Spatial transcriptomics of FFPE pancreatic intraepithelial neoplasias reveals cellular and molecular alterations of progression to pancreatic ductal carcinoma

Alexander T.F. Bell‡[1], Jacob T. Mitchell‡[1-4], Ashley L. Kiemen‡[5], Kohei Fujikura[6], Helen Fedor[1], Bonnie Gambichler[1], Atul Deshpande[1-3], Pei-Hsun Wu[5], Dimitri N. Sidiropoulos[1-3,7], Rossin Erbe[1-4], Jacob Stern[8], Rena Chan[8], Stephen Williams[8], James M. Chell[8], Jacquelyn W. Zimmerman[1-3,12], Denis Wirtz[5,9,10], Elizabeth M. Jaffee[1-3,12], Laura D. Wood[6,12], Elana J. Fertig*[1-3,11,12], Luciane T. Kagohara*[1-3,12]

*Contributed equally

‡ Contributed equally

[1]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD
[2]Convergence Institute, Johns Hopkins University School of Medicine, Baltimore, MD
[3]Bloomberg Kimmel Immunology Institute, Johns Hopkins University School of Medicine, Baltimore, MD
[4]Department of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD
[5]Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, Baltimore, MD
[6]Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD
[7]Cellular and Molecular Medicine, Johns Hopkins University School of Medicine, Baltimore, MD
[8]10x Genomics, Pleasanton, CA
[9]Department of Materials Science and Engineering, The Johns Hopkins University, Baltimore, MD
[10]Johns Hopkins Physical Sciences - Oncology Center, The Johns Hopkins University, Baltimore, MD
[11]Department of Applied Mathematics and Statistics, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD
[12]The Skip Viragh Center for Clinical and Translational Research

Corresponding authors:

Elana J. Fertig, PhD
550 N Broadway, Suite 1101
Baltimore, MD 21205
Phone: 410-955-4268
Email: ejfertig@jhmi.edu

Luciane T. Kagohara, PhD
1650 Orleans St., Room 485
Baltimore, MD 21287
Phone: 410-955-7315
Email: ltsukam1@jhmi.edu

Running title: FFPE Spatial Transcriptomics Analysis of human PanINs

Keywords: pancreatic intraepithelial neoplasia (PanIN), spatial transcriptomics, formalin-fixed paraffin embedded (FFPE), transfer learning, pancreatic cancer (PDAC)

Author's Disclosures: 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. No disclosures were reported by the other authors.

ABSTRACT

Spatial transcriptomics (ST) is a powerful approach for cancers molecular and cellular characterization. Pancreatic intraepithelial neoplasia (PanIN) is a pancreatic ductal adenocarcinoma (PDAC) premalignancy diagnosed from formalin-fixed and paraffin-embedded (FFPE) specimens limiting single-cell based investigations. We developed a new FFPE ST analysis protocol for PanINs complemented with novel transfer learning approaches. The first transfer learning approach, to assign cell types to ST spots and integrate the transcriptional signatures, shows that PanINs are surrounded by PDAC cancer associated fibroblasts (CAFs) subtypes, including the rare antigen-presenting CAFs. Furthermore, most PanINs are of the classical PDAC subtype while one sample expresses cancer stem cell markers. A second transfer learning approach, to integrate ST PanIN data with PDAC scRNA-seq data, identifies a shift between inflammatory and proliferative signaling as PanINs progress to PDAC. Our data support a model of inflammatory signaling and PanIN-CAF interactions promoting premalignancy progression and PDAC immunosuppressive characteristics.

Significance

We developed a novel FFPE spatial transcriptomics analysis pipeline to profile the heterogeneous CAF and malignant epithelial cells in PanINs, premalignant lesions that can progress to invasive PDAC. This study identifies for the first time similar CAF populations residing in PanINs and associated molecular changes that together may be early mediators of premalignant transformation to PDAC.

INTRODUCTION

Spatial molecular technologies can drive pathway discoveries in cancers and their tumor microenvironment (TME) while preserving tissue architecture, thus allowing the characterization of molecular changes resulting from cell-to-cell direct interactions(1). Until recently, these approaches were limited to fresh frozen sample profiling that maintain RNA integrity(2–6). However, the major source of biopsy and surgical tissues from current and retrospective studies are samples stored in paraffin blocks. Moreover, the diagnosis of preinvasive cancer lesions, such as pancreatic intraepithelial neoplasias (PanINs), is almost always limited to formalin-fixed and paraffin embedded (FFPE) tissue. Thus, the recent development of a ST technology that can utilize FFPE samples will provide untapped opportunities to apply high dimensional approaches to evaluate archived FFPE specimens from prior clinical trials as well as standard of care diagnostic samples(7). Nevertheless, a challenge for section preparation exists as the platform limits the size of the sections that can be analyzed. Since most samples are fixed and embedded in larger sizes, methods to prepare smaller sections from FFPE blocks must be developed with the aim of preserving limited valuable clinical samples.

ST approaches have already identified transcriptional signatures associated with spatial interactions that are delineating the cellular phenotypes that underlie tumor biology, evolution, and responses to therapy(1). Even without spatial characterization, complementary single-cell RNA-sequencing (scRNA-seq) technologies have enabled unprecedented characterization of the molecular and cellular pathways in the PDAC TME and of their hypothesized role in cancer progression(8–12). Nevertheless, current scRNA-seq approaches dismantle the tissue architecture and crucial intercellular interactions are

lost. On the other hand, determining complementary inference from ST data is limited by the current lack of single-cell resolution that could be solved by performing integrated single-cell analysis and ST on the same sample(13). Therefore, the identification of the cellular and molecular changes in PDAC progression would ideally be achieved through matched scRNA-seq and ST profiling of both PanINs and PDAC but scRNA-seq approaches are also currently incompatible with FFPE samples because these technologies are dependent on cell viability, making them inaccessible for PanINs profiling. In ST, the possibility of staining the sections prior to library preparation allows pathological examination of cell morphologies to enhance cellular annotations. Nevertheless, manual integration of histologic annotations with ST requires two images, one with the pathology annotations and another with ST mapped clusters, that are compared side-by-side. This process is subjective and based on approximate visual selection thus prone to error and biases. New computational methods that incorporate imaging data are emerging to augment clustering of ST data for cell type annotation, but disregard the functional information represented by the cellular morphology(14,15). The matched stained FFPE image from ST offers the unique opportunity to integrate gene expression data with morphological features associated with cellular phenotypes through machine learning.

We recently developed a machine learning method that provides 3-dimensional (3D) pathologic tissue assessments called CODA, to identify and quantify normal and PanIN cells in the pancreas from hematoxylin and eosin (H&E) stained FFPE sections(16). Here, we report the successful integration of CODA with FFPE ST data through transfer learning of the cellular annotations from CODA to the stained images in Visium to enhance

pancreas cell type annotation by spatial spots and robustly examine cell-type specific molecular phenotypes in PanIN lesions. To further expand our findings and further delineate the transitional pathways that drive normal cell to PanIN to PDAC development, we applied a second transfer learning approach, projectR(17), to integrate CODA and ST data with scRNA-seq data from prior reported studies that have been entered into a novel PDAC atlas (18). Applying transfer learning to multiple types of data provides the opportunity to create a model of PDAC initiation and progression to advanced disease.

In this study, we build up from a recently developed FFPE ST approach(7) with our transfer learning approaches(16,17,19) to examine the transcriptional alterations across the stages of PanIN progression. We developed a novel sample preparation workflow to ensure minimal manipulation of FFPE blocks for ST slide preparation. Subsequently, we adapted our transfer learning approaches used for 3D imaging to evaluate the combined ST data, and successfully identified the cellular components in each sample, including normal and premalignant epithelial cells. We discovered that all except one of the premalignant PanIN samples share similar expression patterns as the classical PDAC subtype. The one PanIN sample that does not express this classical signature expresses cancer stem cell (CSC) markers, suggesting that cells with stemness transcriptional features are present at premalignant stages. Moreover, we observed in PanINs the presence of the same cancer associated fibroblast (CAF) subtypes that are enriched in invasive PDACs, providing new evidence that these cells are already modulating the premalignant developing microenvironment. A common feature to all classical PanINs is the high expression of *TFF1* that gradually increases through premalignancy development. Transfer learning integrating PanIN ST with PDAC scRNA-seq data verified

that *TFF1* up-regulation persists in invasive cancer epithelial cells. We also found that the inflammatory pattern characterized in our PDAC atlas of scRNA-seq data gradually fades with PDAC invasion. The loss of inflammation signals is associated with a concurrent increase in MYC pathway activation in mature PDAC. Overall, this integrated experimental and computational approach provides the means to develop a model of PDAC development and progression through integrated imaging, ST, and scRNA-seq datasets.

RESULTS

Spatial transcriptomics applied to FFPE specimens captures preneoplastic pancreatic tissue architecture

To study the mechanisms of progression from pre-malignant early PanINs to PDAC, we applied ST to a cohort of 4 patients with paired low- (LG) and high-grade (HG) PanINs (total number of lesions = 8). This cohort was designed to enable comparisons of progressive mechanisms within and between patients. The ST slide's dedicated areas for analysis are small (6x6 mm), and the FFPE preserved samples are typically larger and would require coring or scraping of the block to isolate only the PanIN lesion. Extensive manipulation of FFPE blocks for the ST sample preparation would limit subsequent profiling of larger regions with other technologies and additional clinical diagnostics of the broader PDAC surgical specimen in which PanINs are detected. Therefore, we developed a method to score the surface of the FFPE blocks using 5 mm in diameter circular skin biopsy punches. Prior to sectioning, the punches were used to

score the regions of interest (ROI) containing the ~1mm PanIN lesion while preserving the block. The scoring then allowed the non-relevant tissue to detach from the ROI that was collected and placed on the designated area of the ST slide (**Figure 1A**). Initial total RNA quality check indicated that all samples presented some level of RNA degradation (RIN ~2) but with a high concentration of 200bp fragments (DV200 $\geq$ 50%) compatible with the FFPE ST platform. Following ST data pre-processing and filtering, seven out of the eight samples presented high quality data for subsequent downstream analysis. We were able to detect an average of 71,695 reads and 2,537 genes per spatial spot, and an average of 16,266 genes per sample.

The ST data from our PanIN cohort provides combined image and transcriptomics profiling (**Figures 1B and 1C**). We characterized the canonical cellular distribution of PanINs and surrounding pancreas tissue by first applying clustering to the ST profiling data alone. The normal pancreas is composed of multiple cell subtypes with different functions. To execute its most important functions, the pancreas is composed of exocrine cells (acinar cells) that are responsible for the production of digestive enzymes and by endocrine cells (islets of Langerhans) that produce insulin and other hormones. The excretion of enzymes occurs through pancreatic ducts, while insulin and hormones are directly released into the blood stream. PanINs and PDACs differentiation resembles the morphology of normal pancreatic ducts(20). Annotating marker genes differentially expressed in each cluster learned from the transcriptional signal infers these canonical and transformed cell types (**Figure 1C, Supplemental Figure 1)**. Moreover, the distribution of the normal and neoplastic cell types identified from clustering matches their locations in the corresponding section image (**Figures 1B and 1C**). In all samples, it was

possible to map specific clusters to the PanINs that were distinct from the other clusters including the normal ducts. Nevertheless, in some regions the clusters extend beyond cellular boundaries into the adjacent cell types on H&E imaging (**Supplemental Figure 2**). This extended signal could arise either from the intercellular signaling extending the molecular changes beyond cellular margins or from technical artifacts in the ST technology.

To improve cluster annotations based on tissue morphology, we applied a transfer learning method, named CODA, to accurately classify the pancreatic normal and neoplastic cells directly in the ST section imaging associated with the ST data. CODA is a 3D imaging-based approach that uses deep learning semantic segmentation to identify different cell types within the human pancreas (acinar cells, islets of Langerhans, fibroblasts, adipocytes, endothelial cells, ductal cells, neoplastic cells)(16). We adapted CODA to obtain a color-coded image with each color corresponding to a specific cell type from the stained tissue sections (**Figure 1D**). In contrast to the clustering annotations, cells annotated in CODA are at single-cell level and limited to cellular boundaries. Thus, we could apply this approach to estimate the true proportion of cells within each spot (**Supplemental Figure 3**) associated with each spatial barcode in the ST data for further downstream analysis. By determining cell proportions in each ST spot we were able to select those representing a unique cell type, while filtering out spots capturing multiple cell types, to avoid unwanted bias in the comparisons between normal and PanIN clusters, for example.

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

The ability of ST to characterize the PanINs and their surrounding microenvironment provides the unique opportunity to examine the fibroblast population adjacent to these lesions. The PDAC TME is enriched with a heterogeneous population of cancer associated fibroblasts (CAFs). They have been classified into three subtypes based on RNA expression data and include: myofibroblastic CAFs (myCAFs), inflammatory CAFs (iCAFs), and antigen-presenting CAFs (apCAFs)(21–23). These populations of mesenchymal cells play dual roles and can induce or inhibit PDAC progression(21–24). CAFs exert a tumorigenic role by providing metabolites for tumor cell survival, stimulating cell growth pathways through paracrine signaling, and creating an immunosuppressive microenvironment(22). However, a tumor suppressing CAF enriched TME can reduce essential nutrients required for tumor progression and differentiation, and the same CAFs can be functionally repolarized to release chemokines that will recruit immune cells into the tumor(22).

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, in human premalignant lesions their presence is not well described. Here, we used well established markers(23) to map the distribution of myCAFs, iCAFs and apCAFs adjacent to human PanINs. In our cohort, the density of stromal cells varied but were observed adjacent to each premalignant lesion (Figure 1D, pink annotated regions). Since CODA annotates cells based on morphology, it does not have the ability to classify the CAF subtypes as the classification is based on

their transcriptional profile. Instead, additional integration with the transcriptional profiles from the ST data was used to identify these cell types. This approach showed that a CAF common signature (pan-CAF) is consistently expressed across the collagen rich regions annotated by CODA (**Figure 2A**, orange and red spots). The expression of myCAF (**Figure 2B**) and iCAF (**Figure 2C**) markers were detected in all samples overlapping with the regions where pan-CAFs are present.

The presence of a recently described subtype of apCAFs was also investigated. ApCAFs were first identified by scRNA-seq on a PDAC mouse model. Further characterization showed that these cells express MHC-II genes and can present antigens to $CD4^+$ T cells in vitro(23). Subsequently, apCAFS have been shown to present antigen to Tregs which activates their suppressive capability(27). In our study, expression of the apCAF signature was detected in all samples (**Figure 2D**). As expected, a significant proportion of the apCAF positive spots also express $CD45^+$, a marker of immune cells (**Figure 2E**) and that can express the same MHC-II genes. Since CODA cannot annotate immune cells due to their limited size and scant cytoplasm, and ST does not provide single-cell resolution, the confirmation of apCAFs in some samples could not be exactly defined by our analysis. Nevertheless, in some regions of the stroma the expression of apCAF markers does not colocalize with the CD45, indicating that these mesenchymal cells are present in human PanINs.

ST identifies expression of both PDAC classical subtype and cancer stem cell signatures in PanINs

CAFs are mediators of PDAC progression and aggressiveness through interactions with neoplastic cells(22,23,28). The detection of PDAC associated fibroblast subtypes in PanINs prior to establishment of invasive carcinoma suggests that the differentiation of fibroblasts into CAFs is an early event that may influence PanIN progression to PDACs. To test this, we leveraged the automated cell type annotation from CODA with cluster-based annotations to select spots purely associated with ductal cells to compare normal and PanIN ducts. Next, we characterized PanIN cell heterogeneity relative to the established classical and basal-like PDAC subtypes(29). We demonstrate that six out of seven PanINs express the PDAC classical subtype signature (**Figure 3A**). The basal-like signature is not detected in any of the premalignant lesions (**Figure 3B**). This observation supports the hypothesis that PDACs arise with a classical phenotype and likely acquires the basal-like phenotype upon progression and accumulation of molecular aberrations(30).

Only one HG PanIN sample (PanIN-HG3, Figure 1) expressed neither the classical nor the basal-like signatures. Thus, we hypothesized that this sample expresses a third transcriptional phenotype. PDAC progression, resistance to therapies, and immune evasion is in part associated with the presence of populations of cells expressing cancer stem cell (CSC) markers(31). We verified the expression of CSC markers among the PanINs in the cohort. Remarkably, the only sample with significantly high expression of CSC markers is the one that did not express the classical or the basal-like PDAC signatures (**Figure 3C**). The presence of cells expressing CSC markers in PanINs was previously described in a mouse model that mimics PDAC development(32) and in human samples(33), but little is known about the mechanisms leading to CSC genes up-

regulation their role in PDAC progression. Our observation that this stemness signature is not observed in cells expressing the classical subtype suggests that neoplastic cells with stemness features are a distinct population that arise in early pre-malignant stages. Nonetheless, further investigation is needed to understand the pathways driven by stemness and how these cells are interacting with the CAFs and other cells in the TME to modulate PDAC biology.

Differential expression analysis between PanINs and normal ducts identifies *TFF1* expression limited to the classical phenotype

To further define the molecular features of PanINs, spots from all samples annotated to normal and premalignant ducts were merged for each patient and 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 (**Figure 3D**) and their expression pattern discriminated PanINs from normal ducts among the different samples (**Figure 3E**). Among the top 20 up-regulated genes in the premalignant lesions, 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(34–36). The pathway analysis from the differentially expressed genes indicates enrichment for MYC and oxidative phosphorylation pathway mediators. Both signaling pathways have been previously shown to be upregulated in PanINs and PDAC, particulary in association with progression from premalignancy to invasive cancer, metastasis development, and

resistance to therapy(37–39) (**Supplemental Figure 4**). Although predominantly 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 up-regulated in PanINs and PDACs and has been suggested as a potential early detection marker found in bodily fluids(40). In in vitro cell culture models, the secreted form of TFF1 has been shown to increase PDAC and stellate cells motility without a significant impact on proliferation(41). Since stellate cells are considered one of the precursors to some PDAC CAF subtypes(42,43), it is possible that *TFF1* is one of the mediators of intercellular interactions between PanIN and PDAC cells, and CAFs. However, the sample expressing the CSC markers signature does not express *TFF1*, suggesting that the stemness signature and *TFF1* are mutually exclusive (**Supplemental Figure 5**).

TFF1 expression gradually increases from normal ducts through low to high grade PanIN lesions

The characterization of multiple ducts, including those spanning across stages of PanIN differentiation (mixed ducts), provides the opportunity 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. We therefore compared expression changes between lesions classified as LG or HG based on their morphology. The differentiation stages of PanINs cannot be discriminated using CODA and the classification of LG and HG

preneoplastic ducts was performed through pathology examination (KF and LDW) (**Figure 4A**). Using the pathological PanIN classification, we identified mixed ducts containing normal, LG and HG preneoplastic cells (**Figure 4B, 4C** and **4D**).

Next, we used this pathologic classification to evalute MYC pathway targets across PanIN stages in our cohort. Expression of the MYC signature was increased in LG and HG PanINs when compared with normal duct expression (**Supplemental Figure 6**), mimicking what had been previously published in invasive PDACs(37,38). These findings further support our hypothesis that MYC pathway activation is an early event in the malignant transformation process to invasive PDACs. The oxidative phosphorylation pathway was also significantly enriched in our comparison of PanINs and normal ducts. Oxidative phosphorylation is upregulated in PDACs and other tumors and is associated with resistance to chemotherapy(39,44). Here, we observed similar levels of expression of this pathway in both LG and HG lesions that were significantly higher than in normal ducts (**Supplemental Figure 6**).

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 (**Supplemental Figure 7**). In addition, the level of expression of *MUCL3* and *TSPAN1* genes gradually increased from normal ducts through LG and HG lesions (**Figure 4E** and **4F**). The same pattern was observed for *TFF1*. 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 4G**).

Changes in PanIN progression map to transitions in malignancy in PDAC

PanINs are premalignant lesions that can progress to PDAC; this is supported by the detection of common driver mutations in premalignancies that are frequent in invasive cancer(45–47). 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. To identify these alterations, we used a compilation of scRNA-seq PDAC datasets (PDAC Atlas)(18) to project on our PanIN ST data and mapped them according to grades of differentiation. The fine-scale assignment of cell types to ST spots by CODA allowed this further investigation through the projection of transcriptional latent spaces learned from the scRNA-seq expression data. The PDAC Atlas from six previously published scRNA-seq datasets includes gene expression profiles from 25,442 epithelial ductal cells (Cancer: 14,589 cells; Normal: 7,561 cells; Normal_Tumor_Adjacent: 2,375 cells, Unspecified: 917 cells) (**Supplemental Figure 8A**). These epithelial populations were analyzed using CoGAPS(48,49) that learnt eight transcriptional patterns associated with biological processes(18). To study the representation of the patterns learned in the PDAC atlas across PanIN stages, the patterns were projected onto epithelial spots from the ST data (N = 623 spots; normal = 254, LG = 110, HG = 159) using projectR(17,19).

Among the patterns projected from the atlas onto the ST data, Pattern 2 (**Figure 5A** and **Supplemental Figure 8B**) enriched with genes involved in KRAS signaling and estrogen

response, showed a marked increase in projected pattern weights from normal epithelium through LG and HG PanINs (**Figure 5B** and **Supplemental Figure 8C**), corroborating previously reported studies showing up-regulation of pancreatic oncogenic signaling pathways in premalignancy initiation and progression(50,51). Pattern 5 (**Figure 5C** and **Supplemental Figure 8D**), associated with normal metabolic activity, showed a substantial decrease in projected pattern weights with progression of PanIN lesion grade (**Figure 5D** and **Supplemental Figure 8E**), suggesting a decline of normal cellular function in the PanIN cells with the same trend that is present in the primary tumor cells from the PDAC Atlas. A third pattern, Pattern 7 (**Figure 5E** and **Supplemental Figure 8F**), representing an inflammatory state, 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 (**Figure 5F** and **Supplemental Figure 8G**).

Finally, using the PDAC atlas, we verified that PDAC cells express *TFF1* (**Figure 5G**) as do PanIN cells (**Figure 5H**), suggesting that its up-regulation is a marker of PanIN and invasive PDAC progression. *TFF1* expression is almost undetectable in normal ductal cells. Surprisingly, the normal ductal cells adjacent to the tumor express low levels of *TFF1*, suggesting that the transcriptionally normal surrounding ducts are already programmed toward a pre-malignant state. Unfortunately, it was not possible to discriminate PanIN from PDAC cells from the atlas dataset. Due to the dissociation process required for scRNA-seq that disrupts the tissue architecture, PanIN and PDAC cell classification based on their morphology and neoplastic duct formation is not feasible.

DISCUSSION

ST approaches are uncovering new molecular and intercellular interactions that are providing insights into how these complex signaling networks are modeling cancer development and progression(1). In this study, we applied a novel protocol developed for FFPE ST(7) to examine PanINs, premalignancies with potential to develop into invasive PDAC, and uncover the mechanisms of progression to advanced tumors that could provide new opportunities for therapeutic interventions. Besides the novel FFPE ST approach, we also innovated by using two transfer learning methods to overcome limitations of ST and scRNA-seq technologies. First, CODA(16) was used to accurately assign cell types to ST spots from the H&E imaging and identify transcriptional features that could be the result of intercellular interactions. Second, ProjectR(17,19), allowed the integration of scRNA-seq performed on invasive PDACs with ST data from PanINs to learn mechanisms associated with PDAC initiation and progression. The integration of state-of-the-art experimental and computational approaches allowed us to develop a model of molecular and cellular features of PanIN to PDAC development.

Using a technology that preserves tissue architecture combined with gene expression profiling, we observed for the first time the presence of CAFs and the different transcriptional subtypes (myCAF, iCAF and apCAF) in premalignant human lesions. These subtypes were only previously described in PDACs(21,23). 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(52). The presence of these regulatory cells

in the human pancreatic premalignant lesions is not well described, but suggests that CAF-induced TME remodeling is an early event with durable impact on PDAC development. Further studies are necessary to examine what are the specific interactions driven by the different CAF subtypes and how they modulate preinvasive neoplastic 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(29). 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 initially all PDACs develop from the classical phenotype and that during the tumorigenesis there is a diverging point in which some cells will differentiate into the basal-like phenotype, a phenotype that is usually expanded by chemotherapy as resistance develops(29,30). In our PanIN cohort, we detected the classical signature in six out of seven PanINs. This observation corroborates the hypothesis that PDAC initiates as the classical subtype that can differentiate into the basal-like sub-type. The only sample that could not be classified as classical or basal-like expressed a CSC signature. CSCs present a more aggressive behavior and their presence is associated with resistance to therapies, local recurrence, and development of metastasis(53–55). Using ST data, we were able to isolate PanIN ducts and demonstrate that cells expressing CSC features are present in PanINs. One hypothesis, based on these data, is that these cells are the result of interactions of PanINs with CAFs

that already reside in the preneoplastic microenvironment, thereby providing activation signals to stemness related pathways. Additional studies are needed to confirm these interactions.

The differential expression analysis shows that LG and HG PanINs are transcriptionally similar. Among the few genes differentially expressed between these two PanIN grades, the expression of three genes (*MUCL3*, *TSPAN1* and *TFF1*) gradually increased with PanIN progression. *TSPAN1* is a transmembrane protein involved in signal transduction and associated with cell growth, differentiation, and motility in PDACs, but it is role or expression in PanINs have been poorly explored(56,57). The up-regulation of mucin related genes is frequent in PanINs and PDACs and are thought to be involved in the progression of these lesions as they encode proteins that confer survival advantage, contribute to immune evasion, and limit drug uptake(58). *TFF1* is frequently up-regulated in PanIN and PDAC but little is known about its role in tumorigenesis. As mentioned previously, secreted TFF1 could be involved in tumor cell interactions with CAFs(35,41,59). The gradual increase in *TFF1* expression is the most remarkable among the overexpressed genes in our PanIN ST progression analysis. The previous report that high levels of secreted TFF1 can modulate CAF motility combined with our finding that CAFs are present in PanINs, suggests that these are early features of PDAC development. Thus, this interaction should be further investigated to pinpoint the relationship between *TFF1* expressing neoplastic cells and the adjacent CAFs as this could be a target for early PDAC intervention and screening.

Utilizing our novel two stage computational approach we identified three new patterns associated with the development and progression of PanINs and the progression to

invasive PDACs. Specifically, a second transfer learning method was required to uncover PDAC scRNA-seq patterns from the PanIN ST data. Adding to the learned oncogenic and a normal pancreatic function patterns (Patterns 5 and 2, respectively) in the PDAC atlas(18), our approach identified a third pattern enriched for inflammation markers (Pattern 7) that is downregulated in primary PDAC cells. This reduction in inflammatory signals may be one mechanism by which PDACs evade immune recognition, as tumor cells will be producing less immune chemoattractants. We observed that the reduction in this inflammation pattern is gradual with progression from normal ductal cells through LG to HG PanIN lesions. It is possible that downregulation of this pattern may be due to development of immune suppressive CAFs that are already present in PanINs and are functionally similar to those present in PDACs, as demonstrated by Kinny-Koster et al. using a co-culture patient derived organoid model(18). In addition, CAFs can create an immunosuppressive TME by producing and releasing cytokines that inhibits immune cell infiltration and factors that provide cancer cells additional growth and survival advantages(21,23).

Although we used a limited sample number, we were able to corroborate previous findings related to PanINs and discover new features and their potential role in the progression of these lesions to invasive PDAC. We successfully visualized the microenvironment in which PanINs are developing and showed for the first time the presence of CAFs with potential suppressive function in PanINs. Our cohort included samples with varying stromal and acinar cells composition, but due to the limited size we did not observe correlations between PanIN transcriptional profiles with the adjacent cell types. To examine if the CAFs surrounding the PanINs are remodeling the premalignant

microenvironment and influencing their progression, a larger cohort with a more stringent selection criteria that considers patients' clinical features and outcomes (e.g.: tumor stage, metastasis, response to therapies) to correlate with the TME composition would be better suited to unveil the critical features of CAF-PanIN (or PDAC) interactions. Nonetheless, we demonstrate that FFPE ST provides robust results that led to novel findings regarding PDAC initiation and development. We learned a PDAC gene pattern from scRNA-seq that is also detectable in PanINs. Overall, we present a workflow to integrate different datasets and that provides the tools to create a model of PanIN to PDAC progression even though the samples were prepared and analyzed by distinct high-dimensional transcriptomics approaches.

METHODS

Sample selection

FFPE pancreatic ductal adenocarcinoma (PDAC) surgical specimens collected from 2016 to 2018 were examined by experienced pathologists (KF and LDW). PanINs present in the specimens were marked and selected for ST analysis and were classified as low- and high-grade by experienced pathologists (L.D.W. and K.F.). 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.

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 Bionalyzer (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 x 6 mm. For the specimens that were larger than the designated regions of the Visium slides, 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 5μm 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, and scanned using the Nanozoomer scanner (Hamamatsu) at 40x magnification. Human probes hybridization was performed overnight at 50°C. Following probes ligation, the RNA was digested, and the tissue was permeabilized for probes release into the slides, capture,

and extension. The designated area for each sample is covered by probes containing oligo-d(T) that captures 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 H&E 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 a resolution of 1μm.

Nuclear coordinates were generated via detection of two-dimensional hematoxylin intensity peaks. Briefly, the TMA images were downsampled to a resolution of 1 μm/pixel. As the tissues contained only a hematoxylin signal, color deconvolution (generally used

to de-mix the hematoxylin channel from the hematoxylin & eosin image) was not necessary. Instead, the color image was converted to greyscale. 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(60–63). 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.

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(64) 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(65,66). High- and low-grade PanIN spots were subset from the neoplastic Louvain cluster by pathologist (LDW) 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(67). Heatmaps were generated using ComplexHeatmap(68).

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 Kinny-Koster et al.(18). We inferred cellular phenotypes in the epithelial cells using CoGAPS (R, version 3.5.8)(48,49) 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)(69) and Molecular Signatures Database Hallmark gene sets (version 7.5.1)(70,71) using the R package fgsea (version 1.18.0)(72). *TFF1* expression was measured as log-normalized counts. Uniform manifold approximation and projection (UMAP) plots were made using monocle3 (version 1.0.0)(73–79). UMAP plots for epithelial cells from the PDAC atlas 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 = 623 spots; normal = 254, low-grade = 110, high-grade = 159). CoGAPS patterns learned from the PDAC atlas were projected onto scaled SCT expression values from epithelial ST spots using ProjectR (version 1.8.0)(17,19). 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 over layed 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)(60).

Author's Disclosures

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. No disclosures were reported by the other authors.

Author's Contributions

A.T.F. Bell, J.T. Mitchell, A.L. Kiemen: data curation, formal analysis, investigation, visualization, writing original draft and review. *K. Fujikura*: data curation, writing and review. *H. Fedor, B. Gambichler, A. Deshpande, P.H. Wu, D.N. Sidiropoulos, R. Erbe, R. Chan, S. Williams, J. Chell, J. Zimmerman., D. Wirtz*: resources, writing review. *E.M. Jaffee*: resources, writing review and editing. *L.D. Wood*: data curation, resources, writing review and editing. *E.J. Fertig*: conceptualization, data curation, supervision, formal analysis, visualization, writing original draft, review and editing. *L.T. Kagohara*: conceptualization, data curation, supervision, formal analysis, funding acquisition, visualization, writing original draft, review and editing.

Acknowledgements

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. 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., 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.), and 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.W., 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.)

Data and code availability

Submission of spatial transcriptomics data to dbGAP and code to github are in process.

References

1. Davis-Marcisak EF, Deshpande A, Stein-O'Brien GL, Ho WJ, Laheru D, Jaffee EM, et al. From bench to bedside: Single-cell analysis for cancer immunotherapy. Cancer Cell. 2021;39:1062–80.
2. Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature. 2021;596:211–20.
3. Takei Y, Yun J, Zheng S, Ollikainen N, Pierson N, White J, et al. Integrated spatial genomics reveals global architecture of single nuclei. Nature. 2021;590:344–50.
4. Shah S, Lubeck E, Zhou W, Cai L. seqFISH Accurately Detects Transcripts in Single Cells and Reveals Robust Spatial Organization in the Hippocampus. Neuron. 2017;94:752–758.e1.
5. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science. 2019;363:1463–7.
6. Stickels RR, Murray E, Kumar P, Li J, Marshall JL, Di Bella DJ, et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. Nat Biotechnol. 2021;39:313–9.
7. Gracia Villacampa E, Larsson L, Mirzazadeh R, Kvastad L, Andersson A, Mollbrink A, et al. Genome-wide spatial expression profiling in formalin-fixed tissues. Cell Genomics. 2021;1:100065.
8. Peng J, Sun B-F, Chen C-Y, Zhou J-Y, Chen Y-S, Chen H, et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. Cell Res. 2019;29:725–38.
9. Steele NG, Carpenter ES, Kemp SB, Sirihorachai VR, The S, Delrosario L, et al. Multimodal mapping of the tumor and peripheral blood immune landscape in human pancreatic cancer. Nat Cancer. 2020;1:1097–112.
10. Lin W, Noel P, Borazanci EH, Lee J, Amini A, Han IW, et al. Single-cell transcriptome analysis of tumor and stromal compartments of pancreatic ductal adenocarcinoma primary tumors and metastatic lesions. Genome Med. 2020;12:80.
11. Bernard V, Semaan A, Huang J, San Lucas FA, Mulu FC, Stephens BM, et al. Single-Cell Transcriptomics of Pancreatic Cancer Precursors Demonstrates Epithelial and Microenvironmental Heterogeneity as an Early Event in Neoplastic Progression. Clin Cancer Res. 2019;25:2194–205.
12. Raghavan S, Winter PS, Navia AW, Williams HL, DenAdel A, Lowder KE, et al. Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. Cell. 2021;184:6119–6137.e26.
13. Moncada R, Barkley D, Wagner F, Chiodin M, Devlin JC, Baron M, et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. Nat Biotechnol. 2020;38:333–42.
14. Pham DT, Tan X, Xu J, Grice LF, Lam PY, Raghubar A, et al. stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. BioRxiv. 2020;
15. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat Methods. 2021;18:1342–51.

16. Kiemen AL, Braxton AM, Grahn MP, Han KS, Mahesh Babu J, Reichel R, et al. In situ characterization of the 3D microanatomy of the pancreas and pancreatic cancer at single cell resolution. BioRxiv. 2020;
17. Stein-O'Brien GL, Clark BS, Sherman T, Zibetti C, Hu Q, Sealfon R, et al. Decomposing Cell Identity for Transfer Learning across Cellular Measurements, Platforms, Tissues, and Species. Cell Syst. 2019;8:395–411.e8.
18. Kinny-Koster B, Guinn S, Tandurella JA, Mitchell JT, Sidiropoulos DN, Loth M, et al. 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. BioRxiv. 2022;
19. Sharma G, Colantuoni C, Goff LA, Fertig EJ, Stein-O'Brien G. projectR: an R/Bioconductor package for transfer learning via PCA, NMF, correlation and clustering. Bioinformatics. 2020;36:3592–3.
20. Hruban RH, Goggins M, Parsons J, Kern SE. Progression model for pancreatic cancer. Clin Cancer Res. 2000;6:2969–72.
21. Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvise M, et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. J Exp Med. 2017;214:579–96.
22. Helms E, Onate MK, Sherman MH. Fibroblast heterogeneity in the pancreatic tumor microenvironment. Cancer Discov. 2020;10:648–56.
23. Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA, et al. Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. Cancer Discov. 2019;9:1102–23.
24. Mizutani Y, Kobayashi H, Iida T, Asai N, Masamune A, Hara A, et al. Meflin-Positive Cancer-Associated Fibroblasts Inhibit Pancreatic Carcinogenesis. Cancer Res. 2019;79:5367–81.
25. Collins MA, Bednar F, Zhang Y, Brisset J-C, Galbán S, Galbán CJ, et al. Oncogenic Kras is required for both the initiation and maintenance of pancreatic cancer in mice. J Clin Invest. 2012;122:639–53.
26. Hosein AN, Huang H, Wang Z, Parmar K, Du W, Huang J, et al. Cellular heterogeneity during mouse pancreatic ductal adenocarcinoma progression at single-cell resolution. JCI Insight. 2019;5.
27. Huang H, Wang Z, Zhang Y, Pradhan RN, Ganguly D, Chandra R, et al. Mesothelial cell-derived antigen-presenting cancer-associated fibroblasts induce expansion of regulatory T cells in pancreatic cancer. Cancer Cell. 2022;40:656–673.e7.
28. Begum A, McMillan RH, Chang Y-T, Penchev VR, Rajeshkumar NV, Maitra A, et al. Direct Interactions With Cancer-Associated Fibroblasts Lead to Enhanced Pancreatic Cancer Stem Cell Function. Pancreas. 2019;48:329–34.
29. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SGH, Hoadley KA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet. 2015;47:1168–78.
30. Chan-Seng-Yue M, Kim JC, Wilson GW, Ng K, Figueroa EF, O'Kane GM, et al. Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. Nat Genet. 2020;52:231–40.
31. Ishiwata T, Matsuda Y, Yoshimura H, Sasaki N, Ishiwata S, Ishikawa N, et al. Pancreatic cancer stem cells: features and detection methods. Pathol Oncol Res. 2018;24:797–805.

32. Maruno T, Fukuda A, Goto N, Tsuda M, Ikuta K, Hiramatsu Y, et al. Visualization of stem cell activity in pancreatic cancer expansion by direct lineage tracing with live imaging. Elife. 2021;10.
33. Kure S, Matsuda Y, Hagio M, Ueda J, Naito Z, Ishiwata T. Expression of cancer stem cell markers in pancreatic intraepithelial neoplasias and pancreatic ductal adenocarcinomas. Int J Oncol. 2012;41:1314–24.
34. Prasad NB, Biankin AV, Fukushima N, Maitra A, Dhara S, Elkahloun AG, et al. Gene expression profiles in pancreatic intraepithelial neoplasia reflect the effects of Hedgehog signaling on pancreatic ductal epithelial cells. Cancer Res. 2005;65:1619–26.
35. Ayars M, O'Sullivan E, Macgregor-Das A, Shindo K, Kim H, Borges M, et al. IL2RG, identified as overexpressed by RNA-seq profiling of pancreatic intraepithelial neoplasia, mediates pancreatic cancer growth. Oncotarget. 2017;8:83370–83.
36. Buchholz M, Braun M, Heidenblut A, Kestler HA, Klöppel G, Schmiegel W, et al. Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions. Oncogene. 2005;24:6626–36.
37. Sodir NM, Kortlever RM, Barthet VJA, Campos T, Pellegrinet L, Kupczak S, et al. MYC instructs and maintains pancreatic adenocarcinoma phenotype. Cancer Discov. 2020;10:588–607.
38. Maddipati R, Norgard RJ, Baslan T, Rathi KS, Zhang A, Saeid A, et al. MYC levels regulate metastatic heterogeneity in pancreatic adenocarcinoma. Cancer Discov. 2022;12:542–61.
39. Ashton TM, McKenna WG, Kunz-Schughart LA, Higgins GS. Oxidative phosphorylation as an emerging target in cancer therapy. Clin Cancer Res. 2018;24:2482–90.
40. Radon TP, Massat NJ, Jones R, Alrawashdeh W, Dumartin L, Ennis D, et al. Identification of a Three-Biomarker Panel in Urine for Early Detection of Pancreatic Adenocarcinoma. Clin Cancer Res. 2015;21:3512–21.
41. Arumugam T, Brandt W, Ramachandran V, Moore TT, Wang H, May FE, et al. Trefoil factor 1 stimulates both pancreatic cancer and stellate cells and increases metastasis. Pancreas. 2011;40:815–22.
42. Manoukian P, Bijlsma M, van Laarhoven H. The Cellular Origins of Cancer-Associated Fibroblasts and Their Opposing Contributions to Pancreatic Cancer Growth. Front Cell Dev Biol. 2021;9:743907.
43. Helms EJ, Berry MW, Chaw RC, DuFort CC, Sun D, Onate MK, et al. Mesenchymal Lineage Heterogeneity Underlies Nonredundant Functions of Pancreatic Cancer-Associated Fibroblasts. Cancer Discov. 2022;12:484–501.
44. Masoud R, Reyes-Castellanos G, Lac S, Garcia J, Dou S, Shintu L, et al. Targeting mitochondrial complex I overcomes chemoresistance in high OXPHOS pancreatic cancer. Cell Rep Med. 2020;1:100143.
45. Kanda M, Matthaei H, Wu J, Hong S-M, Yu J, Borges M, et al. Presence of somatic mutations in most early-stage pancreatic intraepithelial neoplasia. Gastroenterology. 2012;142:730–733.e9.
46. Löhr M, Klöppel G, Maisonneuve P, Lowenfels AB, Lüttges J. Frequency of K-ras mutations in pancreatic intraductal neoplasias associated with pancreatic ductal adenocarcinoma and chronic pancreatitis: a meta-analysis. Neoplasia. 2005;7:17–23.
47. Murphy SJ, Hart SN, Lima JF, Kipp BR, Klebig M, Winters JL, et al. Genetic alterations associated with progression from pancreatic intraepithelial neoplasia to invasive pancreatic

tumor. Gastroenterology. 2013;145:1098–1109.e1.
48. Fertig EJ, Ding J, Favorov AV, Parmigiani G, Ochs MF. CoGAPS: an R/C++ package to identify patterns and biological process activity in transcriptomic data. Bioinformatics. 2010;26:2792–3.
49. Sherman TD, Gao T, Fertig EJ. CoGAPS 3: Bayesian non-negative matrix factorization for single-cell analysis with asynchronous updates and sparse data structures. BMC Bioinformatics. 2020;21:453.
50. Feldmann G, Beaty R, Hruban RH, Maitra A. Molecular genetics of pancreatic intraepithelial neoplasia. J Hepatobiliary Pancreat Surg. 2007;14:224–32.
51. Lee J, Snyder ER, Liu Y, Gu X, Wang J, Flowers BM, et al. Reconstituting development of pancreatic intraepithelial neoplasia from primary human pancreas duct cells. Nat Commun. 2017;8:14686.
52. Ho WJ, Jaffee EM, Zheng L. The tumour microenvironment in pancreatic cancer - clinical challenges and opportunities. Nat Rev Clin Oncol. 2020;17:527–40.
53. Hermann PC, Sainz B. Pancreatic cancer stem cells: A state or an entity? Semin Cancer Biol. 2018;53:223–31.
54. Valle S, Martin-Hijano L, Alcalá S, Alonso-Nocelo M, Sainz B. The Ever-Evolving Concept of the Cancer Stem Cell in Pancreatic Cancer. Cancers (Basel). 2018;10.
55. Askan G, Sahin IH, Chou JF, Yavas A, Capanu M, Iacobuzio-Donahue CA, et al. Pancreatic cancer stem cells may define tumor stroma characteristics and recurrence patterns in pancreatic ductal adenocarcinoma. BMC Cancer. 2021;21:385.
56. Tian J, Zhang R, Piao H, Li X, Sheng W, Zhou J, et al. Silencing Tspan1 inhibits migration and invasion, and induces the apoptosis of human pancreatic cancer cells. Mol Med Rep. 2018;18:3280–8.
57. Hou F-Q, Lei X-F, Yao J-L, Wang Y-J, Zhang W. Tetraspanin 1 is involved in survival, proliferation and carcinogenesis of pancreatic cancer. Oncol Rep. 2015;34:3068–76.
58. Wang S, You L, Dai M, Zhao Y. Mucins in pancreatic cancer: A well-established but promising family for diagnosis, prognosis and therapy. J Cell Mol Med. 2020;24:10279–89.
59. Sunagawa M, Yamaguchi J, Kokuryo T, Ebata T, Yokoyama Y, Sugawara G, et al. Trefoil factor family 1 expression in the invasion front is a poor prognostic factor associated with lymph node metastasis in pancreatic cancer. Pancreatology. 2017;17:782–7.
60. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184:3573–3587.e29.
61. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019;177:1888–1902.e21.
62. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36:411–20.
63. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015;33:495–502.
64. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 2015;16:278.
65. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide

expression profiles. Proc Natl Acad Sci USA. 2005;102:15545–50.
66. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003;34:267–73.
67. Wickham H. Ggplot2: Elegant graphics for data analysis. 2nd ed. Cham, Switzerland: Springer International Publishing; 2016.
68. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32:2847–9.
69. Stein-O'Brien GL, Carey JL, Lee WS, Considine M, Favorov AV, Flam E, et al. PatternMarkers & GWCoGAPS for novel data-driven biomarkers via whole transcriptome NMF. Bioinformatics. 2017;33:1892–4.
70. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1:417–25.
71. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005;21:3439–40.
72. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. BioRxiv. 2016;
73. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv. 2018;
74. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32:381–6.
75. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017;14:979–82.
76. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature. 2019;566:496–502.
77. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol. 2018;36:421–7.
78. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep. 2019;9:5233.
79. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir ED, Tadmor MD, et al. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell. 2015;162:184–97.

FIGURE LEGENDS

FIGURE 1 – Spatial transcriptomics analysis of FFPE pancreatic intraepithelial neoplasia (PanIN). (A) Pancreatic cancer surgical specimens in FFPE were examined and the regions containing PanIN lesions were identified for scoring using a 5mm skin biopsy punch and sectioning onto the spatial transcriptomics slide. (B) 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 which location resembles the distribution observed in the stained sections. (D) CODA for cell type classification was applied to assign with accuracy the cell type mapped to each spatial spot and to determine the spots that captured more than one cell type.

FIGURE 2 – Spatial distribution of PDAC cancer associated fibroblasts (CAF) subtypes. (A) CAFs localization was mapped using pan-CAF 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.

FIGURE 3 – Pancreatic intraepithelial neoplasia (PanINs) transcriptional features. (A) Six out of seven PanINs, 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) Differential expression analysis identified genes which up-regulation (blue dots) or down-regulation (red dots) in PanINs, relative to normal ducts, discriminate preneoplastic from normal cells (E).

FIGURE 4 – Identification of transcriptional changes associated with pancreatic intraepithelial neoplasia (PanIN) differentiation grade. (A) Normal, low grade (LG) and high grade (HG) PanINs are morphologically distinct and can be classified by pathology examination. (B, C and D) 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, (E) *MUCL3,* (F) *TSPAN1* and (G) *TFF1*, show gradual increase from normal through LG until HG progression.

FIGURE 5 – Integration of pancreatic intraepithelial neoplasia (PanIN) spatial transcriptomics (ST) data with invasive pancreatic cancer single-cell RNA-sequencing (scRNAseq) using transfer learning. (A, C and E) Representation of enriched MSigDB pathways in Pattern 2, Pattern 5 and Pattern 7 of the PDAC atlas. (B, D and F) Violin plots of projected PDAC atlas Pattern 2, Pattern 5 and Pattern 7 weights in PanIN ST spots. (G) UMAP embedding of epithelial cells from the PDAC atlas colored by *TFF1* expression. (H) Violin plots of TFF1 expression in all PanIN ST epithelial spots grouped by epithelial lesion grade. P-values were calculated using two-sample Wilcoxon rank-sum tests. (N: normal, LG: low-grade, HG: high-grade).

Figure 1

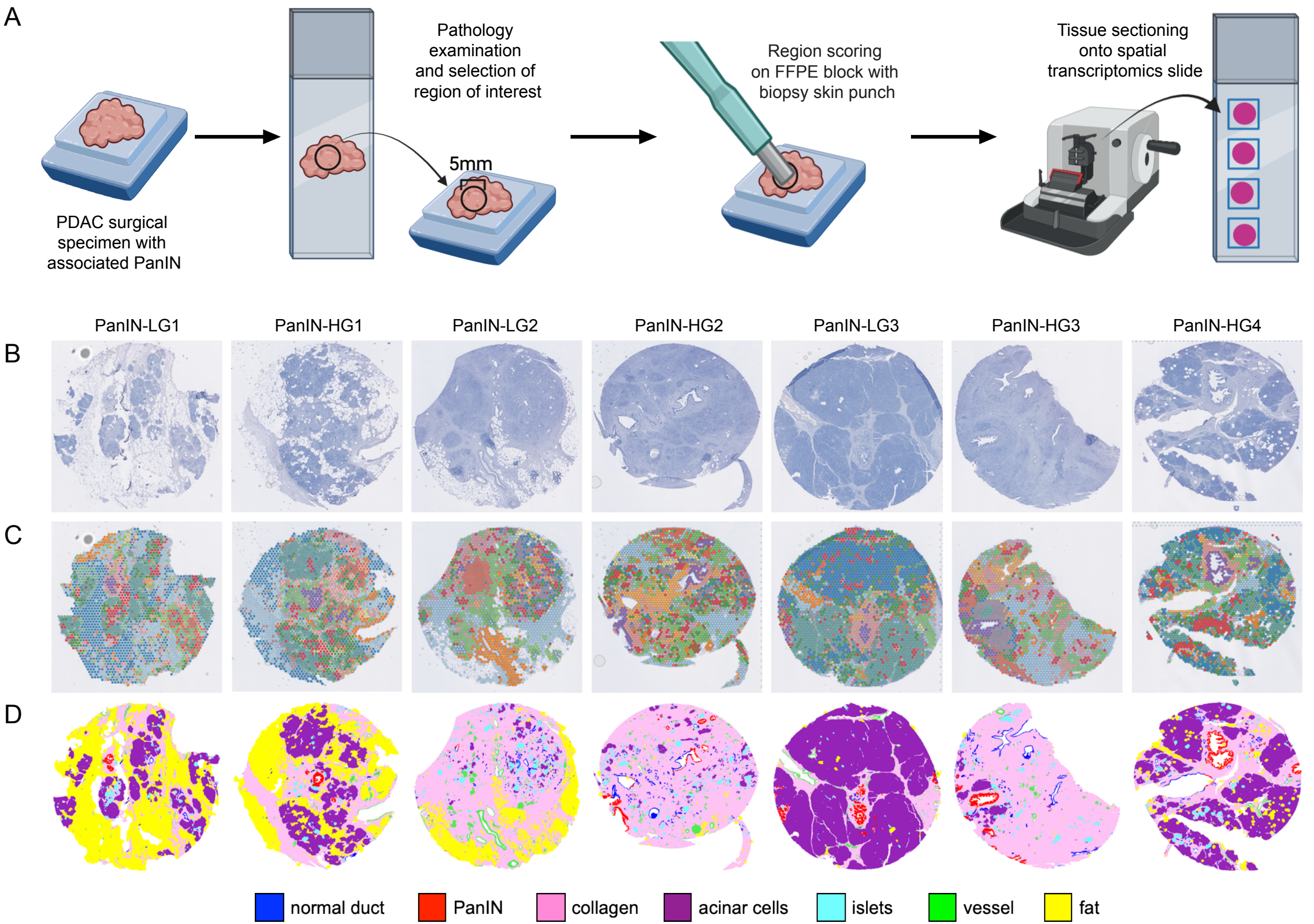

A
PDAC surgical specimen with associated PanIN
Pathology examination and selection of region of interest
5mm
Region scoring on FFPE block with biopsy skin punch
Tissue sectioning onto spatial transcriptomics slide
B
PanIN-LG1
PanIN-HG1
PanIN-LG2
PanIN-HG2
PanIN-LG3
PanIN-HG3
PanIN-HG4
C
D
normal duct
PanIN
collagen
acinar cells
islets
vessel
fat

Figure 2

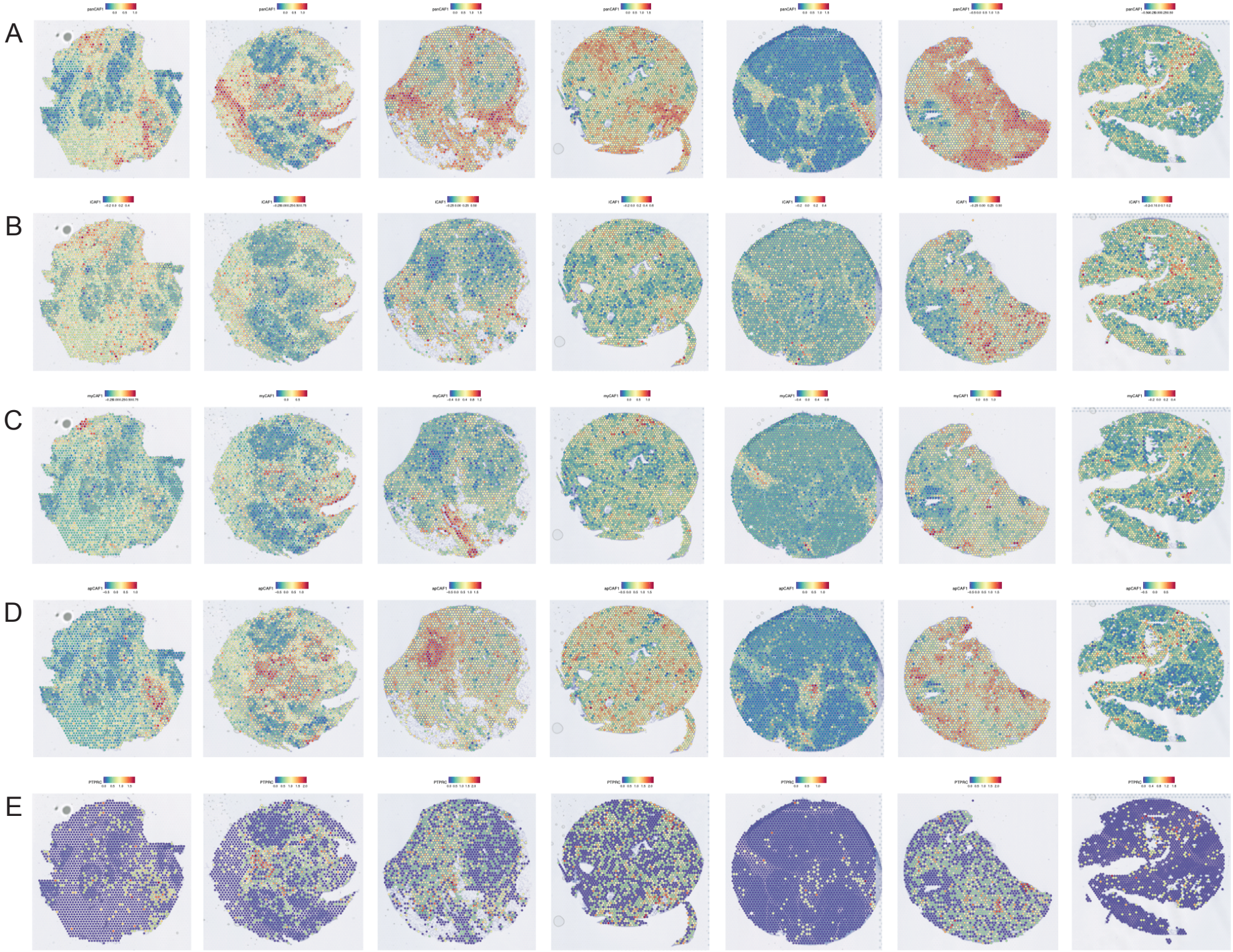
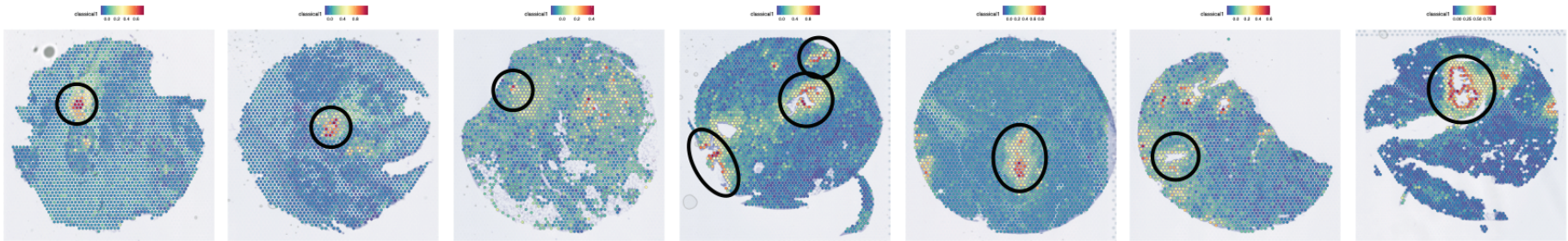


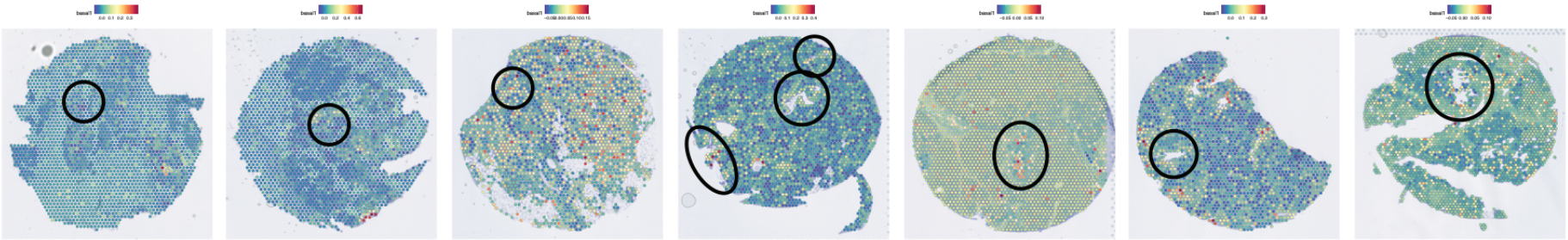
Figure 3

A



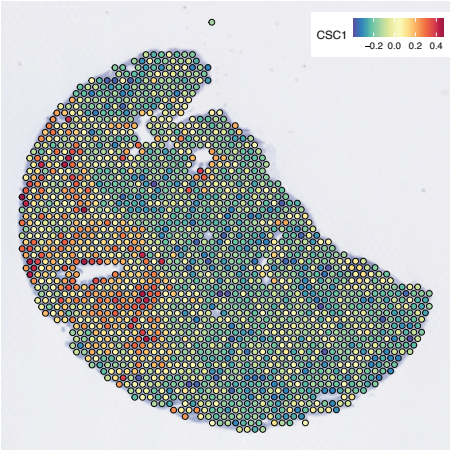
Classical genes: *BTNL8, FAM3D, ATAD4, AGR3, CTSE, LOC400573, LYZ, TFF2, TFF1, ANXA10, LGALS4, PLA2G10, CEACAM6, VSIG2, TSPAN8, ST6GALNAC1, AGR2, TFF3, CYP3A7, MYO1A, CLRN3, KRT20, CDH17, SPINK4, REG4*

B



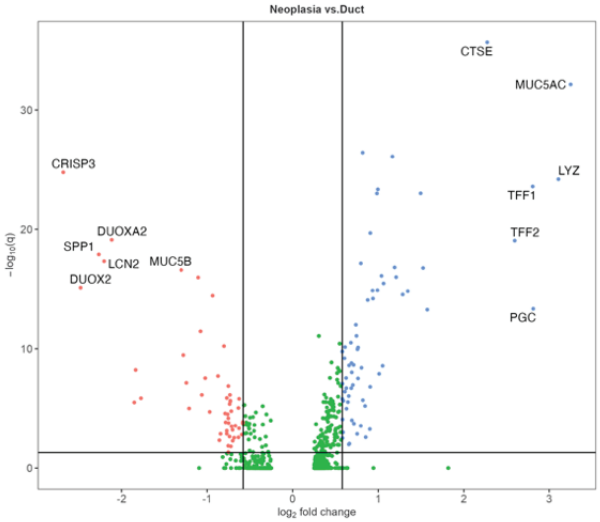
Basal-like genes: *VGLL, UCA1, S100A2, LY6D, SPRR3, SPRR1B, LEMD1, KRT15, CTSL2, DHRS9, AREG, CST6, SERPINB3, KRT6C, KRT6A, SERPINB4, FAM83A, SCEL, FGFBP1, KRT7, KRT17, GPR87, TNS4, SLC2A1, ANXA8L2*

C



CSC signature:
ABCG2, ALDH1A1, CD24, CD44, EPCAM, PROM1, CXCR4, NES, DCLK1, SOX9, NANOG

D



E

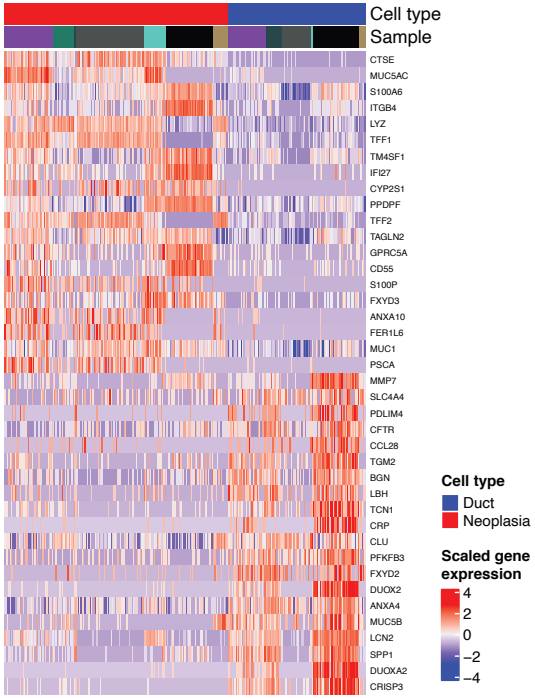


Figure 4

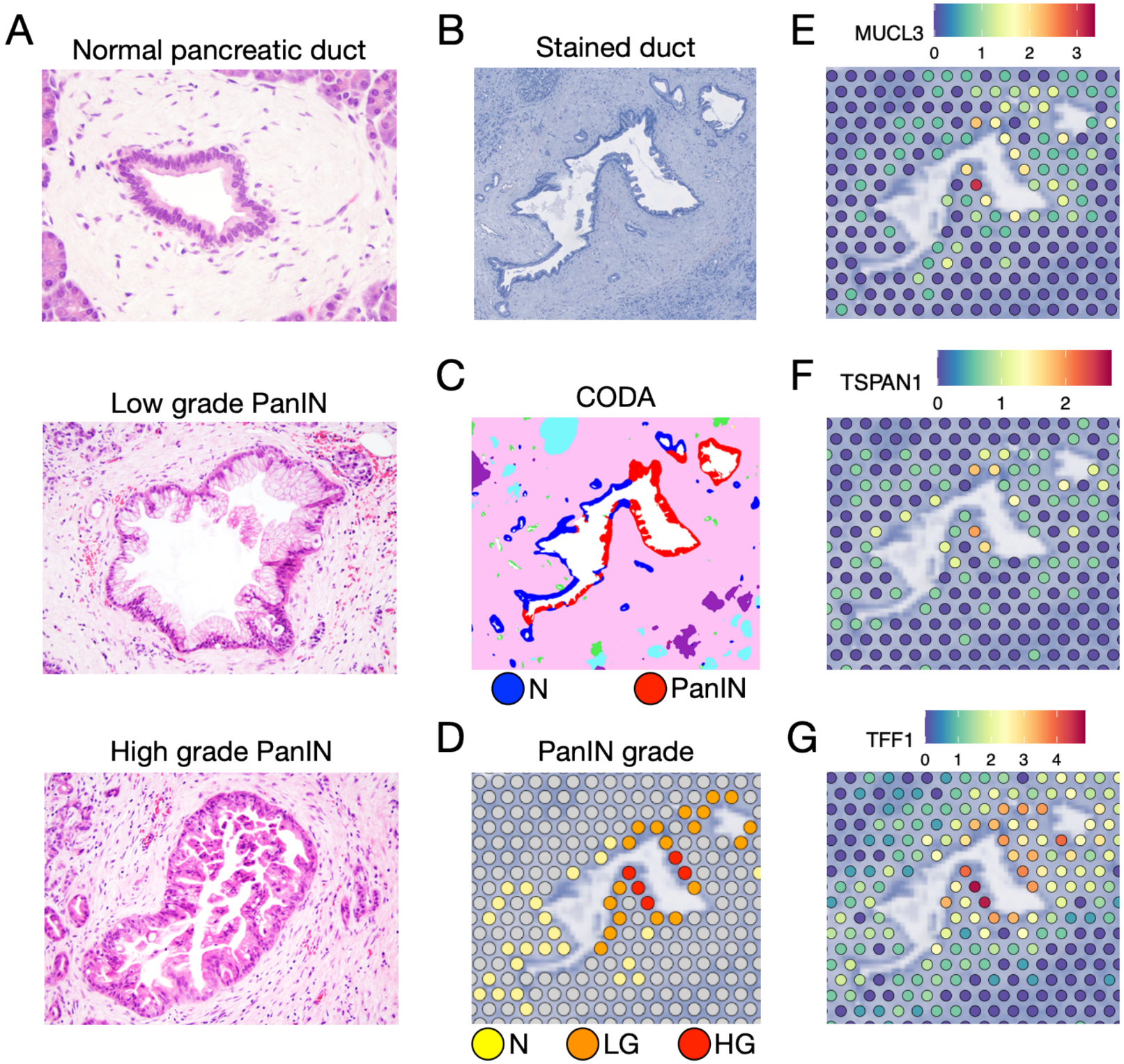
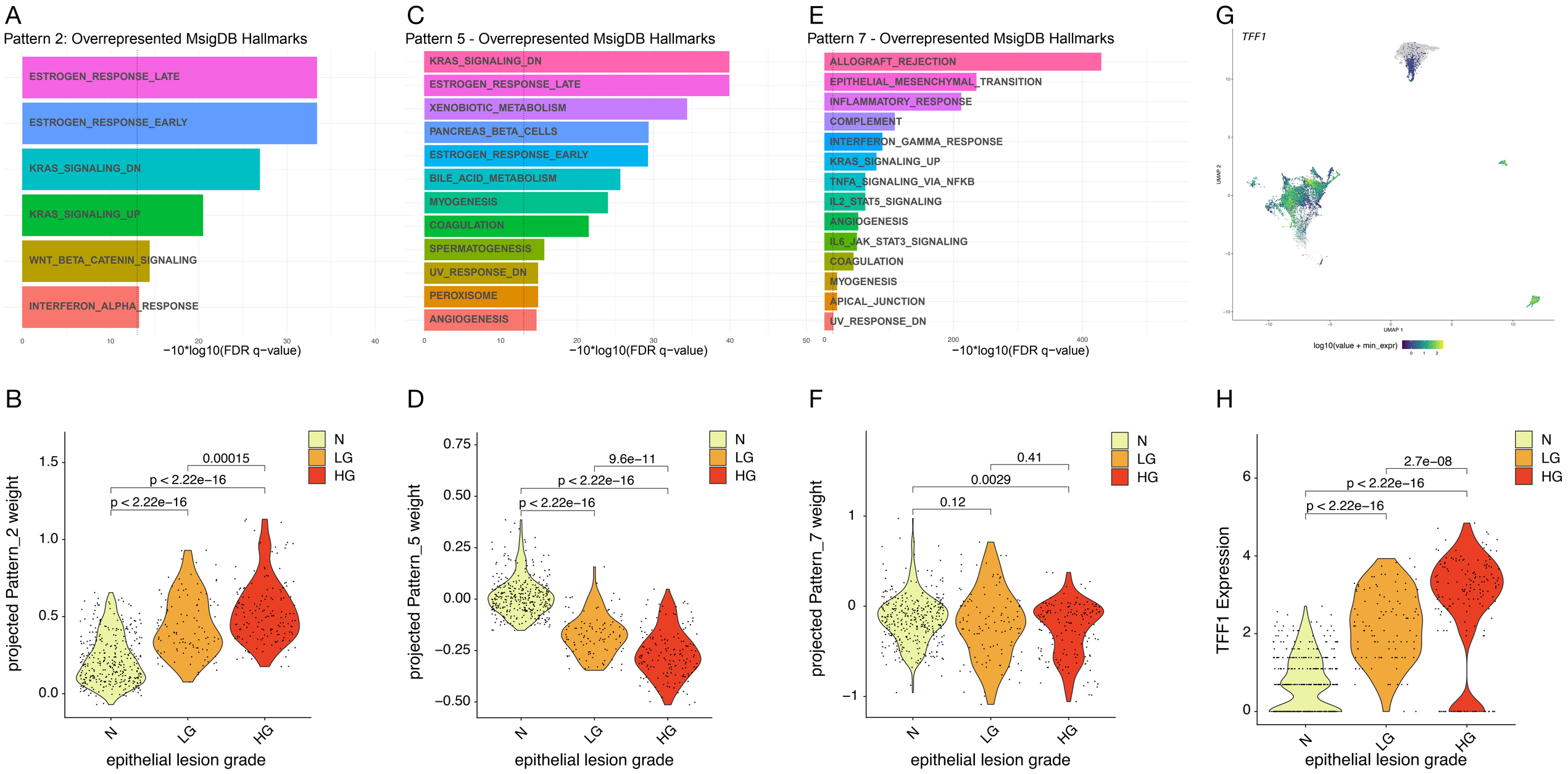

A
Normal pancreatic duct
Low grade PanIN
High grade PanIN
B
Stained duct
C
CODA
N
PanIN
D
PanIN grade
N
LG
HG
E
MUCL3
0
1
2
3
F
TSPAN1
0
1
2
G
TFF1
0
1
2
3
4

FIGURE 5


A
Pattern 2: Overrepresented MsigDB Hallmarks
ESTROGEN_RESPONSE_LATE
ESTROGEN_RESPONSE_EARLY
KRAS_SIGNALING_DN
KRAS_SIGNALING_UP
WNT_BETA_CATENIN_SIGNALING
INTERFERON_ALPHA_RESPONSE
−10*log10(FDR q−value)
C
Pattern 5 - Overrepresented MsigDB Hallmarks
KRAS_SIGNALING_DN
ESTROGEN_RESPONSE_LATE
XENOBIOTIC_METABOLISM
PANCREAS_BETA_CELLS
ESTROGEN_RESPONSE_EARLY
BILE_ACID_METABOLISM
MYOGENESIS
COAGULATION
SPERMATOGENESIS
UV_RESPONSE_DN
PEROXISOME
ANGIOGENESIS
−10*log10(FDR q−value)
E
Pattern 7 - Overrepresented MsigDB Hallmarks
ALLOGRAFT_REJECTION
EPITHELIAL_MESENCHYMAL_TRANSITION
INFLAMMATORY_RESPONSE
COMPLEMENT
INTERFERON_GAMMA_RESPONSE
KRAS_SIGNALING_UP
TNFA_SIGNALING_VIA_NFKB
IL2_STAT5_SIGNALING
ANGIOGENESIS
IL6_JAK_STAT3_SIGNALING
COAGULATION
MYOGENESIS
APICAL_JUNCTION
UV_RESPONSE_DN
−10*log10(FDR q−value)
G
TFF1
UMAP 1
UMAP 2
log10(value + min_expr)
B
projected Pattern_2 weight
0.00015
p < 2.22e−16
p < 2.22e−16
N
LG
HG
epithelial lesion grade
D
projected Pattern_5 weight
9.6e−11
p < 2.22e−16
p < 2.22e−16
epithelial lesion grade
F
projected Pattern_7 weight
0.41
0.0029
0.12
epithelial lesion grade
H
TFF1 Expression
2.7e−08
p < 2.22e−16
p < 2.22e−16
epithelial lesion grade