies for the future. We believe we realized or partially realized all four of these goals, as discussed below.

In both human and murine TNBC CAFs, we observed three subtypes, which in both cases were comprised of a conserved *Rgs5*$^{+}$/*RGS5*$^{+}$ cluster, and less conserved *Spon1*$^{+}$/*COL1A1*$^{+}$ and *Clec3b*$^{+}$/*COL3A1*$^{+}$ clusters. In terms of reconciling discrepancies with prior reports, our results are actually in close agreement with previous analyses of MMTV-PyMT CAFs, except that the report by Pietras's group did not observe *Clec3b*$^{+}$/*COL3A1*$^{+}$ fibroblasts, likely because these authors did not analyze the entire tumor but instead used a negative-selection flow cytometry strategy to purify mesenchymal cells (Bartoschek et al., 2018). Their two main subtypes, comprising over 80% of their CAFs correspond to our *Rgs5*$^{+}$/*RGS5*$^{+}$ and *Spon1*$^{+}$/*COL1A1*$^{+}$ subtypes. Our results are in complete agreement with a subsequent report

using the MMTV-PyMT model, identifying three CAF subtypes with overlapping marker genes (Valdes-Mora et al., 2021). However, we believe it is inaccurate to label the *Clec3b*$^+$/*COL3A1*$^+$ cluster as “iCAF” for inflammatory CAFs, as this can be misleading to suggest that cancer-associated inflammation is associated with this subtype. Instead, our results show that immune markers such as *Cxcl12* and *C3* are expressed at similar levels in normal tissue-resident fibroblasts. In the absence of an established functional role we therefore suggest the term *tumor-distal* or *distal* CAFs since their physical location at some distance away from tumor cells is a very distinct feature.

Our results are also in close agreement with the analysis of human TNBC CAFs by Swarbrick’s group, with the exception that they found two *RGS5*$^+$ clusters as part of their four subtype classification, and which actually correspond to the two *RGS5*$^+$ subclusters we identified. Importantly, they observed that their *COL3A1*$^+$ CAFs were located distal to tumor cells, consistent with our finding that the *Clec3b*$^+$ CAFs are tumor-distal (Wu et al., 2020). Our results are not in close agreement with the report from Scherz-Shouval’s group, who observed two major CAF subtypes, only one of which (pCAFs) corresponds to our three CAFs (*Spon1*$^+$/*COL1A1*$^+$ CAFs) (Friedman et al., 2020). Their other major CAF subtype, sCAFs, is not found in our study or in the other three studies and may be due to their derivation from bone-marrow mesenchymal progenitor cells rather than tissue-resident cells, which could have resulted from the use of a transplantable tumor model, whereas our study and the other three studies used either transgenic mice or human tumors. Additionally, these authors also used a negative-selection flow cytometry strategy to purify fibroblasts, which likely accounts for their inability to detect *Clec3b*$^+$ CAFs (Friedman et al., 2020). Finally, we note that these authors categorically dismissed *Rgs5*$^+$ cells as potential CAFs, based on the assumption that they represented pericytes (Friedman et al., 2020).

Another key goal for our study was determining which markers and other aspects of CAFs were due to their cell-of-origin rather than a consequence of cancer progression. Notably, the

canonical marker *Fap* appears to be expressed at comparable levels in normal progenitors, at least in mouse models of triple negative breast cancer, calling into question its utility as a target in anti-cancer antibody treatments (Hamson et al., 2014), as opposed to targets with a much larger difference in expression between normal fibroblasts versus CAFs such as *Lrrc15*. Additionally, the presence of *Acta2* (smooth muscle actin) in CAFs is almost always interpreted as evidence of activation of fibroblasts into myofibroblasts. However, our results (and others) would suggest that many of the strongly *Acta2*$^+$ CAFs observed are actually perivascular-like *Rgs5*$^+$ CAFs, which we showed do not express more *Acta2* than their corresponding normal counterparts.

An important outcome of comparing CAFs to their normal counterparts came from GO term enrichment analysis which revealed that the most significant alterations of all three CAF subtypes was the dramatic increase in expression of extracellular matrix genes and proteins, including collagens and fibronectin. We also observed significant upregulation of metabolic genes, both mitochondrial respiratory and glycolytic, consistent with a role of CAFs in supplying energy and nutrients to cancer cells (Sahai et al., 2020). Both increased expression of extracellular matrix genes and metabolic activation are known consequences of TGFβ or PDGF induction during fibroblast activation. Indeed, we observed increased production of both TGFβ and PDGF ligands in several tumor epithelial cells and tumor microenvironmental cell types, coupled with corresponding upregulation of receptors for these ligands in CAFs.

We also explored the evolutionary conservation or divergence between murine and human TNBC CAFs. Importantly, apart from *Rgs5*$^+$/*RGS5*$^+$ CAFs, CAF subtypes were not as conserved as might have been expected. In fact, based on analysis of marker expression alone, it is not clear how to align the murine *Spon1*$^+$ and *Clec3b*$^+$ CAF subtypes with the human *COL1A1*$^+$ and *COL3A1*$^+$ subtypes. The reasons why we aligned *Clec3b*$^+$ CAFs with *COL3A1*$^+$ CAFs are their shared tumor location, more distal from tumor cells than other CAFs; as well as

their shared expression of both hyaluronic acid synthase genes and the reticular collagen gene *Col3a1*/*COL3A1*. Similarly, we aligned *Spon1*+ and *COL1A1*+ CAFs based on their selective overexpression of *Col1a1*/*COL1A1*. Nevertheless, the evolutionary divergence of two out of three CAF subtypes presents difficulties for researchers hoping to extrapolate findings from mouse models to human cancer.

To probe this evolutionary divergence, we performed subclustering analysis of the three CAF subtypes in both human and murine TNBCs (Figure 6). This analysis revealed an unexpected degree of heterogeneity within the three CAF subtypes, which we showed also corresponded to spatial location heterogeneity. This heterogeneity gives pause and suggests that the current trend in CAF research to devise strategies to specifically target subtypes or convert one subtype to another may be quixotic. Instead, based on our results we propose a more practical path forward, that is to devise strategies based on CAF molecules that show a large difference in expression between CAFs and normal fibroblasts. Our work revealed a substantial number of potential CAF-specific targets that resemble the promising attributes of *Lrrc15*/*LRRC15*: (*i*) minimal expression in normal fibroblasts, (*ii*) strong induction in at least one CAF subtype, and (iii) suitability for targeting by monoclonal antibodies. Antibodies to two of the five targets that we identified, *CDH11* and *OLFML3*, have already been tested in vivo in mice and showed strong anti-tumor activity (Miljkovic-Licina et al., 2012; Peran et al., 2021). By focusing on specific molecular targets, we believe that CAF-targeted therapy will take its place in the toolkit of the oncologist within a reasonable time frame and chance for success.

Methods

Animals

Animal experiments were carried out at the Division of Laboratory Animal Resources at Stony Brook University in accordance with Institutional Animal Care and Use Committee-approved procedures. Wild-type (FVB) and MMTV-PyMT (FVB) mice were obtained from

Jackson Laboratory. Additional MMTV-PyMT (FVB) and the C3(1)/Tag (FVB) mice were obtained from Dr. Mikala Egeblad (Cold Spring Harbor Laboratory). Four 12-week-old MMTV-PyMT/FVB mice, two corresponding wild-type FVB mice, and four C3(1)/Tag (FVB) mice were used to prepare the mammary tumor single-cell preparations for scRNA-seq and histological slides. In addition, tissue sections from mammary glands of a 52-week Erbb2 transgenic mouse obtained from Dr. Natasha Marchenko were used for histological studies.

Tumor sample dissociation into a single-cell suspension

The mammary tumors were harvested from the mice's left and right mammary glands (four and five). The tissue was finely minced and placed in 15 ml conical tubes with a dissociation solution composed of Collagenase/Hyaluronidase (Stem Cell Technologies) enzymes, 5% fetal bovine serum (FBS) (Gemini Bioproducts), and 1 mg/mL DNase 1 (Stem Cell Technologies). The samples were digested for 1 hour at 37°C with constant low agitation using a thermomixer (Eppendorf). The digested cells were pelleted and washed twice with RPMI media in the presence of 5% FBS. The cell suspensions were filtered using a 70 µm cell strainer (DB) to collect single-cell suspensions and single-cell digested tissue. The cells were resuspended and washed twice in 3 ml Red Blood Cell Lysis Buffer (Roche) to remove any visible blood cells. The final cell suspensions were pelleted by centrifuging at 300Xg for 5 minutes and resuspended in 1 ml RPMI media with 5% FBS and restrained again with a 70 µm strainer to a final concentration of 10,000 cells/ml. The cell viability was examined using trypan blue exclusion (Invitrogen).

scRNA-seq and sequencing library construction using 10X Genomics platform

Approximately 10,000 single cells resulting from the single-cell suspensions of each tumor were loaded into the 10x Genomics microfluidics device along with 10X Genomics gel beads (kit

v2 PN-120237) containing barcoded oligonucleotides, reverse transcription (RT) reagents, and oil, resulting in gel beads in emulsion. The scRNA-seq library preparation followed the manufacturer's protocol (10x Genomics) using the Chromium Single Cell 3-Library. These libraries were paired end sequenced using the Illumina HiSeq 4000 sequencing system.

Bioinformatic processing of scRNAseq data

Cell Ranger Single-data suite 6.01 pipeline was used to demultiplex and identify the assigned barcodes for the mammary tumor and normal mammary samples and process the raw sequencing data of each sample. We used the Cell Ranger analysis pipeline (version 6) to process the 10X Chromium single-cell data and the Seurat R package (version 4.0.0) for downstream analysis and assessment of the scRNA-seq data collected from our experiments.

Tissue preparation of histological sectioning, fixation and staining

The tissue sections preserved for histology purposes fixed in formalin, were switched to 70% ethanol within 24 hours and were sent to the Stony Brook Histology Core for paraffin embedding. Tissue sections were cut to 5 µm thickness and were assessed by RNAScope. Images were captured at 20X and 40X magnification.

RNAScope mRNA In Situ Hybridization Assay

RNAscope® technology was used to perform the assay to spatially resolve our scRNA-seq data. The 2.5 HD Duplex Reagent Kit (CN: 322430, Advanced Cell Diagnostics) and RNAscope® Fluorescent Multiplex Reagent Kit (CN: 320850, Advanced Cell Diagnostics) were used to perform the chromogenic and fluorescent ISH assays respectively. The mRNA expression of our gene markers was detected using the following murine probes: RNAscope® Probe – Mm-Col1a1 (CN: 319371), RNAscope® Probe – Mm-Clec3b C2 (CN: 300031-C2),

RNAscope® Probe – Mm-Spon1 (CN: 300031), RNAscope® Probe – Mm-Pdgfrβ (CN: 411381), RNAscope® Probe – Mm-Acta2 (CN: 319531), RNAscope® Probe – Mm-Fap (CN: 423881), RNAscope® Probe – Mm-Ccl7 (CN: 446821), RNAscope® Probe – Mm-Krt14-C2 (CN: 422521-C2), RNAscope® Probe – Mm-Cxcl12 (CN: 422711), RNAscope® Probe – Mm-Cxcl14 (CN: 459741), RNAscope® Probe – Mm-Cxcr4-C2 (CN: 425901-C2), RNAscope® Probe – Mm-Ccr1-C2 (CN: 402721-C2), RNAscope® Probe – Mm-Rgs5-C3 (CN: 430181-C3). All tissues treated for chromogenic ISH were counterstained using Mayer's Hematoxylin.

Data Deposition and Access.

Single-cell sequencing data generated in the course of this study will be available in NCBI GEO datasets upon publication. To review GEO accession GSE199515, go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE199515 and enter the token mravkaqurxejtin into the box.

Acknowledgements

This project was supported by funding from the NCI (R01 CA217206 to S. Powers) and the National Human Genome Research Institute (R21 HG009255 to S. Powers). The authors thank members of the Powers, Egeblad and Moll labs for discussions. The authors also thank the staff in the Research Histology Core and the Division of Laboratory Animal Resources (DLAR) in the Stony Brook School of Medicine for their technical support, as well as the staff in tTissue Analytics, Biostatistics & Data Science Shared Resources of the Stony Brook Cancer Center.

Author Contributions

SP, APD and ME conceived the study. A.P.D. performed the mouse experiments, sample/library preparation, histology staining and RNAscope analysis with assistance from AN, ME and UMM. ADP and SP performed computational analysis of the data and interpreted the results, with contributions from JL and JP. SP and APD codirected the study. NM provided critical tissue samples, SP wrote the first draft of the manuscript using text and figures prepared by APD. APD revised the first draft. All authors read and accepted the manuscript.

Figure Legends

Figure 1. scRNA-seq characterization of CAFs in the MMTV-PyMT model. (A) U-MAP projection of all clusters of MMTV-PyMT mammary tumors; (B) expression of signature markers within the three CAF subgroups; (C) heat map showing the top ten differentially expressed genes within each of the three CAF subtypes; (D) expression of canonical CAF markers in the three CAF subtypes.

Figure 2. Spatial localization and further molecular characterization of the three MMTV-PyMT CAF subtypes. (A) differential expression of genes for distinct collagen types and hyaluronic acid synthase; (B) images of differentially labeled ISH probes to *Spon1* and *Clec3b* in hematoxylin-stained MMTV-PyMT tumors; (C) quantification of the localization of *Spon1* and *Clec3b* signals to tumor and stromal compartments; (D) localization of *Spon1* and *Clec3b* signals in normal mammary tissue; (E) ISH localization of *Spon1* and *Clec3b* signals in tumors from C3-SV40-T-antigen transgenic mice; (F) ISH localization of *Spon1* and *Clec3b* signals in tumors from MMTV-Erbb2 transgenic mice; (G) localization of *Acta2* and *Spon1* signals in a MMTV-PyMT tumor.

Figure 3. Spatial localization of *Rgs5*$^{+}$ CAFs and integrated analysis of normal mammary tissue and three different mouse models of TNBC. (A) representative localization of *Pdgfrb* and *Spon1* signals in a MMTV-PyMT tumor; (B-C) fluorescent multiplexed ISH localization of *Spon1*, *Clec3b*, and *Rgs5* signals in MMTV-PyMT tumor tissue stained with DAPI to visualize tumor cells; (D) U-MAP projection of different cell types of normal mammary tissue; (E) bar graph showing similarity of with the fibroblast classifications based on > 20 normal tissues (from Buechler et al, 2021); (F) U-MAP projection of different cell types within the integrated dataset of normal mammary tissue and three different murine TNBC mammary tumor types; (G) bar graph showing the relative proportions of the three fibroblast subtypes as indicated; (H) violin plots showing expression of *Spon1*$^{+}$ CAF markers in *Spon1*$^{+}$ CAFs and corresponding normal cells.

Figure 4. Differential expression of CAF markers in normal mammary tissue versus MMTV-PyMT mammary tumors. (A) violin plots showing expression of *Clec3b*$^{+}$ CAF markers in *Clec3b*$^{+}$ CAFs and corresponding normal cells; (B) violin plots showing expression of *Rgs5*$^{+}$ CAF marker genes in *Rgs5*$^{+}$ CAFs and corresponding normal cells; (C-D) violin plots showing expression of canonical CAF markers in tumor CAF subtypes and corresponding normal cells; (E) dotplot showing expression of *Fap* in tumor CAF subtypes and corresponding normal cells; (F) dotplot showing both fold enrichment and FDR of cellular component GO terms in three MMTV-PyMT CAF subtypes relative to corresponding normal cells; (G) dotplot showing both fold enrichment and FDR of biological process GO terms in three MMTV-PyMT CAF subtypes relative to corresponding normal cells.

Figure 5. Reciprocal signaling of TGFβ and PDGF ligands/receptors along with extracellular matrix and metabolic genes in CAF development. (A) violin plots showing expression of extracellular matrix genes in MMTV-PyMT *Clec3b*$^{+}$ CAFs and corresponding normal cells; (B)

violin plots showing expression of mitochondrial respiratory genes n tumor CAF subtypes and corresponding normal cells; (C) violin plots showing expression of glycolyti genes in n tumor CAF subtypes and corresponding normal cells; (D) violin plots showing expression of TGFβ receptor genes in tumor CAF subtypes and corresponding normal cells; (E) heatmap showing the expression of TGFβ target genes in *Spon1*$^{+}$ CAFs; (F) dotplot showing the expression of indicated ligands in different cell types within the tumor microenvironment; (H) representative image of differentially labeled ISH probes to *Tgfbr2* and *Tgfb3* in hematoxylin-stained MMTV-PyMT tumor; (I) schemata of how TGFβ and PDGF ligands mediate the tumor microenvironment activation of CAFs in MMTV-PyMT tumors.

Figure 6. Conservation and divergence of human TNBC CAFs compared to murine TNBC CAFs. (A) UMAP showing clusters of indicated cell types from analysis of three human TNBCs; (B) expression of different markers in three human CAF subtypes; (C) dotplot showing expression of genes encoding hyaluronic acid synthase in CAFs; (D) piecharts showing for each human CAF subtype the percentage of expressed orthologs from different MMTV-PyMT CAF subtype markers; (E) violin plots showing expression of indicated genes in human TNBC CAF subtypes; (F) violin plots showing expression of indicated genes in MMTV-PyMT CAF subtypes; (G) UMAP showing subclusters of murine *Clec3b*$^{+}$ CAFs; (H) dotplot showing expression of indicated genes in *Clec3b*$^{+}$ subclusters; (I) violin plots showing expression of the indicated genes in normal and MMTV-PyMT *Clec3b*$^{+}$ fibroblasts.

Figure 7. Spatial localization of *Rgs5*$^{+}$ CAFs and further molecular characterization of the three MMTV-PyMT CAF subtypes. (A) Dot plots of individual cells within *Clec3b*$^{+}$ CAF subtype showing heterogeneity of *Clec3b* and *Ccl7* expression; (B) two representative images of

differentially labeled ISH probes to *Ccl7* and *Clec3b* in hematoxylin-stained MMTV-PyMT tumors; (C) violin plots showing expression of *Lrrc15* and *Mfap5* in normal and tumor murine fibroblasts; (D) dotplot showing low level expression of *LRRC15* and *MFAP5* in human TNBC CAFs; (E) dotplot showing the expression of eighteen candidate CAF targets in MMTV-PyMT CAFs; (F) dotplot showing the expression of seven candidate CAF targets in human TNBC CAFs; (G) expression of final five CAF targets in human TNBC CAF subtypes.

Supplementary Figure 1. CAF subclusters. (A) UMAP showing subclusters of MMTV-PyMT *Rgs5*$^+$ CAFs; (B) dotplot showing the expression of indicated genes in *Rgs5*$^+$ CAF subclusters; (C) UMAP showing subclusters of human *RGS5*$^+$ CAFs; (D) dotplot showing the expression of indicated genes in human *RGS5*$^+$ CAF subclusters; (E) UMAP showing subclusters of MMTV-PyMT *Spon1*$^+$ CAFs; (F) dotplot showing the expression of indicated genes in *Spon1*$^+$ CAF subclusters; (G) UMAP showing subclusters of human *COL1A1*$^+$ CAFs; (H) dotplot showing the expression of indicated genes in human *COL1A1*$^+$ subclusters; (I) UMAP showing subclusters of human *COL3A1*$^+$ CAFs; (J) dotplot showing the expression of indicated genes in human *COL3A1*$^+$ subclusters.

Supplementary Figure 2. Expression of potential CAF targets. (A) dotplot showing the expression of indicated genes in different MMTV-PyMT tumor cell types; (B) dotplot showing the expression of indicated genes in human CAF subtypes; (C) dotplot showing the expression of indicated genes in human TNBC cell types.

References

Arina, A., Idel, C., Hyjek, E. M., Alegre, M. L., Wang, Y., Bindokas, V. P., Weichselbaum, R. R., and Schreiber, H. (2016). Tumor-associated fibroblasts predominantly come from local and not circulating precursors. Proc Natl Acad Sci U S A *113*, 7551-7556.

Bartoschek, M., Oskolkov, N., Bocci, M., Lovrot, 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.

Brechbuhl, H. M., Finlay-Schultz, J., Yamamoto, T. M., Gillen, A. E., Cittelly, D. M., Tan, A. C., Sams, S. B., Pillai, M. M., Elias, A. D., Robinson, W. A.*, et al.* (2017). Fibroblast Subtypes Regulate Responsiveness of Luminal Breast Cancer to Estrogen. Clin Cancer Res *23*, 1710-1721.

Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods *17*, 159-162.

Buechler, M. B., Pradhan, R. N., Krishnamurty, A. T., Cox, C., Calviello, A. K., Wang, A. W., Yang, Y. A., Tam, L., Caothien, R., Roose-Girma, M.*, et al.* (2021). Cross-tissue organization of the fibroblast lineage. Nature *593*, 575-579.

Chen, Y., McAndrews, K. M., and Kalluri, R. (2021). Clinical and therapeutic relevance of cancer-associated fibroblasts. Nat Rev Clin Oncol *18*, 792-804.

Davidson, S., Efremova, M., Riedel, A., Mahata, B., Pramanik, J., Huuhtanen, J., Kar, G., Vento-Tormo, R., Hagai, T., Chen, X.*, et al.* (2020). Single-Cell RNA Sequencing Reveals a Dynamic Stromal Niche That Supports Tumor Growth. Cell Rep *31*, 107628.

Dominguez, C. X., Muller, S., Keerthivasan, S., Koeppen, H., Hung, J., Gierke, S., Breart, B., Foreman, O., Bainbridge, T. W., Castiglioni, A.*, et al.* (2020). Single-Cell RNA Sequencing Reveals Stromal Evolution into LRRC15(+) Myofibroblasts as a Determinant of Patient Response to Cancer Immunotherapy. Cancer Discov *10*, 232-253.

Elyada, E., Bolisetty, M., Laise, P., Flynn, W. F., Courtois, E. T., Burkhart, R. A., Teinor, J. A., Belleau, P., Biffi, G., Lucito, M. S.*, et al.* (2019). Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. Cancer Discov *9*, 1102-1123.

Friedman, G., Levi-Galibov, O., David, E., Bornstein, C., Giladi, A., Dadiani, M., Mayo, A., Halperin, C., Pevsner-Fischer, M., Lavon, H.*, et al.* (2020). Cancer-associated fibroblast compositions change with breast cancer progression linking the ratio of S100A4(+) and PDPN(+) CAFs to clinical outcome. Nat Cancer *1*, 692-708.

Hamson, E. J., Keane, F. M., Tholen, S., Schilling, O., and Gorrell, M. D. (2014). Understanding fibroblast activation protein (FAP): substrates, activities, expression and targeting for cancer therapy. Proteomics Clin Appl *8*, 454-463.

Hu, H., Piotrowska, Z., Hare, P. J., Chen, H., Mulvey, H. E., Mayfield, A., Noeen, S., Kattermann, K., Greenberg, M., Williams, A.*, et al.* (2021). Three subtypes of lung cancer fibroblasts define distinct therapeutic paradigms. Cancer Cell *39*, 1531-1547 e1510.

Kanzaki, R., and Pietras, K. (2020). Heterogeneity of cancer-associated fibroblasts: Opportunities for precision medicine. Cancer Sci *111*, 2708-2717.

Kobayashi, H., Gieniec, K. A., Lannagan, T. R. M., Wang, T., Asai, N., Mizutani, Y., Iida, T., Ando, R., Thomas, E. M., Sakai, A.*, et al.* (2022). The Origin and Contribution of Cancer-Associated Fibroblasts in Colorectal Carcinogenesis. Gastroenterology *162*, 890-906.

Mayne, R., and Burgeson, R. E. (1987). Structure and function of collagen types, (Orlando: Academic Press).

Miljkovic-Licina, M., Hammel, P., Garrido-Urbani, S., Lee, B. P., Meguenani, M., Chaabane, C., Bochaton-Piallat, M. L., and Imhof, B. A. (2012). Targeting olfactomedin-like 3 inhibits tumor growth by impairing angiogenesis and pericyte coverage. Mol Cancer Ther *11*, 2588-2599.

Peran, I., Dakshanamurthy, S., McCoy, M. D., Mavropoulos, A., Allo, B., Sebastian, A., Hum, N. R., Sprague, S. C., Martin, K. A., Pishvaian, M. J.*, et al.* (2021). Cadherin 11 Promotes

Immunosuppression and Extracellular Matrix Deposition to Support Growth of Pancreatic Tumors and Resistance to Gemcitabine in Mice. Gastroenterology *160*, 1359-1372 e1313.

Plasko, G. R. (2021). The Functional Role of Tetranectin in Overnutrition-Induced Obesity and Hepatosteatosis in Female Mice. In Doctoral dissertation, T.U.o.T.H.S.C.a.S. Antonio, ed.

Purcell, J. W., Tanlimco, S. G., Hickson, J., Fox, M., Sho, M., Durkin, L., Uziel, T., Powers, R., Foster, K., McGonigal, T.*, et al.* (2018). LRRC15 Is a Novel Mesenchymal Protein and Stromal Target for Antibody-Drug Conjugates. Cancer Res *78*, 4059-4072.

Sahai, E., Astsaturov, I., Cukierman, E., DeNardo, D. G., Egeblad, M., Evans, R. M., Fearon, D., Greten, F. R., Hingorani, S. R., Hunter, T.*, et al.* (2020). A framework for advancing our understanding of cancer-associated fibroblasts. Nat Rev Cancer *20*, 174-186.

Valdes-Mora, F., Salomon, R., Gloss, B. S., Law, A. M. K., Venhuizen, J., Castillo, L., Murphy, K. J., Magenau, A., Papanicolaou, M., Rodriguez de la Fuente, L.*, et al.* (2021). Single-cell transcriptomics reveals involution mimicry during the specification of the basal breast cancer subtype. Cell Rep *35*, 108945.

Wang, H., Xiang, D., Liu, B., He, A., Randle, H. J., Zhang, K. X., Dongre, A., Sachs, N., Clark, A. P., Tao, L.*, et al.* (2019). Inadequate DNA Damage Repair Promotes Mammary Transdifferentiation, Leading to BRCA1 Breast Cancer. Cell *178*, 135-151 e119.

Wu, S. Z., Roden, D. L., Wang, C., Holliday, H., Harvey, K., Cazet, A. S., Murphy, K. J., Pereira, B., Al-Eryani, G., Bartonicek, N.*, et al.* (2020). Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. EMBO J *39*, e104063.

Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., Lachmann, A., Wojciechowicz, M. L., Kropiwnicki, E., Jagodnik, K. M.*, et al.* (2021). Gene Set Knowledge Discovery with Enrichr. Curr Protoc *1*, e90.

Yeung, T. L., Leung, C. S., Yip, K. P., Sheng, J., Vien, L., Bover, L. C., Birrer, M. J., Wong, S. T. C., and Mok, S. C. (2019). Anticancer Immunotherapy by MFAP5 Blockade Inhibits Fibrosis and Enhances Chemosensitivity in Ovarian and Pancreatic Cancer. Clin Cancer Res *25*, 6417-6428.

Figure 1

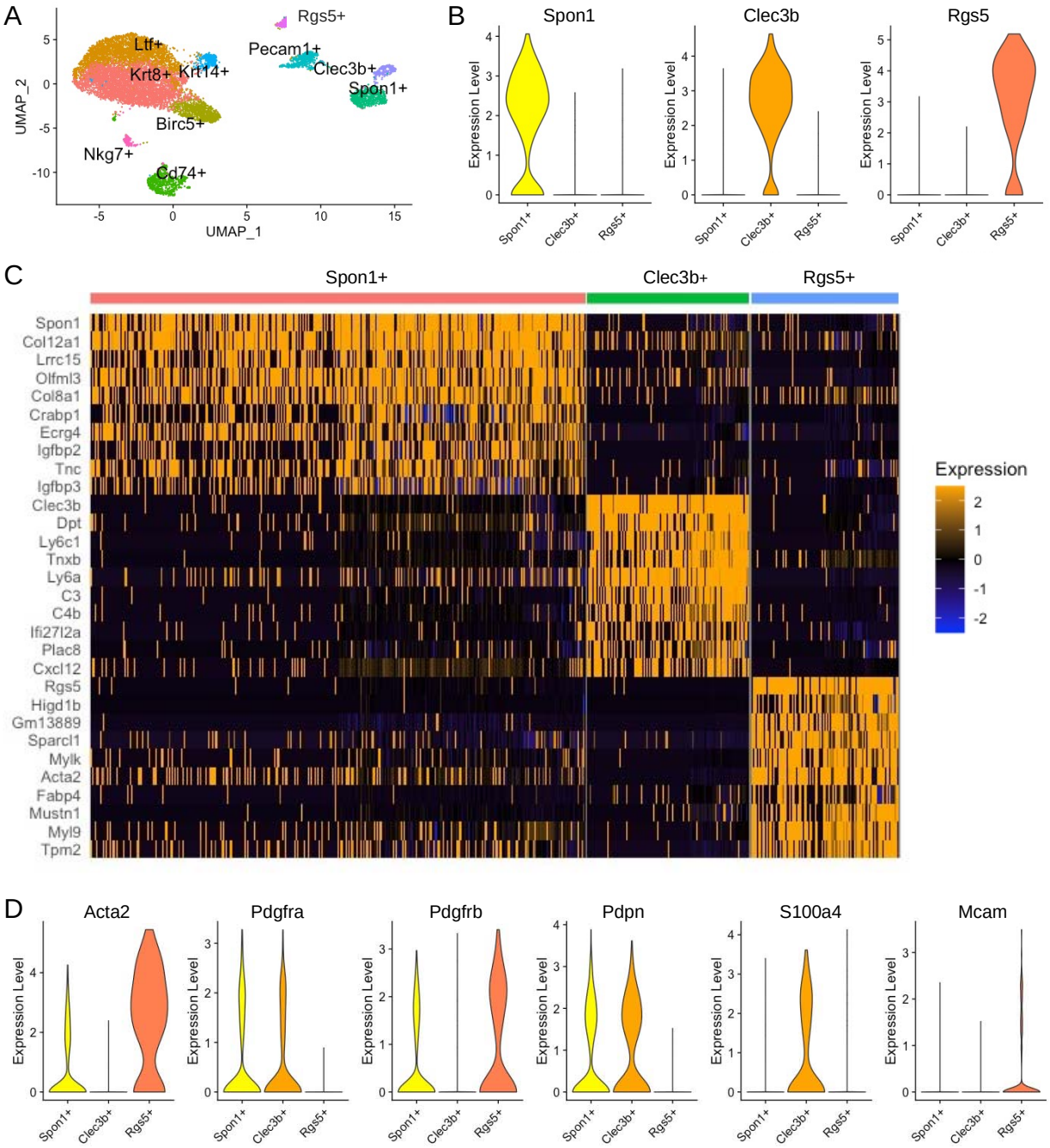

A
Rgs5+
Ltf+
Pecam1+
Krt8+
Krt14+
Clec3b+
Spon1+
Birc5+
Nkg7+
Cd74+
UMAP_2
UMAP_1
B
Spon1
Clec3b
Rgs5
Expression Level
Spon1+
Clec3b+
Rgs5+
C
Spon1+
Clec3b+
Rgs5+
Spon1
Col12a1
Lrrc15
Olfml3
Col8a1
Crabp1
Ecrg4
Igfbp2
Tnc
Igfbp3
Clec3b
Dpt
Ly6c1
Tnxb
Ly6a
C3
C4b
Ifi27l2a
Plac8
Cxcl12
Rgs5
Higd1b
Gm13889
Sparcl1
Mylk
Acta2
Fabp4
Mustn1
Myl9
Tpm2
Expression
D
Acta2
Pdgfra
Pdgfrb
Pdpn
S100a4
Mcam
Expression Level
Spon1+
Clec3b+
Rgs5+

Figure 2

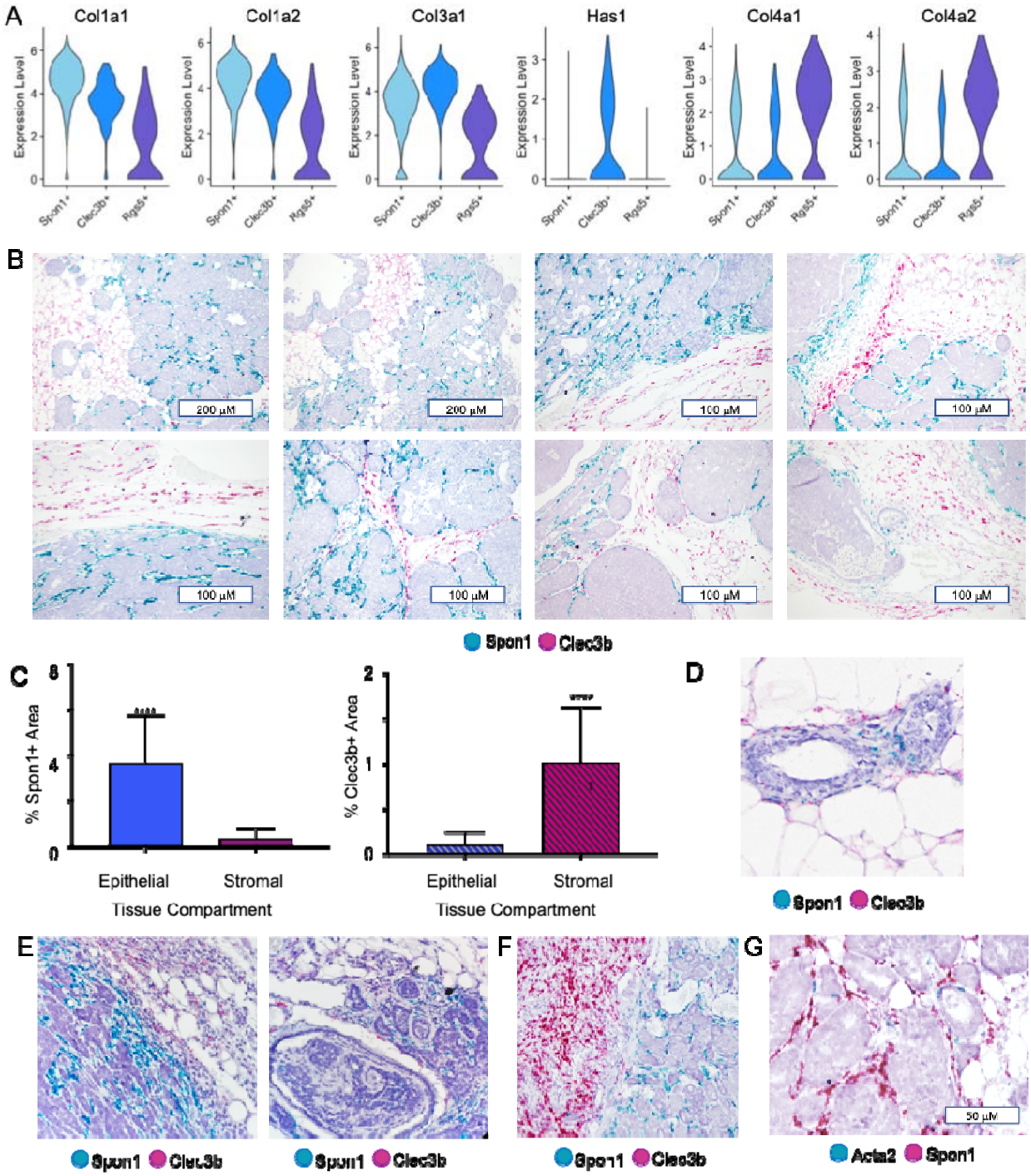

A
Col1a1
Col1a2
Col3a1
Has1
Col4a1
Col4a2
Expression Level
Spon1+
Cloc3b+
Rgs5+
B
200 µM
100 µM
Spon1
Cloc3b
C
% Spon1+ Area
% Cloc3b+ Area
Epithelial
Stromal
Tissue Compartment
D
Spon1
Cloc3b
E
Spon1
Cloc3b
F
Spon1
Cloc3b
G
50 µM
Acta2
Spon1

Figure 3

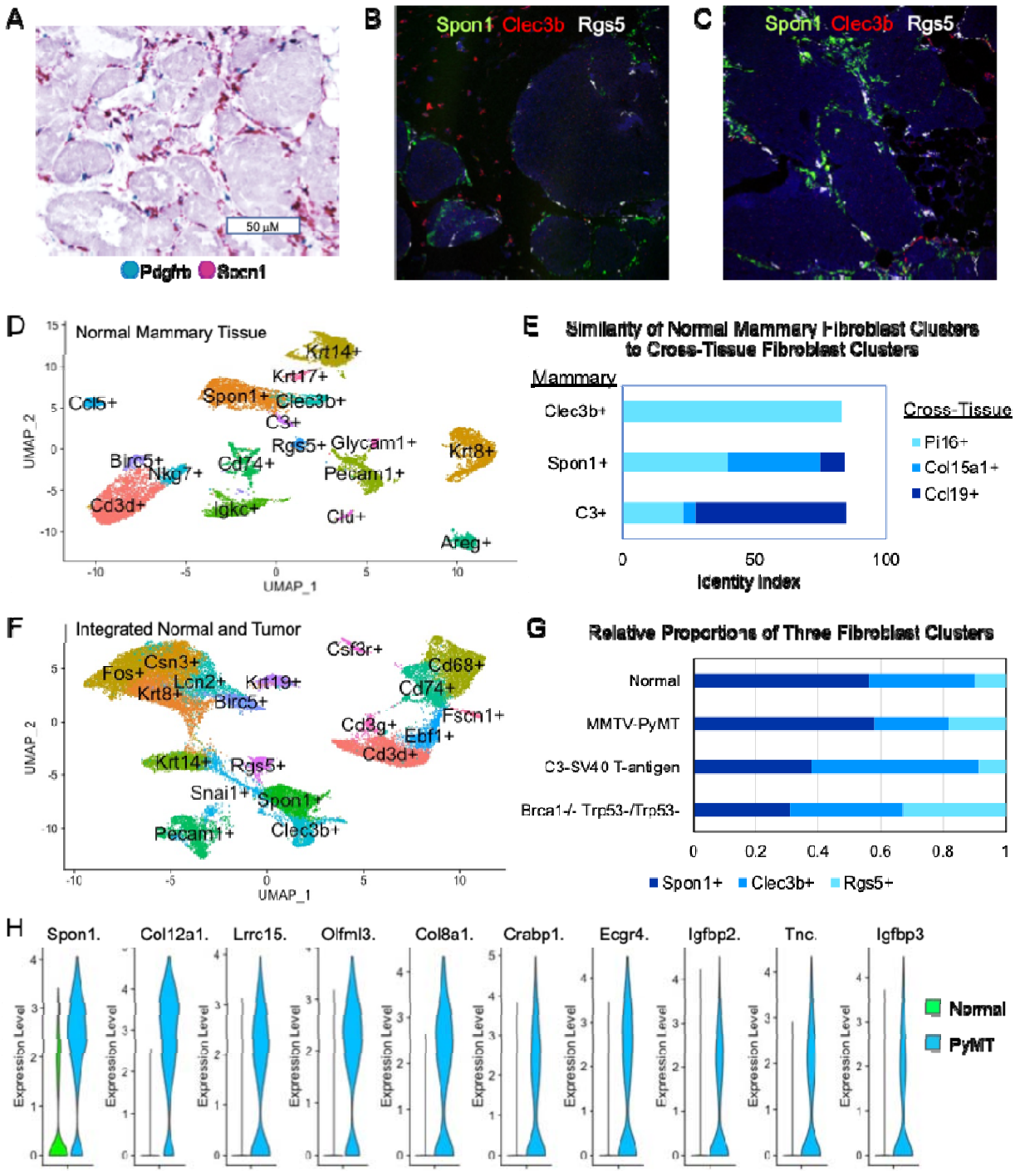

A
50 µM
Pdgfrb
Spon1
B
Spon1 Clec3b Rgs5
C
Spon1 Clec3b Rgs5
D
Normal Mammary Tissue
Krt14+
Krt17+
Ccl5+
Spon1+
Clec3b+
C3+
Rgs5+
Glycam1+
Krt8+
Birc5+
Nkg7+
Cd74+
Pecam1+
Cd3d+
Igkc+
Clu+
Areg+
UMAP_2
UMAP_1
E
Similarity of Normal Mammary Fibroblast Clusters to Cross-Tissue Fibroblast Clusters
Mammary
Clec3b+
Spon1+
C3+
Cross-Tissue
Pi16+
Col15a1+
Ccl19+
0
50
100
Identity Index
F
Integrated Normal and Tumor
Csf3r+
Csn3+
Fos+
Lcn2+
Krt19+
Krt8+
Birc5+
Cd68+
Cd74+
Fscn1+
Cd3g+
Ebf1+
Cd3d+
Krt14+
Rgs5+
Snai1+
Spon1+
Pecam1+
Clec3b+
UMAP_2
UMAP_1
G
Relative Proportions of Three Fibroblast Clusters
Normal
MMTV-PyMT
C3-SV40 T-antigen
Brca1-/- Trp53-/Trp53-
0
0.2
0.4
0.6
0.8
1
Spon1+
Clec3b+
Rgs5+
H
Spon1.
Col12a1.
Lrrc15.
Olfml3.
Col8a1.
Crabp1.
Ecgr4.
Igfbp2.
Tnc.
Igfbp3
Expression Level
Normal
PyMT

Figure 4

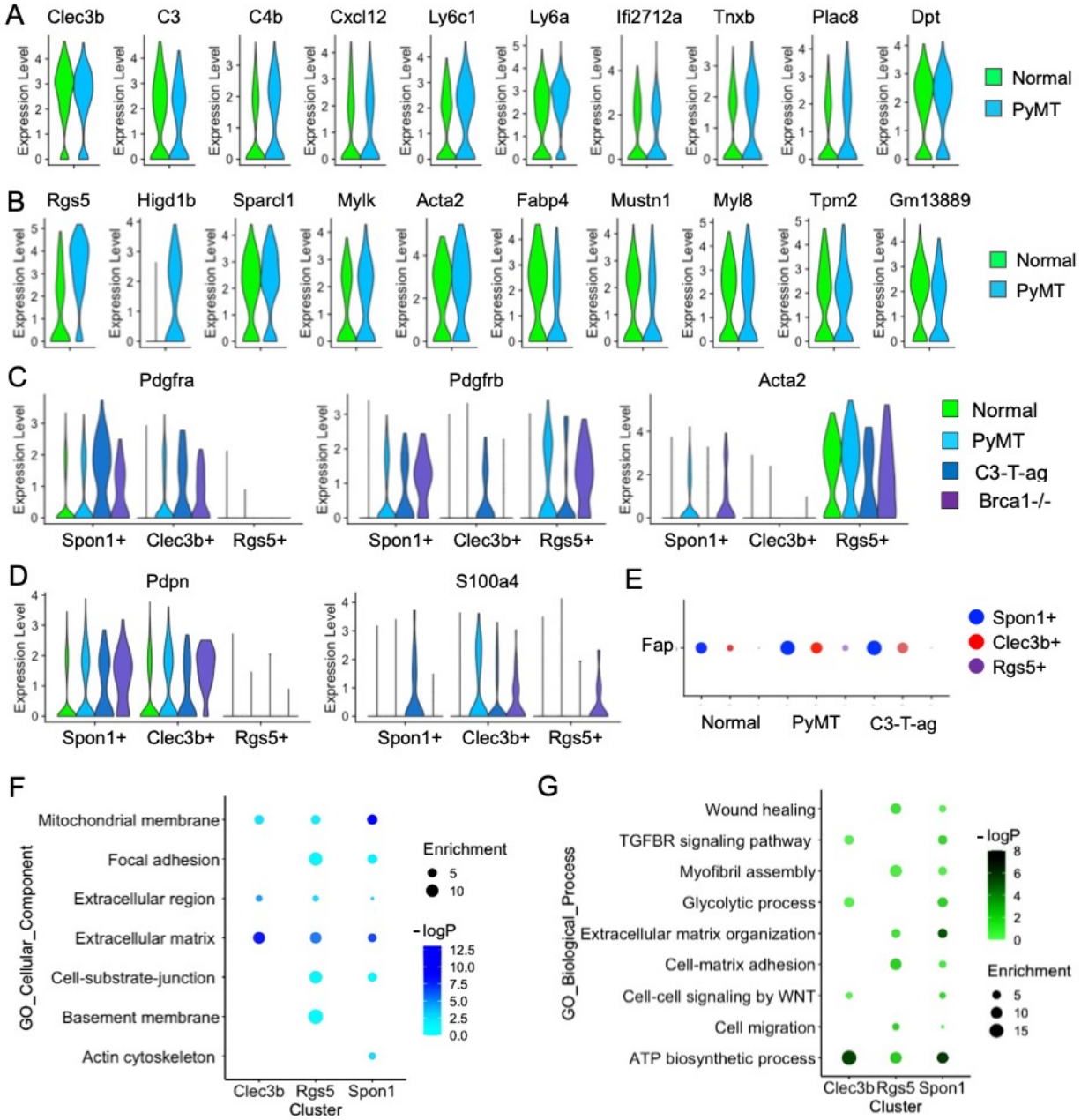

A
Clec3b
C3
C4b
Cxcl12
Ly6c1
Ly6a
Ifi2712a
Tnxb
Plac8
Dpt
Expression Level
Normal
PyMT
B
Rgs5
Higd1b
Sparcl1
Mylk
Acta2
Fabp4
Mustn1
Myl8
Tpm2
Gm13889
Expression Level
Normal
PyMT
C
Pdgfra
Pdgfrb
Acta2
Expression Level
Spon1+
Clec3b+
Rgs5+
Normal
PyMT
C3-T-ag
Brca1-/-
D
Pdpn
S100a4
Expression Level
Spon1+
Clec3b+
Rgs5+
E
Fap
Normal
PyMT
C3-T-ag
Spon1+
Clec3b+
Rgs5+
F
GO_Cellular_Component
Mitochondrial membrane
Focal adhesion
Extracellular region
Extracellular matrix
Cell-substrate-junction
Basement membrane
Actin cytoskeleton
Clec3b
Rgs5
Spon1
Cluster
Enrichment
5
10
-logP
12.5
10.0
7.5
5.0
2.5
0.0
G
GO_Biological_Process
Wound healing
TGFBR signaling pathway
Myofibril assembly
Glycolytic process
Extracellular matrix organization
Cell-matrix adhesion
Cell-cell signaling by WNT
Cell migration
ATP biosynthetic process
Clec3b
Rgs5
Spon1
Cluster
-logP
8
6
4
2
0
Enrichment
5
10
15

Figure 5

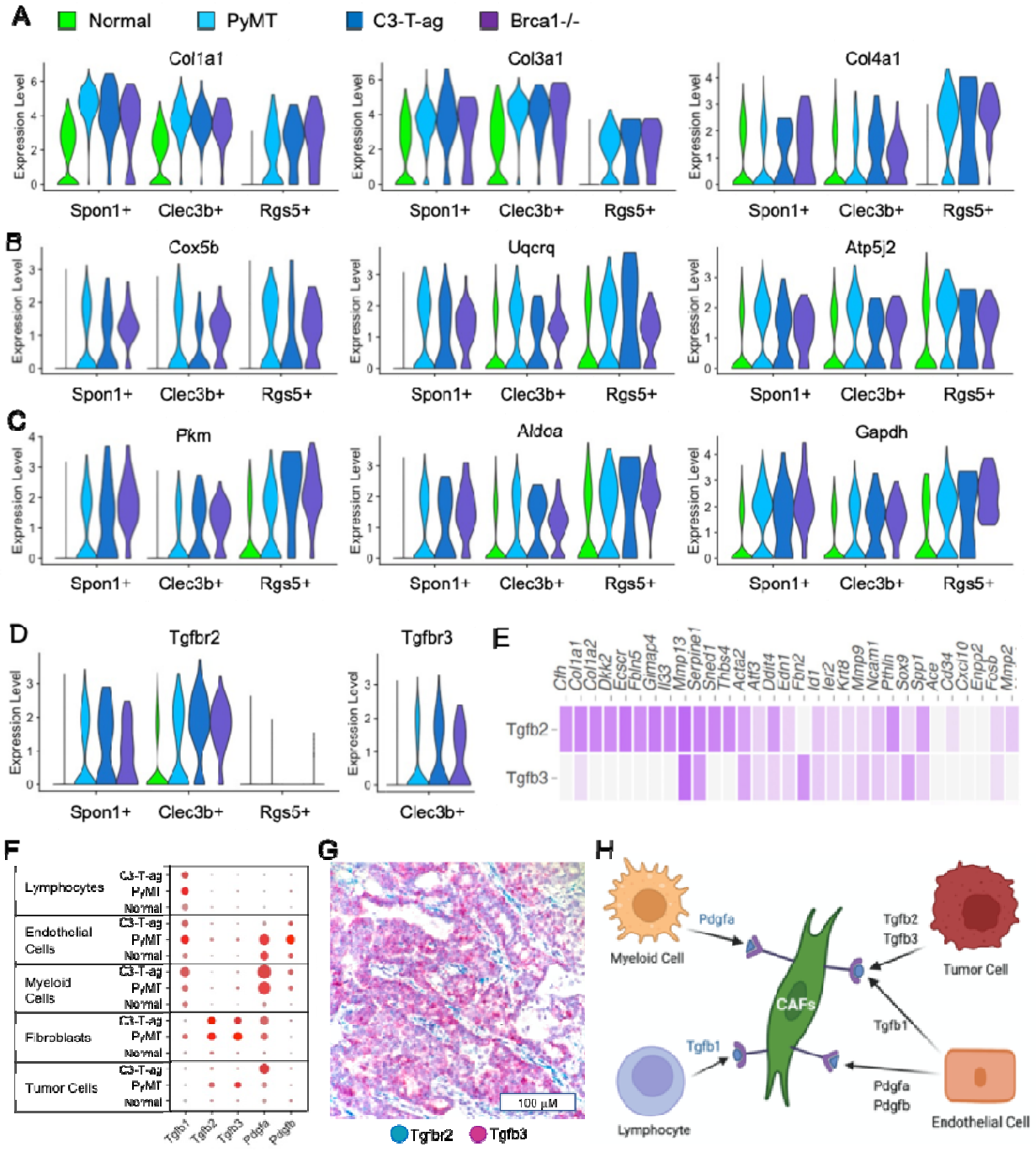

A
Normal
PyMT
C3-T-ag
Brca1-/-
Col1a1
Col3a1
Col4a1
Expression Level
Spon1+
Clec3b+
Rgs5+
B
Cox5b
Uqcrq
Atp5j2
C
Pkm
Aldoa
Gapdh
D
Tgfbr2
Tgfbr3
E
Cfh
Col1a1
Col1a2
Dkk2
Ecscr
Fbln5
Gimap4
Il33
Mmp13
Serpine1
Sned1
Thbs4
Acta2
Atf3
Ddit4
Edn1
Fbn2
Id1
Ier2
Krt8
Mmp9
Ncam1
Prhln
Sox9
Spp1
Ace
Cd34
Cxcl10
Enpp2
Fosb
Mmp2
Tgfb2
Tgfb3
F
Lymphocytes
Endothelial Cells
Myeloid Cells
Fibroblasts
Tumor Cells
C3-T-ag
PyMT
Normal
Tgfb1
Tgfb2
Tgfb3
Pdgfa
Pdgfb
G
100 µM
Tgfbr2
Tgfb3
H
Myeloid Cell
Pdgfa
Tgfb2
Tgfb3
Tumor Cell
CAFs
Tgfb1
Tgfb1
Lymphocyte
Pdgfa
Pdgfb
Endothelial Cell

Figure 6

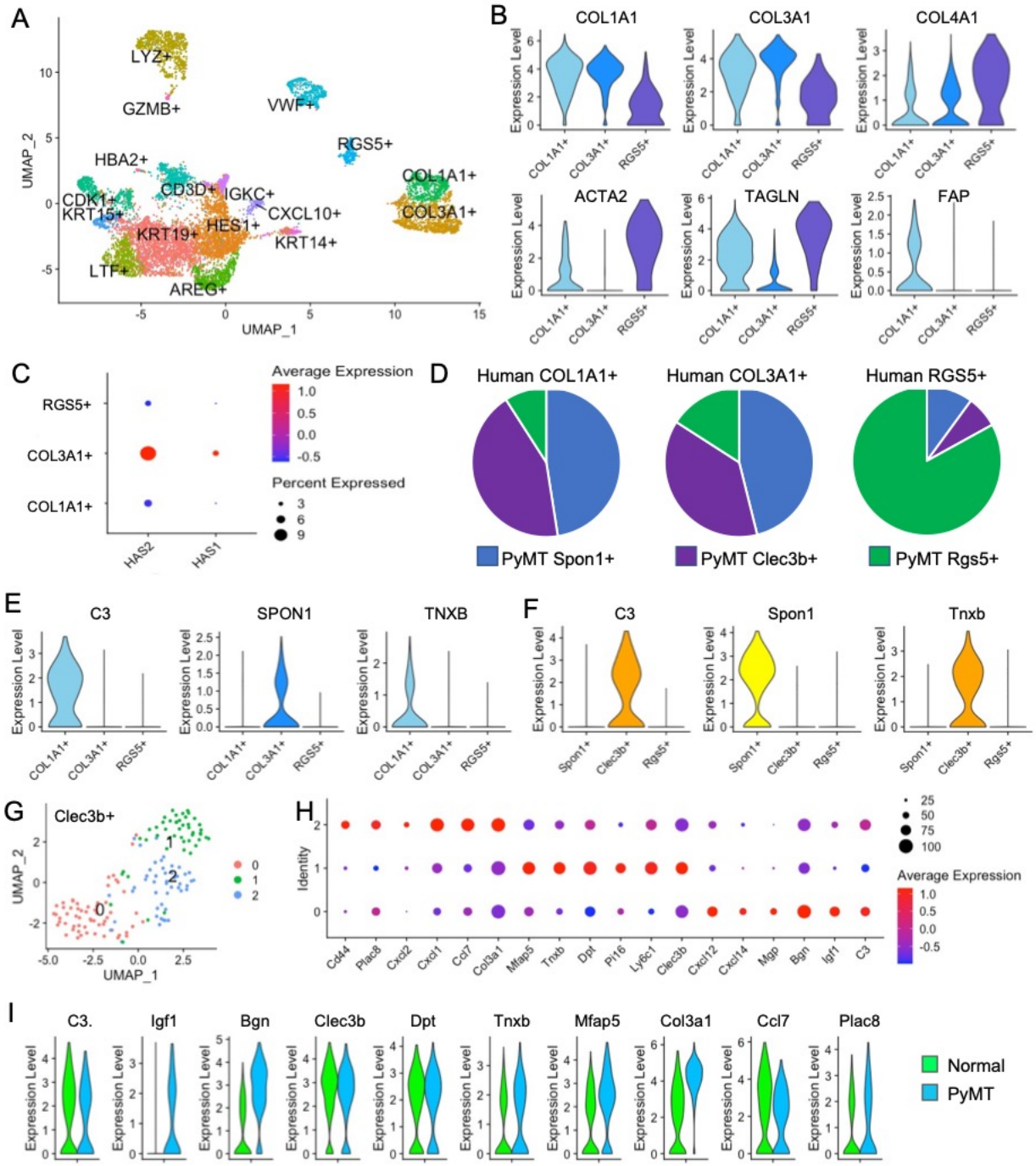
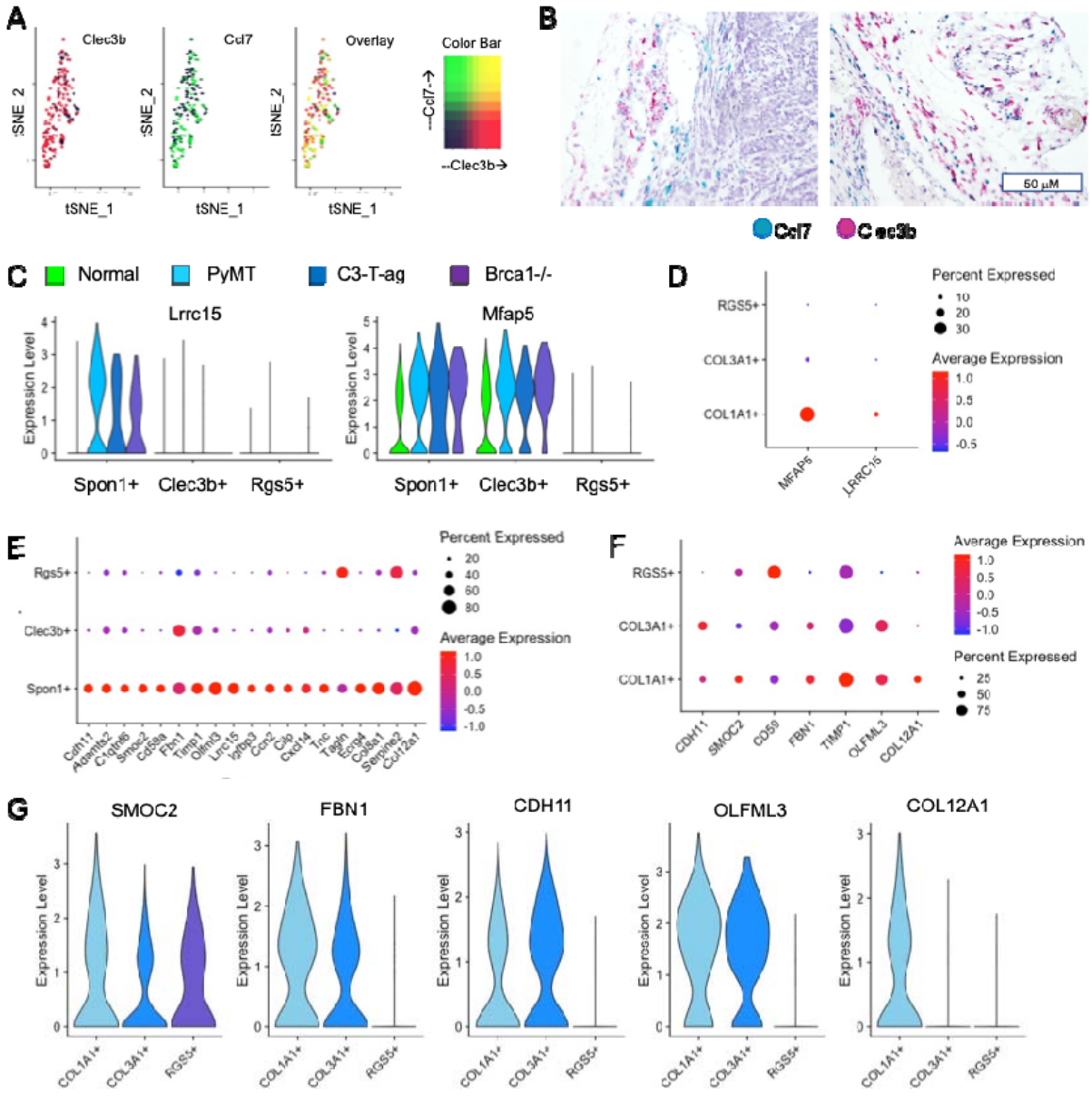

A
LYZ+
GZMB+
VWF+
RGS5+
HBA2+
COL1A1+
CD3D+
IGKC+
CDK1+
KRT15+
CXCL10+
COL3A1+
HES1+
KRT19+
KRT14+
LTF+
AREG+
UMAP_1
UMAP_2
B
COL1A1
COL3A1
COL4A1
ACTA2
TAGLN
FAP
Expression Level
C
RGS5+
COL3A1+
COL1A1+
HAS2
HAS1
Average Expression
Percent Expressed
D
Human COL1A1+
Human COL3A1+
Human RGS5+
PyMT Spon1+
PyMT Clec3b+
PyMT Rgs5+
E
C3
SPON1
TNXB
F
C3
Spon1
Tnxb
G
Clec3b+
UMAP_1
UMAP_2
H
Identity
Cd44
Plac8
Cxcl2
Cxcl1
Ccl7
Col3a1
Mfap5
Tnxb
Dpt
Pi16
Ly6c1
Clec3b
Cxcl12
Cxcl14
Mgp
Bgn
Igf1
C3
Average Expression
I
C3.
Igf1
Bgn
Clec3b
Dpt
Tnxb
Mfap5
Col3a1
Ccl7
Plac8
Normal
PyMT

Figure 7


A
Clec3b
Ccl7
Overlay
Color Bar
tSNE_1
tSNE_2
B
50 µM
Ccl7
Clec3b
C
Normal
PyMT
C3-T-ag
Brca1-/-
Lrrc15
Mfap5
Expression Level
Spon1+
Clec3b+
Rgs5+
D
Percent Expressed
Average Expression
RGS5+
COL3A1+
COL1A1+
MFAP5
LRRC15
E
Percent Expressed
Average Expression
Rgs5+
Clec3b+
Spon1+
F
Average Expression
Percent Expressed
RGS5+
COL3A1+
COL1A1+
CDH11
SMOC2
CD59
FBN1
TIMP1
OLFML3
COL12A1
G
SMOC2
FBN1
CDH11
OLFML3
COL12A1
Expression Level
COL1A1+
COL3A1+
RGS5+