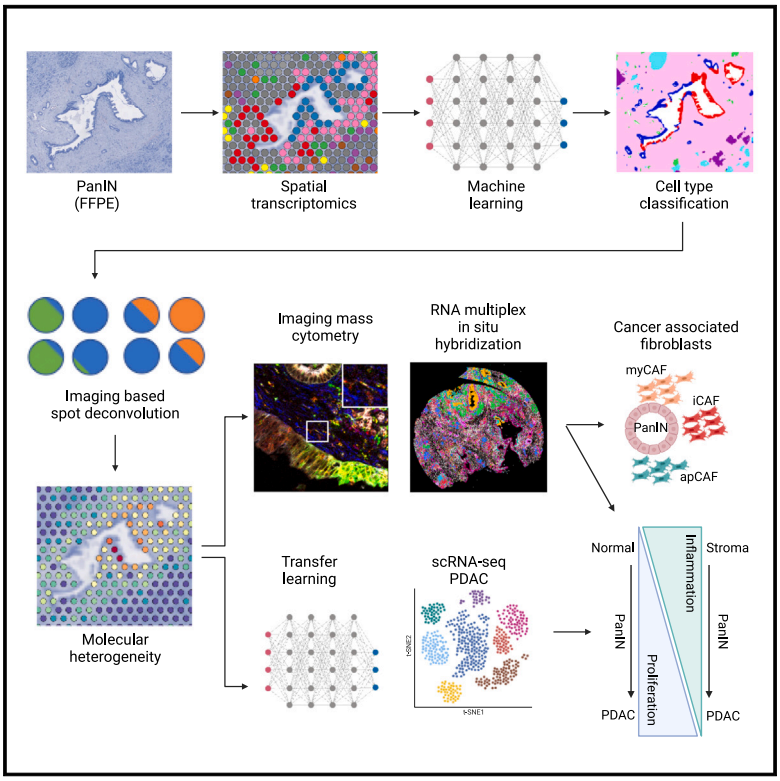
Article

Cell Systems

PanIN and CAF transitions in pancreatic carcinogenesis revealed with spatial data integration

Graphical abstract



Authors

Alexander T.F. Bell, Jacob T. Mitchell, Ashley L. Kiemen, ..., Laura D. Wood, Elana J. Fertig, Luciane T. Kagohara

Correspondence

ejfertig@jhmi.edu (E.J.F.), ltsukam1@jhmi.edu (L.T.K.)

In brief

A new semi-supervised imaging, spatial transcriptomics, and single-cell workflow uncovers transitions of PanIN progression and the surrounding microenvironment. Leveraging this pipeline for panel design for single-cell proteomics and imaging transcriptomics enables the discovery and confirmation of rare cell types and a transition between proliferative and CAF-associated inflammatory signaling in PDAC carcinogenesis.

Highlights

- New workflow for integrated analyses of imaging, spatial, and single-cell datasets
- PanIN premalignant environment presents molecular and cellular features similar to PDAC
- PanIN-to-PDAC progression shows transition from inflammatory to proliferation signaling

Bell et al., 2024, Cell Systems *15*, 753–769
August 21, 2024
https://doi.org/10.1016/j.cels.2024.07.001

CellPress

Cell Systems

CellPress

Article

PanIN and CAF transitions in pancreatic carcinogenesis revealed with spatial data integration

Alexander T.F. Bell,[1,2,13] Jacob T. Mitchell,[1,2,3,4,13] Ashley L. Kiemen,[5,6,13] Melissa Lyman,[1,2] Kohei Fujikura,[6] Jae W. Lee,[1,2,3,6] Erin Coyne,[1,2,3] Sarah M. Shin,[1,2,3] Sushma Nagaraj,[1,2] Atul Deshpande,[1,2,3] Pei-Hsun Wu,[5] Dimitrios N. Sidiropoulos,[1,2,3,7] Rossin Erbe,[1,2,3,4] Jacob Stern,[8] Rena Chan,[8] Stephen Williams,[8] James M. Chell,[8] Lauren Ciotti,[1,2] Jacquelyn W. Zimmerman,[1,2,3,12] Denis Wirtz,[5,9,10] Won Jin Ho,[1,2,3,12] Neeha Zaidi,[1,2,3,12] Elizabeth Thompson,[1,6,12] Elizabeth M. Jaffee,[1,2,3,12] Laura D. Wood,[1,6,12] Elana J. Fertig,[1,2,3,11,12,*] and Luciane T. Kagohara[1,2,3,12,14,*]

[1]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[2]Convergence Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[3]Bloomberg Kimmel Immunology Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[4]Department of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA
[5]Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, Baltimore, MD, USA
[6]Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD, USA
[7]Cellular and Molecular Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[8]10x Genomics, Pleasanton, CA, USA
[9]Department of Materials Science and Engineering, The Johns Hopkins University, Baltimore, MD, USA
[10]Johns Hopkins Physical Sciences - Oncology Center, The Johns Hopkins University, Baltimore, MD, USA
[11]Department of Applied Mathematics and Statistics, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA
[12]The Skip Viragh Center for Clinical and Translational Research, Baltimore, MD, USA
[13]These authors contributed equally
[14]Lead contact
*Correspondence: ejfertig@jhmi.edu (E.J.F.), ltsukam1@jhmi.edu (L.T.K.)
https://doi.org/10.1016/j.cels.2024.07.001

SUMMARY

This study introduces a new imaging, spatial transcriptomics (ST), and single-cell RNA-sequencing integration pipeline to characterize neoplastic cell state transitions during tumorigenesis. We applied a semi-supervised analysis pipeline to examine premalignant pancreatic intraepithelial neoplasias (PanINs) that can develop into pancreatic ductal adenocarcinoma (PDAC). Their strict diagnosis on formalin-fixed and paraffin-embedded (FFPE) samples limited the single-cell characterization of human PanINs within their microenvironment. We leverage whole transcriptome FFPE ST to enable the study of a rare cohort of matched low-grade (LG) and high-grade (HG) PanIN lesions to track progression and map cellular phenotypes relative to single-cell PDAC datasets. We demonstrate that cancer-associated fibroblasts (CAFs), including antigen-presenting CAFs, are located close to PanINs. We further observed a transition from CAF-related inflammatory signaling to cellular proliferation during PanIN progression. We validate these findings with single-cell high-dimensional imaging proteomics and transcriptomics technologies. Altogether, our semi-supervised learning framework for spatial multi-omics has broad applicability across cancer types to decipher the spatiotemporal dynamics of carcinogenesis.

INTRODUCTION

Single-cell RNA sequencing (scRNA-seq) and spatial molecular technologies have enabled unprecedented characterization of the molecular and cellular pathways that comprise the tumor microenvironment (TME).[1,2] This has had a particularly profound impact on the understanding of the complex immunosuppressive pancreatic ductal adenocarcinoma (PDAC) TME and its role in cancer progression and therapeutic resistance.[3–7] Further characterization of premalignancies is critical to delineate the evolutionary mechanisms of malignant transformations, the impact of the complex microenvironment on facilitating carcinogenesis, and the development of nearly universal therapeutic resistance in PDAC. Among the multiple histologically distinct premalignant lesions of the pancreas that potentially progress to PDAC, pancreatic intraepithelial neoplasia (PanIN) is the most frequent and well studied.[8,9] PanINs therefore provide the opportunity to characterize some of these evolutionary

processes in PDAC. Because the diagnosis of PanINs is restricted to formalin-fixed and paraffin-embedded (FFPE) tissue, the studies on the comprehensive characterization of molecular and cellular phenotypes of these atypical cells, and on the other cell types in the surrounding microenvironment, are rare and limited to bulk, target-based, and cell type-focused sequencing.

Whole transcriptome spatial transcriptomics (ST) combines gene expression profiling with spatial information to map the molecular and cellular landscape of cancer.[2] By analyzing gene expression within the tissue spatial context, it is possible to examine tumor heterogeneity and intercellular interactions in the microenvironment. The development of an ST technology that can be performed on FFPE samples opens the opportunity to profile PanIN lesions and the adjacent microenvironment. Computational methods to discover the cellular and molecular changes in the evolution of the TME are an essential component of ST analysis and are being actively developed alongside the experimental approaches.[2] One drawback in ST data is the lack of single-cell resolution.

Even after transcriptional profiles are isolated, relating computationally estimated temporal changes in cellular phenotypes to molecular markers of those phenotypes and their impact on the dynamics of tumor progression remains a challenge. Performing this analysis requires identifying spots associated with cell groups of interest (e.g., from normal, low-grade [LG], and high-grade [HG] PanIN) from which to infer cell state transitions. In principle, clustering or other unsupervised learning analysis of gene expression profiles from the ST data can identify these spots for this analysis. However, the lack of single-cell resolution of the ST technologies can limit the accuracy of these analyses. One solution to this problem is the use of spot deconvolution algorithms to estimate the cell types that are represented in each spatial spot for robust data analysis and interpretation. Many of these deconvolution methods estimate cell type proportions per spot using a scRNA-seq reference that was generated from the same tissue or disease type for this deconvolution.[10,11] Although these methods provide robust estimates, they are not suited for populations of cells that are rare and are not applicable for diseases that cannot be profiled with dissociation-based single-cell technologies (e.g., PanIN cells). Recently, a few methods have been developed to perform spot deconvolution and cell type classification based on imaging analysis. In FFPE ST, the possibility of staining and imaging the sections before library preparations allows pathological examination of cell morphologies to automate single-cell resolved cell type classification and deconvolution of ST data. Machine learning methods have been developed to integrate cell morphology and perform cell classification that will be integrated within ST coordinates to provide the proportion of each specific cell type in a spatial spot or enhance the accuracy of clustering.[12,13] However, gene expression signatures are still required to identify cellular features of interest before the analysis. While these unsupervised learning methods are powerful, leveraging pathology expertise for cellular labeling can enhance spot and region selection for more refined analyses of gene expression changes within specific cellular phenotypes. The laborious, manual annotation for these workflows requires an automated and accurate manner to increase robustness of biological findings from integrated imaging and ST data.

In this study, we demonstrate that three-fold integration with ST, new imaging analysis technologies, and single-cell PDAC atlases provides the opportunity to analyze the dynamic cellular transitions that are associated with different stages of PDAC progression. We present the ST analysis of 14 PanIN lesions from 9 patients, including 5 rarely diagnosed HG PanIN lesions. To analyze the data and overcome the computational limitations to ST analysis, we built a method for three-way integration of imaging, ST, and scRNA-seq data to automate cellular labeling of spots for subsequent supervised and unsupervised learning of transcriptional changes during disease progression in specific cell groups of interest. First, we adapt the recently developed machine learning method, CODA,[14] that automatically identifies and annotates cells in the pancreatic microenvironment to provide cell annotations for ST spots. The resulting gene expression analysis of the epithelial cells (normal and PanIN) and the surrounding cells within the microenvironment revealed that PanINs already express specific transcriptional signatures (e.g., classical subtype) of invasive carcinoma and the presence of fibroblasts resembling cancer-associated fibroblast (CAF) subtypes found within the PDAC TME. Second, multi-omics integration with a reference scRNA-seq atlas of PDAC,[15] using the transfer learning method ProjectR,[16,17] found that a CAF-associated inflammatory and EMT (epithelial mesenchymal transition) pattern gradually decreases during PDAC invasion and is associated with a compensatory increase in proliferation pathways in PDAC carcinogenesis. These approaches further enabled the design of custom panels for proteomics and transcriptomics spatial technologies for confirmation of these findings with single-cell resolution and creation of a spatial multi-omics reference of PanINs. In summary, our experimental and computational pipeline for imaging analysis and multi-omics integration is broadly applicable to analysis of cancer progression in different tumor types.

RESULTS

ST applied to FFPE specimens captures preneoplastic pancreatic tissue architecture

To identify the cellular and molecular features of PanIN that are still present in PDAC, we applied a whole transcriptome FFPE ST protocol to profile a cohort of human PanIN samples and developed a semi-supervised computational pipeline to perform spot deconvolution and integration with a scRNA-seq PDAC reference dataset (Figure 1A). The first step of this pipeline is to assign broad cell type labels to spots at a single-cell resolution from imaging using CODA, a machine learning approach initially trained to identify and classify pancreatic cell types based on their morphologies. The second part of our pipeline leverages transfer learning, using ProjectR, to relate transcriptional patterns from a scRNA-seq PDAC reference to quantify how transcriptional changes in our PanIN dataset relate to changes in advanced-stage cancer (Figure 1A).

In this study, we apply ST profiling to a test cohort of paired LG and HG PanINs diagnosed in the same patients (4 patients' specimens, total number of PanIN lesions = 8) and to a validation cohort of 7 PanINs (5 patients, 6 LG, and 1 HG lesions). In total, we performed ST profiling of 15 PanIN lesions (10 LG and 5 HG). Of note, HG PanIN lesions are rarely diagnosed due to their

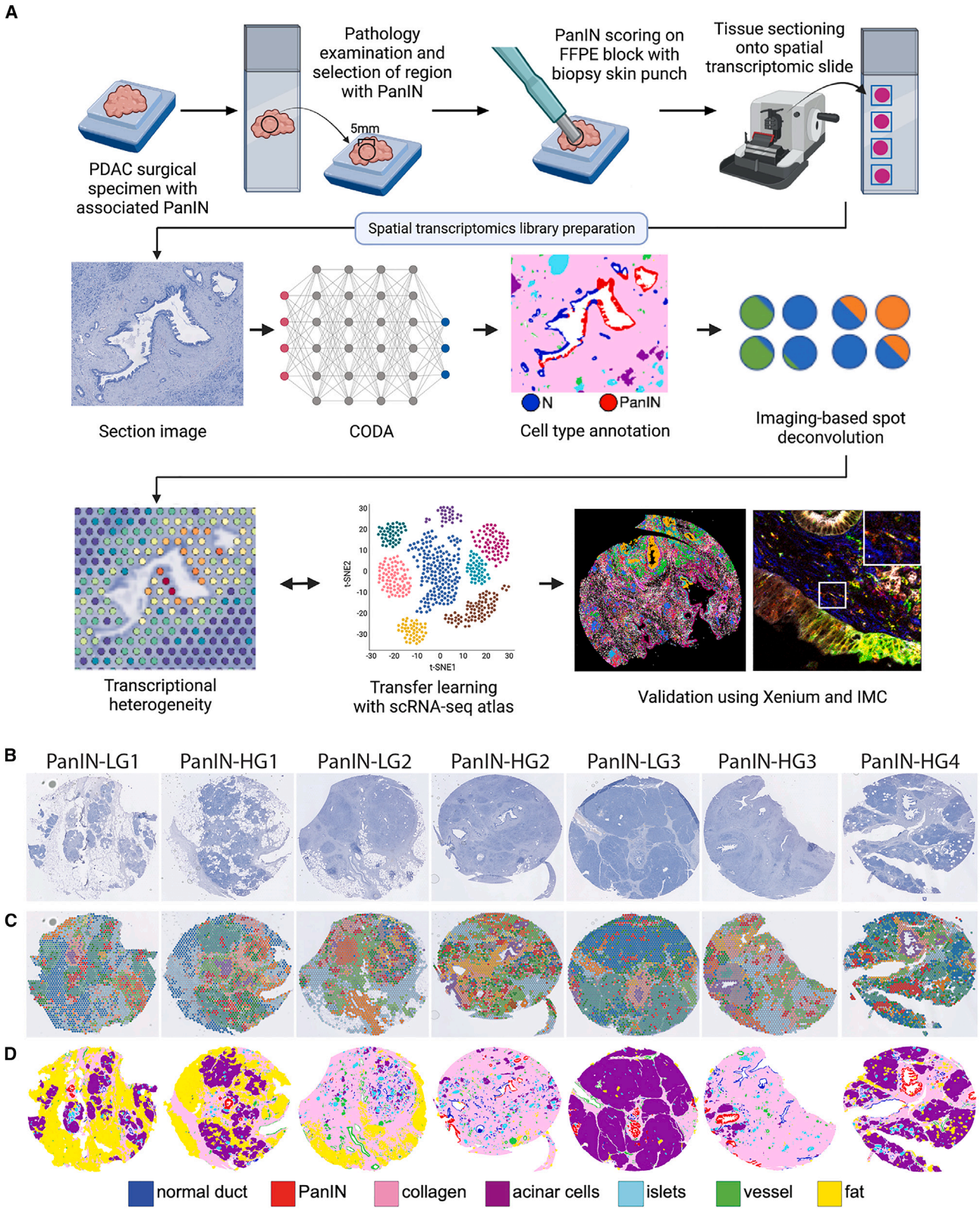


Figure 1. ST analysis of FFPE PanIN

(A) Pancreatic cancer surgical specimens in FFPE were examined, and the regions containing FFPE pancreatic intraepithelial neoplasia (PanIN) lesions were identified for scoring using a 5 mm skin biopsy punch and sectioning onto the spatial transcriptomics (ST) slide. The stained images were used for machine

(legend continued on next page)

challenging differential diagnosis of colonization of the duct by the invasive carcinoma from the HG atypia of these premalignant lesions, making the characterization of even this small number of human HG lesions a valuable reference resource for future studies. Initial total RNA quality check indicated that all samples presented high levels of RNA degradation (RIN ~2) but with a high concentration of 200 bp fragments (DV200 $\geq$ 50%) compatible with the FFPE ST platform. Following ST data generation, preprocessing, and quality check, 14 out of the 15 samples (7 from the paired cohort and 7 from the validation cohort; 9 LG and 5 HG) had sufficiently high-quality data for subsequent analysis.

The ST data from our PanIN cohort provides combined stained imaging and transcriptomics profiling from the same section (Figures 1B, 1C, S1A, and S1B). PanINs are thought to be an intermediate state between the healthy pancreas tissue and PDAC and require pathology examination as the diagnosis is based on morphological identification of the atypical epithelial cells. Therefore, it can be anticipated that the tissues will contain a combination of cell types and histological structures from both states (e.g., ductal cells, acinar cells, neoplastic cells, fibroblasts, acinar cells, islets, immune cells, etc.).[18] We characterized the cellular distribution of PanINs and surrounding pancreas tissue by first applying unsupervised clustering to the ST profiling data. The clustering generated a total of 15 spatially resolved gene expression clusters (Figures 1C and S2–S9). Similar to single-cell analysis, we annotated the clusters through differential expression analysis to identify marker genes associated with each cluster (Figure S2). Using this strategy, for example, we annotated one specific cluster to PanIN based on the expression of *TFF1*, *MUC5AC*, *TFF2*, *MUC6*, and *CTSE*; while another specific cluster, expressing *LCN2, GPX2, TCN1, KRT8*, and *ANXA4*, was annotated to normal ducts. Based on other cell type-specific markers, we were able to annotate 4 fibroblast clusters, 2 immune cell clusters, 4 acini clusters, 2 pancreatic islet clusters, 1 smooth muscle cluster, and 1 neural cell cluster (Figures S2–S9). The spatial distribution of the gene expression clusters recapitulated the overall histological architecture of the samples. The distribution of the normal and neoplastic spots identified from clustering matches the initial pathology identification of these cell types and was further confirmed by pathology examination. Almost all the gene expression spatial clusters extend beyond the histological boundaries observed in the stained sections into the adjacent cells (i.e., the same spatially resolved gene expression cluster mapped to regions of distinct cell types) (Figure S10A). This extended signal, or bleeding, could be a result of technical artifacts in the ST technology leading to the detection of marker genes of one cell type in the space (spot) of the adjacent distinct cell type.[19] This observation combined with the lack of single-cell resolution of the ST data led us to hypothesize that incorporating the cellular labels obtained from imaging analysis into our gene expression analysis could enhance the robustness of the phenotypic characterization of those cells.

The first step of our new analysis pipeline aims to annotate broad cell types of spots in the ST data, leveraging the matched stained imaging for the FFPE-based ST platform. Several methods overcome these observed limitations of the ST data through spot deconvolution with reference scRNA-seq data.[10,11] However, no scRNA-seq reference is available for human PanINs due to the diagnosis in FFPE clinical blocks. Recent computational work with joint analysis of histology features and transcription can serve as powerful alternatives.[12,13] We sought to enhance these approaches by incorporating expert pathological knowledge of cellular labels to delineate relevant cellular features of pancreatic precancer from the ST sections imaging. To focus our transcriptional analysis on the transition from normal through PanIN progression and the structures of the microenvironment, we applied the machine learning method CODA to the stained images of the ST sections to automatically classify the pancreatic cells. CODA is an imaging-based analysis approach that uses deep learning semantic segmentation to identify different cell types that have been trained specifically for the human pancreas, precancer, and cancer (acinar cells, islets of Langerhans, fibroblasts, adipocytes, endothelial cells, ductal cells, and neoplastic cells).[14] In this study, we integrated CODA to the ST analysis to obtain automated cell type annotation combined with ST spots deconvolution (Figures 1D and S1C). In contrast to the clustering analysis, imaging cell type annotations using CODA are at single-cell resolution and, through integration with the ST spots coordinates and dimensions, enable a true estimate of the true proportion of cell types within each ST spot for robust gene expression analysis (Figure S10B). To avoid unwanted bias in the comparisons between normal duct and PanIN clusters, we selected spots that were quantified as representing at least 70% of a unique cell type. Using this threshold for all the cell types, we were able to increase cell type purity from ~25% to ~90% and from ~45% to ~95% for normal duct and PanIN clusters, respectively, for example (Figure S10C). This enables supervised analysis, defining marker genes through differential expression analysis of spots annotated with known cellular features in pancreatic precancer. We also observe elevated expression of cluster-specific marker genes, such as *CTSE*, *INS*, and *PRSS1* in PanIN and islet clusters, respectively (Figure S10D). This observation highlights the importance of spot annotation with machine learning (CODA) prior to differential expression analysis or more advanced multi-omics integration to single-cell reference datasets, as demonstrated through our subsequent analyses of the PanIN microenvironment and progression.

PanIN-associated fibroblasts are a heterogeneous population composed of the same subtypes detected in invasive PDAC

The integration of imaging analysis and ST data provided the unique opportunity to examine the fibroblast population adjacent

learning analysis for cell type identification and spatial spots deconvolution. The ST analysis was integrated with an invasive cancer single-cell dataset. The findings were validated with single-cell resolved transcriptomics and proteomics.

(B) Discovery cohort stained sections were used for pathology examination and identification of PanINs and other pancreatic histological regions.

(C) The unsupervised clustering of the spatial transcriptomics data identified gene expression clusters whose location resembles the distribution observed in the stained sections.

(D) Cell types indicated in the legend were defined automatically from cellular morphologies of the stained sections using the machine learning approach CODA, thereby refining cellular annotations obtained from clustering alone.

to PanIN. While CODA broadly annotates stromal cells, the PDAC TME is enriched with a heterogeneous population of CAFs. They have been classified into three subtypes based on transcriptional profiles, myofibroblastic CAFs (myCAFs), inflammatory CAFs (iCAFs), and antigen-presenting CAFs (apCAFs), and can play dual roles by inhibiting or inducing PDAC progression.[20–23] CAFs exert a tumorigenic role by providing metabolites for tumor cell survival, stimulating cell growth pathways through paracrine signaling, and creating an immunosuppressive microenvironment.[24] However, a tumor-suppressing CAF-enriched TME can reduce essential nutrients required for tumor progression and differentiation, while the same CAFs can be functionally repolarized to release chemokines that will recruit immune cells into the tumor.[24]

MyCAFs and iCAFs have previously been observed in pancreatic premalignant lesions in murine models that recapitulate PDAC development, suggesting that they arise early during tumorigenesis.[25,26] Nevertheless, their presence was not previously described in human premalignant lesions of the pancreas. Here, we leveraged our computational analysis approach to isolate stromal cells in the ST data and further classify these cells from the ST data using established gene markers[22] to map the distribution of myCAFs and iCAFs in the human PanIN microenvironment. In our cohort, the density of stromal cells inferred from CODA varied but were observed adjacent to each premalignant lesion (Figure 1D, pink annotated regions). The further integration of CODA annotations with the ST transcriptional profiles (Figure 2A) showed that a CAF common signature (panCAF) is consistently expressed across the collagen-rich regions annotated by CODA (Figures 2B and S11A, orange and red spots). The expression of myCAF (Figures 2C and S11B) and iCAF (Figures 2D and S11C) markers was detected in all samples overlapping with the regions where panCAFs are present. The presence of a recently described subtype of apCAFs was also investigated using the transcriptional data in our cohort. The apCAFs were first identified by scRNA-seq in a PDAC mouse model and were shown to express major histocompatibility complex (MHC)-II genes and present antigens to $CD4^+$ T *in vitro*, activating their suppressive capability.[22] In our study, expression of the apCAF signature was detected in all samples in the Visium ST data (Figures 2E and S11D).

Since CODA does not have resolution to annotate immune cells because of their limited size and scant cytoplasm, and ST does not provide single-cell resolution, the discrimination of apCAFs from CD45+ immune cells (Figures 2F and S11E) was performed using other spatial approaches with single-cell resolution. The validation of the ST findings was performed using imaging ST *in situ* RNA hybridization (Xenium, 10× Genomics), with a panel to detect 380 transcripts that include epithelial, immune, and CAF markers (Table S1). In contrast to Visium, Xenium provides single-cell resolution using cell segmentation determined by nucleus staining (as described in STAR Methods) and was performed on 3 PanIN lesions from the paired cohort (PanIN-HG1, PanIN-HG2, and PanIN-HG3) that had the premalignant lesions still present on the FFPE blocks. With this technology, transcripts are detected using probes that, after binding to their targets, are conjugated with fluorophores that then are detected, counted, and mapped to each cell identified by a series of multiple scans. This targeted ST approach also recapitulates the sample's architecture, similar to what was observed with the transcriptome wide ST clustering (Figure 3A). Epithelial cells (normal and PanIN), myCAFs, iCAFs, and apCAFs were all detected among the cells spatially profiled with single-cell resolution (Figures 3A and 3B). With Xenium, it was possible to confirm apCAFs in proximity to PanIN lesions (Figures 3A and 3B). The apCAFs were annotated based on module scores of apCAF marker genes and the absence of expression of CD45 (*PTPRC*) (Figure 3C). The module score strategy was used due to the broad expression of MHC II gene markers (Figures 3D and 3E). Among all the cells in the tissue segments profiled (Figure 3F), apCAFs were found to comprise up to 13.9% of the cells detected (Figure 3G). We also verified that the apCAFs co-localize with CD4+ T cells in the regions just adjacent to the PanINs (Figure S12), strong evidence for the interaction between these cell types.

We further sought to confirm the presence of apCAFs using a single-cell resolved proteomics approach. To do so, we performed additional multiplex proteomics analysis of the PanIN samples from the paired cohort with imaging mass cytometry (IMC) using a customized antibody panel (Table S2) that was specifically developed to identify the different CAF subtypes (myCAF, iCAF, and apCAF) in PDAC. The proteomics analysis with IMC corroborates the *in situ* gene expression data, showing cellular co-localization of CAF marker proteins with MHC II proteins. The presence of apCAFs was confirmed in all 5 samples (PanIN-LG1, PanIN-HG1, PanIN-LG2, PanIN-HG2, and PanIN-HG3) profiled by the concomitant expression of panCAF markers (alpha-smooth muscle actin [SMA] and vimentin [VIM]) and MHC II proteins (CD74 and HLA-DR) (Figures 3F and 3G). The presence of apCAFs is not restricted to the PanIN neighbor regions (PanIN region [a], Figures 3F and 3G), but they are also found in regions of fibrosis further from the premalignant lesions (fibrosis region [b], Figures 3F and 3G). For a more accurate quantification of the CAF in the IMC regions of interest, CAFs were identified through the expression of COL, SMA, VIM, and PDPN (Figure 3J). Due to the broad presence of collagen that is detected by the expression of COL, the CAF detection was refined by the co-expression of the panCAF markers and DNA (Figures 3K and 3L). Subsequently, the apCAFs were identified as cells expressing panCAF markers + DNA + CD74 (Figure 3M), panCAF markers + DNA + HLADR (Figure 3N), and panCAF markers + DNA + CD74 + HLADR (Figure 3O). The apCAFs expressing only CD74 represent 0.28% to 6.33% of all CAFs detected, apCAFs expressing only HLADR are 0.50% to 5.56% of the CAFs, and apCAFs expressing both MHC II markers are rare, comprising 0.08% to 1.39% of the CAFs identified (Figure 3P). To determine apCAFs presence the areas with aggregates of immune cells were excluded from the quantifications (Figure 3P) The IMC data confirm the presence of apCAF in areas associated with PanIN and the low frequency of these modulatory cells. The markers used for CAF classification are provided in Table S3.

Altogether, these results demonstrate that the different subtypes of CAFs that are present in PDAC can be detected surrounding PanIN lesions, including the new class of apCAFs, suggesting that the TME modulation by these cells occurs early in pancreas tumorigenesis.

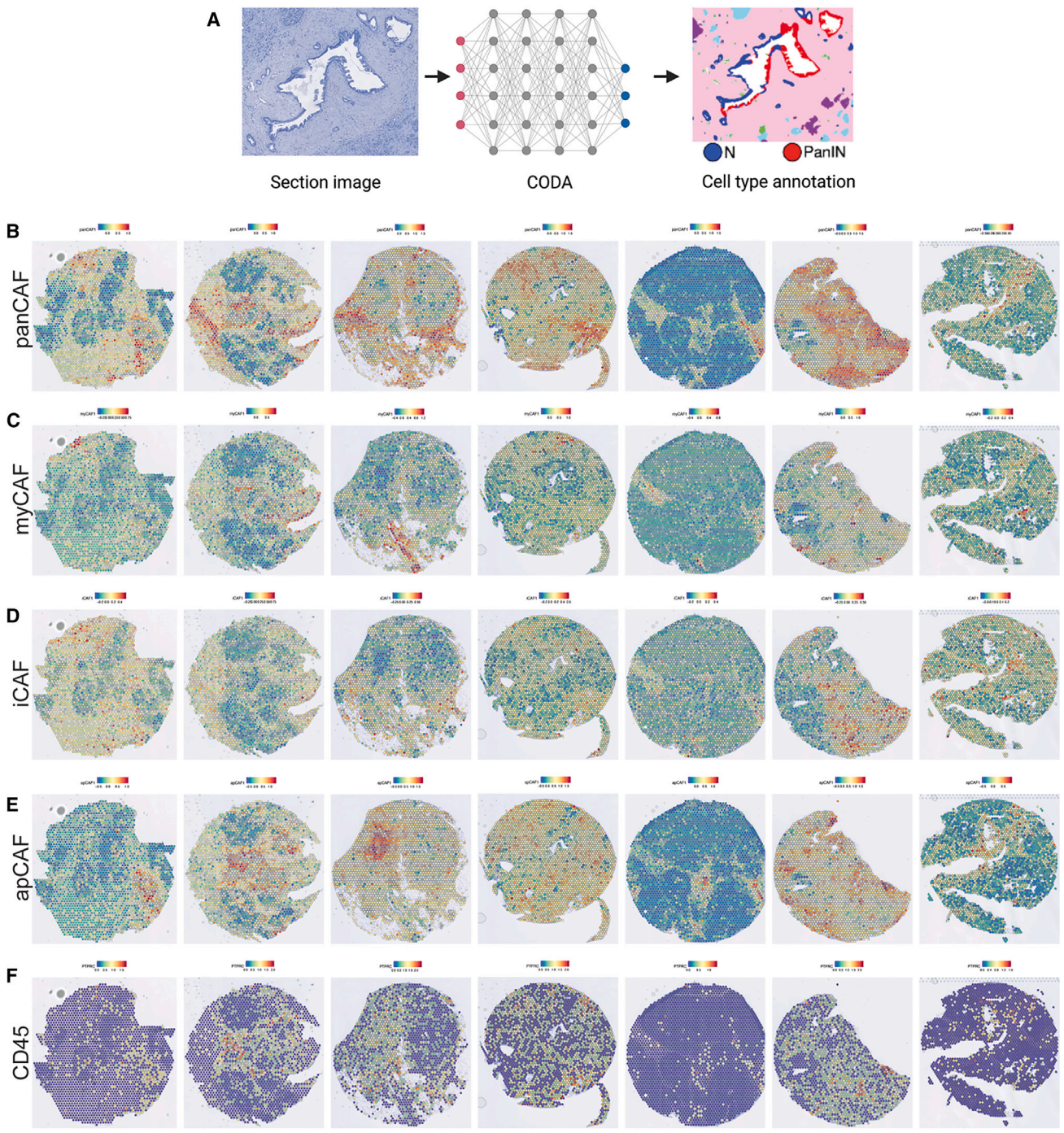


Figure 2. Spatial distribution of PDAC CAF subtypes in the discovery cohort

(A–D) (A) Cancer-associated fibroblast (CAF) localization was mapped using panCAF markers, (B) myofibroblastic-CAF markers, (C) inflammatory-CAF markers, and (D) antigen presenting-CAF markers.

(E) CD45 expression was examined to identify regions where CAFs and immune cells were co-localized.

ST identifies expression of both PDAC classical subtype and CSC signatures in PanINs

PanIN lesions can develop into invasive PDAC. To identify PDAC features that can be detected during the premalignant stage, we leveraged the automated cell type annotation from CODA with cluster-based annotations to classify spots to all pancreatic cell types using a cut-off of 70%. For example, a spot was classified epithelial (normal or PanIN) if CODA quantified that in that coordinate >70% of the cells were epithelial (Figure 4A). Next, we characterized PanIN cell heterogeneity relative to the established classical and basal-like PDAC subtypes.[27] We found that 13 (6 in the test cohort and 7 in the validation cohort) out

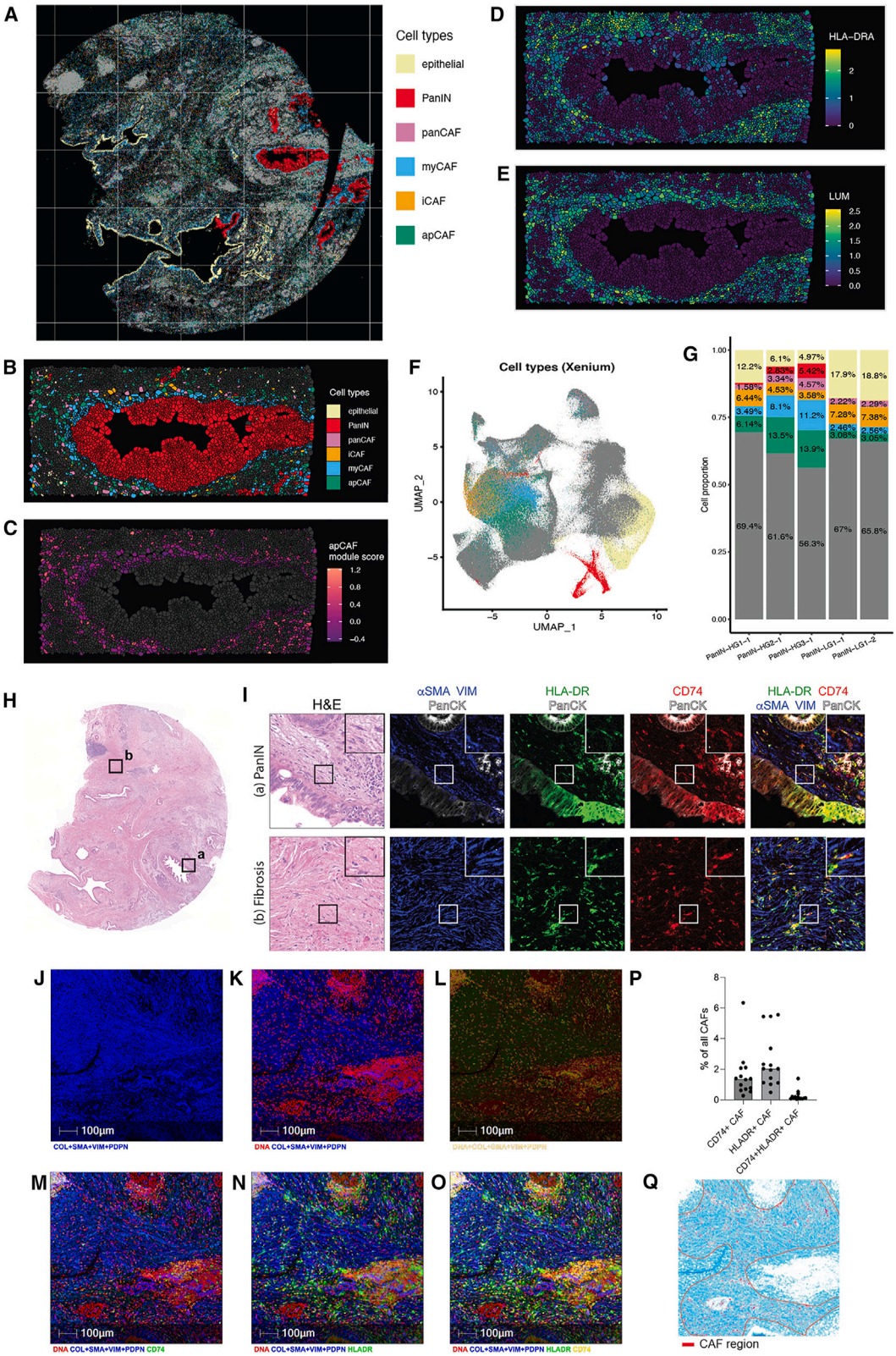


Figure 3. Confirmation of apCAF population using single-cell transcriptomics and proteomics

(A) Xenium clusters spatial distribution from a panel to detect 380 genes recapitulate the sample (PanIN-HG3) architecture. Cells identified as epithelial cells (yellow), PanIN (red), and CAF subtypes apCAF (green), iCAF (orange), and myCAF (blue) are highlighted.

(legend continued on next page)

of 14 PanINs express the PDAC classical subtype signature (Figures 4B and S13A). The basal-like signature is not expressed in any of the premalignant lesions (Figures 4C and S13B). This observation supports the hypothesis that PDACs arise with a classical phenotype and acquire the basal-like phenotype upon progression and accumulation of molecular aberrations.[28] Further studies to determine the classical subtype markers that are critical for the transformation of PanIN into PDAC are necessary to drive the development of early therapeutic interventions and early detection tests to improve patients' survival.

Only one HG PanIN sample from the paired cohort (PanIN-HG3) expressed neither the classical nor the basal-like signatures (Figures 4A, 4B, S12A, and S12B). Thus, we hypothesized that this sample expresses a third transcriptional phenotype. PDAC progression, resistance to therapies, and immune evasion are partially associated with the presence of PDAC cells expressing cancer stem cell (CSC) markers.[29] We verified the expression of CSC markers among the PanINs in the paired and validation cohorts. The only sample with a significantly high expression of CSC markers is the one that did not express the classical or the basal-like PDAC signatures (Figure 4D). The presence of cells with stemness features suggests that some mechanisms of resistance to therapies arise early in PDAC progression.

Differential expression analysis between PanINs and normal ducts identifies gradual increase of *TFF1* expression during PanIN progression limited to the classical phenotype

To further define the molecular features of PanINs, we merged spots from all samples CODA annotated as normal and PanIN ducts for each patient. Differential expression was performed to identify gene expression changes across each patient's premalignant lesions. A total of 118 genes are differentially expressed in PanINs relative to normal ducts in the paired cohort (Figure 4E), and their expression pattern discriminated PanINs from normal ducts among the different samples (Figure 4F). Among the top 20 up-regulated genes in the premalignant lesions of the paired cohort, only 5 genes (*TM4SF1*, *CYP2S1*, *CD55*, *FER1L6*, and *PSCA*) had no known role in pancreas tumorigenesis, suggesting that FFPE ST analysis is robust and corroborates previous gene expression analyses in PanINs.[30–32] The differential expression analysis of the validation cohort also identified sets of genes that discriminate normal ducts from PanIN lesions with common genes aberrantly expressed between the two PanIN cohorts (Figures S13C and S12D). Pathway analysis from the differentially expressed genes in the paired cohort indicates enrichment for MYC and oxidative phosphorylation pathway mediators. Both signaling pathways have been previously shown to be up-regulated in PanINs and PDAC, particularly in association with progression from premalignancy to invasive cancer, metastasis development, and resistance to therapy[33–35] (Figure S14).

Although predominantly consisting of the classical subtype, the differential expression analysis highlighted the inter-sample heterogeneity with only one differentially expressed gene showing up-regulation in all classical samples (*TFF1*). *TFF1* is known to be overexpressed in PanINs and PDACs, and its protein levels have been suggested as a potential early detection marker found in bodily fluids. In *in vitro* cell culture models, the secreted form of TFF1 was shown to increase PDAC and stellate cell motility without a significant impact on cancer cells proliferation.[36] Since stellate cells are considered one of the precursors to some PDAC CAF subtypes,[24,37] it is possible that *TFF1* is one of the mediators of intercellular interactions among PanIN and PDAC cells and CAFs. One interesting observation is that the sample expressing the CSC markers signature does not express high levels of *TFF1*. In contrast to all the other PanIN lesions, this sample lacks expression of classical subtype signature, leading us to hypothesize that the stemness phenotype is independent of *TFF1* expression (Figure S15).

The characterization of multiple ducts, including those across stages of PanIN differentiation (mixed ducts), allows us to trace the cellular changes associated with PanIN progression. Additionally, ST analysis provides the ability to visualize the preneoplastic differentiation stages and concomitantly map the respective gene expression level changes (Figure 5A). We therefore compared expression changes between lesions classified as LG or HG based on their morphology. Since CODA cannot discriminate between LG and HG PanIN, the differential diagnosis was performed by pathology experts (KF, JWL, ET, and LWD) (Figures 5B–5D). Using the pathological PanIN classification, we identified one mixed duct (PanIN-HG2) containing normal ductal cells as well as LG and HG PanIN cells (Figure 5E).

We expanded our differential expression analysis study to uncover additional gene expression changes across PanIN stages. This analysis identified five other genes (*MUCL3*, *C19orf33*, *TSPAN1*, *SCD*, and *ACTB*) that were up-regulated in HG lesions relative to LG lesions (Figure S16). In addition, the level of expression of *MUCL3* and *TSPAN1* gradually increased from

(B) The HG PanIN is surrounded by a heterogeneous population of cells, including apCAFs. The apCAFs were identified based a panCAF module score, absence of CD45 (*PTPRC*) expression, and elevated module scores for marker genes of apCAFs.
(C) The apCAFs were annotated as cells with high apCAF signature module score.
(D and E) (D) Expression of the MHC II gene HLA-DRA and (E) of the CAF marker LUM co-localize with regions of apCAF high module scores.
(F) UMAP representing epithelial, PanIN, panCAF, myCAF, iCAF, and apCAF across the three samples analyzed with Xenium.
(G) Percent composition of cell types in each sample that was profiled with Xenium.
(H) Representative image of the pancreas with pancreatic intraepithelial neoplasia (PanIN) and fibrosis. Regions with PanIN (a) and fibrosis (b) are highlighted.
(I) Representative images of the pancreas with PanIN (top row) and fibrosis (bottom row). H&E and image mass cytometry images of SMA and vimentin (VIM, blue), HLA-DR (green), CD74 (red), and pancytokeratin (PanCK, white) are shown.
(J) CAF detection in the IMC regions of interest was quantified by the expression of panCAF markers (COL, SMA, VIM, and PDPN).
(K and L) DNA presence was used to exclude areas of collagen only from CAFs.
(M–O) (M) CD74 expression, (N) HLADR expression, and (O) CD74 and HLA-DR concomitant expression identified the apCAFs.
(P) Proportion of apCAFs among the CAFs detected within the multiple regions of interested profiled.
(Q) Area enriched for CAFs (marked in red) used to measure the frequency of apCAFs, excluding immune-rich regions.

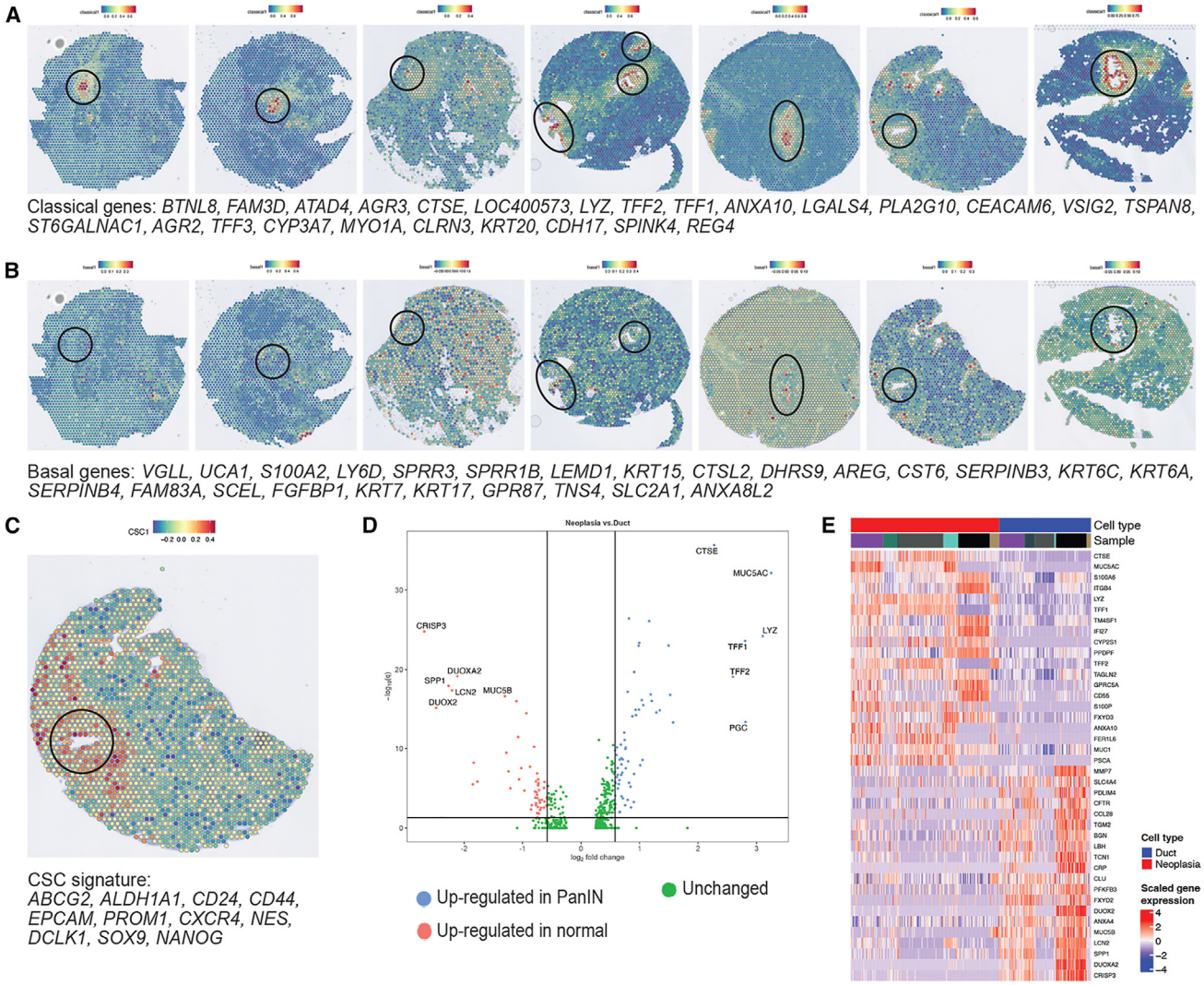


Figure 4. PanINs transcriptional features

(A and B) (A) Six out of seven PanINs (black circles) expressed markers that characterize the classical subtype of pancreatic cancer, while (B) the basal-like signature was not expressed by any of the premalignant lesions.

(C) The only sample that is neither classical nor basal-like expresses cancer stem cell (CSC) markers.

(D and E) (D) Differential expression analysis identified genes whose up-regulation (blue dots) or down-regulation (red dots) in PanINs, relative to normal ducts, discriminate preneoplastic from normal cells (E).

normal ducts through LG and HG lesions (Figures 5F and 5G). The same pattern was observed for *TFF1*, which was found to be up-regulated in the PanIN expressing the classical PDAC genes. This gradual change in expression is best visualized in one of the PanIN samples in which a single duct presents a mix of normal, LG, and HG cells (Figure 5H).

Changes in PanIN progression map to transitions in malignancy in PDAC

The examination of other molecular alterations that are present in PanINs and conserved in PDACs could provide new knowledge about the early transcriptional events of pancreatic carcinogenesis and the mechanisms driving the continuous development into invasive cancer. Our combined set of public domain scRNA-seq PDAC datasets[38] provides a cohort of over 61 samples that include true normal epithelial, tumor-adjacent normal epithelial, and PDAC cells. Although cells in this scRNA-seq data are from advanced PDAC tumors, we hypothesize that comparison of the transcriptional changes between normal and PanIN spots to the scRNA-seq data could quantify the persistence of the premalignant changes in PDAC or which features found in invasive tumors are detected in PanINs. Moreover, it is possible that the scRNA-seq data contain unlabeled PanIN cells from adjacent lesions to the tumor processed during dissociation. Therefore, integrative analysis between the ST data and scRNA-seq could further identify these cells to confirm molecular changes observed across grades of PanIN differentiation in an independent, large-scale reference atlas. Therefore, further computational methods for multi-omics integration of our ST data of PanINs and scRNA-seq data in PDAC can supplement

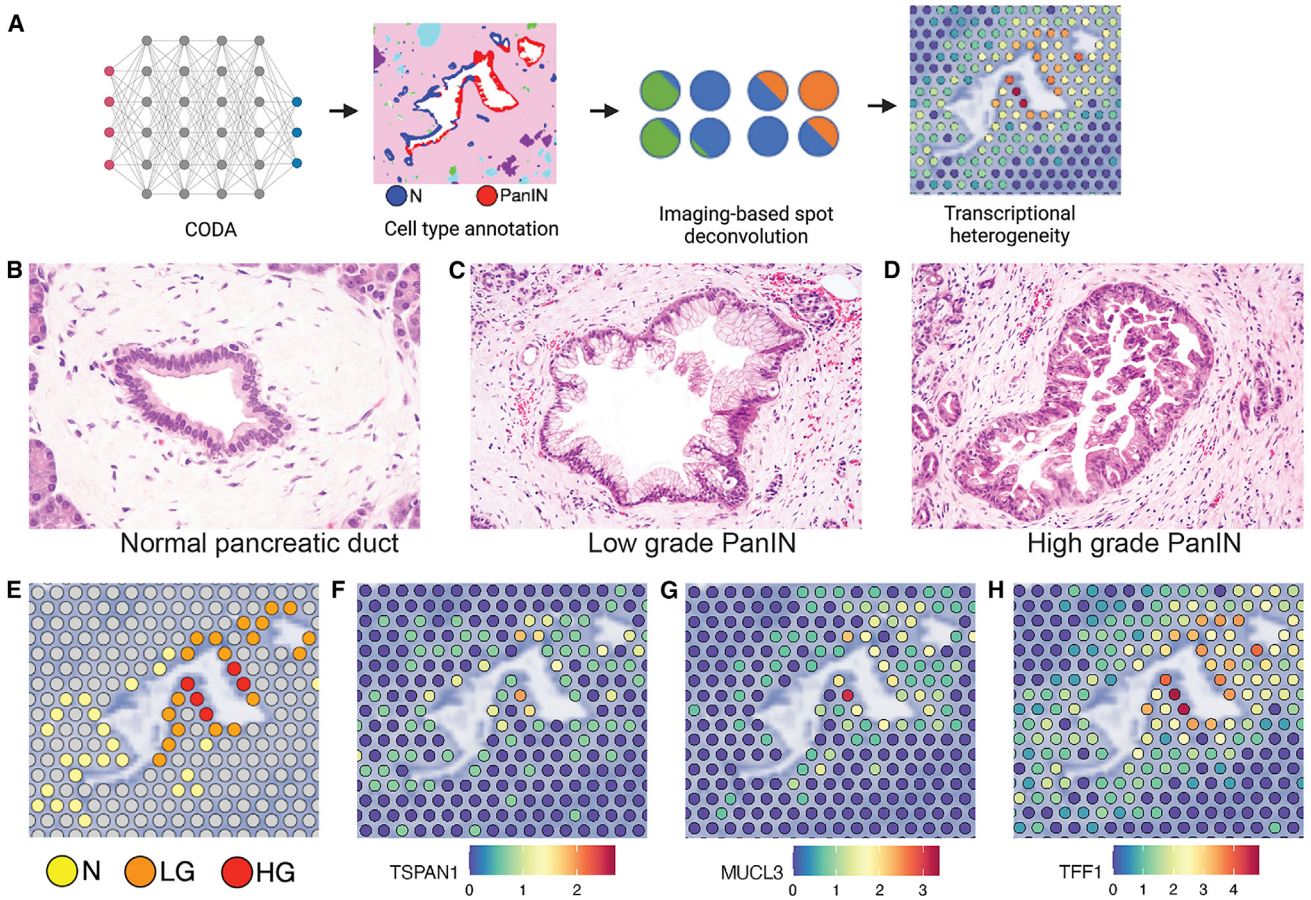


Figure 5. Identification of transcriptional changes associated with PanIN differentiation grade

(A–D) (A) Workflow of CODA annotations to facilitate heterogeneity detection. (B) Normal ducts, (C) low-grade (LG), and (D) high-grade (HG) pancreatic intraepithelial neoplasias (PanINs) are morphologically distinct and can be classified by pathology examination.

(E–H) (E) As a model for PanIN progression, a mixed pancreatic duct containing normal, LG, and HG cells was used to better visualize changes in expression. Top genes from the differential expression analysis, (F) *MUCL3*, (G) *TSPAN1*, and (H) *TFF1*, show gradual increase from normal through LG until HG progression.

our analysis of molecular changes in the epithelial cells that underly carcinogenesis (Figure 6A).

We used the scRNA-seq data of 25,442 epithelial ductal cells from 61 biospecimens collated from six previously published PDAC scRNA-seq datasets to enable tumor progression analysis (Figure S17A). The uniform manifold approximation and projection (UMAP) analysis of these cells identifies a phenotypic switch between true normal epithelial cells, tumor-adjacent normal cells, and within malignant epithelial cells, supporting our hypothesis that these datasets likely contain unannotated PanIN cells. Using this dataset, we verified that *TFF1* expression increases between normal epithelial cells, in tumor-adjacent normal epithelial cells, and again further increases in a subset of malignant PDAC cells (Figure S17B), mirroring the stage-specific increase in its expression observed in PanIN cells. This integrative analysis further supports the association of this gene with PanIN and invasive PDAC progression. *TFF1* expression is almost undetectable in normal ductal cells. Surprisingly, the normal ductal cells adjacent to tumor cells express low levels of *TFF1*, suggesting that the transcriptionally normal surrounding ducts are already programmed toward a premalignant state.

To further delineate the molecular transitions malignant epithelial cells undergo, our complementary study of the PDAC scRNA-seq data applied the Bayesian non-negative matrix factorization method CoGAPS[39,40] to learn transcriptional patterns that delineate transitions in the epithelial cells.[38] In this study, we integrated the patterns learned from the scRNA-seq data with our ST dataset to determine the extent to which they represent stage-related transitions in the transformation from PanIN to PDAC. To enable the integrative analysis between ST and scRNA-seq data, we adapted our transfer learning approach ProjectR[16,17] to spatial data integration by projecting the patterns learned in the scRNA-seq data onto the epithelial spots from the ST data (N = 240 spots; normal = 93, LG = 48, HG = 99). Among the patterns projected from the atlas onto the ST data, a pattern enriched with genes involved in KRAS signaling and proliferation (pattern 2) showed marked increase of pattern weights from normal epithelium through LG and HG PanINs (Figures 6B–6D and S17C). Increased proliferation during tumorigenesis involving the pathways contributing to pattern 2 corroborates previously reported studies showing up-regulation of pancreatic oncogenic signaling pathways in premalignancy

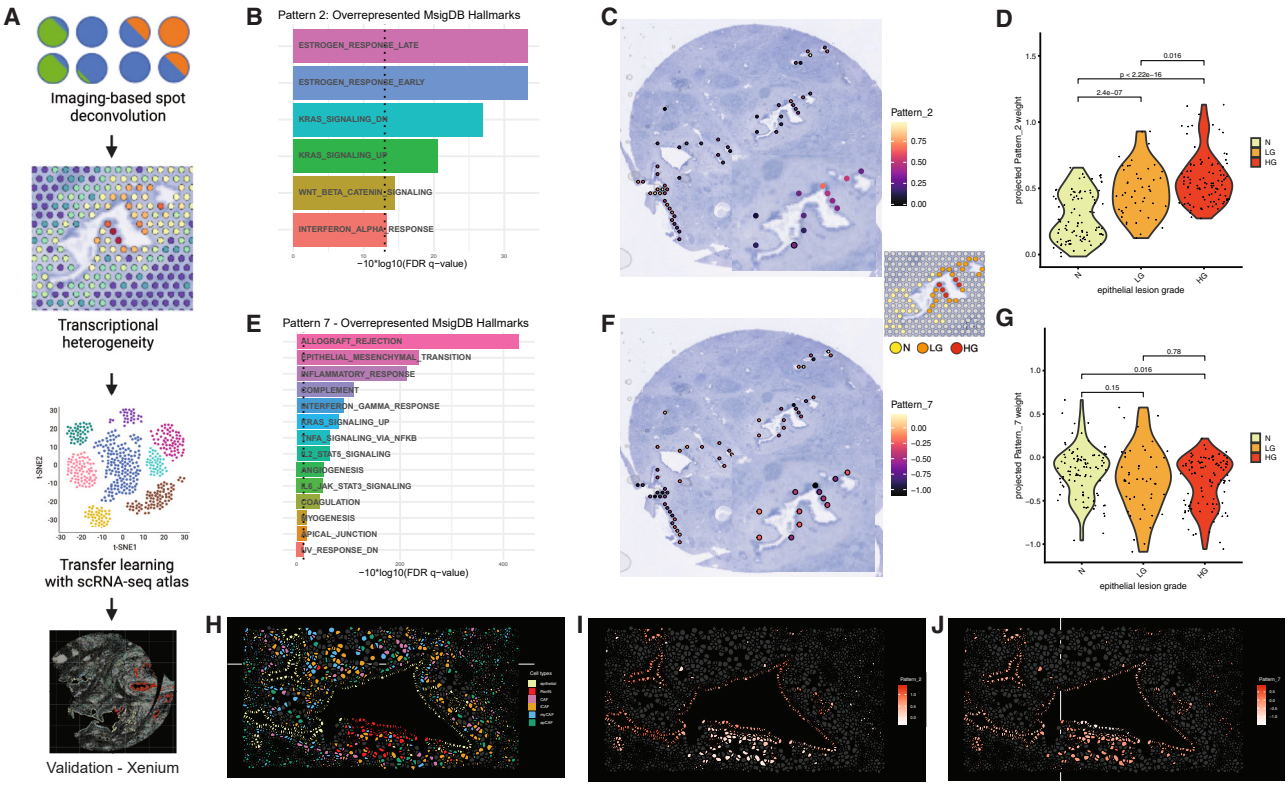


Figure 6. Integration of PanIN ST data with invasive pancreatic cancer scRNA-seq using transfer learning

(A) The deconvolved ST data, after CODA annotation and quantification of cell types per spot, was used to integrate PanIN analysis with that of scRNA-seq from human PDAC and subsequent validation with Xenium.

(B) The PDAC pattern 2 (proliferation) identified as highly expressed in PDAC cells from the atlas.

(C and D) The PDAC pattern 2 shows gradual increase from normal ductal cells through LG to HG PanINs.

(E) The PDAC pattern 7 (inflammation) presents decreased expression in PDAC cells relative to normal epithelium from the atlas.

(F and G) The opposite as observed with pattern 2, pattern 7 (inflammatory) decreases in PanINs relative to normal cells.

(H) In the Xenium data, the visualization of a mixed duct (normal + PanIN) highlights the trend between the PDAC patterns.

(I and J) (I) The projection of PDAC pattern 2 in the mixed duct and of the (J) PDAC pattern 7 confirmed the expression of both patterns using single-cell transcriptomics with RNA *in situ* hybridization and the switch of cells that express one pattern or the other.

initiation and progression.[41,42] Pattern 7, representing a combined inflammatory and EMT state associated with CAF density (Figures 6E and S17D),[38] is enriched in normal ductal cells and dissipates with the development of early-stage PDAC and progression to advanced cancer. Pattern 7 also showed decreasing levels over the course of progression from normal cells to PanIN (independent of the differentiation grade), as demonstrated by the increase in the number of spots with low projected weights (Figures 6F and 6G). This same shift between PDAC patterns 2 and 7 remains when the data is projected within the ST spots that have been classified as normal, LG, and HG by pathologists (Figures S18A–S18D), which retains more spots with low epithelial purity than the CODA classification.

We sought to validate the cellular features at a single-cell level. The Xenium panel was customized to include pattern marker genes from patterns 2 and 7 (10 genes and 103 genes, respectively). First, we sought to validate that our transfer learning approach could also be applied to quantify the occurrence of these patterns defined from scRNA-seq data in Xenium despite the reduced number of features. To test the robustness of transfer learning, we projected the 8 PDAC patterns learned by CoGAPS on the epithelial cells in the PDAC scRNA-seq data only for the 362 genes in common between Xenium and the PDAC scRNA-seq data. Spearman correlations were calculated for each pattern between the original PDAC pattern weights and the new PanIN projected weights. The most highly correlated projected weights with original pattern weights were pattern 5 ($R = 0.84$, $p < 2.2e{-}16$), pattern 7 ($R = 0.83$, $p < 2.2e{-}16$), and pattern 2 ($R = 0.83$, $p < 2.2e{-}16$). We attribute this high correlation to the panel design selected from pattern marker genes, designed specifically to delineate these phenotypes, demonstrating the potential robustness of transfer learning analysis from scRNA-seq to Xenium. Subsequently, we applied ProjectR to identify the PDAC patterns 2 and 7 in the Xenium data. This analysis supports the same trends observed in the ST Visium data with a tradeoff between patterns 2 and 7 during PanIN progression, and that is even more evident when observing these changes within sample PanIN-HG2 mixed duct (Figures 6H–6J), corroborating and further refining the inferred phenotypic changes in the scRNA-seq data of PDAC.

While the transfer learning analysis enables us to relate phenotypic changes between normal and tumor cells to PanIN

progression, we hypothesized that our ability to identify specific LG and HG lesions through ST could refine these epithelial transitions. Therefore, we performed further CoGAPS analysis on the epithelial (normal and PanIN) ST spots annotated by CODA to compare with the patterns learned from the scRNA-seq PDAC atlas. We discover a pattern (ST pattern 3) that increases progressively from normal duct to LG through HG PanIN (Figure S19A) and is similar to the PDAC pattern 2 in terms of enriched pathways and projected distribution in the scRNA-seq data (Figures S19B and S19C). We also uncovered three patterns with the opposite trend (Figure S19A). Pathway analysis reveals these patterns are characterized by genes associated with EMT and inflammation (Figure S19B). Two of these patterns (ST patterns 2 and 5) show partial overlap with PDAC pattern 7 derived from the atlas (Figure S19C). The ST pattern 2 represents an EMT-enriched pattern, while ST pattern 5 is enriched for inflammatory-related genes. These data demonstrate that these ST patterns represent distinct components of the inflammatory/EMT signature captured by PDAC pattern 7, and that they both show relative attenuation in HG PanIN compared with normal duct (ST patterns 2 and 5) and LG PanIN (ST pattern 5 only) (Figures S18A and S18B). Overall, when performed on the PanINs, CoGAPS recovers a gene signature similar to proliferative PDAC pattern 2 and separately recovers the inflammatory (ST pattern 5) and EMT (ST pattern 2) signatures represented by PDAC pattern 7.

The integration of PanIN ST data and the single-cell PDAC data provides further evidence that during PDAC initiation, as PanIN lesions develop to the invasive state, there is a continuous increase in proliferation capability in combination with loss of inflammatory signaling in epithelial cells that is potentially driving an immunosuppressive TME or tumor immune evasion. While these findings have to be further validated, this analysis demonstrates the potential of using transfer learning to integrate spatial and single-cell multi-omics datasets generated through different experimental approaches and custom panel designs.

DISCUSSION

ST technologies are uncovering new molecular and intercellular interactions that provide insights into how these complex signaling networks mediate cancer development and progression.[2] In this study, we applied an FFPE compatible ST approach[43] to profile a novel cohort of PanIN samples progressing from LG to HG lesions. An independent cohort with the same technology, additional single-cell resolution spatial proteomics and imaging transcriptomics profiling, and independent scRNA-seq data of PDAC tumors are used for validation. Our study aimed to uncover the cellular and molecular features potentially associated with the progression from premalignancies to invasive PDAC. For these analyses, we introduced a new imaging, spatial multi-omics, and scRNA-seq integration analysis pipeline to infer phenotypic transitions in carcinogenesis. The major innovation of this pipeline is leveraging two machine learning methods for integrative analysis across imaging, ST profiles, and scRNA-seq data to automatically infer spots associated with disease-relevant cellular features in the stained imaging in the ST analysis pipeline.

The first machine learning method, CODA,[14] enabled the automated assignment of cell types to ST spots using single-cell resolution to classify the cells in each sample based on the imaging of the each ST section. In contrast to other recent integration methods that perform purely unsupervised co-clustering of morphology and gene expression for deconvolution, like Starfysh and iSTAR,[12,13] CODA was specifically trained to embed pathological knowledge to automatically annotate cell types in pancreatic precancer.[44] Unlike computational spot deconvolution methods that rely on prior knowledge about the molecular features of cell types, such as gene expression signatures (e.g., BayesPrism and RCTD),[10,45] our artificial intelligence method for cellular purification requires no prior reference molecular atlas. This is particularly useful when studying histologic features that are difficult to confidently annotate in scRNA-Seq data, such as PanIN and other premalignancies that are molecularly similar to PDAC. Additionally, purifying transcriptomic groups by morphology eliminates the selection bias introduced by predetermined molecular features, and thus preserves within-group heterogeneity and facilitates the study of poorly characterized transcriptomic features. This imaging and ST analysis integration facilitated accurate assignment of cell types, selection of spots using cell purity as a threshold for downstream gene expression analysis, providing a framework for future semi-supervised ST deconvolution methods that incorporate expert pathological knowledge for analysis.

The second machine learning method on our pipeline enables inference of cell state transitions and their relation to reference scRNA-seq data to delineate the molecular mechanisms of PDAC carcinogenesis. Annotating spots to known cellular features from the stained section enables us to significantly improve the overall purity to enhance even differential expression analysis and unsupervised non-negative matrix factorization analysis comparing the molecular changes between normal and PanIN ducts. While focused on epithelial cells, we also applied this imaging enhancement to refine gene expression signatures of pancreatic islets and stromal cells. Beyond these comparative analyses, we sought to further determine which dynamic phenotypic transitions in these cells remain in advanced PDAC and if candidate PanIN cells from tumor-adjacent lesions can be identified in reference scRNA-seq data from dissociated tissue. Therefore, the second computational method in our new ST analysis pipeline, ProjectR,[16,17] allowed the integration of scRNA-seq from invasive PDACs with ST data from PanINs to relate the mechanisms associated with PDAC initiation to subsequent progression. While this study is focused on PanIN progression, this pipeline could enable important future work leveraging different datasets of the distinct pancreatic precursor lesions to compare transcriptional programs and identify the mechanisms of progression into invasive PDAC. Although this study focused on epithelial cells, our combined data-integration pipeline is applicable to any cell type shared between CODA annotations and scRNA-seq data and provides a broader multi-omics framework for the study of carcinogenesis in different tissue types.

Applying our combined experimental and computational approach to PanIN samples, we observed for the first time the presence of CAFs and the different subtypes (myCAF, iCAF, and apCAF) in premalignant human lesions. These subtypes

were only previously described in PDAC in humans.[20,22] Our quantification of apCAFs by ST detected that as many as 13.9% of cells in a section containing PanIN expressed apCAF signature genes and no CD45. Protein validation by IMC confirmed the presence of these apCAFs in proximity to PanIN at lower frequency. Of note, the IMC data were generated using a panel that was specifically designed to detect pancreatic associated fibroblasts and CAFs, limiting the detection of cell types, and the proportion of apCAFs can only be determined relative to the total CAFs present in the regions profiled. In general, CAFs are the most abundant cell type in the PDAC TME and are known to influence tumor cell behavior and to create an immunosuppressive environment.[46] The presence of these regulatory cells in human pancreatic premalignant lesions is not well described, but our findings suggest that CAF-induced TME remodeling is an early event with a durable impact on PDAC development. In a recent publication, Carpenter et al. used ST to profile PanIN diagnosed in normal pancreatic tissue collected from healthy organ donors and observed that the fibroblasts adjacent to the healthy-associated PanINs are transcriptionally different from healthy pancreas fibroblasts and PDAC-associated CAFs, but with similar heterogeneity observed in the latest.[47] This suggests that even in patients without PDAC, the healthy microenvironment adjacent to the PanINs is already being reshaped, the same way we observed in our PDAC-associated samples. Further studies are necessary to examine the specific interactions driven by the different CAF subtypes, how they modulate premalignant cells, and other cellular components of the PDAC TME. Such knowledge is critical to guide the development of new therapeutic interventions that inhibit or revert CAF oncogenic and immunosuppressive activity with the goal of intercepting PDAC development.

ST analysis of the PanINs also identified transcriptional signatures that are known to be associated with PDAC phenotypes. PDACs are classified into classical and basal-like transcriptional subtypes.[27] Classical PDACs present a better prognosis and represent most tumor cells found in early-stage cancers before patients receive treatment. This supports the hypothesis that all PDAC initially develops from the classical phenotype, and there is a diverging point during the tumorigenesis in which some cells will differentiate into the basal-like phenotype. This classical to basal-like transition is usually expanded by chemotherapy as resistance develops.[27,28] Further supporting this hypothesis, is the fact that the PanIN lesions spatially profiled in this study and by Carpenter et al.[47] only express the classical signature. In our cohort, there was only one sample that could not be classified as classical or basal-like but that expressed a CSC signature. CSCs drive aggressive disease, and their presence is associated with resistance to therapies, local recurrence, and development of metastasis.[48–50] The presence of cells expressing CSC markers in PanINs was previously described in a mouse model that mimics PDAC development[51] and in human samples,[52] but little is known about the mechanisms leading to CSC genes upregulation and their role in PanIN initiation and development. Our observation that this stemness signature is not seen in cells expressing the classical subtype suggests that atypical cells with stemness features are a rare, distinct population that arises in early premalignant stages and that these cases will potentially present with distinct behavior and response to the current therapies. Further investigation in a larger cohort is needed to determine the frequency of this rare stem cell-related mechanism of progression, the pathways driven by stemness, and how these cells are interacting with the CAFs and other cells in the TME to modulate PDAC biology. Our study is the first to our knowledge to observe CSC markers expressed by a PanIN lesion, but we note that a recent multiomics study found a population of CD133+ iCAFs that express CSC markers, but these markers were not observed in PDAC.[53] The presence of CSCs and CAFs in PanINs suggests that the features associated with resistance to therapies in PDAC arise early during tumorigenesis. As mentioned previously, further oriented studies are necessary to determine how the interactions between these cell types can modulate additional features of resistance to therapies and progression of PDACs.

Differential expression analysis of the ST data shows that LG and HG PanINs are transcriptionally similar. Among the few genes differentially expressed between these two PanIN grades, *TFF1*, frequently overexpressed in PanIN, demonstrated gradual increase during PanIN progression, but little is known about its role in tumorigenesis. As mentioned previously, secreted TFF1 could be involved in tumor cell interactions with CAFs[36,54] Transcriptional differences were also detected between healthy pancreas PanINs and tumor associate PanINs by Carpenter et al.[47] A few similarities between PanINs in healthy pancreas and tumor-associated PanINs that were common with our study, such as the overexpression *TFF1*. The authors suggest that this latest observation suggests that increased levels of *TFF1* are a feature of PanINs that are lost during the progression to PDAC since they showed that *TFF1* expression is rare in PDAC cells. However, further integrative analysis is needed to make this direct comparison between PanIN progression and advanced PDAC.

Fully relating these atypical cell state transitions inferred in ST data to cancer progression requires relating these transcriptional states across the transition, from normal epithelial through premalignancy to malignant PDAC cells. We demonstrate that transfer learning approaches developed to integrate different datasets can be extended to relate spatial data from premalignancy to reference scRNA-seq data. With this integrative analysis approach to relate mechanisms in advanced PDAC carcinogenesis to premalignancy, we observe that unsupervised learning analysis directly on the epithelial spots in our ST Visium data more specifically separates EMT and inflammatory signaling as two distinct cellular phenotypes in PanIN progression. Still, this two-stage computational approach integrating imaging and ST data of PanINs with scRNA-seq data of PDAC tumors enabled us to identify a transition from inflammatory signaling in neoplastic cells from LG PanIN to cellular proliferation in later stages of carcinogenesis. In our complementary single-cell atlas study that identifies this inflammatory signaling, we further correlated this transition with CAF abundance and validated the ability of CAFs to promote this signaling in a novel human organoid co-culture model.[38]

Although we used a limited sample number with an initial total of 14 PanINs (9 LG and 5 HG), we were able to corroborate previous findings related to PanINs and discover features that are common to invasive PDAC. Due to the small size of this cohort,

the findings cannot be extrapolated or generalized for clinical implications, but we are reassured that our cohorts recapitulate many of the well-characterized features of PanIN since these two cohorts (paired and validation) were prepared at different times using different versions of the commercial reagents and preprocessing software. The paired cohort (3 LG and 4 HG PanIN) was prepared, and the data were preprocessed using prototype reagents and software, while the validation cohort (6 LG and 1 HG PanIN) used more recent versions of both. In this scenario, the replicability of the findings between both cohorts is another certification that the findings of the ST analyses are robust. Another limitation is the lack of single-cell resolution from the ST platform (Visium). Although we were able to dramatically increase average spot purity by integrating Visium with CODA, our unbiased clustering and differential expression results may still contain artifacts originating from undesired intra-spot cell type mixture. However, using additional high-dimensional spatial single-cell transcriptomics (Xenium) and proteomics (IMC), we were able to validate the presence of specific cell types, mainly of apCAFs, that are critical for PDAC biology. These combined spatial multi-omics datasets enabled us to characterize the microenvironment in which PanINs develop and showed for the first time the presence of CAFs surrounding human PanINs and their impact on neoplastic cell signaling. Further analysis of the transcriptomics and proteomics spatial single-cell datasets could uncover new cellular and molecular features of CAF and premalignancy interactions. Our cohort included samples with varying stromal and acinar cell composition, but we did not observe correlations between PanIN transcriptional profiles with the adjacent cell types due to the limited size. To examine if the CAFs surrounding the PanINs remodel the premalignant microenvironment and influence premalignancy progression, a larger cohort with a more stringent selection criteria would be better suited. Patients' clinical features and outcomes (e.g., tumor stage, metastasis, response to therapies) would be critical to unveil the consequences of CAF-PanIN (or PDAC) interactions in PDAC tumorigenesis. Such a specific cohort would allow correlative analysis between clinico-pathological features and TME composition. Nonetheless, we demonstrate that multi-omics analysis enabled by FFPE ST, imaging data analysis, and scRNA-seq data lead to a model that allows the investigation of molecular features that are present in premalignancies and invasive carcinomas of the pancreas. Moreover, this novel hybrid experimental and robust computational pipeline provides broadly applicable tools to create a molecular and cellular model of the pathways that underlie carcinogenesis from multi-modal data spanning distinct high-dimensional transcriptomics and spatial molecular technologies. Our pipeline will facilitate analyses of future datasets aiming to characterize transcriptional changes that are selected based on grade of differentiation to better understand the mechanisms of tumorigenesis in different tissues.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cels.2024.07.001.

ACKNOWLEDGMENTS

The authors would like to thank the Oncology Tissue Services (OTS) Core Facility and the Genetic Resources Core Facility (GRCF) for tissue sections preparation and library sequencing services, respectively. A special thanks to Helen Fedor and Bonnie Gambichler from the OTS core for all their help on perfect sectioning of the samples used for Visium. We are also grateful to Ghezal Beliakoff, Robert Shelanski, Nicole Rapicavoli, and Morgane Rouault for their invaluable contribution to generating and processing the Xenium data. This work was supported by the Sol Goldman Pancreatic Cancer Research Center grant (to L.T.K.), NIH P01-CA247886-01A1 (to E.M.J., E.J.F., and L.T.K.), SU2C/AACR DT-14-14 (to E.M.J.), Lustgarten Foundation Pancreatic Cancer Research grant (to E.M.J., E.J.F., and L.T.K.), the Emerson Cancer Research Fund (to E.M.J. and E.J.F.), an Allegheny Health Network (AHN) grant (to E.J.F.), NIH U01CA212007 (to E.J.F.), NIH U01CA253403 (to E.J.F.), the JHU Discovery Award (to E.J.F.), SPORE GI P50CA062924-24A1 (to E.M.J., E.J.F., and L.T.K.), NCI F31CA268724-01 (to D.N.S.), NIH U54CA268083 (to D.W., P.-H.W., and A.L.K.), NIH U54CA210173 (D.W.), NIH U01AG060903 (to D.W.), Susan Wojcicki and Dennis Troper (to A.L.K.), the Rolfe Foundation for Pancreatic Cancer Research (to A.L.K.), Sanofi (W.J.H.), and NeoTX (W.J.H.).

AUTHOR CONTRIBUTIONS

A.T.F.B., J.T.M., and A.L.K.: data curation, formal analysis, investigation, visualization, writing original draft, and review. K.F., J.W.L., E.R., and S.M.S.: data curation, writing, and review. M.L., S.N., A.D., P.-H.W., D.N.S., R.E., J.S., R.C., S.W., J.M.C., L.C., J.W.Z., M.L., D.W., W.J.H., E.T., and N.Z.: resources and writing review. E.M.J.: resources, writing review, and editing. L.D.W.: data curation, resources, supervision, writing review, and editing. E.J.F.: conceptualization, data curation, supervision, formal analysis, visualization, writing original draft, review, and editing. L.T.K.: conceptualization, data curation, supervision, formal analysis, funding acquisition, visualization, writing original draft, review, and editing.

DECLARATION OF INTERESTS

E.M.J. reports other support from Abmeta; personal fees from Genocea; personal fees from Achilles; personal fees from DragonFly; personal fees from Candel Therapeutics; other support from the Parker Institute; grants and other support from Lustgarten; personal fees from Carta; grants and other support from Genentech; grants and other support from AstraZeneca; personal fees from NextCure; and grants and other support from Break Through Cancer outside of the submitted work. E.J.F. is on the Scientific Advisory Board of Viosera Therapeutics/Resistance Bio and is a consultant to Mestag Therapeutics. W.J.H. reports patent royalties from Rodeo/Amgen and speaking/travel honoraria from Exelixis and Standard BioTools.

Received: July 7, 2023
Revised: February 6, 2024
Accepted: July 8, 2024
Published: August 7, 2024

REFERENCES

1. Lim, B., Lin, Y., and Navin, N. (2020). Advancing Cancer Research and Medicine with Single-Cell Genomics. Cancer Cell *37*, 456–470. https://doi.org/10.1016/j.ccell.2020.03.008.
2. Davis-Marcisak, E.F., Deshpande, A., Stein-O'Brien, G.L., Ho, W.J., Laheru, D., Jaffee, E.M., Fertig, E.J., and Kagohara, L.T. (2021). From bench to bedside: single-cell analysis for cancer immunotherapy. Cancer Cell *39*, 1062–1080. https://doi.org/10.1016/j.ccell.2021.07.004.
3. Peng, J., Sun, B.-F., Chen, C.-Y., Zhou, J.-Y., Chen, Y.-S., Chen, H., Liu, L., Huang, D., Jiang, J., Cui, G.-S., et al. (2019). 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.
4. Steele, N.G., Carpenter, E.S., Kemp, S.B., Sirihorachai, V.R., The, S., Delrosario, L., Lazarus, J., Amir, E.D., Gunchick, V., Espinoza, C., et al. (2020). Multimodal mapping of the tumor and peripheral blood immune landscape in human pancreatic cancer. Nat. Cancer *1*, 1097–1112. https://doi.org/10.1038/s43018-020-00121-4.
5. Lin, W., Noel, P., Borazanci, E.H., Lee, J., Amini, A., Han, I.W., Heo, J.S., Jameson, G.S., Fraser, C., Steinbach, M., et al. (2020). Single-cell transcriptome analysis of tumor and stromal compartments of pancreatic ductal adenocarcinoma primary tumors and metastatic lesions. Genome Med. *12*, 80. https://doi.org/10.1186/s13073-020-00776-9.
6. Bernard, V., Semaan, A., Huang, J., San Lucas, F.A., Mulu, F.C., Stephens, B.M., Guerrero, P.A., Huang, Y., Zhao, J., Kamyabi, N., et al. (2019). Single-Cell Transcriptomics of Pancreatic Cancer Precursors Demonstrates Epithelial and Microenvironmental Heterogeneity as an Early Event in Neoplastic Progression. Clin. Cancer Res. *25*, 2194–2205. https://doi.org/10.1158/1078-0432.CCR-18-1955.
7. Raghavan, S., Winter, P.S., Navia, A.W., Williams, H.L., DenAdel, A., Lowder, K.E., Galvez-Reyes, J., Kalekar, R.L., Mulugeta, N., Kapner, K.S., et al. (2021). Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. Cell *184*, 6119–6137.e26. https://doi.org/10.1016/j.cell.2021.11.017.
8. Distler, M., Aust, D., Weitz, J., Pilarsky, C., and Grützmann, R. (2014). Precursor lesions for sporadic pancreatic cancer: PanIN, IPMN, and MCN. BioMed Res. Int. *2014*, 474905. https://doi.org/10.1155/2014/474905.
9. Hruban, R.H., Maitra, A., Kern, S.E., and Goggins, M. (2007). Precursors to pancreatic cancer. Gastroenterol. Clin. North Am. *36*, 831–849. vi. https://doi.org/10.1016/j.gtc.2007.08.012.
10. Cable, D.M., Murray, E., Zou, L.S., Goeva, A., Macosko, E.Z., Chen, F., and Irizarry, R.A. (2022). Robust decomposition of cell type mixtures in spatial transcriptomics. Nat. Biotechnol. *40*, 517–526. https://doi.org/10.1038/s41587-021-00830-w.
11. Chu, T., Wang, Z., Pe'er, D., and Danko, C.G. (2022). Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. Nat. Cancer *3*, 505–517. https://doi.org/10.1038/s43018-022-00356-3.
12. He, S., Jin, Y., Nazaret, A., Shi, L., Chen, X., Rampersaud, S., Dhillon, B.S., Valdez, I., Friend, L.E., Fan, J.L., et al. (2022). Starfysh reveals heterogeneous spatial dynamics in the breast tumor microenvironment. Preprint at bioRxiv. https://doi.org/10.1101/2022.11.21.517420.
13. Zhang, D., Schroeder, A., Yan, H., Yang, H., Hu, J., Lee, M.Y.Y., Cho, K.S., Susztak, K., Xu, G.X., Feldman, M.D., et al. (2024). Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. Nat. Biotechnol. https://doi.org/10.1038/s41587-023-02019-9.
14. Kiemen, A.L., Braxton, A.M., Grahn, M.P., Han, K.S., Mahesh Babu, J., Reichel, R., Amoa, F., Hong, S.-M., Cornish, T.C., Thompson, E.D., et al. (2020). In situ characterization of the 3D microanatomy of the pancreas and pancreatic cancer at single cell resolution. Preprint at bioRxiv. https://doi.org/10.1101/2020.12.08.416909.
15. Guinn, S., Kinny-Köster, B., Tandurella, J.A., Mitchell, J.T., Sidiropoulos, D.N., Loth, M., Lyman, M.R., Pucsek, A.B., Zabransky, D.J., Lee, J.W., et al. (2024). Transfer Learning Reveals Cancer-Associated Fibroblasts Are Associated with Epithelial-Mesenchymal Transition and Inflammation in Cancer Cells in Pancreatic Ductal Adenocarcinoma. Cancer Res. *84*, 1517–1533. https://doi.org/10.1158/0008-5472.CAN-23-1660.
16. Sharma, G., Colantuoni, C., Goff, L.A., Fertig, E.J., and Stein-O'Brien, G. (2020). projectR: an R/Bioconductor package for transfer learning via PCA, NMF, correlation and clustering. Bioinformatics *36*, 3592–3593. https://doi.org/10.1093/bioinformatics/btaa183.
17. Stein-O'Brien, G.L., Clark, B.S., Sherman, T., Zibetti, C., Hu, Q., Sealfon, R., Liu, S., Qian, J., Colantuoni, C., Blackshaw, S., et al. (2019). Decomposing Cell Identity for Transfer Learning across Cellular Measurements, Platforms, Tissues, and Species. Cell Syst. *8*, 395–411.e8. https://doi.org/10.1016/j.cels.2019.04.004.
18. Hruban, R.H., Goggins, M., Parsons, J., and Kern, S.E. (2000). Progression model for pancreatic cancer. Clin. Cancer Res. *6*, 2969–2972.
19. Ni, Z., Prasad, A., Chen, S., Halberg, R.B., Arkin, L.M., Drolet, B.A., Newton, M.A., and Kendziorski, C. (2022). SpotClean adjusts for spot swapping in spatial transcriptomics data. Nat. Commun. *13*, 2971. https://doi.org/10.1038/s41467-022-30587-y.
20. Öhlund, D., Handly-Santana, A., Biffi, G., Elyada, E., Almeida, A.S., Ponz-Sarvise, M., Corbo, V., Oni, T.E., Hearn, S.A., Lee, E.J., et al. (2017). Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. J. Exp. Med. *214*, 579–596. https://doi.org/10.1084/jem.20162024.
21. Helms, E., Onate, M.K., and Sherman, M.H. (2020). Fibroblast heterogeneity in the pancreatic tumor microenvironment. Cancer Discov. *10*, 648–656. https://doi.org/10.1158/2159-8290.CD-19-1353.
22. Elyada, E., Bolisetty, M., Laise, P., Flynn, W.F., Courtois, E.T., Burkhart, R.A., Teinor, J.A., Belleau, P., Biffi, G., Lucito, M.S., et al. (2019). Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. Cancer Discov. *9*, 1102–1123. https://doi.org/10.1158/2159-8290.CD-19-0094.
23. Mizutani, Y., Kobayashi, H., Iida, T., Asai, N., Masamune, A., Hara, A., Esaki, N., Ushida, K., Mii, S., Shiraki, Y., et al. (2019). Meflin-Positive Cancer-Associated Fibroblasts Inhibit Pancreatic Carcinogenesis. Cancer Res. *79*, 5367–5381. https://doi.org/10.1158/0008-5472.CAN-19-0454.
24. Helms, E.J., Berry, M.W., Chaw, R.C., DuFort, C.C., Sun, D., Onate, M.K., Oon, C., Bhattacharyya, S., Sanford-Crane, H., Horton, W., et al. (2022). Mesenchymal Lineage Heterogeneity Underlies Nonredundant Functions of Pancreatic Cancer-Associated Fibroblasts. Cancer Discov. *12*, 484–501. https://doi.org/10.1158/2159-8290.CD-21-0601.
25. Hosein, A.N., Huang, H., Wang, Z., Parmar, K., Du, W., Huang, J., Maitra, A., Olson, E., Verma, U., and Brekken, R.A. (2019). Cellular heterogeneity during mouse pancreatic ductal adenocarcinoma progression at single-cell resolution. JCI Insight *5*, e129212. https://doi.org/10.1172/jci.insight.129212.
26. Huang, H., Wang, Z., Zhang, Y., Pradhan, R.N., Ganguly, D., Chandra, R., Murimwa, G., Wright, S., Gu, X., Maddipati, R., et al. (2022). Mesothelial cell-derived antigen-presenting cancer-associated fibroblasts induce expansion of regulatory T cells in pancreatic cancer. Cancer Cell *40*, 656–673.e7. https://doi.org/10.1016/j.ccell.2022.04.011.
27. Moffitt, R.A., Marayati, R., Flate, E.L., Volmar, K.E., Loeza, S.G.H., Hoadley, K.A., Rashid, N.U., Williams, L.A., Eaton, S.C., Chung, A.H., et al. (2015). Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat. Genet. *47*, 1168–1178. https://doi.org/10.1038/ng.3398.
28. Chan-Seng-Yue, M., Kim, J.C., Wilson, G.W., Ng, K., Figueroa, E.F., O'Kane, G.M., Connor, A.A., Denroche, R.E., Grant, R.C., McLeod, J.,

et al. (2020). Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. Nat. Genet. *52*, 231–240. https://doi.org/10.1038/s41588-019-0566-9.

29. Ishiwata, T., Matsuda, Y., Yoshimura, H., Sasaki, N., Ishiwata, S., Ishikawa, N., Takubo, K., Arai, T., and Aida, J. (2018). Pancreatic cancer stem cells: features and detection methods. Pathol. Oncol. Res. *24*, 797–805. https://doi.org/10.1007/s12253-018-0420-x.
30. Prasad, N.B., Biankin, A.V., Fukushima, N., Maitra, A., Dhara, S., Elkahloun, A.G., Hruban, R.H., Goggins, M., and Leach, S.D. (2005). Gene expression profiles in pancreatic intraepithelial neoplasia reflect the effects of Hedgehog signaling on pancreatic ductal epithelial cells. Cancer Res. *65*, 1619–1626. https://doi.org/10.1158/0008-5472.CAN-04-1413.
31. Ayars, M., O'Sullivan, E., Macgregor-Das, A., Shindo, K., Kim, H., Borges, M., Yu, J., Hruban, R.H., and Goggins, M. (2017). IL2RG, identified as overexpressed by RNA-seq profiling of pancreatic intraepithelial neoplasia, mediates pancreatic cancer growth. Oncotarget *8*, 83370–83383. https://doi.org/10.18632/oncotarget.19848.
32. Buchholz, M., Braun, M., Heidenblut, A., Kestler, H.A., Klöppel, G., Schmiegel, W., Hahn, S.A., Lüttges, J., and Gress, T.M. (2005). Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions. Oncogene *24*, 6626–6636. https://doi.org/10.1038/sj.onc.1208804.
33. Sodir, N.M., Kortlever, R.M., Barthet, V.J.A., Campos, T., Pellegrinet, L., Kupczak, S., Anastasiou, P., Swigart, L.B., Soucek, L., Arends, M.J., et al. (2020). MYC instructs and maintains pancreatic adenocarcinoma phenotype. Cancer Discov. *10*, 588–607. https://doi.org/10.1158/2159-8290.CD-19-0435.
34. Maddipati, R., Norgard, R.J., Baslan, T., Rathi, K.S., Zhang, A., Saeid, A., Higashihara, T., Wu, F., Kumar, A., Annamalai, V., et al. (2022). MYC levels regulate metastatic heterogeneity in pancreatic adenocarcinoma. Cancer Discov. *12*, 542–561. https://doi.org/10.1158/2159-8290.CD-20-1826.
35. Ashton, T.M., McKenna, W.G., Kunz-Schughart, L.A., and Higgins, G.S. (2018). Oxidative phosphorylation as an emerging target in cancer therapy. Clin. Cancer Res. *24*, 2482–2490. https://doi.org/10.1158/1078-0432.CCR-17-3070.
36. Arumugam, T., Brandt, W., Ramachandran, V., Moore, T.T., Wang, H., May, F.E., Westley, B.R., Hwang, R.F., and Logsdon, C.D. (2011). Trefoil factor 1 stimulates both pancreatic cancer and stellate cells and increases metastasis. Pancreas *40*, 815–822. https://doi.org/10.1097/MPA.0b013e31821f6927.
37. Manoukian, P., Bijlsma, M., and van Laarhoven, H. (2021). The Cellular Origins of Cancer-Associated Fibroblasts and Their Opposing Contributions to Pancreatic Cancer Growth. Front. Cell Dev. Biol. *9*, 743907. https://doi.org/10.3389/fcell.2021.743907.
38. Kinny-Koster, B., Guinn, S., Tandurella, J.A., Mitchell, J.T., Sidiropoulos, D.N., Loth, M., Lyman, M.R., Pucsek, A.B., Seppala, T.T., Cherry, C., et al. (2022). Inflammatory Signaling and Fibroblast-Cancer Cell Interactions Transfer from a Harmonized Human Single-cell RNA Sequencing Atlas of Pancreatic Ductal Adenocarcinoma to Organoid Co-Culture. Preprint at bioRxiv. https://doi.org/10.1101/2022.07.14.500096.
39. Fertig, E.J., Ding, J., Favorov, A.V., Parmigiani, G., and Ochs, M.F. (2010). CoGAPS: an R/C++ package to identify patterns and biological process activity in transcriptomic data. Bioinformatics *26*, 2792–2793. https://doi.org/10.1093/bioinformatics/btq503.
40. Sherman, T.D., Gao, T., and Fertig, E.J. (2020). CoGAPS 3: Bayesian non-negative matrix factorization for single-cell analysis with asynchronous updates and sparse data structures. BMC Bioinformatics *21*, 453. https://doi.org/10.1186/s12859-020-03796-9.
41. Feldmann, G., Beaty, R., Hruban, R.H., and Maitra, A. (2007). Molecular genetics of pancreatic intraepithelial neoplasia. J. Hepatobiliary Pancreat. Surg. *14*, 224–232. https://doi.org/10.1007/s00534-006-1166-5.
42. Lee, J., Snyder, E.R., Liu, Y., Gu, X., Wang, J., Flowers, B.M., Kim, Y.J., Park, S., Szot, G.L., Hruban, R.H., et al. (2017). Reconstituting development of pancreatic intraepithelial neoplasia from primary human pancreas duct cells. Nat. Commun. *8*, 14686. https://doi.org/10.1038/ncomms14686.
43. Gracia Villacampa E., Larsson L., Mirzazadeh R., Kvastad L., Andersson A., Mollbrink A., Kokaraki G., Monteil V., Schultz N., Appelberg K.S., et al. Genome-wide spatial expression profiling in formalin-fixed tissues. Cell Genom. 2021 Dec 8;1:100065. doi: 10.1016/j.xgen.2021.100065.
44. Kiemen, A.L., Braxton, A.M., Grahn, M.P., Han, K.S., Babu, J.M., Reichel, R., Jiang, A.C., Kim, B., Hsu, J., Amoa, F., et al. (2022). CODA: quantitative 3D reconstruction of large tissues at cellular resolution. Nat. Methods *19*, 1490–1499. https://doi.org/10.1038/s41592-022-01650-9.
45. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457. https://doi.org/10.1038/nmeth.3337.
46. Ho, W.J., Jaffee, E.M., and Zheng, L. (2020). The tumour microenvironment in pancreatic cancer - clinical challenges and opportunities. Nat. Rev. Clin. Oncol. *17*, 527–540. https://doi.org/10.1038/s41571-020-0363-5.
47. Carpenter, E.S., Elhossiny, A.M., Kadiyala, P., Li, J., McGue, J., Griffith, B.D., Zhang, Y., Edwards, J., Nelson, S., Lima, F., et al. (2023). Analysis of donor pancreata defines the transcriptomic signature and microenvironment of early neoplastic lesions. Cancer Discov. *13*, 1324–1345. https://doi.org/10.1158/2159-8290.CD-23-0013.
48. Hermann, P.C., and Sainz, B. (2018). Pancreatic cancer stem cells: A state or an entity? Semin. Cancer Biol. *53*, 223–231. https://doi.org/10.1016/j.semcancer.2018.08.007.
49. Valle, S., Martin-Hijano, L., Alcalá, S., Alonso-Nocelo, M., and Sainz, B. (2018). The Ever-Evolving Concept of the Cancer Stem Cell in Pancreatic Cancer. Cancers (Basel) *10*, 33. https://doi.org/10.3390/cancers10020033.
50. Askan, G., Sahin, I.H., Chou, J.F., Yavas, A., Capanu, M., Iacobuzio-Donahue, C.A., Basturk, O., and O'Reilly, E.M. (2021). Pancreatic cancer stem cells may define tumor stroma characteristics and recurrence patterns in pancreatic ductal adenocarcinoma. BMC Cancer *21*, 385. https://doi.org/10.1186/s12885-021-08123-w.
51. Maruno, T., Fukuda, A., Goto, N., Tsuda, M., Ikuta, K., Hiramatsu, Y., Ogawa, S., Nakanishi, Y., Yamaga, Y., Yoshioka, T., et al. (2021). Visualization of stem cell activity in pancreatic cancer expansion by direct lineage tracing with live imaging. eLife *10*, e55117. https://doi.org/10.7554/eLife.55117.
52. Kure, S., Matsuda, Y., Hagio, M., Ueda, J., Naito, Z., and Ishiwata, T. (2012). Expression of cancer stem cell markers in pancreatic intraepithelial neoplasias and pancreatic ductal adenocarcinomas. Int. J. Oncol. *41*, 1314–1324. https://doi.org/10.3892/ijo.2012.1565.
53. Cui Zhou, D., Jayasinghe, R.G., Chen, S., Herndon, J.M., Iglesia, M.D., Navale, P., Wendl, M.C., Caravan, W., Sato, K., Storrs, E., et al. (2022). Spatially restricted drivers and transitional cell populations cooperate with the microenvironment in untreated and chemo-resistant pancreatic cancer. Nat. Genet. *54*, 1390–1405. https://doi.org/10.1038/s41588-022-01157-1.
54. Sunagawa, M., Yamaguchi, J., Kokuryo, T., Ebata, T., Yokoyama, Y., Sugawara, G., and Nagino, M. (2017). Trefoil factor family 1 expression in the invasion front is a poor prognostic factor associated with lymph node metastasis in pancreatic cancer. Pancreatology *17*, 782–787. https://doi.org/10.1016/j.pan.2017.07.188.
55. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
56. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

57. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420. https://doi.org/10.1038/nbt.4096.

58. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502. https://doi.org/10.1038/nbt.3192.

59. Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. *16*, 278. https://doi.org/10.1186/s13059-015-0844-5.

60. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2016). Fast gene set enrichment analysis. Preprint at bioRxiv. https://doi.org/10.1101/060012.

61. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

62. Wickham, H. (2016). Elegant Graphics for Data Analysis, Second Edition (Springer International Publishing), p. Ggplot2.

63. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313.

64. Stein-O'Brien, G.L., Carey, J.L., Lee, W.S., Considine, M., Favorov, A.V., Flam, E., Guo, T., Li, S., Marchionni, L., Sherman, T., et al. (2017). PatternMarkers & GWCoGAPS for novel data-driven biomarkers via whole transcriptome NMF. Bioinformatics *33*, 1892–1894. https://doi.org/10.1093/bioinformatics/btx058.

65. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425. https://doi.org/10.1016/j.cels.2015.12.004.

66. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at arXiv.

67. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386. https://doi.org/10.1038/nbt.2859.

68. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982. https://doi.org/10.1038/nmeth.4402.

69. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.

70. Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427. https://doi.org/10.1038/nbt.4091.

71. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233. https://doi.org/10.1038/s41598-019-41695-z.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Conjugated IMC antibodies	N/A	All information on clones, companies, dilution factors, etc are included in Table S2.
Chemicals, peptides, and recombinant proteins		
Molecular biology grade water	Corning	Catalog #46-000-CI
Xylene, Histological Grade	Milipore Sigma	Catalog #534056
Hematoxylin Solution, Mayer's	Milipore Sigma	Catalog #MHS16
Bluing reagent	Dako	Catalog #CS70230-2
Eosin Y-solution, Alcoholic	Milipore Sigma	Catalog #HT110116
Tris 1M, pH 7.0, RNase-free	Thermo Fisher Scientific	Catalog #AM9850G
PBS 1x, pH 7.4	Corning	Catalog #21-040-CV
Tween 20	Thermo Fisher Scientific	Catalog #28320
KAPA SYBR FAST qPCR Master Mix (2X)	KAPA Biosystems	Catalog #KK4600
SPRIselect Reagent	Beckman Coulter	Catalog #B23318
Ethyl Alcohol, Pure (200 proof, anhydrous)	Millipore Sigma	Catalog #E7023-500ML
Potassium Hydroxide Solution, 8M	Millipore Sigma	Catalog # P4494-50ML
Qiagen Buffer EB	Qiagen	Catalog # 19086
Glycerol solution	Milipore Sigma	49781
Hydrochloric acid solution, 0.1N	Fisher Chemical	SA54-1
TE buffer (pH9.0)	N/A	N/A
Sodium dodecyl sulfate solution	Milipore Sigma	Catalog #71736-500ML
SSC Buffer 20x, Concentrate	Milipore Sigma	Catalog #S6639-1L
Critical commercial assays		
RNeasy extraction kit	Qiagen	Catalog #73504
Bioanalyzer RNA 6000 Pico Kit	Agilent	Catalog #5067-1513
High Sensitivity DNA Kit	Agilent	Catalog #5067-4626
Visium Spatial for FFPE Gene Expression Kit, Human Transcriptome, 16 rxns	10x Genomics	Catalog #1000336
Dual Index Kit TS Set A, 96 rxns	10x Genomics	Catalog #1—251
Software and algorithms		
NDP Scan v3.4	Hamamatsu	N/A
Space Ranger	10x Genomics	N/A
Seurat	N/A	N/A
CODA	N/A	N/A
CoGAPS	N/A	N/A
ProjectR	N/A	N/A
Other		
Epredia HM 355S Automatic Microtome	Fisher Scientific	Catalog #23-900-672
Epredia MX35 Premier Disposable Microtome Blades, Low Profile	Fisher Scientific	Catalog #3052835
DNA LoBind Tubes, 1.5mL	Eppendorf	Catalog #022431021
TempAssure PCR 8-tube strip	USA Scientific	Catalog #1402-4700
MicroAmp Fast Optical 48-well reaction plate	Thermo Fisher Scientific	Catalog #4375816
48-well Optical Adhesive Film	Thermo Fisher Scientific	Catalog #4375323

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Coplin jars	VWR	Catalog #100500-232
Coverslips	Fisher Scientific	Catalog #12-544-EP
2100 Bioanalyzer	Agilent	Catalog # G2939BA
10x Genomics Accessories (Thermocycler Adaptor, Visium Spatial Imaging Test Slide, 10x Magnetic Separator, Slide Alignment Tool)	10x Genomics	Catalog #1000194
C1000 Touch Thermal Cycler	Bio-Rad	Catalog #1851197
Veriti 96-Well Thermal Cycler	Thermo Fisher Scientific	Catalog #4375786
NanoZoomer-XR	Hamamatsu	Catalog #L12225-01
Deposited data and code		
Processed PanI spatial transcriptomics	This paper	GEO: GSE254829
High resolution images for CODA analysis	This paper	Zenodo: doi: https://doi.org/10.5281/zenodo.11243954
Imaging mass cytometry data	This paper	Zenodo: doi: https://doi.org/10.5281/zenodo.11243954
Code for spatial transcriptomics	This paper	Zenodo: doi: https://doi.org/10.5281/zenodo.11478317
Code for CODA analysis	This paper	Zenodo: doi: https://doi.org/10.5281/zenodo.11477585

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by lead contact, Luciane Kagohara (ltsukam1@jhmi.edu).

Materials availability

This study did not generate new materials.

Data and code availability

- **Data**: There are restrictions to the availability of sequencing data. This is a retrospective cohort, and it is not possible to consent these patients with historic samples, particularly those with highly aggressive and rapidly lethal disease. As such, the IRB has requested that we do not publicly share the raw sequencing data from each patient. The data is securely stored within a Johns Hopkins University patient data system. The sequencing data reported in this paper will be shared by the lead contact (Dr. Luciane T. Kagohara – ltsukam1@jhmi.edu) upon request. The data is only available through collaboration following approval of the lead contact and Johns Hopkins University IRB. The processed data from spatial transcriptomics experiments (Visium and Xenium) are deposited in the Gene Expression Omnibus (GEO) (GEO: GSE254829). The high resolution images used for CODA machine learning cell type annotations and the IMC data are deposited in Zenodo (Zenodo: doi: https://doi.org/10.5281/zenodo.11243954).
- **Code**: The code for the spatial transcriptomis analyses are available at Zenodo https://zenodo.org/records/11478318 (Zenodo: doi: https://doi.org/10.5281/zenodo.11478317) and for CODA analysis at https://zenodo.org/records/11477585 (Zenodo: doi: https://doi.org/10.5281/zenodo.11477585).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

FFPE pancreatic ductal adenocarcinoma (PDAC) surgical specimens collected from 2016 to 2020 were examined by experienced pathologists (KF, JWL, ET and LWD) and PanINs present in the specimens were marked and classified as low- and high-grade by experienced pathologists. Only PanIN lesions with a unanimous diagnosis and grading were included in the study. The samples were obtained from the Johns Hopkins University School of Medicine Department of Pathology archives under Institutional Review Board approval (IRB00274690) under a waiver of consent. Samples were distributed into two cohorts: test (PanIN-LG1, PanIN-HG1, PanIN-LG2, PanIN-HG2, PanIN-LG3, PanIN-HG3, PanIN-HG4) and validation (PanIN-R-LG1, PanIN-R-LG2, PanIN-R-LG3, PanIN-R-LG4, PanIN-R-LG5, PanIN-R-HG1, PanIN-R-LG6). For ST analysis, new sections were prepared onto the ST slides and stained

prior to sequencing library preparations. The test cohort sections were stained with hematoxylin and the validation cohort with hematoxylin and eosin (H&E). Pathology revisions of the test cohort were performed by KF, LDW and JWL; while the revisions for the validation cohort were done by ET, LDW and JWL.

METHOD DETAILS

RNA quality control

All samples selected for the study had their RNA quality checked prior to the ST slides preparation. Total RNA was isolated from 20um sections of each sample using the RNase FFPE kit (Qiagen), following manufacturer's instructions. RNA quality was measured using the DV200 assay on the Bioanalyzer (Agilent) to determine the proportion of fragments with ~200bp in the sample. RNA quality was considered good if DV200 > 50%.

Spatial transcriptomics slide preparation

The ST data was generated using the commercial platform Visium FFPE (10x Genomics). The slides are designed to accommodate a total of 4 sections with a maximum size of 6.5 x 6.5 mm. We used a manual method for minimal manipulation of the FFPE blocks from clinical specimens to fit the 6.5x6.5 mm dedicated areas of the ST slides (Visium, 10x Genomics). Briefly, for the specimens that were larger than the designated regions of the Visium slides, we gently scored the surface of FFPE clinical blocks to isolate the area containing PanINs for profiling in the limited area of the Visium slide (Figure 1A). We scored the selected sample area containing the PanIN using skin punches of 5mm in diameter. The skin punches were used directly on the FFPE blocks to delimit the area of interest, so when the block was sectioned in the microtome the PanIN containing region was detached from the rest of the section and could then be placed in the ST capture area of the slides (Figure 1A). A 5 μm section from each sample with 5mm in diameter was used for the ST analysis. Upon preparation, the slides were incubated at 42°C and then stored in a desiccator until use.

Spatial transcriptomics data generation

Using the Visium FFPE (10x Genomics) platform and following manufacturer's validated protocol the samples were deparaffinized, stained with hematoxylin (discovery cohort) or H&E (validation cohort), and scanned using the Nanozoomer scanner (Hamamatsu) at 40x magnification. Human probe hybridization was performed overnight at 50°C. Following probe ligation, the RNA was digested, and the tissue was permeabilized for the release, capture, and extension of the probes. The designated area for each sample is covered by probes containing oligo-d(T) that capture the probes by a poly-A tail sequence present in the probe sequence. The sequencing library preparations were performed as instructed by the manufacturer using the extended probes as the template. All libraries were sequenced with a depth of at least 50,000 reads per spot (minimum of ~250 millions per sample) at the NovaSeq (Illumina). The Visium Human Transcriptome Probe Set v1.0 contains probes to 19,144 genes and after computational preprocessing (filtering for probes off-target activity) provides gene expression information for 17,943 genes.

Cell type annotation using transfer learning from stained imaging

Seven microanatomical components of human pancreas tissue were multi-labelled using a semantic segmentation workflow. The seven components recognized were (1) islets of Langerhans, (2) normal ductal epithelium, (3) vasculature, (4) fat, (5) acinar tissue, (6) collagen, and (7) pancreatic intraepithelial neoplasia (PanIN). Briefly, fifty examples of each tissue type were manually annotated using Aperio ImageScope. Half of the newly generated annotations were used in the training dataset for the convolutional neural network and the other half were used as an independent testing dataset to evaluate model performance. The testing dataset revealed an overall accuracy of 94.0% in classification of tissues in the TMAs. Following training, the tissue images were segmented to to tiles of 1μm each.

Nuclear coordinates were generated via the detection of two-dimensional hematoxylin or H&E intensity peaks. Briefly, the TMA images were down sampled to 1 μm/pixel resolution. To adapt CODA to the hematoxylin only stained images (test cohort), the color image was converted to greyscale. No changes were necessary for the H&E stained sections (validation cohort). The image was smoothed using a Gaussian filter and two-dimensional intensity peaks with minimum radii of 2μm were identified as nuclear coordinates.

Registration of ST data with cell type annotations

The low-resolution image used for the Visium pre-processing with Space Ranger was registered to the high-resolution tissue image used for microanatomical measurements to integrate the two workflows. The registration utilized the fiducial markers present on the ST glass slide to estimate the registration scale factor and translation. As registration was performed on two scans of identical tissue sections, it was assumed that rotation was not necessary. Here, the low-resolution image was registered to the high-resolution image (rather than the other way round) so that the scale factor was always greater than 1 and ensuring that the 1 μm resolution of the tissue micro annotations was preserved. First, the fiducial markers in each pair of images were segmented by identification of small, nonwhite objects surrounding the larger TMAs. Nonwhite objects were determined to be pixels with red-green-blue standard deviations greater than 6 in 8-bit space. These objects were morphologically closed and very small noise (<50 pixels) were removed. The fiducial markers were then determined to be objects in the image within 20% of the median object size (as many fiducial markers existed for each corresponding tissue image). This process resulted in fiducial image masks for the high-resolution and

low-resolution tissue images. With these masks, four possible registrations were calculated to account for the situation where the Visium analysis was performed on the tissue image rotated at a 0-, 90-, 180-, or 270-degree angle. For each registration, the corner fiducial markers of the low-resolution image were rescaled and translated to minimize the Euclidean distance to the fiducial markers of the high-resolution image. Of the four registration results, the registration resulting in the greatest Jaccard coefficient between the high-resolution and low-resolution fiducial masks was chosen. For the eight TMAs, the average Jaccard coefficient of the fiducial masks was 0.94.

The registration information used to overlay the low-resolution tissue image to the high-resolution tissue image was used to convert the coordinates corresponding to the location of the Visium assessment in the low-resolution image into the high-resolution images coordinate system. Once the Visium coordinates were registered to the high-resolution image, the generated tissue microanatomy composition and cellularity were calculated for regions within 25μm of each coordinate. For each Visium coordinate, pixels in the micro-anatomically labelled mask image within 25μm of that coordinate were extracted. Tissue composition was determined by analyzing the % of each classified tissue type within that dot. The cellularity of each dot was determined by counting the number of nuclear coordinates within 25μm of each Visium coordinate. Cellular identity was estimated by determining the microanatomical label at each coordinate where a nucleus was detected (a nucleus detected in the same pixel where the semantic segmentation model detected normal ductal epithelium was labelled an epithelial cell).

Spatial transcriptomics data analysis of PanIN samples

Sequencing data was processed using the Space Ranger software (10x Genomics) for demultiplexing and FASTQ conversion of barcodes and reads data, alignment of barcodes to the stained tissue image, and generation of read counts matrices. The processed sequencing data were inputs for the analyses using the Seurat software.[55–58] Data preprocessing with Seurat involved initial visualization of the counts onto the tissue image to discriminate technical variance from histological variance (e.g.: collagen enriched regions present lower cellularity that reflects in low counts). The filtered data was normalized using the SCTransform approach that uses a negative binomial method to preserve biological relevant changes while filtering out technical artifacts. Following normalization, data from all slides were merged and batch correction was performed with Harmony from harmony_0.1.0. Unsupervised clustering was subsequently performed on the harmony reduction using the Louvain algorithm as implemented by Seurat.[55–58]

Louvain clusters were annotated using the overlap of CODA annotations and quantifications per spot with well-characterized marker genes. Neoplastic and ductal epithelium groups were generated through selecting spots from the respective Louvain cluster that were estimated to be greater than or equal to 70% of the respective cell type on CODA. The data dimensionality was reduced using PCA for clustering and in tissue visualization of the transcriptional clusters. Unsupervised clustering was performed based on the most variable features (genes). Differential gene expression analysis of normal ducts and PanINs, and low and high grade lesions were performed using the MAST test[59] as implemented by Seurat. For comparisons performed across different slides, the slide was assigned as a latent variable and the matrix was prepared using *PrepSCTFindMarkers* to account for the multiple SCT models. Pathway analysis was performed using GSEA v4.2.1.[60,61] High- and low-grade PanIN spots were subset from the neoplastic Louvain cluster by pathologists (KF, JWL, LT and LWD) annotation using a custom-made Shiny app derived from the *SpatialDimPlot* function in Seurat. Violin plots, spatial plots, were generated in Seurat. Volcano plots were generated in ggplot2.[62] Heatmaps were generated using ComplexHeatmap.[63]

High-dimensional RNA in situ hybridization (Xenium, 10x Genomics)

The high-dimensional RNA in situ hybridization was performed at 10x Genomics facilities following manufacturer's instructions. Xenium was performed on 3 samples from the paired cohort (PanIN-HG1, PanIN-HG2 and PanIN-HG3) with PanIN lesions available in same area profiled with Visium. The sections were placed on the Xenium slides and deparaffinized and decrosslinked for optimal probe hybridization. The probes that hybridized to transcripts in the sample were ligated, amplified and conjugated to fluorescent probes in the Xenium Analyzer. The fluorescence captured by the device was preprocessed for visualization using the Xenium Explorer software. Finally, the sections were stained with hematoxylin and eosin and scanned. The data was analyzed using Seurat, following similar steps as described above for the Visium analysis. Briefly, the outputs from Xenium analyzer served as input for Seurat. Briefly, after Xenium data quality check (visualization of transcripts detected distribution in the tissue), the data was normalized (SCTransform) and we performed unsupervised clustering. Each sample was analyzed individually. Individual cells were determined by the Xenium Explorer cell segmentation algorithm. Xenium cell segmentation is an automated process that uses neural network for nuclear segmentation. The DAPI staining signal captured during the imaging is used to generate a segmentation mask. The detected nuclear boundaries are expanded by 15um in all directions and cell boundaries and transcripts are assigned to that detected cell. The detected cells were classified based on known cell type markers included in the panel. To annotate CAF subtypes, a module score was applied for pan-CAF marker genes (*FAP, LUM, DCN, COL1A1*). The distribution of module scores among all cells was modeled as a mixture of 3 gaussian distributions using mixtools v2.0.0. Cells were annotated as CAFs if they had a CAF module score greater than the threshold set at one standard deviation below the mean of the third component gaussian distribution (threshold value: 0.362) and no expression of *PTPRC* (CD45). Among cells annotated as CAFs, cells were annotated as CAF subtypes using module scores for each type (apCAF: *CD74, HLA-DRA, HLA-DPA1, HLA-DQA1, SLPI*; iCAF: *CXCL1, CXCL2, CCL2, LMNA, HAS1, HAS2*; myCAF: *TAGLN, MYL9, TPM2, MMP11, HOPX, TWIST1, SOX4*). Cells were annotated as a CAF subtype based on the highest CAF subtype module score greater than 0. Cells with no module scores above 0 were typed as general "CAF". CD4

T cells were annotated by gating of cells not annotated as epithelial, PanIN, or CAFs for concurrent non-zero expression of *PTPRC, CD4*, and at least one gene encoding CD3 proteins (*CD3D, CD3E*, or *CD3G*).

The Xenium gene panel included a total of 380 genes. This panel was outlined using a commercial panel of 280 genes designed for breast cancer. The large majority of the genes are markers for cell types found in different cancer types (epithelial, immune, stromal and endothelial cells). Then, we customized additional 100 genes that were selected from our Visium analysis and include highly expressed genes in PanIN, and genes from Pattern 2 and 7 (Table S1).

Imaging Mass Cytometry Data Analysis

Immunohistochemical staining was performed with mass cytometry antibodies. The TMA slides were first baked at 60°C for 2 hours, dewaxed in xylene, then rehydrated in an alcohol gradient. The slides were incubated in Antigen Retrieval Agent pH 9 (Agilent® S2367) at 96°C for 30 minutes then blocked with 3% BSA in PBS in RT for 45 minutes. The antibody cocktail listed in Table S2 was prepared at optimized dilutions and used to stain the slides at 4°C overnight. All custom antibodies were prepared to a concentration of 0.25 to 0.5mg/mL and were titrated empirically. Cell-ID™ Intercalator-Ir (Standard Biotools PN 201192A) was used for DNA labelling and Ruthenium tetroxide 0.5% Aqueous Solution (Electron Microscopy Sciences PN 20700-05) was used as counterstain. Images were acquired using the Hyperion Imaging System (Standard BioTools) at the Johns Hopkins Mass Cytometry Facility. Upon image acquisition, representative images were visualized and generated through MCD™ Viewer (Standard BioTools).

Images were acquired with a Hyperion Imaging System (Standard BioTools) at the Johns Hopkins Mass Cytometry Facility. Through MCD Viewer™ (Standard BioTools), multi-layered ome.tiff image stacks were generated and loaded in HALO 3.6. With HALO 3.6, the Area Quantification FL v2.3.4 algorithm was optimized visually and manually thresholded to quantify the positive area of IMC markers Smooth Muscle Actin (SMA), Vimentin (VIM), Collagen (COL), Podoplanin (PDPN), CD74, and HLA-DR. To mark all CAFs, a combination of SMA+VIM+COL+PDPN and DNA was used. The Area Quantification FL v2.3.4 algorithm was also utilized to subset CAFs into phenotypes positive for CD74, HLA-DR, and CD74+HLADR+. To calculate the density of CAF phenotypes, ruthenium counterstain was quantified for tissue area normalization and DNA was quantified for nuclear area normalization. CAF phenotypes were also obtained as percentages over the total CAF population. The resulting data was visualized using GraphPad Prism v10.1.2.

Transfer learning to relate ST data from PanIN to a scRNA-seq atlas of Pancreatic Ductal Adenocarcinoma

We obtained scRNA-seq data for pancreatic epithelial cells from an atlas of 29 tumor samples and 14 non-cancerous samples collated from Peng et al. and Steele et al. as described in Guinn et al.[15] We inferred cellular phenotypes in the epithelial cells using CoGAPS (R, version 3.5.8)[39,40] to infer 8 patterns on the log transformed expression values. Pattern annotation was based on overrepresentation analysis of patternMarker genes identified by CoGAPS (R, version 3.9.5)[64] and Molecular Signatures Database Hallmark gene sets (version 7.5.1)[65] using the R package fgsea (version 1.18.0).[60] *TFF1* expression was measured as log-normalized counts. Uniform manifold approximation and projection (UMAP) plots were made using monocle3 (version 1.0.0).[66–71] UMAP plots for epithelial cells from the scRNA-seq PDAC data were made with cells colored by epithelial cell type, log normalized *TFF1* expression, and Pattern 2, 5, 7 weights.

PanIN ST data was subset to spots annotated as epithelial by CODA (N = 240 spots; normal = 93, low-grade = 48, high-grade = 99). CoGAPS patterns learned from normal and tumor cells in the PDAC scRNA-seq data were projected onto scaled SCT expression values from epithelial ST spots using ProjectR (version 1.8.0).[16,17] Projected pattern weights were plotted as violin plots using Seurat (version 4.1.0). Mean pattern weights were compared across epithelial lesion grades using Wilcoxon rank-sum tests within ggpubr (version 0.4.0). UMAP plots of ST spots and overlayed plots of ST spots colored by epithelial type, log normalized *TFF1* expression, and projected Pattern 2, 5, 7 weights over tissue slices were prepared using Seurat (version 4.1.0). Conversely, CoGAPS patterns learned from the scaled SCT expression values of PanIN ST data were plotted as violin plots using Seurat (version 4.1.0). Pattern weights were compared across epithelial lesion grades using Wilcoxon rank-sum tests within ggpubr (version 0.4.0). Patterns were projected onto the scRNA-Seq atlas using ProjectR (version 1.8.0).