Published in final edited form as:

Gut. 2019 June ; 68(6): 1034–1043. doi:10.1136/gutjnl-2018-317706.

Experimental microdissection enables functional harmonization of pancreatic cancer subtypes

C Maurer^{#,1,2,3}, SR Holmstrom^{#,1,2,3}, J He^{#,1,4,5}, P Laise^{1,4,5}, T Su^{1,3}, A Ahmed^{1,3}, H Hibshoosh^{1,5}, JA Chabot^{1,6}, PE Oberstein⁷, AR Sepulveda^{1,5}, JM Genkinger^{1,8}, J Zhang⁹, AC Iuga^{1,5}, M Bansal¹⁰, A Califano^{1,4,5,†}, and KP Olive^{1,2,3,†}

¹Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, New York, USA.

²Department of Medicine, Division of Digestive and Liver Diseases, Columbia University Medical Center, New York, New York, USA.

³Department of Pathology and Cell Biology, Columbia University Medical Center, New York, New York, USA.

⁴Department of Biomedical Informatics, Columbia University Medical Center, New York, New York, USA.

⁵Department Systems Biology, Columbia University Medical Center, New York, New York, USA.

⁶Department of Surgery, Division of GI/Endocrine Surgery, Columbia University Medical Center, New York, New York, USA.

⁷Department of Medicine, Division of Hematology and Oncology, New York University Langone Medical Center, New York, New York, USA.

⁸Department of Epidemiology, Mailman School of Public Health, New York, New York, USA.

⁹Department of Computer Science and Engineering, University of California, San Diego, California, USA.

¹⁰PsychoGenics Inc, Paramus, New Jersey, USA.

Abstract

[†]Correspondence should be addressed to: Kenneth P. Olive, Columbia University Medical Center, 1130 Saint Nicholas Avenue, ICRC 217A New York, NY 10032, phone: 212-851-4678, kenolive@columbia.edu, Andrea Califano, Columbia University Medical Center 1130 Saint Nicholas Avenue, ICRC 912 New York, NY 10032, phone: 212-851-5183, ac2248@cumc.columbia.edu.

[#]contributed equally

Contributors

Conceptualization: KPO, AC, MB. Computational analysis: HCM, JH. Software: JH, JZ, MB. Investigation: HCM, SRH, JH, JG, ACI, ARS, KPO. Resources: JAC., HH, AA, TS, KPO, AC. Visualization: HCM, JH, KPO. Funding acquisition: HCM, KPO, AC. Project oversight and management: KPO, AC, MB. HCM and KPO wrote the manuscript with feedback from SRH, JH and AC. All authors discussed the results and commented on the manuscript.

Genomic data

Transcriptional data generated in this manuscript have been deposited to GEO and are now publically available through the following link: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93326>

Competing interests:

A.C. is a founder and shareholder of DarwinHealth Inc. and a member of the Tempus Inc. SAB and shareholder. Columbia University is a shareholder of DarwinHealth Inc. K.P.O. is a member of the SAB for Elstar Therapeutics.

Objective: Pancreatic ductal adenocarcinoma (PDA) has among the highest stromal fractions of any cancer and this has complicated attempts at expression-based molecular classification. The goal of this work is to profile purified samples of human PDA epithelium and stroma, and examine their respective contributions to gene expression in bulk PDA samples.

Design: We used laser capture microdissection (LCM) and RNA sequencing to profile the expression of 60 matched pairs of human PDA malignant epithelium and stroma samples. We then used these data to train a computational model that allowed us to infer tissue composition and generate virtual compartment-specific expression profiles from bulk gene expression cohorts.

Results: Our analysis found significant variation in the tissue composition of pancreatic tumors from different public cohorts. Computational removal of stromal gene expression resulted in the reclassification of some tumors, reconciling functional differences between different cohorts. Furthermore, we established a novel classification signature from a total of 110 purified human PDA stroma samples, finding two groups that differ in extracellular matrix- and immune-associated processes. Lastly, a systematic evaluation of cross-compartment subtypes spanning four patient cohorts indicated partial dependence between epithelial and stromal molecular subtypes.

Conclusion: Our findings add clarity to the nature and number of molecular subtypes in PDA, expand our understanding of global transcriptional programs in the stroma, and harmonize the results of molecular subtyping efforts across independent cohorts.

Keywords

pancreatic ductal adenocarcinoma; transcriptional deconvolution; molecular classification; Pancreatic Cancer; Gene Expression; Laser Capture Microdissection; Subtypes; Stroma

Introduction

All carcinomas harbor both transformed malignant cells and non-transformed stromal cells, in varying proportions[1]. Pancreatic ductal adenocarcinoma (PDA) is among the most stroma-rich cancers, with a complex inflammatory microenvironment that typically dominates the tumor parenchyma. Expected to be responsible for over 43,000 deaths per year in the United States, it is a common, aggressive malignancy that responds poorly to therapeutic intervention[2, 3]. Within the stromal compartment of PDA, diverse fibroblast, myeloid, lymphoid, endothelial and other cell lineages contribute to both pro- and anti-tumor processes, including angiogenesis and epithelial differentiation[4], tissue stiffness[5, 6], drug delivery[7], and local immunosuppression[8]. These functions are orchestrated through a host of paracrine signals that pass between and within the epithelial and stromal compartments- communication that is quickly altered upon tissue disruption. Thus, efforts to parse transcriptional programs of PDA should take into account active processes in both compartments, ideally in an *in situ* context.

Despite extensive genomic characterization[9, 10, 11, 12, 13], individual DNA mutations have to date provided limited prognostic or theranostic information for PDA. Indeed, only a small fraction of pancreatic tumors is predicted to harbor “druggable” genetic alterations[11, 13]. As an alternative to genetic biomarkers, transcriptional classifiers for PDA have been

explored using bulk tumor samples[13, 14, 15, 16]. While these studies differ in the number of subtypes described, a shared message is that ductal pancreatic tumors include at least two groups distinguished by markers of epithelial differentiation state, with the more poorly-differentiated subtype (i.e. “Basal-like”, “Squamous”, or “Quasi-Mesenchymal”) exhibiting reduced overall survival relative to well-differentiated subtypes (i.e. “Classical” or “Progenitor”). However, the contributions of stromal cells are handled differently in each instance, leading to some debate as to the merits of different proposed subtypes. To clarify this issue, we endeavored to directly profile gene expression from purified neoplastic epithelium and associated stroma isolated from frozen human PDA samples.

Several techniques may be employed to isolate cellular subsets from bulk tissue including magnetic separation, fluorescence assisted cell sorting (FACS), and laser capture microdissection (LCM). The first two techniques rely on population-specific antibodies to isolate specific cell types, but require disruption of the tumor using prolonged enzymatic digestion, during which time transcriptional profiles are invariably altered. Moreover, PDA diffusely infiltrates the surrounding pancreatic parenchyma[17] so that even tumor samples enriched by FACS for epithelial cell markers can include contributions from normal, atrophic, pre-neoplastic, or metaplastic epithelial cells. Laser capture microdissection (LCM) provides a powerful solution, allowing the isolation of pathologically verified compartment-specific tissue samples based on morphological features, without disrupting the delicate interplay of intercellular communication.

We present here expression profiles of laser capture microdissected malignant epithelium and matched reactive stroma for 60 human pancreatic ductal adenocarcinomas, providing both the opportunity to study each compartment in isolation, and to examine their interplay across samples. Furthermore, we provide a novel stromal classification signature derived from direct analysis of a total of 110 experimentally purified stromal profiles, yielding two prominent subtypes. In contrast with a prior signature derived indirectly using blind-source separation techniques[15], this direct signature highlights the contribution of immune signaling pathways in one subtype (Immune-rich), versus extracellular matrix associated pathways (ECM-rich) in the other.

Using the compartment-specific profiles, we used a machine learning technique called ADVOCATE [18] to model the compartment specificity of every gene expressed in PDA. Applying this information to new bulk PDA expression profiles, we can then infer the epithelial and stromal fractions of that tumor and, critically, generate a pair of virtual compartment-specific gene expression profiles for each bulk tumor, which may then be used by a variety of downstream analytical pipelines. Using this approach, we examined the composition of multiple public PDA expression datasets, and inferred both epithelial and stromal molecular subtypes from over 350 human pancreatic tumors. Critically, we found that consideration of compartment specific molecular subtypes led to a harmonization of results across datasets and the validation of functionally similar subtypes that span human pancreatic cancer.

Results

Transcriptional profiling of isolated pancreatic cancer epithelium and stroma

To study the separate transcriptional programs of intact pancreatic tumor epithelium and stroma, we optimized a robust protocol for maintaining RNA integrity during laser capture microdissection of frozen tumor tissues, yielding total RNA suitable for library preparation and RNA sequencing. We first applied this LCM–RNA–Seq technique to 60 primary PDA specimens that were harvested and frozen intraoperatively by the Columbia University Tumor Bank in collaboration with the Pancreas Center at Columbia University/New York Presbyterian Hospital (see Tables S1,2 for patient characteristics). For each tumor, we generated paired gene expression profiles from the malignant epithelium and nearby reactive stroma, as distinguished by cell morphology (Figure 1A). Extensive quality control metrics confirmed the high quality of resulting RNA libraries (Figures 1B,C and Figures S1A–D) [19, 20]. Critically, samples from the two compartments separated spontaneously along the first component of a Principal Component Analysis (PCA) with virtually no overlap (Figure 1D), and were distinguished by expression of established marker genes for epithelial cells (KRT19, EPCAM, CDH1) versus markers of various stromal cell types, including leukocytes (PTPRC, CD4, CD163), endothelial cells (VWF, ENG, CDH5), and cancer associated fibroblasts (CAFs) (ACTA2, DCN, FAP) (Figure 1E). We observed that technical variance was substantially lower than biological variance (Figures S1E–F) and found that different malignant areas captured from a single tumor clustered closely, suggesting that the intra-tumoral transcriptional heterogeneity of that tumor was less than the inter-tumoral heterogeneity of PDA (Figures S1G–H).

We next validated the paired LCM-RNA-seq profiles by assessing the immunohistochemical staining pattern of proteins that were predicted to be highly compartment-specific at the RNA level (Table S3), making use of data from The Human Protein Atlas (HPA) pathology database [21]. We restricted our analysis to proteins for which the highest-quality antibodies were available (n= 321), based on established HPA criteria (Table S4). Of these, we evaluated the immunostaining patterns for the 50 genes whose LCM–RNA–Seq expression was most differentially expressed for each compartment (Table S5), examining a minimum of six PDA samples per tested protein. This analysis yielded confirmatory staining patterns for 47 of 50 epithelial proteins and 36 of the 50 stromal proteins (Figures 1 F, G). For example, Figures 1 H, I show two members of the galectin protein family, LGALS4 and LGALS1, with inverse staining patterns in the two compartments, consistent with our predictions. Critically, none of the proteins were found expressed in a pattern opposite that predicted; rather, genes lacking supportive IHC staining were simply not detected, perhaps due to post-translational regulation. Thus, through the use of LCM–RNA–Seq, we compiled a comprehensive repertoire of compartment-specific genes serving as a novel, tumor-specific resource for the pancreatic cancer field.

Compartment fraction analysis reveals distinct compositions of public PDA datasets

Multiple large-scale gene expression datasets for PDA have been reported[13, 14, 15, 16], each providing important contributions to our understanding of the disease. However, cross-comparative analysis of these datasets has been challenging, due to differences in expression

profiling platforms, inclusion criteria, sample preparation, and other technical details. As a result, a consistent interpretation of the gene expression profile clusters emerging from these studies is still elusive, especially as it relates to stromal subtypes.

We reasoned that there are three potential sources of relevant heterogeneity in these data: 1) differences in the epithelium/stroma ratio in areas of frank carcinoma; 2) variation in the representation of uncharacterized tissue (e.g., normal pancreas ductal epithelium, pancreatitis, lymph nodes, etc.) in the bulk sample; and 3) technical differences (e.g. expression platform, library preparation method). To manage these issues, we made use of a machine learning algorithm called ADVOCATE [18] to model the epithelial and stromal expression of every gene based on the 60 matched epithelial and stromal PDA LCM-RNA-Seq profiles above. After training, ADVOCATE can perform two functions on new bulk PDA expression profiles: 1) infer the fractions of epithelial and stromal tissues that make up the bulk sample; and 2) generate a pair of complete virtual compartment-specific expression profiles for each bulk tumor. Extensive validation of these functions using *in silico* analyses, mixing approaches, paired LCM/bulk profiles, and histopathological evaluations are presented in a preprint on Bioarxiv [22]. We utilized ADVOCATE to perform a systematic analysis of over 350 published PDA expression profiles.

We began by examining the compartment fractions from the gene expression profiles of three independent cohorts: (a) UNC Chapel Hill (UNC, $n = 125$), (b) the International Cancer Genome Consortium (ICGC, PACA-AU RNA-Seq dataset, $n = 93$), and (c) The Cancer Genome Atlas (TCGA, PAAD dataset, $n = 137$) (see Methods for inclusion criteria). The compartment fractions of these cohorts had not previously been directly compared using a single, common analysis method, perhaps due differences in expression platforms (array versus RNA sequencing), or their available metadata. Using ADVOCATE, we found that the epithelial and stromal fractions varied significantly between the cohorts with 46%, 67% and 55% epithelium for the ICGC, UNC and TCGA cohorts, respectively ($p < 0.001$, one-way ANOVA) (Figure 2A). These results align with “tumor purity” analyses performed on the TCGA and ICGC cohorts using DNA-based techniques [13, 16]. Our findings highlight critical differences in tissue composition between tumor collections curated with different inclusion criteria or enrichment practices.

Prior classification efforts built from individual cohorts have yielded divergent gene signatures that stratify pancreatic cancer into various subgroups, leading to ongoing debate as to their relative merits. We realized that our compartment-specific expression data could provide some context as to the nature of the genes that comprise each classification signature. Therefore, we extracted the list of signature genes overexpressed in each of the 11 proposed subtypes in the Collisson, Moffitt, and Bailey classification schemes (Table S6–10). We then examined the overall expression level and compartment-specificity of these genes in our LCM-RNA-seq dataset (Figure 2B) [13, 14, 15]. We noted that the Bailey classifier was developed using Ensembl gene annotation and that the Bailey–Immunogenic subtype includes numerous recombined immunoglobulin genes that are not designated in the NCBI annotation used for the CUMC dataset (Table S10). Therefore, to assess the compartment-specificity of Bailey classifier genes, we used a version of the CUMC dataset that was remapped using the Ensembl GRCh37 gene annotation.

Examining each of the proposed subtype groups in turn, we noted that the genes used to define the Collisson–Classical, Moffitt–Classical, Moffitt–Basal-like, and Bailey–Progenitor subtypes were all heavily weighted towards epithelial expression, suggesting that regardless of the amount of stroma infiltration, these genes are predominantly providing information about the malignant compartment. Conversely, those used to define the Moffitt–Activated, Moffitt–Normal, and Bailey–Immunogenic subtypes were weighted towards stromal expression, suggesting that these subtypes report on information that is largely independent of the malignant compartment. The Collisson–Quasi-Mesenchymal and Bailey–Squamous gene sets were both well-expressed and represented a mixture of epithelial and stromal identity, consistent with a more poorly differentiated state. Finally, the majority of genes that define the Collisson–Exocrine and Bailey–ADEX subtypes exhibited very low expression in the LCM-RNA-Seq datasets, suggesting that their expression in bulk tissue is derived from cell types that are largely absent from our microdissected samples. Together, these data provide insight into the cellular compartments that contribute to previous molecular gene signatures built from bulk tumor tissue samples.

Transcriptional deconvolution improves functional classification across cohorts

An important feature of a robust classification system is its capacity to identify sample subsets that are functionally similar across independent datasets. Given the uncertainty in the current literature regarding the actual number of these subsets in PDA, we first carried out an unsupervised analysis of CUMC epithelial LCM profiles using multiple independent approaches, all of which favored a two cluster solution (Figure S2A–D). Functional annotation of these groups was in agreement with that of the Basal-like and Classical groups (Figure S2E, Table S11,12) described by Moffitt et al. and the respective UNC classifier genes were significantly enriched towards their counterpart among CUMC samples (Figure S2F). This functional alignment with our LCM data together with a superior compartment–specificity (see Figure 2B) led us to prioritize the UNC tumor classifier lists for molecular subtyping of the epithelium across PDA cohorts.

We next examined the relationship between compartment fraction and inferred epithelial subtype in each cohort (Table S13). In the ICGC and UNC cohorts, Basal-like and Classical tumors were inferred to have similar epithelial fractions. By contrast, in the TCGA cohort, Basal-like tumors were inferred to have a significantly higher epithelial fraction (Figure S2G). This suggested the possibility that subtype calls were confounded by tumor composition. We reasoned that removing the non-epithelial signals from bulk expression profiles might lead to more consistent molecular classification across cohorts with varied tissue composition. We therefore used ADVOCATE to generate virtual epithelial and stromal expression profiles from the bulk samples of each PDA cohort (producing new datasets: vUNC, vTCGA, and vICGC). In each case, virtual profiles displayed clear expression of established cell-specific marker genes (Figure 2C and Figures S3A–B). Notably, bulk samples were distributed between the corresponding virtual epithelium and virtual stroma samples by hierarchical clustering (Figures S3C–E). Strikingly, subtype calls made from deconvolved TCGA expression data yielded two groups whose distributions of epithelial fractions were now balanced (Figure 2D). Moreover, we found the impact of deconvolution to be most apparent by functional analysis (Figures S3F–G; Tables S14–16).

Prior to deconvolution, analysis of the TCGA bulk samples classified by the UNC epithelial signature show substantial mixing of gene sets that are otherwise associated with Basal-like (red) or Classical (blue) tumors (Figure 2E), whereas after deconvolution, these groups closely aligned (Figure 2F). The variable stromal composition of the bulk TCGA dataset was thus interfering with the ability of the UNC epithelial signature to identify functionally meaningful groups of tumors.

We also noted that application of the Moffitt–E classifier to the veICGC dataset revealed excellent alignment with the pancreatic progenitor and squamous subtypes described by Bailey et. al. [13] (SMC = 0.91) (Table S13). Together, these data indicate that removal of stromal expression data from bulk tumor datasets results in the reclassification of many bulk tumor samples, particularly from the TCGA cohort, and this can improve the functional similarity of groups identified by different classification systems.

Identification of Immune-rich and ECM-rich subtypes of PDA stroma

Prior work on classification of pancreatic tumor stroma used indirect inference from bulk tissue profiles and focused primarily on the biology of quiescent or activated fibroblasts. In order to capture the contributions of all of the dozens of distinct cell types present in PDA stroma, we expanded the stromal LCM–RNA–Seq cohort described above to include samples from a total of 110 unique patients. NMF with consensus clustering identified two prominent molecular subtypes among these samples (see Figure S4 for additional details). Clear functional identities were established for these subtypes using gene set variance analysis (GSVA), leading to their designations as: an “Immune-rich” group characterized by numerous immune and interleukin signals; and an “ECM-rich” group, characterized by numerous extracellular matrix-associated pathways (Figure 3A, Tables S17–18). We next extracted a gene signature distinguishing these two stromal subtypes, making use of the compartment specificity analysis described above to filter for stroma-specific genes (Tables S19–21, see Supplementary Methods). Application of this signature to the virtual stroma profiles yielded two prominent clusters, as reflected across the UNC, ICGC, and TCGA cohorts (Figures 3B–D, Tables S22–24). Critically, in each cohort, the two clusters were again characterized by their enrichment for gene sets associated with ECM deposition or immune processes, indicating a robust and consistent performance of this new, stroma-specific “CUMC–S” signature.

Epithelial and stromal subtypes are partially linked and associated with survival differences

Having determined the epithelial and stromal subtypes of all CUMC, UNC, ICGC, and TCGA samples, a comprehensive analysis revealed substantial variation in subtype composition across the four datasets. Within the epithelium, the Basal-like group comprised 29%, 41%, and 27% of cases in the veUNC, veICGC, and veTCGA cohorts, respectively (Figures 3 E–G), and 36% of our epithelial LCM–RNA–Seq profiles (Figure S5A). Within the stroma, the ECM-rich subtype comprised 62%, 52%, and 31% of cases in the vsUNC, vsICGC, and vsTCGA cohorts, respectively (Figures 3 E–G), and 47% of our stromal LCM–RNA–Seq samples (Figure S5A). These observations serve to further highlight the significant heterogeneity between independent collections of pancreatic tumor specimens.

We next assessed the associations of epithelial and stromal subtypes with survival outcomes. Examining the epithelial subtypes, we found that removing stromal gene expression with ADVOCATE increased the survival association between Classical and Basal-like tumors in all three bulk datasets, with a particularly strong effect on TCGA outcomes (Figures 4 A–C) where 45% of the samples were re-classified after deconvolution. For the stromal subtypes, we observed at least a trend towards reduced survival among ECM-rich tumors in all three datasets (a finding made more apparent by deconvolution); however, this only reached significance in the ICGC cohort (Figures 4D–F). Together, these data indicate that (i) variations in tumor composition between different large-scale gene expression datasets can affect the predictive power of established classifier signatures for PDA, and (ii) transcriptional deconvolution can help overcome this hurdle, improving the reproducibility of outcome prediction.

The existence of numerous paracrine signaling pathways whose activity is affected by oncogenic mutations implies that stromal transcriptional programs should be influenced by epithelial identity [23]. We examined this corollary by ascertaining the association of epithelial and stromal subtypes in our experimental LCM dataset as well as in those from the virtual UNC, ICGC, and TCGA datasets. We found that in the ICGC and TCGA cohorts, the ECM-rich stroma subtype was preferentially associated with the Basal-like epithelial subtype; the UNC and CUMC cohorts trended in this direction but did not reach significance. However, a meta-analysis of the 393 samples from all four datasets yielded an Odds Ratio of 2.7 for the association of Basal-like epithelium and ECM-rich stroma (Figure S5B, random effects model: OR 2.7 [1.33 – 5.53], $p < 0.001$), indicating a partial association between epithelial and stromal compartments.

The imperfect alignment of the epithelial and stromal subtypes offered the possibility that combination subtypes might vary in their survival associations as compared to either compartment alone (Figures 4 G–I, Figure S5C–H). Indeed, consideration of combined epithelial and stromal combination subtypes affected the outcome prediction, particularly in the case of the UNC cohort where combination subtyping of deconvolved samples found a particularly poor outcome for Basal-like/ECM-rich tumors relative to Classical/Immune-rich tumors (HR = 3.76 for combined subtyping vs. 2.11 for epithelial subtyping alone, Figure 4I vs. 4C). Together, these data highlight the relationship between Basal-like epithelium with ECM-rich stroma in pancreatic cancer and the strong association of this combination with poor overall survival.

Discussion

The traditional understanding of genetic mutations as drivers of tumor development has led to a focus on the malignant compartment that is exemplified by the term “tumor purity”, which regards the stroma as mere contamination. However, with the understanding that stromal cells play critical roles in both promoting and restraining pancreatic tumor progression[24], the consensus view of the stromal compartment has shifted to that of a critical partner – or foil – to the malignant epithelium. Indeed, in some contexts the stroma can even play a dominant role, as epitomized by the success of stroma-targeted immunotherapy in treating aggressive cancers such as metastatic melanoma and non-small

cell lung cancer. In this light, we sought to study the interplay of PDA epithelium and stroma in their native state, separated by LCM from otherwise intact samples, but matched by patient so that the reciprocal signals active in each compartment might be examined.

A key outcome of this work is to unify our understanding of molecular subtypes in pancreatic ductal adenocarcinoma. To do this, we first examined the properties of subtypes resulting from existing classification schemes. We noted that among more than 60 individual epithelial tumor profiles, there was little evidence for the existence of the Collisson–Exocrine or Bailey–ADEX subtype, as evidenced by the general lack of expression of marker genes associated with these subtypes. Conversely, signature genes for the Bailey–Immunogenic subtype generally well expressed, but predominantly in stromal samples, suggesting that this subtype, which was presented as being mutually exclusive with the epithelial Squamous and Progenitor subtypes, in fact arises from the stromal compartment.

Given the fact that none of the classification signatures were perfectly epithelium specific, we suspected that varying levels of stromal tissue content might impact the assignment of tumors to different molecular subtypes. Indeed, removal of stromal expression signals from bulk expression data resulted in the reclassification of nearly half the TCGA samples using the Moffitt–E signature and improved detection of the functional processes associated with the Classical and Basal–like subtypes in each cohort. Classification efforts may thus benefit from virtual purification of gene expression prior to supervised clustering.

In our effort to establish a novel classification system for PDA stroma, we placed the greatest emphasis on the reproducibility of molecular phenotypes across multiple cohorts. Following this process, we observed with great interest the emergence of two prominent molecular subtypes in the stroma with pronounced enrichment for two different aspects of stromal biology: ECM deposition and remodeling versus immune–related processes. This concept refines the idea of ‘activated’ and ‘normal’ stromal subtypes, which was derived largely from the biology of pancreatic stellate cells [15] and thus did not take into account the substantial contributions of immune cells to the PDA microenvironment. We also note that although the Bailey–Immunogenic tumors in the ICGC cohort are generally identified as Immune-rich by our analysis, there are important distinctions between these classification schemes. Specifically, the Bailey–Immunogenic subtype is one of four mutually-exclusive classes and picks up Classical/Progenitor tumors with a high abundance of immune infiltration. This structure precludes both the possibility of tumors having a low abundance of stroma but for which the stroma has an immunogenic quality and tumors with high stromal abundance lacking an immunogenic character.

By examining all four tumor cohorts, we found a strong association between an ECM–rich stroma and Basal-like epithelium while Immune-rich stroma occurred more often in association with Classical epithelia. The latter finding corroborates the concept that epithelial traits promoting dedifferentiation in PDA, such as the loss of SMAD4 expression, may in fact shape a more matricellular stromal phenotype[23]. Interestingly, a recent study [25] in patient-derived xenografts showed that basal-like and classical tumor cells, respectively, implanted subcutaneously into mice almost unequivocally induced microenvironments dominated by fibrosis (i.e. ECM-rich) and immune infiltration (i.e.

Immune-rich), respectively. We also found that cross-compartment subtypes are associated with differences in outcome, with Basal-like/ECM-rich tumors having a substantially worse overall survival when compared to Classical/Immune-rich tumors (overall HR = 3.76, 3.81, and 2.63 for UNC, ICGC, and TCGA, respectively). Although a direct comparison is not possible, this effect size is in the same general range as other known single variables in pancreatic cancer biology, including lymph node status (HR = 1.5), postoperative CA19-9 level (HR = 3.6) or the number of high penetrance driver genes (HR = 1.4)[26, 27]. Unfortunately, differences in the clinicopathological data reported for each cohort precluded a more sophisticated multivariate model. Nonetheless, we expect that this approach to subtyping will have immediate applications, for example, in interpreting the results of small-scale clinical trials where random inequalities of molecular subtypes could dramatically affect the expected survival between groups or relative to historical controls.

Methods

The information provided here is a succinct summary of the experimental procedures. Detailed information is provided in supplementary information.

Samples studied

Information is provided from a total of 122 PDA patients who underwent surgery at the Columbia Pancreas Center. From these, an implementation of the ADVOCATE algorithm [18] on 60 pairs of epithelial and stromal samples matched by patient. Additional samples were utilized in unsupervised clustering analyses, as detailed in Supplementary Table 25. Patients provided surgical informed consent which was approved by a local ethics committee (IRB # AAAB2667). Samples were frozen intraoperatively by the Columbia University Tumor Bank. Clinical and pathological information on the 122 cases are provided in Supplementary Tables 1 and 2.

Laser capture microdissection and RNA sequencing

Cryosections of OCT-embedded tissue blocks were transferred to PEN membrane glass slides and stained with cresyl violet acetate. Adjacent sections were H&E stained for pathology review. Laser capture microdissection was performed on a PALM MicroBeam microscope (Zeiss), collecting at least 1000 cells per compartment. RNA was extracted and libraries prepared using the Ovation RNA-Seq System V2 kit (NuGEN). Libraries were sequenced to a depth of 30 million, 100bp, single-end reads.

Computational modeling

This manuscript makes use of a novel computational model called ADVOCATE. A description of this approach is being prepared for submission in a separate manuscript. However, we have appended a “Conceptual Approach” document describing the mathematical basis of this method of ADCOATE for the benefit of reviewers of this manuscript. The ADVOCATE software is publically available on Github [18] and a manuscript describing its development is in preparation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank Richard Moffitt for valuable critique of the manuscript. This work was supported by the National Cancer Institute (NCI) Cancer Target Discovery and Development program (1U01CA168426 to A.C.), NCI Research Centers for Cancer Systems Biology Consortium (1U54CA209997 to A.C. and K.P.O.), NCI Outstanding Investigator Award (R35CA197745-02 to A.C.), NCI Cancer Center Support Grant (3 P30 CA13696-40) and NCI Research Project Grant (R01CA157980 to K.P.O.). Financial support was also provided by the Columbia University Pancreas Center. H.C.M. received support from a Mildred Scheel Postdoctoral Fellowship (Deutsche Krebshilfe). P.E.O. received support from the NIH NCATS (KL2TR001874).

Abbreviations:

CPHM	Cox proportional hazards model
CUMC	Columbia University Medical Center cohort
DEG	differentially expressed genes
FACS	fluorescence assisted cell sorting
GSVA	gene set variance analysis
HR	hazard ratio
ICGC	International Cancer Genome Consortium cohort
LCM	laser capture microdissection
NMF	non-negative matrix factorization
PDA	pancreatic ductal adenocarcinoma
TCGA	The Cancer Genome Atlas cohort
UNC	University of North Carolina cohort
ve	virtual epithelial
ves	virtual epithelial/stromal
vs	virtual stromal

References

1. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun* 2015;6:8971. [PubMed: 26634437]
2. Oberstein PE, Olive KP. Pancreatic cancer: why is it so hard to treat? *Therapeutic advances in gastroenterology* 2013;6:321–37. [PubMed: 23814611]
3. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin* 2017;67:7–30. [PubMed: 28055103]

4. Rhim AD, Oberstein PE, Thomas DH, Mirek ET, Palermo CF, Sastra SA, et al. Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. *Cancer Cell* 2014;25:735–47. [PubMed: 24856585]
5. Provenzano PP, Cuevas C, Chang AE, Goel VK, Von Hoff DD, Hingorani SR. Enzymatic targeting of the stroma ablates physical barriers to treatment of pancreatic ductal adenocarcinoma. *Cancer Cell* 2012;21:418–29. [PubMed: 22439937]
6. Jacobetz MA, Chan DS, Neesse A, Bapiro TE, Cook N, Frese KK, et al. Hyaluronan impairs vascular function and drug delivery in a mouse model of pancreatic cancer. *Gut* 2012.
7. Olive KP, Jacobetz MA, Davidson CJ, Gopinathan A, McIntyre D, Honess D, et al. Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science* 2009;324:1457–61. [PubMed: 19460966]
8. Vonderheide RH, Bayne LJ. Inflammatory networks and immune surveillance of pancreatic carcinoma. *Curr Opin Immunol* 2013;25:200–5. [PubMed: 23422836]
9. Jones S, Zhang X, Parsons DW, Lin JC-H, Leary RJ, Angenendt P, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 2008;321:1801–6. [PubMed: 18772397]
10. Biankin AV, Waddell N, Kassahn KS, Gingras M-C, Muthuswamy LB, Johns AL, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* 2012;491:399–405. [PubMed: 23103869]
11. Witkiewicz AK, McMillan EA, Balaji U, Baek G, Lin WC, Mansour J, et al. Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nature communications* 2015;6:6744.
12. Waddell N, Pajic M, Patch A-M, Chang DK, Kassahn KS, Bailey P, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* 2015;518:495–501. [PubMed: 25719666]
13. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* 2016;531:47–52. [PubMed: 26909576]
14. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat Med* 2011;17:500–3. [PubMed: 21460848]
15. 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. [PubMed: 26343385]
16. Cancer Genome Atlas Research Network. Electronic address aadhe, Cancer Genome Atlas Research N. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* 2017;32:185–203 e13. [PubMed: 28810144]
17. Hruban RHP MB.; Klimstra DS. Tumor of the Pancreas Washington DC: American Registry of Pathology, 2007.
18. Laise PH J; Bansal M; Califano A. <https://github.com/califano-lab/ADVOCATE>. Github 2018.
19. Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, et al. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods* 2013;10:623–9. [PubMed: 23685885]
20. Shanker S, Paulson A, Edenberg HJ, Peak A, Perera A, Alekseyev YO, et al. Evaluation of commercially available RNA amplification kits for RNA sequencing using very low input amounts of total RNA. *J Biomol Tech* 2015;26:4–18. [PubMed: 25649271]
21. Pontén F, Jirstrom K, Uhlen M. The Human Protein Atlas-a tool for pathology. *J Pathol* 2008;216:387–93. [PubMed: 18853439]
22. He JM HC.; Holmstrom SR.; Su T.; Ahmed A.; Hibshoosh H.; Chabot JA.; Oberstein PE.; Sepulveda AR.; Genkinger JM.; Zhang J.; Iuga AC.; Bansal M.; Califano A.; Olive KP. Transcriptional deconvolution reveals consistent functional subtypes of pancreatic cancer epithelium and stroma. *Bioarxiv* 2018;10.1101/88779.
23. Laklai H, Miroshnikova YA, Pickup MW, Collisson EA, Kim GE, Barrett AS, et al. Genotype tunes pancreatic ductal adenocarcinoma tissue tension to induce matricellular fibrosis and tumor progression. *Nat Med* 2016;22:497–505. [PubMed: 27089513]

24. Neesse A, Algul H, Tuveson DA, Gress TM. Stromal biology and therapy in pancreatic cancer: a changing paradigm. *Gut* 2015;64:1476–84. [PubMed: 25994217]
25. Nicolle R, Blum Y, Marisa L, Loncle C, Gayet O, Moutardier V, et al. Pancreatic Adenocarcinoma Therapeutic Targets Revealed by Tumor-Stroma Cross-Talk Analyses in Patient-Derived Xenografts. *Cell Rep* 2017;21:2458–70. [PubMed: 29186684]
26. Yachida S, White CM, Naito Y, Zhong Y, Brosnan JA, Macgregor-Das AM, et al. Clinical significance of the genetic landscape of pancreatic cancer and implications for identification of potential long-term survivors. *Clin Cancer Res* 2012;18:6339–47. [PubMed: 22991414]
27. Berger AC, Garcia M Jr., Hoffman JP, Regine WF, Abrams RA, Safran H, et al. Postresection CA 19–9 predicts overall survival in patients with pancreatic cancer treated with adjuvant chemoradiation: a prospective validation by RTOG 9704. *J Clin Oncol* 2008;26:5918–22. [PubMed: 19029412]

Significance of this study

What is already known on this subject?

- Pancreatic ductal adenocarcinoma (PDA) is one of the most aggressive malignancies with currently no targetable genetic alterations. At the pathological level it is a complex mixture of tumor cells, normal pancreatic tissues and stromal cell types, thus impeding the straightforward molecular characterization of transcriptional profiles.
- Previous approaches to molecular subtyping have relied on bulk PDA samples leading to the proposal of anywhere between two to four distinct tumor classes. One study used indirect inference to identify two stromal subtypes associated with the activation state of pancreatic stellate cells.
- While a systematic evaluation of cross-compartment subtypes is lacking for PDA, current evidence suggests that epithelial and stromal programs evolve independently.

What are the new findings?

- We used laser capture microdissection and RNA sequencing to directly sample pathologically verified PDA epithelia and their adjacent stroma for more than 60 patients.
- Tumor epithelia naturally separate into ‘classical’ and ‘basal-like’ subtypes while additional subtypes such as ‘exocrine’ or ‘ADEX’ are not supported.
- Unsupervised class detection among 110 stromal LCM-RNA-Seq profiles detects two groups reflecting immune signaling and matricellular fibrosis, respectively.
- Systematic analysis of epithelial and stromal subtypes on nearly 400 PDA specimens found functional consistency across multiple cohorts.
- Across these same tumors, epithelial and stromal subtypes were partially linked, indicating potential dependence in the evolution of tissue compartments in PDA.

How might it impact on clinical practice in the foreseeable future?

- The ability to robustly assess both epithelial and stromal subtypes for patients will facilitate the discovery of theranostic relationships between molecular composition and treatment studies, and could form the basis of future precision medicine approaches for pancreatic ductal adenocarcinoma.

Figure 1. Compartment-specific gene expression profiling of pancreatic tumors

(A) Images of Cresyl Violet stained human PDA frozen sections before and after laser capture microdissection of malignant epithelial and adjacent stromal cells. (B) RIN values for RNA samples derived from the indicated compartment (N = 60 each). (C) Number of genes and transcripts detected at >1 FPKM in the samples from (B). (D) Principal component analysis of the 60 paired epithelial and stromal LCM expression profiles from (C). Color graduation shows pairing of samples from the same tumor. Three samples discussed later are labeled. (E) Heatmap showing the expression of marker genes for epithelial cells (Epi.), endothelial cells (Endo.), cancer associated fibroblasts (CAF) and immunocytes (Imm). (F, G) Protein validation of genes predicted as epithelium-specific (F) or stroma-specific (G) based on mRNA expression. Bar height and color shading reflect the certainty (t-statistic) of differential expression. The box color below each bar summarizes results of immunohistochemistry on PDA sections from the Human Protein Atlas (HPA).

IHC staining pattern was categorized as strongly or weakly supportive of the predicted compartment (blue/red), indeterminate (grey), absent (white), or opposite the predicted pattern (black). **(H)** An example epithelium-specific gene, LGALS4, showed a protein staining pattern that was strongly consistent with its mRNA expression (at left). Blue and red arrows indicate PDA epithelium and stroma, respectively. **(I)** LGALS1 exhibited a highly stroma-specific expression pattern.

Figure 2. Analysis and classification of pancreatic tumor cohorts and classifiers

(A) Tumor and stroma content analysis of pancreatic tumors from the ICGC (blue), UNC (red), and TCGA (grey) cohorts. (B) Analysis of gene expression across 60 pairs of PDA epithelium and stroma LCM-RNA-seq profiles, highlighting the genes used to determine each subtype from the Collisson, Moffitt, and Bailey classifiers. Top panel displays the compartment– specificity of the signature genes for each subtype based on the t-statistic of their differential expression between PDA epithelium and stroma samples; positive values indicate stromal enrichment. Lower panel depicts the average expression of each signature genes across all LCM-RNA-seq samples, in fragments per million mapped fragments (FPM). (C) Heatmap depicting the differential expression of indicated marker genes in deconvolved virtual epithelial (veTCGA) and stromal (vsTCGA) profiles from the TCGA cohort. (D) Epithelial fraction of TCGA pancreatic tumors allocated to the Basal-like (red) and Classical (blue) subtypes based on analysis of either bulk or virtual epithelial expression

profiles. (E, F) Analysis of gene sets associated with the Moffitt Basal-like (red) and Classical (blue) subtypes based in bulk expression profiles (E) from TCGA versus virtual epithelial profiles (F) of the same tumors. Heatmap depicts gene set variance analysis (GSVA) scores per sample for indicated gene sets. Stratification of bulk TCGA profiles using the Moffitt classifier results in groups that are not differentially enriched in the gene sets classically associated with Basal-like versus Classical subtypes. However, following deconvolution the virtual epithelial profiles stratify into two groups that reflect the functional biology of the Basal-like and Classical subtypes.

Figure 3. Systematic stromal subtyping of PDA

(A–D) Heatmaps of the top 30 DEG between groups obtained by clustering stromal LCM–RNA–Seq samples from CUMC tumors (A), and virtual stromal (vs) profiles from the UNC (B), ICGC (C) and TCGA (D) cohorts, respectively. Clustering was based on the expression of a signature derived from stromal LCM profiles from 110 individual patients (CUMC-S classifier, see Supplementary Methods). Top section of heat-map depicts GSVA scores per sample for indicated gene sets. In each virtual stroma dataset, two groups were identified, one with features indicating elevated extracellular matrix deposition and remodeling (“ECM– rich”, purple) and another enriched in various immune and interleukin pathways (“Immune-rich”, green). (E–G) Multilayered donut plots showing (i) the alignment of epithelial with stromal subtypes for each tumor in each cohort and (ii) the proportion of each epithelial subtype. Separate pie charts summarize the proportion of stromal subtypes per cohort.

Figure 4. Combined epithelial and stromal subtypes associate with overall survival

Kaplan-Meier (KM) survival analysis of patients with resected PDA from the ICGC (n= 93), TCGA (n= 137), or UNC (n= 125) cohorts, stratified by the indicated signatures applied to either bulk expression profiles (thin lines) or transcriptionally deconvolved versions of the same (thick lines). Below each KM plot, horizontal bars indicate the hazard ratios (HR) from a Cox proportional hazards model (CPHM), along with their 80% (blue), 90% (yellow) and 95% (orange) confidence intervals. (A-C) KM plot of patients from the indicated cohorts using the Moffitt-E signature to stratify Basal-like (red) versus Classical (blue) tumors, showing that the detection of a differential prognosis among the epithelial subtypes is generally enhanced by transcriptional deconvolution. Bars indicate HR for Basal-like tumors in virtual epithelial and bulk profiles. (D-F) KM plot of patients from the indicated cohorts using the CUMC-S signature to stratify ECM-rich (purple) versus Immune-rich (green) tumor. Kaplan-Meier (KM) survival analysis depicts overall survival relative to stromal

subtype. Stromal subtypes are statistically associated with outcome in the ICGC cohort with ECM-rich tumors having a worse prognosis. Bars indicate HR for ECM-rich tumors in virtual stromal and bulk profiles. (G-I) KM plot of patients from the indicated cohorts using a combination of the Moffitt-E and CUMC-S signatures. Red lines indicate Basal-like tumors with an ECM-rich stroma while blue lines indicate Classical tumors with an Immune-rich stroma; all other tumors are represented as a grey line. Bars indicate HR for Basal-like::ECM-rich tumors in bulk and virtual epithelial/stroma (ves) profiles.