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

Authors: Benedict Kinny-Köster[*1,2], Samantha Guinn[*3-5], Joseph A. Tandurella[*3-5], Jacob T. Mitchell[3-6], Dimitrios N. Sidiropoulos[3-5], Melanie Loth[3-5], Melissa R. Lyman[3-5], Alexandra B. Pucsek[3-5], Toni T. Seppälä[1,7], Christopher Cherry[5,8,9], Reecha Suri[1], Haley Zlomke[1], Jin He[1], Christopher L. Wolfgang[2], Jun Yu[1], Lei Zheng[3-5], David P. Ryan[10], David T. Ting[10], Alec Kimmelman[11], Anuj Gupta[3], Ludmila Danilova[3-5], Jennifer Elisseef[5,7,8], Laura D. Wood[4,12], Genevieve Stein-O'Brien[3-6], Luciane T. Kagohara[3-5], Richard A. Burkhart[1,3,4], Elizabeth M. Jaffee[3-5], Elana J. Fertig[‡3-5,9,13], Jacquelyn W. Zimmerman[‡3-5]

[*]Contributed equally

[‡] Contributed equally

[1]Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD
[2]Department of Surgery, New York University Grossman School of Medicine, New York, NY
[3]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD
[4]Convergence Institute, Johns Hopkins University School of Medicine, Baltimore, MD
[5]Bloomberg Kimmel Immunology Institute, Johns Hopkins University School of Medicine, Baltimore, MD
[6]Department of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD
[7]Faculty of Medicine and Health Technology, Tampere University and Tays Cancer Center, Tampere University Hospital
[8]Translational Tissue Engineering Center, Wilmer Eye Institute, Johns Hopkins School of Medicine, Baltimore, MD
[9]Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, MD
[10]The Massachusetts General Hospital Cancer Center and Department of Medicine, Harvard Medical School, Boston, Massachusetts.
[11]Department of Radiation Oncology at New York University Grossman School of Medicine, NYU Langone Health, New York, New York.
[12]Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD
[13]Department of Applied Mathematics and Statistics, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD

Running Title: Modeling Fibroblast-Cancer Interactions in Human PDAC

Keywords: pancreatic ductal adenocarcinoma (PDAC), scRNAseq, organoids, cancer associated fibroblasts (CAFs), atlas

Additional Information:

Financial Support: Hopper-Belmont Foundation (JWZ), The Lustgarten Foundation (EMJ), Johns Hopkins University Discovery Award (EJF, LW), NIH/NCI (U01CA253403 to EJF; P01CA247886 to EMJ; P30CA006973; R50 CA243627 to LD; K08CA248710 to RAB), German Research Foundation (KI 2437-2/1 to BKK). TT Seppälä was supported by fellowship grants and research funding from Sigrid Juselius Foundation, Instrumentarium Science Foundation, Emil Aaltonen Foundation, Jane and Aatos Erkko Foundation, Relander Foundation, and the iCAN precision medicine flagship of the Finnish Academy. Stand Up To Cancer–Lustgarten Foundation Pancreatic Cancer Interception Translational Cancer Research Grant (grant number SU2C-AACR-DT26-17 to RAB). Stand Up To Cancer (SU2C) is a division of the Entertainment Industry Foundation and funding is administered by the American Association for Cancer Research, the scientific partner of SU2C). Stand Up To Cancer -Lustgarten Foundation (grant number 2015-002 to DTT)

Conflict of Interest Statement:
T.T.S. is the CEO and co-owner of Healthfund Finland and reports an interview honorarium from Boehringer Ingelheim Finland. E.M.J is a paid consultant for Adaptive Biotech, Achilles, DragonFly, Candel Therapeutics, Genocea, and Roche. She receives funding from Lustgarten Foundation and Bristol Myer Squibb. She is the Chief Medical Advisor for Lustgarten and SAB advisor to the Parker Institute for Cancer Immunotherapy (PICI) and for the C3 Cancer Institute. She is a founding member of Abmeta. E.J.F is on the SAB for Resistance Biology, Consultant for Mestag Therapeutics and Merck.
D.T.T. has received consulting fees from ROME Therapeutics, Tekla Capital, Ikena Oncology, Foundation Medicine, Inc., NanoString Technologies, and Pfizer that are not related to this work. D.T.T. is a founder and has equity in ROME Therapeutics, PanTher Therapeutics and TellBio, Inc., which is not related to this work. D.T.T. receives research support from ACD-Biotechne, PureTech Health LLC, and Ribon Therapeutics, which was not used in this work. D.T.T.'s interests were reviewed and are managed by Massachusetts General Hospital and Mass General Brigham in accordance with their conflict of interest policies.

Corresponding Authors:

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

Jacquelyn W. Zimmerman, MD, PhD
1650 Orleans St., Room 352
Baltimore, MD 21287
Phone: 443-287-2883
Fax: 410-614-8216
Email: jzimme27@jhmi.edu

Word Count: 5,888

Figures: 6

Tables: 1

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is a devastating malignancy driven by a heterogeneous tumor microenvironment enriched with cancer associated fibroblasts (CAFs) that influence its overall immunosuppressive composition. We examined inflammation in PDAC by leveraging our existing patient-derived organoid (PDO) model and a novel PDO-CAF co-culture. We first identified induction of major histocompatibility complex class II expression following treatment with interferon gamma. In parallel, we collated an atlas of 6 published single-cell RNA-sequencing datasets (174,394 cells) combining 61 PDAC (142,807 cells) and 16 non-malignant samples. By combining *in silico* modeling and *in vitro* PDO co-culture, we define a gene expression pattern of inflammatory processes in epithelial tumor cells. Following computational inferences, we examined interactions between epithelial tumor cells and CAFs, focusing on *VEGF-A* and *ITGB1* pathways. This work, integrating computational and biological approaches, highlights the value of convergence to accelerate our understanding of key drivers of PDAC.

Statement of Significance

We established PDO-CAF co-cultures to model tumor cell interactions and validated discoveries using transfer learning into a single-cell RNA-seq atlas of PDAC tumors. This bidirectional approach from human to experimental systems facilitates interrogation of PDAC biology, including the role of inflammation associated with VEGF-A crosstalk between CAFs and tumor cells.

Introduction

Pancreatic ductal adenocarcinoma (PDAC) remains challenging to effectively treat largely due to detection at advanced stages and a heterogeneous tumor microenvironment (TME). Carcinogenesis of the epithelial cell population provokes a strong response from surrounding non-epithelial cell populations and gives rise to a dense TME[1]. Critical infiltrating cell types include mesenchymal cells, such as cancer associated fibroblasts (CAFs), and an array of immune cells of myeloid and lymphoid origin[1]. The recognition of biologic diversity in cell types in the TME, and our understanding of the functional diversity of these distinct cell populations, has historically been limited by available genomic technologies.

Single-cell RNA sequencing (scRNA-seq) provides a platform to study PDAC for cell-type-specific, patient-specific, and generalizable tumor properties. Prior single-cell datasets have provided a survey to help identify individual cell populations and start understanding transcriptional dynamics[2–5]. These data have demonstrated previously underappreciated heterogeneity throughout stages of malignant progression in both humans and mouse models and are are complemented in studies exploring the importance of the TME in driving tumor phenotype[6,7]. Single-cell technologies have furthered our capacity to identify discrete and functionally distinct subpopulations in both the malignant epithelial cells and the supporting stromal cells. In PDAC, CAFs are abundant in the TME with some CAF subtypes implicated in mechanisms of tumor progression and others described as tumor-restraining[8,9]. Myofibroblastic CAFs (myCAFs) and inflammatory CAFs (iCAFs) are two subpopulations that have been well-described[10,11]. More recently, additional subtypes have been proposed, including those expressing major histocompatibility complex class II (MHC-II) genes, termed antigen-presenting CAFs (apCAFs)[10]. While the identification of these subpopulations and their associated characterization has opened new avenues of research in this disease, the mechanisms of intercellular interaction and specifically how CAF functional heterogeneity may impact the phenotype of the tumor cells remains poorly understood. Still, it remains challenging to study intercellular interactions in laboratory tumor models in a manner that enables *in vitro* mechanistic evaluation of *in silico* hypotheses generated from high-throughput datasets.

Currently available biologic models of PDAC are restricted in large part to either mouse models or *in vitro* epithelial cell lines derived from human pancreatic tumor cells. While mouse models can provide a global characterization of the TME, depletion experiments are required to isolate the effect of individual cell types. *In vitro* experiments have traditionally been limited to epithelial tumor cells, but more recently have been adapted to investigate the effect of cell-cell interactions and paracrine crosstalk with other cell types using conditioned media. An emerging area of tumor

modeling is the use of organoids and organoids in co-culture to better recapitulate the tumor and associated microenvironment[12].

With a focus on the impact of CAFs on inflammatory mechanisms and to further interrogate intercellular crosstalk across cell types, we established a dynamic *in vitro* three-dimensional co-culture system of PDOs and patient-derived CAFs. In support of this approach, historical reports have demonstrated MHC-II expression in PDAC cell lines and tissues, suggesting malignant epithelial cells can respond to inflammatory stimuli[13–15]. We now expand on these findings to examine both constitutive and inducible expression of MHC-II in PDO models. Whereas single-cell data can provide an approach to isolate subpopulations with bioinformatics approaches, the PDO co-culture is an additive system for evaluating the transcriptional dynamics in response to external stimuli and investigating core intercellular interactions.

While our PDO co-culture enables a powerful tool for mechanistic studies of interactions between patient-derived tumor and fibroblasts, additional signaling modulation may occur in human PDAC tumors. Therefore, we complemented our PDO co-culture by analyzing, in parallel, six published scRNA-seq datasets generated from PDAC patients to create a large cohort for application of *in silico* methods to evaluate epithelial tumor cell and CAF intercellular interactions[10,16–20].

Results

Epithelial cells in PDO models of PDAC respond to exogenous interferon gamma (IFNγ) stimulation to induce MHC-II gene expression and associated protein expression

We previously demonstrated that PDO models of PDAC malignant epithelial cells have the capacity to recapitulate intratumoral heterogeneity and accurately provide clinically-actionable data for chemotherapeutic selection[21,22]. We now aim to use PDOs to further our understanding of tumor biology and mechanisms of inflammation in PDAC. We first validated the dynamics of our PDOs in preparation for establishing co-cultures by testing the responsiveness of PDOs to exogenous inflammatory stimuli. Previous work has shown that 24 hours of 200ng/mL of IFNγ added into cell media induces MHC-I and PD-L1 upregulation in organoids[23]. We first sought to validate this finding and expand our study to include MHC-II. We screened eleven PDO lines by flow cytometry (gating strategy Fig. S1) for both constitutive and IFNγ-induced cell surface protein expression of MHC-I and PD-L1 (Fig. 1A). At 24 hours, both MHC-I and PD-L1 demonstrated a robust increase in expression in response to IFNγ stimulation. HLA-DR was used to represent MHC-II expression and was still low after 24 hours. Yet, HLA-DR expression was more robustly induced in one representative PDO line following stimulation with IFNγ over a course of 96 hours (Fig. 1B). We further evaluated the endogenous and IFNγ-induced gene expression in five organoid lines after 96 hours of IFNγ supplementation to confirm

reproducibility of this phenotype (Fig. 1C-D). To further evaluate IFNγ-induced changes in MHC-II alleles which lack specific antibodies, we validated HLA-DRB1, HLA-DRA, HLA-DPB1, HLA-DQB1 by qPCR and observed a variable response in gene expression (Fig. 1E). We then examined inducible MHC-II expression in formalin fixed and paraffin-embedded (FFPE) PDAC PDOs using IHC with antibodies against HLA-DR and HLA-DR/DP/DQ to demonstrate the spatial distribution of the MHC-II expression (Fig. 1F-G), further confirming robust MHC-II upregulation in response to IFNγ.

A harmonized atlas of 61 PDAC samples and 16 non-malignant pancreas control samples obtained from integrative analysis of collated PDAC single-cell data recapitulates MHC-II expression in epithelial cells from patient samples

In order to apply these and future data generated in the PDO system to a broader compilation of human samples, we assembled six published human scRNA-seq datasets resulting in 174,394 total cells from 61 PDAC (142,807 cells) and 16 non-malignant pancreatic tissue samples (31,587 cells) (Fig. 2A-B, Fig. S2 with subset of PDAC specimens)[10,16–20]. All samples were of pancreatic origin (no metastatic samples) and derived treatment-naïve patients (Table 1). Of the 61 PDAC samples (2 were described as arising from cystic pancreatic lesions), 52 had localized disease, 6 had distant metastases/stage IV, and stage was unknown for 3 samples. Of the 16 non-malignant control samples, 5 were specified by the authors as normal-adjacent to adenocarcinoma whereas 11 were normal-adjacent to non-malignant pathologies. All control samples and the majority of PDAC samples were obtained from resected surgical specimens except for 10 PDAC samples that originated from fine-needle biopsies (4 of these had metastatic disease).

After filtering out cells based upon biological or technical quality metrics (fraction of mitochondrial reads and number of transcribed genes as defined in the methods), the resulting atlas included 140,250 cells (dropout 19.6%). The median and mean cell counts per patient sample were 1,455 and 1,821, respectively (interquartile range 828 – 2,200). After computational pre-processing, we performed clustering analysis for cell type annotation (supplemental methods). To make inferences on biology from an integrated atlas, it is critical to understand the impact of inter-dataset variation for the mitigation of non-representative findings or technical artifacts. Most of the contributing tissues and cells (35 and 54,813, respectively) originated from Peng et al[16] (Fig. 2C-E) with a mean of 1,566 cells per tissue. The mean number of included cells per sample was highest (4,754 cells) in the dataset by Moncada et al[19] that contributed PDAC samples from three patients (Fig. 2C-E). Mapping of dataset origin on the atlas demonstrated contribution from all six datasets for the majority of clusters, thereby visually verifying our integration strategy (Fig. 2E).

Within PDAC samples, the predominant cell populations were epithelial malignant cells (32,515 or 29.3%), cells of myeloid origin (28,971 cells or 26.1%), and T cells (17,284 cells or 15.6%) (Fig. S2B). The mesenchymal cell populations were composed of 8,953 CAFs (8.1%) and 6,049 stellate cells (5.4%). Altogether, the components of the TME (immune and mesenchymal cell populations combined) contributed 67,690 cells or 60% of the total count. Subpopulations within the myeloid and lymphoid cell clusters, such as macrophages, mast cells, neutrophils, regulatory T cells and natural killer/cytotoxic T cells, were further annotated and are provided in the Supplemental Material (Fig. S3). Furthermore, we classified epithelial cells from PDAC tissues according to a classical and basal gene expression program according to Moffitt et al[21] (Fig. S4) and CAFs according to an iCAF and myCAF gene expression program (Fig. S5). We did not explicitly interrogate the presence of the more recently identified apCAF population[10]. Cell cycle analyses within the PDAC samples revealed high cell cycle activity reflected by increased activity scores and translated phases primarily in the epithelial malignant, acinar, myeloid and CAF populations (Fig. S2C-D). For the epithelial cells from tumor tissues, copy number variation analyses (CNV, Fig. S6) were combined with differential expression analyses to confirm epithelial cell classification (benign or malignant).

To evaluate MHC-II expression in epithelial cells from PDAC tumors, we first queried the PDAC atlas for 12 MHC-II genes such as HLA-DRB1, HLA-DOA, HLA-DMA and identified varying degrees of gene expression in the tumor epithelial cell population (Fig. 2F-G). These findings demonstrate that MHC-II genes are expressed in epithelial populations from dissociated heterogeneous tumor tissue. Overall, MHC-II gene expression was most pronounced in epithelial cells derived from tumor tissue as compared to epithelial cells obtained from non-malignant, control samples (Fig. S7). Additionally, expression of HLA-DO and HLA-DM isoforms suggest functionality of MHC-II proteins given their role as chaperones in antigen presentation. Moreover, these findings are consistent with the identification of MHC-II expression in the PDO model.

Epithelial cells derived from tumor tissue demonstrated enhanced inflammatory signaling relative to epithelial cells originating from non-malignant control pancreas samples

Given the induction of MHC-II expression in epithelial tumor cells from the PDO cultures and in the PDAC atlas, we next investigated the epithelial populations within the atlas to assess patterns of gene expression. To infer the signaling processes that distinguish epithelial cells within PDAC tissues from epithelial cells derived from non-malignant samples, we performed an unsupervised analysis of epithelial cell expression patterns with the single-cell Bayesian non-negative matrix factorization algorithm CoGAPS[24]. To minimize batch effects while also identifying patterns that are robust

across sample cohorts, we performed this analysis on epithelial cells from the two largest datasets, which contain both PDAC and non-malignant samples: Peng et al[16] (18,261 epithelial cells) and Steele et al[17] (7,181 epithelial cells), which captured 61.0% of all epithelial cells. Mapping the epithelial cells from these two datasets on the UMAP from the entire atlas verified that they represent 49.3% of all malignant cells, and 97.1% of all benign cells. We then applied the CoGAPS algorithm to interrogate patterns of gene expression across epithelial populations, identifying 8 distinct patterns (Fig. S8). Overrepresentation analysis between genes identified by the CoGAPS pattern marker statistic[25] and hallmark gene sets from the Molecular Signatures Database[26] was used to annotate each pattern. Pattern 1 identified pathways associated with UV response and TGFβ signaling that did not reach statistical significance (Supplemental Table 1). Pattern 2 identified pathways of estrogen response and KRAS signaling that are relevant for PDAC development. Metabolic pathways including cell cycle, oxidative phosphorylation, and glycolysis were significant in Patterns 3-5. Pattern 6 identified pathways of apoptosis activity, while Pattern 8 identified hypoxia. Notably, genes associated with Pattern 7 demonstrated increased pathway activities in inflammatory, fibrogenic, and malignant progression-associated gene sets, including EMT (Fig. 3A-C).

Based upon the gene signatures, we associated Pattern 7 with the inflammatory process in the epithelial cells as inflammatory hallmark gene sets were amongst the highest ranked (Fig. 3C). Indeed, we observed that Pattern 7 weights were significantly higher in the malignant epithelial cells and the benign epithelial cells of tumor tissue origin relative to epithelial cells in the non-malignant control samples (Fig. 3B). The enrichment of Pattern 7 within the epithelial malignant population confirmed a potentially transformative inflammatory program in malignant epithelial cells. These findings are further supported by MHC-II expression in the tumor epithelial cells, as inducible MHC-II expression is anticipated in response to inflammatory stimuli. We note that the MHC-II expression extends into the malignant epithelial cells independent of the cells with the inflammatory stimuli inferred in Pattern 7, leading us to investigate alternative sources of activation of inflammatory pathways in the epithelial cells in human PDAC.

Fibroblasts have been shown to be a major driver of inflammatory processes in human disease, though much of this association has been established in non-cancer inflammatory diseases[27]. Therefore, we hypothesized CAFs may also drive inflammatory pathways influencing malignant epithelial cells. To test this hypothesis, we annotated the cell type composition in the tumor and control samples to identify subpopulations in the tumor microenvironment. Notably, the non-epithelial populations in the atlas largely originated from tumor tissue (Fig. 3D). When comparing the average cell type composition between the control and tumor samples within the Peng et al[16] dataset, the fraction of fibroblast, myeloid and lymphoid populations were greater in tumor samples, while the proportion of endothelial cells, acinar cells and total epithelial cells was decreased in the tumor samples (Fig. 3E). However, within the Steele et al[17] dataset,

differences between the fibroblast/CAF, myeloid and lymphoid populations were less pronounced (Fig. 3E). We then correlated the mean CoGAPS pattern weights of the inflammatory pattern (Pattern 7) in the epithelial cells with the total proportion of CAFs in the TME in PDAC samples using the datasets from Peng et al[16]and Steele et al[17]to infer the role of CAFs on this inflammatory signaling. We identified a direct association between fibroblast proportion in the tumor sample and mean pattern weight of CoGAPS Pattern 7 in the epithelial cells (Fig. S9). This association was lost when fibroblast populations were further subsetted into iCAFs and myCAFs (Fig. S9). Therefore, testing our hypothesized role of fibroblasts in mediating inflammatory processes in epithelial cells requires further mechanistic validation.

Co-culturing PDOs and CAFs enhances the epithelial inflammatory pattern identified in atlas samples and drives expression of MHC-II genes.

To mechanistically test whether fibroblasts induce the inflammatory pattern in malignant epithelial cells that was identified in the atlas, we established a model system of PDAC PDOs co-cultured with patient-derived CAFs. Each matched model leverages our previously published methods for generating PDOs[21,22] while concurrently extracting CAFs from surgical resection specimens (Fig. S10) for co-culture that maintains viability of both cell-types (Fig. 4A). The three-dimensional cultures were established in Matrigel as the surrogate basement membrane for suspending culture components. We then optimized the culture conditions to maximize both the viability of input cell types during culture and the viability of cells following extraction (Fig. S10).

To compare the signaling processes in our co-culture to the inflammatory signals observed in human PDAC from our atlas, we performed scRNA-seq profiling of PDOs alone, under co-culture, and CAFs alone for one of our matched cultures using multiplexing with the established MULTI-seq protocol[28]. As expected, clustering analysis and visualization of the scRNA-seq data from these conditions separated the fibroblasts from epithelial cells, but the cells from the co-culture and monoculture conditions did not separate into distinct clusters (Fig 4B-C). Despite the lack of separate clusters from the co-culture-derived cells, we did identify transcriptional plasticity in the epithelial cells. The co-culture condition demonstrated a greater number cells expressing markers for both the classical and basal epithelial subtypes (Fig. 4E)[29]. Baseline heterogeneity of the basal and classical programs in PDOs identified using bulk RNA-seq was consistent with the recent report by Krieger et al[30] (Fig. S11). Moreover, induced plasticity was not limited to the epithelial compartment as the co-culture also demonstrated increasing proportions of CAFs expressing gene markers for both iCAFs and myCAFs (Fig. 4F). This further supports the notion that co-culture supports both epithelial and CAF transcriptional plasticity.

We hypothesized that the CAFs in this co-culture model drive inflammatory signaling in a manner similar to that observed in the human system. Therefore, we used the transfer learning method ProjectR[31] to project the inflammatory pattern (Pattern 7) identified in the epithelial cells from the human PDAC atlas onto the epithelial cells of the PDO-CAF co-culture. After 12 hours of co-culture, we identified increased projected gene expression of Pattern 7 in PDAC organoids co-cultured with CAFs relative to those cultured in mono-culture (Fig. 4D), suggesting this inflammatory signal in epithelial cells is stimulated by the presence of CAFs.

We then examined the ability of CAFs to upregulate MHC-II in malignant epithelial cells by first evaluating MHC-II expression from the co-culture scRNA-seq data (Fig. 5A-C). HLA-DRA and HLA-DRB were minimally expressed after 12 hours of co-culture, both in the monoculture and co-culture conditions (Fig. 5 A-C). However, similar to our observations in the PDOs following induction with IFNγ (Fig. 1), expression of both HLA-DRA and HLA-DRB increased in PDOs flow sorted after 24 hours, and more notably after 96 hours of co-culture (Fig. 5D-E). This further implicates CAFs as inflammatory stimuli, particularly following longer co-culture durations. Notably, in the two PDO-CAF co-cultures examined, the HLA-DRA and HLA-DRB expression changes in the PDOs following CAF co-culture were variable, further demonstrating the heterogeneity of patient response to inflammatory stimuli.

Domino intercellular signaling identifies VEGF-A as an epithelial-derived ligand and ITGB1 as a fibroblast-derived ligand driving intercellular communication in the PDAC atlas with validation in PDO co-culture

Upon validating the induction of inflammatory signaling observed in the PDAC atlas in our co-culture PDO model, we sought to perform bidirectional *in silico* and *in vitro* experimental interrogation of the crosstalk between CAFs and epithelial cells that underlies the observed inflammatory signaling in PDAC. In an effort to start to understand interactions between distinct cell type populations, we evaluated the compositional relationships between the epithelial and CAF subtypes. Across datasets, there was not a consistent association between epithelial and CAF subtypes (Fig. S12), leading us to more explicitly model epithelial and CAF interactions. We selected the epithelial and CAF populations of the PDAC tumor atlas derived from the Peng et al[16] (12,120 epithelial cells, 63.8%; 5,823 CAFs, 84.0%) and Steele et al[17] (6,883 epithelial cells, 36.2%; 1,110 CAFs, 16%) datasets and applied Domino, a tool for analyzing intra- and intercellular interactions in scRNA-seq data, to infer intercellular interactions from ligands of one cell type to receptors of another[32]. A global signaling network was generated based on population-specific gene expression of ligands and receptors with established signaling relationships (Fig. 6A). Across both datasets, VEGF-A was expressed by epithelial populations and predicted to target fibroblasts (Fig. 6B-C). We then sought to verify this relationship in

our co-culture model. Cells were flow sorted after 24 or 96 hours of co-culture and qPCR was done to assess VEGF-A expression in the co-culture as compared to the monoculture. At both timepoints, VEGF-A demonstrated increased expression in cells derived from the co-culture relative to those extracted from monoculture, consistent with the Domino predictions (Fig. 6D).

Finally, we evaluated intercellular interactions derived from CAFs that are driving intercellular communication with epithelial organoids. Across basal and classical epithelial subtypes, *ITGB1* was recurrently identified in both the Peng et al[16] and Steele et al[17] datasets as a ligand from both iCAFs and myCAFs directed at the epithelial subpopulations (Fig. 6E-F). While *ITGB1* expression was identified both in patient-derived organoids and CAFs in our co-culture model, it was differentially expressed in the fibroblasts (Fig. 6G). While *ITGB1* expression is not tissue- or cell-type specific, its expression is classically associated with fibroblasts and the function of the extracellular matrix. Thus, the combination of *in silico* and *in vitro* modeling is not only reproducible, but also highlights a modern convergence approach to identify meaningful drivers to the underlying biology of PDAC progression.

Discussion

This study combined novel computational tools and PDO co-culture to examine the role of inflammatory gene expression patterns in epithelial cells in PDAC. This approach was facilitated by the evaluation of dynamic intercellular interactions both *in vitro* and *in silico* by applying an organoid co-culture model and a single-cell atlas evaluating patient specimens in parallel. Specifically, we used the atlas to validate *in vitro* findings of MHC-II expression induced through inflammatory signaling in a novel co-culture PDO system. We then investigated the atlas to validate larger transcriptional programs of inflammation occurring through cellular crosstalk between the epithelium and fibroblasts in the complex PDAC microenvironment. Altogether, these analyses suggest a directed mechanism whereby fibroblasts receive VEGF as a ligand to initiate downstream signaling through the VEGF pathway and induce inflammatory processes independent of endogenous epithelial MHC-II expression. Using our established PDAC PDO models, we demonstrated inducible MHC-II expression in response to exogenous stimuli, with validation in the atlas. However, tumor cells from numerous tumor types can be induced to express MHC-II, but less is understood about the pathologic and therapeutic implications of this induction[14,15,33]. Therefore, delineating the role of MHC-II expression from the specific fibroblast-induced inflammatory signaling found in our study requires further evaluation of crosstalk between additional cell types in the PDAC tumor immune microenvironment.

With the aim to more thoroughly examine the validity of combining our atlas and PDO resources, we next globally explored intercellular interactions computationally. We applied CoGAPS[34,35], a gene expression pattern detection tool, on our PDAC scRNA-seq atlas and identified an inflammatory pattern that was more enriched in the epithelial cells derived from tumor samples, as compared to those from non-malignant control samples. This pattern was also observed in tumor adjacent epithelial cells that were transcriptionally more consistent with "benign" epithelial cells rather than malignant epithelial cells. These findings reiterate that the tissue surrounding a tumor is biologically altered and not the same as tissue obtained from a non-malignant specimen[36]. Moreover, we were able to recapitulate the overrepresentation of this pattern when PDAC PDOs were co-cultured with CAFs, demonstrating that CAFs may impact tumor epithelial cell transcriptional phenotype and overall function. More generally, these data suggest that both CAF and epithelial populations are transcriptionally plastic with changes in cell sub-population distribution after co-culture.

Lastly, we turned to Domino[32] to interrogate direct intercellular interactions and identified bidirectional crosstalk in the epithelial and CAF populations. These interactions were seen in both the Peng et al[16] and the Steel et al[17] datasets, which provided the largest proportion of cells for our collated atlas. *ITGB1* was identified as a CAF-derived ligand that was received by the epithelial cells. While *ITGB1* is a member of the integrin family of proteins that function diversely in cell adhesion and serve as receptors for collagen, it was recently identified as a marker of cytotoxicity potential in CD4+ and CD8+ T cells[37]. Additionally, mutated *ITGB1* was shown to result in an MHC-II-restricted neoantigen in an established sarcoma cell line when evaluated in a subcutaneous xenograft model[38]. Lastly, we identified *VEGF-A* as an epithelial-derived ligand directed towards the fibroblasts. Moreover, this interaction was validated in co-culture, demonstrating the translatability of this and similar algorithms. This finding also supports the inflammatory pattern identified in the tumor epithelial cells in both the atlas and PDO models, as increased VEGF-A expression has been identified in alveolar epithelial cells in response to inflammatory stimuli[39]. Further, CAFs were identified in squamous cell carcinoma of the skin as mediators of tumor-enhancing inflammation and angiogenesis, a signature validated in models of mammary and pancreatic tumors[40]. This particular interaction is of biological interest given the role of VEGF-A across different malignancies as well as the ability to inhibit VEGF-A clinically using bevacizumab[41]. While not currently a standard of care agent in PDAC, this finding further demonstrates the clinical relevance of this approach for future mechanistic evaluations and novel target discovery.

While the strength of combining our computational tools and PDO models is in the dual modalities and flexibility of the organoid model system, there are also shortcomings to this approach. Our collated atlas is limited to 61 patients, with limited representation of some cell types. This presents a challenge when trying to truly understand intercellular dynamics. Similarly, while our organoid co-culture model is established to represent the epithelial and fibroblast

compartments of the tumor, immune cell representation is currently lacking. The data presented here demonstrate the need for immune cell inclusion in future model iterations, particularly when asking questions related to the tumor immune microenvironment or mechanisms of response or resistance to immunotherapy. Additionally, the complexity of the organoid co-culture system and multiple cell type culture condition parameters adds numerous considerations when establishing and validating a complex model system. We have prioritized an additive model in which all components are established with adequate biomass and tested for basic cell type identity prior to co-culture, but this is an additional technical constraint. With these limitations in mind, we advocate for a complementary approach that leverages the high-dimensional data available in an atlas or in additional future cohorts with the capacity for mechanistic inquiries available in the PDO co-culture model system.

In summary, we introduce a bidirectional approach leveraging scRNA-seq data and PDO models to examine patterns of inflammation in PDAC. With that, we identified patterns of inflammation in epithelial cells, both in atlas samples and in PDO models. Additionally, while MHC-II expression is strongly induced in response to IFNγ treatment, these genes are also activated in epithelial cells through co-culture with fibroblasts. Further, the multi-gene signature that CoGAPS detected associated with inflammatory signaling and EMT in epithelial cells (Pattern 7) was enriched in our co-culture PDO system, suggesting that the co-culture system can recapitulate fibroblast induction of inflammatory processes in epithelial cells in PDAC. Similarly, EMT has been previously linked to inflammation in PDAC cells and CAFs through an IFNγ response[6]. Additional studies are needed to distinguish whether this inflammatory process is MHC-II-independent or if inflammatory signaling through fibroblasts induces MHC-II expression to a lesser extent than direct activation in response to IFNγ treatment. Still, our *in silico* modeling using Domino offers intercellular interactions through VEGF-A and ITGB1 as potential mechanisms that underlie the crosstalk between epithelial cells and fibroblasts that contribute to tumor inflammation in PDAC.

Methods

Organoid and CAF co-culture and single-cell analysis

Patients with PDAC undergoing endoscopic biopsy or surgical resection were enrolled in IRB-approved tissue acquisition protocols at Johns Hopkins Hospital and Massachusetts General Hospital (MGH) (NCT03563248). Patient-derived organoids (PDOs) were generated from patient surgical specimens following a combination of mechanical and enzymatic dissociation as previously described[11]. CAFs were extracted from surgical resection specimens after straining remnant tissue through a 70µm cell strainer and washed twice with human organoid wash media (Advanced

DMEM/F12, 10mM HEPES, 1x GlutaMAX, 100µg/mL Primocin, 0.1% BSA) with centrifugation between washes. For co-culture, organoids were combined with patient-matched CAFs in Matrigel (Corning, 356234) at a 1:10 ratio of organoids to CAFs. In parallel, CAFs and PDOs were plated separately in Matrigel. Co-cultures and monocultures were plated in 24-well tissue culture dishes and extracted after 12 hours using Cell Recovery Solution (Corning, 354253) and incubated on ice at 4°C for 45 minutes for Matrigel depolymerization. Cells were then pelleted and washed in human organoid wash media prior to pelleting again. Organoids were dissociated to single cells using TrypLE Express (ThermoFisher Scientific, 12604013) following manufacturer instructions. Single-cells were barcoded using the MULTI-seq protocol as previously described[28]. Single-cell transcriptomics library prep was completed using the 10x Genomics Chromium Single Cell 3' Gene Expression Dual Index Library (V3.1) according to manufacturer specifications. Library preparations quality were analyzed using the 2100 Bioanalyzer (Agilent). Sequencing was completed at the Johns Hopkins Genetic Resources Core Facility (GCRF). Cellranger (V6.0.0) was used to generate the feature-barcode matrices, aligned to the hg38 genome. Multiseq10x (V1.0) was used as the preprocessing pipeline companion to split the MULTI-seq FASTQs into cell barcode, unique molecular identifiers (UMI), and sample barcode sequences. Reads that did not align with >1 mismatch to any reference sequence and reads representing duplicated UMIs on a cell-by-cell basis were removed. Demultiplex (V1.0.2) was used for demultiplexing the data. The 3DGE data were log normalized, linear dimension was reduced using principle component analysis, and differentially expressed genes were identified in Seurat by Wilcoxon Rank Sum Test (V4.0.1). Additional annotations of Moffitt classifiers, denoting classical and basal epithelial subtypes, and CAF subtypes were added to the Seurat object metadata based on the clustering[29]. Co-culture cell types were parsed based on these annotations and the barcode distinctions. Projection of the discovered CoGAPS Pattern 7 onto the 12HR MULTI-seq expression data was completed using ProjectR (V1.8.0). The MULTI-seq expression data and CoGAPs feature loadings were run through the projectR function of the package[31]. The projection results were combined with the MULTI-seq metadata and plotted using ggplot2 (V3.3.5) and Wilcoxon results added using ggpubr (V0.4.0)[31].

Flow Cytometry and Cell Sorting

Organoids were extracted using Cell Recovery Solution (Corning, 354253) and incubated on ice at 4°C for 45 minutes for Matrigel depolymerization. Cells were pelleted and washed in human organoid wash media. Organoids were dissociated to single cells using TrypLE Express and washed in MACS buffer (PBS + 5 mM EDTA + 1% Fetal bovine serum). Cells were resuspended in PBS + Zombie NIR catalog no. 423106 (dilution 1:1000) + Human TruStain FcX

catalog no. 422302 (dilution 1:100) for 10 minutes at room temperature in the dark. Cells were quenched with MACS buffer, spun down, and then resuspended in surface stain for 20 minutes on ice at 4°C in the dark; antibodies were purchased from Biolegend, APC EpCAM catalog no. 324208 (dilution 1:200), PE/Cy7 HLA-A, B, C catalog no. 311429 (dilution 1:200), AF700 HLA-DR catalog no. 307626 (dilution 1:200), PerCP/Cy5.5 PD-L1 catalog no. 329738 (dilution 1:100). Cells were washed twice in MACS buffer. Flow cytometry analyses were performed on the Beckman Coulter Cytoflex.

For co-culture cell sorting, 1mL of 1mg/mL Dispase II (Thermofisher, 17105041) in organoid wash media was added to each coculture and monoculture dome to depolymerize Matrigel for 1 hr at 37°C. Digest was quenched with 1mL of wash media and cells were spun down. Cells resuspended in PBS + Zombie NIR catalog no. 423106 (dilution 1:1000) + Human TruStain FcX catalog no. 422302 (dilution 1:100) for 10 minutes at room temperature in the dark. Cells were quenched with MACS buffer, spun down, and then resuspended in surface stain APC EpCAM catalog no. 324208 (dilution 1:200) and FAP R&D Systems, catalog no FAB3715P-100 (dilution 1:75) on ice at 4°C for 20 minutes in the dark. Cells were washed twice in MACS buffer and filtered through 70um filter. Cell sorting was performed on BD Fusion Sorter.

Quantitative PCR (qPCR)

To evaluate gene expression of MHC-II genes in patient-derived organoids, total RNA extraction using the RNeasy Mini Kit (Qiagen, - Catalog Number: 74104) was completed for each patient-derived organoid line according to manufacturer specifications. cDNA synthesis was performed using Invitrogen TaqMan Reverse Transcription Reagents (Catalog Number: N8080234), following manufacturer's instructions. Real-time quantitative PCR was completed using the ThermoFisher Taqman Gene Expression Assays according to manufacturer's protocol in the QuantStudio 6 Flex System (Applied Biosystems) mRNA targets included: ITGB1 (Hs01127536_m1), VEGFA (Hs00900055_m1), HLA-DRA (Hs00219575_m1), HLA-DRB1 (Hs04192464), HLA-DQB1 (Hs03054971_m1), and HLA-DPB1 (Hs03045105_m1). Relative gene expression was quantified using the $2^{-\Delta\Delta Ct}$ method as previously described[42], and GAPDH (Hs02786624_g1) was used as the endogenous control. Data were analyzed using Applied Biosystems QuantStudio™ Real Time PCR System Software (V1.7.1).

scRNA-seq dataset integration and harmonization for the PDAC atlas

The six different datasets provided gene expression data with different versions (GRCh37 or GRCh38) and nomenclatures (Ensembl identifiers vs. HUGO gene nomenclature) of the human reference genome. Available patient metadata are summarized in Table 1. All analyses were performed in R (V 3.6-4.1) or Python (version 3.8). Unified integration of the measured features revealed 15,219 genes that could be matched between all datasets with assured certainty. Next, cells with unfavorable quality, defined as mitochondrial counts >15% and unique features of <50 or >5,000, were removed. Computational pre-processing was performed with the Monocle3[43] R package. Dimensionality reduction into a unified manifold approximation and projection (UMAP) was based on the first 100 principal components and batch correction was applied per manuscript to account for potential dataset-intrinsic biases (technical or biological) using Batchelor R as utilized by the Monocle3 pipeline[44]. Annotation of cell types is described in detail in the Supplemental Methods. The distributions of epithelial and fibroblast populations and patient-level correlations across epithelial and CAF subtypes are further methodologically detailed and illustrated in Supplemental Fig. 10-12. Plotting was performed with the ggplot2 R package and Excel (Microsoft, Redmond, WA). For high-performance computing tasks, we leveraged the MARCC (Maryland Advanced Research Computing Center, Baltimore, MD) and AWS (Amazon Web Services, Seattle, WA) servers.

CoGAPs analysis of expression patterns

Non-negative matrix factorization (NMF) of transcript counts was conducted using CoGAPS (V 3.5.8)[24,36]. Given a matrix of single-cell data with normalized expression values, CoGAPS factorizes this matrix into two related matrices of gene weights (amplitude matrix) and sample weights (pattern matrix) for random subsets of the data based on the nsets parameter followed by relearning of the amplitude matrix on the full dataset. CoGAPS was run on log2 transformed counts of 15,176 genes from 25,442 cells in Peng et al and Steele et al annotated as epithelial_normal, epithelial_cancer, or epithelial_unspecified[16,17].

Standard parameters were set to 8 Patterns, 50,000 iterations, seed 367, sparse optimized, and distributed: “Single-Cell”. Sparsity parameters were alpha = 0.01, max Gibb mass 100. Distributed CoGAPS parameters were 15 nSets, cut 10, minNS 8, maxNS 23.

Marker genes for each pattern were identified using the patternMarkers function in CoGAPS (V3.9.5) with the “cut” threshold to provide subsets of the top-ranking genes associated with each pattern[45]. Overrepresentation analysis was then conducted using the fora function in the fgsea R package (V1.18.0) to find enrichment of any hallmark gene sets

from the Molecular Signatures Database[26,46] among the pattern markers for each CoGAPS pattern. The universe used in the overrepresentation analysis was all human genes with HGNC symbols in the GRCh38.p13 genome assembly (n = 39,535)[47,48].

Inference of transitions in cellular phenotypes and intercellular interactions

Within each cell group, additional analyses were performed to compute heterogeneity of cellular phenotypes, state transitions, and inter-cellular signaling across the atlas datasets. First, cell cycle scores and phases were computed with tricycle (V1.2.0)[49]. Further unsupervised exploratory analysis of transitions in epithelial cell states was performed with CoGAPS (V3.5.8)[24] analysis across epithelial populations in tumor and normal samples from Peng et al[16] and Steele et al[17]. Single cell CoGAPs was run for 8, 10, and 12 patterns. Eight patterns were selected as the final analysis because 12 patterns returned 10 patterns suggesting an overfitting of the data. Further the 8-pattern run resulted in all 8 patterns that were analogous with the other patterns found in the higher dimensional runs. Finally, the impact of fibroblast cells on epithelial cells was computed by estimating intercellular signaling with Domino (V0.1.1)[32] independently for each of the datasets in the atlas.

For Domino analysis, pyScenic (V0.11.0) for Python was first used to generate the gene regulatory network and co-expression modules, the regulon predictions, and the area under the curve (AUC) matrix of cellular enrichment[50]. This was completed by providing the extracted counts matrix, a list of transcription factors, motif annotations, and cisTarget motifs for the hg38 genome[50]. With the use of the AUC and regulon predictions, a domino object is created and the signaling network built. This allowed for the visualization of global signaling network, gene networks and incoming signaling heatmaps for each narrow subtype annotation, a heatmap of the correlation between transcription factors and receptors, and lastly, the global transcription factor-ligand-receptor network between all subtype annotations[32].

Acknowledgements: The authors would like to thank Dr. Chris McGinnis for his technical expertise in the completion of the MULTI-seq experiment. We would also like to thank the authors of the 6 datasets used in the generation of the atlas: the laboratories of Dr. Hector Alvarez, Dr. Haiyong Han, Dr. Marina Pasca di Magliano, Dr. David Tuveson, Dr. Wenming Wu, and Dr. Itai Yanai. We would like to specifically think Dr. Ela Elyada and Dr. Filip Bednar for their expertise with their specific datasets. Sequencing was completed through the Genetic Resources Core Facility, RRID:SCR_018669. Finally, we are grateful to Aviva Fertig for sharing her mommy with us during a pandemic.

Data Availability Statement: Submission of the RNA-seq data to dbGaP is in process. All analysis scripts are available from: https://github.com/fertiglab/

References

1. Ho, W. J., Jaffee, E. M. & Zheng, L. The tumour microenvironment in pancreatic cancer - clinical challenges and opportunities. *Nat. Rev. Clin. Oncol.* **17**, 527–540 (2020).
2. Peran, I., Madhavan, S., Byers, S. W. & McCoy, M. D. Curation of the Pancreatic Ductal Adenocarcinoma Subset of the Cancer Genome Atlas Is Essential for Accurate Conclusions about Survival-Related Molecular Mechanisms. *Clin. Cancer Res.* **24**, 3813–3819 (2018).
3. Chen, K. *et al.* Single-cell RNA-seq reveals dynamic change in tumor microenvironment during pancreatic ductal adenocarcinoma malignant progression. *EBioMedicine* **66**, 103315 (2021).
4. Hosein, A. N. *et al.* Cellular heterogeneity during mouse pancreatic ductal adenocarcinoma progression at single-cell resolution. *JCI Insight* **5**, (2019).
5. Carstens, J. L. *et al.* Stabilized epithelial phenotype of cancer cells in primary tumors leads to increased colonization of liver metastasis in pancreatic cancer. *Cell Rep.* **35**, 108990 (2021).
6. Ligorio, M. *et al.* Stromal Microenvironment Shapes the Intratumoral Architecture of Pancreatic Cancer. *Cell* **178**, 160-175.e27 (2019).
7. Raghavan, S. *et al.* Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. *Cell* **184**, 6119-6137.e26 (2021).
8. Mizutani, Y. *et al.* Meflin-Positive Cancer-Associated Fibroblasts Inhibit Pancreatic Carcinogenesis. *Cancer Res.* **79**, 5367–5381 (2019).
9. Helms, E., Onate, M. K. & Sherman, M. H. Fibroblast heterogeneity in the pancreatic tumor microenvironment. *Cancer Discov.* **10**, 648–656 (2020).
10. Elyada, E. *et al.* Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discov.* **9**, 1102–1123 (2019).
11. Öhlund, D. *et al.* Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J. Exp. Med.* **214**, 579–596 (2017).
12. Suri, R., Zimmerman, J. W. & Burkhart, R. A. Modeling human pancreatic ductal adenocarcinoma for translational research: current options, challenges, and prospective directions. *Ann. Pancreat. Cancer* **3**, 17–17 (2020).
13. Pandha, H., Rigg, A., John, J. & Lemoine, N. Loss of expression of antigen-presenting molecules in human pancreatic cancer and pancreatic cancer cell lines. *Clin. Exp. Immunol.* **148**, 127–135 (2007).
14. Schmitz-Winnenthal, F. H. *et al.* Specific immune recognition of pancreatic carcinoma by patient-derived CD4 and CD8 T cells and its improvement by interferon-gamma. *Int. J. Oncol.* **28**, 1419–1428 (2006).
15. Scupoli, M. T. *et al.* Expression of MHC class I and class II antigens in pancreatic adenocarcinomas. *Tissue Antigens* **48**, 301–311 (1996).
16. Peng, J. *et al.* Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* **29**, 725–738 (2019).
17. Steele, N. G. *et al.* Multimodal Mapping of the Tumor and Peripheral Blood Immune Landscape in Human Pancreatic Cancer. *Nat Cancer* **1**, 1097–1112 (2020).
18. Lin, W. *et al.* Single-cell transcriptome analysis of tumor and stromal compartments of pancreatic ductal adenocarcinoma primary tumors and metastatic lesions. *Genome Med.* **12**, 80 (2020).
19. Moncada, R. *et al.* Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* **38**, 333–342 (2020).
20. Bernard, V. *et al.* 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 (2019).
21. Seppälä, T. T. *et al.* Patient-derived Organoid Pharmacotyping is a Clinically Tractable Strategy for Precision Medicine in Pancreatic Cancer. *Ann. Surg.* **272**, 427–435 (2020).
22. Seppälä, T. T. *et al.* Precision medicine in pancreatic cancer: Patient derived organoid pharmacotyping is a predictive biomarker of clinical treatment response. *Clin. Cancer Res.* OF1–OF12 (2022).
23. Dijkstra, K. K. *et al.* Generation of Tumor-Reactive T Cells by Co-culture of Peripheral Blood Lymphocytes and Tumor Organoids. *Cell* **174**, 1586-1598.e12 (2018).
24. Sherman, T. D., Gao, T. & Fertig, E. J. CoGAPS 3: Bayesian non-negative matrix factorization for single-cell analysis with asynchronous updates and sparse data structures. *BMC Bioinformatics* **21**, 453 (2020).
25. Stein, -O' & Brien, G. L. PatternMarkers & GWCoGAPS for novel data-driven biomarkers via whole transcriptome NMF. *Bioinformatics* **33**, 1892–1894 (2017).
26. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).
27. Wei, K., Nguyen, H. N. & Brenner, M. B. Fibroblast pathology in inflammatory diseases. *J. Clin. Invest.* **131**, (2021).
28. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).

29. Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**, 1168–1178 (2015).
30. Krieger, T. G. *et al.* Single-cell analysis of patient-derived PDAC organoids reveals cell state heterogeneity and a conserved developmental hierarchy. *Nat. Commun.* **12**, 5826 (2021).
31. Sharma, G., Colantuoni, C., Goff, L. A., Fertig, E. J. & Stein-O'Brien, G. projectR: an R/Bioconductor package for transfer learning via PCA, NMF, correlation and clustering. *Bioinformatics* **36**, 3592–3593 (2020).
32. Cherry, C. *et al.* Computational reconstruction of the signalling networks surrounding implanted biomaterials from single-cell transcriptomics. *Nat Biomed Eng* **5**, 1228–1238 (2021).
33. Xie, Y. *et al.* Naive tumor-specific CD4+ T cells differentiated in vivo eradicate established melanoma. *J. Exp. Med.* **207**, 651–667 (2010).
34. Davis-Marcisak, E. F. *et al.* Transfer learning between preclinical models and human tumors identifies a conserved NK cell activation signature in anti-CTLA-4 responsive tumors. *Genome Med.* **13**, 129 (2021).
35. Stein-O'Brien, G. L. *et al.* Decomposing Cell Identity for Transfer Learning across Cellular Measurements, Platforms, Tissues, and Species. *Cell Syst* **8**, 395-411.e8 (2019).
36. Fertig, E. J., Ding, J., Favorov, A. V., Parmigiani, G. & Ochs, M. F. CoGAPS: an R/C++ package to identify patterns and biological process activity in transcriptomic data. *Bioinformatics* **26**, 2792–2793 (2010).
37. Nicolet, B. P., Guislain, A. & Wolkers, M. C. CD29 enriches for cytotoxic human CD4+ T cells. *J. Immunol.* **207**, 2966–2975 (2021).
38. Alspach, E. *et al.* MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* **574**, 696–701 (2019).
39. Maloney, J. P. & Gao, L. Proinflammatory Cytokines Increase Vascular Endothelial Growth Factor Expression in Alveolar Epithelial Cells. *Mediators Inflamm.* **2015**, 387842 (2015).
40. Erez, N., Truitt, M., Olson, P., Arron, S. T. & Hanahan, D. Cancer-Associated Fibroblasts Are Activated in Incipient Neoplasia to Orchestrate Tumor-Promoting Inflammation in an NF-kappaB-Dependent Manner. *Cancer Cell* **17**, 135–147 (2010).
41. Garcia, J. *et al.* Bevacizumab (Avastin®) in cancer treatment: A review of 15 years of clinical experience and future outlook. *Cancer Treat. Rev.* **86**, 102017 (2020).
42. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).
43. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
44. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
45. Stein-O'Brien, G. L. *et al.* PatternMarkers & GWCoGAPS for novel data-driven biomarkers via whole transcriptome NMF. *Bioinformatics* **33**, 1892–1894 (2017).
46. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
47. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).
48. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
49. *tricycle*. (Github).
50. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

	Peng et al	Steele et al	Lin et al	Elyada et al	Moncada et al	Bernard et al
Year of publication	2019	2020	2020	2019	2020	2019
Country of enrollment	China	USA	South Korea, USA	USA	USA	USA
PDAC pancreas tissue samples						
Treatment-naïve, %	100%	100%	100%	100%	100%	100%
Patients, N	24	16	10	6	3	2
Sex, N						
Female	13	6	4	2	Unknown	0
Male	11	10	6	4	Unknown	2
Age, years median (range)	59 (36-72)	66 (42-80)	66 (41-80)	73 (64-87)	Unknown	61 (59-62)
Diabetes, N	10	6	Unknown	Unknown	Unknown	Unknown
Staging, N						
Localized	24	12	10	5	Unknown	1
Metastatic	0	4	0	1	Unknown	1
Grading, N						
Well differentiated	3	Unknown	0	0	Unknown	0
Moderately differentiated	8	≥6	6	3	Unknown	0
Poorly differentiated	13	Unknown	2	3	Unknown	2
Anaplastic	0	Unknown	2	0	Unknown	0
Tumor location, N						
Head or uncinate	15	Unknown	Unknown	Unknown	Unknown	1
Neck, body or tail	9	Unknown	Unknown	Unknown	Unknown	1
Specimen, N						
Surgery (resected)	24	6	10	6	3	2
Biopsy (FNA)	0	10	0	0	0	0
Survival	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
Non-malignant pancreas tissue samples						
Patients, N	11	3	0	2	0	0
Sex, N						
Female	6	2	---	1	---	---
Male	5	1	---	1	---	---
Histology, N						
Normal	10	1	---	0	---	---
Normal-adjacent to adenocarcinoma	1	2	---	2	---	---
Specimen, N						
Surgery (resected)	11	3	---	2	---	---
Biopsy (EUS-FNA)	0	0	---	0	---	---

PDAC, pancreatic ductal adenocarcinoma
EUS-FNA, endoscopic ultrasound guided-fine needle aspiration

Table 1. Clinical patient data extracted from the six harmonized published datasets integrated in the PDAC atlas.

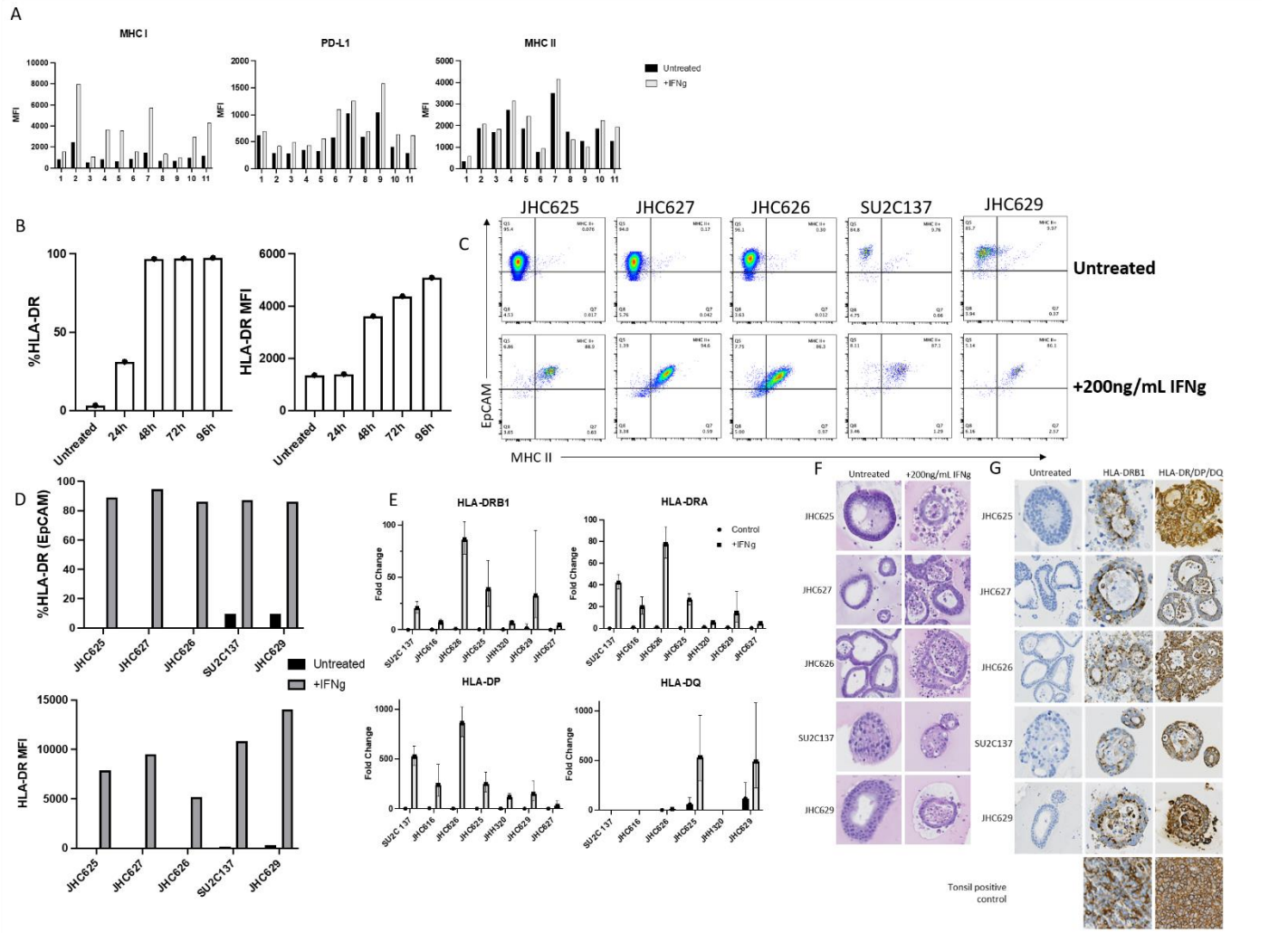

Figure 1. MHC-II expression in PDAC PDOs. (A) Cell surface MHC-I, PD-L1, and MHC-II expression as determined by flow cytometry. Organoids were stimulated with 200ng/mL of IFNγ for 24hr or were left unstimulated. Bar graphs represent Mean Fluorescence Intensity (MFI). (B) HLA-DR (MHC-II) expression as determined by flow cytometry is increased after IFNγ stimulation for 24-96 hours shown both by % increase (left panel) and by increase in MFI (right panel) across one PDO line. (C) Representative flow cytometry dot plots of 5 organoid lines that were treated with 200ng/mL of IFNγ for 96 hours or left unstimulated. (D) Bar graph quantification of plots in C, bar graphs represent % increase in HLA-DR (MHC-II) expression of EpCAM+ cells (top panel) and Mean Fluorescence Intensity (MFI) (bottom panel). (E) RT-PCR readout of organoid lines that were stimulated with 200ng/mL of IFNγ for 96 hours or left unstimulated; genes quantified include HLA-DRA/HLA-DRB1/HLA-DPB1/HLA-DQB1. P-value for HLA-DRA/HLA-DRB/HLA-DP when tested between control and treated groups = 0.0006, p-value for HLADQ = 0.4 (statistical analysis was performed using a two-tailed, unpaired, parametric Mann-Whitney test). (F) H&E slides of PDOs with and without IFNγ induction. Organoids were harvested 96 hours after IFNγ induction. (G) Immunohistochemistry (IHC) demonstrating expression of HLA-DRB1 (middle) and antibody binding to the common beta chain of HLA-DR/DP/DQ (right) after 96 hours of IFNγ stimulation.

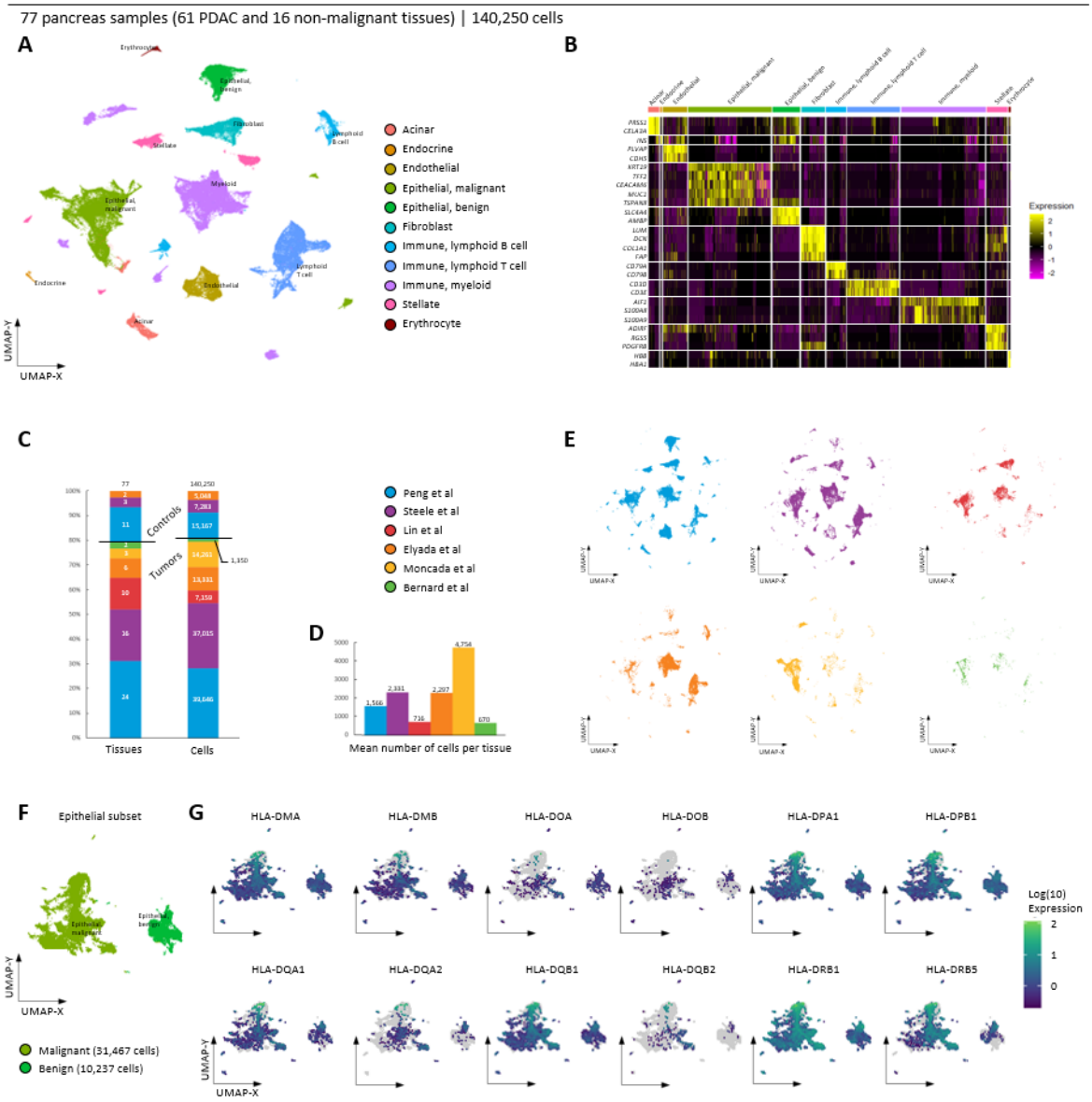

Figure 2. Summary of atlas composition and evaluation of MHC-II gene expression in atlas samples. (A) Complete atlas with assigned cell types. (B) Heatmap of differentially expressed genes used for cell type annotations. (C) Relative contribution of the 77 different samples with 140,250 cells, separated by tumor (below line) and control tissue (above line). (D) Mean number of cells per tissue by dataset origin. (E) Cell mapping by dataset origin from the six manuscripts in the complete atlas. (F) Epithelial cell clusters from patient tumor samples. (G) Expression of 12 MHC-II genes was queried in the atlas demonstrating variable endogenous expression of each gene. UMAP: Uniform Manifold Approximation and Projection.

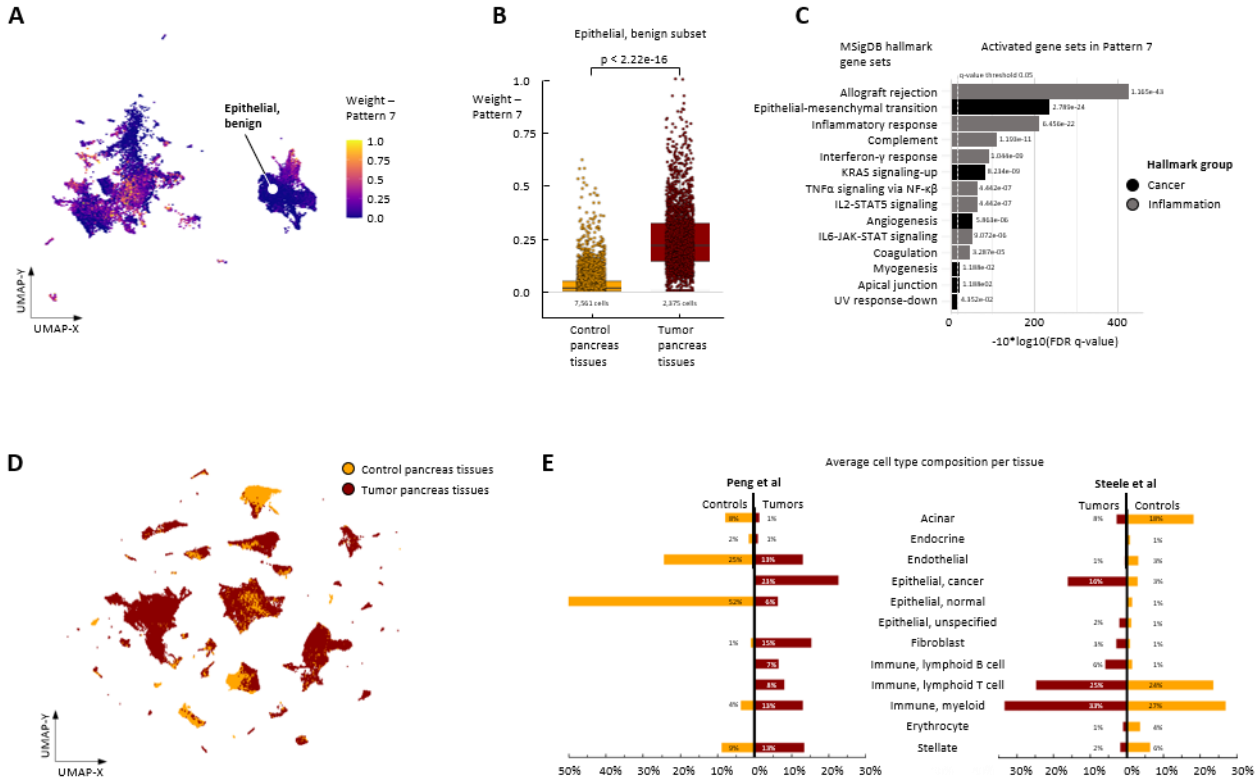

Figure 3. Identification of pattern of inflammation (Pattern 7) in the atlas and cell type distribution in the atlas. (A) Complete atlas subset of the epithelial cell populations (malignant, benign and unspecified) from the Peng et al[16] and Steele et al[17] data with assigned weights of Pattern 7 as resulting from CoGAPS analyses. (B) Boxplot of Pattern 7 weights within the Epithelial, benign cell population demonstrating differences between control and tumor pancreas tissues. $p < 2.22$ e-16, generated by Wilcoxon test. (C) Overrepresented MSigDB hallmark gene sets in cells expressing Pattern 7 genes. (D) Cell mapping by tumor vs. control pancreas tissue in the complete atlas. (E) Differences in cell type composition for selected tumor vs. control pancreas tissues originating from Peng et al[16] and Steele et al[17].

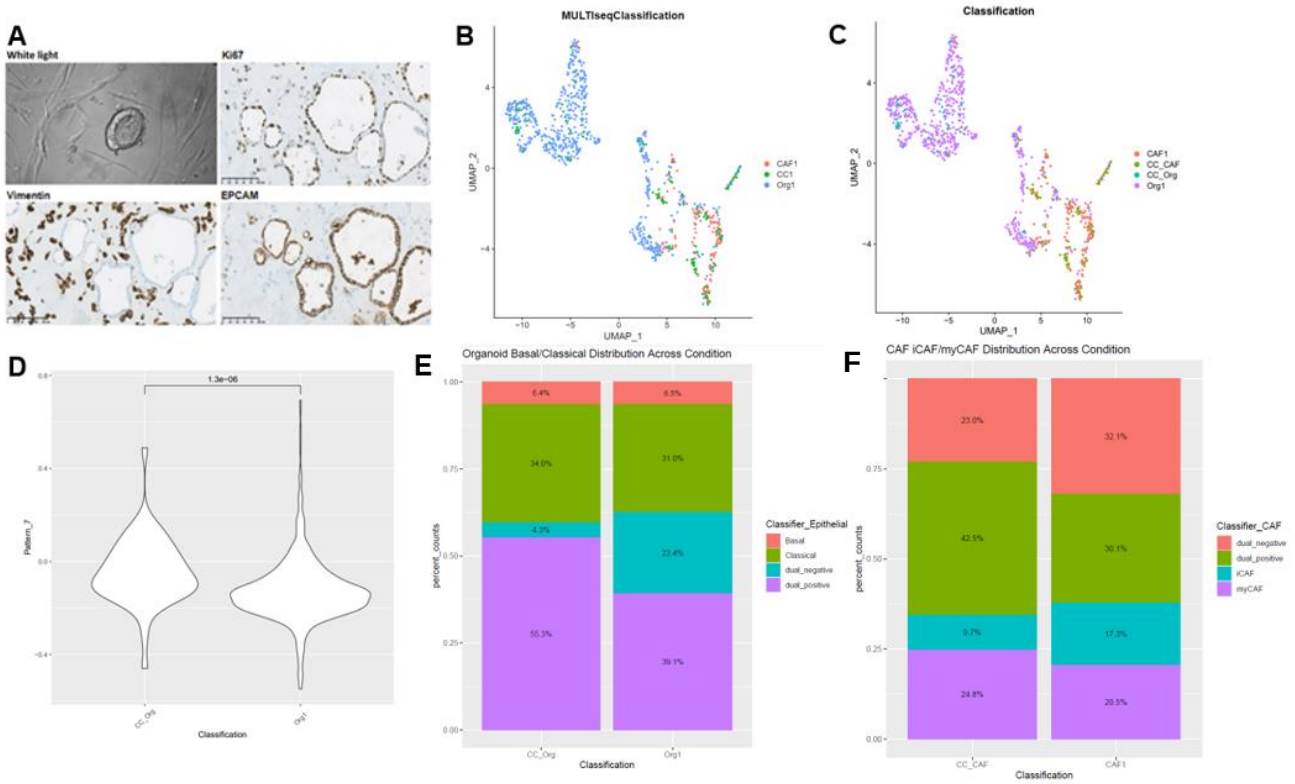

Figure 4. Patient-derived organoids co-cultured with CAFs recapitulate the inflammatory pattern identified in tumor epithelial cells and demonstrate dynamic cellular phenotypes. (A) Representative brightfield image of co-culture at 20x magnification. Representative IHC of co-culture demonstrating proliferation after co-culture (top right), vimentin positive CAFs (bottom left), and EpCAM positive organoids (bottom right). Images obtained at 20x magnification. (B) UMAP demonstrating culture conditions: organoid monoculture (Org1), CAF monoculture (CAF1), co-culture (CC1). (C) UMAP demonstrating cell-type calls after co-culture: Organoid monoculture (Org1), CAF monoculture (CAF1), CAFs from co-culture (CC_CAF), organoids from co-culture (CC_Org). (D) Pattern 7 is composed of inflammatory genes and was enhanced in organoid cells from co-culture relative to organoid cells from monoculture, p=1.3e-6. (E) Co-culture demonstrates dynamic epithelial representation in the co-culture condition with a greater percentage of cells representing both basal and classical markers (dual_positive) present in co-culture. (F) Co-culture demonstrates dynamic CAF representation in the co-culture condition with a greater percentage of cells representing both iCAF and myCAF markers (dual_positive) present in co-culture.

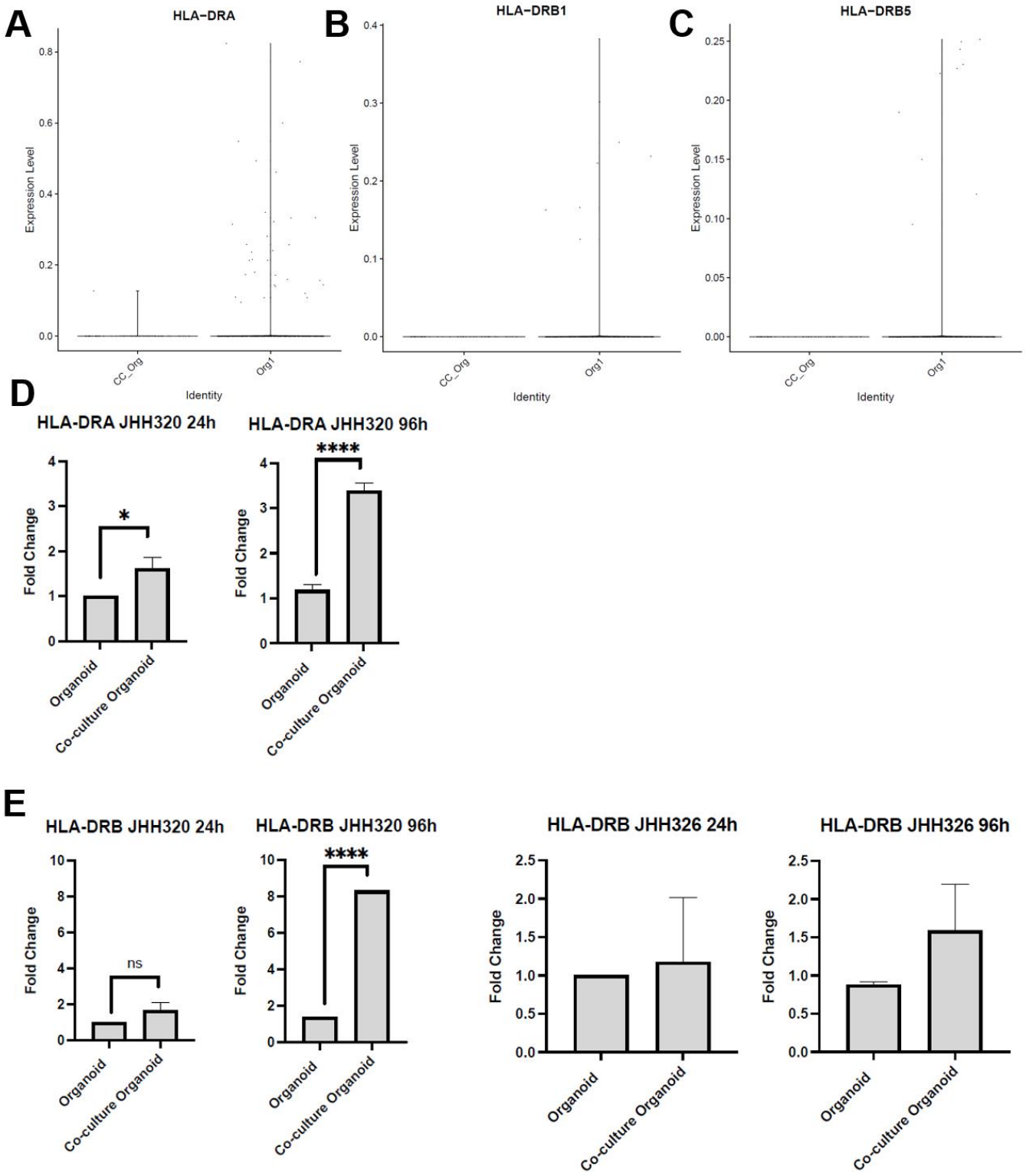

Figure 5. Evaluation of MHC-II expression in the co-culture in response to the presence of CAFs. Expression of (A) HLA-DRA, (B) HLA-DRB1, and (C) HLA-DRB5 after co-culture for 12 hours evaluated using MULTI-seq. Expression is limited in all populations. (D) HLA-DRA and (E) HLA-DRB expression after 24 and 96 hours of co-culture after which cells were flow sorted prior to qPCR. HLA-DRA did not amplify in JHH 326 at either 24 or 96 hours. HLA-DRB demonstrated less consistent amplification limiting statistical analysis. Co-cultures were established from two patients for which there were matched PDOs and CAFs. Plotted are the Fold Change values comparing our PDO co-culture to monoculture using GAPDH as an endogenous control. Comparisons of monoculture and co-culture conditions are statistically supported using the two-tailed students t-test with equal variance in PRISM (V9.2.0 [283]). Significance is measured as: ****, $p<0.0001$; ***, $p<0.001$; **, $p<0.01$; *, $p<0.05$; ns, not significant.

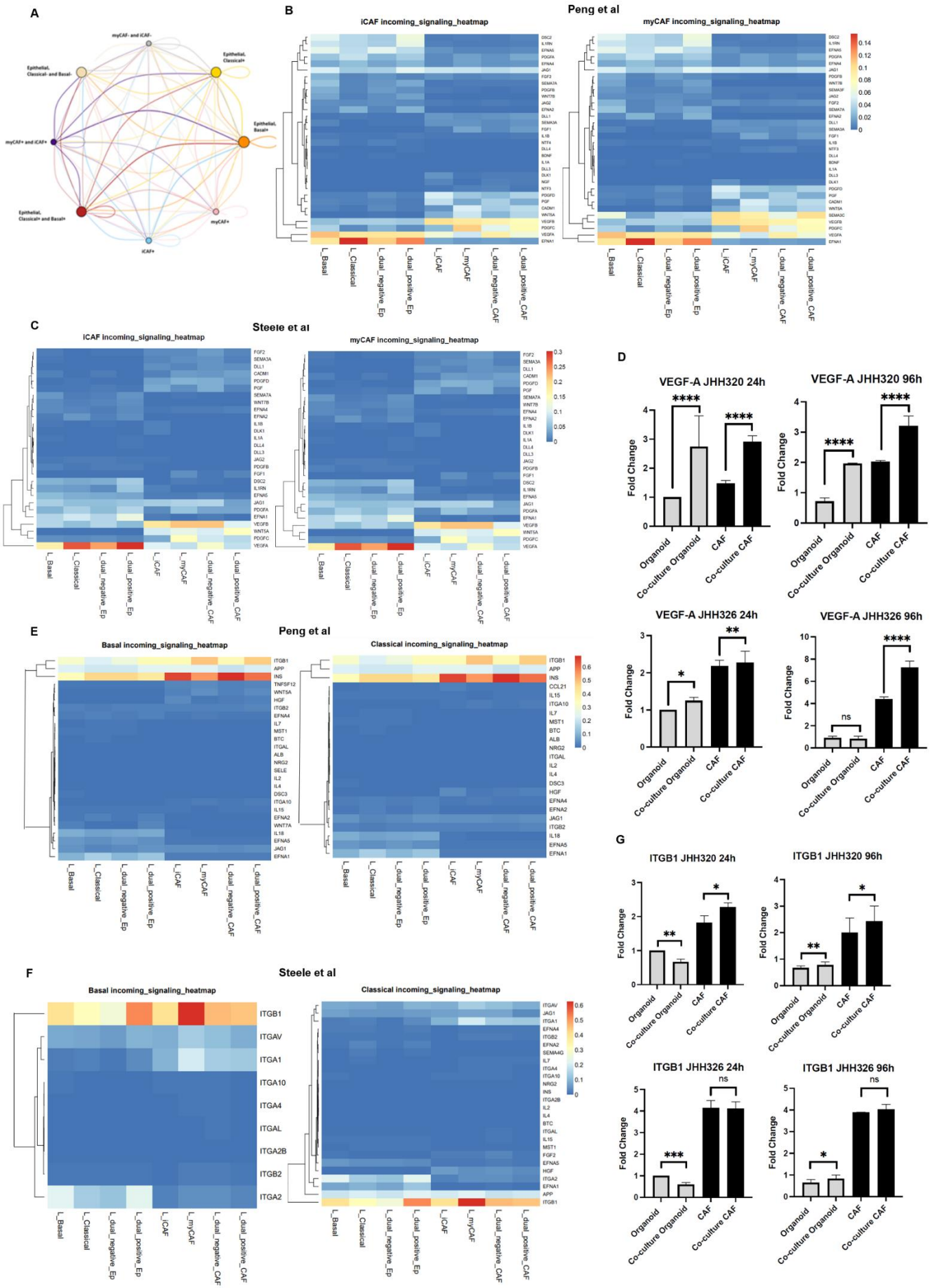

Figure 6. Domino evaluation of intercellular interactions in the atlas with validation in PDO-CAF co-culture. (A) Signaling network between epithelial and CAF subpopulations from tumor pancreas tissues in the Peng et al[16] dataset as derived from the Domino R package. Nodes of the subpopulations are sized according to the amounts of expressed

targeting ligands. The thicknesses of the intercellular connections are scaled based on the strength of signaling with their color indicating the signals' origin (directionality). (B) Heatmap demonstrating *VEGF-A* as a ligand originating in the epithelial populations with the CAFs receiving this signal from the Peng[16] and (C) Steele[17] datasets. (D) Differential expression of *VEGF-A* in monoculture and co-culture CAFs and epithelial cells after 24 or 96 hours of co-culture. (E) Heatmap demonstrating *ITGB1* as a ligand originating in the CAF populations with the epithelial cells receiving this signal from the Peng et al[16] and (F) Steele et al[17] datasets. (G) *ITGB1* expression in monoculture and co-culture CAFs and epithelial cells after 24 or 96 hours of co-culture. Plotted are the Fold Change values comparing our PDO co-culture to monoculture using GAPDH as an endogenous control. Comparisons of monoculture and co-culture conditions are statistically supported using the two-tailed students t-test with equal variance in PRISM (V9.2.0 [283]). Significance is measured as: ****, $p<0.0001$; ***, $p<0.001$; **, $p<0.01$; *, $p<0.05$; ns, not significant.