Stromal cell diversity associated with immune evasion in human triple-negative breast cancer

Sunny Z Wu1,2, Daniel L Roden1,2, Chenfei Wang3, Holly Holliday1,2, Kate Harvey3, Aurélie S Cazet1,2, Kendelle J Murphy1,2, Brooke Pereira1,2, Ghamdan Al-Eryan1,2, Nenad Bartonicek1,2, Rui Hou4, James R Torpy1,2, Simon Junanka1,2, Chia-Ling Chan5, Chuan En Lam5, Mun N Hui1,6, Laurence Gluch7, Jane Beith6, Andrew Parker8, Elizabeth Robbins9, Davendra Segara9, Cindy Mak6, Caroline Cooper10,11, Sanjay Warrier12,13, Alistair Forrest4,14, Joseph Powell5,15, Sandra O'Toole1,2,16, Thomas R Cox1,2, Paul Timpson1,2, Elgene Lim1,2,8, X Shirley Liu3 & Alexander Swarbrick1,2,*

Abstract

The tumour stroma regulates nearly all stages of carcinogenesis. Stromal heterogeneity in human triple-negative breast cancers (TNBCs) remains poorly understood, limiting the development of stromal-targeted therapies. Single-cell RNA sequencing of five TNBCs revealed two cancer-associated fibroblast (CAF) and two perivascular-like (PVL) subpopulations. CAFs clustered into two states: the first with features of myofibroblasts and the second characterised by high expression of growth factors and immunomodulatory molecules. PVL cells clustered into two states consistent with a differentiated and immature phenotype. We showed that these stromal states have distinct morphologies, spatial relationships and functional properties in regulating the extracellular matrix. Using cell signalling predictions, we provide evidence that stromal-immune crosstalk acts via a diverse array of immunoregulatory molecules. Importantly, the investigation of gene signatures from inflammatory-CAFs and differentiated-PVL cells in independent TNBC patient cohorts revealed strong associations with cytotoxic T-cell dysfunction and exclusion, respectively. Such insights present promising candidates to further investigate for new therapeutic strategies in the treatment of TNBCs.

Subject Categories
Cancer; Immunology

Keywords
- cancer-associated fibroblasts; single-cell RNA sequencing; stromal heterogeneity; triple-negative breast cancer; tumour microenvironment

Introduction

Heterotypic interactions between stromal, immune and malignant epithelial cells play important roles in solid tumour progression and therapeutic response. Cancer-associated fibroblasts (CAFs) play an integral part in the tumour microenvironment (TME) and can influence many aspects of carcinogenesis including extracellular matrix (ECM) remodelling, angiogenesis, cancer cell proliferation, invasion, inflammation, metabolic reprogramming and metastasis (Kalluri, 2016). Recent studies have described roles for CAFs in mediating immune suppression and chemoresistance, establishing CAFs as novel and attractive targets for anti-cancer therapies in advanced breast cancer (Brechbuhl et al, 2017; Cazet et al, 2018; Costa et al, 2018; Givel et al, 2018; Su et al, 2018). Despite their well-described roles in cancer biology, CAFs remain enigmatic: limited studies suggest phenotypic heterogeneity, plasticity and functional...
CAs of the tumour parenchyma are routinely studied using a handful of markers including α-smooth muscle actin (α-SMA), fibroblast activation protein (FAP), CD90 (THY-1), platelet-derived growth factor receptor α and β (PDGFRα and PDGFRβ), podoplanin (PDPN) and fibroblast-specific protein 1 (FSP-1, also named S100A4) (Kawase et al., 2008; Kisselbach et al., 2009; Neri et al., 2015; Kalluri, 2016). However, these markers are not necessarily co-expressed, nor specific to the fibroblastic lineage (Costa et al., 2018). For instance, α-SMA not only identifies CAs with a myofibroblast morphology but also serves as a general marker for myoepithelial cells and perivascular cells. α-SMA+ cells in the breast tumour stroma can also arise from different mesenchymal lineages including resident fibroblasts, smooth muscle cells and pericytes (Ronnov-Jessen et al., 1995). In addition, FSP-1 is also expressed in macrophages, other immune cells and even cancer cells (Osterreicher et al., 2011). Thus, a categorical definition of cancer-associated stromal cells and specific cell surface markers remains challenging and is urgently needed (Kalluri, 2016).

Three broad CAF subtypes have been recently profiled in mouse models of pancreatic ductal adenocarcinoma (PDAC) (Ohlund et al., 2017; Biffi et al., 2018; Elyada et al., 2019). These are characterised by a myofibroblast-like (myCAFs) phenotype, inflammatory properties (iCAFs) and antigen-presenting capabilities (apCAFs) (Ohlund et al., 2017; Biffi et al., 2018; Elyada et al., 2019). Although little is known about the mechanistic role and clinical relevance of iCAFs and apCAFs, an accumulation of the myCAF marker α-SMA has been shown to correlate with poor outcome in breast and pancreatic cancer (Yamashita et al., 2012; Sinn et al., 2014). We have shown that targeting Hedgehog-activated CAs, which have a myofibroblast-like phenotype in ECM regulation, results in markedly improved survival, chemosensitivity and reduced metastatic burden in pre-clinical models of TNBC (Cazet et al., 2018). In addition, myofibroblast-like CAs have been shown to contribute to an immunosuppressive microenvironment by attracting T-regulatory cells in breast and ovarian cancer (Costa et al., 2018; Givel et al., 2018). While these studies point towards the therapeutic targeting of myofibroblast-like CAs, genetic ablation of α-SMA+ cells in a mouse model of PDAC resulted in more aggressive tumours and reduced mouse overall survival, indicating complex stromal functionalities across distinct tissue sites (Ozdemir et al., 2015).

Recent advances in single-cell RNA sequencing (scRNA-Seq) have overcome some of the technical hurdles in the investigation of cellular heterogeneity amongst complex tissues such as carcinomas. Recent patient studies have dissected the TME in head and neck squamous cell carcinomas and lung tumours, revealing new insights into stromal and immune subsets associated with disease progression (Puram et al., 2017; Lambrecht et al., 2018). Single-cell studies of human breast cancers have been limited to immune cells, while studies in mouse models have revealed four subclasses of CAs (Bartoschek et al., 2018). Although CAs from human breast carcinomas have been profiled by flow cytometry and bulk sequencing, comprehensive single-cell profiling has yet to be performed in TNBC patients (Costa et al., 2018).

TNBC is an aggressive breast cancer subtype, which is lacking in effective targeted therapeutic options. It is clinically defined by negative status for targetable hormone receptors (oestrogen receptor and progesterone receptor) or HER2 amplification. Studies in mice and humans have demonstrated that TNBC progression can be influenced by stromal cells; however, a comprehensive understanding of the stromal hierarchy is yet to be established (Brechbuhle et al., 2017; Cazet et al., 2018; Costa et al., 2018; Givel et al., 2018; Su et al., 2018). To investigate this in more detail, we performed unbiased high-throughput scRNA-Seq to profile the TME directly in patient tumour tissues. In addition to CAs, we identified stromal cells with a perivascular-like (PVL) profile, which were not necessarily associated with blood vessels. Our study focuses exclusively on CAs and PVL cells, which we collectively refer to as “stroma”. Using orthogonal methods, we found that functions previously ascribed to CAs as unitary cell types are actually performed by specialised subsets of stromal cells with distinct morphological, spatial and functional properties (Bartoschek et al., 2018). In addition, by sampling cells from the entire TME, we were able to predict paracrine signalling between stromal and immune cell subsets. From this, we analysed large patient gene expression datasets to show significant association between inflammatory-like CAs and differentiated-PVL cells with immune evasion. Our human TNBC single-cell datasets provide a new taxonomy of human cancer-associated stromal cells, which we envisage can be used to further develop TME-directed therapies.

Results

Composition of triple-negative breast cancers at cellular resolution

We performed scRNA-Seq on primary breast tumours collected from five patients (Fig EV1A and B) using a marker-free approach. Fresh tissues were dissociated into single-cell suspensions prior to single-cell capture on the Chromium controller (10x Genomics) and sequencing on the NextSeq 500 (Illumina) (Figs 1A and EV1C). In total, we sequenced 24,271 cells, with an average of 4,854 cells per patient (Fig EV1D). A total of 28,118 genes were detected with an average of 1,658 genes expressed, and 6,215 unique molecular identifiers (UMIs) detected per cell (Fig EV1E–H). Data from individual tumours were integrated and clustered using canonical correlation analysis (CCA) in Seurat (Satija et al., 2015).

Epithelial cells (Fig 1B and C) and stromal-immune cells (Fig 1D and E) were first annotated through the expression of canonical cell type gene markers. This revealed four major cell states within the epithelial compartment (Fig 1B and C), including a major cluster of 4,095 cancer cells (16.9% of all cells; EPCAM+, ESR1−) and a second cluster of 614 cancer cells with high proliferation (2.5%; MKI67+). The remaining two smaller epithelial clusters had gene expression features consistent with normal luminal (277 cells, 0.9%; EPCAM+, ESR1+) and myoepithelial cells (212 cells, 0.9%; EPCAM+; KRT5+, KRT14+ and ACTA2−). Neoplastic or normal status of these cell clusters was confirmed by inferring genome copy number alterations over large genomic regions using InferCNV (Appendix Fig S1) (Patel et al., 2014). In addition to marker genes,
stromal and immune clusters were further classified through scoring against published cell type signatures from the XCell database with an area under the curve approach (AUCell) (Fig EV2A; Aibar et al., 2017; Aran et al., 2017). In the immune compartment (Fig 1D and E), we identified 7,990 T lymphocytes (32.9%; CD3D), 1,245 B cells (5.1%; MS4A1), 1,955 plasma cells (8.1%; JCHAIN) and 4,606 myeloid cells (19.0%; CD68). Through re-clustering of the T lymphocytes (Fig EV2B–D), we identified 175 T-follicular helper cells (2.2%; CXCL13 and CD200), 994 T-regulatory cells (12.4%; FOXP3 and BATF), 2,003 other CD4+ T cells (25.1% of all T cells; CD4, IL7R and CD40LG), 3,691 CD8+ T cells (46.2%; CD8A and GZMH), 605 proliferating T cells (7.6%; MKI67), 358 NK Cells (4.5%; GNLY, KLRD1, NCR1, XCL1 and NCAM1) and 164 NKT cells (2.1%; GNLY, KLRD1, NCR1 and CD3D). The remaining cells consisted of 610 endothelial cells (2.5%; PECAM1) and two distinct clusters (with 1,409 and 320 cells, 5.8 and 1.3%, respectively) sharing the expression of common stromal markers including PDGFRB, S100A4 (FSP-1), ITGB1 (CD29) and THY1 (CD90). These non...
endothelial nor immune cells (collectively referred to as stromal in this study) were enriched for a fibroblast cell type signature from XCell (Fibroblasts_FANTOM_1; Fig EV2A). All annotated cell types were detected in each patient, with varying proportions of cell types between cases, indicating no patient specific subpopulations in our integrated dataset (Fig 1F).

Re-clustering stromal cells revealed four distinct sub-clusters in human TNBCs

Although the stromal clusters shared many common markers used to study CAFs, we further inspected their heterogeneity through re-clustering each population (Fig 2A). Sub-clusters were detected across multiple clustering resolutions in the FindClusters function in Seurat (resolutions 0.2, 0.3 and 0.4), with varying proportions from each patient (Fig 2B). The first cluster, which was classified as CAFs through the expression of fibroblast-specific markers (PDGFRα, COL1A1, FAP and PDPN), formed two sub-clusters (Fig 2A–C). The first CAF sub-cluster was comprised of 280 cells (16.2% of all stromal; red cluster) and was classified as myofibroblast-like CAFs (myCAFs) through the elevated expression of activated fibroblast markers (ACTA2, FAP and PDPN) and collagen-related genes (COL1A1 and COL1A2) (Fig 2C and D; Ohlund et al, 2017; Biffi et al, 2018; Elyada et al, 2019). The second CAF sub-cluster comprised of 1,129 cells (65.3%; orange cluster) and resembled inflammatory-CAFs (iCAFs) through the enrichment of the CAF chemokine marker CXCL12 (also known as SDF-1) (Fig 2C and D; Ohlund et al, 2017; Biffi et al, 2018; Elyada et al, 2019). We next compared our CAF clusters to the subsets previously reported in pancreatic cancer (Ohlund et al, 2017; Biffi et al, 2018; Elyada et al, 2019). This was performed by scoring published CAF gene signatures across our stromal clusters using the AUCell method (Fig EV2E; Aibar et al, 2017). This revealed the enrichment of pancreatic myCAF and iCAF signatures in our breast myCAF and iCAF clusters, respectively, suggesting similar phenotypes likely exist across both tissue sites (Fig EV2E). While the signatures were largely conserved, a number of human PDAC CAF markers were detected in opposing cell types, for example Il6 was expressed by PVL cells rather than iCAFs (Fig EV2F). No clusters showed any particular enrichment for signatures of antigen-presenting CAFs, or are unique to pancreas tumours (Fig EV2E).

In contrast, the second stromal cluster was enriched for perivascular markers, including genes associated with pericytes and smooth muscle cells (ACTA2, MCAM, CAV1, TAGLN, MYH11, MYLK and RGSS; Fig 2C and D; Hamzah et al, 2008). MCAM (also known as CD146) has shown to be a robust marker to differentiate perivascular cells from fibroblasts in human tissues (Li et al, 2003; Middleton et al, 2005; Covas et al, 2008; Crisan et al, 2008). PVL cells were further classified as either differentiated-PVL (dpVL; 122 cells in light blue, 7.1%), characterised through the enrichment of myogenic differentiation genes (TAGLN, MYH11 and MYLK), or immature-PVL (imPVL; 198 cells in dark blue; 11.5%), characterised by the elevated expression of genes associated with an immature phenotype (PDGFRα, CD36 and RGSS) (Fig 2C and D; Song et al, 2005). To our surprise, both PVL subsets were also enriched for the human PDAC myCAF signature, suggesting PVL cells share some similarities in gene expression profile with myCAFs (Fig EV2E and F).

All four stromal subsets were detected in all five patients; however, there were differences in the proportions between the patients (Figs 2B, and EV3A and B). The stromal profiles of Patient-1 (P1) and P2 were predominantly comprised of iCAFs, myCAFs were highest in P3, while PVL cells were highly abundant in P4 and P5 (Figs 2B, and EV3A and B).

Next, we identified differentially expressed genes (DEGs) between the four subsets using the MAST method, which compares each subset against all other subsets (Finak et al, 2015). This identified a total of 894, 610, 289 and 289 DEGs (log fold change threshold of 0.1, P-value threshold of $1 \times 10^{-5}$ and FDR threshold of 0.05) by myCAFs, iCAFs, dpVL and imPVL cells, respectively (Fig 2D; Dataset EV1). We performed gene ontology (GO) analysis using the top 250 DEGs from each subset using the clusterProfiler tool (Fig 2E; Dataset EV2) to determine the pathway level differences driving stromal heterogeneity (Yu et al, 2012). This revealed an enrichment of collagen biosynthesis and ECM regulatory pathways in myCAFs, which included fibrillar collagen genes COL1A1 and COL1A2 and ECM remodelling metalloproteinases MMP1 and MMP11 (Fig 2D and E). We identified the enrichment of developmental signalling pathways and chemotactic regulation in iCAFs, including soluble factors such as IGF1, FIGF and PDGF, and the chemokines CXCL12 and CXCL13 (Fig 2D and E). Stem cell markers including ALDH1A1 and ID2, and the growth factor receptor EGFR were also upregulated in iCAFs (Fig 2D). Within the PVL cells, the dpVL cluster was enriched for pathways related to the muscle system and contractility, while the imPVL cluster was enriched for pathways related to focal and substrate adhesion, including the integrin molecule ITGA1 (Fig 2D and E). No stromal clusters expressed canonical markers for proliferation, including MKI67 and AURKA.

As many of the genes and pathways identified were related to cell activation and contractility, we hypothesised that the stromal subclusters resembled cell differentiation stages rather than distinct subpopulations. Cell trajectories were examined using the Monocle method, which revealed subsets of CAFs and PVL cells distributed across pseudotemporal space (Fig EV3C and D; Qiu et al, 2017). For example, COL1A1, ACTA2 and CXCL12 expression transitioned throughout CAF differentiation (Fig EV3C), while CD36, RGSS and MYH11 transitioned throughout PVL differentiation (Fig EV3D). Our findings indicate that the stroma in TNBC is comprised of four major transcriptional states related to cell differentiation, which branch from the two major fibroblast and perivascular-like lineages.

Transcription factor pathways enriched across stromal subclasses

We next sought to investigate if gene regulatory networks could further explain the underlying heterogeneity in stromal subpopulations. To examine the activity of CAF and PVL transcription factors (TFs), we applied the SCENIC method to build gene regulatory networks from scRNA-Seq data and identify activating cis-regulatory elements (Aibar et al, 2017; Moerman et al, 2018). Through applying this to the normalised stromal gene expression matrix, SCENIC identified a total of 190 activated TFs, of which 166 were identified to be significantly different across the four stromal subsets (one-way ANOVA; $P$-value threshold of $1 \times 10^{-5}$). We focused on the top 50 strongest candidates based on their average AUC values (Fig 3; Appendix Fig S2).
Figure 2. Stromal landscape of TNBCs reveals four subpopulations of cancer-associated fibroblasts and perivascular-like cells.

A t-SNE representation of the four subclasses of cancer-associated fibroblasts (CAFs) and perivascular-like cells (PVL), named myofibroblast-like CAFs (myCAFs; 280 cells), inflammatory-like CAFs (iCAFs; 1,129 cells), differentiated-PVL cells (dPVL cells; 122 cells) and immature-PVL cells (imPVL cells; 198 cells).

B Plot showing the proportion of the four stromal subsets across all five patients.

C Log expression of parenchymal gene markers commonly associated with CAFs and perivascular cells.

D Cluster averaged log-normalised expression of the top 300 differentially expressed genes between the four stromal subsets with stromal-related genes of interest annotated. Expression values are scaled per cluster.

E Circle histogram plot of the top gene ontologies enriched in each of the four stromal subsets, with pathways broadly grouped for ECM, development and signalling, muscle contractile features and angiogenesis and adhesion. Scale bar represents the $-\log_{10} q$-value for the enrichment of individual GO terms, as determined using ClusterProfiler.
In examining the top candidate TFs (Fig 3; Appendix Fig S2), ZEB1 and FOXP1 were enriched in myCAFs. A recent study inhibiting stromal ZEB1 in the PyMT mouse model of breast cancer reduced tumour growth, invasion and impaired ECM deposition (Fu et al., 2019). In other tissue contexts, FOXP1 was reported to regulate the fibroblastic potential of stromal cells via the Wnt/beta-catenin pathway, including myCAF marker genes such as ACTA2 and COL1A1 (Shao & Wei, 2018). Known roles of such TFs are consistent with the predicted ECM-regulating phenotype of myCAFs. The EGR2 and TCF7L2 regulons were enriched in iCAFs (Fig 3). EGR2 is known to regulate the expression of immunomodulatory molecules in mesenchymal stem cells (Tamama & Barbeau, 2012). The TCF family including TCF7L2 (also known as TCF4) are Wnt-regulated TFs that are highly expressed during early development (Hrckulak et al., 2016). As iCAFs also expressed the stem cell markers ALDH1A1 and ID2, we hypothesised that they resemble a stem or progenitor-like state.

For PVL cells, MEF2C was a highly enriched driver in both subsets (Fig 3). Myocyte enhancer factor 2 (MEF2) is a well-defined regulator for the development of vascular smooth muscle cells (Creemers et al., 2006; Gordon et al., 2009). We identified KLF2 enriched in dPVL cells, and NR2F2 enriched in imPVL cells (Fig 3). KLF2 is required for smooth muscle cell migration and maturation in blood vessel formation, consistent with the predicted differentiation state of dPVL cells (Wu et al., 2008). Furthermore, NR2F2, also known as COUP-TFI, is highly expressed by myogenic precursors and is known to inhibit muscle development, which is consistent with the predicted immature state of imPVL cells (Lee et al., 2017).

Figure 3. Polarised gene regulatory states between cancer-associated fibroblasts and perivascular-like subclasses.
A. Polarised gene regulatory states underlying stromal subclasses. Heatmap shows the averaged regulon activity (area under the curve; AUC) for the top 50 highest TFs regulons as estimated using SCENIC. All regulons are statistically enriched across the four subsets (P < 1 × 10⁻⁵ One-way ANOVA). Heatmap is clustered using Euclidean distance and complete linkage.
B. Candidate transcriptional drivers of each CAF and PVL subset. Violin plots showing the log-normalised gene expression (left) of the TF and its respective AUC regulon activity (right). TFs ZEB1 and FOXP1 enriched in myofibroblast-like CAFs, EGR2 and TCF7L2 enriched in inflammatory-like CAFs, MEF2C enriched in PVL cells and NR2F2 enriched in immature-PVL cells.
In summary, we identified unique and novel TF drivers in each of the four stromal subclasses, providing further insights into the transcriptional drivers underlying stromal heterogeneity.

Validation of stromal subsets in primary breast cancer tissue

To validate the existence of the four stromal subclasses described above in TNBC patient tissue (Fig 4A), we first performed fluorescence-activated cell sorting (FACS) isolation on scRNA-Seq matched human tissue sections (Fig 4B). Our gating strategy used EPCAM, CD45 and CD31 as negative markers to exclude epithelial, immune and endothelial cells, respectively (Fig 4B). We additionally used PDGFRβ to positively select all stromal populations and avoid contaminations from cancer stem cells and breast myoepithelial cells which have low EPCAM expression (Prater et al, 2014; Hyun et al, 2016). Based on our initial scRNA-Seq findings, we determined PDGFRα and CD146 (MCAM) as good markers to discriminate CAFs and PVL cells, respectively. Following the initial isolation and culturing of CAFs (PDGFRβ+/PDGFRα+/CD146+) and PVL cells (PDGFRβ+/PDGFRα−/CD146−), we next performed simultaneous FACS analysis of additional stromal markers to validate the presence of the four stromal subsets in culture. We show that myCAFs and iCAFs could be distinguished by FAPHIGH/CD90HIGH and FAPLOW/CD90LOW expression, respectively (Figs 4B, and EV3E and F), while imPVL cells could be discriminated from dPVL cells by CD36+ expression (Fig 4B). We validated the gene expression of cultured bulk and sorted CAF fractions using quantitative real-time PCR (qPCR) (Fig EV3G). As controls, PDGFRα and PDGFRβ were expressed in both the FAP-high and FAP-low populations. Consistent with the FACS sorting strategy and scRNA-Seq findings, FAP and ACTA2 were enriched in FAPHIGH (myCAF) sorted cells, while CXCL12 and EGFR were enriched in FAPLOW (iCAF) sorted cells (Fig EV3G). We next performed immunofluorescence (IF) to further validate additional markers and explore potential morphological differences. Here, α-smooth muscle actin expression was used to identify myCAFs from iCAFs (Fig 4C; Appendix Fig S3), and CD36 to distinguish imPVL from dPVL cells (Fig 4D; Appendix Fig S3). From our observations, myCAFs and dPVL cells had a more elongated morphology in comparison to iCAFs and imPVL cells (Fig 4C; Appendix Fig S3), which is consistent with the predicted differentiation state of each subset. Importantly, we defined a novel gating strategy that allowed us to purify the four stromal subsets for subsequent in vitro functional characterisation.

Myofibroblast-like CAFs have elevated capabilities for collagen secretion and alignment

From the above results, we predicted myCAFs to be the predominant subset synthesising ECM components. To investigate this, we generated cell-derived matrices (CDMs) to compare the ability of each human stromal subset to lay down collagen, as previously described (Cukierman et al, 2001). Purified stromal subsets were seeded and cultured onto glass for 7 days. To assess Collagen I deposition, we used second harmonic generation (SHG) microscopy, which is a sensitive method for quantifying fibrillar collagen density and orientation in an unlabelled manner. This revealed FAPHIGH myCAFs had a significant increase in SHG signal intensity compared to FAPLOW iCAFs, while PVL cells had a significantly lower SHG signal compared to both CAF subsets (Fig 4E). Higher densities of stromal collagen are a hallmark of breast tumour growth, invasiveness and risk of disease development (McCormack & dos Santos Silva, 2006; Levental et al, 2009; Huo et al, 2015). Our findings also indicate that PVL cells do not adopt fibroblast-like traits in contributing to the collagenous TME. Further analyses of collagen fibre orientation also revealed that in addition to increased amounts, the orientation of the collagen fibres deposited by myCAFs was more uniformly aligned compared to iCAFs and PVL cells (indicated by the higher, narrow peak in Fig 4F). It has been previously shown that tumour-associated collagen signatures (TACS), characterised by the alignment of collagen fibres, is a good factor for predicting breast cancer survival (Conklin et al, 2011). In further parallels to pancreatic cancers, FAP-overexpressing fibroblasts have been shown to produce more parallel aligned fibres, enhancing the directionality and velocity of cancer cell invasion (Lee et al, 2011). Importantly, these data highlight that the regulation of the ECM, namely in collagen density and orientation, is mainly regulated by the specialised myCAF subsets. In summary, our findings demonstrate that the stromal subclasses described here are functionally distinct and provide a novel strategy for their purification from breast cancers.

Stromal subclasses are spatially distinct

To investigate the spatial localisation of CAFs and PVL cells, we performed immunohistochemistry (IHC) with markers identified by scRNA-Seq on data matched patient tissues. We also wanted to validate that CAFs and PVL cells localise to the intratumoural regions of tumour specimens and are not from adjacent normal tissue or blood vessels. We stained serial 4 µm sections and identified stromal cell types using a combination of markers identified previously by scRNA-Seq and DGE (Fig 2C): pan-stromal (PDGFRβ+), myCAFs (PDGFRβ+, a-SMAHIGH and CD146+), iCAFs (PDGFRβ+, a-SMA−, CD34HIGH and CD146+) and PVL cells (PDGFRβ+, a-SMAHIGH, CD34− and CD146+). As CD34 and CD146 are commonly used markers of the endothelium but are mutually exclusive in CAFs and PVL cells, we used their co-localisation in combination with PDGFRβ staining and morphology (rings surrounding lumen) to identify endothelial cells (Middleton et al, 2005). This IHC strategy revealed regions where myCAFs (α-SMAHIGH) were located in close proximity to the invasive tumour interface, while iCAFs (CD34HIGH) were relatively distal to this interface (Fig 4G). In these particular cases, no PVL cells were present in these regions and CD146 was completely restricted to blood vessels (Fig 4G). In distal regions which were enriched for iCAFs, we also identified a high co-localisation of tumour-infiltrating lymphocytes as identified by morphology (Fig 4G).

By definition, vascular smooth muscle cells (vSMCs) and pericytes should be localised around arteries and veins to facilitate vascular development and stability. To examine whether PVL cells are vessel-associated, we used co-IF staining for CD31 and CD146 to mark endothelial cells and PVL cells, respectively. We readily detected PVL cells at non-blood vessel regions in the stroma of 4 out of 5 matched patient tissue sections (all cases except P3), including P4 where it was highly abundant (Figs 4H and I, and EV4A). Consistent with the cell proportions identified by scRNA-Seq, PVL cells were highly abundant in P4, and lowly detected in P3 (Fig 2B). PVL
**Figure 4.**

Panel A: Schematic illustration of CAFs and PVL cells with differentially expressed markers.

Panel B: Heatmap analysis of PDGFβ, CD146, and CD31 expression in CAFs and PVL cells.

Panel C: Immunofluorescence images showing CD34, α-SMA, and merged channel for myCAFs, iCAFs, and PVL cells.

Panel D: Immunofluorescence images showing CD146, α-SMA, and CD36 expression in myCAFs, iCAFs, and PVL cells.

Panel E: Normalized small scale intensity analysis for myCAFs, iCAFs, and PVL cells.

Panel F: Heatmap showing fiber alignment frequency for myCAFs, iCAFs, and PVL cells.

Panel G: Histological images showing PDGFβ, α-SMA, and merged channel for invasive front and distal region.

Panel H: Immunohistochemistry images showing PDGFβ, CD146, α-SMA, and CD34 expression in PVL cells.

Panel I: Immunofluorescence images showing DAPI, CD31, and merged channel for BV and PVL cells.
cells were highly dispersed throughout the tumour stroma with no obvious co-localisation to the invasive malignant borders. Importantly, our findings suggest that these smooth muscle-like cells, like CAFs, can be readily identified disseminated throughout the stroma, independent of blood vessels.

To understand how the four stromal subpopulations correspond to their normal tissue counterparts, we repeated the staining of PDGFRβ, CD34, α-SMA and CD146 on healthy breast tissue collected from four women. This revealed a high abundance of iCAF-like fibroblasts (PDGFRβ⁺, α-SMA⁺, CD34⁺HIGH and CD146⁻) surrounding ductal regions, while myCAF-like fibroblasts (PDGFRβ⁺, α-SMA⁺HIGH and CD146⁻) were sparsely detected across all four cases (Fig EV4B and C). While this small panel of markers do not highlight the large transcriptional changes that may occur upon CAF activation, it does suggest that the broad iCAF-like and myCAF-like fibroblast subsets are resident cell types which are reactivated during carcinogenesis. For PVL cells, IHC staining of CD146 was completely restricted to blood vessels (Fig EV4C). This further confirmed using co-IF staining for CD31 and CD146 on the normal tissue cases, where CD146 was completely restricted to CD31-positive blood vessels (Fig EV4D). Our findings suggest that disseminated PVL cells are a distinct feature in a subset of TNBCs.

Distinct ligand–receptor expression predicts diverse stromal crosstalk to the tumour microenvironment

We next sought to investigate how spatially distinct stromal subclasses may interact with other cells within the TME. Here, we annotated our scRNA-Seq dataset using a published set of curated human ligand–receptor pairs (Ramilowski et al., 2015). We used these annotations to construct a cell-to-cell communication network and predict intratumoural signalling between the four stromal clusters, and the surrounding neoplastic, immune and endothelial microenvironment. This revealed diverse stromal signalling profiles (Fig 5A), with myCAFs and iCAFs having the highest overall predicted ligand activity out of all the cell types (Fig 5B). The “interaction strength”, or the weight of each edge, was defined as the product of expression levels of the corresponding ligand and receptor. All ligand–receptor pairs with an arbitrary “interaction strength” cut-off of 0.1 were classified as candidate signalling molecules, which revealed a total of 570, 482, 437 and 357 unique predicted interactions between stromal clusters with cancer epithelial cells, endothelial cells, myeloid cells (Appendix Fig S4A–C) and T-cell subpopulations, respectively (Appendix Fig S4D; Dataset EV3).

Consistent with the enrichment of growth factor signalling gene ontologies in iCAFs (Fig 2E), we identified a strong upregulation of crosstalk via the FGFR (FGFR1, 2 and 3), PDGF (PDGFAB and PDGFCC), HGF and IGF1 pathways to their cognate receptors across cancer cells and endothelial cells (Fig 5C; Appendix Fig S4A and B). These factors are known to be highly expressed in breast tumours and associated with breast cancer proliferation, invasion and inducing cancer stem cell (CSC) phenotypes (Palmieri et al., 2003; Alarco et al., 2007; de ostrovich et al., 2008; Kuang et al., 2017). Different ligands from these pathways were also identified from myCAFs and dPVL cells, suggesting that neoplastic phenotypes could also be influenced by different stromal cells (Appendix Fig S4A). As we identified iCAFs to be located more distal to the invasive tumour interface, we hypothesise that these secreted factors function from a distance. For signalling to the endothelial compartment, iCAFs and PVL cells were both enriched for well-characterised growth factors involved in angiogenesis (Appendix Fig S4B). Classical angiogenic pathways including VEGFs (VEGF, also known as VEGFD), PDGs (PDGFCC), IGFs (IGF1 and IGF2) and Notch signalling (DLL1) were enriched in signals emanating from iCAFs (Appendix Fig S4B). These pathways suggest that the inflammatory-CAF phenotype is also associated with tumour neovascularisation (Samani et al., 2017; Wang et al., 2019). In addition, PVL-derived signals were enriched for the canonical ANGPT1/ANGPT2-TIE1 pathway, which are known stimuli that can induce the sprouting of new vessels during the formation of new endothelial tubes (Fagiani & Christofori, 2013).

Given the reported immunoregulatory properties of mesenchymal cells (Costa et al., 2018; Givel et al., 2018), we next focused on the signalling of stromal cytokines and checkpoint molecules to immune populations. Here, we identified an enriched interaction between iCAFs and myeloid cells via the complement cascade activation interaction C5-C5AR1 (Fig 5D; Appendix Fig S4C). C5
activation in the TME acts as a chemotactic factor for the recruitment of immunosuppressive myeloid cells to suppress T-cell activities (Markiewski et al., 2008). In addition, myCAFs and iCAFs were enriched for TGFβ1-TGFBR1 and TGFβ2-TGFBR1 interactions with myeloid cells, respectively (Fig 5D; Appendix Fig S4C). As TGFβ-activated myeloid cells have been shown to enhance breast cancer progression and metastasis in vivo, it suggests that both CAF subsets could influence myeloid phenotypes (Li et al., 2012). While the TGFβ1 receptor was predominantly enriched on myeloid clusters, it is worth noting that its expression was also detected by cancer and endothelial clusters (Fig 5D). Although PVL cells had lower ligand expression profiles compared to CAFs, several immunomodulatory cytokine interactions were predicted between PVL cells and myeloid cells, including an enrichment of the CCL8-CCR1, IL6-IL6R and CCL2-CCR1 pathways (Fig 5D; Appendix Fig S4C). CCL2 produced by the microenvironment in other cancers has been shown to be essential for the recruitment of T-Reg and tumour-associated macrophages, supporting an additional role of PVL cells in recruiting immunosuppressive cells (Chang et al., 2016).

For the signalling to the lymphocyte compartment, iCAFs had a strong upregulation of the chemo-attractant pathways CXCL12-CXCR4 and CXCL13-CXC5 with T- and B cells (Fig 5E; Appendix Fig S4D). CAF-derived CXCL12 has been shown to recruit and regulate the activity of CD4+/CD25+ T-Reg in breast cancers, suggesting iCAFs may have a direct role in recruiting immunosuppressive populations (Costa et al., 2018; Givel et al., 2018). CXCL12 and CXCL13 signalling axes have also been shown to mediate lymphocyte recruitment to tertiary lymphoid structures (TLS) (Sautes-Fridman et al., 2016). MyCAFs were also enriched for secreted immunoregulatory molecules and checkpoints including CXCL9-CXCR3, CXCL11-CXCR3 and CD274-PDCD1 (PD1-PD1) with T cells (Fig 5E; Appendix Fig S4D). Lastly, only few candidates were identified between PVL cells with T cells, including the enrichment of CCL21-CCR7, which is associated with immune tolerance in favour of tumour progression (Fig 5E; Appendix Fig S4D) (Shields et al., 2010). It is evident from our signalling predictions that diverse immunoregulatory molecules are expressed in the stroma, highlighting that immune evasion can be regulated by distinct stromal subpopulations in TNBC.

**Inflammatory-CAFs associated with cytotoxic T-lymphocyte dysfunction**

To further investigate the influence of stromal subsets on immune evasion, we explored the association between distinct stromal gene signatures and immune content in three large independent TNBC patient cohorts with associated bulk gene expression data.

---

**Figure 5. Predicted stromal crosstalk to cancer and immune cells.**

Overview of the predicted stromal paracrine signalling conserved across the five TNBC patients. The scRNA-Seq dataset was annotated by ligand–receptor pairs as curated in Ramlowski et al (2015).

A. Circos plot summary of the stromal ligand–receptor interactions. Outer sectors are weighted according to the number of annotated ligand–receptor interactions per cell type. Links between sectors are weighted according to the “Interaction Strength”, calculated as a product of ligand and receptor expression. Links are coloured by the respective stromal subsets, myCAFs (red), iCAFs (orange), dPVL cells (blue) and imPVL cells (light blue).

B. Summary of the total ligands and receptors annotated per cell type.

C–E. Imputed gene expression of selected candidate signalling molecules identified between the four stromal subsets and malignant (C) epithelial, (D) myeloid and (E) T cells. Expression of ligands in stromal clusters is represented on the left, with cognate receptors on target cells on the right.
(METABRIC, GSE8812 and GSE21653) (Sabatier et al, 2011; Curtis et al, 2012; Jezequel et al, 2015). Using a computational model called tumour immune dysfunction and exclusion (TIDE), we examined two primary mechanisms of immune evasion. The first examines factors driving the “dysfunction” of cytotoxic T lymphocytes (CTLs), while the second examines factors preventing the infiltration of CTLs to the tumour, known as “exclusion” (described below) (Jiang et al, 2018). TIDE first estimates CTL levels in each sample within a bulk sequencing cohort using the averaged expression of CTL-specific genes (See Materials and Methods). Patients are then stratified into high and low CTL groups based on comparisons to the mean CTL level within the cohort. For dysfunction, we then evaluated whether gene signatures from each of the stromal subsets influence the beneficial effect of CTL levels on patient prognosis (Jiang et al, 2018). This analysis revealed a strong enrichment of genes from the iCAF dysfunction signature that were significantly associated with CTL dysfunction in all three bulk tumour cohorts (Fig 6A). In patients with a low iCAF dysfunction signature level, a significant survival benefit was associated with high CTL levels (Figs 6B and EV5A). This is consistent with previous clinical observations in TNBCs where lymphocyte infiltration is a robust prognostic factor for improved disease-free survival and overall survival benefit (Loi et al, 2013). Remarkably, in patients with a high iCAF dysfunction signature level, CTL levels were not associated with prognosis in any of the three cohorts (Figs 6B and EV5A), suggesting a role for stromal iCAFs in driving dysfunctional CTLs in TNBC. Other stromal subset signatures did not show a significant enrichment of prognostic genes in the context of CTL dysfunction.

To investigate whether CTLs in each patient were indeed dysfunctional, we scored a published T-cell exhaustion gene signature in our CD8+ T-cell populations from each patient using an AUC approach (Fig 6C; Blackburn et al, 2009). This gene set includes canonical markers of exhausted T cells including PDCD1 (PD-1), LAG3, TIGIT and CTLA4 (Blackburn et al, 2009). This revealed heterogeneity for exhausted CD8+ T-cell populations in all five patients (Fig 6C), with P2 and P4 having the highest average exhausted gene signature score. In contrast, the exhaustion signature was not enriched in any other cell population with the exception of the myeloid cell cluster (Fig 6C). Myeloid cells, which can include tumour-associated macrophages and myeloid derived suppressor cells, are known to hold immunosuppressive properties and can also express inhibitory molecules associated with T-cell suppression (Jiang et al, 2015).

**Discussion**

Our study describes a detailed taxonomy of human stromal subclasses in TNBC at cellular resolution. The activated tumour stroma is classically described using a broad ‘CAF’ classification. Here, we provide evidence that it is also comprised of functionally distinct perivascular-like cells which are not necessarily associated with the endothelium. We show that stromal heterogeneity diverges to four distinct states: myofibroblast-like, inflammatory-like CAFs and developmental CAF subsets (Bartoschek et al, 2012; Jezequel et al, 2015), and dPVL and imPVL clusters independently (Figs 6E and EV5C). This revealed that dPVL cells were the only subset with a significant negative correlation with CTL level in three of four cohorts, suggesting they are the primary subset associated with T-cell exclusion (Figs 6E and EV5C). To further explore this correlation in our five patient, tumour-infiltrating lymphocytes (TILs) and CTLs were scored in matched tumour sections by a specialist breast pathologist. Total TILs were estimated using standard H&E-based assessment (Fig 6F), while stromal CTLs were accurately quantified by CD8 staining and scored as previously described (Fig 6G; Salgado et al, 2015). The latter measurements were performed as TILs can also be comprised of non-CTL populations including CD4+ T cells, T-Regs and B cells. TILs and CTL scoring revealed that 2 out of 5 patients (P4 and P5) had very low CTL infiltration (< 5% TILs and < 50 CD8+ T cells per 1 mm2), whereas P3 had a very high infiltration (> 70% TILs and > 200 CD8+ T cells per 1 mm2) (Fig 6F–H). In support of dPVL cells as drivers of T-cell exclusion, only 4% of stromal cells from P3 were annotated as dPVL cells, while P4 and P5 had the two largest proportions of dPVL profiles with 35.5 and 26.8%, respectively (Fig 2B). Furthermore, no disseminated PVL cells could be readily detected in P3 using co-IF (Fig EV4A). While small numbers, our findings are consistent with the proposal that specialised stromal subclasses are associated with immune evasion.
expression of epithelial lineage exclusive keratins (Appendix Fig S5B and C). Our comparisons suggest that developmental CAFs are either unique to mouse models or are cancer cells whose expression

Figure 6. Inflammatory-CAFs and differentiated-PVL cells associated with immune evasion in TNBC patient cohorts.

Significant associations between iCAF and dPVL gene signatures with cytotoxic T-lymphocyte (CTL) dysfunction and exclusion in multiple TNBC patient cohorts, respectively, as determined using the tumour immune dysfunction and evasion (TIDE) method.

A iCAF T-cell dysfunction gene signature highlighting genes significantly associated with CTL dysfunction in two out of three independent patient cohorts (METABRIC, GSE21653 and GSE58812).

B Representative cohort (METABRIC) showing the prognostic value of iCAF T-cell dysfunction signature in the context of CTLs for a total of 233 patients. Kaplan–Meier present two groups of patients, “low CTL” (blue line) and “high CTL” (red line), as estimated according to the average expression of CTL-specific genes and stratified as compared to the mean. Tumours with low iCAF T-cell dysfunction signatures (top) show patients with high CTL levels have a better survival outcome. In contrast, this survival benefit is lost in tumours with a high iCAF T-cell dysfunction signature (bottom). P-values were defined from the Cox proportional hazard (Cox–PH) model.

C Dysfunctional CTLs detected in all five TNBC patients determined through scoring a T-cell exhaustion signature. UMAP featureplot of the exhaustion signature across all stromal and immune cells as in Fig 1D.

D Bulk stromal signature associates with CTL exclusion. Pearson correlation was computed between all inferred CTL levels (y axis) and the respective correlation between the bulk sample and the single-cell cluster (x axis). Signature of all stromal cells divided over all cells correlated negatively with CTL levels, while control CD4+ and CD8+ gene signatures show a positive correlation. P-values were computed using a two-sided t-test for correlation and were adjusted using the Benjamini–Hochberg procedure.

E dPVL cells associated with CTL exclusion. Repeated analysis in the same manner as in (D), instead with myCAF, iCAF, dPVL and imPVL clusters divided over all stromal cells independently, highlighting that CTL exclusion is mainly driven by dPVL cells. Representative cohort GSE58812 is shown.

F–H dPVL profiles and CTL exclusion consistent in our study. (F) Patients with the highest dPVL profiles by scRNA-Seq (P4 and P5) show the lowest tumour-infiltrating lymphocyte (TIL) pathology counts. (G, H) Accurate quantification of CTLs (G) and representative immunohistochemistry staining for CD8 on matched patient tumour sections (H). n = 5 stromal 1 mm2 regions were counted per tumour. The central band, boxes and whiskers represent the median, lower/upper quartile and min/max CTL counts per 1 mm2, respectively. P3 is shown as an example of a low dPVL profile with high CTLs. In contrast, P4 has a high dPVL profile with low CTLs. Statistical significance was determined using pairwise comparison with Student’s t-test with P-values denoted by asterisks: *P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001.

Source data are available online for this figure.
of EPCAM is down-regulated, which was a negative marker used for CAF isolation in these studies (Prater et al, 2014; Hyun et al, 2016).

Despite well-characterised roles in cancer progression, the cellular origins of CAFs remain poorly understood. Our results support the notion that dispersed stromal cells can also arise from perivascular cells, likely delaminated from vascular structures. Although PVL cells clustered distinctly from CAFs and express perivascular markers including MCAM (CD146), CAV1, RCSS5, MYH11 and TAGLN (SM-22-Alpha), they also expressed an array of markers commonly used to classify CAFs, including ACTA2 (α-SMA), PDGFRB, THY1 (CD90), S100A4 (FSP-1) and ITGB1 (CD29) (Hamzah et al, 2008; Costa et al, 2018). Similar PVL subsets were identified in a previous mouse model of breast cancer (Bartoschek et al, 2018). The authors defined these cells as “vascular-like CAFs” through the expression of vessel development markers such as CD146 (Bartoschek et al, 2018). Although the authors hypothesised that vascular-like CAFs are derived from perivascular cells such as pericytes, the concept of pericyte-to-fibroblast transition has been debated (Hosaka et al, 2016). Our findings from functional assays suggest PVL cells do not possess the defining fibroblast trait of collagen deposition and remain phenotypically distinct from the fibroblast lineage (Fig 4E).

The functional role of perivascular cells in breast cancer is poorly understood. A very early study found that 4 out of 10 breast tumours showed substantial infiltration of vascular smooth muscle cells based on staining for markers including α-SMA, smooth muscle myosin and calponin (Ronnov-Jessen et al, 1995). This finding went without further exploration until this manuscript, where we validate their existence using state-of-the-art scRNA-Seq and staining of CD146 in matched patient tissue. From our TF analysis, we predict the MEF2 regulon to be a strong activating TF of the PVL subclass. MEF2C is a well-defined regulator for establishing vSMCs during development, highlighting a likely vSMC origin of the described PVL cells (Creemers et al, 2006; Gordon et al, 2009). As observed during wound healing, we hypothesise that vSMCs could be stimulated by malignant factors or mitogens, allowing them to migrate from the vessel basement membrane into the stroma (Banerjee et al, 2006; Louis & Zahradka, 2010). This is further supported by in vitro studies showing that breast cancer derived PDGFRs can induce the recruitment and migration of vSMCs (Banerjee et al, 2006). As perivascular cells play an important part during angiogenesis and blood vessel stability, it is also possible that their displacement in tumours is stimulated by, or a driver of, dysregulated angiogenesis or hypoxia. Although it is yet to be studied in the context of perivascular cells, studies have reported that the imPVL marker CD36 is enriched in normal tissue regions and is associated with good survival outcome in breast cancer (DeFilippis et al, 2012). However, the origin and functional role of PVL subpopulations remain to be defined by future studies. The staining of CD146 exclusively associated with blood vessels of normal breast tissue suggests that detached PVL cells are a distinct feature of breast cancers.

Importantly, our findings suggest that previous studies characterising CAFs with a small number of markers have likely also studied PVL cells. For example, subsets discriminated by CD146 have been characterised in endocrine-resistant breast cancers (Brechbuhl et al, 2017). Patients with a CD146+ stroma demonstrated good responses to tamoxifen therapy through the maintenance of oestrogen receptor (ER)-dependent proliferation in cancer cells. Our findings suggest that PVL cells rather than CAFs are a biomarker for ER-directed therapeutic response in ER-positive breast cancers (Brechbuhl et al, 2017), a prediction that requires more detailed validation. Another elegant study reported a subset of chemoresistance-promoting CAFs, marked by α-SMA+, GPR77+ and CD10+ expression (Su et al, 2018). Due to the shared expression of α-SMA between myCAFs and PVL cells, our findings also raise the question whether PVL cells could also contribute to chemoresistance in a subset of patients (Su et al, 2018). Although we did not find an enrichment of GPR77+ CD10+ α-SMA+ cells in any CAF subclasses, this may be explained by the treatment status of our samples.

Lastly, we found a strong enrichment of immunomodulatory pathways in the predicted signalling between stromal cells and immune cells. We identified an array of important candidates in patient tissue for future experimental studies for functional relevance. It is important to acknowledge; however, that transcript signalling predictions are not always concordant with protein expression. Although no CAF subsets in previous mouse studies were distinguishable by immunomodulatory properties (Bartoschek et al, 2018), there are several reports of predicted CAF-immune interactions in human tissue. We found that iCAFs expressed an array of immunomodulatory molecules to cognate receptors on T cells. In other studies, CAFs have been implicated in the recruitment and activity of T-Regs through the regulatory molecules CXCL12, CD40, B7H3, DPP4 and CD73 (Costa et al, 2018). In addition, iCAFs also expressed several molecules known to regulate myeloid cells, including complement C5, IL6 and TGFβ (Markiewski et al, 2008; Li et al, 2012). Myeloid cells, including tumour-associated macrophages and myeloid derived suppressor cells, are well characterised in contributing to an immunosuppressive TME. Most importantly, gene signatures generated from iCAFs were strongly associated with CTL dysfunction in TNBC patient cohorts. We also report a novel dPVL stromal subset strongly associated with CTL exclusion. We identified an enrichment of dysfunctional/exhausted T cells which correlated with their respective stromal profiles, though we acknowledge that our study consists of small patient numbers. In patients with the highest dPVL profile, we found consistently low TIL and CD8 counts in matched pathology. Considering the proposed origin of detached PVL cells from the vascular structure, we hypothesise that this may be related to reduced lymphocyte extravasation from dysregulated tumour blood vessels. In support of this, previous studies restoring vascular integrity in tumours through vessel normalisation and increased perivascular coverage find an influx of CD8+ T cells in tumour tissue (Hamzah et al, 2008; Johansson-Percival et al, 2015). In addition, signalling between CD4+ T cells and pericytes have also been reported to play a reciprocal role in tumour vessel normalisation (Tian et al, 2017). It is possible that the association between dPVL cells and CTL exclusion in patient cohorts reflect tumours with low vascular integrity and may act as a biomarker for patients suitable for vessel normalisation therapeutic strategies.

While our findings point to the targeting of stromal cells, future work investigating the transcriptional changes in stromal cells between healthy breast tissue and cancer is required to understand the stromal states that are cancer-specific vs. reactivated resident cell types. In support of the latter possibility, a recent study showed that there are minimal proteomic differences between normal fibroblasts and CAFs in prostate cancer models (Nguyen et al, 2019). We find that iCAF- and myCAF-like fibroblasts exist in cancer-free normal breast tissue. It is important to note that...
desmoplasia is often observed in cancer-free tissues, particularly in high risk women with high mammographic density (Li et al., 2005). This can be influenced by several physiological factors such as weight, pregnancy and menopausal status, highlighting important factors that need to be considered in future projects examining the normal breast tissue microenvironment such as the human cell atlas project (Li et al., 2005). These differences may exist from distinct epigenetic states between CAFs and normal fibroblasts, indicating another layer of complexity that remains to be explored in the four breast cancer stromal subsets identified in our study (Pidsley et al., 2018). The integration of future assays combining scRNA-Seq with chromatin states will be important in elucidating the epigenetic regulation of cancer-associated stromal cells. Identifying specific activation markers in comparison to healthy tissue is also an important prerequisite for the development of precise cancer therapeutic strategies with low toxicities. Future in vitro and in vivo studies will be important in understanding how stromal cells are dynamically reprogrammed and how the subclasses described here restrain or promote tumour growth and invasion.

Clinical trials for mainstream immune checkpoint therapies including anti-PD1 have shown limited efficacy in the treatment of advanced TNBC. This hints at alternate mechanisms of immune evasion and novel therapeutic strategies are desperately needed to improve immunotherapy for TNBC. Our findings suggest that co-targeting stromal subpopulations could elicit a more effective immune response in a subset of patients through inhibiting CTL dysfunction and exclusion. This remains to be experimentally tested. In conclusion, we have comprehensively profiled four functionally distinct stromal subclasses in human TNBC, not previously described in breast cancer, mouse models or other cancer types. Importantly, we described subsets of CAFs and PVL cells with clinical relevance, presenting as candidates to further investigate. While our dataset captures a majority of the expected cell types from the TME, certain cell types such as adipocytes are under-represented due to biases from standard tissue dissociation protocols. Integration of alternative methods such as single-nuclei sequencing and spatial transcriptomics in future cancer cell atlas studies will be crucial for a comprehensive understanding of the TME. Our findings in only five patients also highlight the potential of applying scRNA-Seq methods to larger scale patient cohorts for the identification of new disease relevant cell states and their gene expression features.

Materials and Methods

Ethics approval and consent for publication

Patient tissues used in this work were collected under protocols x13-0133, x16-018 and x17-155. HREC approval was obtained through the SLHD (Sydney Local Health District) Ethics Committee; RPAH (Royal Prince Alfred Hospital) zone, and the St Vincent’s hospital Ethics Committee. Site-specific approvals were obtained for all additional sites. Written consent was obtained from all patients prior to collection of tissue and clinical data stored in a de-identified manner, following pre-approved protocols. Consent into the study included the agreement to the use of all patient tissue and data for publication.

Tissue dissociation

Fresh surgically resected tissue was washed with RPMI 1640 (Thermo Fisher Scientific) and minced with scissors. Samples were enzymatically dissociated using Human Tumor Dissociation Kit (Miltenyi Biotec) according to manufacturer’s protocol (https://www.miltenyibiotech.com/AU-en/products/macs-sample-preparation/tissue-dissociation-kits/tumor-dissociation-kit-human.html#ref). Following incubation, the sample was then resuspended in RPMI 1640 and filtered through MACS® SmartStrainers (70 μM; Miltenyi Biotec), and the resulting single-cell suspension was centrifuged at 300 × g for 5 min. Red blood cells were lysed with Lysing Buffer (Becton Dickinson) for 5 min, and the resulting suspension was centrifuged at 300 × g for 5 min. Viability was assessed to be > 80% using Trypan Blue (Thermo Fisher). Viability enrichment was performed using the EasySep Dead Cell Removal (Annexin V) Kit (StemCell Technologies) as per manufacturers protocol. Dissociated cells were resuspended in a final solution of PBS with 10% foetal calf serum solution prior to loading on the 10x Chromium platform. Previously reported stress pathways associated with tissue dissociation, such as FOS/JUN and reactive oxygen species, were not identified as part of the candidate genes, pathways and transcription factor networks differentiating the stromal subsets in our study (O’Flanagan et al., 2019).

Single-cell RNA sequencing on the 10x Chromium platform

High-throughput droplet-based scRNA-Seq was performed on the single-cell suspensions using the Chromium Single Cell 3’ v2 Library, Gel Bead and Multiplex Kit and Chip Kit (10x Genomics) according the to manufacturer’s instructions, with a target of 5,000 cells per lane. scRNA-Seq libraries were sequenced on the Illumina NextSeq 500 platform with pair-end sequencing and dual indexing according to the recommended Chromium platform protocol; 26 cycles for Read 1, 8 cycles for i7 index and 98 cycles for Read 2.

Data processing

Sample demultiplexing, reference mapping, barcode processing and gene counting were performed using the Cell Ranger Single Cell Software v2.0 (10x Genomics). Reads were aligned to the GRCh38 human reference genome. Raw count matrices were exported and filtered using the EmptyDrops package in R (Lun et al., 2018). EmptyDrops distinguishes “real” barcodes from “noise” by calculating deviations of each cell against a generated ambient background RNA profile. Filtered barcodes were then processed using the Seurat v2.0 package in R (Satija et al., 2015). Additional conservative cut-offs were further applied based on the number of genes detected per cell (> 200), and the percentage of mitochondrial unique molecular identifier (UMI) counts (< 10%). Individual Seurat objects were then integrated using the canonical correlation analysis (CCA) function RunMul ticCA according the developer guidelines (Butler et al., 2018). The top 2,000 most variable genes from each sample were combined for CCA vector identification. The first 20 CC dimensions were used for the alignment of subspaces and UMAP projection.
Cluster annotation

Cell clusters were annotated using canonical cell type markers for epithelial (EPCAM), myoepithelial (EPCAM$^{0^+}$, ACTA2, KRT5 and KRT14), basal (KRT5 and KRT14), mature luminal (ESR1), endothelial (PECAM1), immune (PTPRC), T cells (CD3D, CD8A and CD4), T-regulatory cells (FOXP3), B cells (MS4A1), plasmablasts (UCHAIN), myeloid cells (CD68) and stromal cells (PDGFRB and COL1A1). Malignant epithelial cells were distinguished from entrapped normal epithelial cells by inferring copy number variations using the inferCNV package as previously described (Patel et al., 2014). In addition, an area under the curve (AUC) approach using published cell type signatures from the XCELL database was performed using AUCell (Aibar et al., 2017; Aran et al., 2017). AUCell scores single cells with input gene signatures and analyses its activity and distribution across the entire dataset to explore the relative expression of the gene set of interest. AUCell utilises raw gene counts and thus is independent of normalisation bias. CAFs, PVL cells and T cells were independently re-clustered using the Seurat v3 method. Re-clustering was performed across resolutions 0.2, 0.3, 0.4 and 0.5 to identify stable clusters.

Differential gene expression and pathway enrichment

The MAST method was used to perform differential gene expression through the FindAllMarkers function in Seurat (log fold change threshold of 0.1, P-value threshold of $1 \times 10^{-5}$ and FDR threshold of 0.05) (Finak et al., 2015). The top 250 DEGs from each cluster were then passed on to the ClusterProfiler package for functional enrichment (Yu et al., 2012). The compareClusters function was used with the enrichGO databases CC, MF and BP sub-ontologies using the human org.Hs.eg.db database.

Pseudotime cell trajectory analysis

The Monocle 2 method was applied to infer cell trajectories for CAFs and PVL cells using default parameters, as recommended by developers’ (Qiu et al., 2017). CAFs from Patient-2 and PVL cells from Patient-1 were extracted for Monocle analysis due to adequate cell numbers and representations of each respective subset. Gene expression matrices from each cell type were first exported from Seurat into Monocle 2 to construct a CellDataSet. Variable genes defined by the differentialGeneTest function (q-val cut-off < 0.001) were used for cell ordering and dimensionality reduction with the setOrderingFilter and reduceDimension functions, respectively.

Gene regulatory analysis using SCENIC

Investigation of gene regulatory networks using SCENIC was performed using a faster python implementation of the tool (pySCENIC) as described by the developers on the 1,729 stromal cells (Aibar et al., 2017; Moerman et al., 2018). SCENIC explores gene regulatory networks by identifying TF co-expression modules and binding motif enrichment. The normalised expression matrix generated from Seurat was first filtered for genes as previously described (sum of gene expression $＞ 3 \times 0.005 = 1.729$) (Lambrechts et al., 2018). Genes detected in at least 0.5% of cells were kept. This resulted in 12,100 genes for pySCENIC input (Lambrechts et al., 2018). Analysis was performed using the hg38 mc9nr motif collection with a TSS ± 10 kb (hg38__refseq-r80__1okb_up_and_-down_tss.mc9nr) for the arboreo and RcisTarget steps. Gene regulations were clustered and plotted using the pheatmap function in R.

Flow cytometry and FACS isolation of stromal cells

Cell sorting and flow cytometry experiments were performed at the Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research. Flow cytometry was performed on a Becton Dickinson CantoII or LSRII SORP flow cytometer using BD FACSDIVA software, and the results were analysed using FlowJo software (Treestar Inc.). FACS experiments were performed on a FACS AriaIII sorter using the BD FACS sorter software. All antibody details used in this study can be found in Appendix Table S1. Cryopreserved single-cell suspensions from Patient-4 were thawed, washed with RPMI and incubated with an anti-CD16/CD32 antibody (1:200, BD Biosciences #564220) in FACS buffer (PBS containing salts, 2% FBS) for 10 min to block nonspecific antibody binding. For the isolation of the different stromal subpopulations for subsequent experiments, cells were pelleted and resuspended in FACS buffer containing the following antibodies: anti-EPCAM (1:100; BioLegend #324203), anti-CD31 (1:100; BioLegend #303103), anti-CD45 (1:100; BioLegend #304005), anti-PDGFRb (1:100; BioLegend #323605), anti-PDGFRα (1:100; BioLegend #323507) and anti-CD146 (1:100; BioLegend #342011) for 20 min on ice. All epithelial, immune and endothelial cells were excluded together on the FITC channel marking EPCAM, CD45 and CD31. In addition, we performed positive selection using PDGFRb. CAFs and PVL cells were discriminated using PDGFRα and CD146, respectively. CAFs and PVL cells were isolated and cultured into dishes (Corning® LifeSciences) coated with collagen (0.15 mg/ml) in RPMI 1640 supplemented with 20% (v/v) FBS, 50 µg/ml gentamicin and 1× antibiotic/antimycotic (15-240-096, Gibco®) in a 5% O2, 5% CO2 incubator at 37°C. Cell sorting was repeated on cultured CAFs and PVL cells using the previously described experimental conditions with anti-PDGFRα (1:100; BioLegend #323507), anti-CD146 (1:100; BioLegend #342011), anti-FAP (1:100; R&D Systems #FAB3715P-025), anti-CD90 (1:100; BioLegend #328113) and anti-CD36 (1:100; BioLegend #322011). FAPHIGH expression was used to discriminate myCAFs from FAPLOW iCAFs, while CD36 expression was used to identify imPVL cells from dPVL cells.

Immunofluorescence

Primary cells were grown on glass coverslips coated with collagen in the same manner as the CDMs as described below. Media was removed and cells were rinsed with PBS for 5 min. Cells were fixed in 4% paraformaldehyde (ProSciTech) diluted in PBS for 15 min at room temperature then washed three times with PBS for 5 min. Cells were permeabilised with ice cold methanol for 10 min at −20°C followed by three 5 min PBS washes. Cells were blocked in blocking buffer (3% BSA + 0.1% Tween-20 in PBS) for 1 hr at room temperature. Primary antibody was diluted in blocking buffer at the following dilutions: anti-CD34 (1:100; Abcam #MA1-10202), anti-FAPβ (1:200; Abcam #ab53066), anti-sMA (1:500; Abcam #ab21027), anti-CD36 (1:100; BioLegend #336203), anti-CD146 (1:200; Abcam #ab75769) and anti-PDGFRb (1:250; Abcam #ab32570). Coverslips were inverted and incubated on droplets of
diluted primary antibody on paraffin in a humidified chamber overnight at 4°C. The following day cells were washed three times for 5 min in PBS. Cells were incubated with fluorescent secondary antibody (Jackson ImmunoResearch) diluted 1:500 in blocking buffer for 1 h at room temperature in a light proof container then washed two times with PBS for 5 min. Nuclei were stained with 1 µg/ml Hoechst 33342 (Sigma) in PBS for 5 min at room temperature followed by two 2 min PBS rinses. Coverslips were mounted with Prolong Diamond antifade mountant (Thermo Fisher Scientific) and allowed to dry overnight at room temperature. Fluorescent images were captured using a Leica DMI Sp8 confocal microscope.

Immunofluorescence was performed on 4 µm FFPE tissue sections prepared as described below for IHC. Antigen retrieval was performed for 20 min in a 100°C water bath in target retrieval buffer, pH 9 (Agilent Technologies). Slides were blocked for 1 h at room temperature in PBS containing 3% BSA and 5% goat serum. Slides were incubated with primary antibodies diluted in blocking buffer: anti-CD31 (1:50; Agilent Technologies #M0823) and anti-CD146 (1:600; Abcam #ab75769). Secondary antibody staining, nuclear counterstaining and microscopy were performed as described above.

Quantitative real-time PCR analysis

RNA was extracted from sorted CAF cells using the Qiagen miRNeasy Mini Kit (Qiagen) and was reverse transcribed using the Transcripter First Strand cDNA synthesis kit (Roche). TaqMan probes used to analyse mRNA expression levels were performed for 20 min in a 100 µl reaction mixture, using the QuantStudio 7 Flex RT–PCR machine (Thermo Fisher Scientific). TaqMan probes used were FAP (Hs00990791_m1), ACTA2 (Hs00990791_m1), CXCL12 (Hs00171022_m1), CXCR4 (Hs01076078_M1), PDGFRA (Hs00998018_m1), PDGFRB (Hs01019589_m1) and ACTB (Hs99999903_M1). Relative gene expression was calculated using the ΔΔCt method.

Cell-derived matrices

Cell-derived matrices (CDMs) were established as previously described (Cukierman et al., 2001). A total of 1.5 × 10^5 cells/well were allowed to expand until confluent and ascorbic acid (50 mg/ml) was added to culture medium on days 1, 3 and 5. To maintain the structure interact of the matrix architecture, CDMs were imaged using second harmonic generation (SHG) at day 7 with cells still present in the matrix.

Second harmonic generation imaging

Second harmonic generation (SHG) imaging was achieved using an inverted Leica DMS 6000 SP8 confocal microscope with a Ti-Sapphire femtosecond laser cavity (Coherent Chameleon Ultra II) excitation source, operating at 80 MHz and tuned to a wavelength of 880 nm, as previously described (Timpson et al., 2011; Conway et al., 2017; Vennin et al., 2017). SHG intensity was detected using a 440/20 nm RLD HyD detectors. For CDMs, three representative fields of view (512 µm × 512 µm) were imaged over a 3D z-stack (80 µm with a 2.52 µm step size, and 30 µm with a 1.51 µm step size, respectively), with a line average of 4 at 25× magnification. Rotation images were acquired on the z-level of maximum intensity with a line average of 64 at 63× magnification.

Collagen fibre orientation analysis

Collagen fibre orientation analysis in SHG images from CDMs was carried out as previously described (Mayorca-Guiliani et al., 2017; Cazet et al., 2018). Briefly, the distribution of orientation of collagen within images was assessed based on methodology published by Rezakhanlou et al (2012). The local orientation and isotropic properties of individual pixels making up collagen fibres were derived from structure tensors evaluated by computing the continuous spatial derivatives in the x and y directions using a cubic B-spline interpolation to obtain the local predominant orientation. Graphical outputs show a hue–saturation–brightness (HSB) colour-coded map indicating the angles of the oriented structures within the image. Orientation distribution peaks were then aligned. The shape of the distribution indicates the degree of alignment within the image, where wide and broad shapes suggested little coherency in alignment, and tight peaks with small standard deviations implied aligned structures.

Immunohistochemistry and image alignment

In-house FFPE blocks were made of patient tissues by fixing in 10% neutral buffered formalin for 24 h and processing for paraffin embedding. Where tissue was limited, diagnostic tumour FFPE blocks were accessed for analysis. FFPE blocks were sectioned at 4 µm. These were used for histological analysis, using a standard haematoxylin and eosin stain, and for immunohistochemical analysis on the Leica BOND RX Autostainer. Details of antibodies and staining conditions are described in Appendix Table S1. H&E and IHC slides were imaged using the Aperio CS2 Digital Pathology Slide Scanner. IHC images were imported into FIJI as a virtual stack. Each layer was then aligned using least squared mode (linear feature correspondences), propagating to the first and last layers for rigid transformation. All other parameters were set to default in FIJI.

Cell signalling predictions using ligand–receptor annotation

Genes from the scRNA-Seq data were annotated based on a published set of human ligand–receptor pairs derived from supporting literature (Ramilowski et al., 2015). We used this knowledge to construct a cell-to-cell communication network between the four stromal clusters and other epithelial, immune and endothelial clusters. To investigate conserved signalling modules in TNBCs, we applied this to the cluster averaged expression levels of all ligands and receptors in the integrated dataset of five patients. The “interaction strength,” or the weight of edges between two clusters, was defined as the product of expression values from the ligand and its cognate receptor. All “interaction strengths” greater than an arbitrary cut-off of 0.1 were considered as cell signalling candidates and kept for subsequent analyses (Dataset EV3). The total number of interaction pairs identified per cluster was used to generate summaries of this data (Fig 5A and B). The top 100 candidates between the four stromal subsets and each target population were clustered using hierarchical clustering (complete and Euclidean distance) and rescaled for visualisation in ggplot2. For the visualisation purposes only, the ligand and receptor expression values in Fig 5C–E were imputed using the MAGIC method to better represent the structure of genes with low expression and dropout (van Dijk et al., 2018).
Raw count matrices and cluster IDs identified by Seurat (as previously described) were used as input to MAGIC and run with default parameters as recommended by the developers.

**T-cell dysfunction and exclusion analysis**

To investigate the immunomodulatory roles of different stromal subsets, we performed T-cell dysfunction and exclusion analysis using similar strategy from TIDE (Jiang et al., 2018). We first used the average expression level of CD8A, CD8B, GZMA, GZMB and PRF1 to estimate the cytotoxic T-lymphocyte (CTL) level in each sample from the bulk sequencing cohort. Patients with a higher and lower CTL level compared to the mean CTL level within the cohort were stratified into high and low CTL groups, respectively. For CTL dysfunction analysis, TIDE evaluates whether gene signatures from each of the stromal subsets influences the beneficial effect of CTL levels on patient prognosis. This is performed using the interaction coefficient $d$ from Cox proportional hazard (Cox-PH) model to evaluate how the interaction between a candidate gene and the CTL affects the death hazard. Genes with a higher TIDE dysfunction score suggest antagonistic interactions with regard to CTL levels, where the survival benefit of patients with high CTL is lost, and thus suggesting an association with CTL dysfunction. This method was used to calculate the TIDE T-cell dysfunction score from the differentially expressed genes across the four stromal subsets in TNBC patients from the METABRIC cohort (Curtis et al., 2012) and two independent TNBC cohorts (Sabatier et al., 2011; Jezquel et al., 2015). A total of 233, 84 and 107 patients were evaluated for the METABRIC, GSE21653 and GSE38812 cohorts, respectively.

For T-cell exclusion analysis, we examined Pearson correlations between all CTL levels (indicated on the $y$ axis in Fig 6E) and the respective correlation score between the bulk tumour sample and single-cell cluster of interest (indicated on the $x$ axis in Fig 6E). Here, gene signatures for the single-cell cluster of interest were defined by the averaged gene expression of all single cells in the cluster, divided over the averaged gene expression of all cells detected in the dataset. This method was used to define signatures in this section, as opposed to a DEG list in the previous CTL dysfunction analysis. This was first performed for all stromal cells, CD4 + and CD8 + T-cell clusters divided over all detected cells independently, as shown in Fig 6D. We next repeated this for the myCAF, iCAF, dPVL and imPVL clusters divided over all detected cells independently, as shown in Fig 6E. In each of the breast cancer cohorts, a higher correlation suggests a positive association between the single-cell cluster of interest and CTL levels (as observed in CD4 + and CD8 + T cells shown in Fig 6D), while a negative correlation suggest a negative association (as observed in dPVL cells shown in Fig 6E). This correlation indicates a potential afluence of each stromal subset on T-cell infiltration in tumours. For T-cell exclusion analysis, we examined the three aforementioned TNBC cohorts, as well as the TNBC cohort from The Cancer Genome Atlas (https://www.cancer.gov/tcga) (Cancer Genome Atlas, 2012).

**Statistical analysis**

For the identification of significant differentially expressed genes in scRNA-Seq, the MAST method was applied with a $P$-value threshold of $1 \times 10^{-5}$ and FDR threshold of 0.05 (Finak et al., 2015). Comparisons for enriched AUCell regulon scores from the SCENIC pipeline were made using a one-way analysis of variance (ANOVA) with a $P$-value threshold of $1 \times 10^{-5}$ (Aiabar et al., 2017; Moerman et al., 2018). Quantitative analysis of collagen abundance from SHG was determined using unpaired two-tailed Student’s $t$-test with equal standard deviation after normalisation of the orientation peak distributions. $P$-values were denoted by asterisks: *$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$. For collagen orientation, statistical significant was determined using a Kruskal–Wallis test with Dunn’s post hoc multiple comparisons test with a $P$-value threshold of $1 \times 10^{-4}$. For TIDE analysis of CTL dysfunction, $P$-values were derived from a two-sided $t$-test for each correlation. The Benjamini–Hochberg procedure was used for adjusting $P$-values to correct for multiple testing. Statistical significance of CDB IHC staining was determined using pairwise comparison of each patient against all others (i.e., base-mean) using a Student’s $t$-test.

**Data availability**

The scRNA-Seq data from this study have been deposited in the European Nucleotide Archive (ENA) under the accession code PRJEB35405 (http://www.ebi.ac.uk/ena/data/view/PRJEB35405). This depository includes the demultiplexed paired ended reads (R1 and R2), Illumina indices and bam files processed using the Cellranger software. The scRNA-Seq analysis scripts can be found on the website: https://github.com/sunnyzwu/stromal_subclasses. All relevant data are available from the authors upon request.

**Expanded View** for this article is available online.

**Acknowledgements**

This work was supported by funding from John and Deborah McMurtrie, the National Breast Cancer Foundation (NBCF) of Australia, The Petre Foundation and The Sydney Breast Cancer Foundation. A.S. is the recipient of a Senior Research Fellowship from the National Health and Medical Research Council of Australia. S.Z.W is supported by the Australian Government Research Training Program Scholarship. S.O.T. is supported by the NBCF (PRAC 16-006), the IIRS 19 084 and the Sydney Breast Cancer Foundation and the Family and Friends of Michael O'Sullivan. T.R.C. is supported by an NHMRC RD Wright Biomedical Career Development Fellowship, a Susan G Komen Career Catalyst Award and a Cancer Institute NSW (CINSW) fellowship. S.J. is supported by a research fellowship from the NBCF. X.S.L. is supported by the Breast Cancer Research Foundation (BCRF-19-100) and the National Institute of Health of the United States (RO1CA234018). We would like to thanks the following people for their assistance in the experimental part of this manuscript; Ms. Gillian Lehrbach from the Garvan Tissue Culture Facility; Ms. Anaïs Zaratzian from the Garvan Histopathology Facility for tissue processing and IHC staining; The Garvan-Weizmann Centre for Cellular Genomics, including Mr. Eric Lam, Ms. Hira Saeed and Ms. Melissa Armstrong for the expertise in flow sorting, and Mr. Dominik Kaczorowski for his help in next-generation sequencing.

**Author contributions**

AS conceived the project and directed the study with input from all authors. SZW and AS wrote the manuscript with input from all authors. Clinical
The Authors declare that they have no conflict of interest.

**Conflict of interest**
The authors declare that they have no conflict of interest.

**References**


Tamura K, Barbeau DJ (2012) Early growth response genes signaling supports strong paracrine capability of mesenchymal stem cells. Stem Cells Int 2012: 428403


