Yeo *et al. Discover Oncology* (2026) 17:866
https://doi.org/10.1007/s12672-026-05091-0

Discover Oncology

ANALYSIS Open Access

A tumor microenvironment integrated (TMI) staging system for pancreatic ductal adenocarcinoma based on cancer-associated fibroblast and molecular consensus signatures

Samuel Jianjie Yeo[1], Ian Tate Ek Ern Sim[1], Hui Yi Tay[2], David Tai[3,4], Ruoyu Shi[5] and Jason Yongsheng Chan[2,3,4*]

*Correspondence:
Jason Yongsheng Chan
jason.chan.y.s@singhealth.com.sg
[1]Lee Kong Chian School of Medicine, Singapore, Singapore
[2]Cancer Discovery Hub, National Cancer Centre Singapore, Singapore, Singapore
[3]Division of Medical Oncology, National Cancer Centre Singapore, 30 Hospital Blvd, Singapore 168583, Singapore
[4]Duke-NUS Medical School, Singapore, Singapore
[5]Department of Pathology, National University Hospital, Singapore, Singapore

Abstract

Background Pancreatic ductal adenocarcinoma (PDAC) is an aggressive solid tumor with poor prognosis. Existing prognostication systems rely on clinical staging, but systems of tumor classification through molecular signatures and the association between cancer-associated fibroblast (CAF) subtypes and clinical outcomes suggest that prognostic systems could be improved by integrating these data.

Methods We clustered a discovery cohort ($n = 320$) of PDAC into basal and nonbasal tumor subtypes by applying a molecular signature previously defined by O'Kane et al. We performed deconvolution using data of annotated cells from separate single-cell RNA-seq studies to estimate the fractions of different CAFs in each case in the discovery cohort and further generated a fraction threshold for myofibroblastic CAFs (myCAFs) enrichment delineating poor prognosis. We integrated these molecular classifications and metastasis status to generate a tumor microenvironment integrated (TMI) staging system, which we subsequently compared against the 7th American Joint Committee on Cancer (AJCC) clinical staging system using an external validation cohort ($n = 156$). We finally explored the signaling pathways within the single-cell dataset to identify key signaling pathways associated with myCAFs.

Results TMI staging outperformed AJCC staging in predicting overall survival within both the discovery cohort (Harrell's C-index: 0.558 vs. 0.547) and validation cohort (Harrell's C-index: 0.535 vs. 0.500). In terms of differences, TMI was better able to account for overall survival rate variations, while AJCC staging remained stronger in hazard consistency, a surrogate of the similarity of survival rates within subgroups of each defined stage. We further identified fibronectin 1 (FN1) as a key ECM signaling protein within PDAC tumors, both in myCAF autocrine signaling as well as between myCAFs and ductal epithelial cells (including tumor cells), that are enriched in both basal (TMI-Poor) and nonbasal, myCAF-enriched (TMI-Intermediate) cases, and that delineate a poorer prognosis.

Conclusion The integration of tumor subtyping and stromal myCAF fractions with clinical staging improves prognostication in PDAC. Among the ECM signaling proteins,

FN1 shows promise in explaining the relationship between basal subtypes and myCAF-enrichment, with poor prognosis.

Keywords Single-cell sequencing, RNAseq, Prognosis, CellChat, The Cancer Atlas Genome, Clinical proteomic tumor analysis consortium, Fibronectin 1, Cancer-associated fibroblasts

1 Background

Pancreatic ductal adenocarcinoma (PDAC) is frequently detected in its late stages due to the absence of notable symptoms, resulting in an abysmal overall survival [1, 2]. The current treatment approach of PDACs is largely based on tumor staging. Although the American Joint Committee on Cancer (AJCC) pathologic staging system provides a fair estimate on prognosis based primarily on clinical and surgical characteristics, recent work has revealed significant heterogeneity in the prognoses of PDACs of different molecular characteristics [3]. Multiple molecular subtyping systems of PDAC have been proposed, most of which segregate into two dominant molecular subtypes – a consensus classical subtype characterized by the expression of terminal epithelial and adhesion-related proteins native to the pancreas, particularly GATA-binding factor 6 (GATA6), and a consensus basal subtype characterized by the repression of pancreas-specific transcriptional programmes, expression of epithelial-to-mesenchymal genes as well as higher rates of metastasis [4–8]. This classification has yielded prognostic benefit, as subtypes correlating with the consensus basal subtype have been correlated with a significantly poorer prognosis, suggesting that its inclusion into existing staging systems may improve their prognostic accuracy [5, 6, 9].

Concurrently, significant progress has been made in elucidating the stromal components of PDAC, which includes a complex mix of tumor-modifying immune cells and cancer-associated fibroblasts (CAFs) among various cell types within an extracellular matrix (ECM) that interact through secreted growth factors, immunomodulatory cytokines and glycoproteins [10–13]. Single-cell RNA sequencing (scRNA-seq) has enabled the identification of numerous putative CAF subtypes, but two dominant subtypes have been consistently identified across most studies of human PDAC tumors – inflammatory CAFs (iCAFs) and cancer-associated myofibroblasts (myCAFs) [10, 14]. These subtypes have been consistently identified in both human and mouse models, with myCAFs showing particularly high alpha-smooth muscle actin (α-SMA/ACTA2) expression and colocalization with neoplastic cells, and iCAFs showing an extensive secretion of inflammatory cytokines like interleukin 6 (IL6), leukaemia inhibitory factor (LIF) and CXCL1 [11, 12]. The deconvolution of bulk RNA-seq datasets to estimate myCAF fractions have suggested poorer survival outcomes associated with myCAFs, while the prognostic relevance with iCAFs remain unclear [15, 16]. Other studies have also reported novel CAF subtypes not only modulating metastasis risk and prognosis, but also response to immunotherapy [17, 18]. In many cases, these prognostic changes are linked to ligand-receptor interactions between these CAF subtypes, such as complement-secreting CAFs (csCAFs), antigen-presenting CAFs (apCAFs), periostin-positive CAFs ($POSTN^+$ CAFs) and metabolic CAFs (meCAFs), and various adaptive and innate immune cells that can restrict or enable tumor proliferation [12, 13, 19, 20]. While these findings suggest that the molecular subtyping and characterization of different CAF stromal components convey prognostic information, the integration of such information into existing clinical

staging systems is not well reported in literature. In this work, we attempt to incorporate various molecular findings with clinical characteristics of PDAC to propose a tumor microenvironment integrated (TMI) staging system for prognostication.

2 Materials and methods

2.1 Study cohorts for development and validation of staging systems

For creating and validating a staging system, two cohorts were assembled. The first cohort was an discovery cohort comprising 320 cases of pancreatic adenocarcinomas (PDAC) retrieved via the Genome Data Commons (GDC) and comprising data from two cohorts – The Cancer Genome Atlas (TCGA) (n = 162) and Clinical Proteomic Tumor Analysis Consortium (n = 158) (CPTAC) (Table 1) [21–23]. The second cohort

Table 1 Characteristics of discovery cohort

Characteristics	*N* (%)
Total number	320 (100)
Mean age of diagnosis (standard deviation)	65.1 (10.9)
Female sex	147 (45.9)
Ethnicity ^	
White (including Hispanics)	234 (73.1)
Black or African American	8 (2.5)
Asian	35 (10.9)
Others / not reported	43 (13.4)
Alive at last follow-up	105 (32.8)
AJCC Staging ^	
Stage I	34 (10.6)
Stage II	221 (69.1)
Stage III	47 (14.7)
Stage IV	16 (5.0)
Not reported	2 (0.6%)
AJCC T stage	
T1	12 (3.8)
T2	64 (20.0)
T3	230 (71.9)
T4	8 (2.5)
TX	2 (0.6)
Not reported	4 (1.3)
AJCC N stage	
N0	77 (24.1)
N1	209 (65.3)
N2	27 (8.4)
NX	5 (1.6)
Not reported	2 (0.6)
AJCC M stage	
M0	182 (56.9)
M1	16 (5.0)
MX	122 (38.1)
Primary tumor site ^	
Head of pancreas	242 (75.6)
Body of pancreas	26 (8.1)
Tail of pancreas	24 (7.5)
Overlapping lesion of pancreas	2 (0.6)
Not otherwise specified	26 (8.1)

^Samples with unreported characteristics not included in percentage proportions

was a validation cohort comprising 156 individuals from GSE62452 and GSE183795 of the Genome Expression Omnibus (GEO) [24, 25].

To build the discovery GDC cohort, the Genome Data Commons (GDC) data portal (https://portal.gdc.cancer.gov/analysis_page?app=Downloads) was used to search for cases of pancreatic carcinomas using the following search filters: pancreas tissue type, RNA-seq experimental strategy, open-access data, primary tumor, solid tissue type. 367 pancreatic carcinoma cases were identified and retrieved, including both RNA-seq-derived transcription data as well as the following clinical metadata of each case: age at diagnosis, living status, time from diagnosis to most recent follow-up or death, AJCC pathologic staging, AJCC tumor staging, AJCC nodal staging, AJCC metastasis staging and site of resected tumor. Next, we removed 47 cases where either clinical metadata for living status and time from diagnosis to most recent follow-up or death was incomplete or histological diagnoses were not that of "Infiltrating duct carcinoma, NOS" or "Adenocarcinoma, NOS", thus creating the final cohort of 320 cases for our discovery analysis.

To build the validation cohort, a thorough search of GEO datasets (https://www.ncbi.nlm.nih.gov/gds/) was conducted to identify additional PDAC datasets containing cases with both gene expression and clinical survival as well as staging data available [26]. Two studies – GSE62452 (n = 130) and GSE183795 (n = 139) – which contained cases from all 4 AJCC stages were selected. Cases were filtered to remove non-PDAC histology cases and cases lacking survival or staging data (number removed: GSE183795, 6; GSE62452, 66). Among the remaining cases, 41 overlapped between both datasets. To preserve a sufficient case number for each dataset, overlapped cases were removed from GSE183795, thus leaving 64 cases in GSE62452 and 92 cases in GSE183795 that formed the 156 cases composing the final validation cohort.

The discovery cohort data was downloaded via the GDC with expression data and clinical metadata extracted on R (version 4.3.0) in RStudio (version 2024.04.0 + 735.pro3) while validation cohort data was downloaded using the GEOquery package (version 2.72.0) [27–29]. Within the discovery cohort, data was converted to FPKM-normalized and log2 + 1 of counts prior to scaling, the latter to handle genes with zero transcripts. In the validation cohort, as the downloaded expression data was already log normalized, expression data was converted back into a linear, normalized format.

2.2 Modified Moffitt classification of discovery and validation cohorts.

The modified Moffitt's classifier (mMC) is a 47-gene modification of a gene signature originally derived from non-negative matrix factorization (NMF) and clustering of PDAC expression data that delineates PDACs into classical and basal subtypes [4, 6, 30]. As with the method utilized by O'Kane et al., the log2-normalized and scaled data was used to incorporate the mMC into the discovery cohort cases. The ComplexHeatmap package (version 2.20.0) was applied to mMC gene expression for the hierarchical clustering of cases into basal and nonbasal subtypes [31]. This same process was applied separately to each set of cases originating from the study datasets (TCGA, CPTAC, GSE183795, GSE62452).

2.3 CIBERSORTx-based deconvolution of cell fractions and establishment of CAF enrichment status

To establish CAF enrichment status, we first established cell fractions within bulk-sequenced cases via deconvolution before deriving cut-offs for myCAF-, iCAF- and csCAF- enriched and depleted cases. CIBERSORTx (https://cibersortx.stanford.edu/) is an online tool employing support vector regression that harnesses single-cell gene expression data of phenotypically annotated cells to estimate the fractions of the same cell phenotypes in bulk gene expression data [32]. Single-cell expression data (of which the derivation is detailed in the next section) with cell phenotypes annotated was uploaded to CIBERSORTx to generate a signature matrix, which was subsequently applied separately to the discovery and validation cohorts on default settings with permutations set to 100 to derive cell fractions. The discovery subcohorts (TCGA, CPTAC) were separately uploaded to and processed by CIBERSORTx as raw counts, while validation subcohorts (GSE62452, GSE183795) were separately uploaded and processed in a normalized linear form. To determine cell fraction thresholds for enriched and depleted statuses of each CAF subtype, Delong's method for computing the area under the curve (AUC) of the receptor operating curve (ROC) was employed via the pROC package (version 1.18.5) on the CAF fractions of the combined TCGA and CPTAC discovery cohort to find the optimum thresholds based on survival data [33]. These thresholds were then applied to both discovery and validation cohorts to delineate enriched or depleted status for each CAF subtype.

2.4 Construction of a tumor microenvironment integrated (TMI) staging system

TMI is a staging system for pancreatic ductal adenocarcinomas that uses both molecular and clinical data for prognostication (Supplementary Figure S4) into three stages. Basal subtype cases or cases involving metastatic PDAC are classified as "Poor", while the remaining cases are classified as "Intermediate" if they are myCAF-enriched or "Good" if myCAF-depleted. Derivation of this staging was based on Cox proportional hazards modelling of T, N and M staging together with myCAF enrichment and modified Moffitt classification to determine the most significant variables (Supplementary Table 1).

2.5 Single-cell RNA-sequencing dataset

We obtained single-cell data of histologically-confirmed primary PDAC resected from 23 subjects across 5 studies and processed on 10x Genomics platforms deposited in the Genome Expression Omnibus (GEO): GSE179705, GSE231535, GSE197177, GSE154778 and GSE217845 [34–38]. The GEO data was in a post-processed form, having been aligned to GRCh38 (except for GSE154778 which was aligned on GRCh37) using the Cell Ranger pipeline by their original study teams, and was imported in the standard 10x Genomics feature-barcode matrices format (https://www.10xgenomics.com/support/software/cell-ranger-arc/latest/analysis/outputs/feature-barcode-matrices). AJCC staging data was available for each participant at the point of diagnosis and tumor resection. These studies yielded a mixed set of both treatment-naïve (n = 3), treatment-experienced (n = 3) and treatment-unknown (n = 17) subjects with a range of AJCC stages.

2.6 Single-cell RNA-sequencing analysis

We used Seurat (version 5.1.0) for the further processing and integration of single-cell data [39]. We performed quality control on all samples collectively by removing cells with a mitochondrial transcript proportion above 20% (suggestive of dead, apoptotic cells) and utilizing the scDblFinder package (version 1.17.3) to remove potential multiplets, clusters of cells incorrectly processed, sequenced and interpreted as single cells [40]. Prior to clustering and analyses, we first performed log normalization, identified variable genes (n = 2000) with high cell-to-cell expression differences that can be used to distinguish phenotypes, performed data scaling to prevent high-expression genes from dominating downstream processes and ran dimensionality reduction using principal component analysis to identify principal components that can be used to determine "anchors," or pairings of cells with similar phenotypes across experiments. As data were generated via different versions of 10x Genomics Chromium platforms, we performed batch correction across all experiments using Seurat's anchor-based canonical correlation analysis (CCA)-integration method, which transforms the data into a shared space and accounts for variability between experimental chemistries, conditions and individuals. All such steps were performed as per the workflow instructed in the official Seurat vignette "Integrative analysis in Seurat v5" (https://satijalab.org/seurat/articles/seurat5_integration) using default parameters.

For clustering and visualization, we first transformed integrated data into reduced dimensional space using Seurat's uniform manifold approximation and projection (UMAP) method with dimensions set at 5. To then identify broad classes of cells, we then performed low-resolution clustering based on the relations of each cell's expression in the reduced dimensionality space as evaluated by the k-nearest neighbor algorithm. We assigned cell identities through the manual identification of each cluster based on common markers: *COL1A1*, *COL3A1*, *LUM* and *DCN* for fibroblasts [12, 41]; *CD3E*, *CD4* and *CD8A* for T-cells [42]; *CD68* and *CD163* for macrophages (myeloid cells) [43, 44]; *PECAM1*, *VWF* and *CD34* for endothelial cells [45]; *RGS5* for pericytes [46]; *KRT8*, *KRT18*, *KRT19* and *MMP7* for ductal epithelial cells, which include tumor cells [47–50].

To identify specific CAF subtypes, we re-clustered all identified fibroblasts using the above methodology, adjusting the resolution of clustering till we arrived at three clusters which were validated from previous reports. These clusters were annotated with the following markers: *ACTA2*, *POSTN* and *HOPX* for cancer-associated myofibroblasts (myCAFs) [12]; *IL6* and *CXCL8* for inflammatory CAFs (iCAFs) [20, 51]; *C3* and *C7* for complement-secreting CAFs (csCAFs).

2.7 Further processing of single-cell RNA-sequencing data

To build the signature matrix for CIBERSORTx using scRNA-seq data, we uploaded transcript counts for the set of 47 genes comprising the mMC defined by O'Kane et al. for each of a cluster-proportional down sample of cells comprising the scRNA-seq data. (4) To infer basal and nonbasal subtypes akin with the discovery and validation cohorts, we calculated the FPKM of "pseudo-bulk" counts formed by summing transcripts across all cells of each sample and applied a logarithmic transformation, before using the ComplexHeatmap package to perform hierarchical clustering.

We utilized CellChat (version 2.0) to explore signaling pathways enriched within our single-cell dataset [52]. Pathway analyses were performed for the total dataset, as well

as for the separate basal and nonbasal sets of samples as inferred prior, to obtain relative outgoing and incoming signaling strengths for the top five pathways by the clusters defined in our initial Seurat analysis.

2.8 Statistical analysis

R and the SciPy package (version 1.13.0) were used to perform all statistical tests in this study [53]. The Mann-Whitney U (Wilcoxon) test was used for comparisons between our continuous variables and both Fisher's exact test as well as the Chi-squared test were used for comparisons between categorical variables, with statistical significance set at $p < 0.05$. Cox proportional hazards models, Harrell's concordance-index (C-index) calculations and Kaplan-Meier survival curves were generated on the survival package (version 3.8.6), with the ggplot2 package (version 4.0.2), ggfortify (version 0.4.19) and ggpubr (version 0.6.2) used to generate Kaplan-Meier survival plots [54, 55].

3 Results

3.1 Single-cell analysis validates the presence of three dominant CAF subtypes in pancreatic ductal adenocarcinoma

To unify molecular subtyping with clinical scoring systems, we planned to use single-cell data to impute CAF fractions and molecular consensus subtypes (basal or nonbasal) in bulk RNA-seq samples in our discovery cohort (Fig. 1A). Thus, we first examined the stromal make-up of CAFs in PDAC. Data from twenty-three surgically resected PDAC samples were retrieved from five GEO experiments (Table 2). All samples were of primary PDAC tumors and not sorted by surface markers prior to sequencing to obtain an unbiased make-up of the single-cell landscape. Fifteen tumors were AJCC stages III and IV while eight were AJCC pathologic stage II tumors (Fig. 1B). Of the 89,567 cells we collated following quality control, our clustering revealed 17,722 CAFs marked by a high expression of fibroblast-specific markers such as *COL1A1*, *COL13A1*, *LUM* and *DCN*, which have been employed in other studies as markers of CAFs. Of the non-CAF clusters, a significant portion of ductal epithelial cells, which include tumor cells, endothelial cells, pericytes, and immune cells were identified (Fig. 1C and D).

Further clustering of CAFs revealed three evenly sized clusters – 0, 1 and 2 – (Supplementary Figure S1). We annotated cluster 0 as myCAFs based on the expression of *ACTA2*, *POSTN* and *HOPX*, genes reported to be extensively expressed by myCAFs and myCAF subtypes [12]. We identified cluster 2 as complement-secreting CAFs (csCAFs) due to their remarkable expression of *C3* and *C7* [20, 51]. While the expression of complement has been reported by iCAFs and even been used to annotate them, CAFs secreting complement have reported as a distinct phenotype [20, 51, 56]. Further, although all clusters did not show as high an expression of interleukins as we would expect based on other single-cell studies, we noted differences in the expression of various interleukins between csCAFs and cluster 1, which we identified as iCAFs, including a higher expression of *IL-6* and *CXCL8* by iCAFs [11, 12, 57]. Finally, although other CAF subtypes with unique markers and secretory profiles have been reported, we were unable to isolate significant numbers of these phenotypes from single-cell analyses [18, 19]. These data suggest that myCAFs, iCAFs and csCAFs are the dominant CAF subtypes in PDAC.

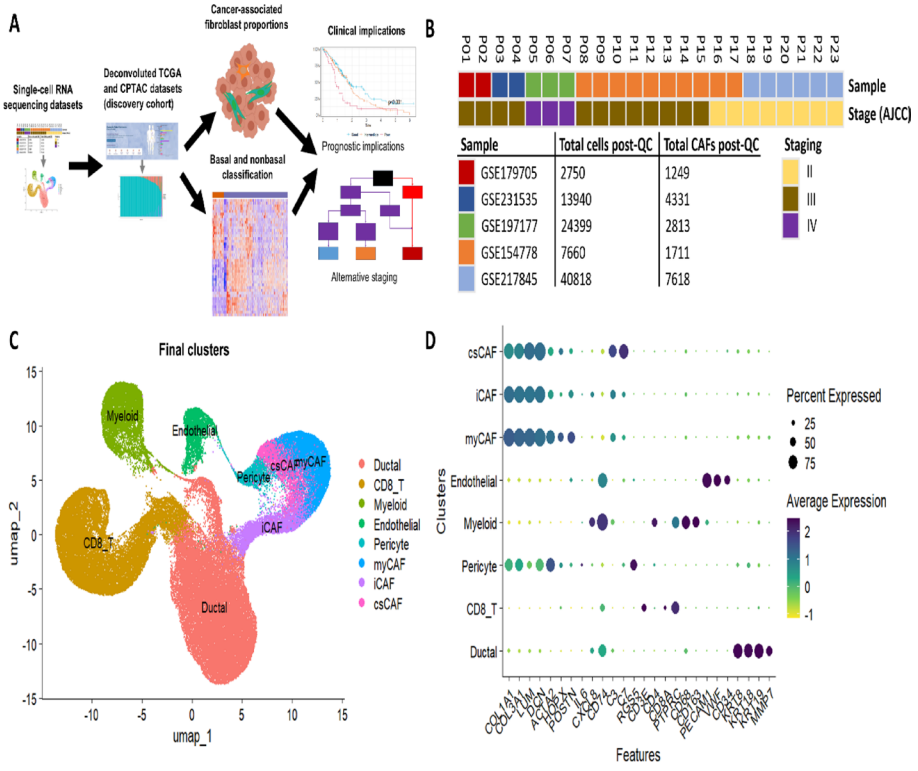

Fig. 1 Study design and analysis of scRNA-seq data in PDAC. **A** Overview of study design and workflow. scRNA-seq data of PDAC samples were collated, integrated, and clustered. Transcriptomic data of each cluster was subsequently used to deconvolute bulk transcriptomic data of PDAC samples from Genome Data Commons and obtain proportions of various cell types identified from the single-cell analysis. Transcriptomic data of deconvoluted cell types and proportions of different populations of cancer-associated fibroblasts (CAFs), as well as molecular consensus signatures, were examined for their prognostic implications. **B** Overview and characteristics of samples incorporated in the scRNA-seq analysis. The AJCC staging, total number of cells after quality control (QC) and total number of CAFs for twenty-three PDAC samples (each from a unique patient) across five studies are displayed. **C** 2-dimensional uniform manifold approximation and projection (UMAP) plot of the single-cell dataset. All single-cell data is plotted along the first two UMAP axes and colored by cell type clustered by Seurat and identified manually from transcriptomic profiles. **D** Dotplot of the single-cell data showing the expression of key gene markers in each cell cluster. The proportion of cells in a cluster expressing each gene is illustrated by the size of the respective circle and the mean expression of each gene per cell that expresses the gene within each cluster is indicated by the color of the circle in accordance with the color gradient legend

3.2 Molecular subtyping of TCGA samples

We molecularly subtyped discovery cohort cases (n = 320) through two approaches – firstly based on the mMC *O'Kane et al.*, and secondly based on CAF fractions [4, 30]. Tumor subtyping revealed that the majority of cases were nonbasal (300 of 320, 93.8%), in alignment with previous investigations (Fig. 2A) [4]. We validated tumor subtyping using three surrogate markers (*GATA6*, *HMGA2* and *KRT5*) previously reported to distinguish classical-like and basal-like tumors (Fig. 2C) [4, 58, 59]. *GATA6* showed a higher expression among nonbasal cases ($\log_2$ fold change 2.07, $p < 0.0001$), while *HMGA2* ($\log_2$ fold change 2.51, $p < 0.0001$) and *KRT5* ($\log_2$ fold change 5.72, $p < 0.0001$) were expectedly more highly expressed among basal cases, validating the robust tumor subtyping of the discovery dataset.

Next, we imputed cell fractions of phenotypes annotated in our single-cell analysis within the discovery cohort using CIBERSORTx. Among the CAF subtypes which were imputed, myCAFs formed a majority, with iCAFs forming a minority and csCAFs being nearly absent (Fig. 2B). The fraction of myCAFs was marginally elevated in stage III and

Table 2 Characteristics of single-cell RNA-seq samples

Characteristics	*N* (%)
Total number	23 (100)
Sex	
Female	7 (30.4)
Male	11 (47.8)
Not reported	5 (21.7)
Mean age (standard deviation) ^	65.4 (9.7)
Specimen retrieval modality	
Whipple's procedure	1 (4.3)
Distal pancreatectomy	3 (13.0)
Total pancreatectomy	1 (4.3)
Unspecified surgical resection	10 (43.5)
Not reported	2 (8.7)
AJCC Staging ^	
Stage I	0 (0.0)
Stage II	8 (34.8)
Stage III	7 (30.4)
Stage IV	3 (13.0)
Not reported	5 (21.7)
AJCC T stage	
T1	1 (4.3)
T2	6 (26.1)
T3	3 (13.0)
T4	0 (0.0)
Not reported	15 (65.2)
AJCC N stage	
N0	0 (0.0)
N1	2 (8.7)
N2	6 (26.1)
Not reported	15 (65.2)
AJCC M stage	
M0	15 (65.2)
M1	3 (13.0)
Not reported	5 (21.7)
Primary tumor histological grade	
Grade 1	1 (4.3)
Grade 2	12 (52.2)
Grade 3	3 (13.0)
Grade 4	2 (8.7)
Unspecified	5 (21.7)
Primary tumor site	
Head or uncinate process of pancreas	7 (30.4)
Body or tail of pancreas	1 (4.3)
Unspecified	15 (65.2)
Chemotherapy received	
FOLFIRINOX	2 (8.7)
Gemcitabine + nab-paclitaxel	1 (4.3)
None	3 (13.0)
Unspecified	17 (73.9)

^ Reported only for 18 of 23 patients

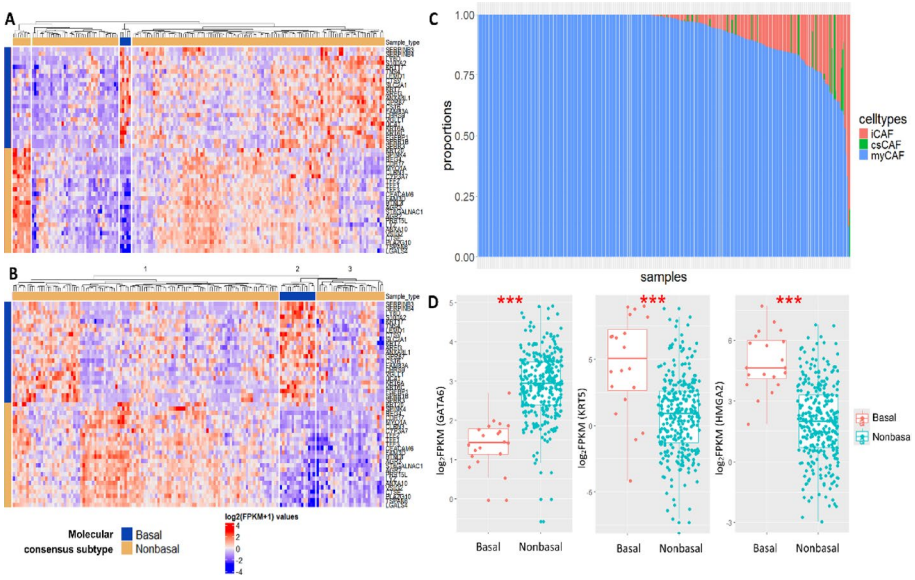

Fig. 2 Composition of discovery dataset based on hierarchical clustering, deconvolution, and expression of key markers. Hierarchical clustering heatmap of (**A**) 162 unique PDAC cases from the TCGA cohort and (**B**) 152 unique PDAC cases from the CPTAC cohort using a 47-gene signature defined by O'Kane et al. into two classes – basal and nonbasal. The top dendrogram reflects the similarity in gene signatures between different samples. The color of the left and top bar adjacent to the heatmap reflects the association of genes and samples respectively to the classes (dark blue: basal; gold: nonbasal). Expression intensity is reflected by a color gradient. **C** Stacked bar chart showing the deconvoluted proportions of each CAF type and pericytes across TCGA and CPTAC cases. Deconvolution was performed using CIBERSORTx on TCGA and CPTAC cases separately with our single-cell clustering acting as a reference of expression across cell types. **D** Box-and-whisker plots of the expression of three genes – *GATA6*, *HMGA2* and *KRT5*, reported to effectively distinguish basal and classical type PDACs across all 320 TCGA and CPTAC cases

IV tumors compared with stage I and II tumors (fold change 1.14, $p = 0.0035$), while the iCAF fraction was greater in stage I and II tumors (fold change 1.47, $p = 0.019$) (Fig. 3A-B). On the other hand, AJCC tumor, nodal and metastasis staging were largely not significantly associated with differences in myCAF or iCAF fractions apart from T3/4 tumors showing a higher iCAF fraction relative to T1/2 tumors (fold change 1.45, $p = 0.044$) (Fig. 3A-B, Supplementary Figure S3A-B). However, both myCAF fractions (fold change 1.20, $p = 0.014$) and iCAF fractions (fold change 2.72, $p = 0.00069$) were elevated in basal cases relative to nonbasal cases, suggesting that CAFs may form a significantly larger component of the stroma in basal PDAC (Fig. 3A-B). Together, these findings suggest that CAF fractions may have various associations with PDAC tumor subtypes but are largely independent of clinical staging parameters.

3.3 Myofibroblast fractions and molecular consensus subtypes may be useful prognostic tools

To determine whether CAF fractions and molecular signatures might be useful in prognostication, we separately stratified the discovery cohort based on molecular consensus subtype (basal, nonbasal), myCAF fraction, and iCAF fraction. myCAF and iCAF proportions were inversely correlated (Spearman's rho = -0.143, $p = 0.0105$). We determined the optimal thresholds for CAF fractions predicting overall survival (OS) by optimizing the AUC of the ROC curves for each stratum (Supplementary Figure S2A-D). A total of 178 of 320 (55.6%) cases were classified as myCAF-enriched, and 184 of 320 (57.5%) were classified as iCAF-depleted (Table 3). While OS did not significantly differ between

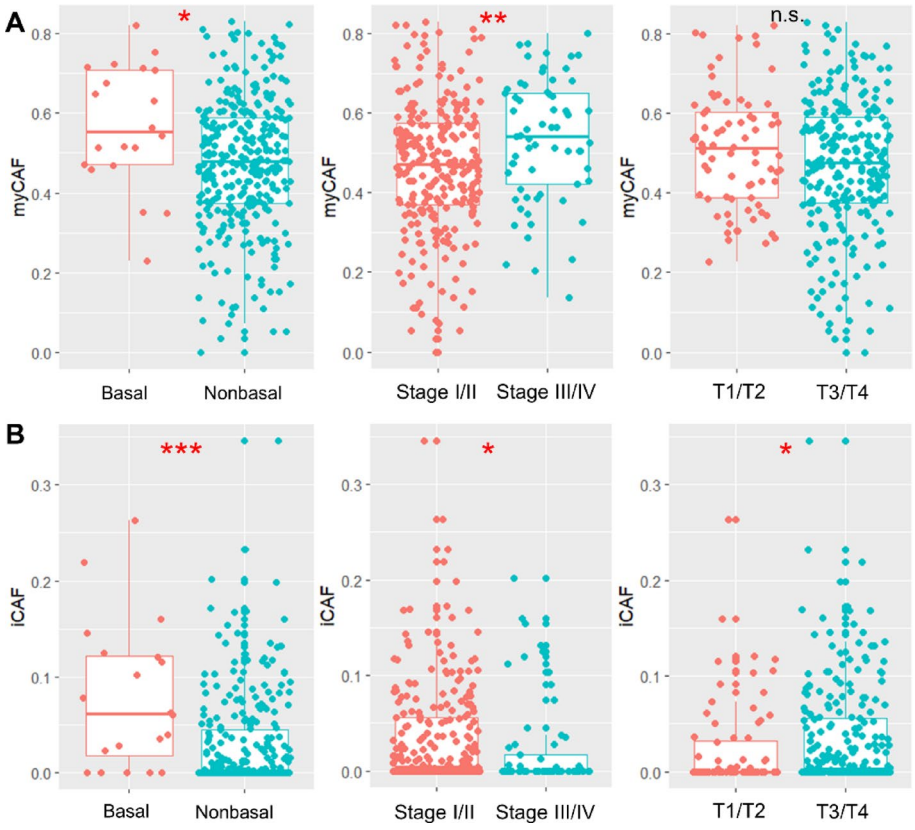

Fig. 3 Box-and-whisker plots of (**A**) myCAF and (**B**) iCAF proportions of PDAC samples in the discovery cohort (TCGA and CPTAC). myCAF and iCAF proportions were higher in basal subtypes. myCAF proportions were higher in advanced stage disease, whereas iCAF proportions were higher in early-stage disease. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, n.s. non-significant

iCAF-enriched and iCAF-depleted cases (HR: 1.03 [0.78–1.36], $p = 0.8$), myCAF-enriched cases demonstrated a significantly poorer OS compared with myCAF-depleted cases (HR: 1.33 [1.01–1.75], $p = 0.042$) (Fig. 4A and B). The basal subtype was associated with a lower OS than the nonbasal subgroup (HR: 2.17 [1.32–3.58], $p = 0.002$) (Fig. 4C).

On this basis, we reasoned that incorporating information on myCAF fractions and mMC subtyping might augment the prognostic accuracy of existing staging systems. We therefore modelled myCAF enrichment status, mMC status, as well as AJCC tumor, nodal and metastasis statuses of discovery cohort cases with staging data in a multivariable Cox proportional hazards model. This modelling revealed that basal subtype (HR: 1.85 [1.10–3.14], $p = 0.0215$) was the only significant risk factor modifying OS (Supplementary Table 1). Given the relatively larger size of the nonbasal subgroup, we reasoned that some of the remaining 4 variables might be helpful in predicting prognosis within this subgroup. By further modelling these 4 variables on nonbasal samples only ($n = 300$), we found that myCAF enrichment (HR: 1.34 [1.00–1.80], $p = 0.0495$) had the most significant effects in predicting poor OS, while nodal status (HR: 1.39 [0.969–2.00], $p = 0.0733$) or metastasis (HR: 1.85 [0.939–3.66], $p = 0.0752$) showed only nominal statistical significance and tumor size was insignificant (HR: 1.06 [0.814–1.38], $p = 0.665$) (Supplementary Table 1). Given these findings, we surmised that mMC subtyping, nodal status, metastasis and myCAF enrichment might be useful variables for prognosticating PDAC.

Table 3 Characteristics of myCAF enriched and depleted cases

Characteristics	myCAF-enriched (%)	myCAF-depleted (%)	*p*-value
Total number	178 (100)	142 (100)	
Mean age of diagnosis (standard deviation)	66.4 (10.8)	64.0 (10.9)	0.048
Female sex	85 (47.8)	62 (43.7)	0.466
Ethnicity ^			0.289
White (including Hispanics)	127 (71.3)	107 (75.4)	
Black or African American	3 (1.7)	5 (3.5)	
Asian	18 (10.1)	17 (12.0)	
Others	28 (15.7)	12 (8.5)	
Not reported	2 (1.1)	1 (0.7)	
Alive at last follow-up	45 (25.3)	60 (33.7)	0.001
AJCC Staging ^			
Stage I	20 (11.2)	14 (9.9)	0.223
Stage II	114 (64.0)	107 (75.4)	
Stage III	32 (18.0)	15 (10.6)	
Stage IV	11 (6.2)	5 (3.5)	
Not Reported	1 (0.6)	1 (0.7)	
AJCC T stage			
T1	9 (5.1)	3 (2.1)	0.326
T2	37 (20.8)	27 (19.0)	
T3	124 (69.7)	106 (74.6)	
T4	6 (3.4)	2 (1.4)	
Not reported	2 (1.1)	4 (2.8)	
AJCC N stage			
N0	28 (25.9)	55 (24.2)	0.629
N1	64 (59.3)	150 (66.1)	
N2	14 (13.0)	13 (5.7)	
NX	2 (1.9)	6 (2.6)	
AJCC M stage			
M0	167 (93.8)	137 (96.5)	0.409
M1	11 (6.2)	5 (3.5)	
MX	58 (32.6)	64 (45.1)	
Primary tumor site ^			
Head of pancreas	126 (70.8)	116 (81.6)	0.031
Body of pancreas	15 (8.4)	11 (7.7)	
Tail of pancreas	18 (10.1)	6 (4.2)	
Overlapping lesion of pancreas	0 (0)	2 (1.4)	
Not otherwise specified	19 (10.7)	7 (4.9)	

In addition, using the subset of our discovery cohort which had received some form of chemotherapy ($n = 88$), we attempted to determine the ability of these molecular classifiers to predict outcomes following chemotherapy. However, neither mMC subtyping (HR: 1.93 [0.60–6.22], $p = 0.3$) nor myCAF enrichment (HR: 0.98 [0.57–1.70], $p > 0.9$) predicted OS (Supplementary Figure S3C-D). Given the relatively small sample of basal cases ($n = 4$) as well as lack of access to annotations that distinguished whether chemotherapy was delivered neoadjuvant or adjuvant, as well as actual treatment response data, future studies will be be necessary to truly determine the potential of these molecular classifiers as predictors of chemotherapeutic response.

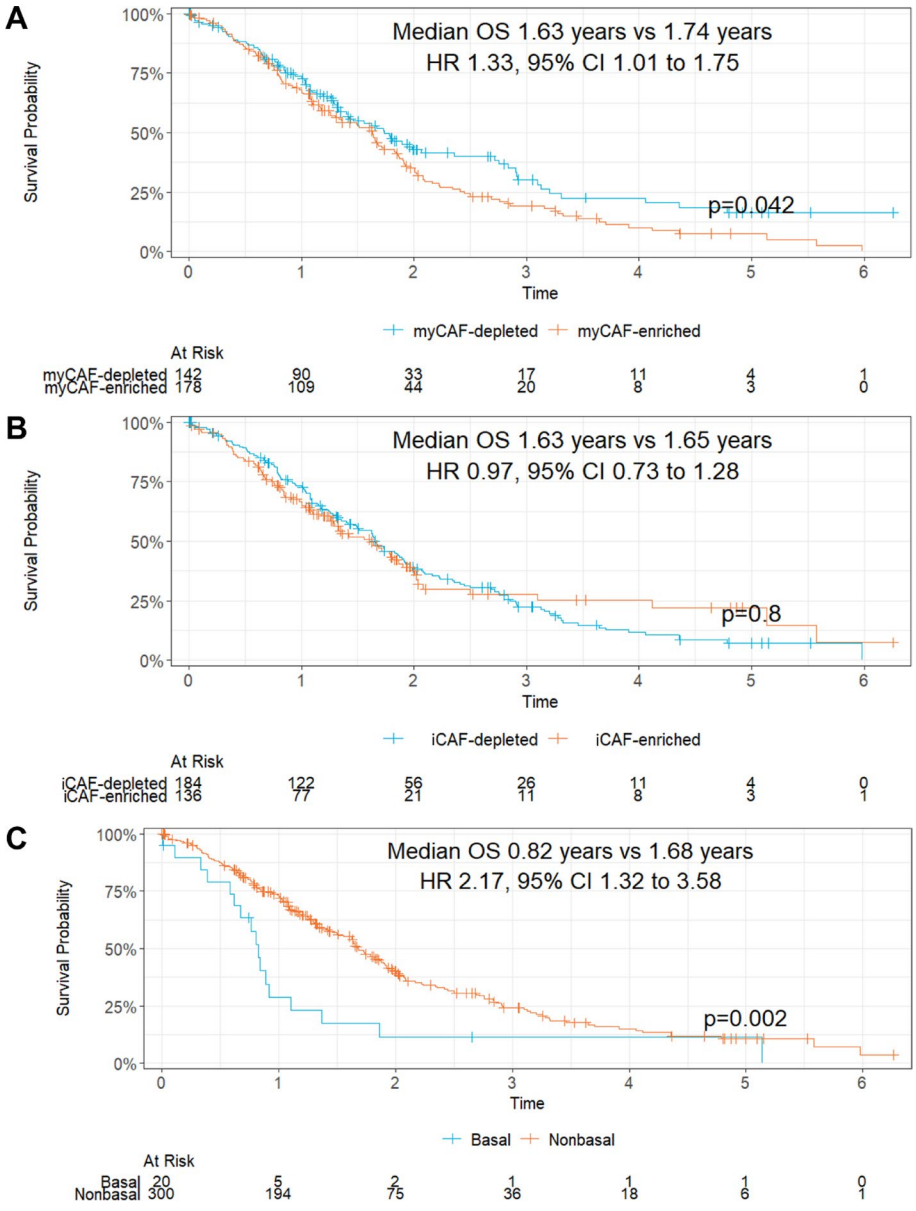

Fig. 4 Kaplan-Meier survival plots of individuals with PDAC registered in the discovery cohort. Overall survival by **A** myCAF fraction, **B** iCAF fraction, **C** and basal and nonbasal classification in accordance with the modified Moffitt gene signature

3.4 Tumor-Microenvironment Integrated (TMI) staging system effectively prognosticates PDAC survival durations

Based on the potential for mMC, metastasis and myCAF enrichment to prognosticate PDAC, we established a TMI staging system (Fig. 5A, Supplementary Figure S4). Nodal status was not included in TMI as subsequent validation data lacked nodal staging information. The TMI staging system first assigns any cases of the basal subtype or with metastasis as "Poor" ($n = 33$). Nonbasal cases are then divided into myCAF-enriched cases grouped as "Intermediate" ($n = 154$) and myCAF-depleted cases which are grouped as "Good" ($n = 133$) (Fig. 5B). Two cases were filtered out from these 320 cases for subsequent model comparisons as their AJCC stages were unknown. According to our TMI staging system, TMI-Good cases bear a median survival of 1.78 years and 3-year OS of

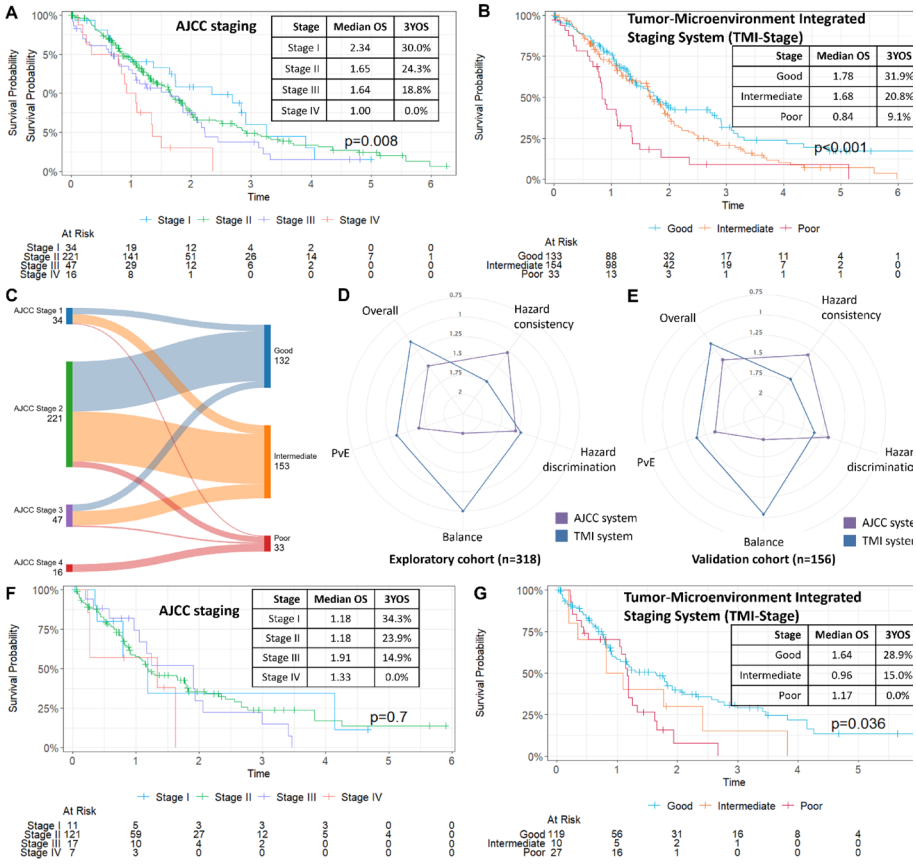

Fig. 5 Performance of the Tumor-Microenvironment Integrated (TMI) staging system, a system of staging and prognostication which integrates the AJCC staging system together with CAF proportion and molecular consensus classification of PDAC. Overall survival of patients with PDAC in the discovery (TCGA and CPTAC) dataset segregated by **A** AJCC staging and **B** TMI staging systems. Median OS reflected for each stage in the table at the top-right. **C** Alluvial diagram showing the reassignment of PDAC data between the AJCC and TMI staging systems. Radar maps of ranked performance metrics of the AJCC and TMI staging systems for (**D**) the discovery cohort and (**E**) the validation cohort. Performance metrics are defined as per Xie et al. and a lower rank (further outward) indicates better performance. Overall survival of patients with PDAC in the validation cohort segregated by **F** AJCC staging and **G** TMI staging systems. Median OS reflected for each stage in the table at the top-right

31.9%, compared to that of the TMI-Intermediate (median OS 1.68 years, 3-year OS 20.8%) and TMI-Poor (median OS 0.84 years, 3-year OS 9.10%) cases, log rank $p < 0.001$) (Fig. 5B).

We next looked for and validated performance differences between TMI staging and AJCC staging systems. Log-rank testing of AJCC (p = 0.008) and TMI (p < 0.001) systems suggested that both models contained significant inter-staging differences on Kaplan-Meier curves. TMI (0.558 ± 0.0205) showed superiority to AJCC (0.547 ± 0.0191) in ranking the risk of death according to Harrell's C-index. Additionally, we used recursive partitioning analysis (RPA) via autoRPA, a web-based tool that has previously been applied to stage and analyze nasopharyngeal and hepatocellular carcinoma staging systems, to compare TMI and AJCC staging within the discovery cohort using four indicators (Fig. 5D) [60]. AutoRPA uses four metrics – hazard discrimination, the extent of difference in duration to outcome between stages, hazard consistency, the homogeneity of durations to outcome within each stage, PvE, the percentage of variance in durations to outcome that is accounted for by model staging, and balance, the proportionality of case division between stages. Both TMI and AJCC showed similar hazard

discrimination, and while AJCC had a superior performance in hazard consistency, TMI showed superior performance in both PvE and balance.

To validate these findings and determine whether TMI and our methodology might apply as well with microarray data, we applied TMI to a validation cohort of 156 PDAC cases from a series of microarray studies. Similar to our discovery cohort, our validation cohort comprised primarily of nonbasal cases (135 of 156, 86.5%) but was notably sparse for myCAF-enriched cases (12 of 156, 7.69%) which we reasoned might be due to platform differences in detecting low-abundance transcripts between microarray and RNA-seq that CIBERSORTx (Supplementary Figure S5A-B). Consequently, we observed a significantly lopsided distribution in staging as well. "TMI-Good" ($n = 119$) cases were the most numerous (median OS 1.64 years, 3-year OS 28.9%), while "TMI-Intermediate" ($n = 10$) cases (median OS 0.96 years, 3-year OS 15.0%) and "TMI-Poor" ($n = 27$) cases (median OS 1.17 years, 3-year OS 0.00%) were few (Fig. 5F-G). Surprisingly, while log-rank testing showed a statistically significant difference in OS between TMI stages ($p = 0.036$), this was not the case for AJCC stages ($p = 0.7$). Analysis via autoRPA showed that TMI was comparable to AJCC, showing superiority to AJCC across both PvE and balance, while AJCC maintained an advantage in hazard discrimination and hazard discrimination (Fig. 5E). Conversely, Harrell's C-index was higher in TMI (0.535 ± 0.0223) than AJCC (0.500 ± 0.0223). Together, these results suggest that TMI has the capacity to prognosticate PDAC cases.

3.5 Fibronectin 1 signaling is a top pathway outgoing from myofibroblasts that prognosticates poor survival

Based on the difference in prognoses between different tumor subtypes, we sought to identify differences in molecular signaling between basal and nonbasal tumors. Akin to the bulk sequencing data, we performed hierarchical clustering on our single-cell dataset with gene expression summated across all cells per sample using the mMC gene panel and identified 3 of the 23 single-cell samples as basal tumors (Supplementary Fig. 6). Using CellChat, we first looked for the top signaling pathways within our overall single-cell dataset and determined the intensity of signaling from each cell cluster. The top three signaling pathways comprised the ECM proteins of collagen, fibronectin 1 (FN1) and laminin, and the next strongest signaling pathways included the transmembrane protein CD99 and neurite-growth promoting factor 2 (NEGF2/MK). Looking at an overview of all cell types, all signaling pathways were primarily outgoing from myCAFs and received by ductal cells, which includes mainly the tumor population, except for CD99 and MK which were received primarily by myCAFs (Fig. 6A). Basal tumors showed a higher intensity of molecular signaling for collagen, laminins and FN1 compared with nonbasal tumors, with signaling predominantly from myCAFs and pericytes to ductal cells (Figs. 6B-E). A subset of collagen and FN1 signaling additionally redirected back to myCAFs, suggestive of autocrine signaling by ECM proteins that might shape the tumor microenvironment.

We therefore looked at associations between CAF proportions and ECM protein expression and identified a unique positive association between summated FN1 expression in the single-cell samples ($R = 0.48$, $p = 0.023$) and the proportion of myCAFs making up the total CAF population (Supplementary Figure S7A). FN1 expression was also significantly higher in basal (myCAF-depleted or enriched) cases, and in nonbasal,

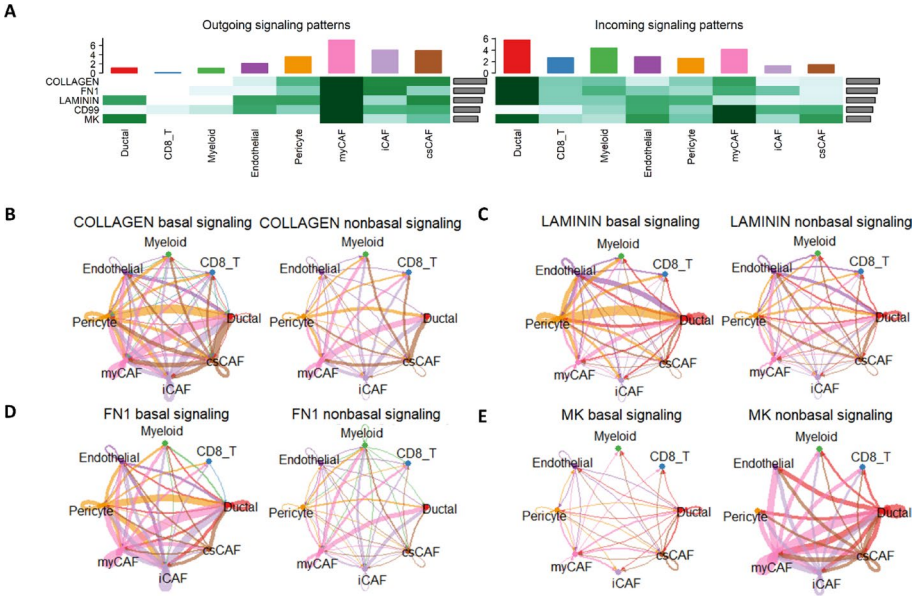

Fig. 6 CellChat analysis of signaling pathways in the single-cell dataset. **A** Outgoing and incoming signaling patterns of the top five gene classes in CellChat across all single-cell clusters. The intensity of green reflects the relative signaling strength of a signaling pathway across cell types (across the row), the colored bar on top reflects the total signaling strength of a cell type relative to others and the right grey bar reflects the total signaling strength of signaling pathways across all cell types relative to other signaling pathways. **B** Circle plot of collagen signaling between cell types in basal and nonbasal single-cell samples. Signaling is a summation across all collagen genes. **C** Circle plot of laminin signaling between cell types in basal and nonbasal single-cell samples. Signaling is a summation across all laminin genes. **D** Circle plot of fibronectin 1 (FN1) signaling between cell types in basal and nonbasal single-cell samples. **E** Circle plot of midkine (MK) signaling between cell types in basal and nonbasal single-cell samples. For each arrow in Fig. 6A and E, the arrowhead delineates the cell type which is receiving the signaling. The width of the arrow illustrates the strength in signaling

myCAF-enriched cases over nonbasal, myCAF-depleted cases, and by extension equally elevated in both TMI-Poor and TMI-Intermediate cases relative to TMI-Good cases (Supplementary Figure S7B-C). Looking at signaling pathways specifically in basal tumors, we also noticed that FN1 remained most heavily expressed by myCAFs, though other CAF types also contributed, and that a significant subset of FN1 signaling in basal tumors was directed toward iCAFs (Fig. 6D, Supplementary Figure S7D-E). This suggests that even in basal, myCAF-depleted tumors, myCAFs continue to drive elevated FN1 expression. Most notably, FN1-enriched cases also showed a significantly poorer OS compared with FN1-depleted cases (Supplementary Figure S7F). Put together, these findings suggest FN1 might be a myCAF-specific positive regulator promulgating myCAF-enriched microenvironments that interact with both tumor and other stromal components to worsen prognosis. Interestingly, not only was FN1 positively associated with the M2 macrophage marker CD163 in bulk expression (R = 0.36, $p < 0.0001$), but FN1 was also a notable incoming pathway to myeloid cells within nonbasal single-cell tumors, further raising the possibility of a role for FN1-expressing myCAFs in establishing an immunosuppressive niche comprising M2 macrophages as other reports have similarly posited (Supplementary Figure S7E, S7G) [13, 61].

4 Discussion

The prognostication of PDAC is currently a clinical task achieved at the time of staging scans and gross tumor measurement [3]. We utilized a rational approach in developing TMI, employing Cox proportional hazards modelling to identify the most statistically significant variables for use. Our Cox proportional hazards model and median OS of the resultant TMI-Poor stage highlighted the usefulness of the mMC signature in distinguishing basal samples for prognostication, reflecting poor survival regardless of metastasis status. The enrichment of myCAF also demonstrated good use in providing a distinction in survival between TMI-Good and TMI-Intermediate cases of the remaining nonbasal cases, being more significant than any of the remaining individual clinical variables. Our performance metrics indicate that TMI staging provides a better estimate of survival when mMC signatures and the stromal make-up of the myCAF subtype are incorporated. Particularly in the exploratory discovery cohort, many cases from AJCC stage II and stage III, which bore a similar median OS, were re-classified into the TMI-Intermediate subgroup based on myCAF enrichment, while the TMI-Poor subgroup was comparable to AJCC stage IV in median OS. TMI may therefore prove valuable as an accurate tool to manage expectations and prepare for potential disease progression, especially for more advanced PDACs which constitute the majority of cases.

Nonetheless, TMI faced drawbacks in hazard consistency when compared to AJCC in stratifying the discovery cohort. The combination of a subset of AJCC stage I and II cases into TMI-Good could be the cause of this, creating heterogenous subgroups in TMI-Good with distinct overall survivals. Additionally, although statistically distinct, the median OS of TMI-Good and TMI-Intermediate cases did not have much magnitude of difference in the discovery cohort, though this was not the case in the validation cohort. Moreover, in both cohorts, TMI-Good never achieved a median OS as high as that of AJCC stage I. These findings indicate that while myCAF enrichment might be useful in characterizing intermediate prognosis cases just as well as AJCC stages II and III, incorporation of the clinical T1 staging variable might still yield prognostic benefit for a small subset of good prognosis cases where tumor size remains a better predictor of overall survival than molecular characteristics.

Previous work has also demonstrated that CAF fractions have clinical implications on responses to treatment regimens involving gemcitabine as well as anti-PD-1 immunotherapy, underscoring their potential use in the selection of appropriate patient populations for specific chemotherapies [62–65]. Although the viability of surgical resection remains the key clinical consideration in the management of PDAC patients, chemotherapy still plays a major role in the treatment of both resectable and unresectable PDAC [66–71]. Although the lack of difference in OS between myCAF enrichment status or tumor subtype suggests that these molecular classifiers might be less applicable for predicting chemotherapeutic response, our dataset lacked information of the specific chemotherapy regimens and whether treatment was neoadjuvant or adjuvant in each case. In vitro, CAFs have been reported to reprogramme macrophages to shield against FOLFIRINOX-induced cell death [65]. Conversely, a comparison between upfront resected pancreatic carcinomas and pancreatic carcinomas with neoadjuvant FOLFIRINOX therapy prior to resection using two surrogates for myCAF density – activated stroma index and ACTA2 expression – reported that cases treated with neoadjuvant FOLFIRINOX in the highest quartile of the activated stroma index had the highest survival rates,

alongside cases with relatively higher ACTA2 expression [62]. Therefore, it may be worth exploring myCAF fractions as a means to determine patient suitability for FOLFIRINOX neoadjuvant therapy. Similarly, CAFs may also play a role in resistance to immunotherapies such as PD-1 inhibitors, which are being increasingly used to treat metastatic PDAC that are mismatch repair deficient or microsatellite instability-high. MEK and STAT3 inhibitors, which reduce iCAF polarization and increase Ly6a/Cd34-expressing CAFs, have been shown to augment the anti-tumor effect of PD-1 blockade and survival in a PDAC mouse model [72]. A finer understanding of CAF-driven mechanisms of chemo- and immuno-resistance may therefore be fruitful to define the CAF make-ups of PDAC that worsen or benefit prognosis.

In TMI staging, we made use of the association we observed from the discovery dataset between myCAF-enriched signatures and poorer survival. Tumors with higher myCAF fractions, which included more basal subtype tumors as well as nonbasal, myCAF-enriched subtype tumors, bore significantly poorer OS compared to nonbasal, myCAF-depleted tumors. Interestingly, of the top signaling pathways we identified in our single-cell analysis, the ECM protein FN1, which we found relatively enriched in both basal tumors as well as nonbasal, myCAF-enriched tumors, is known to augment transforming growth factor beta (TGFβ) expression to drive myofibroblastic polarization, suggesting that FN1 might be a key signaling pathway in the polarization of CAFs into myCAFs to facilitate tumors with a poorer prognosis [11, 73–75]. This might explain our findings that the proportional makeup of CAFs by myCAFs was positively related to FN1 expression in the single-cell dataset and the poor survival of FN1-high cases within the discovery cohort. The fact that basal tumors had similarly elevated levels of overall FN1 expression (largely originating from myCAFs as well) compared to nonbasal, myCAF-enriched tumors might also align with the possibility that FN1 signaling to ductal tumor cells drives the poor prognosis. Recent work supports this position, as in vitro co-cultures of PDAC cells and CAFs show that FN1 is heavily upregulated in metastatic CAF subsets and tied to increased invasiveness in migration assays [61]. It is also worth considering that this mechanism might involve immunosuppression, as FN1 has been observed as a key marker of $POSTN^+$ CAFs which have been spatially linked with immunosuppressive M2-like $SPP1^+$ macrophages [13]. Ultimately, further work will be necessary to determine key pathways, of which FN1 might constitute one such example, that provide an integrated explanation to PDAC aggression and worsened prognosis in both basal tumors and nonbasal, myCAF-enriched tumors.

Regardless, the relationship between the myCAF stromal population and pancreatic cancer progression and prognosis remains complex and uncertain. Historically, the dense stroma formed by myCAFs was thought to impede drug delivery and thus, chemotherapeutic response, but mouse models later demonstrated increased aggression and metastasis in tumors where myCAFs were ablated through the deletion or inhibition of the embryonic signaling *Sonic hedgehog* (*SHH*), likely by a variety of mechanisms not limited to the alteration of local angiogenesis and immune cell function [76–79]. Since then, myCAF and by extension, CAF heterogeneity has been a major avenue of exploration, with different CAF subpopulations shown to operate through distinct mechanisms to yield varying prognoses [14]. For instance, the selective depletion of *leucine-rich-repeat-containing protein 15* ($LRRC15)^+$ myCAFs was shown to improve the response rates to anti-PD-L1 immunotherapy [80]. In a separate study, the ablation of a

subpopulation of CD90$^-$ myCAFs by ERBB2 inhibition was further found to reduce lung and diaphragmatic metastases in orthotropic mouse models of PDAC [81]. Therefore, further work will be necessary to elucidate the functional heterogeneity of myCAFs in modifying PDAC aggression and identify markers defining key subpopulations contributing to tumor aggression.

In a similar vein, while novel CAF subtypes have been identified with immunomodulatory effects and linked with changes in prognosis in retrospective cohort studies, their heterogeneity and plasticity might raise the issue of how best to arrive at key CAF subtypes to therapeutically target [11–13, 19, 81, 82]. The plasticity of CAFs may enable different subtypes to be isolated and cultured in large quantities in vitro, but their actual fractions might be substantially smaller, especially for subtypes which heavily overlap [83]. For example, outside of LRRC15$^+$ and epidermal growth factor receptor (EGFR)$^+$ myCAFs, Pleckstrin Homology-Like Domain Family A Member 1 (PHLDA1)$^+$ CAFs are technically defined from matrix CAFs expressing both fibroblast activation protein (FAP) as well as POSTN, among other genes, and colocalize in spatial transcriptomic assays with POSTN$^+$ CAFs, which were originally reported to drive immunosuppressive macrophage recruitment in PDAC [13, 84, 85]. Further, discretely classifying CAFs becomes challenging in cases when markers overlap. For instance, endothelial-like CAFs (endoCAFs) are a very recently defined CAF subtype in PDAC marked by FAP and CD144 that promote tumor proliferation and metastasis via the CD144-driven beta-catenin-STAT3 oncogenic axis [86]. Relating FAP$^+$ CAFs with FAP$^+$ POSTN$^+$ matrix CAFs, FAP$^+$ CD144$^+$ endothelial-like CAFs (endoCAFs) and POSTN$^+$ CAFs would consequently require attention to their functional effects to tumor cells and consideration of how their phenotypes might vacillate depending on microenvironmental factors [85–87]. Therefore, an integration of known CAF subtypes into overarching families alongside their functional outcomes to tumor aggression might be the next step.

5 Limitations

Although we were able to validate our results across both bulk RNA-seq and microarray expression data, it is unclear whether TMI can be translated to immunohistochemistry (IHC) analyses that are a frequently employed tool in the workup of pancreatic tumors [88]. Previous work has affirmed the use of IHC staining to detect from resected tumors CAF subtypes, such as iCAFs and antigen-presenting CAFs [12]. Moreover, surrogate markers like GATA6 in IHC staining has been validated to align with molecular subtypes and predict clinical outcomes [89–93]. Additional validation will be needed to investigate quantitative IHC in estimating CAF fractions to determine if it might offer the same predictive benefit that would warrant integration into regular histological workflows.

6 Conclusions

In conclusion, TMI staging suggests the potential incorporation of myCAF proportions and molecular consensus signatures in prognostication. Biologically, pro-metastatic signaling of FN1 from myCAFs to tumor cells and other CAFs may in part explain this phenomenon. Further validation on additional independent patient cohorts will be necessary to integrate it into existing practice and investigate its application to the selection of patients for adjuvant or neoadjuvant chemotherapy.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s12672-026-05091-0.

Supplementary Material 1.

Acknowledgements

The results published here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

Author contributions

SJY and ITS analyzed the data and drafted the manuscript; HYT provided bioinformatics support. DT and RS provided clinical insights. JYC designed the study, interpreted the results, and revised the manuscript; and all authors read and approved the final version of the manuscript.

Funding

This work was supported by the Singapore Ministry of Health's National Medical Research Council Research Transition Award (TA21jun-0005), Clinician Scientist Individual Research Grant (CIRG25jan-0007), Large Collaborative Grant (OFLCG-23May0039), and TETRAD II Collaborative Centre Grant (CG21APR2002), SingHealth Duke-NUS AM/ACP-Designated Philanthropic Fund Grant Award (08/FY2023/EX/27-A65), Tanoto Foundation, NCCS Cancer fund, as well as the Khoo Bridge Funding Award (Duke-NUS-KBrFA/2025/0090) provided by Duke-NUS Medical School and the "Estate of Tan Sri Khoo Teck Puat".

Data availability

All datasets employed in this study are freely available from Genome Expression Omnibus (GEO): GSE179705, GSE231535, GSE197177, GSE154778, GSE217845, GSE183795 and GSE62452.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 6 January 2026 / Accepted: 22 April 2026

Published online: 25 April 2026

References

1. American Cancer Society. (2025) Cancer Facts & Figs. 2025. American Cancer Society, Atlanta.
2. Siegel RL, Kratzer TB, Wagle NS, et al. Cancer statistics, 2026. CA: A Cancer. J Clin. 2026;76:e70043. https://doi.org/10.3322/caac.70043.
3. Allen PJ, Kuk D, Castillo CF, et al. Multi-Institutional validation study of the American Joint Commission on Cancer (8th edition) changes for T and N staging in patients with pancreatic adenocarcinoma. Ann Surg. 2017;265:185–91. https://doi.org/10.1097/SLA.0000000000001763.
4. O'Kane GM, Grünwald BT, Jang G-H, et al. GATA6 Expression Distinguishes Classical and Basal-like Subtypes in Advanced Pancreatic Cancer. Clin Cancer Res. 2020;26:4901–10. https://doi.org/10.1158/1078-0432.CCR-19-3724.
5. Bailey P, Chang DK, Nones K, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature. 2016;531:47–52. https://doi.org/10.1038/nature16965.
6. Moffitt RA, Marayati R, Flate EL, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet. 2015;47:1168–78. https://doi.org/10.1038/ng.3398.
7. Collisson EA, Sadanandam A, Olson P, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med. 2011;17:500–3. https://doi.org/10.1038/nm.2344.
8. Ohara Y, Tang W, Liu H, et al. SERPINB3-MYC axis induces the basal-like/squamous subtype and enhances disease progression in pancreatic cancer. Cell Rep. 2023;42:113434. https://doi.org/10.1016/j.celrep.2023.113434.
9. Chan-Seng-Yue M, Kim JC, Wilson GW, et al. Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. Nat Genet. 2020;52:231–40. https://doi.org/10.1038/s41588-019-0566-9.
10. Sahai E, Astsaturov I, Cukierman E, et al. A framework for advancing our understanding of cancer-associated fibroblasts. Nat Rev Cancer. 2020;20:174–86. https://doi.org/10.1038/s41568-019-0238-1.
11. Biffi G, Oni TE, Spielman B, et al. IL1-Induced JAK/STAT Signaling Is Antagonized by TGFβ to Shape CAF Heterogeneity in Pancreatic Ductal Adenocarcinoma. Cancer Discov. 2019;9:282–301. https://doi.org/10.1158/2159-8290.CD-18-0710.
12. Elyada E, Bolisetty M, Laise P, et al. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. Cancer Discov. 2019;9:1102–23. https://doi.org/10.1158/2159-8290.CD-19-0094.
13. Wu J, Dai T, Li Z, et al. Integrating multi-modal transcriptomics identifies cellular subtypes with distinct roles in PDAC progression. Cell Oncol (Dordr). 2025;48:1571–92. https://doi.org/10.1007/s13402-025-01100-6.
14. Brichkina A, Polo P, Sharma SD, et al. A Quick Guide to CAF Subtypes in Pancreatic Cancer. Cancers (Basel). 2023;15:2614. https://doi.org/10.3390/cancers15092614.

15. Hu B, Wu C, Mao H, et al. Subpopulations of cancer-associated fibroblasts link the prognosis and metabolic features of pancreatic ductal adenocarcinoma. Ann Transl Med. 2022;10:262. https://doi.org/10.21037/atm-22-407.
16. Li B, Pei G, Yao J, et al. Cell-type deconvolution analysis identifies cancer-associated myofibroblast component as a poor prognostic factor in multiple cancer types. Oncogene. 2021;40:4686–94. https://doi.org/10.1038/s41388-021-01870-x.
17. Wang Y, Liang Y, Xu H, et al. Single-cell analysis of pancreatic ductal adenocarcinoma identifies a novel fibroblast subtype associated with poor prognosis but better immunotherapy response. Cell Discov. 2021;7:36. https://doi.org/10.1038/s41421-021-00271-4.
18. Dominguez CX, Müller S, Keerthivasan S, et al. Single-Cell RNA Sequencing Reveals Stromal Evolution into LRRC15 + Myofibroblasts as a Determinant of Patient Response to Cancer Immunotherapy. Cancer Discov. 2020;10:232–53. https://doi.org/10.1158/2159-8290.CD-19-0644.
19. Ge W, Yue M, Lin R, et al. PLA2G2A+ cancer-associated fibroblasts mediate pancreatic cancer immune escape via impeding antitumor immune response of CD8 + cytotoxic T cells. Cancer Lett. 2023;558:216095. https://doi.org/10.1016/j.canlet.2023.216095.
20. Chen K, Ma Y, Huang L, et al. Complement-secreting CAFs are associated with better prognosis in pancreatic cancer: single-cell multiomics. Gut. 2026. https://doi.org/10.1136/gutjnl-2025-335683.
21. The Cancer Genome Atlas Program (TCGA) - NCI. (2022). https://www.cancer.gov/ccg/research/genome-sequencing/tcga. Accessed 8 Jul 2025.
22. CPTAC-PDA. In: The Cancer Imaging Archive (TCIA). https://www.cancerimagingarchive.net/collection/cptac-pda/. Accessed 19 Mar 2026.
23. Heath AP, Ferretti V, Agrawal S, et al. The NCI Genomic Data Commons. Nat Genet. 2021;53:257–62. https://doi.org/10.1038/s41588-021-00791-5.
24. Yang S, Tang W, Azizian A, et al. Dysregulation of HNF1B/Clusterin axis enhances disease progression in a highly aggressive subset of pancreatic cancer patients. Carcinogenesis. 2022;43:1198–210. https://doi.org/10.1093/carcin/bgac092.
25. Yang S, He P, Wang J, et al. A novel MIF signaling pathway drives the malignant character of pancreatic cancer by targeting NR3C2. Cancer Res. 2016;76:3838–50. https://doi.org/10.1158/0008-5472.CAN-15-2841.
26. Clough E, Barrett T, Wilhite SE, et al. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. Nucleic Acids Res. 2023;52:D138–44. https://doi.org/10.1093/nar/gkad965.
27. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2025.
28. Posit team. (2025) RStudio: Integrated Development Environment for R.
29. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics. 2007;23:1846–7. https://doi.org/10.1093/bioinformatics/btm254.
30. Aung KL, Fischer SE, Denroche RE, et al. Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. Clin Cancer Res. 2018;24:1344–54. https://doi.org/10.1158/1078-0432.CCR-17-2994.
31. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32:2847–9. https://doi.org/10.1093/bioinformatics/btw313.
32. Steen CB, Liu CL, Alizadeh AA, Newman AM. Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. Methods Mol Biol. 2020;2117:135–57. https://doi.org/10.1007/978-1-0716-0301-7_7.
33. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics. 1988;44:837–45. https://doi.org/10.2307/2531595.
34. Oh K, Yoo YJ, Torre-Healy LA, et al. Coordinated single-cell tumor microenvironment dynamics reinforce pancreatic cancer subtype. Nat Commun. 2023;14:5226. https://doi.org/10.1038/s41467-023-40895-6.
35. Zhang S, Fang W, Zhou S, et al. Single cell transcriptomic analyses implicate an immunosuppressive tumor microenvironment in pancreatic cancer liver metastasis. Nat Commun. 2023;14:5123. https://doi.org/10.1038/s41467-023-40727-7.
36. Lin W, Noel P, Borazanci EH, et al. Single-cell transcriptome analysis of tumor and stromal compartments of pancreatic ductal adenocarcinoma primary tumors and metastatic lesions. Genome Med. 2020;12:80. https://doi.org/10.1186/s13073-020-00776-9.
37. Caronni N, La Terza F, Vittoria FM, et al. IL-1β + macrophages fuel pathogenic inflammation in pancreatic cancer. Nature. 2023;623:415–22. https://doi.org/10.1038/s41586-023-06685-2.
38. GEO Accession viewer. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE179705. Accessed 19 Mar 2026.
39. Hao Y, Stuart T, Kowalski MH, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42:293–304. https://doi.org/10.1038/s41587-023-01767-y.
40. Germain P-L, Lun A, Garcia Meixide C, et al. Doublet identification in single-cell sequencing data using scDblFinder. F1000Res. 2022;10:979. https://doi.org/10.12688/f1000research.73600.2.
41. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019;177:1888–902. https://doi.org/10.1016/j.cell.2019.05.031. e21.
42. Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. Nat Rev Immunol. 2012;12:191–200. https://doi.org/10.1038/nri3158.
43. Lau SK, Chu PG, Weiss LM. CD163: A Specific Marker of Macrophages in Paraffin-Embedded Tissue Samples. Am J Clin Pathol. 2004;122:794–801. https://doi.org/10.1309/QHD6YFN81KQXUUH6.
44. Pulford KAF, Sipos A, Cordell JL, et al. Distribution of the CD68 macrophage/myeloid associated antigen. Int Immunol. 1990;2:973–80. https://doi.org/10.1093/intimm/2.10.973.
45. Müller AM, Hermanns MI, Skrzynski C, et al. Expression of the Endothelial Markers PECAM-1, vWf, and CD34 *in Vivo* and *in Vitro*. Exp Mol Pathol. 2002;72:221–9. https://doi.org/10.1006/exmp.2002.2424.
46. Craig-Schapiro R, Li G, Chen K, et al. Single-cell atlas of human pancreatic islet and acinar endothelial cells in health and diabetes. Nat Commun. 2025;16:1338. https://doi.org/10.1038/s41467-024-55415-3.
47. Shi H-Q, Li X, Chen Z, et al. KRT19 is regulated by miR-642a-5p and promotes pancreatic cancer progression through the Wnt/β-catenin pathway. iScience. 2024;27:110782. https://doi.org/10.1016/j.isci.2024.110782.
48. Walsh N, O'Donovan N, Kennedy S, et al. Identification of pancreatic cancer invasion-related proteins by proteomic analysis. Proteome Sci. 2009;7:3. https://doi.org/10.1186/1477-5956-7-3.
49. Xiong F, Guo T, Wang X, et al. Keratin 8 Is an Inflammation-Induced and Prognosis-Related Marker for Pancreatic Adenocarcinoma. Dis Markers. 2022;2022:8159537. https://doi.org/10.1155/2022/8159537.

50. Crawford HC, Scoggins CR, Washington MK, et al. Matrix metalloproteinase-7 is expressed by pancreatic cancer precursors and regulates acinar-to-ductal metaplasia in exocrine pancreas. J Clin Invest. 2002;109:1437–44. https://doi.org/10.1172/JCI15051.
51. Chen K, Wang Q, Li M, et al. Single-cell RNA-seq reveals dynamic change in tumor microenvironment during pancreatic ductal adenocarcinoma malignant progression. eBioMedicine. 2021;66:103315. https://doi.org/10.1016/j.ebiom.2021.103315.
52. Jin S, Plikus MV, Nie Q. CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics. Nat Protoc. 2025;20:180–219. https://doi.org/10.1038/s41596-024-01045-4.
53. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17:261–72. https://doi.org/10.1038/s41592-019-0686-2.
54. Therneau TM, Elizabeth A, Cynthia C. until (2026) survival: Survival Analysis.
55. Wickham H. ggplot2. Cham: Springer International Publishing; 2016.
56. Tao M, Liu W, Chen J, et al. Transcriptome Landscape of Cancer-Associated Fibroblasts in Human PDAC. Adv Sci. 2025;12:2415196. https://doi.org/10.1002/advs.202415196.
57. Öhlund D, Handly-Santana A, Biffi G, et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. J Exp Med. 2017;214:579–96. https://doi.org/10.1084/jem.20162024.
58. Chen Q, Chen Z, Zhang J, et al. Dual and triple gene combinations of KRT5, KRT17, and S100A2 identify basal-like subtype of pancreatic ductal adenocarcinoma and correlate with survival outcome. FASEB J. 2024;38:e23867. https://doi.org/10.1096/fj.202302484RR.
59. Dobersch S, Yamamoto N, Schutter A, et al. HMGA2 and protein leucine methylation drive pancreatic cancer lineage plasticity. Nat Commun. 2025;16:4866. https://doi.org/10.1038/s41467-025-60129-1.
60. Xie Y, Luo X, Li H, et al. autoRPA: A web server for constructing cancer staging models by recursive partitioning analysis. Comput Struct Biotechnol J. 2020;18:3361–7. https://doi.org/10.1016/j.csbj.2020.10.038.
61. Zhu X, Li Y, Liu H, Xiao Z. FN1 from cancer-associated fibroblasts orchestrates pancreatic cancer metastasis via integrin-PI3K/AKT signaling. Front Oncol. 2025;15:1595523. https://doi.org/10.3389/fonc.2025.1595523.
62. Heger U, Martens A, Schillings L, et al. Myofibroblastic CAF Density, Not Activated Stroma Index, Indicates Prognosis after Neoadjuvant Therapy of Pancreatic Carcinoma. Cancers. 2022;14:3881. https://doi.org/10.3390/cancers14163881.
63. Ge W, Yue M, Wang Y, et al. A Novel Molecular Signature of Cancer-Associated Fibroblasts Predicts Prognosis and Immunotherapy Response in Pancreatic Cancer. Int J Mol Sci. 2023;24:156. https://doi.org/10.3390/ijms24010156.
64. Lansbergen MF, Dings MPG, Manoukian P, et al. Transcriptome-based classification to predict FOLFIRINOX response in a real-world metastatic pancreatic cancer cohort. Translational Res. 2024;273:137–47. https://doi.org/10.1016/j.trsl.2024.08.002.
65. Hussain Z, Bertran T, Finetti P, et al. Macrophages reprogramming driven by cancer-associated fibroblasts under FOLFIRINOX treatment correlates with shorter survival in pancreatic cancer. Cell Commun Signal. 2024;22:1. https://doi.org/10.1186/s12964-023-01388-7.
66. Varadhachary GR, Wolff RA, Crane CH, et al. Preoperative gemcitabine and cisplatin followed by gemcitabine-based chemoradiation for resectable adenocarcinoma of the pancreatic head. J Clin Oncol. 2008;26:3487–95. https://doi.org/10.1200/JCO.2007.15.8642.
67. Evans DB, Varadhachary GR, Crane CH, et al. Preoperative gemcitabine-based chemoradiation for patients with resectable adenocarcinoma of the pancreatic head. J Clin Oncol. 2008;26:3496–502. https://doi.org/10.1200/JCO.2007.15.8634.
68. Tempero MA, Malafa MP, Al-Hawary M, et al. Pancreatic Adenocarcinoma, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2021;19:439–57. https://doi.org/10.6004/jnccn.2021.0017.
69. Wainberg ZA, Bekaii-Saab T, Boland PM, et al. First-line liposomal irinotecan with oxaliplatin, 5-fluorouracil and leucovorin (NALIRIFOX) in pancreatic ductal adenocarcinoma: A phase I/II study. Eur J Cancer. 2021;151:14–24. https://doi.org/10.1016/j.ejca.2021.03.028.
70. Von Hoff DD, Ervin T, Arena FP, et al. Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. N Engl J Med. 2013;369:1691–703. https://doi.org/10.1056/NEJMoa1304369.
71. Conroy T, Hammel P, Hebbar M, et al. FOLFIRINOX or Gemcitabine as Adjuvant Therapy for Pancreatic Cancer. N Engl J Med. 2018;379:2395–406. https://doi.org/10.1056/NEJMoa1809775.
72. Datta J, Dai X, Bianchi A, et al. Combined MEK and STAT3 inhibition uncovers stromal plasticity by enriching for cancer-associated fibroblasts with mesenchymal stem cell-like features to overcome immunotherapy resistance in pancreatic cancer. Gastroenterology. 2022;163:1593–612. https://doi.org/10.1053/j.gastro.2022.07.076.
73. Wu C, Von Schalscha T, Sansanwal D, et al. Targeting Pancreatic Cancer Cell Stemness by Blocking Fibronectin-Binding Integrins on Cancer-Associated Fibroblasts. Cancer Res Commun. 2025;5:195–208. https://doi.org/10.1158/2767-9764.CRC-24-0491.
74. Griggs LA, Hassan NT, Malik RS, et al. Fibronectin Fibrils Regulate TGF-β1-induced Epithelial-Mesenchymal Transition. Matrix Biol. 2017;60–61:157–75. https://doi.org/10.1016/j.matbio.2017.01.001.
75. Dallas SL, Sivakumar P, Jones CJP, et al. Fibronectin regulates latent transforming growth factor-beta (TGF beta) by controlling matrix assembly of latent TGF beta-binding protein-1. J Biol Chem. 2005;280:18871–80. https://doi.org/10.1074/jbc.M410762200.
76. Rhim AD, Oberstein PE, Thomas DH, et al. Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. Cancer Cell. 2014;25:735–47. https://doi.org/10.1016/j.ccr.2014.04.021.
77. Özdemir BC, Pentcheva-Hoang T, Carstens JL, et al. Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. Cancer Cell. 2014;25:719–34. https://doi.org/10.1016/j.ccr.2014.04.005.
78. Lee JJ, Perera RM, Wang H, et al. Stromal response to Hedgehog signaling restrains pancreatic cancer progression. Proc Natl Acad Sci U S A. 2014;111:E3091–3100. https://doi.org/10.1073/pnas.1411679111.
79. Olive KP, Jacobetz MA, Davidson CJ, et al. Inhibition of Hedgehog Signaling Enhances Delivery of Chemotherapy in a Mouse Model of Pancreatic Cancer. Science. 2009;324:1457–61. https://doi.org/10.1126/science.1171362.
80. Krishnamurty AT, Shyer JA, Thai M, et al. LRRC15 + myofibroblasts dictate the stromal setpoint to suppress tumour immunity. Nature. 2022;611:148–54. https://doi.org/10.1038/s41586-022-05272-1.

81. Mucciolo G, Araos Henríquez J, Jihad M, et al. EGFR-activated myofibroblasts promote metastasis of pancreatic cancer. Cancer Cell. 2024;42:101–e11811. https://doi.org/10.1016/j.ccell.2023.12.002.
82. Avery D, Govindaraju P, Jacob M, et al. Extracellular Matrix Directs Phenotypic Heterogeneity of Activated Fibroblasts. Matrix Biol. 2018;67:90–106. https://doi.org/10.1016/j.matbio.2017.12.003.
83. Kazakova AN, Lukina MM, Anufrieva KS, et al. Exploring the diversity of cancer-associated fibroblasts: insights into mechanisms of drug resistance. Front Cell Dev Biol. 2024;12. https://doi.org/10.3389/fcell.2024.1403122.
84. Neuzillet C, Nicolle R, Raffenne J, et al. Periostin- and podoplanin-positive cancer-associated fibroblast subtypes cooperate to shape the inflamed tumor microenvironment in aggressive pancreatic adenocarcinoma. J Pathol. 2022;258:408–25. https://doi.org/10.1002/path.6011.
85. Wang R, Qin G-H, Jiang Y, et al. Integrated multiomics analysis identifies PHLDA1 + fibroblasts as prognostic biomarkers and mediators of biological functions in pancreatic cancer. Front Immunol. 2025;16:1592416. https://doi.org/10.3389/fimmu.2025.1592416.
86. Sun X, Cai W, Li H, et al. Endothelial-like cancer-associated fibroblasts facilitate pancreatic cancer metastasis via vasculogenic mimicry and paracrine signalling. Gut. 2025;74:1437–51. https://doi.org/10.1136/gutjnl-2024-333638.
87. Feig C, Jones JO, Kraman M, et al. Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti–PD-L1 immunotherapy in pancreatic cancer. Proc Natl Acad Sci U S A. 2013;110:20212–7. https://doi.org/10.1073/pnas.1320318110.
88. Verbeke C, Webster F, Brosens L, et al. Dataset for the reporting of carcinoma of the exocrine pancreas: recommendations from the International Collaboration on Cancer Reporting (ICCR). Histopathology. 2021;79:902–12. https://doi.org/10.1111/his.14540.
89. Duan K, Jang G-H, Grant RC, et al. The value of GATA6 immunohistochemistry and computer-assisted diagnosis to predict clinical outcome in advanced pancreatic cancer. Sci Rep. 2021;11:14951. https://doi.org/10.1038/s41598-021-94544-3.
90. Shibayama T, Hayashi A, Toki M, et al. Combination immunohistochemistry for CK5/6, p63, GATA6, and HNF4a predicts clinical outcome in treatment-naïve pancreatic ductal adenocarcinoma. Sci Rep. 2024;14:15598. https://doi.org/10.1038/s41598-024-65900-w.
91. van Eijck CWF, Real FX, Malats N, et al. GATA6 identifies an immune-enriched phenotype linked to favorable outcomes in patients with pancreatic cancer undergoing upfront surgery. Cell Rep Med. 2024;5:101557. https://doi.org/10.1016/j.xcrm.2024.101557.
92. Mc Laughlin RA, Salvatore L, Barone D, et al. GATA6 expression in advanced pancreatic ductal adenocarcinoma (PDAC) as a prognostic and predictive biomarker in the Canadian COMPASS trial and the Comprehensive Cancer Center, Fondazione Policlinico Universitario Agostino Gemelli IRCCS. JCO. 2024;42:4150–4150. https://doi.org/10.1200/JCO.2024.42.16_suppl.4150.
93. Grant RC, Duan K, Jackson R, et al. GATA6 immunohistochemistry and prognosis after surgical resection of pancreatic adenocarcinoma: results from the ESPAC-4 trial. ESMO Gastrointest Oncol. 2025;8:100138. https://doi.org/10.1016/j.esmogo.2025.100138.

Publisher's Note