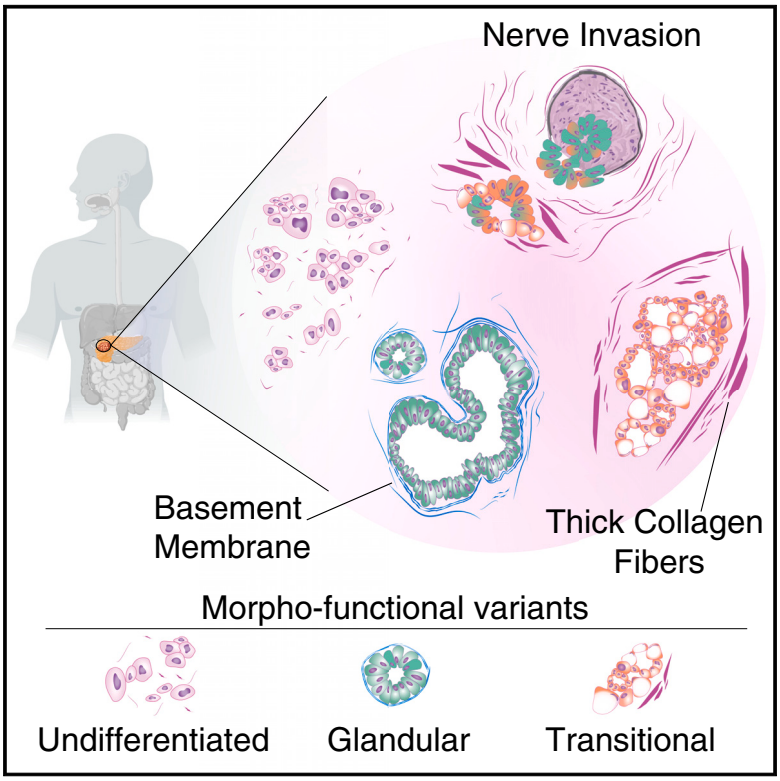
Cancer Cell

Article

Mapping functional to morphological variation reveals the basis of regional extracellular matrix subversion and nerve invasion in pancreatic cancer

Graphical abstract



Highlights

- Integrated transcriptional and morphological maps reveal three coexisting PDAC variants
- Combining transcriptional and histological features has high prognostic impact
- A variant with partial myofibroblast-like properties subverts ECM organization
- Local ECM subversion impacts tissue mechanics, PDAC cell states and nerve invasion

Authors

Pierluigi Di Chiaro, Lucia Nacci, Fabiana Arco, ..., Iros Barozzi, Giuseppe R. Diaferia, Gioacchino Natoli

Correspondence

pierluigi.dichiaro@ieo.it (P.D.C.), giuseppe.diaferia@ieo.it (G.R.D.), gioacchino.natoli@ieo.it (G.N.)

In brief

Di Chiaro et al. assign distinct transcriptional and functional programs to intratumor morphological variants that coexist in all pancreatic ductal adenocarcinomas (PDAC), providing a valuable resource for translational studies. These variants impose distinct local subversions of the extracellular matrix that in turn affect tumor invasive properties and specifically neural infiltration.

Di Chiaro et al., 2024, Cancer Cell *42*, 662–681
April 8, 2024
https://doi.org/10.1016/j.ccell.2024.02.017

CellPress

CellPress

Cancer Cell

Article

Mapping functional to morphological variation reveals the basis of regional extracellular matrix subversion and nerve invasion in pancreatic cancer

Pierluigi Di Chiaro,[1,9,*] Lucia Nacci,[1,9] Fabiana Arco,[1] Stefania Brandini,[1] Sara Polletti,[1] Andrea Palamidessi,[2] Benedetta Donati,[3] Chiara Soriani,[1] Francesco Gualdrini,[1] Gianmaria Frigè,[1] Luca Mazzarella,[1,4] Alessia Ciarrocchi,[3] Alessandro Zerbi,[5,6] Paola Spaggiari,[5] Giorgio Scita,[2,7] Simona Rodighiero,[1] Iros Barozzi,[8] Giuseppe R. Diaferia,[1,10,*] and Gioacchino Natoli[1,10,11,*]

[1]Department of Experimental Oncology, IEO, European Institute of Oncology IRCCS, Milano, Italy
[2]IFOM, The FIRC Institute for Molecular Oncology, Via Adamello 16, 20139 Milan, Italy
[3]Laboratory of Translational Research, Azienda Unità Sanitaria Locale-IRCCS di Reggio Emilia, Reggio Emilia, Italy
[4]Division of Gastrointestinal Medical Oncology and Neuroendocrine Tumors, IEO, European Institute of Oncology, IRCCS, Milano, Italy
[5]IRCCS Humanitas Research Hospital, Rozzano, Milano, Italy
[6]Humanitas University, Pieve Emanuele - Milano, Italy
[7]Department of Oncology and Haemato-Oncology, University of Milan, Milano, Italy
[8]Center for Cancer Research, Medical University of Vienna, Vienna, Austria
[9]These authors contributed equally
[10]These authors contributed equally
[11]Lead contact
*Correspondence: pierluigi.dichiaro@ieo.it (P.D.C.), giuseppe.diaferia@ieo.it (G.R.D.), gioacchino.natoli@ieo.it (G.N.)
https://doi.org/10.1016/j.ccell.2024.02.017

SUMMARY

Intratumor morphological heterogeneity of pancreatic ductal adenocarcinoma (PDAC) predicts clinical outcomes but is only partially understood at the molecular level. To elucidate the gene expression programs underpinning intratumor morphological variation in PDAC, we investigated and deconvoluted at single cell level the molecular profiles of histologically distinct clusters of PDAC cells. We identified three major morphological and functional variants that co-exist in varying proportions in all PDACs, display limited genetic diversity, and are associated with a distinct organization of the extracellular matrix: a glandular variant with classical ductal features; a transitional variant displaying abortive ductal structures and mixed endodermal and myofibroblast-like gene expression; and a poorly differentiated variant lacking ductal features and basement membrane, and showing neuronal lineage priming. *Ex vivo* and *in vitro* evidence supports the occurrence of dynamic transitions among these variants in part influenced by extracellular matrix composition and stiffness and associated with local, specifically neural, invasion.

INTRODUCTION

Within a dense fibrotic stroma accounting for up to 90% of the tumor mass, pancreatic ductal adenocarcinoma (PDAC) contains variable proportions of tumor cells of different size, shape, and subcellular features organized in growth patterns ranging from pseudo-glandular structures to compact nests.[1] While morphological heterogeneity is not part of standard pathological reporting, it has prognostic impact,[1] indicating links to the underlying tumor biology.

At the molecular level, PDAC intratumor heterogeneity is mirrored by the variability of gene expression programs revealed by single cell RNA-sequencing (scRNA-seq).[2–4] Notably, high PDAC cell transcriptional heterogeneity correlates with poor overall survival,[5] suggesting that the coexistence of heterogeneous tumor cells may enable the rapid adaptation to therapy or/and the selection of aggressive, treatment-resistant tumor cells. However, scRNA-seq does not clarify the relationship between gene expression programs and morphological patterns, thus limiting our understanding of the molecular circuitry associated with, and underlying morphological heterogeneity.

Different PDAC molecular classification systems converge on the existence of a *classical* (or *pancreatic progenitor*) signature characterized by the preferential expression of endodermal or epithelial differentiation genes, and a series of partially overlapping, yet distinct signatures associated with loss of endodermal lineage fidelity, and referred to as *quasi-mesenchymal*, *basal* or *squamous*, with the different names reflecting the most over-represented gene expression program identified in individual studies.[6–9] While the *quasi*-mesenchymal signature includes genes involved in epithelial-to-mesenchymal transition,[10] the basal and squamous signatures comprise genes encoding

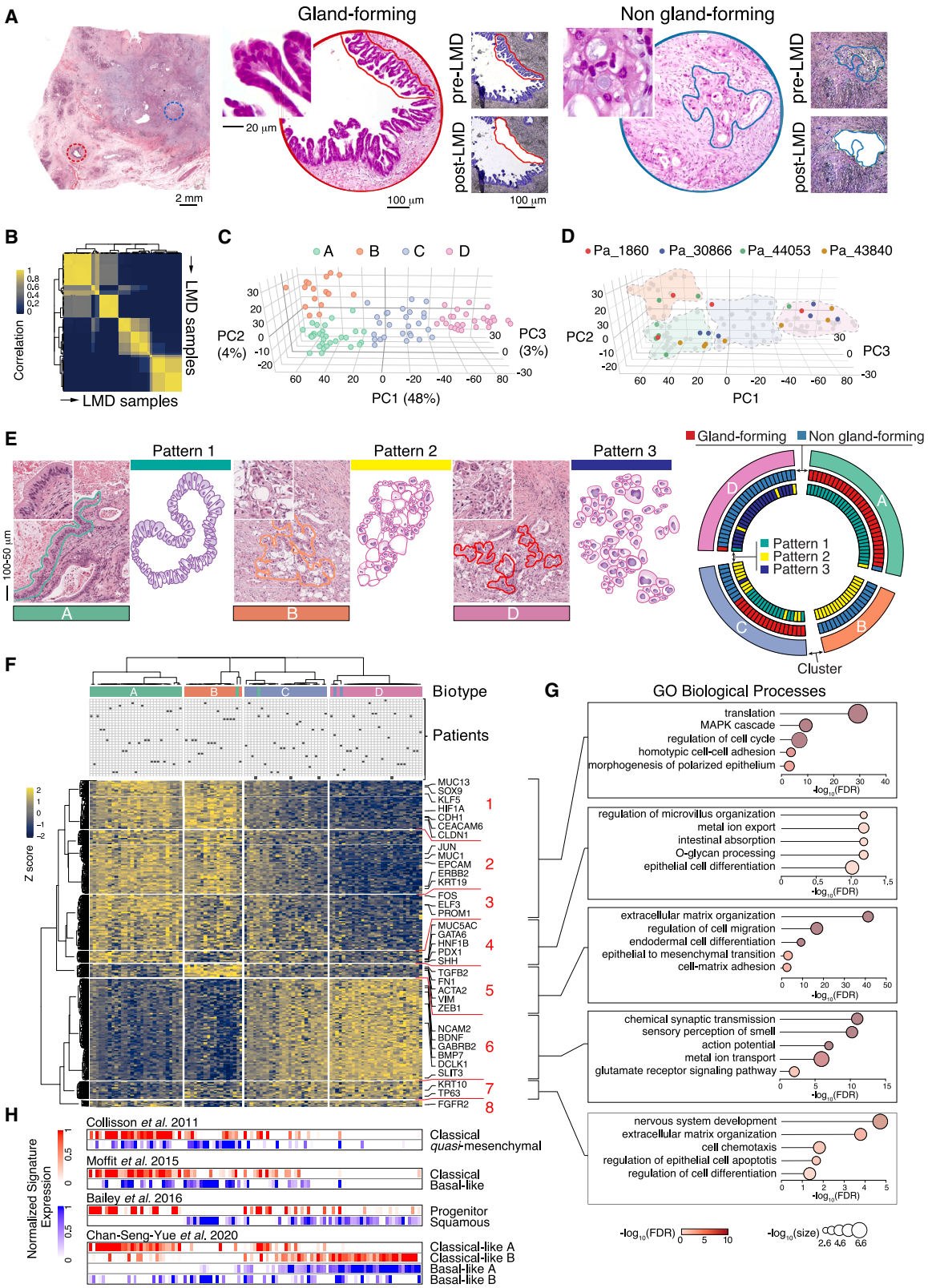


Figure 1. Three distinct morpho-biotypes coexisting in PDACs

(A) Hematoxylin and eosin (H&E) images of a gland-forming (red circle) and a non-gland-forming area (blue circle) in the same PDAC. The same regions were matched on a consecutive section prepared for laser microdissection and imaged pre- (upper panels) and post-cut (lower panels).

(legend continued on next page)

regulators of basal cells, the stem cells of glandular and stratified epithelia, and genes expressed by their differentiating progeny. Bulk tumor analyses, however, average the transcriptomes from different coexisting tumor cells, possibly explaining the limited clinical impact of such classification schemes,[7] which are inferior to morphological classifications at predicting prognosis.[11]

When considering mutational profiles, PDACs display genetic alterations in four main genes (*KRAS*, *TP53*, *CDKN2A*, and *SMAD4*) in addition to less frequent mutations or copy number alterations in genes encoding chromatin regulators and transcription factors.[12] However, the contribution of genetic diversity to heterogeneity of PDAC cells, albeit detectable,[4] is lower than that of tumor microenvironment (TME)-derived instructive signals.[3,13,14] Hence, heterogeneous PDAC cell states may be primarily influenced by local tissue signals, rather than being hardwired in distinct cancer cell genomes.

In order to bridge the knowledge gap between intratumor morphological variation and the molecular bases of PDAC heterogeneity, we aimed to map transcriptional and functional variation to the distinct morphological patterns coexisting in individual PDACs. We identified three main morphological and functional variants that coexist in various proportions in all PDACs. Hence, PDACs are ensembles of morphologically discrete, spatially organized groups of tumors cells with distinct gene expression programs and functional properties, with direct implications for spatial tumor architecture, local organization of the extracellular matrix and perineural invasion.

RESULTS

Identification of coexisting PDAC morpho-biotypes

We used laser micro-dissection (LMD) to isolate from treatment-naive PDACs (Table S1) multiple morphologically distinguishable areas of ~200–500 cells each, with either gland-forming or non-gland-forming morphology[11] (Figure 1A) in conditions that maximized dissection precision.

The microdissected areas (n = 102; 1–7 per patient) were subjected to RNA-seq, resulting in an average of 9,513 expressed genes per sample (Table S2). We applied consensus clustering[15] on the top variable genes followed by principal component analysis (PCA) to obtain an unsupervised classification of tumor areas (Table S3). We identified four clusters with high intra-group correlations (Figure 1B) that were well separated by the first three principal components (Figure 1C). In spite of the high inter-patient variability, tumor areas with similar gene expression profiles could be detected across patients (Figure 1D).

Retrospective examination of histological data allowed us to assign these four gene expression clusters to morphologically different areas that we categorized into three groups that broadly matched the morphological patterns previously reported to describe PDAC heterogeneity and to possess prognostic value when combined into a grading score[16] (Figures 1E and S1A).

Pattern 1 areas, associated with cluster A, consisted of ductal-like structures encompassing tubular or elongated glands as well as papillary and finger-like patterns. Tumor cells in these areas ranged from columnar cells resembling mucin-producing foveolar cells of the stomach, to cuboidal cells similar to those of the pancreato-biliary ducts, indicating loss of lineage fidelity.

Pattern 2 areas, associated with cluster B, included ill-defined glands with incomplete borders, irregular formation of rudimentary lumina (*cribriform* architecture) and vacuolated tumor cells in which cytoplasmic mucin pushed the nucleus to the cell periphery (*signet-ring cells*).

Pattern 3, associated with cluster D, consisted of *non-glandular agglomerates* of cancer cells arranged into sheets or nests, including: (1) packets of round cells with variable compaction and minimal keratinization, (2) large pleomorphic and anaplastic cells with prominent nuclear atypia, hyperchromatic nuclei, and eosinophilic cytoplasm, and (3) nests of cells with solid or glassy-looking cytoplasm reminiscent of squamous cells with varying degrees of keratinization.

The gene expression cluster C showed a variegated association with both glandular and non-glandular areas (discussed in the following).

Hence, different gene expression programs could be assigned to morphological variants coexisting in the same tumor. PDAC areas with both distinct morphology and transcriptional programs are henceforth indicated as *morpho-biotypes* (or biotypes), a term conveying the dual notion of morphological and functional diversity.

Differentially expressed genes (DEGs) were then clustered into eight modules, allowing us to highlight transcriptional differences and similarities among the morpho-biotypes (Figure 1F; Table S3). Immunofluorescence staining of selected gene products is shown in Figure S1B.

Morpho-biotypes A and B shared high-level expression of genes in modules #1–3, which were enriched with functional terms indicating high protein synthesis (ribosome biogenesis and translation), as well as terms pertaining to cell adhesion and epithelial polarization (Figure 1G; Table S4). Genes in these modules encoded: (1) mucins (e.g., *MUC1*); (2) CEACAMs (carcinoembryonic antigen-related cell adhesion molecules); (3) epithelial adhesion molecules (*CDH1*, E-cadherin; *EPCAM*,

(B) Correlation matrix obtained by consensus clustering of the top 25% variable genes from the LMD-seq dataset (n = 102).
(C and D) 3D-PCA plots showing the distribution of the samples in the first three principal components. (C) Samples were colored by the four identified gene expression clusters (A–D) shown in (B). (D) Multiple samples from four different patients were colored to show their distribution in the four highlighted clusters.
(E) Left: association of the microdissected areas to the morphological patterns derived from Adsay et al.[16] (schematic drawing depicting typical features). Right: circos plot of the identified sample clusters (A–D), the morphological components (gland- or non-gland-forming) derived from Kalimuthu et al.[11] and the morphological patterns (patterns 1–3) from Adsay et al.[16]
(F) Heatmap showing the four sample groups derived from differentially expressed genes clustered by graph-based clustering approach. Samples were re-clustered by consensus clustering. Color-coded morpho-biotype annotations are shown for each sample. Gene expression modules (#1–8) are indicated on the right.
(G) Representative functional classes enriched in the indicated gene cluster.
(H) Gene expression scores in each sample using the indicated bulk gene signatures. Also see Figures S1–S3, Tables S1–S4.

epithelial cell adhesion molecule); (4) proteins involved in apical cell membrane organization (e.g., *PROM1*, prominin-1/CD133); and (5) transcription factors associated with endodermal and/or pancreatic differentiation and highly expressed in well-differentiated PDACs (SOX9, KLF5, and ELF3).[17,18]

However, morpho-biotype A differed from B because of the expression of a group of genes (module #4) associated with endodermal differentiation and functions, such as those encoding pancreatic lineage-determining transcription factors (GATA6, HNF1B, and PDX1) and Sonic Hedgehog (SHH), the main inducer of the desmoplastic reaction in PDAC.[19,20] Conversely, morpho-biotype B selectively expressed genes (module #5) encoding proteins involved in extracellular matrix synthesis, cell migration and epithelial-to-mesenchymal transition (EMT) such as *TGFB2* (transforming growth factor β2), *VIM* (vimentin), *FN1* (fibronectin 1) and the EMT driver ZEB1.[21,22]

Overall, morpho-biotype A (henceforth *glandular* morpho-biotype) was characterized by both a ductal architecture and the expression of pancreato-biliary and endodermal lineage genes. Morpho-biotype B, while sharing with the glandular morpho-biotype the expression of many such genes, showed a distinctive cribriform pattern coupled with the downregulation of pancreatic lineage-determining transcription factors. Instead, genes involved in EMT and encoding extracellular matrix components were strongly upregulated in this morpho-biotype. We named this morpho-biotype *transitional* to emphasize the induction of mesenchymal gene expression within cells retaining the expression of glandular genes.

At the opposite end of the spectrum, morpho-biotype D displayed the extensive downregulation of genes defining the glandular and transitional morpho-biotypes (modules #1–5), and instead the upregulation of three large gene modules (#6–8) encoding:

(1) *Master regulators of stem/progenitor cells,* including the transcription factor TP63, which is overexpressed in squamous cell carcinomas[23] and in PDAC[24]; and the protein kinase DCLK1, which marks stem cells in PDAC[25] and in normal pancreas.[26] Keratins of squamous epithelia (e.g., *KRT10*) were selectively expressed in this morpho-biotype;

(2) *Neuronal activity-related proteins* involved in *synaptic transmission* (GABA receptors, glutamate receptor components, and opioid receptors), *axonal migration* (SLIT-ROBO signaling components and neural adhesion molecules), *neurotrophic activity* (neuregulin1 and NTRK family members), *voltage-gated ion channels* and *chemosensing* (olfactory receptor genes).

Noticeably, a large fraction of genes (63/153) of the *Smell perception* program was enriched in module #6 (Figure 1G and S2A-C), suggesting their coordinated activation. However, neuronal genes in this morpho-biotype were expressed at very low levels, thus excluding a full deployment of a neuronal program. Moreover, the co-expression of squamous and neuronal markers observed by immunofluorescence (Figure S2D), along with the morphological resemblance of these cells to poorly differentiated tumor areas,[1] motivated us to name this morpho-biotype *undifferentiated*. This choice was coherent with the assumption that neuronal gene expression represents an abortive lineage priming event in undifferentiated cancer cells incapable of endodermal differentiation.

Some tumor areas displayed a gene expression profile intermediate between the glandular and transitional morpho-biotypes on the one hand, and the undifferentiated one on the other (group C, Figure 1E). This *hybrid* gene expression profile arised either from the co-expression in individual cells of markers of different biotypes, as reported before,[2,3,27] or from the presence in a single area of cells expressing markers of alternative morpho-biotypes, with either sharp or gradual transitions between them (Figure S1C). This *hybrid* profile may not represent a distinct entity but instead the consequence of extensive cellular plasticity.[28]

We next analyzed the relationship between the morpho-biotypes and the PDAC molecular subtypes identified by bulk tumor analyses.[6–9] *Classical* (or *progenitor-like*) signatures were enriched in the *glandular* morpho-biotype and in the hybrid group (Figure 1H). The *classical B* signature,[4] however, was enriched in the undifferentiated morpho-biotype (Figure 1H) because of many genes not related to endoderm or pancreas, such as a group of 36 brain-enriched genes (including *GRIN2B*, *RPGRIP1*, and *HECW2*).

Mesenchymal and *basal-like* signatures were enriched in the *transitional* morpho-biotype. The *undifferentiated* morpho-biotype was not significantly enriched for any of the signatures reported in Collisson et al. and Moffit et al.,[8,9] which include only 62 and 50 genes, respectively. Instead, it was mildly enriched for an extensive *squamous* signature reported by Bailey et al.[6] However, this signature could not discriminate between transitional and undifferentiated morpho-biotypes (Figure 1H).

The existence of distinct morpho-biotypes was validated with orthogonal spatial transcriptomics techniques. First, using a predesigned gene panel for the GeoMx platform, we identified three clusters that were each differentially enriched for morpho-biotype-specific signatures, and an additional one corresponding to the hybrid phenotype (Figures S3A–S3D). Second, multiplexed single-molecule fluorescence in situ hybridization (FISH) with probes corresponding to 32 genes differentially expressed in the biotypes was consistent with the LMD-seq data, confirming the existence of the three morpho-biotypes (Figures S3E–S3H).

Prognostic impact of PDAC morpho-biotypes

We next used a machine learning approach to identify a set of genes needed for the robust identification of morpho-biotypes. A Random Forest (RF) Classifier coupled with a Recursive Feature Elimination (RFE) (Figure 2A) yielded a set of 457 predictor genes (Figure 2B) that was enriched for genes differentially expressed among the morpho-biotypes (Figures 2C and S4A). Unsupervised clustering using these genes reproduced initial clustering based on top variable genes, confirming the gene signature's predictive power (Figure 2D). The resulting model demonstrated 91% accuracy in classifying a testing set of 51 samples, highlighting its robustness (Figures 2D; Table S5).

The classifier was then applied to a different clinical dataset[4] to infer morpho-biotype composition of individual PDACs based on bulk tumor RNA-seq data (Figure 2E). This analysis indicated on the one hand the presence of a dominant morpho-biotype

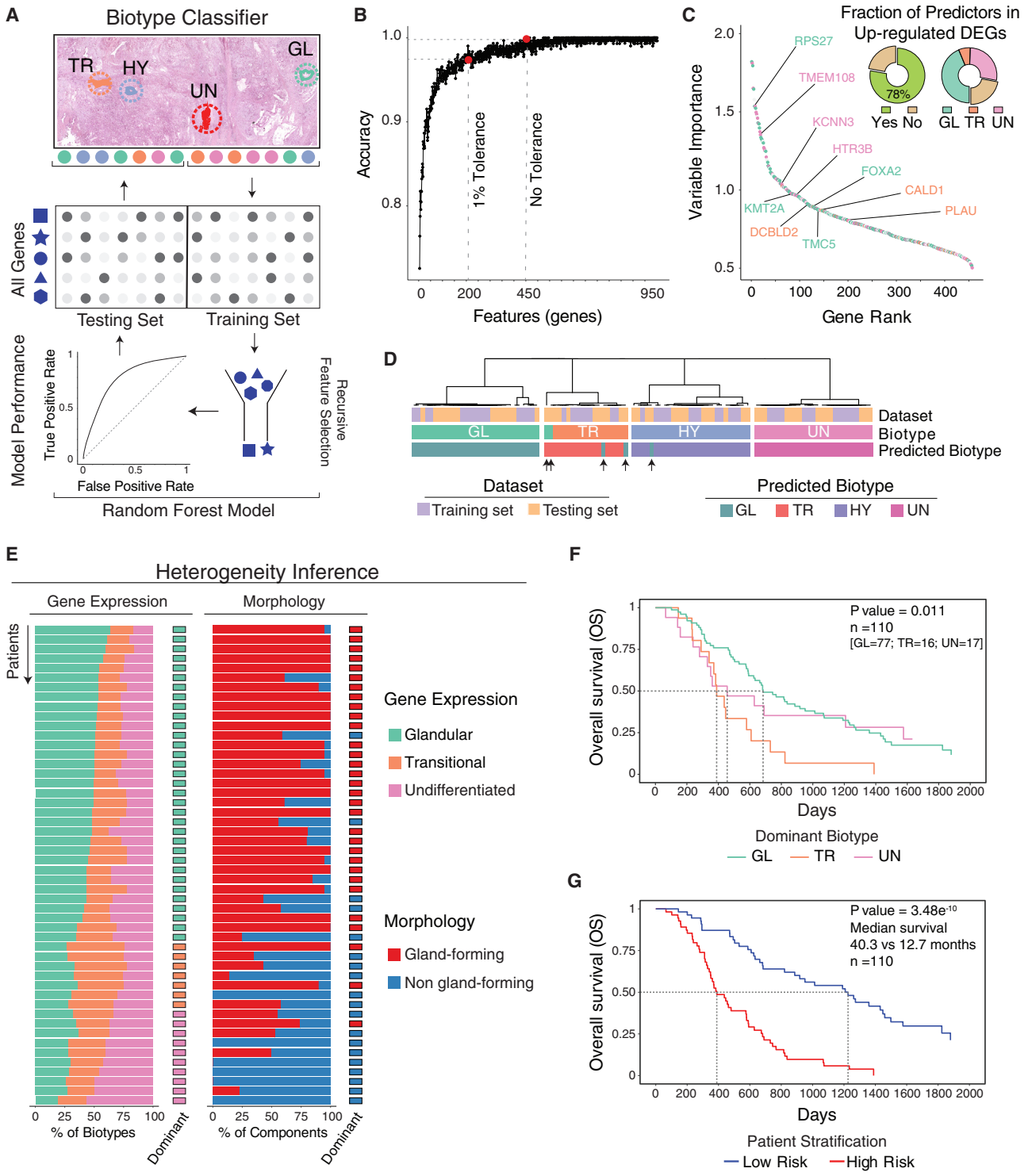


Figure 2. A Random Forest-based classifier discriminating PDAC morpho-biotypes

(A) Random Forest workflow for the prediction of PDAC morpho-biotypes from LMD data.

(B) RFE analysis showing the optimal number of features (red dots and vertical black lines) at different percentage of tolerance in the model accuracy of the training set.

(C) Variable importance plot of the 457 genes selected by RFE and ranked from the most to the least important in the model. Each gene (dot) is colored according to the biotype in which it was differentially expressed: glandular (green); transitional (orange); undifferentiated (pink); N/A (gray). Three selected genes for each biotype are highlighted. The donut charts show the fraction of the 457 genes matching all the upregulated DEGs (left) or the upregulated DEGs stratified into the three biotypes (right).

(legend continued on next page)

and on the other the possibility to detect the co-occurrence of the others in different proportions in individual tumors. The inferred morpho-biotype composition was consistent with the reported morphological classification of the same cohort.[11] Overall, the model identified the dominant morpho-biotype and determined the composition of individual lesions, highlighting the co-existence of the three morpho-biotypes, and its correlation with the presence of different proportions of gland- and non-gland-forming areas.

We next determined the prognostic impact of the presence of different morpho-biotypes in individual PDACs. First, using the aforementioned classifier, we assigned each tumor of the same cohort to one of three different classes based on the dominant morpho-biotype, and then determined the prognostic correlates of this assignment (Figure 2F). In line with the current literature,[7] tumors dominated by the transitional morpho-biotype showed significantly worse prognosis (HR = 2.15, 95% CI = 1.01,4.59; p value < 0.05) than those dominated by a glandular morpho-biotype. Tumors with a dominant undifferentiated morpho-biotype showed a non-significant trend toward worse prognosis than those with a glandular dominance (Figure 2F).

Survival analysis of patients showing different levels of expression of morpho-biotype signatures showed that only the transitional signature and its combination with the undifferentiated one had significant, yet small prognostic impact (Figure S4B). We reasoned that since the morpho-biotypes co-exist in individual PDACs, we may achieve better prognostic stratification by combining the expression of genes characteristic of different morpho-biotypes. We trained an elastic-net regression approach to extract an optimized combination of genes for improved patient outcome prediction starting from the initial 457 genes, eventually obtaining a set of 23 genes. The expression of these genes and their respective model coefficients, which estimate the risk, were then used to calculate a risk score for each patient and to stratify the cohort into high- and low-risk groups with significantly different overall survival (Figure 2G).

As an additional validation, we applied this 23 genes set to two independent cohorts, confirming a significant association of the risk scores with prognosis (Figures S4C and S4D). Finally, a stringent 4-fold cross validation on the initial cohort confirmed the strong prognostic value of the risk groups (Figure S4E).

The prognostic genes derived previously included genes belonging to all three morpho-biotypes but the level of expression of the individual genes of the glandular or undifferentiated morpho-biotypes showed a complex relationship with patient stratification, since some of them were associated with better prognosis when higher expressed and others when expressed at lower levels (Figure S4F). Overall, the prognostic power of combining genes associated with the three morpho-biotypes underscores the relevance of the co-existence of the three morpho-biotypes for clinical assessment.

Finally, the prognostic model and its associated stratification strategy may help identify a high-risk group of patients who may benefit from combinatorial therapies since morpho-biotypes differentially expressed targets of currently approved drugs (Figure S5; Table S6).

Deconvolution of PDAC morpho-biotypes at single cell level

LMD-seq is a population analysis in which non-tumor cells may in principle be co-isolated with cancer cells. Therefore, we aimed to identify morpho-biotypes signatures at single-cell level.

First, we analyzed the enrichment in the morpho-biotypes of functional signatures extracted from single nucleus RNA-seq data[5] (Figures 3A and 3B). The glandular morpho-biotype displayed the expected enrichment of a "*classical-like*" signature. Conversely, in the transitional morpho-biotype, there was a robust enrichment of a *mesenchymal* signature[5] that included genes encoding extracellular matrix proteins (Figure 3B).

The enrichment of the "*basaloid*" signature in the glandular and the transitional morpho-biotypes is explained by the high expression in both morpho-biotypes of a large set of ribosomal genes included in this signature. Similarly, the enrichment of a "*squamoid*" signature in the glandular morpho-biotype was mainly due to the inclusion in this signature of endodermal genes (e.g., *MUC5AC* and *MUC5B*) that are not associated with squamous differentiation.

Notably, the *undifferentiated* morpho-biotype overlapped with *neural-like* and *neuro-endocrine* signatures previously identified in patients who underwent neo-adjuvant therapy.[5] The enrichment of the "*neuroendocrine*" signature was driven by the expression in the undifferentiated morpho-biotype of the neural (rather than the endocrine) genes in this signature.[5]

We next used a scRNA-seq dataset[4] to verify the enrichment of morpho-biotype signatures in tumor cells identified on the basis of inferred copy number variants.[29] The morpho-biotype signatures showed complementary patterns of enrichment in cancer cells from all patients (Figures 3C and S6A). We next assigned single tumor cells from individual patients to the different morpho-biotypes based on the expression of morpho-biotype-specific signatures (Figure 3D). While the composition of each tumor was heterogeneous, cells of undifferentiated morpho-biotype were a smaller fraction than the glandular and the transitional ones. A large fraction of cells in each patient could not be assigned to any morpho-biotypes. Although this observation may relate to technical issues, it is consistent with previous studies reporting single cells that did not express programs corresponding to classical or basal signatures.[3] Importantly, the level of expression of genes of the glandular and transitional morpho-biotypes signatures was consistently higher than that of genes of the undifferentiated morpho-biotype (Figures S6B and S6C). The uncommon occurrence of undifferentiated morpho-biotype cells,

(D) Unsupervised clustering of all 102 LMD samples using consensus clustering. Dataset (training and testing set) and the predicted biotype annotations are shown for each sample. Asterisks indicate the samples in the testing set that were not correctly predicted by the model.
(E) Left: inference of morpho-biotype composition of individual PDACs from bulk RNA-seq data. The dominant biotype for each patient is shown on the right. Right: percentage of gland- and non-gland-forming morphological components in individual PDACs based on Kalimuthu et al.[11] The dominant morphological component for each patient is shown on the right. Only patients who died before the end of the study are shown (n = 50).
(F and G) Overall patient survival stratified by morpho-biotype (F) or risk group (G) (PACA-CA dataset, n = 110). Log rank p values are shown. Dashed lines indicate the median of survival for each group. Also see Figures S4, S5, Tables S5, and S6.

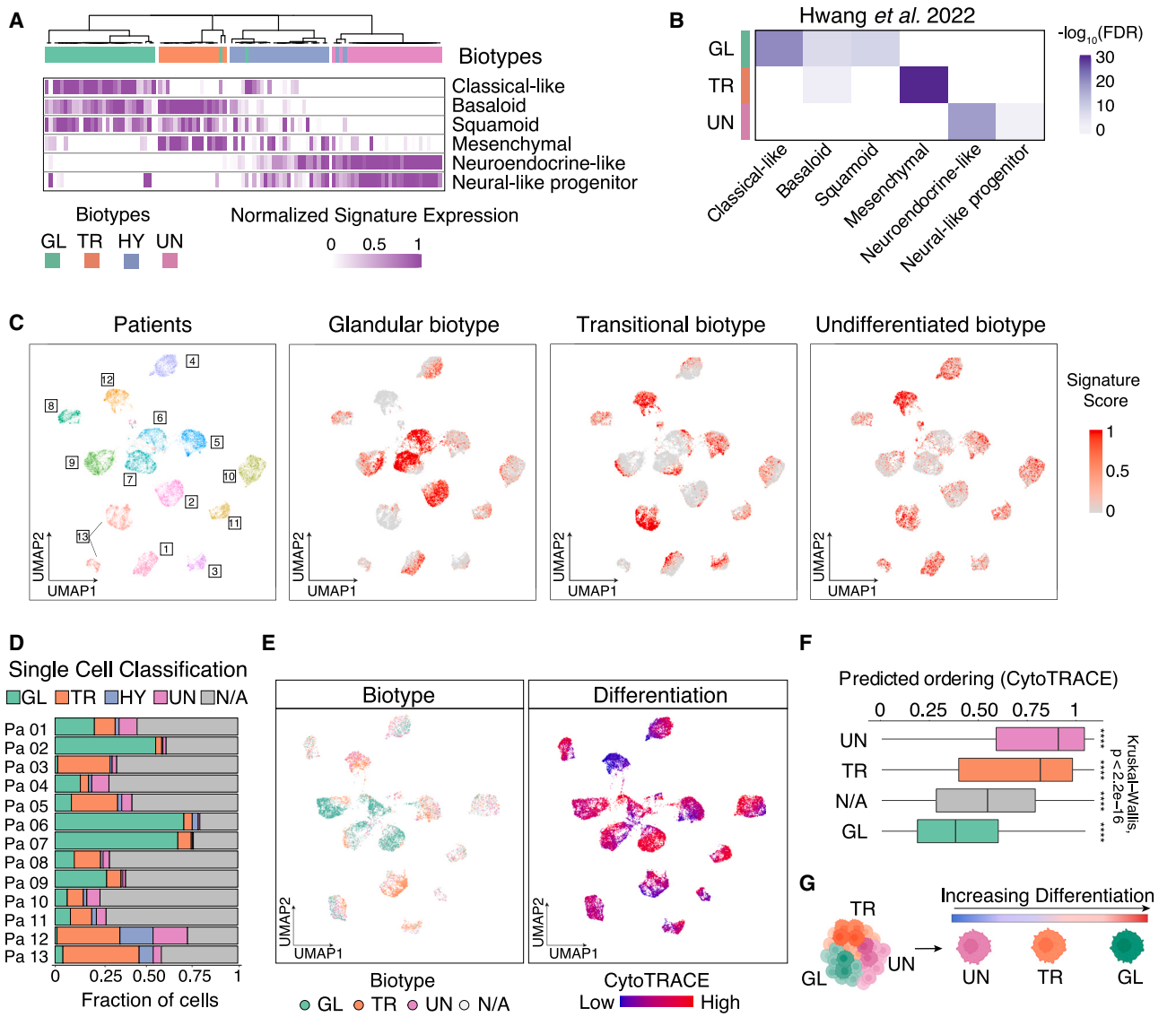


Figure 3. Deconvolution of PDAC morpho-biotypes at single cell level

(A) Gene expression scores were calculated for each sample in the LMD dataset (clustered as in Figure 1E) using the indicated single cell gene signatures.[5]

(B) Enrichment of the indicated tumor single cell transcriptional programs[5] in each morpho-biotype-specific gene signature.

(C) UMAPs of malignant cells from 13 primary PDACs labeled by patient (left) or by gene expression score obtained for each morpho-biotype-specific gene signature (right).

(D) Stacked bar plot showing the fraction of single cells assigned to different morpho-biotypes based on gene expression scores. N/A: not assigned.

(E) UMAP of single malignant cells colored by morpho-biotype (left) or CytoTRACE (right) values (as proxies of differentiation).

(F) CytoTRACE values for morpho-biotype-assigned cells (GL, n = 5498; TR, n = 3757; UN, n = 1286, N/A, n = 12,000). Kruskal-Wallis p value and two-sided Wilcoxon rank-sum tests (group *vs.* rest) are shown.

(G) Schematic representation of differentiation state of PDAC cells of the three morpho-biotypes. See also Figure S6.

coupled with low marker gene expression, may explain why this signature was previously missed in bulk transcriptome data and even in scRNA-seq datasets, in which low-expressed genes are inadequately detected.[30]

We next analyzed the differentiation state of PDAC morpho-biotype cells using CytoTRACE,[31] which uses the number of expressed genes per cell as an indicator of developmental potential, with stem/progenitor cells expressing a high number of genes at low levels, and differentiated cells expressing at high level a comparatively small number of differentiation-specific genes. The glandular and the undifferentiated morpho-biotypes were at the opposite ends of the spectrum, being characterized by the highest and the lowest differentiation state, respectively, with the transitional morpho-biotype showing intermediate properties (Figures 3E and 3F). Therefore, cells of the undifferentiated morpho-biotype resemble progenitor-like cells poised to undergo neuronal lineage priming but retaining the ability to sporadically differentiate toward alternative fates (such as

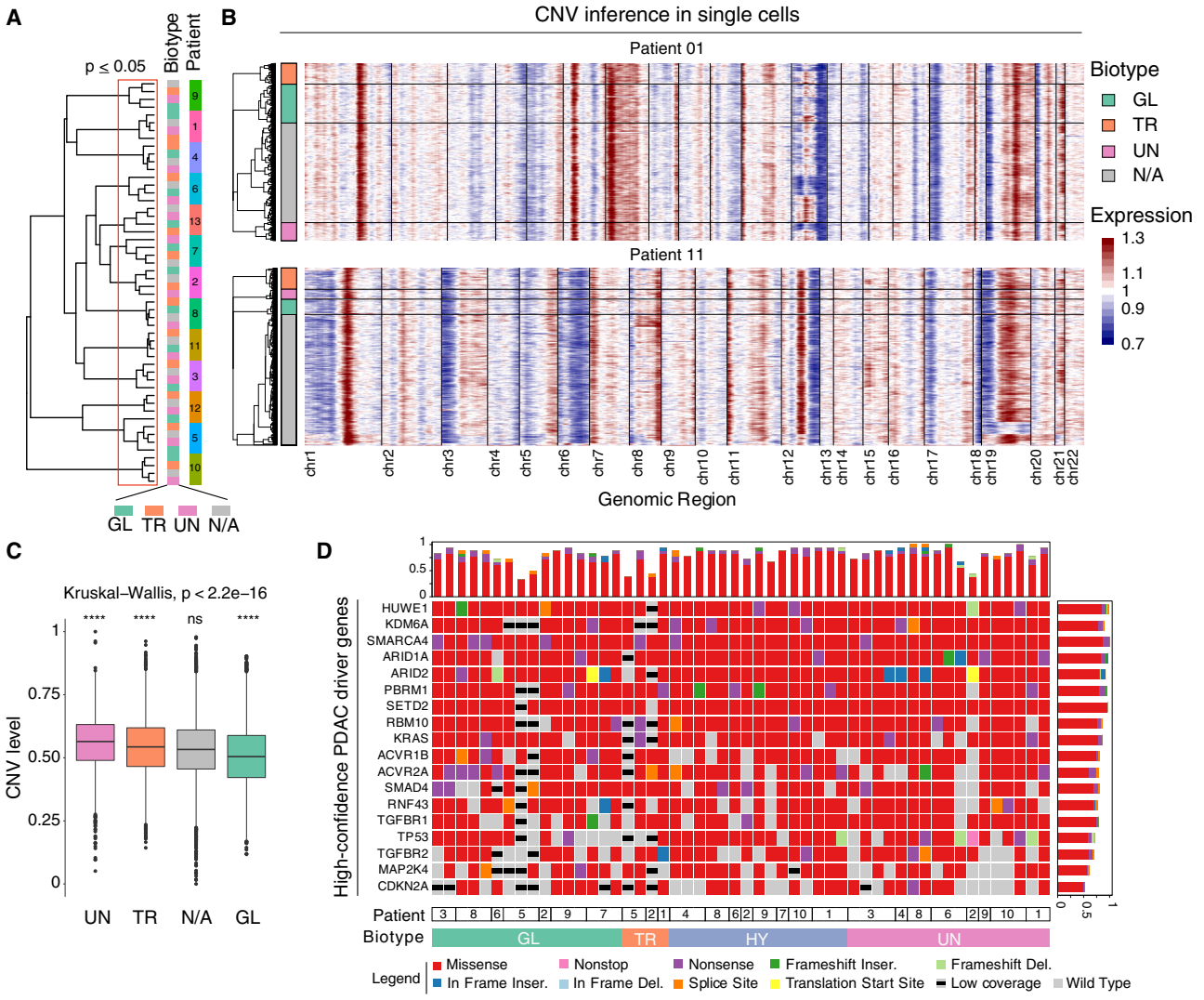


Figure 4. Genetic variation in PDAC morpho-biotypes

(A) Hierarchical clustering of the mean of single-cell CNVs values for the indicated morpho-biotypes and unclassified cells (N/A) in each patient. The red rectangle indicates cluster stability ($p \leq 0.05$).

(B) Heatmaps of inferred copy number variations (CNV) of single malignant cells from two representative patients clustered by morpho-biotype. Inferred amplifications are shown in red while inferred deletions are shown in blue.

(C) Distribution of CNV levels for the glandular (GL, n = 5,498), transitional (TR, n = 3,757), and undifferentiated (UN, n = 1,286) morpho-biotypes and unclassified cells (n = 12,000). Kruskal-Wallis p value and two-sided Wilcoxon rank-sum tests (group *vs.* rest) are shown. Morpho-biotypes were ordered based on decreasing median CNV level.

(D) Oncoprint showing non-synonymous somatic alterations of PDAC driver genes in 52 laser microdissected tumor areas obtained from 10 PDACs and grouped by morpho-biotype. Low coverage annotation indicates a low number of reads on more than half of all mutational sites called for a gene in a certain sample. Bar plots represent the fraction of somatic variants by gene (right) or sample (top). See also Table S7.

metaplastic squamous differentiation) (Figure 3G). Non-classifiable cells showed a differentiation state similar to the glandular ones, indicating that they may have been caught in an incomplete, yet advanced differentiation stage.

Genetic alterations in PDAC morpho-biotypes

Different morpho-biotypes may be driven by microenvironmental signals acting on cancer cells[3,14] and/or arise from the selection of genetically distinct clones. We used scRNA-seq data to infer copy number variations (CNV) and link them to cells assigned to different morpho-biotypes. An average morpho-biotype-specific CNV pattern was generated for each patient from the single-cell CNVs assigned to each morpho-biotype. Consistent with the high inter-patient heterogeneity, the average CNV profiles clustered by patient rather than by morpho-biotype (Figures 4A and 4B). However, a higher number of CNVs was observed in cells of the undifferentiated morpho-biotype than in the others, albeit differences were of small magnitude (Figure 4C).

Next, we generated mutational profiles of laser-microdissected areas assigned to the different morpho-biotypes based

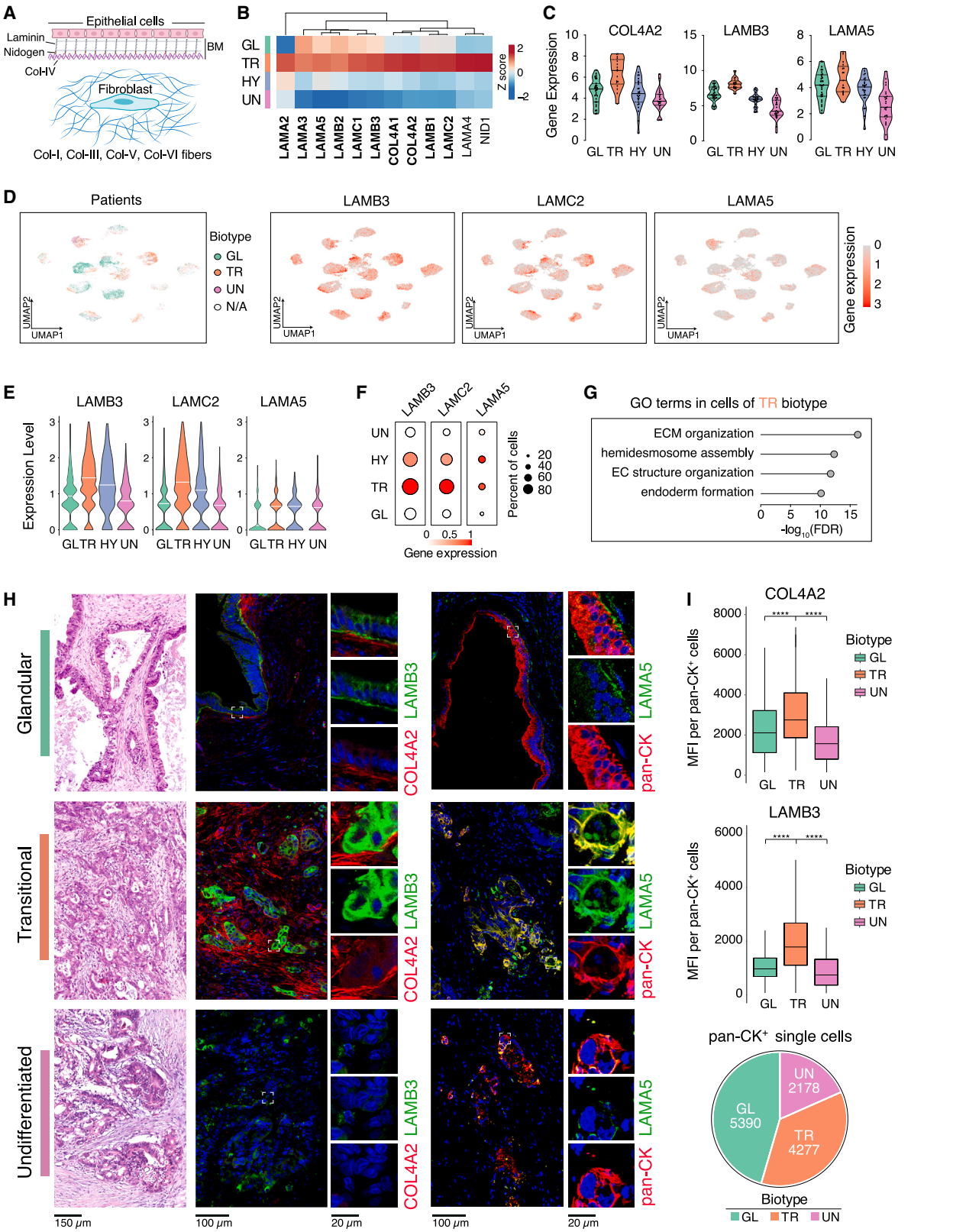


Figure 5. Production and assembly of basement membranes in PDAC morpho-biotypes

(A) Schematic representation of the BM and its main components (top) and the extracellular matrix (ECM) containing collagen fibers produced by fibroblasts (bottom). (B) Heatmap showing the main BM components[36] expressed in PDAC morpho-biotypes. Differentially expressed genes are labeled in bold.

(legend continued on next page)

on gene expression data. To this aim, we used a custom amplicon-based panel[32] covering 467 genes selected based on a revision of the literature updated to the PanCancer analysis[33,34] (Frigè et al., data not shown). When considering both a restricted set of PDAC driver genes[6] and the entire gene panel (Figure 4D; Table S7), no significant differences were observed among morpho-biotypes, suggesting they may not arise because of the selection and expansion of clones with distinct driver genes.

Differential basement membrane production in PDAC morpho-biotypes

The high expression of ECM genes in the transitional morpho-biotype (Figure 3B) suggests a possible direct influence of transitional tumor cells on architecture and composition of the surrounding tissue. We thus analyzed morpho-biotype-specific changes occurring in two distinct ECM compartments, the basement membrane (BM) and the interstitial matrix (Figure 5A).

The BM, which provides a barrier against tumor invasion,[35] is an interwoven network of polymerized laminins and type IV collagen, further crosslinked by additional proteins, such as nidogen and perlecan, and anchored to the surface of epithelial cells.[36] The polarized secretion of BM components at the basal side of epithelial cells is enforced[37] and required to maintain apico-basal cell polarity.[38,39]

In LMD-seq data, laminins and type IV collagens were almost undetectable in the *undifferentiated* morpho-biotype, well expressed in the *glandular* morpho-biotype and strongly increased in the *transitional* one (Figures 5B and 5C). To exclude any impact of contaminating fibroblasts, which may contribute to BM secretion, we used a scRNA-seq dataset[4] to examine the expression of BM components in malignant cells of different biotypes. Transitional cells exhibited significantly higher expression of BM components than the glandular ones (Figures 5D–5F). Moreover, BM components were nearly undetectable in malignant cells of the undifferentiated morpho-biotype (Figures 5D–5F). Consistent with this observation, functional terms related to ECM were enriched in transitional cells (Figure 5G).

Staining of type IV collagen and laminins (LAMB3 and LAMA5) revealed a homogeneous and continuous BM in *glandular* areas (Figure 5H). *Transitional* areas exhibited not only a strongly increased abundance of BM components, but also a massive cytoplasmic accumulation of laminins in cells that lost apico-basal polarity (Figure 5H). Hence, this biotype displays concomitant alterations in cell polarity and secretion of ECM components. Single-cell quantification of representative stainings is shown in Figure 5I.

These data indicate that the transitional and undifferentiated morpho-biotypes display a compromised BM organization, ranging from its complete absence to significant structural subversion, which may facilitate invasion of the surrounding tissue.

Myofibroblast-like programs in PDAC transitional cells

We next examined the *interstitial* matrix[40] that is actively produced by myofibroblasts generated during wound healing and/or in response to mechanical stress from fibroblast precursors and to a lesser extent from *trans*-differentiated endothelial or epithelial cells.[41,42]

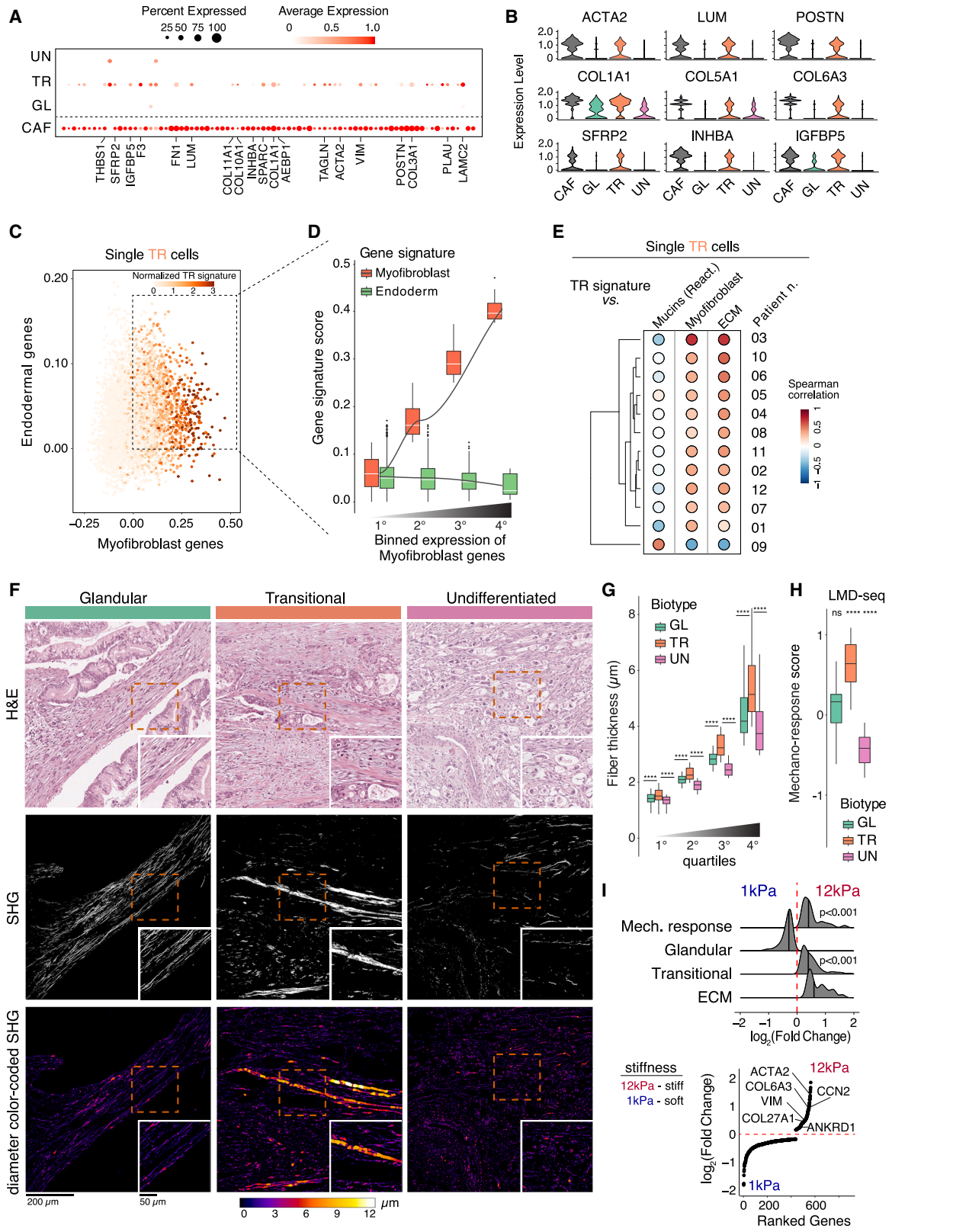
Using a single-nucleus RNA-seq dataset,[5] we measured the expression of myofibroblast genes in CAFs and tumor cells. Cells of the transitional morpho-biotype exhibited elevated expression of genes of the myofibroblast lineage and also highly expressed by PDAC CAFs but not by cells of the glandular or undifferentiated morpho-biotypes (Figures 6A and 6B). This set encompassed the myofibroblast marker gene *ACTA2*, which encodes αSMA (an actin isoform that enhances cell contractile activity and resistance to tensile stress),[41] as well as genes encoding signaling molecules (e.g., *SFRP2*, *INHBA*, and *IGFBP5*) and components of the interstitial matrix, including fibrous collagens (*COL1A1*, *COL5A1*, and *COL6A3*), proteoglycans (*LUM*), and ECM cross-linkers (e.g., *POSTN*). Overall, transitional cells displayed myofibroblast-like properties.

However, the transitional morpho-biotype retained the expression of many endodermal genes (Figure 1). To gain further insight into this transcriptional plasticity, we analyzed the relative enrichment of endodermal and myofibroblast gene signatures. Transitional cells exhibited a range of transcriptional programs that varied from the coexistence of low levels of endodermal and myofibroblast gene signatures, to the exclusive and high expression of myofibroblast genes with minimal residual endodermal gene expression (Figures 6C and 6D). Hence, the acquisition of myofibroblast-like properties by PDAC cells is a gradual rather than a switch-like transition from glandular cells expressing endodermal programs.

When considering individual patients' data, transitional cells in each patient were characterized by variable enrichments of endodermal, myofibroblast and ECM signatures, although in most cases mesenchymal signatures were significantly more enriched (Figure 6E).

Based on these data, we tested whether the elevated production of ECM components by PDAC cells of the transitional morpho-biotype may affect the organization of the surrounding stroma. We exploited the second harmonic generation (SHG) signal obtained by two-photon illumination[43] to measure the

(C) Expression levels stratified by morpho-biotype of representative differentially expressed genes validated at the protein level (H and I). Gene expression represents the $\log_2$ (normalized read counts per kb).
(D) UMAPs of single tumor cells from scRNA-seq data in 13 primary PDACs labeled by morpho-biotype (left); expression of selected genes encoding BM components is shown (right).
(E and F) Normalized (E) and average (F) gene expression of selected BM components in PDAC cells classified by morpho-biotype.
(G) Gene Ontology classes enriched in transitional morpho-biotype cells.
(H) H&E sections and IF images of the PDAC tumor areas showing the BM in the different morpho-biotypes. Areas representative of the three morpho-biotypes in one PDAC patient were immunostained as indicated. Large IF images show the merged colors with DAPI-counterstained nuclei. White boxes indicate the area shown at higher magnification on the right.
(I) Quantification of MFI (mean fluorescence intensity) for COL4A2 (top) and LAMB3 (middle) in single pan-CK$^+$ cells. The pie chart at the bottom shows the distribution of these cells in each tumor area, categorized by biotype (n = 3 patients; 3–10 areas per patient). Two-sided Wilcoxon rank-sum tests (TR *vs*. GL; TR *vs*. UN) are shown.



(legend on next page)

local supramolecular fibrillar architecture of the collagen-rich matrix in PDAC areas of the three morpho-biotypes. When collagen fibers are hit by two low-energy incident photons, a scattering photon with twice the frequency of the incident light is generated,[44] enabling a direct measurement of collagen network organization. Tissue regions surrounding PDAC areas assigned to the three morpho-biotypes contained fibers with a broad range of thickness, which reflects the heterogeneity of the PDAC TME (Figure 6F). However, compared to glandular, and even more so to undifferentiated areas, transitional areas exhibited a substantial increase in thick collagen fibers in proximity to tumor cells (Figures 6F and 6G).

Thick collagen fibers, which are hallmarks of stiff tissues, can exert large stresses to which cells respond by mounting a mechanoresponsive gene expression program.[45] This mechanoresponse signature was significantly enriched in the transitional morpho-biotype compared to the glandular and undifferentiated ones (Figure 6H). A typical mechanoresponse to elevated stiffness and mechanical stresses is the increased deposition of fibrous ECM. While representing a protective mechanism of parenchymal cells subjected to persistent mechanical stress, this response triggers a feedforward loop that further exacerbates tissue mechanics and worsens fibrosis.[46] To determine whether a similar phenomenon occurs in PDAC, we cultured the primary human PDAC cell line MDA-PATC69 on collagen-coated hydrogels of defined rigidity (1–12 kPa) that mimic the range of stiffness recorded in PDAC.[47] Culture of MDA-PATC69 cells on the 12 kPa stiff matrix induced not only the mechanoresponse gene signature but also the transitional program and the expression of ECM components and *ACTA2* (Figure 6I).

Overall, the induction of the transitional morpho-biotype signature may occur in response to the local sensing of a stiff ECM, resulting in a vicious circle that exacerbates local fibrosis.

Switch of transcriptional programs during perineural invasion

The data shown previously linking the transitional program to ECM production and invasion prompted us to evaluate if morpho-biotypes were differentially enriched at sites of local invasion. Perineural invasion is detected in 93–100% of PDACs and is a critical dissemination route mediating the invasion of the superior mesenteric artery.[48]

The perineurium consists of concentric layers of flat cells and is encircled by a collagen-rich epineurium accounting for the tensile state of the nerves.[49] We employed LMD-seq to analyze PDAC cells located outside, in proximity to, or inside nerves (Figure 7A). Although the undifferentiated signature included neural-like terms, it did not show any differential enrichment outside vs. inside nerves, excluding its relevance for neural invasion. Conversely, we detected a robust increase in the transitional signature in PDAC areas immediately outside of the perineurium and a strong decrease of the same signature inside nerves (Figures 7B and 7C). This was associated with a mild increase of the glandular signature inside nerves. The induction of the transitional signature correlated with the high content of thick collagen fibers in the epineurium (Figure 7D) suggesting that it may have ensued as a response to local changes in tissue mechanics. Similarly, downregulation of the transitional signature inside nerves correlated with the steep decrease in collagen fiber thickness near the perineurial edges and the absence of fibers in the inner nerve region (Figure 7D).

While correlative, these data suggest the hypothesis that the activation of the transitional program is driven by the contact of PDAC cells with areas with high collagen fiber content, eventually triggering a positive feedback loop further increasing collagen deposition and stiffness (Figure 7E). The downregulation of this program once cancer cells enter the soft intraneural environment may reveal the intrinsic plasticity of biotype-specific gene expression programs and their adaptation to different microenvironments.

sFRP2 secretion by transitional PDAC cells enhances migration and invasion

Finally, a valuable outcome of this study would be the identification of regulators of PDAC properties with practical relevance. In pursuit of a proof of concept in this context, we directed our attention to the transitional morpho-biotype due to its correlation with unfavorable prognosis.

Figure 6. Myofibroblasts-like properties of transitional cells

(A) Average expression level of fibroblast signature genes in CAFs and single tumor cells assigned to individual biotypes. Genes shown were first selected by intersecting genes upregulated in the transitional biotype (LMD-seq dataset) and tissue fibroblast signature genes[5,77]; then the list was restricted to the top variable genes in snRNAseq dataset.

(B) Normalized expression of selected genes encoding myofibroblast differentiation markers (top), fibrous collagens (middle) and soluble regulatory factors (bottom) in CAFs and in single tumor cells classified by morpho-biotype.

(C) Scatterplot of single transitional cells (n = 5809) showing the relationship between myofibroblast (x axis) and endodermal (y axis) gene signatures. Each dot represents a single tumor cell colored based on the transitional gene signature score.

(D) Co-expression of myofibroblast (red) and endoderm (green) gene expression programs in single TR cells with endodermal and myofibroblast gene scores >0 (as in (C)). Cells were divided into four groups of equal number by dividing the expression of myofibroblast genes into four equal bins.

(E) Spearman correlation matrix for single TR cells from 12 patients obtained from pairwise comparisons of the transitional gene signature vs. an endodermal signature ("glycosylation of mucins", Reactome), a myofibroblast signature,[5] and a PDAC ECM signature.[68]

(F) H&E sections (top), second harmonic generation (SHG) images (middle), and diameter color-coded SHG images (bottom) of the PDAC tumor areas of the different morpho-biotypes. Orange boxes indicate the area shown at higher magnification on the bottom-right.

(G) Quantification of fiber thickness derived by SHG images for tumor areas stratified by morpho-biotype (n = 6 patients; 5–13 areas per patient). The fiber thickness values were divided into quartiles based on increasing signal. Two-sided Wilcoxon rank-sum tests (TR *vs*. GL; TR *vs*. UN) are shown.

(H) Gene expression scores of the mechanoresponse program[45] calculated for each microdissected sample and stratified by morpho-biotype. Two-sided Wilcoxon rank-sum tests (group *vs*. rest) are shown.

(I) Log_2 fold change distributions of core enriched genes of the mechanoresponse program,[45] the glandular and transitional biotype and a PDAC ECM signature[68] in MDA-PATC69 cells cultured on stiff (12 kPa) or soft (1 kPa) matrices. Statistically significant enrichments are indicated. Bottom: log_2 fold change values of enriched genes ranked in ascending order.

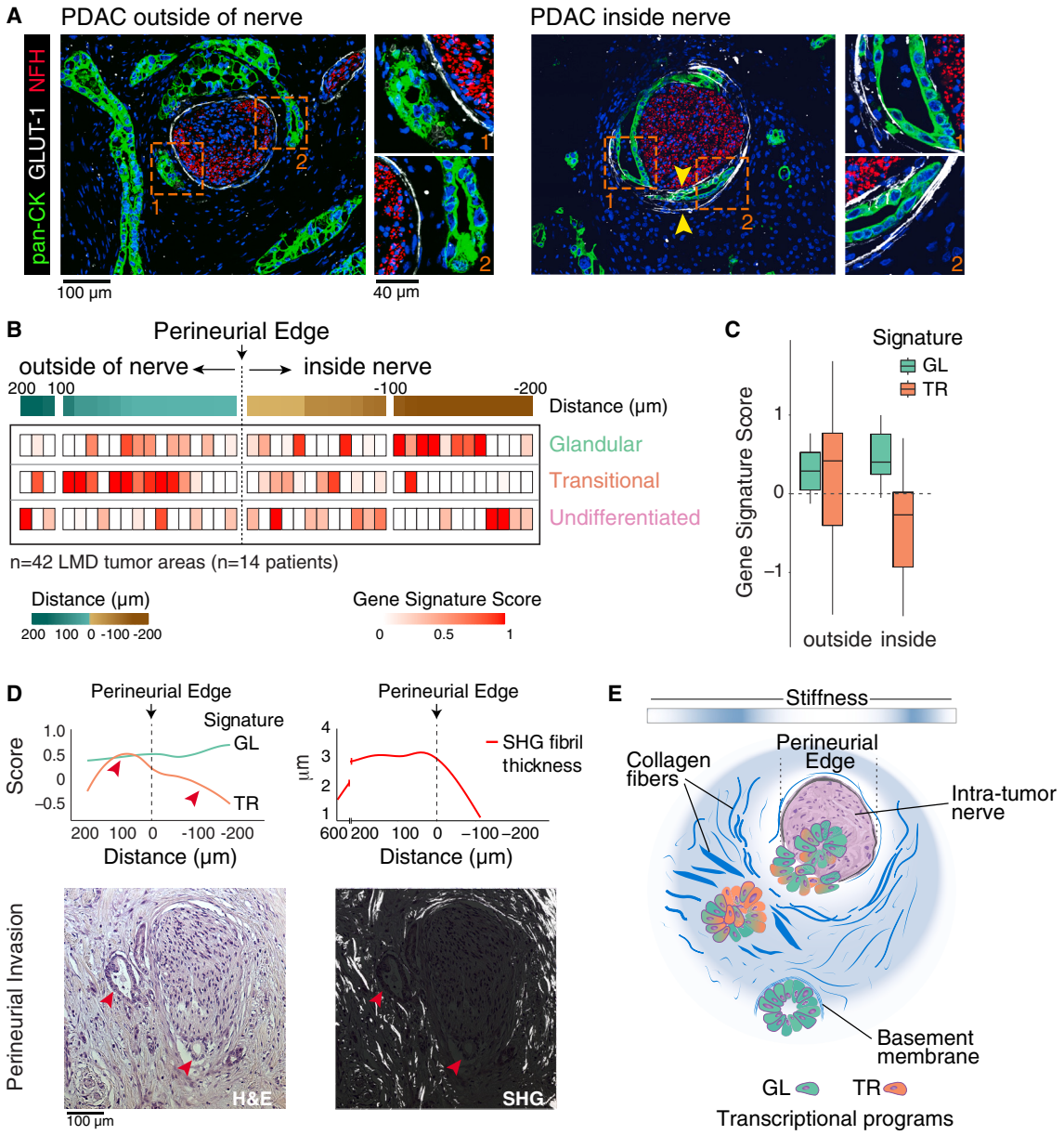


Figure 7. Induction of the transitional program at the periphery of intratumor nerves

(A) IF images showing tumor cells in proximity to a nerve or invading the perineural space and in contact with nerve fibers. Human PDAC slices were immunostained for pan-CK (green), Neurofilament heavy chain (NFH, red), and glucose transporter 1 (GLUT-1, white). Large immunofluorescence images show merged colors with DAPI-counterstained nuclei. Orange dashed boxes indicate the area shown at higher magnification on the right. Arrowheads indicate tumor cells infiltrating and delaminating the perineurium.

(B) Enrichment of morpho-biotype signatures in laser microdissected tumor areas at the indicated distance relative to the perineurial edge of the closest intratumor nerve. n = 42 areas from n = 14 treatment-naive patients.

(C) Quantification of the transitional and glandular gene signature scores in each sample outside and inside nerves.

(D) Relationship between glandular and transitional gene signature scores *vs*. distance with the perineurium and thickness of collagen fibers. The smoothed trend line of the glandular and transitional gene signature scores (left) and of the fiber thickness (right) at different distances from the perineurial edge are shown on top of a representative H&E (bottom left) and SHG (bottom right) image. Arrows indicate the tumor areas at different distances from the perineurium.

(E) Schematic representation of the relationship between transcriptional programs and nerves.

The genes upregulated in PDAC cells of the transitional morpho-biotype (Table S3) comprised regulators of mesenchymal properties, including secreted molecules. Among them, sFRP2 (secreted Frizzle-related protein 2), an antagonist of Wnt-induced signaling[50,51] currently explored as therapeutic target for fibrotic disorders and cancer,[52,53] was of particular

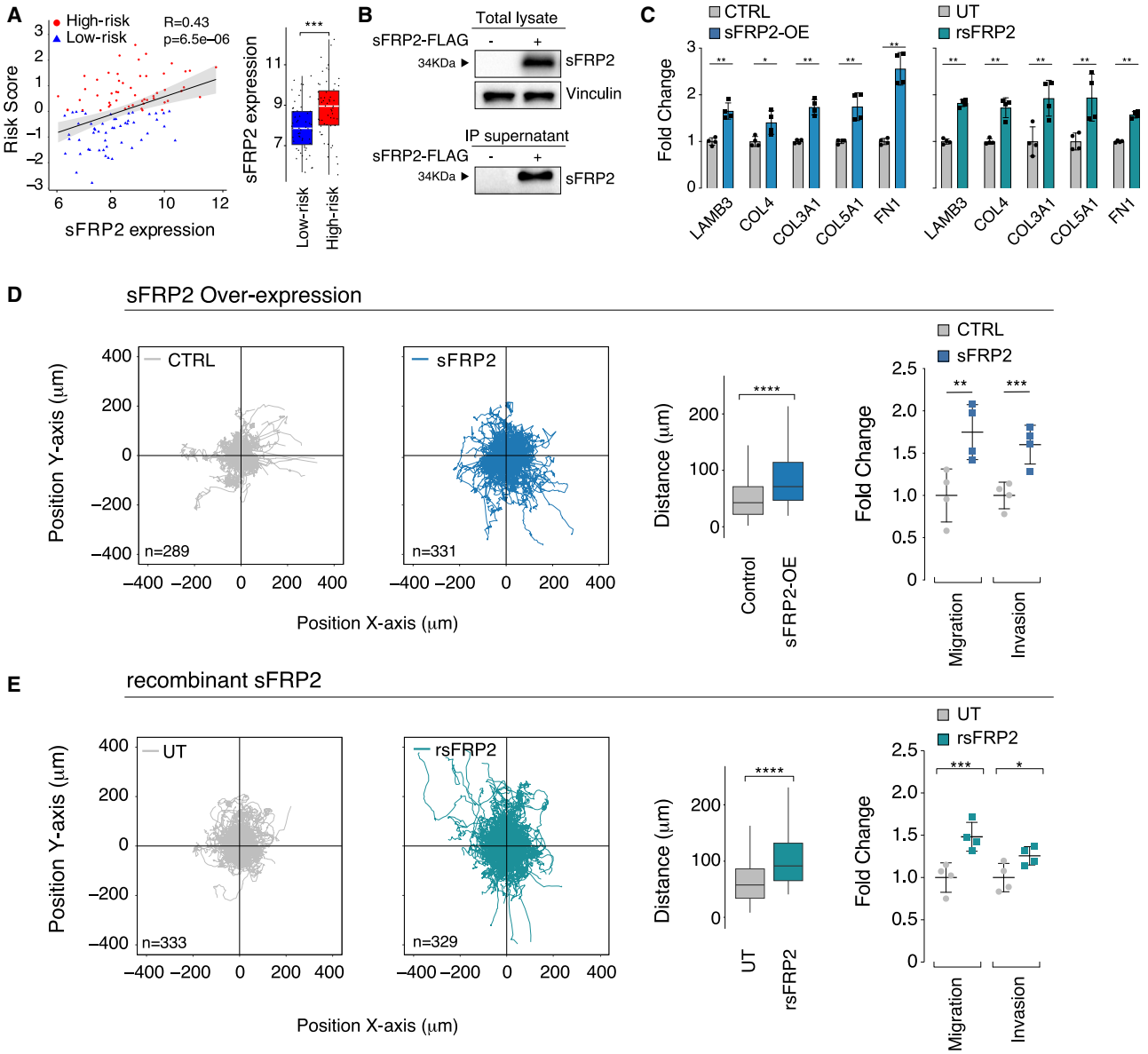


Figure 8. Functional effects of sFRP2 on invasion and migration of PDAC cells

(A) Left, scatterplot and linear regression line showing the correlation between the expression of *SFRP2* (x axis) and risk score (y axis) calculated for each patient in the PACA-CA cohort (n = 110).[4] Right: *SFRP2* expression categorized by patient risk group (as in Figure 2).

(B) sFRP2 western blot of total lysate (top) and anti-FLAG immunoprecipitation from culture supernatant (bottom) of MDA-PATC69 cells transfected with either control or sFRP2-FLAG vector. Vinculin: loading control.

(C) ELISA measurement of ECM components in the MDA-PATC69 PDAC cell line transfected with control or sFRP2-FLAG vector (left bars) or treated with recombinant sFRP2 (right bars). UT: untreated. Means ± SD are shown (two biological replicates with two technical duplicates each). Two-tailed t tests are shown.

(D) Random migration, chemotaxis migration and matrigel invasion assays in MDA-PATC69 cells transfected with control vector (CTRL) or overexpressing (OE) sFRP2. Trajectory plots with maxima distance quantification are reported (n = 2 sets). Two-sided Wilcoxon rank-sum tests are shown. Chemotaxis migration and Matrigel invasion assays are expressed as fold change compared to CTRL (two biological replicates with two technical duplicates). Means ± SD and two-tailed t tests are shown.

(E) Random migration, chemotaxis migration and matrigel invasion assays were carried out as in (D) in MDA-PATC69 cells stimulated with rsFRP2 or left untreated (UT). Two-sided Wilcoxon rank-sum tests are shown. Chemotaxis migration and matrigel invasion assays are expressed as fold change compared to UT (two biological replicates with two technical duplicates). Means ± SD and two-tailed t tests are shown.

interest. sFRP2 is over-expressed by CAFs in tumors with stromal signatures and high mesenchymal gene expression,[54–56] contributing to CAF-instigated induction of invasive properties of cancer cells and poor prognosis in melanoma.[55] Similarly, high *SFRP2* expression in the epithelial component of PDACs correlated with poor prognosis (Figure 8A). This finding

prompted us to investigate if sFRP2 production by PDAC cells could enhance their migratory and invasive capacity. Indeed, even if sFRP2 can also be efficiently produced by CAFs, its high affinity to ECM components greatly limits its diffusion rate, thus preventing long-distance effects.[57] We over-expressed human sFRP2 into the MDA-PATC69 cell line, which shows undetectable endogenous expression. After verifying proper secretion of sFRP2 (Figure 8B), we analyzed its effects on ECM production. In parallel, we tested the effects of treatment with recombinant sFRP2 (rsFRP2). Cell-based ELISA assays showed a significantly increased deposition of all the ECM proteins analyzed (Figure 8C). Moreover, PDAC cells endogenously expressing sFRP2 (Figure 8D) or treated with rsFRP2 showed increased migratory and invasive capabilities (Figure 8E).

These results suggest that the production of sFRP2 by transitional PDAC cells, may contribute to local ECM remodeling and the acquisition by PDAC cells of migratory and invasive capabilities.

DISCUSSION

This study aimed to intersect the heterogeneous morphological patterns, which not only coexist in most PDACs but also possess prognostic value,[1,11,16] with the distinct gene expression programs and genomic alteration profiles underlying PDAC heterogeneity.

We introduced the concept of *morpho-biotypes* to highlight the notion that PDAC cells with different gene expression programs are organized into morphologically distinct and spatially discrete tumor areas, indicating either varied cell states influenced by microenvironmental signals or the localized expansion of distinct clones. However, mutational profiles were not morpho-biotype-specific and we observed only a slightly higher burden of chromosomal alterations in the undifferentiated morpho-biotype than in the more differentiated glandular and transitional ones. Hence, microenvironmental signals, such as local stiffness, may have a dominant role in instructing the properties of distinct morpho-biotypes. Importantly, the morpho-biotypes must themselves be considered broad patterns encompassing morphologically heterogeneous variants underpinned by a vast molecular heterogeneity.

The terminological choices we made warrant some clarifications. Differently from other common cancers, no unified terminology and classification system exists for PDAC. Moreover, because of technical biases during sample preparation, variations in tumor purity and variable inclusion of tumor stroma in the analysis, a substantial lack of consistency among PDAC molecular classification schemes exists, as reported in previous systematic analyses.[58,59] Because of such issues, objective difficulties arise in unambiguously assigning clinical samples to molecular signatures identified from bulk or single cell tumor samples, as it becomes clear by comparing some of the most frequently cited publications in this field. For example, 50% of the tumors classified as "*squamous*" in Bailey et al.[6] had instead previously been classified as "*basal*" in the study by Moffit et al.,[9] and the remaining ones were composed of a mixture of other subtypes, including the *quasi-mesenchymal* variant reported in Collison et al..[8] Although the overlap of the "*squamous*" signature genes with those of the "*basal*" and "*quasi-mesenchymal*" signatures is high,[58] the squamous signature includes several hundred genes, and hence is likely to contain genes belonging to several distinct functional programs. Along the same line, another recent study,[5] showed that "squamoid" or "mesenchymal" programs defined on the basis of snRNA-seq data did not exhibit any significant overlap with the squamous or *quasi-*mesenchymal signatures identified in bulk transcriptome analyses.

Overall, these considerations unveil a confusing scenario whereby the terms "*basal*" and "*mesenchymal*" and less commonly "*squamous*" are often interchangeably (and thus inappropriately) used to refer to all the non-endodermal (or "*non-classical*") PDAC subtypes, without properly representing their profound heterogeneity. Moreover, the names chosen for each of these signatures do not always reflect the underlying biology and properties of the different PDAC cell variants.

Based on these considerations, the partially novel terminology chosen in our study was meant as an attempt to provide a rational framework to describe the diverse transcriptional and functional programs expressed by coexisting and morphologically different tumor areas.

The term "*glandular*", rather than "*classical*" or "*progenitor-like*" is meant to indicate the combination of a defined morphological pattern and a gene expression programs typical of glandular endoderm.

The term "*transitional*" evokes the epithelial-to-mesenchymal *transition* and is meant to describe cells that display gene expression programs intermediate between those of endodermal and mesenchymal- or myofibroblast-like cells, a concept extensively validated in our study. The previously used term *quasi-mesenchymal*[8] also appears suitable to describe this variant, although it falls short of delivering the notion that cells of this variant still express endodermal genes and hence are in a *transitional* state.

Finally, the term "*undifferentiated*" refers to the lack of histologically evident differentiation as well as the presence of a gene expression program characterized by the broad but low-level expression of a large number of genes, a distinguishing property of undifferentiated cells (including stem and progenitor cells).[31] The expression of neural genes by this undifferentiated population of cells is, in our interpretation, an event compatible with lineage priming, not a *bona fide* neuronal differentiation. The lineage-priming concept has been originally developed in the hematopoietic system[60] and then extended to other tissues[61] to indicate a phenomenon whereby stem or progenitor cells occasionally express at very low levels a subset of the genes associated to the differentiation pathways they can commit to.

The identification of a neuronal-like gene expression profile in a tumor of an endodermal organ may appear surprising. However, the endodermal origin of a subset of PDX1-positive neural progenitors contributing to the enteric nervous system has been reported in humans and mice,[62–64] indicating a developmental trajectory that may be pathologically reactivated in tumors. Moreover, neurogenic programs have been associated with prostate cancer cell stemness[65] and poor prognosis of the neuroendocrine prostate cancer subtype.[66] The fact that the undifferentiated biotype also overlaps with squamous signatures is compatible with the presence of undifferentiated cancer cells

undergoing priming toward the neuronal lineage but also capable of stochastically initiate alternative fates such as squamous differentiation. Indeed, cells of this biotype frequently expressed TP63, which maintains stemness and proliferative capacity of basal cells and promotes plasticity and propension to multilineage priming.[67] In this regard, the observation that basal, squamous, and neuronal genes may be co-expressed within the same tumor regions, points at the existence of pathological niches in which stem/progenitor cells (such as basal-like cells) unable to generate differentiated ducts may instead be primed toward alternative squamous-like or neural-like programs.

Expression of this neuronal-like program at low levels and in a small proportion of tumor cells may explain why it was overlooked in both bulk and single-cell analyses. However, a similar signature was detected after neoadjuvant treatment in a single nucleus RNA-seq dataset,[5] suggesting extensive loss of differentiation after chemotherapy. It is important to stress that although the undifferentiated tumor areas usually represent a rather small proportion of the tumor, they are pervasively represented in clinical samples, which may have clinical implications particularly in terms of response to therapy.

Our data also provide insight into regional heterogeneity of the PDAC TME, which has previously been linked to alternative differentiation, functional properties and abundance of CAFs.[13] We found morpho-biotype-specific and possibly tumor cell-driven subversions of both BM and interstitial matrix. In particular, the transitional morpho-biotype displayed an increased, yet disorganized production of BM components, associated with loss of apico-basal polarity and cytoplasmic retention of laminins. As a result, transitional areas were embedded in abundant but disorganized type IV collagen and not surrounded by a defined BM. A total lack of BM was observed in the undifferentiated morpho-biotype because of the lack of expression of BM components.

Transitional cells displayed a myofibroblast-like gene expression program associated with the production of ECM components and the generation of thick collagen fibers in the surrounding tissue. Notably, the production of matrisomal proteins by PDAC cells, but not by the stroma, correlates with poor prognosis,[68,69] and is associated with prometastatic effects,[70] indicating that the acquisition of myofibroblast-like properties is relevant for disease progression. Previous data linked loss of TGFβ signaling in *SMAD4*-mutant PDACs to increased deposition of collagen fibers in peritumoral areas, leading to increased stiffness and shorter survival.[47] Our data, while not contradicting these findings, indicate that within the same tumor, heterogeneity of collagen fibers in peritumoral areas can occur in the absence of detectable genetic variation. An important issue in this context is the relative contribution of CAFs (and specifically myofibroblast-like CAFs, myCAFs) and PDAC cells to the subversion of the architectural organization of the stroma in proximity of transitional areas. Available data indicate that αSMA-positive myCAFs surround periglandular areas.[71,72] Similarly, spatially confined communities containing "classical" PDAC cells are enriched in myCAFs.[5] Hence, these data, while only correlative, suggest that the reorganization of the interstitial stroma in proximity of transitional areas is unlikely to be explained by myCAFs, suggesting instead a role of myofibroblast-like PDAC cells.

In the specific instance of neural invasion, a clear correlation existed between the activation of the transitional gene expression program and contact with the stiff epineurial layer. Conversely, the downregulation of this program inside nerves correlated with the absence of intraneural collagen fibers. An inference of this model is the existence of a feedforward loop in which, on the one hand, the increased mechanical stress exerted by a rigid matrix induces the transitional program, and on the other the increased synthesis of matrisomal protein exacerbates local fibrosis.[73] The desmoplastic matrix stimulates tumor invasive properties and dissemination by various mechanisms,[39] such as promoting active migration of tumor cells along the immobilized gradients of fibronectin and collagen fibers (haptotaxis).[73,74]

Finally, our results suggest the possibility to implement artificial intelligence (AI)-driven approaches to connect histological patterns with gene expression signatures,[75,76] eventually enabling a detailed mapping of biotype composition of each patient, which may assist in therapeutic choices. To accomplish this goal, it will be crucial to create experimental models reproducing the characteristics of PDAC biotypes.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human specimens
  - Clinical data
  - Primary human PDAC cells
- METHODS DETAILS
  - Histological review
  - Laser microdissection (LMD)
  - LMD-seq library preparation
  - Gene panel mutational analysis
  - Immunofluorescence
  - Two-photon second harmonic generation (SHG)
  - Cloning and plasmid construction
  - Cell culture, transfection and treatments
  - Western Blot and immunoprecipitation
  - Random migration and Boyden chamber assays
  - Cell-based ELISA
  - RNA-seq library preparation
  - GeoMx digital spatial profiling (DSP)
  - Multiplexed RNA smFISH
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - RNA-seq data analysis
  - Gene set enrichment analysis (GSEA)
  - LMD-seq data analysis
  - Subtype identification using consensus clustering
  - Differential gene expression analysis
  - Unsupervised clustering using k-nearest neighbors
  - Functional enrichment analysis and gene signature score

- Supervised learning using Random Forest
- Prognostic model
- GeoMx digital spatial profiling data analysis
- Gene panel mutational data analysis
- Mutation calling pipeline
- Single-cell RNAseq data analysis
- Single cell classification
- CytoTRACE
- CNV estimation
- Multiplexed RNA smFISH data analysis
- Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.ccell.2024.02.017.

ACKNOWLEDGMENTS

We thank Andrea Viale (MD Anderson Cancer Center, Houston, USA) and Silvia Monticelli (IRB, Bellinzona, CH) for comments on the manuscript. This study was supported by AIRC, the Italian Association for Research on Cancer (AIRC Investigator Grants #20251 to G.N. and #18621 to G.S.; AIRC 5x1000 Grants #21147 to G.N. and #22759 to G.S.), the ERC Synergy Grant #101071470 to G.S. and the Italian Ministry of Health (grant GR-2016-02361721 to G.R.D.). This work was also partially supported by the Italian Ministry of Health with the "Ricerca Corrente" and "5x1000" funds to the IEO IRCCS; and with "Ricerca Corrente" and "Ricerca Corrente di Rete" (ACCORD) 2020–2022 funds to Alleanza Contro il Cancro. P.D.C. and F.A. were supported by AIRC fellowships for Italy. The single cell RNA-seq datasets from PDAC samples of the PACA-CA cohort used in this research[4] were provided by the Ontario Institute for Cancer Research (O.I.C.R.), Toronto, Ontario, and generated through funding provided by the Government of Ontario.

AUTHOR CONTRIBUTIONS

Experimental design and conceptualization: P.D.C., L.N., G.R.D., and G.N. Experimental work and data generation: L.N., P.D.C., F.A., S.B., A.P., G.F., G.R.D., S.P., and B.D. Data analysis: P.D.C., F.G., and C.S. Clinical samples and pathology: A.Z. and P.S. Supervision: L.M., A.C., G.S., S.R., I.B., G.R.D., and G.N. Funding acquisition: G.R.D. and G.N. Manuscript writing: P.D.C., L.N., G.R.D., and G.N. with inputs from all authors.

DECLARATION OF INTERESTS

L.M. reports research funding from Bristol-Myers Squibb and Roche, and consulting fees from Tethis, which are all unrelated to the content of this study.

Received: December 21, 2022
Revised: December 7, 2023
Accepted: February 27, 2024
Published: March 21, 2024

REFERENCES

1. Verbeke, C. (2016). Morphological heterogeneity in ductal adenocarcinoma of the pancreas - Does it matter? Pancreatology *16*, 295–301. https://doi.org/10.1016/j.pan.2016.02.004.
2. Juiz, N., Elkaoutari, A., Bigonnet, M., Gayet, O., Roques, J., Nicolle, R., Iovanna, J., and Dusetti, N. (2020). Basal-like and classical cells coexist in pancreatic cancer revealed by single-cell analysis on biopsy-derived pancreatic cancer organoids from the classical subtype. FASEB J *34*, 12214–12228. https://doi.org/10.1096/fj.202000363RR.
3. Raghavan, S., Winter, P.S., Navia, A.W., Williams, H.L., DenAdel, A., Lowder, K.E., Galvez-Reyes, J., Kalekar, R.L., Mulugeta, N., Kapner, K.S., et al. (2021). Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. Cell *184*, 6119–6137.e26. https://doi.org/10.1016/j.cell.2021.11.017.
4. Chan-Seng-Yue, M., Kim, J.C., Wilson, G.W., Ng, K., Figueroa, E.F., O'Kane, G.M., Connor, A.A., Denroche, R.E., Grant, R.C., McLeod, J., et al. (2020). Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. Nat. Genet. *52*, 231–240. https://doi.org/10.1038/s41588-019-0566-9.
5. Hwang, W.L., Jagadeesh, K.A., Guo, J.A., Hoffman, H.I., Yadollahpour, P., Reeves, J.W., Mohan, R., Drokhlyansky, E., Van Wittenberghe, N., Ashenberg, O., et al. (2022). Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment. Nat. Genet. *54*, 1178–1191. https://doi.org/10.1038/s41588-022-01134-8.
6. Bailey, P., Chang, D.K., Nones, K., Johns, A.L., Patch, A.M., Gingras, M.C., Miller, D.K., Christ, A.N., Bruxner, T.J.C., Quinn, M.C., et al. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. Nature *531*, 47–52. https://doi.org/10.1038/nature16965.
7. Collisson, E.A., Bailey, P., Chang, D.K., and Biankin, A.V. (2019). Molecular subtypes of pancreatic cancer. Nat. Rev. Gastroenterol. Hepatol. *16*, 207–220. https://doi.org/10.1038/s41575-019-0109-y.
8. Collisson, E.A., Sadanandam, A., Olson, P., Gibb, W.J., Truitt, M., Gu, S., Cooc, J., Weinkle, J., Kim, G.E., Jakkula, L., et al. (2011). Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat. Med. *17*, 500–503. https://doi.org/10.1038/nm.2344.
9. Moffitt, R.A., Marayati, R., Flate, E.L., Volmar, K.E., Loeza, S.G.H., Hoadley, K.A., Rashid, N.U., Williams, L.A., Eaton, S.C., Chung, A.H., et al. (2015). Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat. Genet. *47*, 1168–1178. https://doi.org/10.1038/ng.3398.
10. Nieto, M.A., Huang, R.Y.J., Jackson, R.A., and Thiery, J.P. (2016). Emt: 2016. Cell *166*, 21–45. https://doi.org/10.1016/j.cell.2016.06.028.
11. N Kalimuthu, S., Wilson, G.W., Grant, R.C., Seto, M., O'Kane, G., Vajpeyi, R., Notta, F., Gallinger, S., and Chetty, R. (2020). Morphological classification of pancreatic ductal adenocarcinoma that predicts molecular subtypes and correlates with clinical outcome. Gut *69*, 317–328. https://doi.org/10.1136/gutjnl-2019-318217.
12. Hayashi, A., Hong, J., and Iacobuzio-Donahue, C.A. (2021). The pancreatic cancer genome revisited. Nat. Rev. Gastroenterol. Hepatol. *18*, 469–481. https://doi.org/10.1038/s41575-021-00463-z.
13. Grünwald, B.T., Devisme, A., Andrieux, G., Vyas, F., Aliar, K., McCloskey, C.W., Macklin, A., Jang, G.H., Denroche, R., Romero, J.M., et al. (2021). Spatially confined sub-tumor microenvironments in pancreatic cancer. Cell *184*, 5577–5592.e18. https://doi.org/10.1016/j.cell.2021.09.022.
14. Tu, M., Klein, L., Espinet, E., Georgomanolis, T., Wegwitz, F., Li, X., Urbach, L., Danieli-Mackay, A., Küffer, S., Bojarczuk, K., et al. (2021). TNF-alpha-producing macrophages determine subtype identity and prognosis via AP1 enhancer reprogramming in pancreatic cancer. Nat. Cancer *2*, 1185–1203. https://doi.org/10.1038/s43018-021-00258-w.
15. Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Mach. Learn. *52*, 91–118.
16. Adsay, N.V., Basturk, O., Bonnett, M., Kilinc, N., Andea, A.A., Feng, J., Che, M., Aulicino, M.R., Levi, E., and Cheng, J.D. (2005). A proposal for a new and more practical grading scheme for pancreatic ductal adenocarcinoma. Am. J. Surg. Pathol. *29*, 724–733. https://doi.org/10.1097/01.pas.0000163360.40357.f1.
17. Diaferia, G.R., Balestrieri, C., Prosperini, E., Nicoli, P., Spaggiari, P., Zerbi, A., and Natoli, G. (2016). Dissection of transcriptional and cis-regulatory control of differentiation in human pancreatic cancer. EMBO J. *35*, 595–617. https://doi.org/10.15252/embj.201592404.
18. Milan, M., Diaferia, G.R., and Natoli, G. (2021). Tumor cell heterogeneity and its transcriptional bases in pancreatic cancer: a tale of two cell types and their many variants. EMBO J. *40*, e107206. https://doi.org/10.15252/embj.2020107206.

19. Bailey, J.M., Swanson, B.J., Hamada, T., Eggers, J.P., Singh, P.K., Caffery, T., Ouellette, M.M., and Hollingsworth, M.A. (2008). Sonic hedgehog promotes desmoplasia in pancreatic cancer. Clin. Cancer Res. *14*, 5995–6004. https://doi.org/10.1158/1078-0432.CCR-08-0291.

20. Yauch, R.L., Gould, S.E., Scales, S.J., Tang, T., Tian, H., Ahn, C.P., Marshall, D., Fu, L., Januario, T., Kallop, D., et al. (2008). A paracrine requirement for hedgehog signalling in cancer. Nature *455*, 406–410. https://doi.org/10.1038/nature07275.

21. Balestrieri, C., Alfarano, G., Milan, M., Tosi, V., Prosperini, E., Nicoli, P., Palamidessi, A., Scita, G., Diaferia, G.R., and Natoli, G. (2018). Co-optation of Tandem DNA Repeats for the Maintenance of Mesenchymal Identity. Cell *173*, 1150–1164.e14. https://doi.org/10.1016/j.cell.2018.03.081.

22. De Craene, B., and Berx, G. (2013). Regulatory networks defining EMT during cancer initiation and progression. Nat. Rev. Cancer *13*, 97–110. https://doi.org/10.1038/nrc3447.

23. Nylander, K., Vojtesek, B., Nenutil, R., Lindgren, B., Roos, G., Zhanxiang, W., Sjöström, B., Dahlqvist, A., and Coates, P.J. (2002). Differential expression of p63 isoforms in normal tissues and neoplastic cells. J. Pathol. *198*, 417–427. https://doi.org/10.1002/path.1231.

24. Hayashi, A., Fan, J., Chen, R., Ho, Y.-j., Makohon-Moore, A.P., Lecomte, N., Zhong, Y., Hong, J., Huang, J., Sakamoto, H., et al. (2020). A unifying paradigm for transcriptional heterogeneity and squamous features in pancreatic ductal adenocarcinoma. Nat. Cancer *1*, 59–74. https://doi.org/10.1038/s43018-019-0010-1.

25. Bailey, J.M., Alsina, J., Rasheed, Z.A., McAllister, F.M., Fu, Y.Y., Plentz, R., Zhang, H., Pasricha, P.J., Bardeesy, N., Matsui, W., et al. (2014). DCLK1 marks a morphologically distinct subpopulation of cells with stem cell properties in preinvasive pancreatic cancer. Gastroenterology *146*, 245–256. https://doi.org/10.1053/j.gastro.2013.09.050.

26. Westphalen, C.B., Takemoto, Y., Tanaka, T., Macchini, M., Jiang, Z., Renz, B.W., Chen, X., Ormanns, S., Nagar, K., Tailor, Y., et al. (2016). Dclk1 Defines Quiescent Pancreatic Progenitors that Promote Injury-Induced Regeneration and Tumorigenesis. Cell Stem Cell *18*, 441–455. https://doi.org/10.1016/j.stem.2016.03.016.

27. Williams, H.L., Dias Costa, A., Zhang, J., Raghavan, S., Winter, P.S., Kapner, K.S., Ginebaugh, S.P., Väyrynen, S.A., Väyrynen, J.P., Yuan, C., et al. (2023). Spatially Resolved Single-Cell Assessment of Pancreatic Cancer Expression Subtypes Reveals Co-expressor Phenotypes and Extensive Intratumoral Heterogeneity. Cancer Res. *83*, 441–455. https://doi.org/10.1158/0008-5472.CAN-22-3050.

28. Halbrook, C.J., Lyssiotis, C.A., Pasca di Magliano, M., and Maitra, A. (2023). Pancreatic cancer: Advances and challenges. Cell *186*, 1729–1754. https://doi.org/10.1016/j.cell.2023.02.014.

29. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401. https://doi.org/10.1126/science.1254257.

30. Haque, A., Engel, J., Teichmann, S.A., and Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med. *9*, 75. https://doi.org/10.1186/s13073-017-0467-4.

31. Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan, F., Kuo, A.H., Hsieh, R.W., Cai, S., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. Science *367*, 405–411. https://doi.org/10.1126/science.aax0249.

32. Gregorc, V., Mazzarella, L., Lazzari, C., Graziano, P., Vigneri, P., Genova, C., Toschi, L., Ciliberto, G., Bonanno, L., Delmonte, A., et al. (2021). Prospective Validation of the Italian Alliance Against Cancer Lung Panel in Patients With Advanced Non-Small-Cell Lung Cancer. Clin. Lung Cancer *22*, e637–e641. https://doi.org/10.1016/j.cllc.2020.12.007.

33. Huang, K.L., Mashl, R.J., Wu, Y., Ritter, D.I., Wang, J., Oh, C., Paczkowska, M., Reynolds, S., Wyczalkowski, M.A., Oak, N., et al. (2018). Pathogenic Germline Variants in 10,389 Adult Cancers. Cell *173*, 355–370.e14. https://doi.org/10.1016/j.cell.2018.03.039.

34. Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371–385.e18. https://doi.org/10.1016/j.cell.2018.02.060.

35. Chang, J., and Chaudhuri, O. (2019). Beyond proteases: Basement membrane mechanics and cancer invasion. J. Cell Biol. *218*, 2456–2469. https://doi.org/10.1083/jcb.201903066.

36. Jayadev, R., and Sherwood, D.R. (2017). Basement membranes. Curr. Biol. *27*, R207–R211. https://doi.org/10.1016/j.cub.2017.02.006.

37. Devergne, O., Tsung, K., Barcelo, G., and Schüpbach, T. (2014). Polarized deposition of basement membrane proteins depends on Phosphatidylinositol synthase and the levels of Phosphatidylinositol 4,5-bisphosphate. Proc. Natl. Acad. Sci. USA *111*, 7689–7694. https://doi.org/10.1073/pnas.1407351111.

38. Bryant, D.M., and Mostov, K.E. (2008). From cells to organs: building polarized tissue. Nat. Rev. Mol. Cell Biol. *9*, 887–901. https://doi.org/10.1038/nrm2523.

39. Kai, F., Drain, A.P., and Weaver, V.M. (2019). The Extracellular Matrix Modulates the Metastatic Journey. Dev. Cell *49*, 332–346. https://doi.org/10.1016/j.devcel.2019.03.026.

40. Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K.J., and Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. Nat. Commun. *11*, 5120. https://doi.org/10.1038/s41467-020-18794-x.

41. Gabbiani, G. (2003). The myofibroblast in wound healing and fibrocontractive diseases. J. Pathol. *200*, 500–503. https://doi.org/10.1002/path.1427.

42. Klingberg, F., Hinz, B., and White, E.S. (2013). The myofibroblast matrix: implications for tissue repair and fibrosis. J. Pathol. *229*, 298–309. https://doi.org/10.1002/path.4104.

43. Campagnola, P.J., Millard, A.C., Terasaki, M., Hoppe, P.E., Malone, C.J., and Mohler, W.A. (2002). Three-dimensional high-resolution second-harmonic generation imaging of endogenous structural proteins in biological tissues. Biophys. J. *82*, 493–508. https://doi.org/10.1016/S0006-3495(02)75414-3.

44. Campagnola, P. (2011). Second harmonic generation imaging microscopy: applications to diseases diagnostics. Anal. Chem. *83*, 3224–3231. https://doi.org/10.1021/ac1032325.

45. Cordenonsi, M., Zanconato, F., Azzolin, L., Forcato, M., Rosato, A., Frasson, C., Inui, M., Montagner, M., Parenti, A.R., Poletti, A., et al. (2011). The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. Cell *147*, 759–772. https://doi.org/10.1016/j.cell.2011.09.048.

46. Tschumperlin, D.J., Ligresti, G., Hilscher, M.B., and Shah, V.H. (2018). Mechanosensing and fibrosis. J. Clin. Invest. *128*, 74–84. https://doi.org/10.1172/JCI93561.

47. Laklai, H., Miroshnikova, Y.A., Pickup, M.W., Collisson, E.A., Kim, G.E., Barrett, A.S., Hill, R.C., Lakins, J.N., Schlaepfer, D.D., Mouw, J.K., et al. (2016). Genotype tunes pancreatic ductal adenocarcinoma tissue tension to induce matricellular fibrosis and tumor progression. Nat. Med. *22*, 497–505. https://doi.org/10.1038/nm.4082.

48. Bapat, A.A., Hostetter, G., Von Hoff, D.D., and Han, H. (2011). Perineural invasion and associated pain in pancreatic cancer. Nat. Rev. Cancer *11*, 695–707. https://doi.org/10.1038/nrc3131.

49. Peltonen, S., Alanne, M., and Peltonen, J. (2013). Barriers of the peripheral nerve. Tissue Barriers *1*, e24956. https://doi.org/10.4161/tisb.24956.

50. Bovolenta, P., Esteve, P., Ruiz, J.M., Cisneros, E., and Lopez-Rios, J. (2008). Beyond Wnt inhibition: new functions of secreted Frizzled-related proteins in development and disease. J. Cell Sci. *121*, 737–746. https://doi.org/10.1242/jcs.026096.

51. Cruciat, C.M., and Niehrs, C. (2013). Secreted and transmembrane wnt inhibitors and activators. Cold Spring Harb. Perspect. Biol. *5*, a015081. https://doi.org/10.1101/cshperspect.a015081.

52. Mastri, M., Shah, Z., Hsieh, K., Wang, X., Wooldridge, B., Martin, S., Suzuki, G., and Lee, T. (2014). Secreted Frizzled-related protein 2 as a target in antifibrotic therapeutic intervention. Am. J. Physiol. Cell Physiol. *306*, C531–C539. https://doi.org/10.1152/ajpcell.00238.2013.

53. van Loon, K., Huijbers, E.J.M., and Griffioen, A.W. (2021). Secreted frizzled-related protein 2: a key player in noncanonical Wnt signaling and tumor angiogenesis. Cancer Metastasis Rev. *40*, 191–203. https://doi.org/10.1007/s10555-020-09941-3.

54. Puram, S.V., Tirosh, I., Parikh, A.S., Patel, A.P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C.L., Mroz, E.A., Emerick, K.S., et al. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell *171*, 1611–1624.e24. https://doi.org/10.1016/j.cell.2017.10.044.

55. Kaur, A., Webster, M.R., Marchbank, K., Behera, R., Ndoye, A., Kugel, C.H., 3rd, Dang, V.M., Appleton, J., O'Connell, M.P., Cheng, P., et al. (2016). sFRP2 in the aged microenvironment drives melanoma metastasis and therapy resistance. Nature *532*, 250–254. https://doi.org/10.1038/nature17392.

56. Vincent, K.M., and Postovit, L.M. (2017). A pan-cancer analysis of secreted Frizzled-related proteins: re-examining their proposed tumour suppressive function. Sci. Rep. *7*, 42719. https://doi.org/10.1038/srep42719.

57. Lee, J.L., Lin, C.T., Chueh, L.L., and Chang, C.J. (2004). Autocrine/paracrine secreted Frizzled-related protein 2 induces cellular resistance to apoptosis: a possible mechanism of mammary tumorigenesis. J. Biol. Chem. *279*, 14602–14609. https://doi.org/10.1074/jbc.M309008200.

58. Birnbaum, D.J., Finetti, P., Birnbaum, D., Mamessier, E., and Bertucci, F. (2017). Validation and comparison of the molecular classifications of pancreatic carcinomas. Mol. Cancer *16*, 168. https://doi.org/10.1186/s12943-017-0739-z.

59. Lautizi, M., Baumbach, J., Weichert, W., Steiger, K., List, M., Pfarr, N., and Kacprowski, T. (2022). The limits of molecular signatures for pancreatic ductal adenocarcinoma subtyping. NAR Cancer *4*, zcac030. https://doi.org/10.1093/narcan/zcac030.

60. Hu, M., Krause, D., Greaves, M., Sharkis, S., Dexter, M., Heyworth, C., and Enver, T. (1997). Multilineage gene expression precedes commitment in the hemopoietic system. Genes Dev. *11*, 774–785. https://doi.org/10.1101/gad.11.6.774.

61. Delorme, B., Ringe, J., Pontikoglou, C., Gaillard, J., Langonné, A., Sensebé, L., Noël, D., Jorgensen, C., Häupl, T., and Charbord, P. (2009). Specific lineage-priming of bone marrow mesenchymal stem cells provides the molecular framework for their plasticity. Stem Cell. *27*, 1142–1151. https://doi.org/10.1002/stem.34.

62. Brokhman, I., Xu, J., Coles, B.L.K., Razavi, R., Engert, S., Lickert, H., Babona-Pilipos, R., Morshead, C.M., Sibley, E., Chen, C., and van der Kooy, D. (2019). Dual embryonic origin of the mammalian enteric nervous system. Dev. Biol. *445*, 256–270. https://doi.org/10.1016/j.ydbio.2018.11.014.

63. Seaberg, R.M., Smukler, S.R., Kieffer, T.J., Enikolopov, G., Asghar, Z., Wheeler, M.B., Korbutt, G., and van der Kooy, D. (2004). Clonal identification of multipotent precursors from adult mouse pancreas that generate neural and pancreatic lineages. Nat. Biotechnol. *22*, 1115–1124. https://doi.org/10.1038/nbt1004.

64. Smukler, S.R., Arntfield, M.E., Razavi, R., Bikopoulos, G., Karpowicz, P., Seaberg, R., Dai, F., Lee, S., Ahrens, R., Fraser, P.E., et al. (2011). The adult mouse and human pancreas contain rare multipotent stem cells that express insulin. Cell Stem Cell *8*, 281–293. https://doi.org/10.1016/j.stem.2011.01.015.

65. Zhang, D., Park, D., Zhong, Y., Lu, Y., Rycaj, K., Gong, S., Chen, X., Liu, X., Chao, H.P., Whitney, P., et al. (2016). Stem cell and neurogenic gene-expression profiles link prostate basal cells to aggressive prostate cancer. Nat. Commun. *7*, 10798. https://doi.org/10.1038/ncomms10798.

66. Wang, H.T., Yao, Y.H., Li, B.G., Tang, Y., Chang, J.W., and Zhang, J. (2014). Neuroendocrine Prostate Cancer (NEPC) progressing from conventional prostatic adenocarcinoma: factors associated with time to development of NEPC and survival from NEPC diagnosis-a systematic review and pooled analysis. J. Clin. Oncol. *32*, 3383–3390. https://doi.org/10.1200/JCO.2013.54.3553.

67. Claudinot, S., Sakabe, J.I., Oshima, H., Gonneau, C., Mitsiadis, T., Littman, D., Bonfanti, P., Martens, G., Nicolas, M., Rochat, A., and Barrandon, Y. (2020). Tp63-expressing adult epithelial stem cells cross lineages boundaries revealing latent hairy skin competence. Nat. Commun. *11*, 5645. https://doi.org/10.1038/s41467-020-19485-3.

68. Tian, C., Clauser, K.R., Öhlund, D., Rickelt, S., Huang, Y., Gupta, M., Mani, D.R., Carr, S.A., Tuveson, D.A., and Hynes, R.O. (2019). Proteomic analyses of ECM during pancreatic ductal adenocarcinoma progression reveal different contributions by tumor and stromal cells. Proc. Natl. Acad. Sci. USA *116*, 19609–19618. https://doi.org/10.1073/pnas.1908626116.

69. Tian, C., Öhlund, D., Rickelt, S., Lidström, T., Huang, Y., Hao, L., Zhao, R.T., Franklin, O., Bhatia, S.N., Tuveson, D.A., and Hynes, R.O. (2020). Cancer Cell-Derived Matrisome Proteins Promote Metastasis in Pancreatic Ductal Adenocarcinoma. Cancer Res. *80*, 1461–1474. https://doi.org/10.1158/0008-5472.CAN-19-2578.

70. Swietlik, J.J., Bärthel, S., Falcomatà, C., Fink, D., Sinha, A., Cheng, J., Ebner, S., Landgraf, P., Dieterich, D.C., Daub, H., et al. (2023). Cell-selective proteomics segregates pancreatic cancer subtypes by extracellular proteins in tumors and circulation. Nat. Commun. *14*, 2642. https://doi.org/10.1038/s41467-023-38171-8.

71. Öhlund, D., Handly-Santana, A., Biffi, G., Elyada, E., Almeida, A.S., Ponz-Sarvise, M., Corbo, V., Oni, T.E., Hearn, S.A., Lee, E.J., et al. (2017). Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. J. Exp. Med. *214*, 579–596. https://doi.org/10.1084/jem.20162024.

72. Elyada, E., Bolisetty, M., Laise, P., Flynn, W.F., Courtois, E.T., Burkhart, R.A., Teinor, J.A., Belleau, P., Biffi, G., Lucito, M.S., et al. (2019). Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. Cancer Discov. *9*, 1102–1123. https://doi.org/10.1158/2159-8290.CD-19-0094.

73. Han, W., Chen, S., Yuan, W., Fan, Q., Tian, J., Wang, X., Chen, L., Zhang, X., Wei, W., Liu, R., et al. (2016). Oriented collagen fibers direct tumor cell intravasation. Proc. Natl. Acad. Sci. USA *113*, 11208–11213. https://doi.org/10.1073/pnas.1610347113.

74. Oudin, M.J., Jonas, O., Kosciuk, T., Broye, L.C., Guido, B.C., Wyckoff, J., Riquelme, D., Lamar, J.M., Asokan, S.B., Whittaker, C., et al. (2016). Tumor Cell-Driven Extracellular Matrix Remodeling Drives Haptotaxis during Metastatic Progression. Cancer Discov. *6*, 516–531. https://doi.org/10.1158/2159-8290.CD-15-1183.

75. He, B., Bergenstråhle, L., Stenbeck, L., Abid, A., Andersson, A., Borg, Å., Maaskola, J., Lundeberg, J., and Zou, J. (2020). Integrating spatial gene expression and breast tumour morphology via deep learning. Nat. Biomed. Eng. *4*, 827–834. https://doi.org/10.1038/s41551-020-0578-x.

76. Wang, Y., Kartasalo, K., Weitz, P., Ács, B., Valkonen, M., Larsson, C., Ruusuvuori, P., Hartman, J., and Rantalainen, M. (2021). Predicting Molecular Phenotypes from Histopathology Images: A Transcriptome-Wide Expression-Morphology Analysis in Breast Cancer. Cancer Res. *81*, 5115–5126. https://doi.org/10.1158/0008-5472.CAN-21-0482.

77. Buechler, M.B., Pradhan, R.N., Krishnamurty, A.T., Cox, C., Calviello, A.K., Wang, A.W., Yang, Y.A., Tam, L., Caothien, R., Roose-Girma, M., et al. (2021). Cross-tissue organization of the fibroblast lineage. Nature *593*, 575–579. https://doi.org/10.1038/s41586-021-03549-5.

78. Diaferia, G.R., Cirulli, V., and Biunno, I. (2013). SEL1L regulates adhesion, proliferation and secretion of insulin by affecting integrin signaling. PLoS One *8*, e79458. https://doi.org/10.1371/journal.pone.0079458.

79. Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome

profiling in single cells. Nat. Methods *10*, 1096–1098. https://doi.org/10.1038/nmeth.2639.

80. Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181. https://doi.org/10.1038/nprot.2014.006.
81. Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development *145*, dev165753. https://doi.org/10.1242/dev.165753.
82. Perkel, J.M. (2019). Starfish enterprise: finding RNA patterns in single cells. Nature *572*, 549–551. https://doi.org/10.1038/d41586-019-02477-9.
83. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36. https://doi.org/10.1186/gb-2013-14-4-r36.
84. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.
85. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527. https://doi.org/10.1038/nbt.3519.
86. Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573. https://doi.org/10.1093/bioinformatics/btq170.
87. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97. https://doi.org/10.1093/nar/gkw377.
88. Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, e21800. https://doi.org/10.1371/journal.pone.0021800.
89. Merritt, C.R., Ong, G.T., Church, S.E., Barker, K., Danaher, P., Geiss, G., Hoang, M., Jung, J., Liang, Y., McKay-Fleisch, J., et al. (2020). Multiplex digital spatial profiling of proteins and RNA in fixed tissue. Nat. Biotechnol. *38*, 586–599. https://doi.org/10.1038/s41587-020-0472-9.
90. Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). In Cell Detection with Star-Convex Polygons. held in Cham, 2018//. A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, eds. (Springer International Publishing)), pp. 265–273.
91. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
92. Kim, M.P., Evans, D.B., Wang, H., Abbruzzese, J.L., Fleming, J.B., and Gallick, G.E. (2009). Generation of orthotopic and heterotopic human pancreatic cancer xenografts in immunodeficient mice. Nat. Protoc. *4*, 1670–1680. https://doi.org/10.1038/nprot.2009.171.
93. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.
94. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: Open source software for digital pathology image analysis. Sci. Rep. *7*, 16878. https://doi.org/10.1038/s41598-017-17204-5.
95. Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. *4*, 1521. https://doi.org/10.12688/f1000research.7563.2.
96. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.
97. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
98. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics *31*, 2032–2034. https://doi.org/10.1093/bioinformatics/btv098.
99. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.
100. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219. https://doi.org/10.1038/nbt.2514.
101. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. *20*, 296. https://doi.org/10.1186/s13059-019-1874-1.
102. Tosti, L., Hang, Y., Debnath, O., Tiesmeyer, S., Trefzer, T., Steiger, K., Ten, F.W., Lukassen, S., Ballke, S., Kühl, A.A., et al. (2021). Single-Nucleus and In Situ RNA-Sequencing Reveal Cell Topographies in the Human Pancreas. Gastroenterology *160*, 1330–1344.e11. https://doi.org/10.1053/j.gastro.2020.11.010.
103. Muraro, M.J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning, E.J.P., and van Oudenaarden, A. (2016). A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst. *3*, 385–394.e3. https://doi.org/10.1016/j.cels.2016.09.002.
104. Peng, J., Sun, B.F., Chen, C.Y., Zhou, J.Y., Chen, Y.S., Chen, H., Liu, L., Huang, D., Jiang, J., Cui, G.S., et al. (2019). Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. Cell Res. *29*, 725–738. https://doi.org/10.1038/s41422-019-0195-y.
105. Dries, R., Zhu, Q., Dong, R., Eng, C.H.L., Li, H., Liu, K., Fu, Y., Zhao, T., Sarkar, A., Bao, F., et al. (2021). Giotto: a toolbox for integrative analysis and visualization of spatial expression data. Genome Biol. *22*, 78. https://doi.org/10.1186/s13059-021-02286-2.
106. Codeluppi, S., Borm, L.E., Zeisel, A., La Manno, G., van Lunteren, J.A., Svensson, C.I., and Linnarsson, S. (2018). Spatial organization of the somatosensory cortex revealed by osmFISH. Nat. Methods *15*, 932–935. https://doi.org/10.1038/s41592-018-0175-z.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-cytocheratin 19 (CK19)	Biocare Medical	Cat# CM242A; Lot 100711
Rabbit polyclonal anti-cytocheratin 19 (KRT19)	Novusbio	Cat# NBP1-78278; RRID:AB_11034002
Mouse monoclonal anti-pan-cytokeratin Alexa Fluor 488 conjugated	Novusbio	Cat# NBP2-33200AF488
Rabbit polyclonal anti-FLNB	Sigma-Aldrich	Cat# HPA004747; RRID:AB_1848599
Rabbit polyclonal anti-GATA6 (H-92)	Santa Cruz Biotech	Cat# sc-9055; RRID:AB_2108768
Rabbit polyclonal anti-HNF1B	Sigma-Aldrich	Cat# HPA002083; RRID:AB_1080232
Rabbit polyclonal anti-CEACAM6	Abcam	Cat# 3241-1; RRID: AB_2291545
Mouse monoclonal anti-CDH11 (clone 5B2H5)	ThermoFisher Scientific	Cat# 32-1700; Lot: UC283815; RRID: AB_2533068
Mouse monoclonal anti-cytokeratin HMW (34βE12)	Biocare Medical	Cat# CM 127 A; Lot:080118; RRID: AB_10580896
Mouse monoclonal anti-LAMB3 (clone 3353)	Sigma-Aldrich	Cat# AMAb91160; RRID: AB_2665825
Rabbit polyclonal anti-COL4A2	Atlas Antibodies	Cat# HPA069337
Mouse monoclonal anti-TP63	Dako	Cat# M7317 Clone DAK-p63
Mouse monoclonal anti-MUC5AC	Dako	Cat# M7316 Clone CLH2
Rabbit polyclonal anti-GABBR2	Sigma-Aldrich	Cat# HPA013820; Lot: R99228; RRID: AB_1849364
Mouse monoclonal anti-CDH1	Dako	Cat# M3612; RRID: AB_2076672
Rabbit polyclonal anti-PLAU	Sigma-Aldrich	Cat# HPA008719; RRID: AB_1858635
Goat polyclonal anti-COL4A2	Millipore	Cat# AB769; RRID: AB_92262
Mouse monoclonal anti-LAMA5	Sigma-Aldrich	Cat# AMAb91124; RRID: AB_2665809
Rabbit polyclonal anti-KCNMB3	Sigma-Aldrich	Cat# HPA015665; RRID: AB_1852142
Rabbit polyclonal anti-KCNJ5	Sigma-Aldrich	Cat# HPA017353; RRID: AB_1852135
Mouse monoclonal anti-SFRP2	OriGene	Cat# TA807387; RRID: AB_2628429
Mouse monoclonal anti-VINCULIN	Sigma-Aldrich	Cat# V9131; RRID: AB_477629
Rabbit polyclonal anti-GLUT-1	Sigma-Aldrich	Cat# HPA031345; RRID: AB_2673835
Chicken polyclonal anti-NFH	Millipore	Cat# AB5539; RRID: AB_11212161
Rabbit polyclonal anti-COL3A1	Sigma-Aldrich	Cat# HPA007583; RRID: AB_1078548

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Rabbit polyclonal anti-COL5A1	Sigma-Aldrich	Cat# HPA030769; RRID: AB_2673603
Rabbit polyclonal anti-FN1	Sigma-Aldrich	Cat# F3648; RRID: AB_476976
Biological samples		
Human PDAC FFPE samples	This paper; internal stock	N/A
Chemicals, peptides, and recombinant proteins		
1mg/mL proteinase K	Thermo Fisher Scientific	Cat# AM2546
Proteinase K inhibitor	Calbiochem	Cat# 539470
0.5U/ul RNase-free DNase I	Roche	Cat# 04716728001
Vector TrueVIEW Autofluorescence Quenching Kit	VectorLabs	#SP-8400
Recombinant human sFRP-2 protein	R&D Systems	Cat# 6838-FR-025/CF
Tn5	In-house produced	N/A
Matrigel	BD Biosciences	Cat# 356231
Collagen	Sigma-Aldrich	Cat# C8919
Critical commercial assays		
SMART-seq Stranded kit	Takara Bio USA, Inc	Cat # 634444, protocol version 051018
GeoMx-NGS Slide Preparation	Nanostring	Cat #MAN-10115-01
GeoMx DSP-Genomics Library Preparation	Nanostring	Cat #MAN-10117-01
Ion AmpliSeq™ Direct FFPE DNA Kit	ThermoFisher Scientific	MAN0014881
NEBNext Ultra II Fs DNA Library kit for Illumina	New England BioLabs	NEB #E7805
HCR RNA-FISH Bundle	Molecular Instruments	N/A
Matrigen Softwell 12 Well Plate Collagen	CELL guidance systems	Cat# SW12-COL-1-EA, SW12-COL-12-EA,
P1 Primary Cell 4D-Nucleofector X kit	Lonza Bioscience	Cat# V4XP-1012
Transwell Permeable supports	Costar	Cat# 3422
Deposited data		
RNAseq from tumor areas captured by LMD in PDAC patient samples	This paper	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE208732
DNAseq from tumor areas captured by LMD in PDAC patient samples	This paper	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE209951
RNAseq from tumor areas during PNI captured by LMD in PDAC patient samples	This paper	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE209951
RNAseq from a primary human PDAC cell line	This paper	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE209951
single-cell transcriptomic data from PDAC patient samples	Chan-Seng-Yue et al.,[4]	https://ega-archive.org/datasets/EGAD00010001811
single-cell transcriptomic data from PDAC patient samples	Hwang et al.,[5]	https://singlecell.broadinstitute.org/single_cell/study/SCP1089
PACA-CA clinical and transcriptomic data	ICGC portal	https://dcc.icgc.org
PACA-AU clinical and transcriptomic data	ICGC portal	https://dcc.icgc.org
TCGA clinical and transcriptomic data	GDC portal	https://portal.gdc.cancer.gov
Experimental models: Cell lines		
Primary human PDAC cell line MDA-PATC69	Michael Kim, MD Anderson Cancer Center, Houston, USA	N/A
Software and algorithms		
TopHat2	Kim et al.,[78]	http://ccb.jhu.edu/software/tophat/index.shtml
fgsea R package	Bioconductor	https://bioconductor.org/packages/release/bioc/html/fgsea.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GenomicRanges R package	Bioconductor	https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html
Kallisto	Bray et al.,[79]	https://pachterlab.github.io/kallisto/
DEseq2	Love et al.,[80]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Tximport R package	Soneson et al.,[81]	https://bioconductor.org/packages/release/bioc/html/tximport.html
ConsensusClusterPlus R package	Wilkerson and Hayes,[82]	https://bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html
igraph R package	CRAN	https://igraph.org/r/
EnrichR	Kuleshov et al.,[83]	https://maayanlab.cloud/Enrichr/
REVIGO	Supek et al.,[84]	http://revigo.irb.hr/
ClusterProfiler R package	Yu et al.,[85]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
Tidymodels R package	GitHub	https://github.com/tidymodels
caret R package	GitHub	https://topepo.github.io/caret
vip R package	GitHub	https://github.com/koalaverse/vip
Survminer R package	CRAN	https://rpkgs.datanovia.com/survminer
GeomxTools R package	Bioconductor	https://www.bioconductor.org/packages/release/bioc/html/GeomxTools.html
bwa mem	Li and Durbin[86]	https://github.com/lh3/bwa
Sambamba	Tarasov et al.,[87]	https://github.com/biod/sambamba
Broad Picard tools	GitHub	http://broadinstitute.github.io/picard
Broad GATK4	McKenna et al.,[88]	https://gatk.broadinstitute.org/hc/en-us
ANNOVAR	openbioinformatics	https://annovar.openbioinformatics.org/en/latest
Seurat R toolkit	Hao et al.,[89]	https://satijalab.org/seurat/
infercnv R package	GitHub	https://github.com/broadinstitute/inferCNV
CytoTRACE R package	Gulati et al.,[31]	https://cytotrace.stanford.edu
pvclust R package	GitHub	https://github.com/shimo-lab/pvclust
Starfish Python library	Perkel[90]	https://github.com/spacetx/starfish
Giotto R package	Dries et al.,[91]	https://giottosuite.readthedocs.io/en/master/
Fiji	GitHub	https://fiji.sc
Qupath	GitHub	https://qupath.github.io
R v4	The R Project	https://www.r-project.org/
Python v3.7	Python	https://www.python.org
Other		
LMD7 laser microdissection system	Leica Microsystems	RRID: SCR_020232
PEN membrane slides	Leica Microsystems	Cat# 11600289
Aperio ScanScope XTsystem	Leica Biosystems	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents may be directed to and will be fulfilled by the Lead Contact, Gioacchino Natoli (gioacchino.natoli@ieo.it).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- De-identified raw and processed datasets derived from human samples generated in this study were deposited in the Gene Expression Omnibus (GEO) database under the accession number GSE209952. All deposited data are publicly accessible as of the date of publication.

- This paper analyzes existing, publicly available data. The links and accession numbers for these datasets are listed in the key resources table.
- This paper does not report any original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human specimens

Human PDAC specimens were provided by the Humanitas Research Hospital IRCCS (Rozzano, Italy) with written consent for tissue donation and under a protocol approved by the institutional ethical committee. All patients had a premortem diagnosis of PDAC based on histopathological analysis on the resected material and they did not receive any neo-adjuvant therapy before surgery.

A summary of clinical features of our patient cohort is provided in Table S1.

Clinical data

Additional cohorts of primary treatment-naive PDAC patients were used in this study. Data referring to the Canadian (PACA-CA) and the Australian (PACA-AU) cohorts were downloaded from the ICGC database, while those enrolled in the TCGA study cohort were obtained from the GDC database. For single cell analyses, primary treatment-naive PDACs were obtained from Chan-Seng-Yue et al.[4] and Hwang et al.[5] (https://singlecell.broadinstitute.org/single_cell/study/SCP1089/).

Primary human PDAC cells

The MDA-PATC69 cell line (kindly gifted by Michael Kim, MD Anderson Cancer Center, Houston, USA), generated from a xenotransplanted human tumor fragment,[92] was cultured in Dulbecco's Modified Eagle Medium (Euroclone) supplemented with 10% North American foetal bovine serum (HyClone), 1% Pen/Strep and 2mM L-glutamine. The cell line was authenticated by the IEO Tissue Culture Facility using the GenePrint10 System (Promega) and was routinely screened for Mycoplasma contamination.

METHODS DETAILS

Histological review

Hematoxylin and Eosin (H&E)-stained slides cut from formalin-fixed paraffin-embedded (FFPE) sections were reviewed by a gastrointestinal pathologist (P.S.) on the basis of biomarker studies and cytological appearance. H&E slides were scanned using an Aperio ScanScope XTsystem (Leica Biosystems) at 20x magnification and morphological distinct areas were drawn on the digital sections. Annotated areas were matched and aligned to the consecutive sections mounted on specific slide for microdissection (see below).

Laser microdissection (LMD)

4 μm thick FFPE sections were mounted on polyethylene naphthalate membrane (PEN) slides (Leica cat. 11600289) previously UV-photoactivated in an UV crosslinker for 30 min (BLX-254, Bio-Link). FFPE sections were deparaffinized with two changes of xylene and then partially rehydrated in an ethanol gradient up to 75% EtOH.

Sections were then counterstained for 30 s with freshly prepared Cresyl Violet (0.8% Cresyl Violet in 60% EtOH and 4 mM Tris–HCl, pH 8.0), washed twice in 75% EtOH and air dried completely before proceeding to the microdissection. Single or multiple areas were drawn to collect 200-500 cells per sample using a UV-based LMD7 laser microdissection system (Leica Microsystems) at 20x magnification. Whenever possible, multiple samples of morphologically distinct areas were collected in each patient. Microdissected areas were collected by gravity into strip-tube caps and immediately processed for RNA or DNA library preparation.

LMD-seq library preparation

Microdissected areas were digested with 10 mU of Proteinase K (ThermoFisher, #AM2546) in 6 μl of digestion buffer (0.05% Triton X-100, 30 mM Tris-HCl pH 8.0) for 1 hour at 60°C. 1 ul of 5 mM Proteinase K inhibitor (Calbiochem, #539470) was added to stop the reaction and samples were stored at -80°C until library preparation.

Samples were then processed using the SMART-seq Stranded kit (Cat. 634444, protocol version 051018; Takara Bio USA, Inc.) according to the manufacturer's protocol with minor modifications. Briefly, the fragmentation step was omitted and RNA molecules were copied into first strand cDNA by reverse transcription using random primers and a template-switching oligo (TSO). Sequencing adapters and indexes were added to single-stranded cDNA with 5 cycles of PCR. cDNA originating from ribosomal RNA was depleted using scZapR in the presence of the mammalian-specific scR-Probes following manual instructions. The resulting ribo-depleted library fragments were amplified with 14 cycles of PCR, purified with AMPure beads and profiled for size distribution on an Agilent 2200 TapeStation with High Sensitivity D5000 reagent kits. Libraries were quantified by QuantiFluor assay run on a Glomax Explorer instrument (Promega) and sequenced on an Illumina NovaSeq 6000 platform to obtain 50 bp paired-end reads.

Gene panel mutational analysis

Areas selected for microdissection corresponded to sections consecutive to those used for RNA-seq. Microdissected areas were digested with 12 μl of Direct Reagent (Ion AmpliSeq™ Direct FFPE DNA Kit, MAN0014881, ThermoFisher Scientific) for 30 min at 65°C. The digested samples were subjected to DNA target enrichment by 15 cycles of PCR amplification using the Ion AmpliSeq Library Kit Plus (MAN0017003, ThermoFisher Scientific) and a custom pool of primers (GerSom). GerSom is a custom amplicon-based panel covering 467 genes selected among cancer-predisposing (n=172) and driver or actionable genes (n=295, of which 62 in common with the hereditary ones), for a total sequencing space of 2.2 Mb. Panel design details are being separately submitted (Frigè et al., in preparation) and are available upon request. After partial amplicon digestion and phosphorylation, the fragments were purified and used for adapter ligation and library amplification using the NEBNext Ultra II Fs DNA Library kit for Illumina (NEB # E7805, New England BioLabs) according to the manufacturer's instructions. Quality-checked and quantified libraries (see above) were pooled and sequenced (150 bp paired-end reads) on an Illumina NovaSeq 6000 platform.

Immunofluorescence

FFPE sections were deparaffinized, rehydrated and subjected to heat-induced antigen retrieval in EDTA buffer (1 mM EDTA pH 8.0, 0.05% Tween-20) using pressure cooker (TintoRetriever, BioSB). After blocking, sections were incubated overnight at 4°C with the following primary antibodies: anti-pan-Cytokeratin Alexa Fluor 488 conjugated (2.5 μg/ml, Novusbio #NBP2-33200AF488), anti-KRT10 (1:100, Biocare Medical, cytokeratin HMN #CM127A,C), anti-FLNB (0.5 μg/ml, Sigma-Aldrich, #HPA004747), anti-TP63 (1:75, Dako, #M7317), anti-HNF1B (0.4 μg/ml, Sigma-Aldrich, #HPA002083), anti-MUC5AC (1:50, Dako, #M7316), anti-GATA6 (10 μg/ml, Santa Cruz, #sc-9055X), anti-CDH1 (1:50, Dako, #M3612), anti-CDH11 (10 μg/ml, ThermoFisher Scientific, #32-1700), anti-GABBR2 (1 μg/ml, Sigma-Aldrich, #HPA031684), anti-PLAU (0.5 μg/ml, Sigma-Aldrich, #HPA008719), anti-LAMB3 (10 μg/ml, Sigma-Aldrich, #AMAb91160), anti-Collagen IV alpha 2 (3 μg/ml, Atlas Antibodies, #HPA069337), anti-LAMA5 (5 μg/ml, Sigma-Aldrich, #AMAb91124), anti-KCNMB3 (1 μg/ml, Sigma-Aldrich, #HPA015665), and anti-KCNJ5 (0.5 μg/ml, Sigma-Aldrich, #HPA017353), anti-NFH (1:4000, Millipore, #AB5539), and anti-GLUT-1 (1:200, Sigma-Aldrich, #HPA031345).

After washing, anti-rabbit or anti-chicken Rhodamine-Red-X conjugated and anti-mouse Alexa Fluor® 647 conjugated fluorophore-labeled F(ab)2 donkey secondary antibodies were used (Jackson Immunoresearch). Sections were DAPI counterstained and 3x3 Z-stacked large images (9-5 μm depth, 11-7 stacks) were acquired using a Yokogawa Spinning Disk Field Scanning Confocal System (CSU-W1, Nikon Europe BV, Amsterdam, Netherlands) equipped with 405, 488, 561, 640, 785 nm lines of solid-state lasers, 40x/1.15NA water immersion objective lens and a Prime BSI sCMOS camera (Teledyne Photometrics, Tucson, AZ).

For quantification of immunofluorescence intensities, raw images were first pre-processed with a maximum intensity projection along the z direction in each channel using Fiji.[93] For the single cell segmentation, images were imported into QuPath software (v 0.4.3)[94] and the nuclear region (DAPI channel) of each cell was segmented using the deep-learning algorithm Stardist,[90] and the cytoplasmic region was approximated as a band of 3 μm of thickness around the identified nuclei. For pan-CK staining, an object classifier in Qupath was trained, based on manual annotations of pan-CK positive and pan-CK negative cells, and used to identify only tumor cells. Fluorescence intensity measurements of the markers of interest were extracted for each segmented cell.

Two-photon second harmonic generation (SHG)

Hematoxylin and Eosin (H&E)-stained slides cut from formalin-fixed paraffin-embedded (FFPE) sections were used for a two-photon excitation (2PE) technique using a pulsed infrared laser (Chameleon Ultra II; Coherent) at 880 nm. Collagen SHG images were acquired by a confocal microscope (Leica; TCS SP5) on an upright microscope (DM6000 CFS) equipped with blue (argon, 488 nm), yellow (561 nm solid state laser), and red (633 nm solid state laser) excitation laser lines with a 20x water immersion objective lens (HCX Apochromat L 20x NA 1.00 W; Leica) and controlled by Leica LAS AF software (Leica). Image analysis was performed using a custom-made macro in Fiji.[93] Briefly, collagen fibers in SHG images were first pre-processed with a Median and a Top Hat filter and then identified using the Ridge Detection plugin in Fiji. For the quantification of fibers thickness, images were processed with an Enhance Local Contrast (CLAHE) function and an auto local threshold (method=Bernsen). Then the ''Local Thickness (masked, calibrated, silent)'' plugin (https://imagej.net/imagej-wiki-static/Local_Thickness) was executed.

For the perineurial invasion analysis, the distance of each collagen fiber from the nearest nerve was added to the measurements using the Spatial Analysis function ''Signed distance to annotations 2D'' in QuPath software.[94]

Cloning and plasmid construction

The sFRP2 sequence was cloned into the EcoRI sites of the pcDNA3.1 vector (Invitrogen) adding FLAG tag at the C-terminal of the protein using NEB Gibson Assembly Cloning Kit and the following primers:

Fw: 5'-CCGCCAGTGTGCTGGAATTCGCCACCATGCTGCAGGGCCCTG-3';

Rev: 5' GTGTGATGGATATCTGCAGAATTCTCACTTGTCATCGTCGTCCTTGTAGTCGCACTGCAGCTTGCGGA-3'. The construct was verified by Sanger sequencing.

Cell culture, transfection and treatments

For recombinant protein experiments, MDA-PATC69 cells were treated with 200ng/mL of recombinant human sFRP2 protein (R&D Systems, #6838-FR-025/CF) 24 hours after seeding.

For gene overexpression experiments, 1x10e6 MDA-PATC69 cells were transfected with 2μg of the sFRP2-Flag or the empty vector (pcDNA3.1, ThermoFisher Scientific) using the P1 Primary Cell 4D-Nucleofector X kit (Lonza Bioscience #V4XP-1012) and program EA-104 of the 4D-Nucleofector system (Lonza Bioscience) following manufacturer's instructions. After nucleofection, cells were grown for 48 h in culture medium before proceeding with further assays.

For mechano-response experiments, 5x10e4 cells were plated onto collagen-coated hydrogels with defined stiffness (1 and 12 kPa) bound to 12-well plates (Matrigen Softwell, CELL guidance systems) and harvested after 48 h.

Western Blot and immunoprecipitation

Cells were lysed in RIPA buffer (50mM Tris-HCl pH 8, 150mM NaCl, 1% NP-40, 0.1% Na-Deoxycholate, 0.1% SDS) containing protease inhibitors (Complete EDTA-free tablets, Roche) and 1mM PMSF. 10 μg of clarified cell extracts was resolved on SDS–polyacrylamide gel and blotted onto nitrocellulose membranes. After blocking, membranes were probed with an anti-sFRP2 (OriGene, TA807387, 1 μg/ml) or an anti-vinculin antibody (Sigma-Aldrich, V9131, 1:10000), washed and probed with the respective secondary HRP antibodies. Chemiluminescent HRP signal was revealed using Clarity ECL (Bio-Rad). Images were acquired using a Chemidoc imaging system (Bio-Rad).

2 ml of 0.22μm-filtered conditioned medium from PATC69 cells nucleofected with the empty vector or with the sFRP2-Flag vector were incubated with 5μg of anti-FLAG antibody (Sigma-Aldrich, F4799) pre-bound to 50μl of G protein-coupled paramagnetic beads (Dynabeads) overnight at 4°C (with 1mM PMSF, 1μg/ml aprotinin and 1μg/ml leupeptin). Beads were washed extensively with RIPA buffer and once with 1x TE 50mM NaCl. Immunoprecipitates were eluted in 2x Laemmli buffer, resolved by SDS–PAGE and immunoblotted with the anti-sFRP2 antibody.

Random migration and Boyden chamber assays

For random migration assays, 2x10e4 untreated or nucleofected PATC69 cells were left adhering for 24 h in a 12-well plate. Before starting time-lapse imaging, media was replaced and supplemented with Nuclight Rapid Red (1:2000, Incucyte) to live-stain nuclei and cells stimulated with recombinant sFRP2 protein. Bright field and 647 nm fluorescence images were acquired every 15 min for 72h with Leica DFC9000 GTC camera using the objective HC PL Fluotar 10×/0,30, mounted on a Leica Thunder imager system microscope equipped with an incubator chamber (OKOlab) maintained at 37°C in a 5% CO2 atmosphere. Tracking of migrating cells was performed using TrackMate plug-in (Fiji software) with LoG detector and Lack Tracker algorithms.

For Boyden chamber assays, 1x10e5 serum-starved nucleofected cells or cells treated with recombinant sFRP2 for 48h were seeded on top of 8mm-pore polycarbonate membranes (Costar) precoated with 20 μg Collagen (for migration assay, non gelling type I collagen, C8919, Sigma-Aldrich) or 35μg Matrigel (for invasion assay, growth factor reduced 356231, BD biosciences). Cells were allowed to migrate for 24 hours towards a serum gradient and fixed in 4% PFA. Nuclei were stained with DAPI and counted using the local maxima algorithm by Fiji software.

Cell-based ELISA

2x10e3 untreated or nucleofected MDA-PATC69 cells were left adhering for 24 h in a 96-well plate. Culture media was replaced and untreated cells stimulated with recombinant sFRP2 protein. Cells were cultured for 96h to allow ECM deposition and lysed with 0.075% ammonium hydroxide to decellularize the plate.[78] The wells were rinsed in PBS, blocked with 5% BSA and incubated overnight at 4°C with the following antibodies: anti-Collagen IV (0.4 μg/ml, Millipore, AB769), anti-COL3A1 (0.5 μg/ml, Sigma-Aldrich, #HPA007583), anti-COL5A1 (0.5 μg/ml, Sigma-Aldrich, #HPA030769), anti-LAMB3 (1 μg/ml, Sigma-Aldrich, #AMAb91160), anti-Fibronectin (1:10000, Sigma-Aldrich, F3648). After extensive washing, wells were incubated with the respective HRP-conjugated secondary antibodies and signal revealed with tetramethylbenzidine (TMB) substrate. The colorimetric reaction was allowed to develop for 1 hour at room temperature. Reaction was stopped with 0.6M sulfuric acid and 450nm absorbance measured at Glomax Explorer instrument (Promega). Signals were normalized to the DNA content assessed by QuantiFluor assay (Promega) from the cell lysate and empty wells coated with 5% BSA were used as background controls for each antibody.

RNA-seq library preparation

Total RNA was extracted from MDA-PATC69 cells cultured on substrates of different stiffness (see above) using the Zymo Quick-RNA kit (Zymo Research, R1055). RNA-seq was carried out using the SMART-seq2 protocol[80] with minor modifications. Briefly, 10 ng of total RNA were reverse transcribed with template-switching using oligo(dT) primers and an LNA-containing template-switching oligo (TSO). The resulting cDNA was pre-amplified, purified and tagmented with Tn5 transposase produced in-house using a described protocol.[79] cDNA fragments generated after tagmentation were gap-repaired, enriched by PCR and purified to create the final cDNA library.

Libraries were quantified by QuantiFluor assay run on a Glomax Explorer instrument (Promega) and sequenced on an Illumina NovaSeq 6000 platform to obtain 50 bp paired-end reads.

GeoMx digital spatial profiling (DSP)

5 mm glass-mounted FFPE sections were baked at 65°C for 1 h and manually prepared using the manufacturer supplied V2.0 protocol (MAN-10115-01).

Slides were hybridized to UV-photocleavable barcode-conjugated RNA probe set (Cancer Transcriptome Atlas Panel) overnight at 37°C using a Hyb EZ II hybridization oven (Advanced Cell Diagnostics). After stringent washes with 2X SSC and formamide at 37°C, slides were stained with fluorescently labeled morphology markers (Pan-cytokeratin, CD68 and Syto13) for 1 hour to distinguish tumor cells from the tumor microenvironment. Slides were loaded onto the GeoMx digital spatial profiling (DSP) instrument for imaging, tumor area selection, barcode cleavage and region of interest (ROI) collection (3–4 ROIs per PDAC patient). Library preparation was performed according to manufacturer instructions (Nanostring DSP-Genomics Library Preparation Protocol MAN-10117-01). Pooled libraries were sequenced on an Illumina NextSeq2000 platform (75 bp paired-end reads).

Multiplexed RNA smFISH

5 μm glass-mounted FFPE sections were baked at 65°C for 1 hour, deparaffinized, rehydrated and subjected to heat-induced antigen retrieval in TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0) using a steamer at 98°C for 20 min. After blocking in 5% BSA, a rabbit polyclonal antibody against KRT19 (10ug/ml, Novusbio #NBP1-78278) was incubated for 1h at room temperature and revealed with a Pacific Orange-conjugated secondary antibody (1:50, ThermoFisher, #P31584). After DAPI counterstaining, four representative heterogeneous tumor areas were acquired as Z-stacks using a Yokogawa Spinning Disk Field Scanning Confocal System (CSU-W1, Nikon Europe BV, Amsterdam, Netherlands) equipped with 405, 488, 561, 640, 785 nm lines of solid-state lasers and 60x/1.4NA oil immersion objective lens and a Prime BSI sCMOS camera (Teledyne photometrics, Tucson, AZ). After washing in PBS, samples were subjected to 8 rounds of hybridization with 4 probe sets at 4nM each according to the manufacturer's protocol (Molecular Instruments). After overnight hybridization and washes, probes were amplified using a mixture of metastable paired hairpins (HCR amplifiers)[81] conjugated with Alexa488, Alexa594, Alexa647 and Cy7 fluorophores at 60nM each in Amplification Buffer (Molecular Instruments) for 45 min at room temperature. After washing, background autofluorescence quenching (VectorLabs, #SP-8400) and DAPI counterstaining the same tumor areas were imaged as before.

DNA probes were digested with 0.5 U/ml RNase-free DNase I (Roche, #04716728001) for 1h at RT to perform a multi-day hybridization routine. After probe displacement, samples were first washed several times in Wash Buffer and then hybridized with a new probe set as described above.

For quantification of the RNA transcripts, the images acquired after each round of hybridization were processed using starfish.[82] To correct for microscope stage drift during the multiple rounds of imaging, raw images were first registered by taking a maximum intensity projection along the z direction in each channel and then were aligned using a sub-pixel image cross-correlation method. Each image was subjected to the background subtraction using white top-hat filter with a radius of 3 pixels and to the spot enhancement by filtering with a Laplacian of Gaussian (LoG) filter in order to increase the signal-to-noise.

RNA Transcripts were detected using local maxima finding method and distinguished from the background noise with an intensity threshold. The minimum intensity threshold was determined plotting dot intensities over threshold values and then choosing the closest value beyond which the number of detected dots showed an evident drop. After a manual cell segmentation using the DAPI and KRT19 images, RNA detected spots were assigned to the cells and were counted in each cell to generate single-cell spatial expression data.

QUANTIFICATION AND STATISTICAL ANALYSIS

RNA-seq data analysis

Paired-end reads with a minimum of 30 bp (after quality trimming and adapter clipping) were aligned to the human genome build GRCh38/hg38 using TopHat2[83] with parameter "–b2-very-sensitive". The RefSeq gene annotations was provided as the reference transcriptome. Only uniquely mapped reads were retained. At the gene level, the expression counts were estimated using summarizeOverlaps() function of GenomicRanges R package (https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html) with "IntersectionStrict" mode. Read counts were normalized using DEseq2's median of ratios[84] and then they were Log_2-transformed and used as proxy of expression for downstream analyses.

Gene set enrichment analysis (GSEA)

Gene Set Enrichment Analysis (GSEA) was performed using the fgsea R package (https://bioconductor.org/packages/release/bioc/html/fgsea.html) with the following parameters eps=0.0, minSize = 15 and maxSize = 2000. It was run using genes ranked based on the log_2 fold change[84] between stiff (12kPa) and soft (1kPa) groups. Biotype-specific signatures, Mechanoresponse[45] and ECM[68] programs were used as gene sets to compute the enrichments. Statistical significance was estimated by an FDR value ≤ 0.05.

LMD-seq data analysis

After the pre-processing steps (quality trimming and adapter clipping), the first 3 nt of the second read derived from the G-overhang of the template-switching oligonucleotide were discarded. Paired-end reads with a minimum of 20 bp were aligned to human genome build GRCh38/hg38 using Kallisto[85] with default parameters and in the correct orientation (–rf-stranded). The GENCODE version 33 gene annotations was provided as the reference transcriptome. Transcript-level expression was loaded using tximport R package[95] and read counts were normalized using DEseq2's median of ratios.[84] Normalized counts for protein-coding genes were Log_2-transformed and used as proxy of expression for downstream analyses. Genome-wide tracks were linearly scaled according the sample size factor.

Subtype identification using consensus clustering

The top 25% most variable genes (different iterations from top 30% to top 5% were also tested) in the LMD gene expression dataset were identified using rowVars R function and then used to cluster samples into different subtypes. Consensus clustering was performed using ConsensusClusterPlus R package[86] with Pearson's correlation distance, *k*-means clustering algorithm and 1,000 re-samplings (100% observation and 100% feature re-samplings). Considering the relative change in area under the cumulative distribution function (CDF) curve for consensus values for different clusters, the number of clusters was determined when there was no substantial increase in consensus.

Differential gene expression analysis

Differential gene expression analysis was performed in group-wise fashion across the different subtypes using the DEseq2 R package.[84] The patient ID was used as covariate in the model. p values for each tested gene were computed using the Wald test, and Benjamini-Hochberg correction was applied to correct for multiple hypotheses testing. The cutoff for selecting significantly differentially expressed genes was set to adjusted p value (or FDR) $\leq$ 0.01 and a linear fold change $\geq$ 2 (in either direction).

Unsupervised clustering using k-nearest neighbors

K-nearest neighbors (KNN) clustering was used to stratify genes based on their pattern of expression across samples. We first calculated the standardized gene expression considering all the differentially expressed genes. We then performed a Principal Component Analysis (PCA) to identify the top principal components (PCs) that explained the majority of the variance (more than 50%) in the data. We then built a KNN graph of selected genes (k = 115) based on the top 3 PCs and performed community detection on this neighborhood graph using the Louvain graph clustering method in igraph R package (https://igraph.org/r/) with default parameters.

Functional enrichment analysis and gene signature score

Functional enrichment analyses were performed using EnrichR tool.[87] We restricted the analysis to Biological Process categories and selected GO terms with enrichment of adjusted p value (FDR) $\leq$ 0.1. REVIGO[88] was used in tiny mode to summarize GO terms and reduce redundancy.

Transcriptional gene signatures reported previously[4–6,8,9] and biotype-specific signatures were used to calculate a score for all samples using Seurat's AddModuleScore function. Default parameters were used except for the number of control features that was set equal to the number of gene signature features. We calculated the standardized gene signature score and plotted as heatmap using the ComplexHeatmap R package.

Enrichment analyses of previously reported transcriptional single cell programs[5] in each biotype-specific gene signature (up-regulated DEGs from LMD-seq) were generated using ClusterProfiler's enricher function.[96] The adjusted p value (FDR) was plotted as heatmap using the ComplexHeatmap R package.

Supervised learning using Random Forest

A multi-class classifier was implemented to classify all the samples from the RNA-seq. Normalized gene expression matrix (as described above) was pre-processed by centering and scaling samples using the Tidymodels R package (https://github.com/tidymodels). The entire cohort was divided in two equally sized training and testing dataset using a stratified sampling method that preserved the same proportions of samples in each class as in the original data. The classifier was built using a Random Forest (RF) classification method setting a high number of trees (n = 500) and using the *k*-fold cross validation to avoid overfitting issues. Moreover, the number of features sampled for each tree (mtry) and the minimum number of data points in a node to allow further splitting (min_n) were tuned to optimize the hyperparameters of the model. Given the large amount of features used, the RFE function in caret R package was used as selection procedure to identify and remove the weakest features. The vip R package (https://github.com/koalaverse/vip) was then used with default parameters to measure the variable Importance (Gini index) as significant impact of features on the model.

To evaluate the model performance, the resulting model was used to predict samples from the test set (Table S5).

To infer the heterogeneity from the PACA-CA cohort as external dataset,[4] the predict function in the caret R package was used with the following parameter type=''prob''. The resulting model accurately predicts the classification of each patient based on the biotype scores. The dominant biotype was determined that yielded the highest biotype score, while biotype heterogeneity was inferred displaying biotype scores as the probabilities that a patient is classified as glandular, transitional and undifferentiated. To validate the overlap between biotype heterogeneity and morphological heterogeneity, probabilities of each different histological pattern from each patient of the same cohort were obtained from Kalimuthu et al..[11]

Prognostic model

A machine learning-based approach was implemented to develop a prognostic model and to extrapolate prognostic genes starting from the 457 predictor genes of the Random Forest-based model described above using clinical RNA-seq bulk data. Gene expression and clinical data of primary PDAC patients enrolled in the study cohorts, PACA-CA and PACA-AU, were downloaded from the ICGC database while those enrolled in the TCGA study cohort were obtained by the GDC database. Moreover, samples from the PACA-CA cohort that were used in Chan-Seng-Yue et al.[4] were used for this analysis (n=110 patients with clinical data out of the 118 of the initial cohort). Normalized gene expression matrix was standardized using the

Tidymodels R package (https://github.com/tidymodels). The univariate Cox proportional hazards regression was performed to identify significant genes potentially correlated with the survival of PDAC patients (Cox Univariate, $p \leq 0.05$). Using the genes selected above, the elastic-net (LASSO) regression against the overall survival was used to evaluate simultaneously the survival correlation of multiple features. The model was build using the *k*-fold cross validation to avoid overfitting issues. A stringent four-fold cross-validation on the initial cohort (including a feature selection step to identify the prognostic genes) was performed. After the establishing of the optimal multivariate model using the concordance index (C-index), genes with nonzero model coefficients were further selected to calculate the molecular risk score (combination of gene expression values weighted by the model coefficients) for each patient. A cut-off value was determined as the median risk score. Kaplan-Meier survival curves were generated using the Survminer R package (https://rpkgs.datanovia.com/survminer).

GeoMx digital spatial profiling data analysis

FASTQ files were trimmed, aligned and deduplicated using GeoMx NGS Pipeline one Illumina BaseSpace cloud environment as described previously.[89] The GeomxTools R package was used to perform pre-processing and to assess the quality of the samples and probes. Samples were retained according to a set of criteria based on the sequencing results (percent of mapping $\geq$80 and minimum number of reads per segment $\geq$ 1,000). Outlier probes were removed on a gene-wise basis to exclude those showing an average count across all segments accounting for less than 10% of the counts of all the probes for the gene. Moreover, a probe was also excluded according to the Grubb's test in at least 20% of the segments (if the counts of that probe was always higher or lower than the other probes for the same gene). Gene expression was calculated as the geometric mean of the remaining probes and read counts were normalized using DEseq2's median of ratios.[84] Standardized scores of transcriptional gene signatures of three biotypes (up-regulated DEGs from RNA-seq) were calculated for each sample using Seurat's AddModuleScore function. To cluster all samples in distinct groups, we used the ConsensusClusterPlus R package[86] with Pearson's correlation distance, *k*-means clustering algorithm and 1,000 re-samplings (100% observation and 100% feature re-samplings).

Gene panel mutational data analysis

After pre-processing (quality trimming and adapter clipping), paired-end reads with a minimum length of 20 bp were aligned to human genome (GRCh38/hg38 build) using bwa mem aligner[97] with default parameters. Then, Sambamba[98] was used to remove unmapped reads and PCR duplicates were marked by Picard Tools (http://broadinstitute.github.io/picard). The alignments were subjected to a recalibration of base quality scores using the BaseRecalibrator function (part of GATK4[99] providing known sites of variation (ExAc data). The resulting BAM files were used to estimate final metrics for each sample using the CollectHsMetrics function of the Picard tools and for the following mutation calling pipeline.

Mutation calling pipeline

Mutect2 (part of GATK4)[100] was run individually on each tumor sample for the single nucleotide variant calling without a matched normal sample as reference and using the following parameters:

- 1000g_pon.hg38.vcf.gz was downloaded from the Broad Institute Bundle and was used as normal reference (–panel-of-normals option);
- af-only-gnomad.hg38.vcf.gz was used as the source of germline variants with estimated allele frequency (–germline-resource option);
- targeted regions derived from Frigè et al. (ms in preparation) were specified (-L option).

The LearnReadOrientationModel function of GATK4 was then used to estimate the substitution errors occurring as a result of FFPE artefacts. The resulting orientation bias model was fed into the FilterMutectCalls function of GATK4. The resulting SNVs and INDELs marked as PASS were retained and were further filtered considering a minimum depth of 20 reads and a minimum of 5 reads on the alternate allele.

ANNOVAR (https://annovar.openbioinformatics.org/en/latest) was used to annotate the called variants using the NCBI Refseq release 200 gene annotations. Variants annotated in the gnomAD database with a minor allele frequency (MAF) greater than 0.05 were removed.

To ensure enough coverage in the different samples, only mutational sites covered with at least 20 reads in more than 75% of the samples were retained.

Single-cell RNAseq data analysis

Single-cell RNA-seq profiles derived from primary treatment-naive PDACs were mined from Chan-Seng-Yue et al.[4] and from Hwang et al.[5] (https://singlecell.broadinstitute.org/single_cell/study/SCP1089/). Only for the last dataset metadata with cell type annotations for each single cell were downloaded and used for downstream analysis. Count matrices were imported and further processed for downstream analysis, using the Seurat R toolkit.[91] Low quality cells (defined as showing either >25% reads mapping to mitochondrial genes or <1,000 genes expressed) were excluded. Data was then normalized using SCTransform.[101] Principal Component Analysis (PCA) was performed using the normalized expression of the top 2,000 highly variable genes. The top PCs (determined as the number of PCs whose percentage of variance increase among consecutive PCs is less than 0.1%) were then used to build a *k*-nearest

neighbors' graph (k=20) of single-cell profiles, and to perform community detection using the Leiden clustering method (resolution = 0.8), for the identification of distinct cell clusters.

Different cell populations were identified using known cell type-specific gene markers from fibroblasts, immune cells and endocrine/exocrine pancreatic cells.[102,103] To confirm the presence of clusters enriched for cancer cells, somatic CNVs were inferred from the single-cell gene expression profiles using the InferCNV R package (https://github.com/broadinstitute/inferCNV). This step was run on all malignant cells using a set of high confidence non-neoplastic cells (identified in the same data) as reference. InferCNV was run with parameters –cutoff 0.1 and –noise_filter 0.2.

Single cell classification

Cancer cells from individual patients were scored using the gene signatures of three biotypes we previously identified (up-regulated DEGs from LMD-seq). Seurat's AddModuleScore function was used as described above to estimate the expression of each gene signature at the single-cell level, properly accounting for the average expression of the genes in each signature, by using a background set of randomly selected genes with comparable expression levels. We then assigned a probability of observing each specific score by chance, given a signature and a single cell, by performing a Permutation test. Briefly, we randomly re-shuffled the gene expression matrix without replacement 100 times, and each time we computed the scores for each signature, for each single-cell. Given a real score, an empirical p value was then associated to it as the number of times a score equal or higher than the real score was observed in the those generated from the permuted data. Each single cell was then classified as a specific biotype, if the p value of the score for the associated signature was ≤ 0.05.

CytoTRACE

CytoTRACE[31] was used to predict the developmental potential of single cells ranking them from less differentiated to more differentiated on the basis of their number of expressed genes. After calculating the total number of detected genes for each single cell, CytoTRACE correlates this information with gene expression to establish a gene count signature (GCS). Subsequently, it normalizes this signature on a scale from 0 (indicating more differentiated cells) to 1 (indicating less differentiated cells).

CNV estimation

To identify CNV patterns specific to each biotype, CNVs for each single cell were estimated using the inferCNV R package (https://github.com/broadinstitute/inferCNV) and a set of high confidence non-neoplastic cells as reference, as described above. We then computed the mean of CNVs values across all cells of the same biotype, separately for each patient. We then performed hierarchical clustering using Euclidean distance and the Ward.d2 method. The pvclust R package (https://github.com/shimo-lab/pvclust) was then used to assess the cluster stability via bootstrapping (1,000 boostraps). The CNV scoring method[104] was used to quantify the overall level of CNVs in each single cell. To this end, CNVs values estimated by inferCNV were first re-standardized to a mean of zero, and then scaled from -1 to 1. In order to account for both deletions and duplications, the sum of squared values was calculated as the overall CNV score for each cell.

Multiplexed RNA smFISH data analysis

Single-cell expression matrix and spatial coordinates were imported into the Giotto[105] R package and were further processed for downstream analysis. Low quality cells ($\leq$5 genes detected per cell and at least 1 molecule as expression threshold) were excluded. The data were normalized per gene and per cell using osmFISH.[106] Z-scored gene expression was used for the Principal Component Analysis (PCA) and for the finding of the top 20 PCs. Cell cluster identification was performed by k-nearest neighbors graph (k=30) and community detection using the Leiden graph clustering method (res = 0.7). UMAP manifold was used for visualization of the cluster groups. Markers from each cluster were identified according to the findMarkers_one_vs_all function, intersected with the genes up-regulated in LMD-seq and then used to assign cell cluster to the three biotypes. To infer tumor areas based on the spatial coordinates of the identified cell types, single cells belonging to the identified cluster groups were mapped back to their original histological space. Moreover, gene expression of selected genes was plotted by spatFeatPlot2D function.

Statistical analysis

Unless indicated otherwise, statistical analyses were performed in the statistical computing environment using R v4.