



# Single-cell RNA sequencing reveals the effects of chemotherapy on human pancreatic adenocarcinoma and its tumor microenvironment

Received: 2 June 2022

Accepted: 25 January 2023

Published online: 13 February 2023

Check for updates

Gregor Werba , Daniel Weissinger<sup>2,7</sup>, Emily A. Kawaler<sup>2,7</sup>, Ende Zhao<sup>2</sup>, Despoina Kalfakakou , Surajit Dhara<sup>2</sup>, Lidong Wang<sup>2</sup>, Heather B. Lim<sup>2</sup>, Grace Oh<sup>1,2</sup>, Xiaohong Jing<sup>2</sup>, Nina Beri<sup>2,3</sup>, Lauren Khanna<sup>3</sup>, Tamas Gonda<sup>3</sup>, Paul Oberstein , Cristina Hajdu<sup>4</sup>, Cynthia Loomis<sup>4</sup>, Adriana Heguy , Mara H. Sherman , Amanda W. Lund , Theodore H. Welling<sup>1</sup>, Igor Dolgalev , Aristotelis Tsirigos & Diane M. Simeone

The tumor microenvironment (TME) in pancreatic ductal adenocarcinoma (PDAC) is a complex ecosystem that drives tumor progression; however, in-depth single cell characterization of the PDAC TME and its role in response to therapy is lacking. Here, we perform single-cell RNA sequencing on freshly collected human PDAC samples either before or after chemotherapy. Overall, we find a heterogeneous mixture of basal and classical cancer cell subtypes, along with distinct cancer-associated fibroblast and macrophage subpopulations. Strikingly, classical and basal-like cancer cells exhibit similar transcriptional responses to chemotherapy and do not demonstrate a shift towards a basal-like transcriptional program among treated samples. We observe decreased ligand-receptor interactions in treated samples, particularly between TIGIT on *CD8*+ T cells and its receptor on cancer cells, and identify TIGIT as the major inhibitory checkpoint molecule of *CD8*+ T cells. Our results suggest that chemotherapy profoundly impacts the PDAC TME and may promote resistance to immunotherapy.

Pancreatic ductal adenocarcinoma (PDAC) is a highly lethal cancer with a five-year survival rate of approximately 11%. With increasing incidence, it is projected to become the second leading cause of cancer-related deaths in the United States by 2030. Current treatment options are limited. Only 15% to 20% of patients qualify for upfront surgery with curative intent; the remaining patients present with unresectable locally advanced disease or distant metastases. Nearly all

patients receive adjuvant, neoadjuvant, or palliative chemotherapy<sup>3</sup>. Recent advances in immunomodulating drugs, such as checkpoint inhibitors, have yet to show encouraging results in PDAC<sup>4,5</sup>. A deeper understanding of pancreatic cancer biology is needed to develop better therapeutics and improve outcomes.

Although recent genomic and transcriptomic studies on bulk tumor samples identified key molecular drivers (notably *KRAS*,

<sup>1</sup>Department of Surgery, NYU Langone Health, New York, New York 10016, USA. <sup>2</sup>Perlmutter Cancer Center, NYU Langone Health, New York, New York 10016, USA. <sup>3</sup>Department of Medicine, NYU Langone Health, New York, New York 10016, USA. <sup>4</sup>Department of Pathology, NYU Langone Health, New York, New York 10016, USA. <sup>5</sup>Department of Cell, Developmental and Cancer Biology, Oregon Health Sciences University, Portland, Oregon 97239, USA. <sup>6</sup>Department of Dermatology, NYU Langone Health, New York, New York 10016, USA. <sup>7</sup>These authors contributed equally: Gregor Werba, Daniel Weissinger, Emily A. Kawaler. e-mail: [Aristotelis.Tsirigos@nyulangone.org](mailto:Aristotelis.Tsirigos@nyulangone.org); [Diane.Simeone@nyulangone.org](mailto:Diane.Simeone@nyulangone.org)

*TP53*, *CDKN2A*, and *SMAD4*) and cancer cell subtypes (basal and classical), these findings have yet to translate into treatment advances<sup>6–9</sup>. Recent investigations into the PDAC tumor microenvironment (TME) revealed that this complex ecosystem is a critical mediator of therapeutic resistance and tumor progression<sup>10,11</sup>. Single-cell RNA sequencing (scRNA-seq) has emerged as a valuable tool in precision medicine by enabling in-depth characterization of the TME and serial assessment on scant amounts of tissue<sup>12–17</sup>. While recent single-cell studies have provided insights into PDAC, there is an unmet need to characterize the effects of chemotherapy on the TME<sup>18–22</sup>.

For a comprehensive view of the PDAC TME, and its response to chemotherapy in a real-time clinical context, here we perform scRNA-seq on freshly collected tumor specimens from 27 PDAC patients. We observe significant heterogeneity in the malignant epithelial compartment, with most tumors displaying a mixture of basal-like and classical-like cells. Furthermore, we examine the composition of CAFs and discrete groups of immune cells in the TME and execute comparative analyses on naive and chemotherapy-treated samples, which we confirm in an external dataset. Our data show that chemotherapy profoundly alters the PDAC TME, inducing features consistent with tumor progression and refractoriness to immunotherapy by down-regulation of inhibitory checkpoint molecules in CD8+ T cells.

## Results

### Single-cell analysis reveals the transcriptomic landscape in PDAC

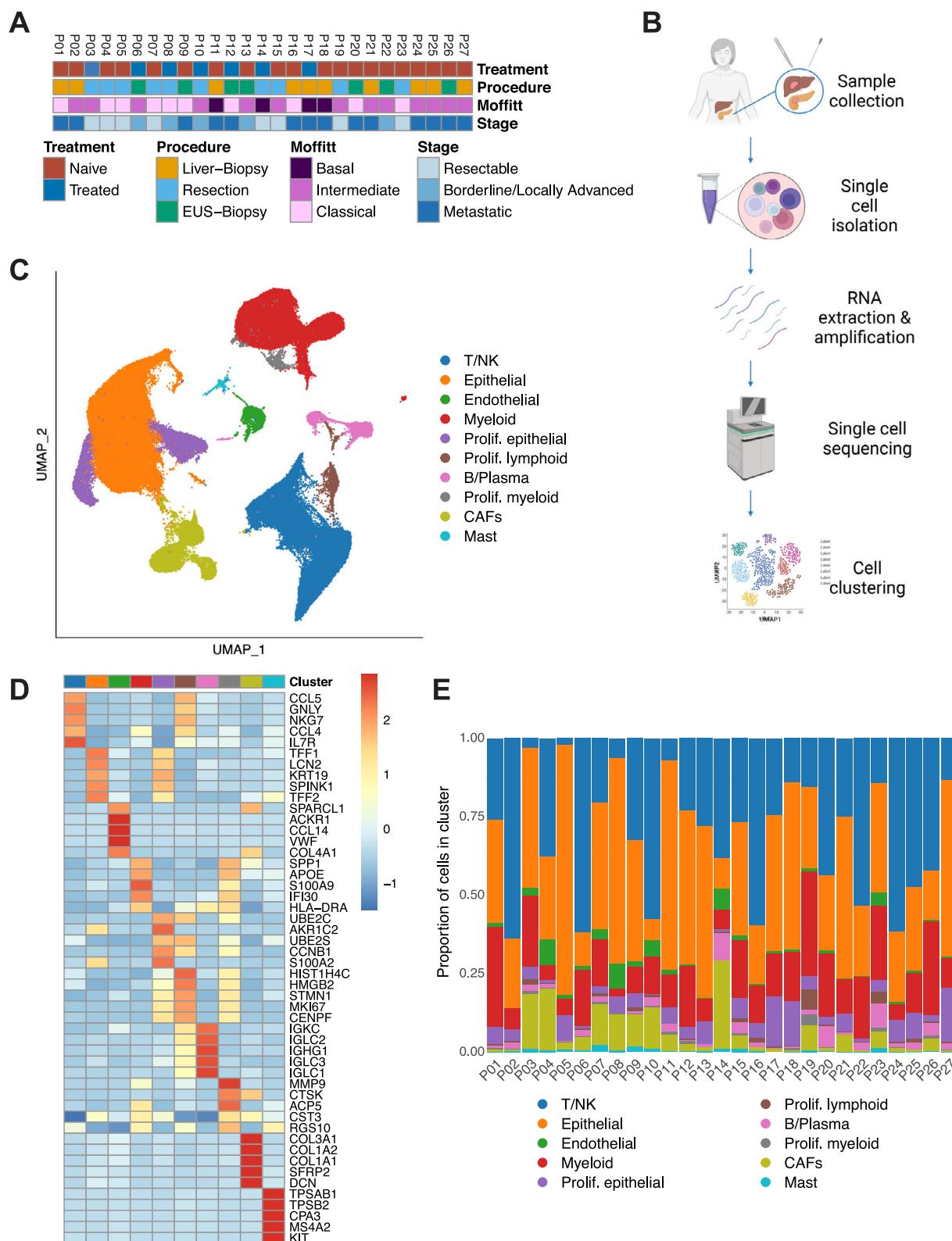
PDAC specimens were acquired via surgical resection of primary pancreatic lesions ( $n=10$ ) or by endoscopic/interventional radiology-guided biopsies of either primary pancreatic lesions ( $n=7$ ) or liver metastases ( $n=10$ ) (Fig. 1A, Supplementary Table 1). Of the 27 total patients, seven received chemotherapy (FOLFIRINOX-based  $n=4$ , gemcitabine/abraxane-based  $n=3$ ) prior to tissue collection, while 20 patients had not been treated at the time of specimen acquisition (Fig. 1A, Supplementary Table 1). Evaluating the treated patients using the RECIST criteria, we found that two had stable disease, two partial response, three progressive disease, and one was nonevaluable (Supplementary Table 2). We prepared single-cell suspensions from tumor specimens and sequenced them using the 10x Genomics Chromium system (Fig. 1B, see Methods for details). Histopathological assessment was done on each tumor specimen to verify malignancy, and samples underwent mutational profiling (Supplementary Data 1, 2). After standard processing and quality filtering of the raw sequencing data (see Methods), a total of 139,446 cells were retained for analysis. Unsupervised, graph-based clustering and visualization via uniform manifold approximation and projection (UMAP) revealed ten distinct cell clusters that were annotated using canonical markers and gene expression profiles (Fig. 1C, D, Supplementary Fig. 1A, Methods). Three major clusters were identified as epithelial, T/natural killer (NK), and myeloid cells. We further identified CAFs, endothelial cells, B/plasma cells, and a small mast cell population among our clusters. The proportional distribution of these major cell types was similar across all groups, with the exception of biopsies from liver metastases and primary PDAC which had quantitatively lower CAFs and endothelial cells than resection samples. We also identified an elevated amount of proliferating cancer cells in liver biopsies (Fig. 1E; Supplementary Fig. 1B). To ensure that tissue processing did not significantly change gene expression in our samples, we applied hypoxia and dissociation signatures<sup>23,24</sup> to our dataset. This analysis showed that all scSeq samples contained very few hypoxic cells, with no significant differences between procedures (Supplementary Fig. 1C), and demonstrated low dissociation signatures independent of procedure (Supplementary Fig. 1D).

### The epithelial compartment reveals heterogeneous malignant subtype composition

The epithelial compartment evinced sample-specific differences while the mesenchymal and immune compartments clustered together across all samples (Supplementary Fig. 2A). This suggests that non-epithelial cells of the PDAC TME share more common transcriptional signatures across all patients than their malignant counterparts within the same tumor. Since some epithelial cells also clustered across samples, we further assessed the proportion of malignant cells in the total epithelial compartment (Fig. 2A). To this end, we performed copy number variation (CNV) analysis. The vast majority of epithelial cells showed several CNV events, indicating malignancy and distinguishing them from normal epithelial cells (Fig. 2B, Supplementary Fig. 2B). We further classified malignant cells as either basal or classical based on their Moffitt subtype signature expression<sup>7</sup>. (Any signature gene sets referenced in this paper can be found in Supplementary Data 3). All samples but one were comprised of a mixture of basal and classical cancer cells in varying proportions (Fig. 2C, D). Tumor subtype composition did not correlate with the amount of any major non-malignant cell types (Supplementary Fig. 2C).

We validated the Moffitt transcriptional subtypes on formalin-fixed, paraffin-embedded tissue (FFPE) sections using multiplex immunohistochemistry (IHC). We used the canonical markers cytokeratin 17 (CK17) for the basal subtype, GATA6 for the classical subtype, and cytokeratin 19 (CK19) as a pan-cancer cell marker. Representative samples with a higher percentage of basal or classical transcriptional subtypes were strongly positive for CK17 or GATA6, respectively (Fig. 2E, upper and lower panels), while a sample with a more even mixture of both subtypes showed an interspersed, more equal expression of GATA6 and CK17 (Fig. 2E, middle panels). Another patient sample with a mixture of both subtypes showed two distinct growth patterns with a CK17<sup>high</sup> region next to a GATA6<sup>high</sup> region (Supplementary Fig. 2D). In general, GATA6 staining was more omnipresent, and CK17<sup>high</sup> cells often stained positive for GATA6. Of note, most samples included cells staining positive for both CK17 and GATA6 (Supplementary Fig. 2E), consistent with a previous report that hypothesized an intermediate cell state<sup>25</sup>.

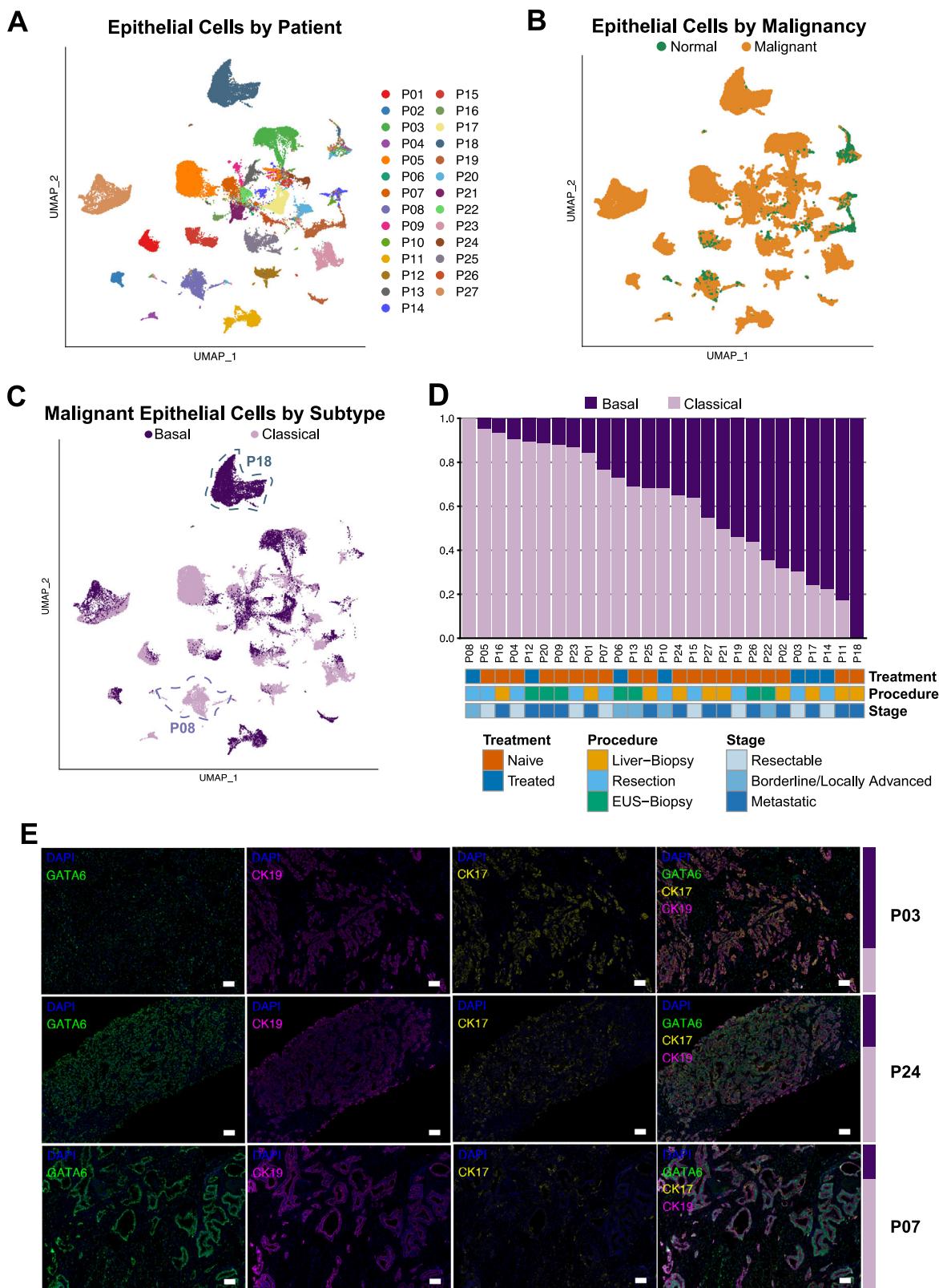
To test this intermediate cell state hypothesis in our dataset, we applied the basal, classical, and intermediate gene signatures from that report, which were derived from treatment-naive liver metastases, to malignant cells from our treatment-naive PDAC liver metastatic samples. Although their basal and classical signatures matched closely with ours, we did not see clear evidence of the intermediate signature (Supplementary Fig. 2F). We also assessed the bulk RNA-derived subtype classifications of Collisson, Bailey and Puleo<sup>26–28</sup>. In our dataset, the Moffitt-basal signature showed overlap with the Collisson-quasi-mesenchymal and Puleo-pure-basal-like, while the Bailey-squamous signature did not show a strong correlation with the basal signatures. Similarly, the Moffitt-classical signature overlapped with the Collisson-classical and Puleo-pure-classical signatures, while the Bailey-progenitor signature did not show a strong correlation with the classical signatures. The less commonly found subtype signatures of Bailey (ADEX and immunogenic), Collisson (exocrine-like), and Puleo (desmoplastic) did not show strong expression in our dataset of CNV-confirmed cancer cells, suggesting that these might be derived from non-epithelial cells within tumor samples (Supplementary Fig. 2G, H). These findings are supported by a TCGA publication from 2017 which examined sample purity and subtype classifications<sup>9</sup>. Further characterization of basal and classical cells by differential gene expression showed typical basal and classical genes upregulated in the respective subtypes (Supplementary Fig. 2I). Pathway enrichment analysis revealed enrichment of EMT, angiogenesis, and extracellular matrix (ECM) interaction—features related to migration and invasion—in basal cells, while classical cells were enriched for metabolic pathways, especially lipid metabolism (Supplementary Fig. 2J). These results emphasize the heterogeneity of malignant epithelial cells in PDAC and

**Fig. 1 | Single-cell analysis reveals the transcriptomic landscape in PDAC.**

A Clinicopathologic characteristics for each patient sample including treatment status, stage, Moffitt subtype, and procedure. EUS endoscopic ultrasound.

B Workflow depicting sample acquisition, processing, and analysis. Designed with BioRender©. C UMAP, embedding of the expression profile of 139,446 cells that passed quality control from  $n = 27$  patient samples. Distinct clusters are annotated

and color-coded. CAF cancer-associated fibroblast, NK natural killer, UMAP uniform manifold approximation and projection. D Heatmap of the five most differentially expressed genes within each cluster. Colors correspond to the cluster colors in Fig. 1C. E Proportional distribution of each cell cluster by individual sample.



**Fig. 2 | The epithelial compartment reveals heterogeneous malignant subtype composition.** **A** Non-batch corrected UMAP of all epithelial cells, showing clustering by patient sample of most epithelial cells and clustering of some epithelial cells across samples. UMAP uniform manifold approximation and projection. **B** Non-batch corrected UMAP of malignant and normal epithelial cells as determined by InferCNV. **C** UMAP embedding of malignant epithelial cells labeled with Moffitt subtypes. Exemplary samples high in basal (P18) and classical (P08) cells are annotated. **D** Proportion of cells of each Moffitt subtype by sample. EUS

endoscopic ultrasound. **E** Representative ( $n = 27$ ) images of multiplex immunofluorescence across the subtype spectrum: higher in basal (upper panel, P03), mixed (middle panel, P24), and higher in classical (lower panel, P07) subtype tumors. Basal-to-classical ratio for each sample is shown on the right as a colored bar (dark = basal, light = classical). Channels (always including DAPI (blue)): GATA6 (green), CK19 (cytokeratin 19) (violet), CK17 (cytokeratin 17) (yellow), and merged. Scale bar = 100  $\mu$ m.

suggest that single-marker analysis may be insufficient to accurately identify cancer cell subtypes within a patient's tumor.

### Charting the CAF landscape in the PDAC tumor microenvironment

We next analyzed the mesenchymal compartment of the PDAC TME. Mesenchymal cell reclustering revealed seven distinct clusters (Fig. 3A). The two main clusters consisted of myofibroblastic CAFs (myCAFs) and inflammatory CAFs (iCAFs) (Fig. 3A, B)<sup>29</sup>. We did not find a strong antigen-presenting CAF (apCAF) signature (Fig. 3C). Notably, the apCAF signature was originally derived from mouse CAFs and may not translate to the human CAFs in our cohort. As apCAF features in human CAFs have been linked to co-expression of *CD74* and human leukocyte antigen (HLA) molecules, we checked for expression of these genes in our CAF populations. A small number of cells in both the iCAF and myCAF subpopulations co-expressed *CD74* and HLA genes (Fig. 3C, Supplementary Fig. 3A)<sup>19</sup>. Other clusters were identified as pericytes and Schwann cells, while a single sample provided a cluster of chondrocyte-like cells. Additionally, we detected a small proportion of cells expressing epithelial markers (Fig. 3A, B). We ruled out these cells being cancer cells undergoing EMT due to a lack of CNVs, low expression of EMT-related transcription factors (TFs) *SNAIL1* and *TWIST1*, and low values of an EMT-TF independent partial EMT (pEMT) program signature<sup>30</sup> (Supplementary Fig. 3B). These cells were also not consistent with mesothelial cells, as mesothelin (*MSLN*) and other markers for mesothelial cells were minimally expressed; we therefore labeled this cluster "epithelial-like" (Supplementary Fig. 3C).

The myCAF to iCAF ratio varied across samples, while the proportional distribution of both subpopulations in relation to Moffitt subtype distribution was not significantly different (Fig. 3D). In myCAFs, *SDCI*<sup>31</sup>, a gene linked to CAF-mediated regulation of breast cancer cell migration, was upregulated, while in iCAFs, *CCL19*<sup>32</sup>, a gene linked to cytotoxic T cell recruitment by CAFs in lung cancer, was highly expressed (Supplementary Fig. 3D, Supplementary Data 4). Additionally, iCAFs showed enrichment for genes involved in the complement cascade and EGF signaling, whereas myCAFs demonstrated enrichment for ECM organization and smooth muscle contraction, supporting the notion of CAF heterogeneity and distinct functional characteristics of CAFs in the TME (Fig. 3E).

### Charting the T/NK landscape in the PDAC tumor microenvironment

To better characterize tumor-infiltrating lymphocytes, we analyzed the T/NK population and identified 13 distinct clusters (Fig. 4A). *CD4*+ T cells clustered into four distinct subtypes: (i) *CCR7*+ *CD4*+ (high expression of naïve/central memory markers), (ii) *IL7R*+ *CD4*+ (high expression of activation markers), (iii) *FOXP3*+ *CD4*+ (high expression of regulatory T cell genes), and (iv) *CXCL13*+ *CD4*+ (high expression of follicular helper genes) (Fig. 4B). *CD8*+ T cells clustered into three main subpopulations: (i) *GZMH*+ *CD8*+ (high in cytotoxic gene expression), (ii) *GZMK*+ *CD8*+ (expressing effector memory genes), and (iii) *ITGA1*+ *CD8*+ (expression of tissue residency and exhaustion markers). We also identified a small *ISG15*+ cluster, shared between *CD4*+ and *CD8*+ T cells, with high interferon (IFN)-related gene expression. NK cells separated into a *GNLY*+ cluster highly expressing cytotoxic genes and an *XCL1*+ cluster with lower cytotoxic gene expression and higher expression of genes previously linked to tissue-resident NK cells<sup>33,34</sup>. Smaller clusters were identified as mast cells and plasma cells (Fig. 4A, B). T/NK subpopulations were heterogeneously distributed across samples and did not correlate with basal/classical subtyping (Supplementary Fig. 4A, B).

Because our initial T/NK clustering did not identify distinct exhausted T cell subpopulations, we refined the resolution of the *CD8*+ T cell clustering (Fig. 4C). At this finer granularity, we discerned a small *CTLA4*+ *CD8*+ T cell cluster with high levels of exhaustion

markers (*CTLA4*, *HAVCR2*, *LAG3*, *TIGIT*). Additional distinct *CD8*+ T cell clusters included (i) *GZMK*+ *CD8*+ (high expression of effector memory genes), (ii) *GZMH*+ *CD8*+ (high expression of cytotoxic genes), (iii) *ITGA1*+ *CD8*+ (high expression of tissue residency genes), (iv) *SELL*+ *CD8*+ (expression of naïve/central memory genes), and (v) *ISG15*+ *CD8*+ (expressing IFN activation genes) (Fig. 4C, D). We also found an epithelial-like cluster, an erythrocyte-like cluster, and an *HSPA6*+ cluster; the latter was mainly derived from one patient and did not show other distinct *CD8*+ subpopulation features (Fig. 4C, D).

To further assess exhaustion in our *CD8*+ T cell subpopulations, we scored each subset on a dysfunction and cytotoxicity scale<sup>35</sup>. As expected, *GZMH*+ *CD8*+ T cells scored highest in cytotoxicity, and *CTLA4*+ *CD8*+ T cells highest in exhaustion (Fig. 4E). Except for *CTLA4*+ *CD8*+ T cells, the exhaustion state was similar between subsets (Fig. 4E). Exhausted *CD8*+ T cells represented only 2.8% of all *CD8*+ T cells (Fig. 4F). Moreover, exhausted T cells, along with most other *CD8*+ T cell subsets, were heterogeneously distributed across samples, independent of the Moffitt subtype composition of their epithelial compartments (Supplementary Fig. 4C). Of note, *ITGA1*+ *CD8*+ T cells were the only subpopulation that was significantly negatively correlated with the basal subtype (Supplementary Fig. 4D).

### *C1QC*+ and *SPP1*+ TAMs comprise two distinct subpopulations within the myeloid cell compartment

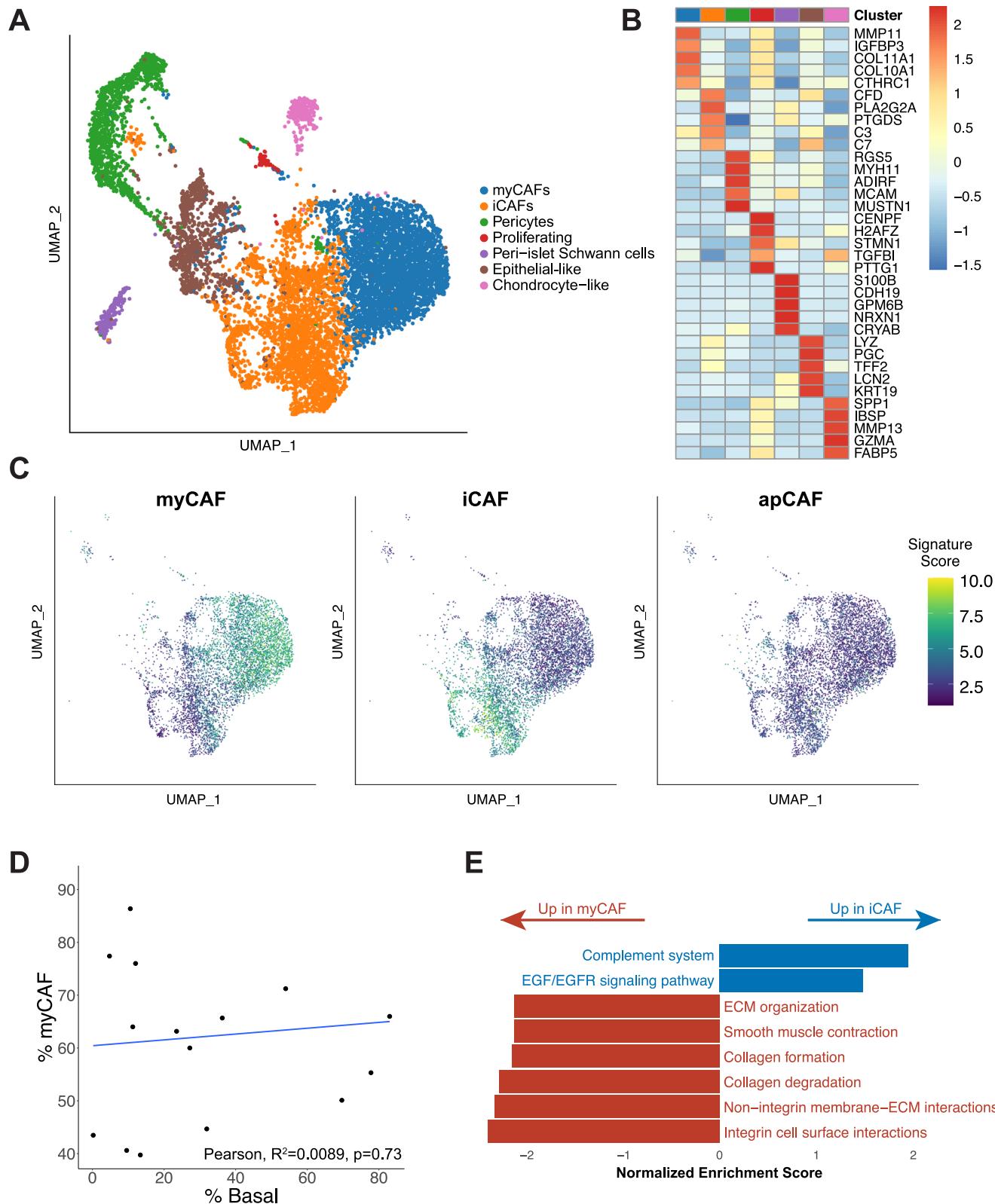
Myeloid cells are an important component of the larger TME. We identified twelve distinct myeloid clusters (Fig. 5A). TAMs, representing the majority of myeloid cells, clustered into *SPP1*+ and *C1QC*+ subpopulations (Fig. 5A, B). We also identified myeloid-derived suppressor cells (MDSCs), monocytes, and dendritic cells (DCs), which further clustered into conventional DCs 1-3 (cDC1-3) and plasmacytoid DCs (pDCs) (Fig. 5A, B). We did not detect any trends in myeloid composition across samples or in relation to Moffitt subtypes (Supplementary Fig. 5A, B).

To better understand what distinguishes *SPP1*+ and *C1QC*+ TAMs, we further characterized these two populations. Historically, TAMs have been classified as either M1 (pro-inflammatory) or M2 (pro-tumorigenic), so we assessed M1 and M2 signature expression in our TAMs<sup>13</sup>. While *C1QC*+ TAMs showed higher M1 signature expression than *SPP1*+ TAMs, both *SPP1*+ and *C1QC*+ subsets displayed strong M2 signature expression and therefore could not be clearly distinguished based on the M1/M2 classification (Fig. 5C). *SPP1*+ TAMs showed increased expression of *S100A8* and *S100A9*, genes previously linked to immunosuppressive macrophages<sup>36</sup>, while *C1QC*+ TAMs had increased expression of *HLA-DRB5* and *CXCL9*, which is involved in the anticancer immune response<sup>37</sup> (Supplementary Fig. 5C, Supplementary Data 4). In general, *SPP1*+ TAMs were enriched for genes involved in EMT, metabolism of carbohydrates, and hypoxia, while *C1QC*+ TAMs showed enrichment for IFN signaling and antigen presentation (Fig. 5D). Consistent with the gene set enrichment analysis (GSEA) results and a recent colon cancer study<sup>38</sup>, *SPP1*+ TAMs were enriched for angiogenesis-related genes and *C1QC*+ TAMs for phagocytosis (Fig. 5E).

To estimate whether the TAM subtype may affect overall survival in PDAC, we evaluated *SPP1*+ and *C1QC*+ TAM signature expression in the Cancer Genome Atlas (TCGA) PDAC bulk tumor dataset<sup>9</sup>. While a similar analysis of the colorectal cancer (CRC) TCGA dataset implied worse overall survival for patients with a *SPP1*<sup>high</sup>/*C1QC*<sup>low</sup> TAM combination<sup>38</sup>, we found that estimated TAM composition did not correlate with overall survival in the TCGA PDAC data (Supplementary Fig. 5D).

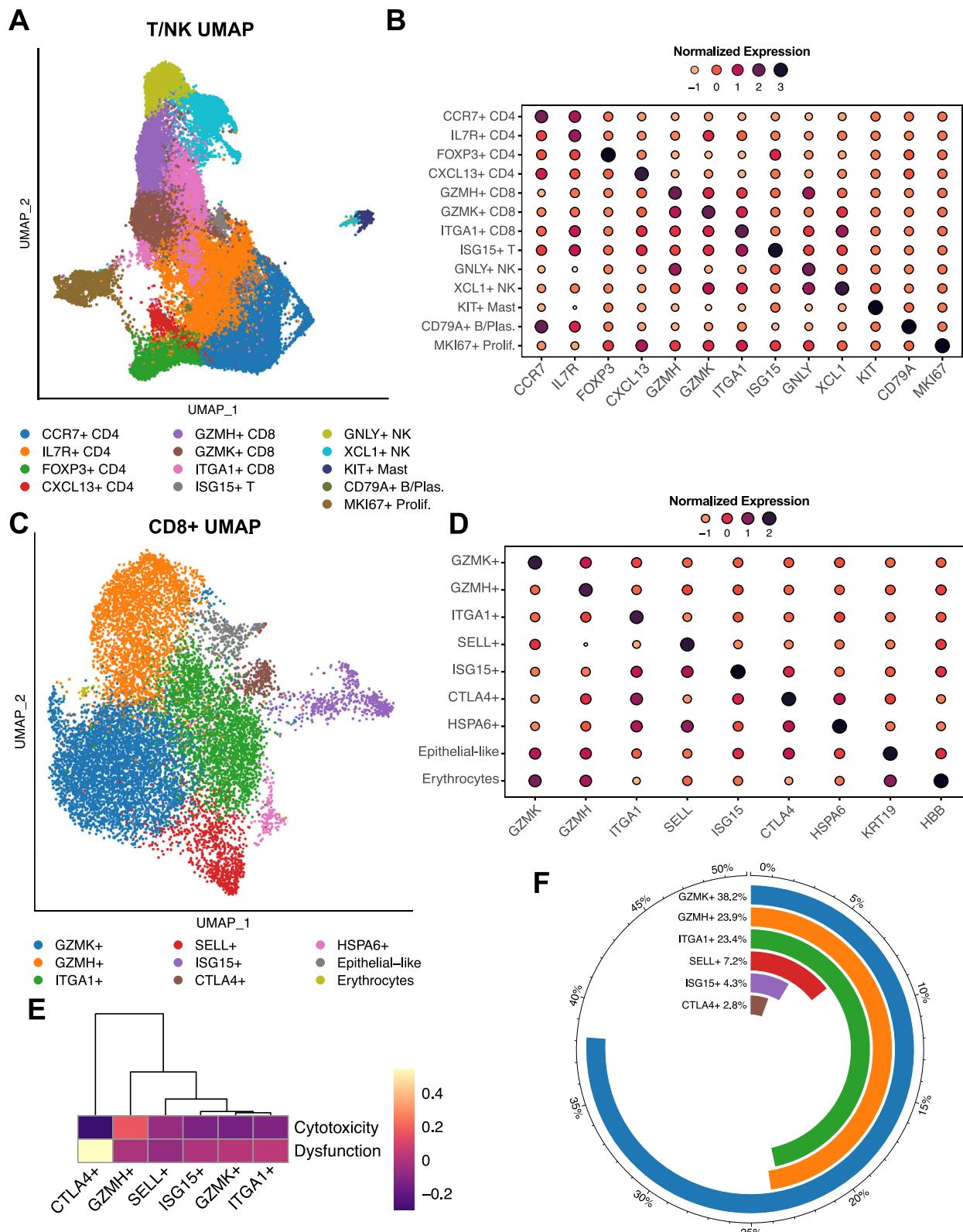
### Chemotherapy induces transcriptional changes in cancer cells independent of subtype

To analyze the effect of chemotherapy on the PDAC TME, we focused on primary pancreatic tumor specimens collected before treatment (treatment-naïve,  $n=11$ ) and after treatment ( $n=6$ ). Specimens



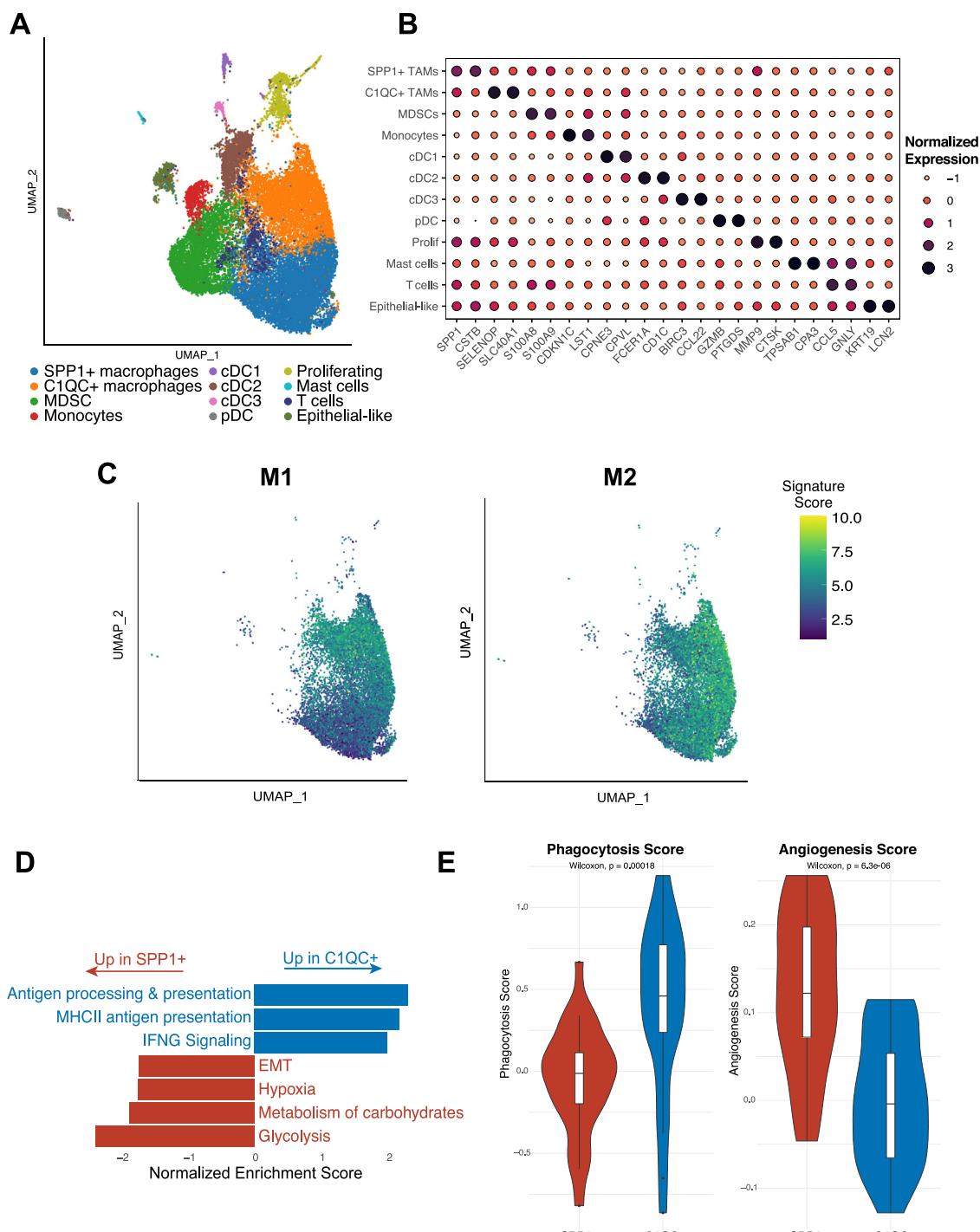
**Fig. 3 | Charting the CAF landscape in the PDAC tumor microenvironment.** A UMAP overview of the mesenchymal compartment. CAF cancer-associated fibroblast, iCAF inflammatory CAF, myCAF myofibroblastic CAF, UMAP uniform manifold approximation and projection. B Heatmap of the five most differentially expressed genes within each cluster. Colors correspond to the cluster colors in panel (A, C). iCAF, myCAF, and apCAF signatures from Elyada et al.<sup>19</sup> overlaid on the UMAP embedding of our mesenchymal compartment. The iCAF and myCAF signatures are derived from human fibroblasts; the apCAF signature is derived from KPC mouse fibroblasts. apCAF antigen-presenting CAF. D Assessment of

correlation (two-sided Pearson correlation) between myCAF/iCAF composition and basal/classical composition among our samples ( $n=16$ ). Samples are plotted as a function of the percentage of their CAF cells that are myCAFs and the percentage of their malignant epithelial cells that are basal. Samples with fewer than 40 total CAF cells were excluded; all but one of these samples were biopsies, which were not expected to pick up many mesenchymal cells. E Selected gene set enrichment analysis results from a comparison between the cells in the iCAF and myCAF compartments. ECM extracellular matrix.



**Fig. 4 | Charting the T/NK landscape in the PDAC tumor microenvironment.** A UMAP overview of the T/NK compartment. NK natural killer, UMAP uniform manifold approximation and projection. **B** The most differentially expressed gene within each T/NK cluster. **C** UMAP of the more granular CD8+ T cell reclustering

reveals additional subpopulations. **D** The most differentially expressed gene within each CD8+ T cell cluster. **E** Hierarchical clustering of the CD8+ T cell subsets based on dysfunction and cytotoxicity scores. **F** Composition of the CD8+ T cell subset (excluding the two non-T cell clusters and the single-patient cluster).



**Fig. 5 | C1QC+ and SPP1+ TAMs comprise two distinct subpopulations within the myeloid cell compartment.** A UMAP overview of the myeloid compartment. DC dendritic cell, MDSC myeloid-derived suppressor cell, UMAP uniform manifold approximation and projection. B For each myeloid cluster, the two genes most differentially expressed between that cluster and the rest of the myeloid compartment. TAM, tumor-associated macrophages. C M1 (left) and M2 (right) polarization signatures in the macrophage populations. D Selected gene set enrichment

analysis results from a comparison between *SPP1* + and *C1QC* + TAMs. EMT, epithelial-mesenchymal transition. E Phagocytosis and angiogenesis scores for *SPP1* + and *C1QC* + TAMs. One sample with fewer than 40 total TAM cells was excluded. ( $n = 26$  samples; two-sided Wilcoxon rank-sum test; box plots centered around the median with hinges at 1<sup>st</sup> and 3<sup>rd</sup> quartiles and whiskers from hinge to max value or 1.5\*IQR, whichever is smallest).

collected from liver biopsies were excluded from this part of the analysis in order to minimize potential confounding effects of mixing tumors from multiple sites. We also obtained an external dataset from Cui Zhou et al.<sup>39</sup> to use as a confirmatory set to validate our findings; this dataset contains primary pancreatic tumor specimens collected before treatment ( $n = 7$ ) and after treatment ( $n = 14$ ).

As our results have shown that most samples are a mixture of basal and classical cancer cells, which could imply subtype plasticity, we investigated the effect of chemotherapy on subtype distribution. Among the top differentially expressed genes was *TFI3*, part of the classical subtype signature<sup>7,27</sup>, which was significantly higher in naive samples than treated samples (Supplementary Fig. 6A, Supplementary

Data 4). To validate the differences in RNA expression we performed IHC analyses of TFF3 in select corresponding tissue slides (Supplementary Fig. 6B, C). Since a single marker might not be sufficient to characterize samples as basal or classical, we used representative genes for both Moffitt subtypes<sup>7</sup> to create a basal and classical signature score which allowed us to place each cell on a basal-classical spectrum rather than assigning discrete phenotypes. Overall, treated samples did not show a shift towards a more basal-like expression pattern (Fig. 6A), although the treated sample P03 was distinctly more basal than any other sample (Fig. 6B). P03 was, notably, the only treated sample with poorly differentiated histology. Generally, cancer cells clustered by sample on this subtype spectrum, highlighting intertumoral heterogeneity (Fig. 6B). We saw this behavior in our confirmatory dataset (Supplementary Fig. 6D, E), which also contained one distinctly basal treated patient; when this patient was removed, the median classical score in the treated samples was actually higher than in the untreated samples. Looking beyond Moffitt signature genes, we also measured the heterogeneity of a sample's expression profile by calculating the correlations between every pair of its cancer cells. By this metric, treated and naive samples displayed no difference in intratumor heterogeneity, either per sample (Fig. 6C) or per cell (Fig. 6D).

To identify cell states, we used GSEA to check for common gene expression patterns of cancer cells in each sample. Here, cancer cells from treated samples were enriched in pathways related to angiogenesis and EMT, processes that can contribute to tumor recurrence and progression, as compared to naive samples (Fig. 6E). In the confirmatory dataset, we continued to see pathways related to angiogenesis and EMT enriched in the treated samples (Supplementary Fig. 6F). Analysis of CAF and TAM subpopulation distributions between naive and treated groups yielded no significant changes (Supplementary Fig. 6G).

### Chemotherapy reduces expression of inhibitory checkpoint molecules and ligand-receptor interactions in PDAC

We applied CellPhoneDB<sup>40</sup> to infer potential ligand-receptor interactions (LRIs) between malignant epithelial cells, TAM and CAF subpopulations, CD8+ and CD4+ T cells, MDSCs, and DCs in naïve and treated samples (Fig. 7A; more detailed results in Supplementary Data 5), and found a general decrease of cell-cell interactions in the chemotherapy-treated group in our dataset (Fig. 7B). We noted an unexpectedly high volume of putative interactions between CAFs and TAMs, which is highlighted in Supplementary Fig. 7A. Before replicating this analysis in our confirmatory dataset, we removed one sample, which had an extreme number of SPP1+ cells (Supplementary Fig. 7B) and was producing highly unusual results. The general overview of LRIs in the confirmatory dataset matched that of our original dataset, with the decrease in LRIs between naïve and treated samples being even more prevalent across compartments (Supplementary Fig. 7C, D).

As CAFs are known to recruit and polarize macrophages, we checked for recruitment-associated ligand-receptor interactions between CAFs and TAMs<sup>41</sup>. There was evidence for only two ligand-receptor pair interactions in this analysis: low levels of the CSF1-CSF1R interaction that did not change with treatment, and an interaction between CXCL12 (on iCAF<sup>s</sup> but not myCAF<sup>s</sup>) and CXCR4 (on both TAM subpopulations). CXCL12-CXCR4, the strongest potential ligand-receptor interaction pair in naïve samples, significantly weakened with treatment in both our dataset and the confirmatory dataset (Fig. 7C, Supplementary Fig. 7E–H). Notably, it appears that the change in the volume of the CXCL12-CXCR4 interaction is driven by a decrease in CXCR4 in TAMs with treatment (Supplementary Fig. 7F–H). The CXCL12-CXCR4 interaction between perivascular CAFs and TAMs has recently been linked to metastasis<sup>42</sup> and may be one way in which conventional chemotherapy affects crosstalk in the TME.

Since checkpoint inhibitory drugs have yet to show benefit in PDAC clinical trials (except in the 1% of PDAC cases with mutations in

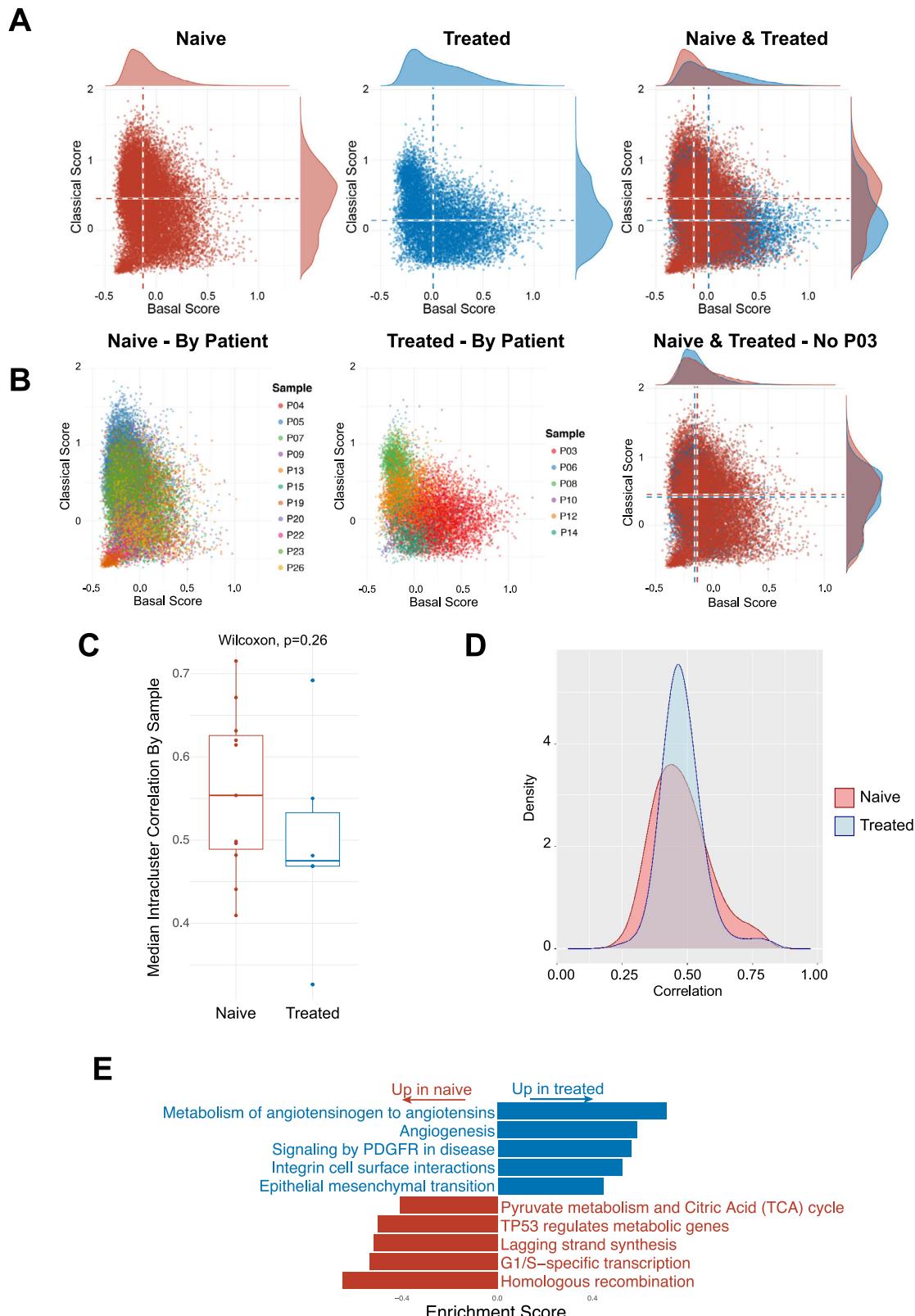
mismatch repair genes<sup>43</sup>), we further evaluated treatment-related checkpoint molecule expression in CD8+ T cells. We found TIGIT to be the highest and most broadly expressed inhibitory checkpoint molecule, while PDCD1 was expressed at low levels and in a minority of CD8+ T cells (Fig. 7D). In addition, expression of TIGIT, PDCD1, and other checkpoint molecules was lower in CD8+ T cells of chemotherapy-treated samples than naive samples (Fig. 7D). The findings in Fig. 7D are replicated very closely in the confirmatory set (Supplementary Fig. 7I). To further validate our findings, we used multiplex IHC to stain for CD8+ T cells, TIGIT, and PD-1 in corresponding tissue slides of treatment-naïve and treated and pancreatic samples (Supplementary Fig. 7J). Using representative samples, we found that TIGIT was expressed in most CD8+ T cells in naive samples and to a significantly lesser extent in CD8+ T cells in treated samples (Supplementary Figure 7K). In addition, most CD8+ T cells did not show expression of PD-1 in either naive or treated samples (Supplementary Fig. 7L). Of note, we found cells in the TME that are TIGIT positive but CD8 negative, representing expression in other immune cell types.

We also assessed whether treatment alters checkpoint molecule interactions between CD8+ T cells, TAMs, and malignant cells. We evaluated all CD8+ T cells due to the scarcity of exhausted T cells in our dataset. Moreover, CD8+ T cells can exhibit different states of exhaustion, and the larger CD8+ population is more representative of response regarding checkpoint molecule interactions with chemotherapy<sup>44</sup>. We observed more potential ligand-receptor interactions between TAMs and CD8+ T cells than between cancer cells and CD8+ T cells in both our samples (Fig. 7E; Supplementary Fig. 7M, N) and the confirmatory dataset (Supplementary Fig. 7O–P). Comparing naïve and treated samples, we noticed that all of the putative inhibitory ligand-receptor interactions were significant only in naïve samples, while immune stimulatory interactions between TAMs and CD8+ cells were significant in both treated and naïve samples (Fig. 7E). The only significant PD-1 (PDCD1) interaction occurred in naïve samples, between PD-1 on CIQC+ TAMs and PD-L2 (PDCD1LG2) on CD8+ T cells (Fig. 7E); this interaction was not present in the confirmatory dataset (Supplementary Fig. 7O). We only detected three significant interactions between CD8+ T cells and cancer cells, with TIGIT-PVR being the strongest. TIGIT-PVR was also the only interaction that was still significant in treated samples, though the effect was weaker than in naïve samples (Fig. 7E). Of those three, only PVR-TIGIT was found in the confirmatory dataset (Supplementary Fig. 7O). Of note, we did not detect any PD-1 related interactions between CD8+ T cells and cancer cells. We next performed multiplex IHC to assess coexpression of TIGIT+ on CD8+ T cells and PVR+ on CK19+ cancer cells at the protein level. We detected both cell types expressing the ligand-receptor pair in close proximity to each other, supporting the potential for ligand-receptor interactions between those cell types (Supplementary Fig. 7Q).

### Discussion

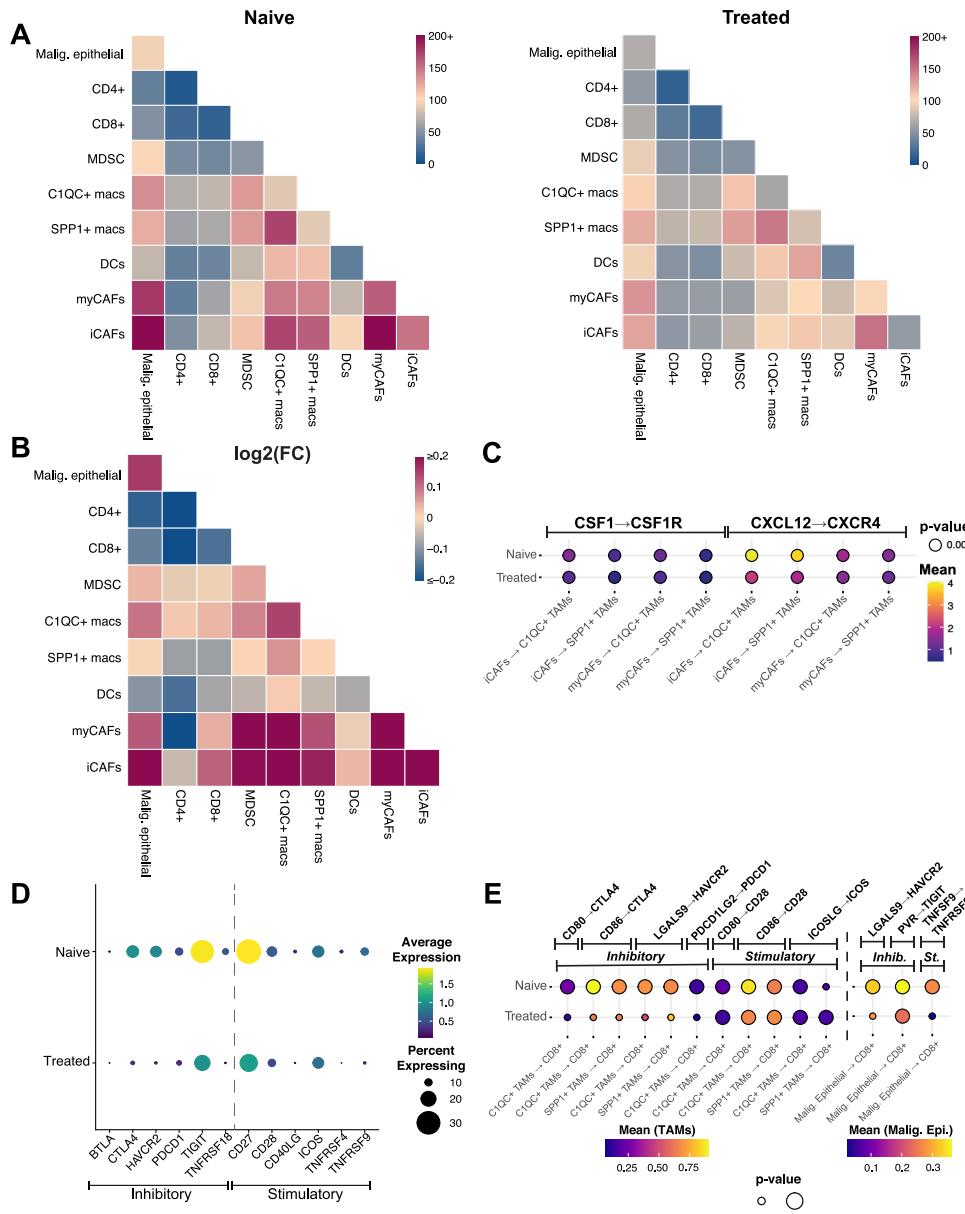
In this study we systematically characterized the composition of human PDAC at the single-cell transcriptomic level and assessed the effects of chemotherapy on the TME. Individual samples demonstrated significant heterogeneity in their Moffitt subtypes and TME composition. Analysis of the TME revealed discrete cell subpopulations with distinct features, further highlighting the complexity of the PDAC TME. We found, and confirmed in an external dataset, that chemotherapy profoundly alters the PDAC TME and might lead to further resistance to immunotherapy by reduced inhibitory checkpoint molecule expression and interactions involving CD8+ T cells.

Consistent with previous reports on subtype heterogeneity, our initial per-sample analysis of cancer cell composition revealed a heterogeneous mixture of cells with classical or basal transcriptional signatures in most samples and a variety of subclonal growth patterns<sup>22,25,45,46</sup>. A recent study demonstrating plasticity of pancreatic cancer cells and subtype switching dependent on culture conditions



**Fig. 6 | Chemotherapy treatment induces transcriptional changes in cancer cells independent of subtype.** **A** Basal and classical signature expression scatterplot of cancer cells (treated, naïve and combined). **B** Basal and classical signature expression scatterplot of cancer cells (treated and naïve by patient, combined by treatment); sample P03 has been removed from the combined plot. **C** Box plot showing the median intracluster correlation of cancer cells by sample for treated

( $n=6$ ) and naïve ( $n=11$ ) groups (two-sided Wilcoxon rank-sum test,  $p=0.26$ ; box plots centered around the median with hinges at 1<sup>st</sup> and 3<sup>rd</sup> quartiles and whiskers from hinge to max value or 1.5\*IQR, whichever is smallest). **D** Density plots of cancer cell correlation within samples for treated and untreated groups. **E** Gene set enrichment analysis of overall cancer cells, treated and untreated.



**Fig. 7 | Chemotherapy reduces expression of inhibitory checkpoint molecules and ligand-receptor interactions in PDAC. A** Heatmap displaying an overview of all inferred LRLs between displayed cell types (cancer cells, MDSCs, DCs, *C1QC* + TAMs, *SPP1* + TAMs, iCAFs, myCAFs, *CD8* + T cells, *CD4* + T cells). Left: untreated samples, right: treated samples. CAF cancer-associated fibroblast, DC dendritic cell, iCAF inflammatory CAF, LRL ligand-receptor interaction, MDSC myeloid-derived suppressor cell, myCAF myofibroblastic CAF, TAM tumor-associated macrophage. **B** Heatmap displaying the fold change between the number of inferred LRLs in naïve samples.

and treated samples for each pair of cell types. Red indicates more inferred LRLs in naïve samples and blue indicates more inferred LRLs in treated samples. **C** Dot plot of CellphoneDB output (see Methods) for recruitment-associated LRLs between iCAF/myCAF and *SPP1* + /*C1QC* + TAMs comparing untreated and treated samples. **D** Differential gene expression of checkpoint molecules in *CD8* + T cells between treated and untreated samples. **E** Dot plot of CellphoneDB output (see Methods) for checkpoint molecule LRLs between *CD8* + T cells, *SPP1* + or *C1QC* + TAMs, and cancer cells.

in vitro also identified a new intermediate subtype state<sup>25</sup>. Although we did not observe this specific intermediate state gene signature (the number of samples in each dataset may limit the generalizability of such a signature), we did discover cancer cells in numerous patient samples that expressed both basal and classical histological markers to varying degrees. Since both subtypes exhibit different biology and metabolic features<sup>47,48</sup>, using scRNA-seq to evaluate individual subtype composition prior to therapy might help understand how these elements dictate responses to different therapeutic approaches.

In recent years, investigation of CAFs has led to the identification of distinct CAF subpopulations in human and mouse PDAC<sup>19</sup>. While iCAFs and myCAF were present in our dataset, neither unsupervised clustering nor signature analysis revealed a defined apCAF population.

Our finding is also consistent with a recent publication on single nucleus sequencing in human PDAC<sup>49</sup>. Nevertheless, a selection of iCAFs and myCAF did co-express CD74 and HLA genes, consistent with another recent report<sup>22</sup>. Plasticity between CAF subtypes, as hypothesized by multiple groups, might explain the differences in CAF subpopulation findings<sup>19,29</sup>.

Myeloid cell analysis led to the identification of *SPP1* + and *C1QC* + TAMs as the major TAM subpopulations in human PDAC. In our data, M1/M2 polarization did not fully match these recently described subpopulations, suggesting that M1/M2 classification does not capture the complexity of TAM subpopulations in the human TME<sup>50–52</sup>. Further characterization showed enrichment for EMT and a high angiogenesis score in *SPP1* + TAMs, while *C1QC* + TAMs were enriched for antigen

presentation and phagocytosis. These findings are consistent with recent scRNA-seq studies in CRC<sup>38,53</sup>. A *SPP1*<sup>high</sup>/*C1QC*<sup>low</sup> TAM signature has been associated with worse overall survival in the TCGA CRC dataset, and *SPP1*+ TAMs have been linked to CRC liver metastases<sup>38,54</sup>. In our analysis of both TAM signatures in the TCGA PDAC dataset, TAM subtype combination did not correlate with worse overall survival. Since the TCGA PDAC dataset has only bulk transcriptomics and a limited variety of clinical stages, the role of *SPP1*+ and *C1QC*+ TAMs in PDAC will require further evaluation in a larger, non-bulk RNA clinical cohort.

For a more unambiguous analysis of chemotherapy effects on the TME we compared naive and treated samples from primary tumors only. Recent studies have correlated the classical subtype with better therapy response than the more aggressive basal subtype and proposed that chemotherapy might shift the Moffitt subtype from classical to basal<sup>55,56</sup>. However, when assessing for a basal and classical multi-gene signature we found heterogeneity between tumors in regard to Moffitt subtype composition but no shift toward a basal subtype in chemotherapy-treated samples, nor a significant difference in intratumor heterogeneity. Taken together, the heterogeneity within each tumor and differences in individual gene expression imply that PDAC subtype categorization should not only rely on single marker analysis. Since individual genes might not reveal biological changes in cancer cells with treatment, we also checked for gene expression programs with GSEA and noticed a Moffitt-independent shift in enrichment of angiogenesis and EMT genes in treated samples. EMT is a hallmark feature of tumor invasion, and different EMT programs have been linked to PDAC progression, dissemination, and resistance to chemotherapy<sup>57–59</sup>. While the complexity of EMT makes therapeutic approaches targeting it challenging, novel approaches would be highly impactful<sup>60</sup>. Tumor angiogenesis and microvessel density have been linked to metastasis in PDAC<sup>61–63</sup>; while anti-angiogenic drugs have thus far not yielded clinical results, the question of how chemotherapy affects the microvasculature in PDAC warrants further investigation<sup>64</sup>.

The complexity of PDAC is not only determined by the cancer cell heterogeneity but also the intricate composition of and communication between cells in the TME. In our analysis, we found a high volume of putative ligand receptor interactions between TAMs and CAFs. There is limited data on crosstalk between CAFs and TAMs in the TME, even though these interactions likely play an important role in PDAC biology and may reveal potential targets for therapy<sup>41</sup>. We found CXCL12-CXCR4, specific to iCAFs and TAMs respectively, to be the most significant potential interaction. Notably, it has been demonstrated that CXCL12 secretion by CAFs in PDAC promotes immune evasion by T cell exclusion<sup>65</sup>, and CXCL12-CXCR4 interaction between perivascular CAFs and TAMs promotes metastasis in a breast cancer model<sup>42</sup>. In our dataset, CXCL12-CXCR4 interaction between iCAFs and TAMs was significantly lower in treated samples, exemplifying how conventional chemotherapy may alter the PDAC TME in ways we could not previously observe.

Although effective in a number of other malignancies, checkpoint inhibition therapy has performed poorly in PDAC. While the underlying biology is still being uncovered, few studies have explored checkpoint expression in human PDAC samples at the single cell level. In our analysis, exhausted *CD8*+ T cells constituted only 2.8% of all *CD8*+ T cells. This corresponds with previous studies showing PDAC to be a relatively immune-cold tumor with low PD-1 expression and might be a reason why checkpoint inhibition has failed in clinical trials for PDAC<sup>4,5,66</sup>. The strongest interaction we observed between *CD8*+ T cells and cancer cells was TIGIT-PVR, which significantly decreased with treatment. Recently, Steele et al. found elevated TIGIT expression levels in tumor-infiltrating *CD8*+ T cells compared to *CD8*+ T cells from normal pancreatic tissue, but no changes in PD1 between the two groups<sup>21</sup>. Functionally, Freed-Pastor et al. used a PDAC mouse model to show that the TIGIT-PVR axis is critical for maintaining immune evasion and that TIGIT blockage elicited an immune response to PDAC<sup>67</sup>. *TIGIT* was the most

expressed inhibitory checkpoint molecule on *CD8*+ T cells and its expression, as with other checkpoint molecules, decreased with treatment. Lately, another study using single-nucleus sequencing in PDAC also found *TIGIT* overexpressed in *CD8*+ T cells in naive samples<sup>49</sup>. These findings and ours imply that TIGIT may be a stronger immunotherapy candidate than PD-1 in PDAC. The higher *TIGIT* expression in treatment-naïve samples further suggests that blocking TIGIT may be more effective given as first-line monotherapy or administered concurrently with chemotherapy, and that immune blockade as second-line treatment following chemotherapy may lessen its antitumor effects. While a recent phase III trial of tiragolumab as first line therapy in small cell lung cancer was unsuccessful<sup>68</sup>, the Morpheus-Pancreatic Cancer trial (NCT03193190) is investigating the effect of TIGIT blockage in metastatic PDAC and will hopefully provide more insight into the clinical relevance of TIGIT inhibition in PDAC. Overall, our results provide an in-depth view into human PDAC composition, its TME, and the effects of chemotherapy, and may provide insight into more effective strategies to treat this recalcitrant disease.

## Methods

### Patient cohort and data collection

A total of 27 PDAC patients were recruited for study participation at the Perlmutter Cancer Center at New York University (NYU) Langone Health between May 1, 2020, and June 30, 2021. Informed written consent for the collection of blood, tissue, and clinical information was obtained from all patients using a study protocol approved by the NYU Langone Health Institutional Review Board. Standard clinicopathological variables including sex, age, tumor site, disease stage, type of procedure, tumor stage (TNM classification), and treatment type were collected for each patient as part of a prospective PDAC database. Clinical germline and somatic tumor analyses (Tempus xT 596 gene panel, Tempus Labs, Chicago, IL; Invitae 91 gene panel, Invitae Corporation, San Francisco, CA; FoundationOne CDx 324 gene panel, Foundation Medicine, Cambridge, MA) were performed. A supplement providing individual UMAP, H&E as well as clinical and mutational information is provided as Supplementary Data 2.

### Sample acquisition and processing

Of the 27 samples, 10 were obtained following surgical resection, 7 were obtained from endoscopic ultrasound (EUS)-guided biopsies, and 10 were obtained by interventional radiology (IR)-guided biopsies from metastatic liver lesions. For EUS- and IR-guided biopsies, 2-3 cores were used for scRNA-seq analysis. The pathologic diagnosis of PDAC was confirmed in all samples by a board-certified pancreatic pathologist. Within two hours of tissue acquisition, the specimen was placed in refrigerated RPMI 1640 medium (Gibco, Cat. No. 11875101) and transported to the Center for Biospecimen Research & Development at NYU, where the sample was provided for tissue processing.

### Tissue processing and single cell isolation

Tissue samples were washed in ice-cold phosphate-buffered saline (PBS; Corning, Cat. No. 21-040-CM) and grossly assessed for adequacy (solid tissue component, free of large hemorrhagic or necrotic areas). Next, the tissue was manually minced with a scalpel blade and filtered through a 40 µm mesh filter (Falcon; Cat. No. 352340) to remove any excess debris or blood. The tissue was then transferred to a sterile gentleMACS C tube (Miltenyi Biotec, Cat. No. 130096334), resuspended in 3–5 mL of Miltenyi Tumor Dissociation Kit buffer (Cat. No. 130-095-929), and mechanically minced to submillimeter particles. The suspension was then digested utilizing the gentleMACS Octo Dissociator (Miltenyi Biotec, Cat. No. 130-095-937) for 25–30 min using the “37C\_h\_TDK\_3” program for tumor tissue. Dissociated cells were then quenched with ice-cold RPMI 1640 containing 10% fetal bovine serum (FBS; Gibco, Cat. No. 10082147) and filtered through a 40 µm mesh filter. Red blood cells were removed by treatment with 3–5 mL of ACK lysis buffer (Gibco, Cat.

No. A1049201) and a subsequent washing step in RPMI 1640 with 10% FBS. Final cell counts were established with 0.4% trypan blue solution and a Countess II Automated Cell Counter (Applied Biosystems, Cat. No. A27977). The final concentration was titrated to 300–500 viable cells/ $\mu$ l for further analysis.

### Single-cell library preparation

Single-cell suspensions were processed for 10x Genomics by the Genome Technology Center at the NYU School of Medicine per the manufacturer's guidelines. The libraries were prepared using the following reagents: Chromium Single Cell 3' Reagent Kits (v3)—the Single Cell 3' Library & Gel Bead Kit v3 (PN-1000075), the Chromium Next GEM Chip G Single Cell Kit (PN-100120), and the i7 Multiplex Kit (PN-120262) (10x Genomics)—by following the Single-Cell 3' Reagent Kits (v2) User Guide (manual part no. CG00052 Rev A) and the Single-Cell 3' Reagent Kits (v3) User Guide (manual part no. CG000183 Rev C). DNA libraries were run using the NovaSeq 6000 system (Illumina, San Diego, CA) to achieve roughly 500 million paired-end reads per sample.

### Single-cell RNA sequencing and data processing

Sequencing results were de-multiplexed and converted to FASTQ format using Illumina bcl2fastq software. The 10x Genomics Cell Ranger 5.0.1 software suite<sup>69</sup> was used to perform sample de-multiplexing, barcode processing, and single-cell 3' gene counting. The cDNA insert was aligned to the hg38/GRCh38 reference genome. Only confidently mapped, non-PCR duplicates with valid barcodes and unique molecular identifiers were used to generate the gene-barcode matrix. Further analysis including the identification of highly variable genes, dimensionality reduction, standard unsupervised clustering algorithms, and the discovery of differentially expressed genes was performed using Seurat<sup>70</sup> and scooter<sup>71</sup>.

Raw data were initially filtered to remove ambient RNA using SoupX version 1.5.2<sup>72</sup>; and cells were subsequently filtered to include only high-quality cells (as defined by >500 detectable genes, >1500 unique molecular identifiers, and <15% of transcripts coming from mitochondrial genes). Cells with >1% of transcripts representing erythroid genes (HBA1, HBA2, HBB, HBM, and ALAS2) were also excluded from the analysis. The data were normalized to the total expression, multiplied by a scaling factor of 10,000, and log-transformed. Likely doublets/multiplets were identified and removed using scDblFinder 1.6.0<sup>73</sup>.

To account for biological and technical batch differences between individual patients and scRNA-seq libraries, the Seurat anchor-based integration method for merging datasets that identify pairwise correspondences between cell pairs across datasets<sup>74</sup> was utilized to transform them into a shared space. The 2000 most variable genes based on standardized variance were selected for canonical correlation analysis as an initial dimensional reduction. The integration anchors were then identified based on the first 30 dimensions and used to generate a new dimensional reduction for further analysis.

To visualize the data, the dimensionality of the scaled integrated data matrix was further reduced to project the cells in two-dimensional space using principal component analysis followed by UMAP<sup>75</sup>. The 30 nearest neighbors were used to define the local neighborhood size with a minimum distance of 0.3. The resulting principal components were also used as a basis for partitioning the dataset into clusters using a smart local moving community detection algorithm<sup>76</sup>. A range of resolutions (0.1–10) was utilized to establish a sufficient number of clusters to separate known populations based on the expression of established markers.

### Cell identification and clustering

Single-cell clusters were identified based on common marker genes for various cell types, the most prominent of which are listed here. Full dataset (Fig. 1C): T/NK (*CD3E*), epithelial (*KRT19*), endothelial (*VWF*, *PECAMI*), myeloid (*CD68*), proliferating epithelial (*KRT19*, *MKI67*),

proliferating lymphoid (*CD3E*, *MKI67*), B/plasma (*CD79A*), proliferating myeloid (*CD68*, *MKI67*), mesenchyme (*DCN*), mast (*KIT*). Mesenchyme (Fig. 3A): myCAF (*ACTA2*, *MMP11*, *COL10AI*), iCAF (*C3*, *C7*, *CFD*, *PTGDS*), pericyte (*ACTA2*<sup>+</sup>, *RGSS*<sup>+</sup>, *CSPG4*, *NOTCH3*), proliferating (*MKI67*), peri-islet Schwann cells (*SOX10*, *S100B*, *PLP1*), epithelial-like (*KRT19*, *EPCAM*), chondrocyte-like (SPP1, IBSP, MMP13). T/NK (Fig. 4A): *CCR7*+*CD4*+T cells (*CCR7*), *IL7R*+*CD4*+T cells (*IL7R*), *FOXP3*+*CD4*+T cells (*FOXP3*), *CXCL13*+*CD4*+T cells (*CXCL13*), *GZMH*+*CD8*+T cells (*GZMH*), *GZMK*+*CD8*+T cells (*GZMK*), *ITGA1*+*CD8*+T cells (*ITGA1*), *ISG15*+T cells (*ISG15*), *GNLY*+NK cells (*GNLY*, *NCAMI*), *XCL1*+NK cells (*XCL1*, *NCAMI*), *KIT*+mast (*KIT*), *CD79*+B/Plasma (*CD79A*), *MKI67*+proliferating (*MKI67*). *CD8*+T cells (Fig. 4C): *GZMK*+*CD8*+T cells (*GZMK*), *GZMH*+*CD8*+T cell (*GZMH*), *ITGA1*+*CD8*+T cells (*ITGA1*), *SELL*+*CD8*+T cells (*SELL*), *ISG15*+*CD8*+T cells (*ISG15*), *CTLA4*+*CD8*+T cells (*CTLA4*), *HSPA6*+*CD8*+T cells (*HSPA6*), epithelial-like (*KRT19*), erythrocytes (*HBB*). Myeloid (Fig. 5): SPP1+macrophage (SPP1, MARCO), CIQC+macrophage (CIQA, CIQB, CIQC), MDSC (*S100A8*, *S100A9*, *S100A12*), monocyte (FCGR3A, CDKN1C), cDC1 (*CLEC9A*, *XRCA1*), cDC2 (*CD1C*, *FCER1A*), cDC3 (*CCL19*, *CCL22*, *CCR7*), pDC (*LILRA4*, *PLD4*), proliferating (*MKI67*), mast (*KIT*), T (*CD3E*), epithelial-like (*KRT18*, *KRT19*).

### Identification of malignant and pancreatic epithelial cells

To detect large-scale chromosomal CNVs, InferCNV version 1.8.1<sup>77</sup> was run at a sample level with the parameters recommended for 10x Genomics data. For each sample, all epithelial cells were tested for copy number aberrations. Non-epithelial (CAF, endothelial, immune) cells were used as the reference normal set. For samples with over 1,000 non-epithelial cells, they were randomly subsampled to 1000. Additionally, 10% of non-epithelial cells were used in the observation set. We used a 101-gene window in subclustering mode with HMM enabled and a six-group clustering of observations.

### Ligand-receptor interaction

CellPhoneDB v.2.1.7<sup>40</sup> was used to estimate ligand-receptor interaction. We used the statistical method with 10 iterations and no subsampling; other parameters were left as default. CPDB calculates statistical significance using permutation testing; see original paper for details.

### Gene signature scores

Gene signature scores were created using Seurat's AddModuleScore function. Lists of genes used for each score can be found in Supplementary Data 3.

### Differential gene expression

Differential gene expression was analyzed using the Wilcoxon rank-sum test with Bonferroni correction that is included in Seurat's FindMarkers function.

### Gene set enrichment analysis

Gene set enrichment analysis was performed using the GSEAPreranked function of GSEA version 4.2.0<sup>78</sup>. Genes were pre-ranked for each comparison using the output of our differential gene expression; first by taking the -log<sub>10</sub>(adjusted p-value) and then multiplying it by the sign of the fold change (for example, a gene with a fold change of -2.3 and an adjusted p-value of 1e-160 would have a rank metric of -160). GSEAPreranked parameters were set to the defaults. The gene chip annotation file used was Human\_Gene\_Symbol\_with\_Remapping\_MSigDB.v7.4.chip, and genes were collapsed and remapped to gene symbols. Gene sets used included Hallmarks (h.all.v7.5.1), KEGG (c2.cp.kegg.v7.5.1), Reactome (c2.cp.reactome.v7.5.1), Canonical Pathways (c2.cp.v7.5.1), and GO (c5.go.v7.5.1).

### Survival analysis of TCGA data

The TCGA PDAC (PAAD) clinical and gene expression data<sup>9</sup> were downloaded, using TCGABiolinks version 3.15<sup>79</sup>, to examine the

prognostic utility of specific cell subpopulation signatures. Gene expression was z-score transformed. For each signature, two distinct groups (low and high expression) were formed, corresponding to mean expression values below the 45th and above the 55th quantile values, respectively. The survival analysis was performed using the R package *survminer*<sup>80</sup>. Cox proportional hazards models from the R package *survival*<sup>81</sup> were used to adjust for tumor stage, age, and sex.

### Histology and multiplex IHC

For histopathological assessment, tumor sections were cut and slides prepared by routine processes, followed by staining with hematoxylin & eosin. For multiplex immunofluorescence and imaging, five-micron formalin fixed paraffin-embedded sections were stained with Akoya Biosciences® Opal™ multiplex automation kit reagents (Leica Cat #ARD1001EA) on a Leica BondRX® autostainer, according to the manufacturers' instructions. In brief, slides underwent sequential epitope retrieval with Leica Biosystems Epitope Retrieval 2 solution (ER2, EDTA based, pH 9, Cat. AR9640), primary and secondary antibody incubation, and tyramide signal amplification with Opal® fluorophores as shown in Supplementary Table 3. Primary and secondary antibodies/polymers were removed during sequential epitope retrieval steps while the relevant fluorophores remained covalently attached to the antigen. Slides were counterstained with spectral DAPI (Akoya Biosciences, FP1490) and mounted with ProLong Gold Antifade (ThermoFisher Scientific, P36935). Semi-automated image acquisition was performed on a Vectra® Polaris multispectral imaging system. After whole slide scanning at 20x in the motif mode, regions of interest were selected for spectral unmixing and image processing using InForm® version 2.4.10 software from Akoya Biosciences.

### IHC Quantification

Semi-automated image acquisition was performed on a Vectra® Polaris multispectral imaging system. Eleven cases were used for this quantification. After whole slide scanning at 20X the tissue was manually outlined to select fields for spectral unmixing and image analysis using InForm® version 2.6 software from Akoya Biosciences. Selection of areas was performed by a trained pathologist on H&E-stained sequential sections and copied onto the fluorescent image on Akoya Biosciences Phenochart software. A training algorithm was first used to segment images into tissue categories: "CK19 +" or "CK19 -" (to exclude areas of the image with no tissue). Tissue segmentations were reviewed for different samples during training iterations. Following tissue classification cells were segmented based on nuclear signal (DAPI). Cells were phenotyped after segmentation using inForm's trainable algorithm based on the R *glmnet* package<sup>82</sup>.

Four algorithms were created to classify cells as CD8 + or 'other', CK19 + or 'other', TIGIT + or 'other', and PDI + or 'other'. Phenotypes were reviewed for different samples during training iterations. Data was exported as text containing sample names, field of acquisition coordinates, individual cell information including coordinates and identified phenotype. Each image was analyzed with all four algorithms so that every cell was classified four times. Concatenation and aggregation of all phenotyping information was performed in R using the *phenoptrReports* package<sup>83</sup>.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The raw scRNA-seq data generated for this project is available on GEO with accession number [GSE205013](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE205013). TCGA data<sup>9</sup> is publicly available on the GDC Data Portal [<https://portal.gdc.cancer.gov/>]. The confirmatory dataset from Cui Zhou et al.<sup>39</sup> is publicly available at the Human Tumor Atlas Network (HTAN) under the HTAN WUSTL Atlas [<https://humantumoratlas.org/explore?selectedFilters=%5B%7B%22group%22%3A%22AtlasName%22%2C%22value%22%3A%22HTAN+WUSTL%22%7D%20%26%20%22file%22%3A%22%22>]

and on dbGaP with accession number [phs002371.v2.p1](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC802371/). The remaining data are available within the article and its supplementary files.

### References

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **72**, 7–33 (2022).
2. Rahib, L., Wehner, M. R., Matisian, L. M. & Nead, K. T. Estimated projection of US cancer incidence and death to 2040. *JAMA Netw. Open* **4**, e214708 (2021).
3. Vincent, A., Herman, J., Schulick, R., Hruban, R. H. & Goggins, M. Pancreatic cancer. *Lancet* **378**, 607–620 (2011).
4. O'Reilly, E. M. et al. Durvalumab with or without tremelimumab for patients with metastatic pancreatic ductal adenocarcinoma: a phase 2 randomized clinical trial. *JAMA Oncol.* **5**, 1431–1438 (2019).
5. Bear, A. S., Vonderheide, R. H. & O'Hara, M. H. Challenges and opportunities for pancreatic cancer immunotherapy. *Cancer Cell* **38**, 788–802 (2020).
6. Collisson, E. A., Bailey, P., Chang, D. K. & Biankin, A. V. Molecular subtypes of pancreatic cancer. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 207–220 (2019).
7. Moffitt, R. A. et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**, 1168–1178 (2015).
8. Wood, L. D., Canto, M. I., Jaffee, E. M. & Simeone, D. M. Pancreatic cancer: pathogenesis, screening, diagnosis and treatment. *Gastroenterology* (2022) <https://doi.org/10.1053/j.gastro.2022.03.056>
9. Cancer Genome Atlas Research Network. Integrated genomic characterization of pancreatic ductal adenocarcinoma. *Cancer Cell* **32**, 185–203.e13 (2017).
10. Ho, W. J., Jaffee, E. M. & Zheng, L. The tumour microenvironment in pancreatic cancer - clinical challenges and opportunities. *Nat. Rev. Clin. Oncol.* **17**, 527–540 (2020).
11. Beatty, G. L., Werba, G., Lyssiotis, C. A. & Simeone, D. M. The biological underpinnings of therapeutic resistance in pancreatic cancer. *Genes Dev.* **35**, 940–962, <https://doi.org/10.1101/gad.348523.121> (2021).
12. Moncada, R. et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* **38**, 333–342 (2020).
13. Azizi, E. et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308.e36 (2018).
14. Zhang, Y. et al. Single-cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer. *Cancer Cell* **39**, 1578–1593.e8 (2021).
15. Chan, J. M. et al. Signatures of plasticity, metastasis, and immunosuppression in an atlas of human small cell lung cancer. *Cancer Cell* **39**, 1479–1496.e18, <https://doi.org/10.1016/j.ccr.2021.09.008> (2021).
16. Wu, H. et al. Single-cell transcriptomics of a human kidney allograft biopsy specimen defines a diverse inflammatory response. *J. Am. Soc. Nephrol.* **29**, 2069–2080 (2018).
17. Kim, N. et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* **11**, 2285 (2020).
18. Peng, J. et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* **29**, 725–738, <https://doi.org/10.1038/s41422-019-0195-y> (2019).
19. Elyada, E. et al. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discov.* **9**, 1102–1123 (2019).

20. Wang, Y. et al. Single-cell analysis of pancreatic ductal adenocarcinoma identifies a novel fibroblast subtype associated with poor prognosis but better immunotherapy response. *Cell Disco.* **7**, 36 (2021).
21. Steele, N. G. et al. Multimodal mapping of the tumor and peripheral blood immune landscape in human pancreatic cancer. *Nat. Cancer* **1**, 1097–1112 (2020).
22. Lee, J. J. et al. Elucidation of tumor-stromal heterogeneity and the ligand-receptor interactome by single-cell transcriptomics in real-world pancreatic cancer biopsies. *Clin. Cancer Res* **27**, 5912–5921 (2021).
23. Abou Khuzam, R. et al. An eight-gene hypoxia signature predicts survival in pancreatic cancer and is associated with an immuno-suppressed tumor microenvironment. *Front. Immunol.* **12**, 680435 (2021).
24. O'Flanagan, C. H. et al. Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. *Genome Biol.* **20**, 1–13 (2019).
25. Raghavan, S. et al. Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. *Cell* **184**, 6119–6137.e26 (2021).
26. Puleo, F. et al. Stratification of pancreatic ductal adenocarcinomas based on tumor and microenvironment features. *Gastroenterology* **155**, 1999–2013.e3 (2018).
27. Collisson, E. A. et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat. Med.* **17**, 500–503 (2011).
28. Bailey, P. et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52 (2016).
29. Öhlund, D. et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J. Exp. Med.* **214**, 579–596 (2017).
30. Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624.e24 (2017).
31. Yang, N., Mosher, R., Seo, S., Beebe, D. & Friedl, A. Syndecan-1 in breast cancer stroma fibroblasts regulates extracellular matrix fiber organization and carcinoma cell motility. *Am. J. Pathol.* **178**, 325–335 (2011).
32. Cheng, H.-W. et al. CCL19-producing fibroblastic stromal cells restrain lung carcinoma growth by promoting local antitumor T-cell responses. *J. Allergy Clin. Immunol.* **142**, 1257–1271.e4 (2018).
33. Crinier, A. et al. High-dimensional single-cell analysis identifies organ-specific signatures and conserved nk cell subsets in humans and mice. *Immunity* **49**, 971–986.e5 (2018).
34. de Andrade, L. F. et al. Discovery of specialized NK cell populations infiltrating human melanoma metastases. *JCI Insight* **4**, e133103 (2019).
35. Li, H. et al. Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma. *Cell* **176**, 775–789.e18 (2019).
36. Kwak, T. et al. Distinct populations of immune-suppressive macrophages differentiate from monocytic myeloid-derived suppressor cells in cancer. *Cell Rep.* **33**, 108571 (2020).
37. House, I. G. et al. Macrophage-derived CXCL9 and CXCL10 are required for antitumor immune responses following immune checkpoint blockade. *Clin. Cancer Res.* **26**, 487–504, <https://doi.org/10.1158/1078-0432.ccr-19-1868> (2019).
38. Zhang, L. et al. Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. *Cell* **181**, 442–459.e29 (2020).
39. Cui Zhou, D. et al. Spatially restricted drivers and transitional cell populations cooperate with the microenvironment in untreated and chemo-resistant pancreatic cancer. *Nat. Genet.* **54**, 1390–1405 (2022).
40. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
41. Gunaydin, G. CAFs interacting with TAMs in tumor microenvironment to enhance tumorigenesis and immune evasion. *Front. Oncol.* **11**, 668349 (2021).
42. Arwert, E. N. et al. A unidirectional transition from migratory to perivascular macrophage is required for tumor cell intravasation. *Cell Rep.* **23**, 1239–1248 (2018).
43. Hu, Z. I. et al. Evaluating mismatch repair deficiency in pancreatic adenocarcinoma: challenges and recommendations. *Clin. Cancer Res* **24**, 1326–1336 (2018).
44. Beltra, J.-C. et al. Developmental relationships of four exhausted CD8 T cell subsets reveals underlying transcriptional and epigenetic landscape control mechanisms. *Immunity* **52**, 825–841.e8 (2020).
45. Chan-Seng-Yue, M. et al. Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. *Nat. Genet.* **52**, 231–240 (2020).
46. Grünwald, B. T. et al. Spatially confined sub-tumor microenvironments in pancreatic cancer. *Cell* **184**, 5577–5592.e18 (2021).
47. Espiau, R. P., Courtois, S., Parejo, A. B. & Sancho, P. Molecular and metabolic subtypes correspondence for pancreatic ductal adenocarcinoma classification. *J. Clin. Med. Res.* **9**, 4128 (2020).
48. O'Kane, G. M. et al. GATA6 expression distinguishes classical and basal-like subtypes in advanced pancreatic cancer. *Clin. Cancer Res* **26**, 4901–4910 (2020).
49. Hwang, W. L. et al. Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment. *Nat. Genet.* **54**, 1178–1191 (2022).
50. Mills, C. D., Kincaid, K., Alt, J. M., Heilman, M. J. & Hill, A. M. M1/M2 macrophages and the Th1/Th2 paradigm. *J. Immunol.* **164**, 6166–6173 (2000).
51. Anderson, C. F. & Mosser, D. M. A novel phenotype for an activated macrophage: the type 2 activated macrophage. *J. Leukoc. Biol.* **72**, 101–106 (2002).
52. Italiani, P. & Boraschi, D. From monocytes to M1/M2 macrophages: phenotypical vs. functional differentiation. *Front. Immunol.* **5**, 514 (2014).
53. Cheng, S. et al. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792–809.e23 (2021).
54. Liu, Y. et al. Immune phenotypic linkage between colorectal cancer and liver metastasis. *Cancer Cell* **40**, 424–437.e5 (2022).
55. Aung, K. L. et al. Genomics-driven precision medicine for advanced pancreatic cancer: early results from the COMPASS trial. *Clin. Cancer Res* **24**, 1344–1354 (2018).
56. Porter, R. L. et al. Epithelial to mesenchymal plasticity and differential response to therapies in pancreatic ductal adenocarcinoma. *Proc. Natl. Acad. Sci. USA* (2019) <https://doi.org/10.1073/pnas.1914915116>
57. Aiello, N. M. et al. EMT subtype influences epithelial plasticity and mode of cell migration. *Dev. Cell* **45**, 681–695.e4 (2018).
58. Arumugam, T. et al. Epithelial to mesenchymal transition contributes to drug resistance in pancreatic cancer. *Cancer Res* **69**, 5820–5828 (2009).
59. Rhim, A. D. et al. EMT and dissemination precede pancreatic tumor formation. *Cell* **148**, 349–361 (2012).
60. Bulle, A. & Lim, K.-H. Beyond just a tight fortress: contribution of stroma to epithelial-mesenchymal transition in pancreatic cancer. *Signal Transduct. Target Ther.* **5**, 249 (2020).
61. Bielenberg, D. R. & Zetter, B. R. The contribution of angiogenesis to the process of metastasis. *Cancer J.* **21**, 267–273 (2015).
62. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).

63. Takagi, K., Takada, T. & Amano, H. A high peripheral microvessel density count correlates with a poor prognosis in pancreatic cancer. *J. Gastroenterol.* **40**, 402–408 (2005).
64. Annese, T., Tamma, R., Ruggieri, S. & Ribatti, D. Angiogenesis in pancreatic cancer: pre-clinical and clinical studies. *Cancers* **11**, 381 (2019).
65. Feig, C. et al. Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti-PD-L1 immunotherapy in pancreatic cancer. *Proc. Natl Acad. Sci. USA* **110**, 20212–20217 (2013).
66. Vonderheide, R. H. The immune revolution: a case for priming, not checkpoint. *Cancer Cell* **33**, 563–569 (2018).
67. Freed-Pastor, W. A. et al. The CD155/TIGIT axis promotes and maintains immune evasion in neoantigen-expressing pancreatic cancer. *Cancer Cell* **39**, 1342–1360.e14 (2021).
68. Mullard, A. Roche's anti-TIGIT drug suffers a phase III cancer setback. *Nat. Rev. Drug Discov.* **21**, 327 (2022).
69. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
70. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
71. Dolgalev, I. & Yeaton, A. Scooter: streamlined scRNA-seq analysis pipeline. (2021).
72. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, 151 (2020).
73. Germain, P.-L., Lun, A., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Research* **10**, 979, <https://doi.org/10.12688/f1000research.73600.1> (2021).
74. Stuart, T. et al. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21, <https://doi.org/10.1016/j.cell.2019.05.031> (2019).
75. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* (2018) <https://doi.org/10.1038/nbt.4314>.
76. Waltman, L. & van Eck, N. J. A smart local moving algorithm for large-scale modularity-based community detection. *Eur. Phys. J. B* **86**, 471 (2013).
77. Tickle, T., Tirosh, I., Georgescu, C., Brown, M. & Haas, B. *inferCNV* of the Trinity CTAT Project. (Bioconductor, 2019). <https://doi.org/10.18129/B9.BIOC.INFERCNV>
78. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
79. Colaprico, A. et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* vol. **44**, e71–e71, <https://doi.org/10.1093/nar/gkv1507> (2016).
80. Kassambara, A., Kosinski, M. & Biecek, P. *survminer*. (2021).
81. Therneau, T. M. *A Package for Survival Analysis in R*. (2022). <https://CRAN.R-project.org/package=survival>
82. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, 1–22 (2010).
83. Johnson, K. S. *PhenoptrReports: Create reports using Phenoptics data*. (2022). <https://akoyabio.github.io/phenoptrReports/>

## Acknowledgements

We thank Valeria Mezzano Robinson, Shanmugapriya Selvaraj and Branka Brukner Dabovic from the Experimental Pathology Core at NYU Langone Health for support with multiplex IHC staining and imaging. We thank Peter Meyn and Yutong Zhang from the Genome Technology

Center Core at NYU Langone Health. We thank the Center for Biospecimen Research and Development at NYU Langone Health for support with tissue acquisition. These shared resources are partially supported by the Cancer Center Support Grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center. We thank the Ding Lab at WUSTL for providing us with their data. This project was supported by NIH R01CA245005 (DMS) and a research scholar grant from the German Research Foundation (WE7047/1-1; DW). This investigation was also supported by the National Institutes of Health under Ruth L. Kirschstein National Research Service Award T32CA193111. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

## Author contributions

G.W., D.W., E.A.K., I.D., A.T. and D.M.S. developed the study concept and were responsible for the study design. D.W., E.A.K. and G.W. wrote the manuscript, which was then edited by all co-authors. D.M.S., T.G., X.J. and C.H. helped acquire human biospecimens. G.W., D.W., E.Z., L.W., G.O. and A.H. processed samples for scSeq. D.W., E.A.K., G.W., I.D. and D.K. analyzed the scRNA-Seq data. E.A.K., I.D. and D.K. were responsible for the bioinformatics analysis pipelines. X.J., G.W., E.Z., H.B.L., and D.W. collected patient information and maintained the databases. C.H., C.L., D.W., L.W., G.W. and D.M.S. analyzed the H&E and multiplex IHC. S.D., T.G., N.B., L.K. and P.O. provided study guidance and feedback on the manuscript. M.H.S., A.W.L. and T.H.W. assisted with study design and analysis. D.M.S. and A.T. directed the study. All co-authors approved the final version of the manuscript before submission.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-36296-4>.

**Correspondence** and requests for materials should be addressed to Aristotelis Tsirigos or Diane M. Simeone.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023, corrected publication 2023