nature biotechnology

ARTICLES

https://doi.org/10.1038/s41587-019-0392-8

Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas

Reuben Moncada[1], Dalia Barkley[1], Florian Wagner[1], Marta Chiodin[1], Joseph C. Devlin[1], Maayan Baron[1], Cristina H. Hajdu[2], Diane M. Simeone[2,3,4] and Itai Yanai[1,5*]

Single-cell RNA sequencing (scRNA-seq) enables the systematic identification of cell populations in a tissue, but characterizing their spatial organization remains challenging. We combine a microarray-based spatial transcriptomics method that reveals spatial patterns of gene expression using an array of spots, each capturing the transcriptomes of multiple adjacent cells, with scRNA-Seq generated from the same sample. To annotate the precise cellular composition of distinct tissue regions, we introduce a method for multimodal intersection analysis. Applying multimodal intersection analysis to primary pancreatic tumors, we find that subpopulations of ductal cells, macrophages, dendritic cells and cancer cells have spatially restricted enrichments, as well as distinct coenrichments with other cell types. Furthermore, we identify colocalization of inflammatory fibroblasts and cancer cells expressing a stress-response gene module. Our approach for mapping the architecture of scRNA-seq-defined subpopulations can be applied to reveal the interactions inherent to complex tissues.

Recent technological advances have enabled a view into cancer at unprecedented molecular resolution[1]. scRNA-seq has emerged as a powerful tool for the unbiased and systematic characterization of the cells present in a given tissue[2,3]. Indeed, the application of scRNA-seq to patient tumors has uncovered multiple cellular subpopulations and has highlighted intercellular cross-talk within the tumor microenvironment[4–10]. In particular, over the past 5 yr many independent reports have found evidence for intratumoral transcriptional heterogeneity among cancer cells[4–19]. While these results clearly highlight the power of scRNA-seq, tissue dissociation before sequencing leads to the loss of spatial information, thus limiting our understanding of cellular interactions and organization in the tumor microenvironment.

Methods providing spatially resolved transcriptomic profiling[20–24] have also been introduced and are complementary to scRNA-seq. To resolve the cell type spatial composition of tissues, the integration of in situ hybridization (ISH) gene expression atlases with scRNA-seq data has been useful to map rare subpopulations and cell types using a small subset of genes[25,26]. Such ISH atlases used for tissue with defined morphology do not exist for solid tumors, however, which have variable tissue architecture and gene expression patterns. Moreover, the throughput of ISH methods remains limited to a subset of the transcriptome, thus preventing comprehensive expression analysis in a single experiment.

The recently developed spatial transcriptomics (ST) method[27] overcomes the throughput limitation of ISH methods, allowing for unbiased mapping of transcripts over entire tissue sections using spatially barcoded oligo-deoxythymidine microarrays (Fig. 1a). ST has already been used to study the mouse olfactory bulb[27], breast cancer[27], melanoma[28], prostate cancer[29], gingival tissue[30], adult human heart tissue[31], mouse and human spinal cord tissue[32], as well as model plant species[33]. Similar to other previously reported spatially resolved transcriptomic tools[20,34,35], however, a main limitation of ST is its lack of cellular resolution; each spot captures the transcriptomes of 10–200 cells depending on the tissue context[36].

While scRNA-seq and ST data each have limitations, we reasoned that these could be overcome by an integration of the two data modalities to enable comprehensive and unbiased tissue analysis. Here, we present an integration of scRNA-seq with the ST method to study pancreatic ductal adenocarcinoma (PDAC). In this approach, a tumor is first divided and a single-cell suspension is generated from one portion and processed for scRNA-seq to identify the cell populations present in the tissue. From the remaining tissue, cryosections are processed using the ST method to provide an unbiased map of expressed transcripts across the tissue. We then integrate these two datasets by introducing multimodal intersection analysis (MIA). Our approach infers the enrichment of specific cell types in a given tissue region by computing the degree of overlap between genes specifically mapped to that region and the cell type-specific genes identified by the scRNA-seq data. Studying primary PDAC tumors from different patients, we identified enrichments of specific cell types and subpopulations across spatially restricted regions of the tissue. Our approach for combining these two complementary and powerful technologies is easily scalable to any architecturally complex tissue and has the potential to provide meaningful biological insight across a range of fields.

Results

Identifying cell populations in PDAC with scRNA-seq. We first processed fresh primary PDAC tumors from two untreated

[1]Institute for Computational Medicine, NYU Langone Health, New York, NY, USA. [2]Department of Pathology, NYU Langone Health, New York, NY, USA. [3]Department of Surgery, NYU Langone Health, New York, NY, USA. [4]Perlmutter Cancer Center, NYU Langone Health, New York, NY, USA. [5]Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY, USA. *e-mail: itai.yanai@nyulangone.org

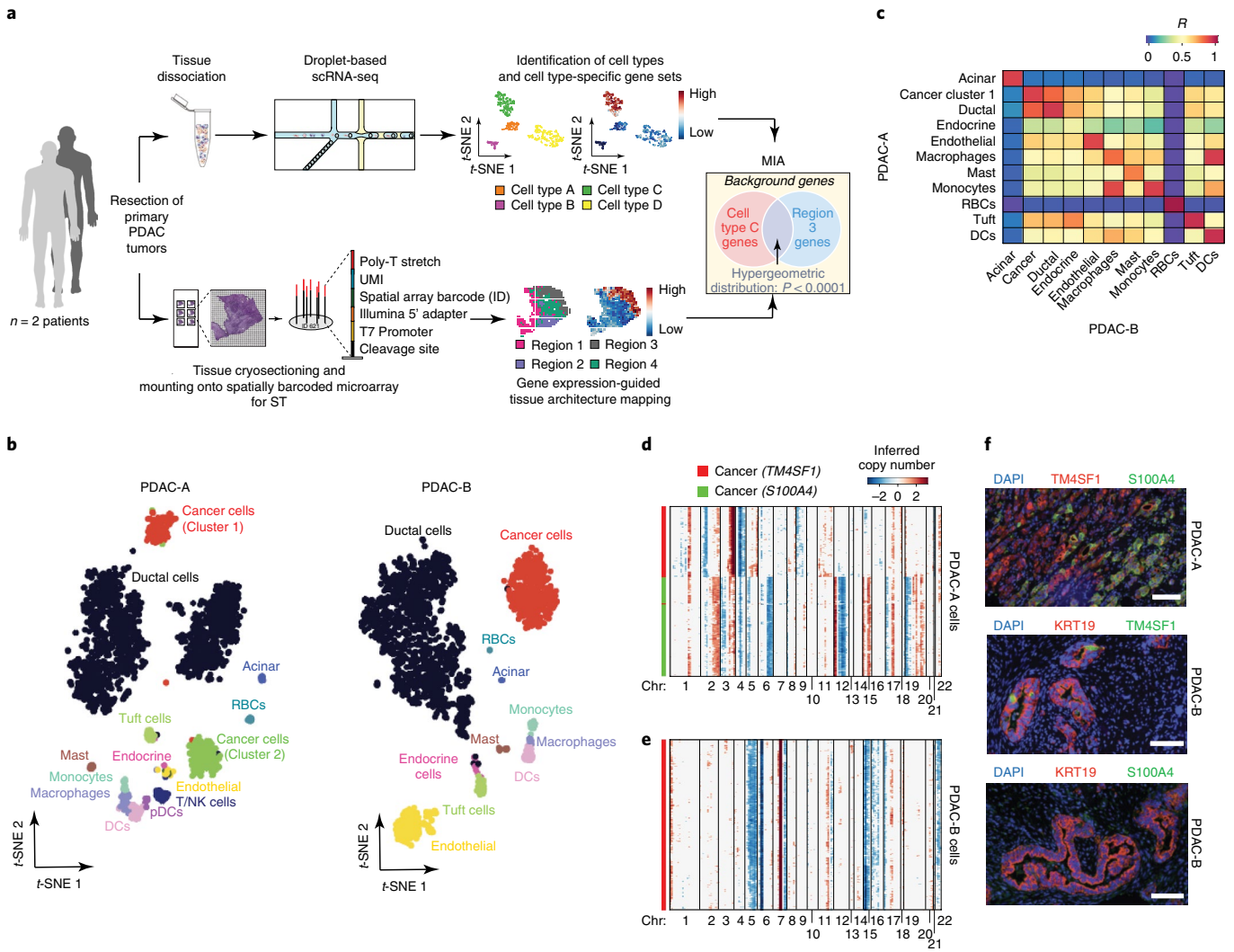

Fig. 1 | scRNA-seq analysis of two tumors from patients with PDAC. a, Schematic of the experimental design and analysis. Surgically resected PDAC tumors are split and processed in parallel for scRNA-seq and ST. After clustering, the cell type of each cluster is inferred according to specifically expressed genes. Cryosections of the remaining tissue were used for ST analysis in which each spot captures the transcriptomes of the cells at a specific location in the tissue. Applying our multimodal intersection analysis (MIA) across the two datasets reveals the spatial distribution of the cell populations and subpopulations. **b**, *t*-SNE projection of 1,926 cells from the PDAC-A tumor (left panel) and 1,733 cells from PDAC-B (right panel). Clusters are colored and labeled according to their inferred cell type identities. **c**, Correspondence between PDAC-A and PDAC-B cell types computed by Pearson's correlation on the average cell type transcriptomes of shared cell types (see Methods). **d**,**e**, CNV profiles inferred from scRNA-seq on PDAC-A (**d**) and PDAC-B (**e**). Red and blue indicate chromosomal amplifications and deletions, respectively. **f**, Double immunofluorescence staining of markers for cancer cell subpopulations ($n = 2$). Top panel, double staining of TM4SF1 (cancer cluster 1) and S100A4 (cancer cluster 2) in PDAC-A FFPE tissue. Note the mutual exclusion of TM4SF1 and S100A4 signals. Middle, KRT19 and TM4SF1 staining in PDAC-B tissue. Bottom, KRT19 and S100A4 staining in PDAC-B tissue. Note colocalization of KRT19 and TM4SF1 signals, but a lack of S100A4 signal in PDAC-B as expected. Scale bars, 100 μm. Chr, chromosome; DC, dendritic cell; ID, spatial barcode identification; NK, natural killer; pDC, plasmacytoid dendritic cell; RBC, red blood cell.

patients—PDAC-A and PDAC-B—for parallel scRNA-seq and ST analysis (Fig. 1a and see Methods). The scRNA-seq data consisted of cells with approximately 2,500–3,300 unique molecular identifiers (UMIs) and approximately 1,400–1,700 uniquely expressed genes per cell (Supplementary Fig. 1). To infer cell type identities, we used a recursive hierarchical clustering scheme (see Methods) that applies our recently developed *k*-nearest neighbors (KNN) smoothing algorithm[37] to reduce the noise inherent to scRNA-seq data[38]. We identified 15 and 11 distinct populations in the PDAC-A and PDAC-B tumors, respectively, providing an in-depth perspective of the PDAC tumor microenvironment (Fig. 1b). The gene expression profiles of shared cell types showed strong correlation between the patient samples, validating our annotation of clusters across samples (Fig. 1c).

To distinguish the PDAC cancer cells from the nonmalignant ductal cells, we used scRNA-seq-based copy number variation (CNV) analysis (as performed previously[4]; see Methods). For each identified cell type, we inferred chromosomal amplifications and deletions (with the cells of other clusters as the background) and detected two clusters in PDAC-A and one cluster in PDAC-B that displayed aberrant CNV profiles (Fig. 1d,e). Notably, the deletion along chromosome 6 in both samples and the amplification on chromosome 7 in the PDAC-B profile are consistent with the most common chromosomal abnormalities seen in PDAC from cytogenetic

data[39]. As a negative control, we removed a random subset of ductal cells from the background and inferred the relative CNV profiles of these cells. We found that while these profiles are noisy, they do not provide evidence for aberrant CNVs (Supplementary Fig. 2a). The CNV profiles of the PDAC-B cancer cells do show a degree of variation, perhaps reflecting minor heterogeneities within a single clone since transcriptionally they form a single cluster of malignant cells (Fig. 1b).

The two PDAC-A cancer clusters—cancer clusters 1 and 2—exhibited high expression of *TM4SF1* (cluster 1) or *S100A4* (cluster 2) while the PDAC-B cancer cluster was high in *TM4SF1* expression (Supplementary Fig. 2b), suggesting similarity between PDAC-A cancer cluster 1 and the PDAC-B cancer cluster. To further validate whether the *TM4SF1*- and *S100A4*-expressing populations identified in the scRNA-seq data represent cancer cell populations, we performed double immunofluorescence staining of TM4SF1 and S100A4 on formalin-fixed paraffin embedded (FFPE) tissue originating from the same patients (Fig. 1f and Supplementary Fig. 2c,d). We found TM4SF1 and S100A4 signals in malignant PDAC-A ducts as identified by morphology (Supplementary Fig. 2c,d, top panels), but not in nonmalignant ducts (bottom panels). Consistent with the transcriptomic data, the PDAC-A cancer cells displayed mutually exclusive staining for TM4SF1 and S100A4 (Fig. 1f, top). When we double-stained PDAC-B tissue for KRT19—a marker for both malignant and nonmalignant pancreatic ductal cells—and TM4SF1, or KRT19 and S100A4, we found colocalization of the KRT19 and TM4SF1 signals (Fig. 1f, middle) but no colocalization of the KRT19 and S100A4 signals (Fig. 1f, bottom). This validates expression of TM4SF1 and not S100A4 in the malignant cells of PDAC-B. Taken together with the distinct CNV profiles for the cancer clusters, we confirm the presence of genetically and transcriptionally distinct cancer cell populations in the PDAC-A sample.

ST of PDAC tissue. To generate unbiased transcriptomic maps of the tissue sections, we mounted cryosections of the two unfixed PDAC tissues originating from the same tumor sample onto the spatially barcoded ST microarray slides (see Methods). After hematoxylin and eosin (H&E) staining and brightfield imaging, we annotated the slide for distinct histological features (Fig. 2a,b). In the PDAC-A tumor section, we defined four main regions: cancer cells and desmoplasia, nonmalignant duct epithelium, stroma and normal acini-rich pancreatic tissue (Fig. 2a and Supplementary Fig. 3a–d). The PDAC-B tissue section (Fig. 2b) did not appear to contain normal pancreatic tissue; however, we noted the presence of interstitial space adjacent to the cancer (Supplementary Fig. 3e,f).

The samples were then processed for ST analysis, including complementary DNA synthesis, amplification by in vitro transcription, library construction and sequencing[19]. We demultiplexed the sequenced reads and identified their spatial location within the tissue using the ST location-specific barcodes of the array. We used the H&E images to estimate that each ST spot captured approximately 20–70 cells (Supplementary Fig. 4). We expect these numbers to be highly variable depending on the tissue histology[36]. For example, highly fibrotic regions with dense connective tissue (and therefore lower cellular content) may have low cell number per spot compared with regions with high cellular density such as the acinus of the pancreas (Supplementary Fig. 4). We detected approximately 2,400 UMIs and approximately 1,000 unique genes per spot for both ST datasets (Supplementary Fig. 5a–f). Other published works using the ST method (particularly on human tissue) report similar statistics for the average number of UMIs per spot (Supplementary Table 1). In both sample datasets, we found that the spatial expression of many variably expressed genes (see Methods) matched the annotated histological regions (Fig. 2c,d).

We classified the ST data into regions by first performing principal component analysis (PCA) on the most variably expressed genes across all ST spots (Supplementary Fig. 5g,h and see Methods). After clustering the spots of each ST array based on the principal component scores (see Methods), we found that the resulting clusters were consistent with the independent histological annotations (Fig. 2e,f and Supplementary Fig. 5g,h), supporting the ability to identify distinct spatial regions within a section based on ST gene expression alone.

Multimodal intersection analysis (MIA). To integrate the scRNA-seq and ST datasets, we developed MIA. This analysis proceeds by first delineating sets of cell type-specific and tissue region-specific genes and then determining whether their overlap is higher (enrichment) or lower (depletion) than expected by chance. In the scRNA-seq data, we defined the gene sets by identifying for each cell type those genes whose expression is statistically higher in the cells annotated to that cell type in comparison with expression in the remaining cells ($P < 10^{-5}$, two-tailed Student's *t*-test; see Methods). For the ST data, we then identified genes with significantly higher expression in each spatial region relative to the others ($P < 0.01$, two-tailed Student's *t*-test).

With the gene sets extracted across the scRNA-seq and ST modalities, MIA next computes the overlap between each pair of cell type-specific and region-specific gene sets and performs a hypergeometric test to assess significant enrichment or depletion. As an example, we found that the fibroblast-specific genes in PDAC-A overlapped significantly with the set of genes specific to the cancer region of the ST data (Fig. 2g; $P < 10^{-10}$). Extending this analysis to all pairs of cell types and tumor regions produces an 'MIA map'. Thus, while the scRNA-seq data led to identification of fibroblasts, MIA further revealed that these were enriched in the cancer region, as opposed to the stromal region, suggesting that the fibroblasts captured by scRNA-seq represent activated fibroblasts[40] (Fig. 2h). As expected, we found that the duct epithelium region was mainly enriched with ductal cells, while the pancreatic tissue was enriched with acinar cells and endocrine cells ($P < 10^{-10}$, for all specified enrichments). Similarly, a MIA map was generated for the PDAC-B sample (Fig. 2i). Testing a range of thresholds for marker gene selection for both modalities, we found that the MIA approach is robust at identifying enrichments and depletions of cell types across spatial regions (Supplementary Fig. 5i,j). To assess the robustness to the number of genes required for MIA maps, we found that when the number of detected genes in the cancer region was downsampled below 100 genes, an enrichment with fibroblast-specific genes dropped below significance ($P > 0.05$; Supplementary Fig. 5k). These results support a broad utility of MIA maps to provide spatial and functional annotations for the scRNA-seq-defined cell populations.

Identification and mapping of cell type subpopulations across tissue regions. One of the most useful aspects of scRNA-seq is its ability to reveal distinct subpopulations within cell types. We thus sought to characterize intra-population heterogeneity by identifying subpopulations within cell types and then applying MIA to query for spatially restricted mapping across the tissue. From both tumor samples, a majority of the scRNA-seq cells consisted of *KRT19*-expressing ductal cells (Fig. 1b and Supplementary Fig. 2b), one of the two main constituent cell types of the pancreatic exocrine system[41]. We identified a total of four ductal subpopulations: a ductal population expressing *APOL1* and hypoxia-response-related genes including *ERO1A* (ref. [42]) and *CA9* (ref. [43]); a terminal ductal population expressing *TFF1*, *TFF2* and *TFF3* (ref. [44]); a centroacinar ductal population expressing *CRISP3* and *CFTR* (ref. [44]); and antigen-presenting ductal cells expressing major histocompatibility complex (MHC) class II genes *CD74*, *HLA-DPA1*, *HLA-DQA2*, *HLA-DRA*, *HLA-DRB1* and *HLA-DRB5* and complement pathway components *C1S*, *C4A*, *C4B*, *CFB* and *CFH* (Fig. 3a–d). Although MHC class

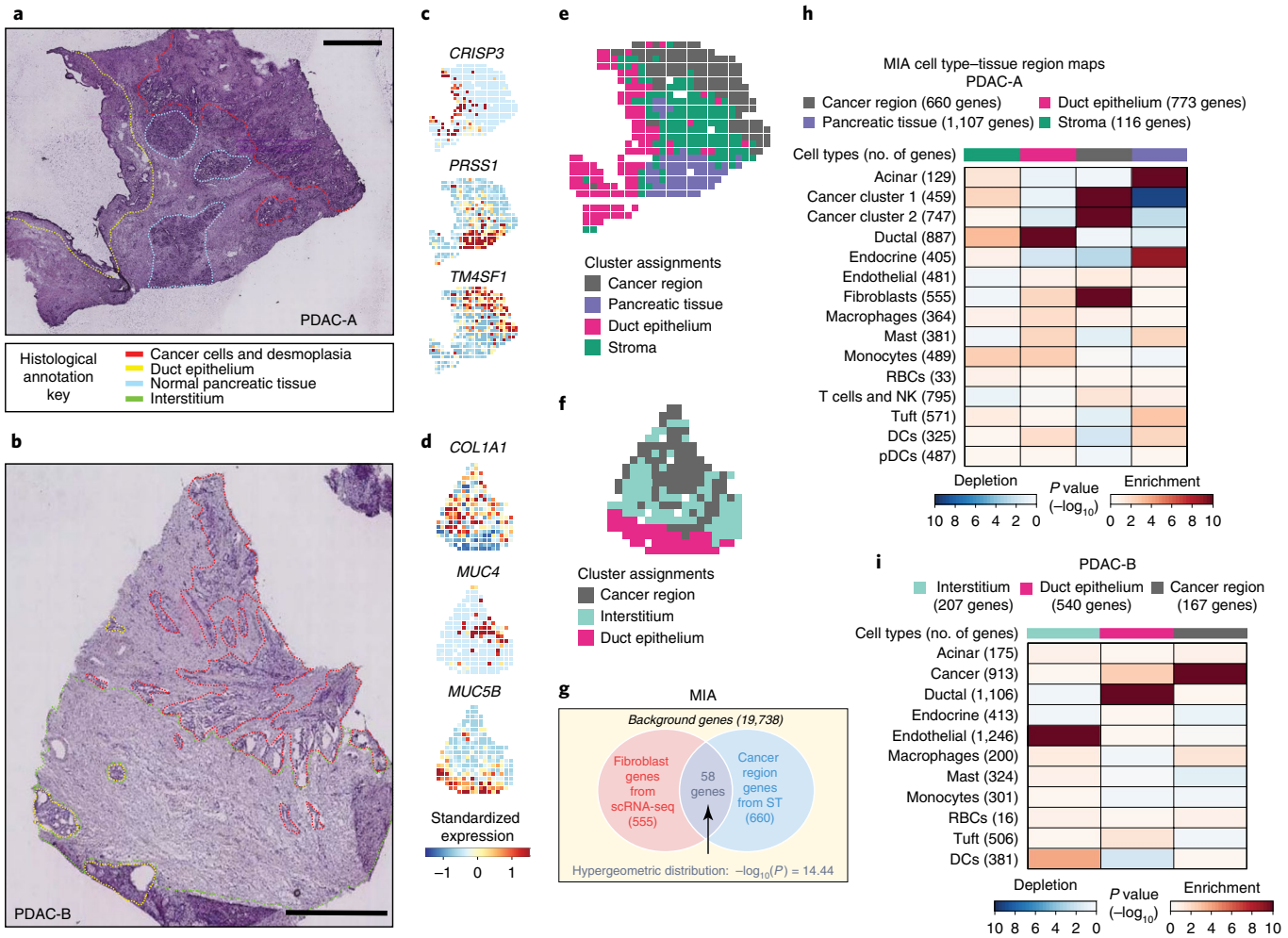

Fig. 2 | ST of PDAC and mapping of cell types. a, Annotated PDAC-A tumor cryosection on the ST slide. The annotations indicate a region high in cancer cells and desmoplasia (red), duct epithelium (yellow) and normal pancreatic tissue (blue). Scale bar, 1 mm. **b**, Annotated PDAC-B tumor cryosection on the ST slide. Annotated regions include a cancer cell-rich region (red), duct epithelium (yellow) and interstitium (green). Scale bar, 1 mm. **c**,**d**, Standardized expression levels of three genes in the PDAC-A ST (**c**) and PDAC-B ST (**d**) datasets. **e**,**f**, Clustering of the PDAC-A (**e**) and PDAC-B (**f**) ST spots. Color indicates the clustering assignments. **g**, MIA. The hypergeometric distribution is used to infer the significance of the intersection of genes specifically expressed in a given cell type (fibroblasts in this case) and genes specifically expressed in a given tissue region (cancer region). Applying this analysis systematically for all pairs of cell types and tissue regions allows for insight into the spatial distribution of the cell types in the tumor. The numbers of genes used in the calculation are shown. **h**, The PDAC-A MIA map of all scRNA-seq-identified cell types and ST-defined regions. Each element in the matrix is computed as described in **g** for all pairs of cell types and tissue regions using the same 19,738 background genes. The numbers of cell type- and tissue region-specific genes used in the calculation are shown. Red indicates enrichment (significantly high overlap); blue indicates depletion (significantly low overlap). The bar on top indicates the regions delineated in **e**. **i**, The PDAC-B MIA map for the ST clusters identified in **f**. Matrix elements were computed as described in **g** and **h**. The bar on top indicates the regions delineated in **f**.

II molecules are primarily expressed on the surface of professional antigen-presenting cells (B-cells, macrophages, dendritic cells), epithelial cells in the liver, gastrointestinal and respiratory tracts are known to express MHC class II (refs. [45,46]). Because of their presence in the tumor, it is likely that these antigen-presenting ductal cells play a role in modulating the inflammatory response within the tumor microenvironment by promoting T-cell activation[45,46]. Notably, past work studying pancreatic ductal cells at single-cell resolution reported the presence of terminal and centroacinar ductal subpopulations[44], but did not detect populations displaying the particular antigen-presenting or *APOL1*-high/hypoxic gene signatures present in the PDAC tumor ductal cells described here. When we performed double immunofluorescence of archival FFPE patient tissue, we found colocalization of subpopulation markers with the duct marker KRT19, confirming the presence of these ductal cell subpopulations (Fig. 3e–h and Supplementary Fig. 6).

As in our cell type analysis, we used marker genes specific to each ductal subpopulation to determine the enrichment of these ductal subpopulations across the tissue region using MIA. As expected, we found that all ductal subpopulations in PDAC-A were enriched in the duct region of the tissue. In contrast, only the hypoxic and terminal ductal cell populations were significantly enriched in the cancer region ($P < 10^{-4}$; Fig. 3i). It is possible that the transcriptional phenotypes exhibited by these ductal cells reflect environmental signals from the surrounding tissue; that is, ductal cells in the cancer region may express hypoxia-response genes due to low oxygen content. Additionally, the hypoxic ductal cells appeared to be depleted in the pancreatic tissue ($P < 10^{-3}$; Fig. 3i). In contrast, the ductal subpopulations in PDAC-B were all exclusively enriched in the ducts of the tissue ($P < 10^{-4}$; Fig. 3j).

We also found that the PDAC-A macrophages comprised two subpopulations, one of which expressed *IL1B* corresponding to an

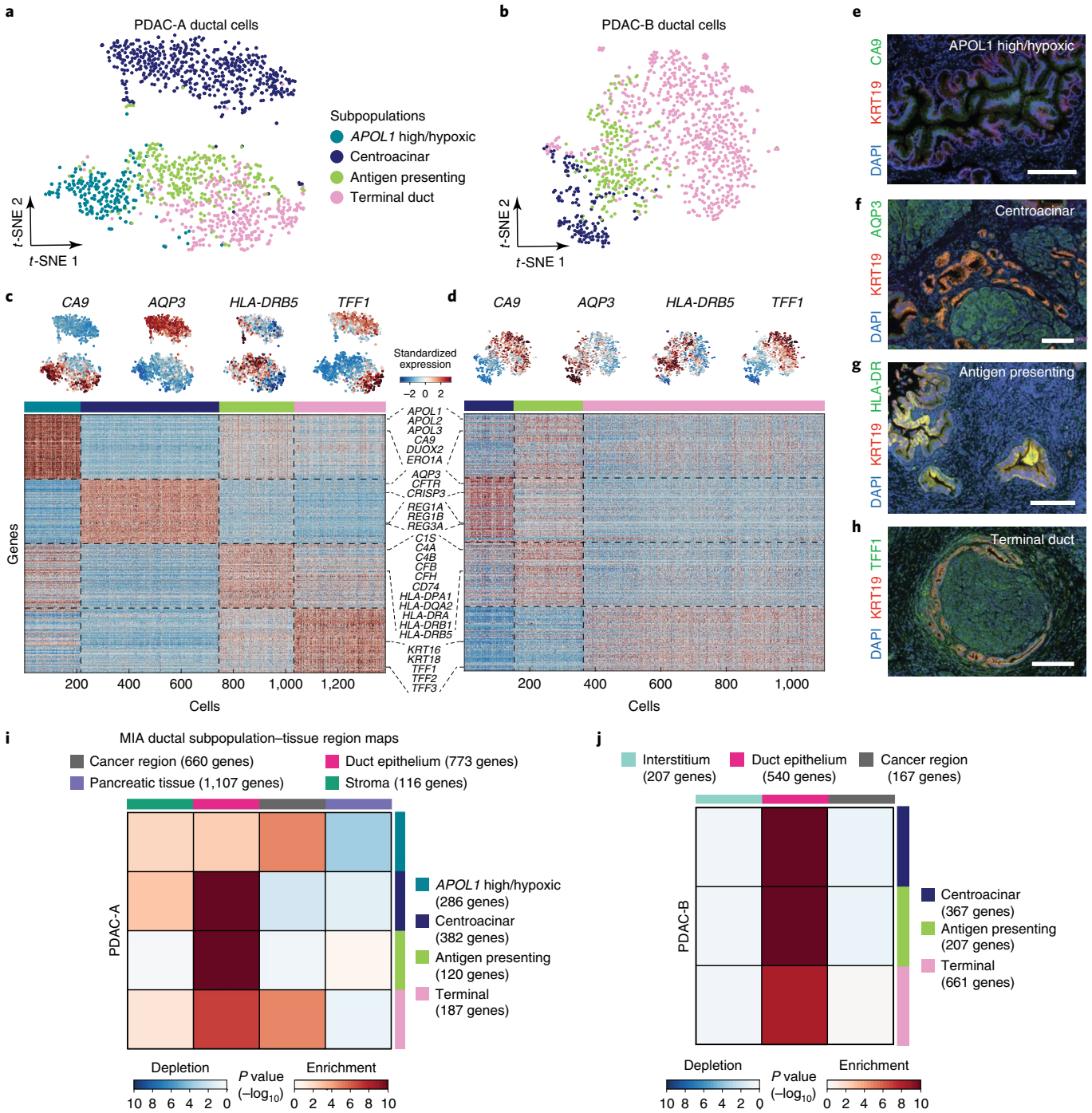

Fig. 3 | MIA mapping of ductal subpopulations across tissue regions. a,**b**, Identifying subpopulations of ductal cells in PDAC-A (**a**) and PDAC-B (**b**). Colors in the *t*-SNE projections indicate the identified subpopulations. **c**,**d**, Expression levels for genes with subpopulation-specific patterns in PDAC-A (**c**) and in PDAC-B (**d**) ductal cells. Top, standardized expression of subpopulation-specific marker genes projected onto *t*-SNE space. Bottom, heatmaps showing the standardized expression of the top 200 genes with subpopulation-specific expression. The genes are ordered identically across the two heatmaps. The color bars above the respective heatmaps reflect the subpopulation color coding indicated in panels **a** and **b**. **e**–**h**, Double immunofluorescence staining ($n = 2$) of KRT19 (ducts) and subpopulation markers CA9 (**e**, APOL1-high/hypoxic ductal cells), AQP3 (**f**, centroacinar ductal cells), HLA-DR (**g**, antigen-presenting ductal cells) and TFF1 (**h**, terminal ductal cells). Scale bar, 250 µm. **i**,**j**, MIA maps of PDAC ductal subpopulations across the ST regions identified in Fig. 2 for PDAC-A (**i**) and PDAC-B (**j**). Indicated are the numbers of subpopulation-specific and tissue region-specific genes used for *P* value calculation using the hypergeometric distribution.

inflammatory M1 state[47], and the other resembling the M2 alternatively activated state based on its expression of *CD163* and *MS4A4A* (ref. [48]) (Supplementary Fig. 7a). When we tested for the enrichment of these macrophage subpopulations across the tissue with MIA, we found that they appeared to have opposite patterns of enrichment across the tissue (Supplementary Fig. 7b). The M2-like macrophages were most enriched in the ducts, consistent with their function as tissue resident macrophages, while the M1 macrophages were more enriched in the stroma and cancer regions, reflecting an inflammatory environment in these regions (Supplementary Fig. 7b).

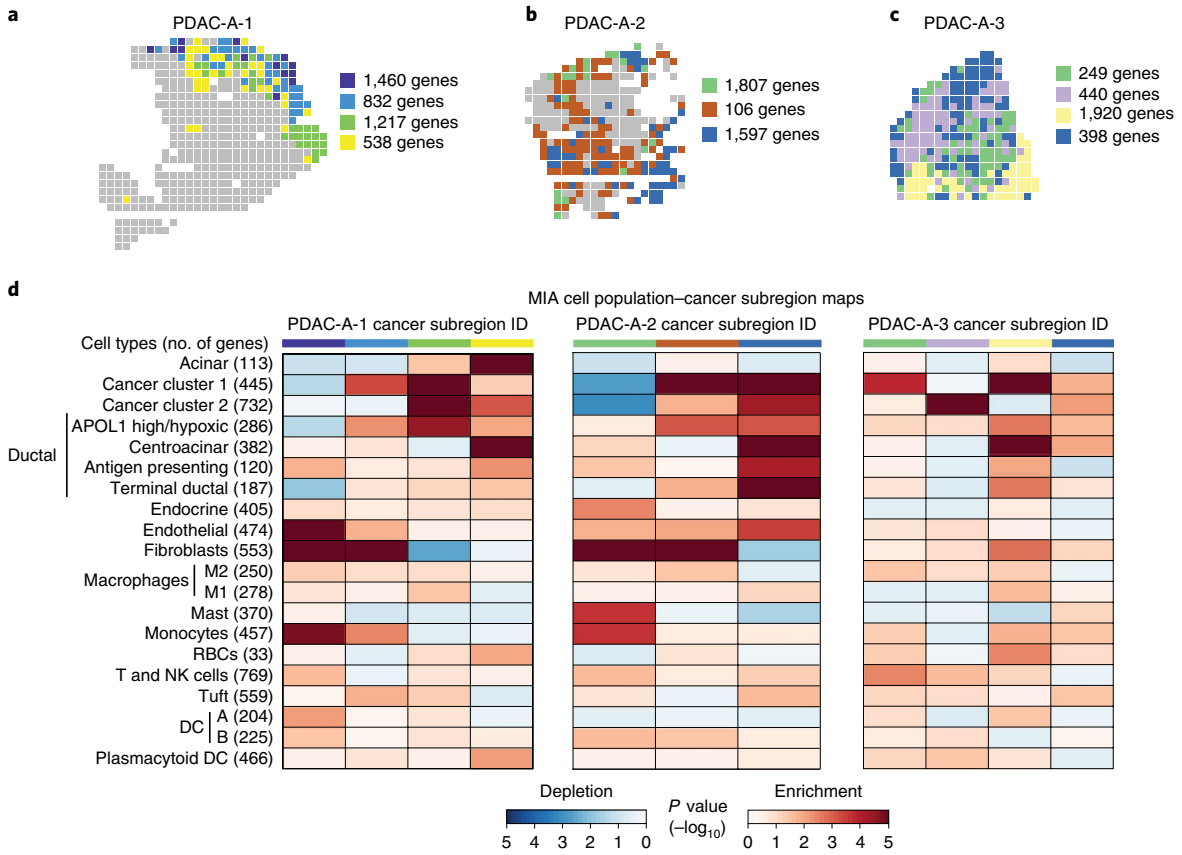

Fig. 4 | Cancer subregions reveal differential cell type and subpopulation enrichments. a–**c**, Identifying subregions in the DAC-A-1 (**a**), PDAC-A-2 (**b**), and PDAC-A-3 (**c**) ST cancer regions. Colors indicate the cluster assignments for ST spots annotated to be in the cancer regions, identified as in Fig. 2. Analysis of PDAC-A-3 ST data includes all tissue regions based on high cancer cell content (Supplementary Fig. 8a). **d**, MIA maps of the cancer identified subregions in PDAC-A and the cell types and cell subpopulations identified in Figs. 1 and 3. Color bars above the MIA maps reflect ST region subclustering assignments in panels **a**–**c**. Indicated are the numbers of cell type-specific and cancer subregion-specific genes used for *P* value calculation using the hypergeometric distribution.

We also found two subpopulations of dendritic cells, A and B, with subpopulation B expressing higher levels of complement pathway genes and MHC class II (Supplementary Fig. 7c). Dendritic cell subpopulation A was most enriched in pancreatic tissue, while subpopulation B appeared most enriched in the ducts of the tissue (Supplementary Fig. 7d). Based on the MIA maps, it is likely that these subpopulations play unique roles in the tissue based on their differential localization.

Mapping distinct cancer populations across PDAC tissue sections. We found two cancer cell populations in the PDAC-A scRNA-seq data that appeared to be both genetically and transcriptionally distinct (Fig. 1b–e). Although we found both cancer populations to be highly enriched in the corresponding ST cancer region (Fig. 2h), we asked whether they colocalized with different cell types within this region. To address this question, we processed for ST analysis two additional tissue sections originating from different regions of the same PDAC-A tumor (Supplementary Fig. 8a–d). Our protocol ensures that the scRNA-seq data are representative of all ST sections: the tumor was first divided into three parts, then each of these was split between ST and scRNA-seq (see Methods). Comparing regions across PDAC-A replicates, we determined the overlap among the list of genes specific to the respective PDAC-A sample regions, and found significant overlap between the cancer and stromal regions across the ST datasets (Supplementary Fig. 8e). For the cancer-rich regions in each tissue section, we used hierarchical clustering to further divide the cancer-rich regions into transcriptionally coherent subregions (Fig. 4a–c). After defining sets of genes specific to each subregion, we again applied our MIA mapping approach to study the overlap between the gene sets associated with the subregions and the sets of genes defined for each cell type and subpopulation from the scRNA-seq data (Fig. 4d). For example, in cancer subregion 2 of the PDAC-A-1 sample, we found enrichment of only the cancer cluster 1 subpopulation together with fibroblasts. Consistently across the three PDAC-A tissue sections, we observed an enrichment of fibroblasts in tissue regions with high enrichment of cancer cluster 1 but weak or no enrichment of cancer cluster 2 (Fig. 4d). This pattern suggests that the cancer cluster 1 cells may illicit a particular stromal response in the tissue, or that the cancer cluster 2 cells are mutually exclusive with fibroblasts in the tissue.

Deconvolving cell state relationships in the tumor microenvironment. Recent work on scRNA-seq data of cancer cells has revealed unique cell states in multiple cancer types, including in glioblastoma[4,11], melanoma[6], and head and neck cancer[17]. Since ST provides spatial information, we asked whether we could map cancer cell states to distinct spatial tissue regions and characterize their interactions with other cell types using our MIA approach. For this we generated a third PDAC scRNA-seq dataset (Supplementary Fig. 9) and used the combined scRNA-seq datasets to define three gene expression modules among the PDAC cancer cells (Fig. 5a and see Methods). Based on the genes of each module, we annotated these

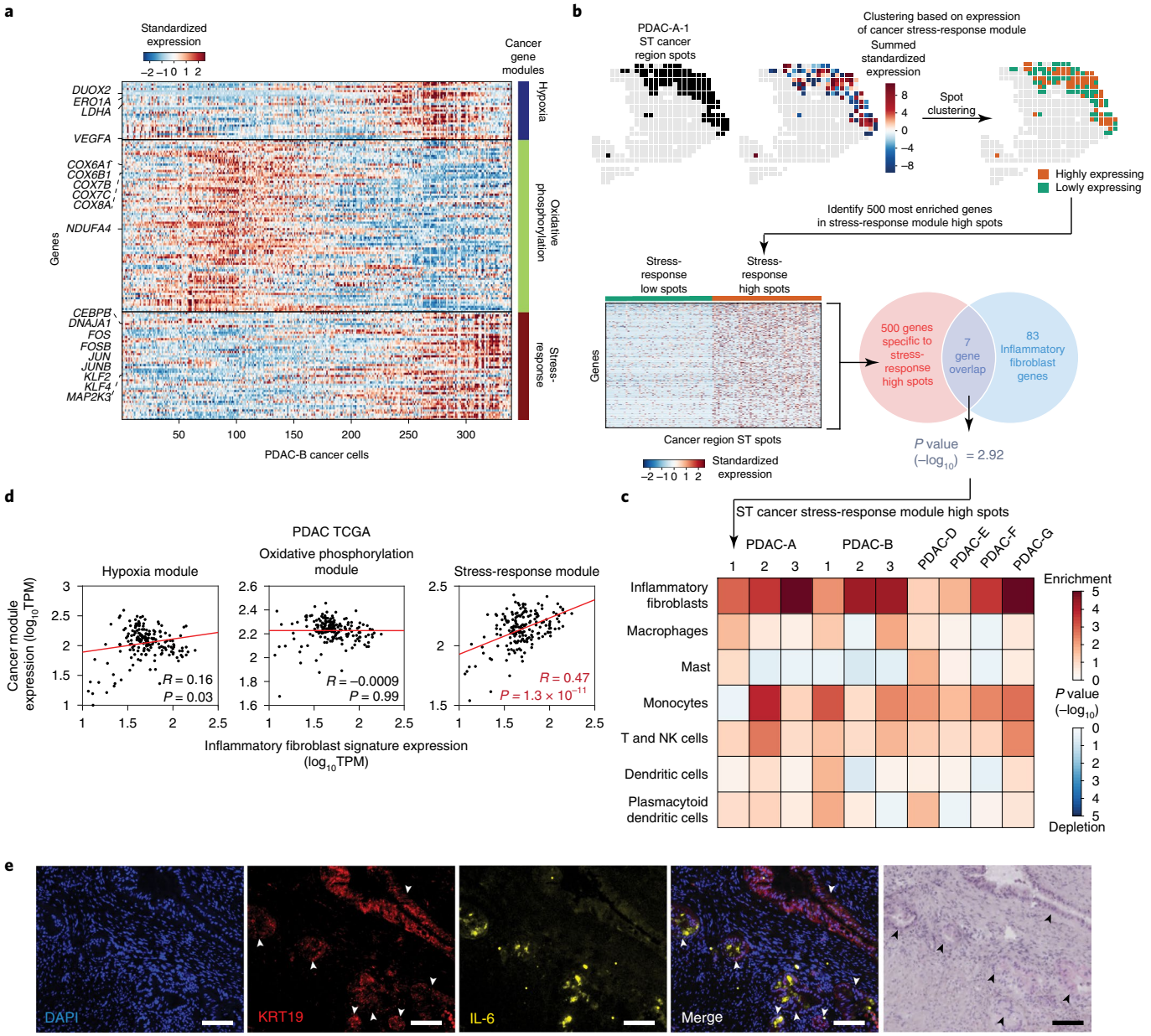

Fig. 5 | Mapping the relationship between cancer cell states and stromal subtypes using MIA. a, Heatmap indicating the expression of three NMF-defined gene modules across the PDAC-B cancer cells. Modules were defined based on all three PDAC scRNA-seq datasets (see Methods). **b**, Analysis workflow for studying coenrichment of cancer cell gene modules and cell type-specific genes in the ST cancer region. The example shown is for the cancer stress-response module and inflammatory fibroblasts in the PDAC-A-1 ST with 19,738 genes as the background for calculating the hypergeometric distribution. **c**, MIA map indicating coenrichment of genes of the indicated cell types and genes associated with the cancer stress-response module across ten ST samples. **d**, Plots indicate the expression of the inflammatory fibroblast gene signature and the cancer gene expression modules in the 183 PDAC TCGA datasets. Red line represents linear least squares regression of expression data. Pearson's correlation was used to calculate *R* and *P* value. **e**, Immunofluorescence localization of IL-6 cytokine in PDAC tissue sections. KRT19 signal identifies cancer cells in the tissue, as indicated by the white arrows ($n = 3$). Far right panel shows H&E staining of the same region of tissue in an adjacent tissue section (8 µm). Black arrows confirm the presence of cancer cells. Scale bar, 100 µm.

as hypoxia-response, oxidative phosphorylation and stress-response modules. In particular, the stress-response module was of interest to us as this has also been studied by our group and others[6,17,19]. Furthermore, several genes in this module have been implicated in the regulation of cell growth[49] and invasive phenotypes[50], suggesting a role for this gene module in tumor progression.

To adapt MIA to study the stress-response cancer cell state, we identified the genes with expression associated in those spots with high expression of the stress module. We first distinguished two clusters of ST cancer spots based on high and low expression of the stress-response module (Fig. 5b and see Methods). Next, we identified the genes that are specifically expressed in this cancer subregion and intersected them with cell type-specific genes using MIA. Using this analysis workflow, we found enrichment with the inflammatory fibroblasts in the PDAC-A-1 ST ($P = 0.0012$; Fig. 5b and see Methods). Similarly, we found strong enrichment between the genes of the stress-module high spots and the inflammatory fibroblasts in nine other ST arrays: two additional PDAC-A samples,

three from the patient PDAC-B, as well as four samples from four other PDAC tumors, thus spanning a cohort of six patient samples (Fig. 5c and Supplementary Figs. 10 and 11). We also detected significant enrichments of monocytes and T/natural killer cells in the stress-response module-high regions, though these enrichments were not as strong or as consistent as the inflammatory fibroblast enrichments.

The Cancer Genome Atlas (TCGA) project has made available the bulk transcriptomes of 183 PDAC tumors. We reasoned that if the colocalization of stress-response cancer cells and inflammatory fibroblasts reflects a functional relationship between these cells, a correlation should also be detectable at the level of entire tumors. Therefore, we computed the correlation between the expression of each identified cancer gene module and the inflammatory fibroblast gene signature. We found that only the stress-response gene module was significantly correlated with the inflammatory fibroblast signature in these datasets (Fig. 5d). To provide further evidence for the signaling relationship between inflammatory fibroblasts and cancer cells, we performed immunofluorescence on PDAC tissue sections for IL-6, which is primarily expressed by inflammatory fibroblasts compared with other cell types in the PDAC microenvironment[51]. By immunofluorescence staining of cancer cells with KRT19, we found evidence for colocalization between the presence of IL-6 and malignant cells (Fig. 5e). Through H&E staining of the tissue section adjacent to the section used for immunofluorescence, we confirmed that the epithelial cells throughout the tissue section indeed correspond to malignant cells (Fig. 5e, far right panel). Taken together, these results implicate a relationship between the inflammatory fibroblasts and cancer cells expressing a stress-response gene module.

Finally, we sought to build MIA maps of other tumors types by re-analyzing publicly available data using our approach. Studying metastatic melanoma tumors, we first analyzed the scRNA-seq data from Tirosh et al. (ref. [6]) to delineate cell types and subpopulations (Supplementary Fig. 12a,b). Using marker genes included in Tirosh et al., we identified three T-cell subsets[6] (Supplementary Fig. 12b). We then clustered the ST data of melanoma lymph node metastases from two patients analyzed in Thrane et al.[28], and found strong correspondence between the tissue histological annotations and the ST clustering assignments (Supplementary Fig. 12c,d, top). Next, we generated MIA maps of the Thrane et al. ST data using population-specific markers identified in the Tirosh et al. scRNA-seq dataset (Supplementary Fig. 12c,d, bottom). In both samples, we found that the stromal tissue compartment largely consisted of fibroblasts and endothelial cells. The colocalization of fibroblasts and endothelial cells was also seen in the PDAC MIA maps (Fig. 4c,d). In both melanoma MIA maps, we also find macrophages to be spatially restricted relative to the annotated melanoma region; in the melanoma 1 ST sample the macrophages are restricted to the melanoma region periphery, and in the second ST sample the macrophages are restricted to a particular region (melanoma region 1) within the larger annotated melanoma area. In both metastatic melanoma samples, we also find that the melanoma regions have a significant depletion of $CD8^+$ T cells, which are known be useful biomarkers for prognosis and response to therapy[52]. Collectively, these results demonstrate the utility of MIA maps for generating testable hypotheses regarding the relationships of cell populations with spatially restricted architectures.

Discussion

Here, we have presented a method for the identification and spatial mapping of distinct cell types, subpopulations and cell states within heterogeneous samples. The method begins with the characterization of cell types and subpopulations present in a tissue by scRNA-seq, and, in parallel, the identification of transcriptomic regions by ST. We then take advantage of the systematic and unbiased nature of both data modalities to detect cell population enrichment across coherent transcriptomic regions using an approach we term MIA. While other methods integrate across data modalities[53,54], MIA provides a unique approach for integration on initial independent annotations. By applying MIA to ten PDAC ST samples, we mapped the location of distinct cell types (Fig. 2) and subpopulations (Fig. 3 and Supplementary Fig. 7) in the tumor microenvironment, and the relationships among cell types within cancer subregions (Fig. 4).

Given the elucidation of cancer cell states through scRNA-seq studies[5–7,10,13–21] and that extensive communication between stromal cells and cancer cells is known to influence transcriptional states in PDAC[55], we sought to test whether we could identify unique PDAC cancer cell states and link these states to the localization of other cell types in the microenvironment. Of particular interest was the interplay between cancer cell states and fibroblast subtypes, including inflammatory fibroblasts[51,56]. By generating MIA maps across ten ST samples, we found evidence that the stress-response cancer cell state colocalizes with the inflammatory fibroblasts (Fig. 5). Interestingly, the inflammatory fibroblasts have been identified as the major source of IL-6 cytokine in PDAC; this cytokine participates in a number of signaling cascades with factors encoded by our identified stress-response genes[57–59]. We found further support for this association by an analysis of the PDAC TCGA data and through immunofluorescence experiments. These results highlight the applicability of our MIA approach to reveal biological insights from systematic gene expression explorations.

Established methods for mapping transcripts (ISH, fluorescence in situ hybridization) or proteins (immunohistochemistry) are limited to a few antibodies or in situ probes, even with multiplexing. In this respect, ST has the advantage of providing an unbiased map of expressed transcripts in a given tissue section[28–31,33]. More recent spatially resolved methods can detect hundreds of RNA species at single-cell resolution[20], but these methods do not provide transcriptome-wide measurements and consequently cannot identify all of the genes that are specifically expressed in a spatially defined tissue region. Although the sensitivity of ST may not enable the determination of specific cell type enrichment in a single ST spot, our MIA approach aims to mitigate this constraint by determining enrichment of scRNA-seq-identified cell types and cell states across clusters of spots for which gene signatures can be reliably identified. Since ST spots are clustered based on the transcriptomic data independently of tissue histology, MIA maps provide an unbiased perspective into the organization of distinct populations in local tissue niches. This allows for the inference of functional relationships between scRNA-seq-defined populations based on their colocalization in space, and ultimately provides a more comprehensive characterization of cell types in their native environment than can be gained from either modality alone. As the sequencing depth achieved by this method is expected to increase, ST is well-positioned to study histologically variable tissues such as tumors, which lack ISH maps to guide cell location inference[25,26].

Our approach to integrating scRNA-seq and ST using MIA has important limitations. First, the ST array is about $6 \times 6.5\,mm^2$ in size; thus, in many cases the array is not large enough to cover the entire tissue, nor does each spot achieve single-cell resolution. With the improvement of the ST technology, the method will be empowered with more spots on the array and increased sensitivity, as well as an improved protocol to further reduce the diffusion of transcripts across spots. Though tissue optimization experiments are already performed to minimize transcript diffusion[27], these steps do not completely rule out this possibility; this is particularly true for tissues with highly variable tissue architecture, for which any one tissue permeabilization condition necessary for ST may not be necessarily optimal for all histological features captured on the ST array. Despite these limitations, ST remains very accessible to researchers as it requires only a few additional steps compared with bulk

RNA-seq analysis of homogenized tissue, and, unlike other spatially resolved transcriptomic approaches[20,21], these additional steps do not require specialized equipment outside of standard laboratory equipment[27].

Given the strength of scRNA-seq for the identification of cell states and subpopulations, the unique perspective offered from MIA maps can aid in assigning potential functional roles for identified cell states and subpopulations based on spatial localization (relative to the tissue or to other cell types present). In the case of tumors for which the precise composition of different tumor subclassifications is likely to vary from individual to individual, the subpopulation composition and spatial localization can be ascertained for a given patient and could in the future be of prognostic value.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-019-0392-8.

Received: 7 April 2019; Accepted: 11 December 2019;
Published online: 13 January 2020

References

1. Offit, K. A decade of discovery in cancer genomics. *Nat. Rev. Clin. Oncol.* **11**, 632–634 (2014).
2. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).
3. Papalexi, E. & Satija, R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* **18**, 35–45 (2017).
4. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
5. Tirosh, I. et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).
6. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
7. Venteicher, A. S. et al. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science (80-.)* **355**, eaai8478 (2017).
8. Darmanis, S. et al. Single-cell RNA-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. *Cell Rep.* **21**, 1399–1410 (2017).
9. Chung, W. et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat. Commun.* **8**, 15081 (2017).
10. Horning, A. M. et al. Single-cell RNA-seq reveals a subpopulation of prostate cancer cells with enhanced cell cycle-related transcription and attenuated androgen response. *Cancer Res.* **78**, 853–864 (2017).
11. Dirkse, A. et al. Stem cell-associated heterogeneity in glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment. *Nat. Commun.* **10**, 1787 (2019).
12. Filbin, M. G. et al. Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science (80-.)* **360**, 331–335 (2018).
13. Lawson, D. A. et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* **526**, 131–135 (2015).
14. Savage, P. et al. A targetable EGFR-dependent tumor-initiating program in breast cancer. *Cell Rep.* **21**, 1140–1149 (2017).
15. Sharma, A. et al. Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. *Nat. Commun.* **9**, 4931 (2018).
16. Dalerba, P. et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* **29**, 1120–1127 (2011).
17. 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).
18. Rambow, F. et al. Toward minimal residual disease-directed therapy in melanoma. *Cell* **174**, 843–855 (2018).
19. Baron, M. et al. Cancer archetypes co-opt and adapt the transcriptional programs of existing cellular states. Preprint at *bioRxiv* https://doi.org/10.1101/396622 (2018).
20. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
21. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016).
22. Long, X., Colonell, J., Wong, A. M., Singer, R. H. & Lionnet, T. Quantitative mRNA imaging throughout the entire drosophila brain. *Nat. Methods* **14**, 703–706 (2017).
23. Lee, J. H. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science (80-.)* **343**, 1360–1363 (2014).
24. Nichterwitz, S. et al. Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nat. Commun.* **7**, 12139 (2016).
25. Achim, K. et al. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).
26. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
27. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
28. Thrane, K., Eriksson, H., Maaskola, J., Hansson, J. & Lundeberg, J. Spatially resolved transcriptomics enables dissection of genetic heterogeneity in stage III cutaneous malignant melanoma. *Cancer Res.* **78**, 5970–5979 (2018).
29. Berglund, E. et al. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat. Commun.* **9**, 2419 (2018).
30. Lundmark, A. et al. Gene expression profiling of periodontitis-affected gingival tissue by spatial transcriptomics. *Sci. Rep.* **8**, 9370 (2018).
31. Asp, M. et al. Spatial detection of fetal marker genes expressed at low level in adult human heart tissue. *Sci. Rep.* **7**, 12941 (2017).
32. Maniatis, S. et al. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science* **364**, 89–93 (2019).
33. Giacomello, S. et al. Spatially resolved transcriptome profiling in model plant species. *Nat. Plants* **3**, 17061 (2017).
34. Junker, J. P. et al. Genome-wide RNA tomography in the zebrafish embryo. *Cell* **159**, 662–675 (2014).
35. Chen, J. et al. Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nat. Protoc.* **12**, 566–580 (2017).
36. Saiselet, M. et al. Transcriptional output, cell types densities and normalization in spatial transcriptomics. Preprint at *bioRxiv* https://doi.org/10.1101/503870 (2018).
37. Wagner, F., Yan, Y. & Yanai, I. K-nearest neighbor smoothing for high-throughput single-cell RNA-Seq data. Preprint at *bioRxiv* https://doi.org/10.1101/217737 (2018).
38. Ziegenhain, C. et al. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* **65**, 631–643.e4 (2017).
39. Griffin, C. A. et al. Consistent chromosome abnormalities in adenocarcinoma of the pancreas. *Cancer Res.* **55**, 2394–2399 (1995).
40. Shiga, K. et al. Cancer-associated fibroblasts: their characteristics and their roles in tumor growth. *Cancers (Basel)* **7**, 2443–2458 (2015).
41. Motta, P. M., Macchiarelli, G., Nottola, S. A. & Correr, S. Histology of the exocrine pancreas. *Microsc. Res. Tech.* **37**, 384–398 (1997).
42. May, D. et al. Ero1-Lα plays a key role in a HIF-1-mediated pathway to improve disulfide bond formation and VEGF secretion under hypoxia: implication for cancer. *Oncogene* **24**, 1011–1020 (2005).
43. Sedlakova, O. et al. Carbonic anhydrase IX, a hypoxia-induced catalytic component of the pH regulating machinery in tumors. *Front. Physiol.* **4**, 400 (2014).
44. Baron, M. et al. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst.* **3**, 346–360 (2016).
45. Mehrfeld, C., Zenner, S., Kornek, M. & Lukacs-Kornek, V. The contribution of non-professional antigen-presenting cells to immunity and tolerance in the liver. *Front. Immunol.* **9**, 635 (2018).
46. Wosen, J. E., Mukhopadhyay, D., Macaubas, C. & Mellins, E. D. Epithelial MHC class II expression and its role in antigen presentation in the gastrointestinal and respiratory tracts. *Front. Immunol.* **9**, 2144 (2018).
47. Rőszer, T. Understanding the mysterious M2 macrophage through activation markers and effector mechanisms. *Mediators Inflamm.* **2015**, 816460 (2015).
48. Sanyal, R. et al. MS4A4A: a novel cell surface marker for M2 macrophages and plasma cells. *Immunol. Cell Biol.* **95**, 611–619 (2017).
49. Lopez-Bergami, P., Lau, E. & Ronai, Z. Emerging roles of ATF2 and the dynamic AP1 network in cancer. *Nat. Rev. Cancer* **10**, 65–76 (2010).
50. Hyakusoku, H. et al. JunB promotes cell invasion, migration and distant metastasis of head and neck squamous cell carcinoma. *J. Exp. Clin. Cancer Res.* **35**, 6 (2016).
51. Öhlund, D. et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J. Exp. Med.* **214**, 579–596 (2017).

52. Obeid, J. M., Hu, Y., Erdag, G., Leick, K. M. & Slingluff, C. L. The heterogeneity of tumor-infiltrating CD8+ T cells in metastatic melanoma distorts their quantification. *Melanoma Res.* **27**, 211–217 (2017).
53. Welch, J. D. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887.e17 (2019).
54. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
55. Ligorio, M. et al. Stromal microenvironment shapes the intratumoral architecture of pancreatic cancer. *Cell* **178**, 160–175.e27 (2019).
56. 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).
57. Akira, S. et al. A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family. *EMBO J.* **9**, 1897–1906 (1990).
58. Schuringa, J.-J., Timmer, H., Luttickhuizen, D., Vellenga, E. & Kruijer, W. c-Jun and c-Fos cooperate with Stat3 in IL-6-induced transactivation of the IL-6 response element (IRE). *Cytokine* **14**, 78–87 (2001).
59. Venugopal, R. & Jaiswal, A. K. Nrf2 and Nrf1 in association with Jun proteins regulate antioxidant response element-mediated expression and coordinated induction of genes encoding detoxifying enzymes. *Oncogene* **17**, 3145–3156 (1998).

Methods

Tumor sample handling and dissociation to a single-cell suspension for scRNA-seq. Three PDAC tumors were delivered in RPMI (Fisher Scientific) on ice directly from the operating room to the laboratory after clearing pathology (~2 h). Two tumor samples (PDAC-A and PDAC-B) were processed for ST and scRNA-seq in parallel: tumors were rinsed in ice-cold PBS and cut into ~4–5-mm^3 pieces from which 1-mm-thick slices were taken and set aside in ice-cold PBS. The remaining ~3–4-mm^3 pieces were embedded in optimal cutting temperature compound (OCT) and frozen in isopentane cooled with liquid N_2. The 1-mm tissue slices were further minced with scalpels to <1 mm^3. Tissue was collected and pelleted by centrifuging at 300*g* for 3 min at 4 °C. PBS was aspirated and 5 ml of 0.25% prewarmed trypsin-EDTA with 10 U µl^{-1} DNaseI (Roche) was added and put into a 37 °C water bath for 30 min with gentle inversion every 5 min. The resulting suspension was filtered through a 100-µm cell strainer to remove larger chunks of undigested tissue. Enzymatic digestion was quenched with the addition of FBS to a final concentration of 10%. Cells were pelleted by centrifuging the suspension at 800*g* for 3 min at 4 °C and washed twice with 5 ml of ice-cold PBS. After a final spin at 300*g* for 3 min, the cells were resuspended in PBS to a final concentration of 10,000 cells per ml. The PDAC-C tumor sample (Supplementary Fig. 9) was only processed for scRNA-seq, involving enzymatic digestion as described in this section. The resulting viability was >95% as shown by trypan blue exclusion.

inDrop library preparation and scRNA-seq. From each single-cell suspension, between 6,000 and 12,000 cells were encapsulated using the inDrop platform and the reverse transcription reaction was performed as previously described[44]. Each library was prepared from aliquots consisting of 2,000–2,500 encapsulated cells. Libraries were diluted to 4 nM and paired-end sequencing was performed on an Illumina NextSeq platform. Between 139 million and 145 million paired-reads were generated for each library (estimated 2,000–2,500 encapsulated cells), corresponding to ~58,000 paired-reads per cell before computational filtering of low-quality cells.

scRNA-seq initial processing. Raw sequencing data obtained from the inDrop method were processed using a custom-built pipeline (https://github.com/flo-compbio/singlecell). Briefly, the 'W1' adapter sequence of the inDrop reverse transcription primer was located in the barcode read (read 2), by comparing the 22-mer sequences starting at positions 9–12 of the read with the known W1 sequence ('GAGTGATTGCTTGTGACGCCTT'), allowing at most two mismatches. Reads for which the W1 sequence could not be located in this way were discarded. The start position of the W1 sequence was then used to infer the length of the first part of the inDrop cell barcode in each read (8–11 base pairs), as well as the start position of the second part of the inDrop cell barcode (8 base pairs). Cell barcode sequences were mapped to the known list of 384 barcode sequences for each read, allowing at most one mismatch. The resulting barcode combination was used to identify the cell from which the fragment originated. Finally, the UMI sequence was extracted, and reads with low-confidence base calls for the six bases comprising the UMI sequence (minimum Phred score less than 20) were discarded. The reads containing the messenger RNA sequence were mapped by STAR 2.5.1 with parameter '—outSAMmultNmax 1' and default settings otherwise[60]. Mapped reads were split according to their cell barcode and assigned to genes by testing for overlap with exons of protein-coding genes. Only single-cell transcriptomes with ≥1,000 UMIs, ≤20% mitochondrial transcripts and ≤30% ribosomal transcripts were kept. UMI counts were normalized by the total transcript count and scaled by the median count (transcripts per median, TPM). Expression was transformed using the Freeman–Tukey transform as described previously[37].

KNN smoothing of scRNA-seq data. To reduce the noise inherent to scRNA-seq data[38], we applied KNN smoothing[37]. Briefly, all single-cell expression profiles were normalized by the total transcript count and scaled by the median number of total transcripts across all cells, the Freeman–Tukey transformation was applied to all expression values and the k closest neighbors of each cell were identified using Euclidean distance. The expression profile of each cell was then combined with those of its neighbors, thus obtaining its smoothed expression profile.

Hierarchical clustering of scRNA-seq data and marker gene selection. For clustering and identification of cell types, we used a recursive clustering scheme involving data smoothing, clustering, removal of identified and repeating until all clusters are identified. The data were first smoothed using KNN smoothing[37], normalized and scaled by the median transcript count (TPM) and transformed using the Freeman–Tukey transformation $y=\sqrt{x}+\sqrt{(x+1)}$, where 'x' is the TPM-normalized expression of any given gene for any given cell, and 'y' is the Freeman–Tukey transformed expression. Cells were then clustered with Ward's criterion using the most variable genes (defined as Fano factor and mean expression above mean-dependent threshold). From resulting clusters, we obtained a list of marker genes by examining genes that were differentially expressed ($P<10^{-5}$, two-tailed Student's t-test; effect size >0.2, Cohen's d). For each gene with $P<10^{-5}$, we examined for which cell type the effect size was highest and determined that gene to be specific for this cell type. While clustering the PDAC-B data, we identified a population of ductal cells with low UMI counts and high mitochondrial content. Because of these reasons, and the lack of a similar population in PDAC-A and PDAC-C, we suspected these cells to be low-quality cells that may have arisen as an artifact of the dissociation process. These cells were therefore removed from the PDAC-B dataset.

Correlation between cell types in the scRNA-seq data. To determine the similarity between cell types identified in PDAC-A and PDAC-B, the transcriptomes of all cells in each cluster were averaged (after TPM normalization and applying the Freeman–Tukey transform, see above). The top most variably expressed genes within each scRNA-seq were identified using the Fano factor (see above), and the union between these two gene sets was used to compute the Pearson's correlation coefficient between the averaged cell type profiles.

Identification of ductal subpopulations. To subcluster the ductal subpopulations, the ductal cells were first isolated from the expression matrix and smoothed using KNN smoothing[37] with $k=64$. Cells were then clustered with hierarchical clustering using the top variably expressed genes (see section 'Hierarchical clustering of scRNA-seq data and marker gene selection'). To annotate subpopulations, differentially expressed genes between each subpopulation were identified using a two-tailed Student's t-test ($P<10^{-5}$). The top 200 differentially expressed genes for each subpopulation were used for the heatmap visualization shown in Fig. 3c,d.

Dimensionality reduction (PCA and *t*-SNE). Dimensionality reduction methods were performed on the Freeman–Tukey-transformed data (after normalizing UMI counts to the median as described in 'scRNA-seq initial processing') using variable genes (defined as Fano factor and mean expression above mean-dependent threshold). t-distributed stochastic neighbor embedding (t-SNE) was performed using the following parameters: perplexity = 30 and initial dimension = number of principal components explaining >90% of the variance[61].

CNV analysis. To estimate CNV profiles, PDAC-A, PDAC-B and PDAC-C scRNA-seq gene expression matrices were smoothed with KNN smoothing[37], using $k=10$. Next, genes were sorted based on their chromosomal location and a moving average of gene expression was calculated using a window size of 0.1 multiplied by the number of genes of each chromosome. The expression was then centered to zero by subtracting the mean. A subset of 200 randomly selected ductal cells was removed as a negative control for the analysis, leaving all remaining cells as the background.

Tissue preparation, cryosectioning, fixation, staining and brightfield imaging for ST. Patients at New York University (NYU) Langone Health consented preoperatively to participate in the study. PDAC tissue was gently washed with cold PBS and cut into 4–5-mm^3 pieces. From each piece of fresh tissue, a 1-mm-thick portion of tissue was removed for preparing a single-cell suspension (see above). The surface from which the 1-mm-thick tissue for scRNA-seq was cut was then placed cut side down into a plastic mold (such that the side of the tissue cut for preparing the single-cell suspension is then cryosectioned for ST). The OCT-filled mold was then snap frozen in chilled isopentane. Cryosections were cut at 10-µm thickness, mounted onto the ST arrays and stored at −80 °C until use. For processing, the tissue was first warmed to 37 °C for 1 min and fixed for 10 min with 3.6% formaldehyde in PBS. Next, the tissue was dehydrated with isopropanol for 1 min followed by staining with H&E. Slides were mounted in 80% glycerol and brightfield images were taken on a Leica SCN400 F whole-slide scanner at 40× resolution.

ST barcoded microarray slide information. Library preparation slides used were purchased from the Spatial Transcriptomics team (https://www.spatialtranscriptomics.com; lot 10002 for PDAC-A ST, lot 10003 for PDAC-B ST and lot 10010 for PDAC-D, -E, -F, -G ST). Each of the spots printed onto the array is 100 µm in diameter and 200 µm from center to center, covering an area of 6.1 × 6.5 mm^2. Spots are printed with approximately 2×10^8 oligonucleotides containing an 18-mer spatial barcode, a randomized 7-mer UMI and a poly-20TVN transcript capture region[27] (Fig. 1a).

On-slide tissue permeabilization, cDNA synthesis and probe release. After brightfield imaging, the ST slide was prewarmed to 42 °C and attached to a microarray slide module (Grace Biolabs). The sections were prepermeabilized with 0.2 mg ml^{-1} BSA and 200 units of collagenase diluted in HBSS buffer for 20 min at 37 °C and washed with 100 µl of 0.1× SSC buffer. Tissue was permeabilized with 0.1% pepsin in HCl for 4 min at 42 °C and washed with 100 µl of 0.1× SSC buffer. Reverse transcription was carried out overnight (~18–20 h) at 42 °C by incubating permeabilized tissue with 75 µl of cDNA synthesis mix containing 1× First strand buffer (Invitrogen), 5 mM dithiothreitol, 0.5 mM of each dNTP, 0.2 µg µl^{-1} BSA, 50 ng µl^{-1} Actinomycin D, 1% dimethylsulfoxide, 20 U µl^{-1} Superscript III (Invitrogen) and 2 U µl^{-1} RNaseOUT (Invitrogen). Tissue was then digested away from the slide by incubating the tissue with 1% 2-mercaptoethanol in RLT buffer (Qiagen) for 1 h at 56 °C with interval shaking, followed by digestion with

proteinase K (Qiagen) diluted 1:8 in PKD buffer (Qiagen) at 56 °C for 1 h with interval shaking. Slides were rinsed in 2× SSC with 0.1% SDS for 10 min at 50 °C, then in 0.2× SSC for 1 min at room temperature and finally in 0.1× SSC for 1 min at room temperature. Probes were cleaved from the slide by incubating arrays with 70 µl of cleavage mix (8.75 µM of each dNTP, 0.2 µg µl^{-1} BSA, 0.1 U µl^{-1} USER enzyme (New England Biolabs)) and incubated at 37 °C for 2 h with interval mixing. After incubation, 65 µl of cleaved probes was transferred to 0.2-ml low-binding tubes and kept on ice.

ST library preparation and sequencing. Libraries were prepared from cleaved probes as previously described[27], with the following changes: briefly, after in vitro transcription, a second reverse transcription reaction was performed using random hexamers, eliminating the need for a primer ligation step[62]. The remaining purified cDNA was indexed using the following program: 98 °C for 3 min, followed by 25 cycles of 98 °C for 20 s, 60 °C for 30 s and 72 °C for 5 min. Average lengths of the indexed, purified libraries were assessed using a 2100 Bioanalyzer (Agilent) and concentrations were measured using a Qubit dsDNA HS Assay Kit (Life Technologies) according to the manufacturer's instructions. Libraries were diluted to 4 nM and paired-end sequencing was performed on an Illumina NextSeq sequencer, with 31 cycles for read 1 and 55 cycles for read 2. Between 100 million and 200 million raw read-pairs were generated for each sequenced library.

ST spot selection and image alignment. After probe cleavage, the slide was then placed into a microarray cassette and incubated with 70 µl of hybridization solution (0.2 µM Cy3-A-probe (Cy3-AGATCGGAAGAGCGTCGTGT), 0.2 µM Cy3 Frame probe (Cy3-GGTACAGAAGCGCGATAGCAG), in 1× PBS) for 10 min at room temperature. The slide was subsequently rinsed in 2× SSC with 0.1% SDS for 10 min at 50 °C, followed by 1-min washes with 0.2× SSC and 0.1× SSC at room temperature. Fluorescent images were taken on a Hamamatsu NanoZoomer whole-slide fluorescence scanner. Brightfield images of the tissue and fluorescent images were manually aligned with Adobe Photoshop CS6 to identify the array spots beneath the tissue.

ST library sequence alignment and annotation. Raw ST sequencing data were processed using a custom-built pipeline (https://github.com/yanailab/celseq2). The pipeline was adapted to ST sequencing data using the following three steps: (1) tagging and demultiplexing: the left-most 25 nucleotides (nt) of the R1 sequence contain the 18 nt for the spot-specific barcode and then 7 nt for the UMI. The R2 sequence contains the transcript sequence, and its left-most 35 nt are used. The name of every read in the R2 file is tagged with the spot-specific barcode and UMI sequence that are extracted from the paired R1 read. The R2 file is demultiplexed to create the 1,007 spot-specific FASTQ files. If the detected spot-specific barcode of a read is not present in the predefined barcodes list, the read is excluded from the downstream analysis. (2) Alignment of demultiplexed FASTQ files using Bowtie2 version 2.3.1 (ref. [63]). (3) Counting UMIs using HTSeq version 0.9.1 (ref. [64]). The reads that are aligned to a user-defined feature are collapsed to count only once if they have same UMI. Bowtie alignment was performed with default parameters and UMI counting with HTSeq in 'union' mode.

ST data analysis. UMI counts in each spot were normalized by the total transcript count and then scaled by the median number transcript count across all spots. A pseudocount of 1 was added before $\log_{10}$ transformation. For PCA of ST data, the 200–700 most-variable genes were selected (defined by the Fano factor above a mean-dependent threshold). Principal component scores for the first six components were then plotted for each spot corresponding to PDAC tissue. For clustering of spots, hierarchical clustering was performed on principal component scores using Ward's criterion on the first five to ten components. To extract marker genes specific to each of the resulting region clusters, a two-tailed Student's *t*-test was used, $P < 0.01$.

Immunofluorescence staining of FFPE tissue. FFPE tissues were cut at 5-µm thickness and dried overnight. After deparaffinizing the slides, antigen retrieval was carried out by boiling the samples for 10–20 min in Tris-EDTA or citrate buffer in a microwave oven. Primary antibodies were diluted 1:100 in Tris-buffered saline + 0.5% BSA and incubated with slides overnight in a wet chamber kept at 4 °C. Secondary antibodies (Molecular Probes, Invitrogen) were diluted 1:200 in Tris-buffered saline and incubated with slides for 1 h at room temperature before mounting and imaging. Fluorescent images were taken on a Hamamatsu NanoZoomer whole-slide fluorescence scanner.

Determination of cell type enrichment/depletion by MIA. We queried the significance of the overlap between ST genes and cell type marker genes using the hypergeometric cumulative distribution, with all genes as the background to compute the *P* value. In parallel, we test for cell type depletion by computing $-\log_{10}(1-P)$.

Clustering ST cancer region into subregions. To further cluster the PDAC-A ST cancer regions, the ST spots enriched with cancer cells (determined using MIA) were first separated from the raw data and normalized as described above. The top 200–500 variably expressed genes were identified after Fano factor calculation and this gene list was intersected with the genes specific to the cancer cluster 1 and cluster 2 populations from the scRNA-seq data. The standardized expression of this new gene list was used to cluster the cancer region using hierarchical clustering with Ward's criterion. Genes specific to each resulting cancer subregion were identified using a two-tailed Student's *t*-test, $P < 0.05$.

Identifying cancer-specific gene modules using nonnegative matrix factorization (NMF) and mapping of cancer gene modules in ST data. After TPM normalization, the expression matrices for the *TM4SF1*-expressing cancer populations in PDAC-A, PDAC-B and PDAC-C were centered individually by removing the mean expression for each gene. Negative values were then set to zero, and the three expression matrices were combined. Sparse nonsmooth NMF was performed using the nsNMF function implemented in the R NMF package[65,66] with a rank of 20. We then obtained gene signatures from the basis matrix as previously described[66] by sorting the genes for each column and keeping only the consecutive genes with higher coefficients in that column than in all other columns. We then removed signatures with fewer than 20 genes.

To cluster the ST cancer region spots based on the expression of the stress-response gene module (related to Fig. 5b,c), we first summed the standardized, $\log_{10}$-transformed expression of the genes in this module across the cancer region ST spots. We then annotated spots with an expression level above the median as the highly expressing module spots, and the spots below the median as the lowly expressing module spots. Next, we identified genes specific to the highly expressing module spots with a two-tailed *t*-test, and took the 500 most-significant genes by *P* value for subsequent MIA cell type enrichment. For determining the enrichment of inflammatory fibroblasts in these regions, we obtained the scRNA-seq-defined inflammatory fibroblast signature from Elyada et al.[56].

Processing and analysis of PDAC TCGA data. TCGA data were retrieved using the 'TCGA2STAT' function in R. Data were normalized by the total number of counts in each sample, scaled by a factor of 10^6 (TPM) and $\log_{10}$-transformed after applying a pseudocount of 1. The average expression of all genes in the cancer gene modules and inflammatory fibroblast gene signature after TPM-normalization and $\log_{10}$-transformation was used for plotting and for computing the Pearson's correlation coefficient. The linear least squares regression line was plotted using the 'lsline' function in MATLAB.

Ethical compliance. The study was approved by the NYU School of Medicine Institutional Review Board and the Ethics Committee. The pancreas samples were obtained from patients that signed the universal consent form for tissue collection for research studies. The tissue used was not needed for clinical or diagnostic purposes.

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

Data availability

The scRNA-seq and ST data reported in this manuscript have been deposited to the Gene Expression Omnibus under accession number GSE111672.

References

60. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
61. Van Der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
62. Hashimshony, T. et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome. Biol.* **17**, 77 (2016).
63. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
64. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
65. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).
66. Carmona-Saez, P., Pascual-Marqui, R. D., Tirado, F., Carazo, J. M. & Pascual-Montano, A. Biclustering of gene expression data by non-smooth non-negative matrix factorization. *BMC Bioinformatics* **7**, 78 (2006).

Acknowledgements

We thank C. Loomis, Z. Dewan and B. Dabovic from the NYU Experimental Pathology Core, and B. Zeck and L. Chiriboga from the NYU Center for Biospecimen Research and Development for technical assistance, A. Weil from the NYU CBRD for sample acquisition, and members of the Yanai laboratory for constructive comments.

Author contributions

R.M. performed the spatial transcriptomics and scRNA-seq as well as the data analysis. F.W. contributed to scRNA-seq and spatial transcriptomics analysis. M.C. contributed to spatial transcriptomics and scRNA-seq processing, and immunofluorescence experiments. M.B. and D.B. contributed expertise in scRNA-seq processing and analysis. J.C.D. contributed to the spatial transcriptomics analysis. C.H.H. contributed histology analysis.

D.M.S. contributed sample acquisition. I.Y. conceived the project, and contributed to the data analysis and interpretation of the results. R.M. and I.Y. drafted the manuscript.

Competing interests

I.Y. is a shareholder of OneCell Medical Ltd.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41587-019-0392-8.

Correspondence and requests for materials should be addressed to I.Y.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Itai Yanai

Last updated by author(s): Nov 25, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	NextSeq500 was used for sequencing of ST and scRNA-Seq libraries. Leica SCN400 was used for brightfield imaging, and images were processed using Aperio Imagescope 12.3.3
Data analysis	STAR 2.5.1 was used for alignment of scRNA-Seq data with parameter "—outSAMmultNmax 1" and otherwise default parameter settings. Bowtie2 version 2.3.1 was used for alignment of raw ST data. R version 3.6.1 and MATLAB version 2017b were used for data analysis. HTSeq v0.9.1 was used for UMI counting. Analysis pipeline used to process raw Spatial Transcriptomics (ST) data is publicly available on https://github.com/yanailab/celseq2. Analysis pipeline used to process raw single-cell RNA-Seq data is publicly available: https://github.com/flo-compbio/singlecell. Aperio Imagescope 12.3.3 was used for processing of brightfield images. NDP view verison 2.3 was used for processing of fluorescence images. Photoshop CS6 was used for alignment of images. Further details regarding the steps taken by each analysis pipeline are described in the Methods section.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence data that supports the findings of this study have been deposited in the Gene Expression Omnibus (GEO) with the accession code GSE111672. Additional data files will be made available upon reasonable request to the corresponding author. The reviewer password to access this data is: alwzeqcqpdgrdkt

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For single-cell RNA-Seq, 4977 cells were analyzed. This was not pre-determined or calculated; after sequencing of 22,000 single-cell (encapsulated in drops) transcriptomes, 4977 cells were left after computationally filtering cells on the basis of percentage of mitochondrial transcripts (< 20%) ribosomal transcripts (< 30%), and number of transcripts per cell (>= 1000 transcripts). This number was sufficient to identify 15, 11, and 7 cell type clusters for PDAC-A, PDAC-B, and PDAC-C tumors, respectively. For Spatial Transcriptomics, three tissue sections from the PDAC-A patient sample, three tissue sections from the PDAC-B patient sample, and four tissue sections from four additional patient samples was used for the study.
Data exclusions	As described above, although 22,000 cell transcriptomes were sequenced, only 4977 were retained after computational filtering of the datasets; the excluded data was removed on the basis of percentage of mitochondrial transcripts (> 20%) ribosomal transcripts (> 30%), and number of transcripts per cell (<= 1000 transcripts). These criteria were pre-determined prior to data processing.
Replication	Analysis was performed for six tumors in total, where triplicate sections of ST were studied for two tumors with matching scRNA-Seq data.
Randomization	Not relevant to study. The six pancreatic cancer tumor samples were provided from consenting patients.
Blinding	C.H.H., the pathologist who annotated the histology feature across the tissue sections, was blinded from any further details concerning the spatial transcriptomics analysis of the tissue sections.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	anti-AQP3 (Abcam cat. ab85903, lot no. NA) 1:100 anti-CA9 primary antibody (Novus Biologicals, cat. NB100-417 lot no. NA) 1:100 anti-HLA-DR (GeneTex, cat GTX40718 lot no. NA) 1:100 anti-TFF1 (Novus, cat. NBP2-34623-0.1mg, clone GE2 lot no. 1PABX190116) 1:100 anti-KRT19 (ThermoFisher, cat PA5-12431, log UA2703141) 1:100 anti-TM4SF1 (ThermoFisher, cat PA5-51746, lot no. UA2700063) 1:100 anti-S100A4 (Thermofisher, cat PA518601, lot no. UA2699861) 1:100 anti-IL-6 (Thermofisher, cat AHC0762, clone 8H12, lot TI273093) 1:500
Validation	anti-AQP3 (Abcam cat. ab85903) has been validated for IHC at 1/300 - 1/2000 dilution, as stated on the manufacturer's website. anti-CA9 (Novus Biologicals, cat. NB100-417) has been validated for IHC at 1:1000 dilution, as stated on the manufacturer's website: https://www.novusbio.com/products/carbonic-anhydrase-ix-ca9-antibody_nb100-417. anti-HLA-DR (GeneTex, cat GTX40718) has been validated for IHC at 1:100 - 1:1000 dilution as stated on the manufacturer's website: https://www.genetex.com/Product/Detail/HLA-DR-antibody-N2C3/GTX113459. anti-TFF1 (Novus, cat. NBP2-34623) has been validated for IHC at 0.5-1.0 µg/ml, as stated on the manufacturer's website: https://www.novusbio.com/products/tff1-ps2-antibody-ge2-same-as-r47-94-_nbp2-34623. anti-KRT19 (ThermoFisher, cat PA5-12431) has been validated for IF/IHC at 1:100-1:1000 dilution, as stated on the manufacturer's website: https://www.thermofisher.com/antibody/product/Cytokeratin-19-Antibody-Polyclonal/PA5-29548. anti-TM4SF1 (ThermoFisher, cat PA5-51746) has been validated for IHC at 1:500-1:1000 dilution, as stated on the

manufacturer's website: https://www.thermofisher.com/antibody/product/TM4SF1-Antibody-Polyclonal/PA5-51746. anti-S100A4 (Thermofisher, cat PA518601) has been validated for IHC at 1-2 µg/ml, as stated on the manufacturer's website: https://www.thermofisher.com/antibody/product/S100A4-Antibody-Polyclonal/PA5-18601. anti-IL-6 (Thermofisher, cat AHC0762) has been validated for IHC at 1:10-1:100 dilution, as stated on the manufacturer's website: https://www.thermofisher.com/antibody/product/IL-6-Antibody-clone-8H12-Monoclonal/AHC0762

Human research participants

Policy information about studies involving human research participants

Population characteristics	Only tumor samples from patients with no treatment history were collected.
Recruitment	All samples were collected from patients with no reported treatment history. The tissue used was not needed for clinical or diagnostic purposes.
Ethics oversight	The study was approved by the NYU School of Medicine Institutional Review Board and the Ethics Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.