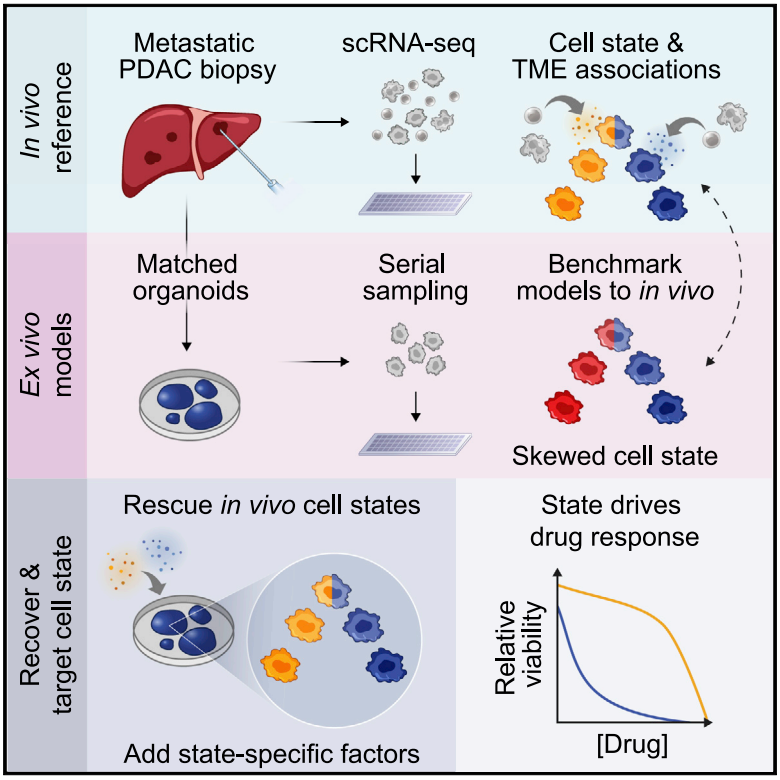
Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer

Cell

Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer

Graphical abstract



Highlights

- scRNA-seq of metastatic pancreatic cancer and matched organoid models
- *Ex vivo* to *in vivo* comparisons reveal loss of cell state heterogeneity in models
- Cell state is shaped by the microenvironment *in vivo* and can be controlled *ex vivo*
- Cell state drives drug response

Authors

Srivatsan Raghavan, Peter S. Winter, Andrew W. Navia, ..., William C. Hahn, Andrew J. Aguirre, Alex K. Shalek

Correspondence

pswinter@mit.edu (P.S.W.), andrew_aguirre@dfci.harvard.edu (A.J.A.), shalek@mit.edu (A.K.S.)

In brief

Systematic profiling of metastatic pancreatic cancer biopsies and matched organoid models provides a view of cellular states, their regulation by the tumor microenvironment, and the ability to modulate these states to impact drug responses.

Raghavan et al., 2021, Cell *184*, 6119–6137
December 9, 2021
https://doi.org/10.1016/j.cell.2021.11.017

CellPress

Cell

CellPress
OPEN ACCESS

Article

Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer

Srivatsan Raghavan,[1,2,3,4,20] Peter S. Winter,[1,2,5,6,7,20,*] Andrew W. Navia,[1,2,5,6,7,20] Hannah L. Williams,[1,3,20] Alan DenAdel,[8,9] Kristen E. Lowder,[1,2] Jennyfer Galvez-Reyes,[2,5,6] Radha L. Kalekar,[1,2] Nolawit Mulugeta,[2,5,6] Kevin S. Kapner,[1,2] Manisha S. Raghavan,[1,2] Ashir A. Borah,[2] Nuo Liu,[2,5,6] Sara A. Väyrynen,[1,3] Andressa Dias Costa,[1,3] Raymond W.S. Ng,[1,2] Junning Wang,[1] Emma K. Hill,[1] Dorisanne Y. Ragon,[1] Lauren K. Brais,[1] Alex M. Jaeger,[6] Liam F. Spurr,[1,2] Yvonne Y. Li,[1,2] Andrew D. Cherniack,[1,2,3] Matthew A. Booker,[10] Elizabeth F. Cohen,[10] Michael Y. Tolstorukov,[10] Isaac Wakiro,[1] Asaf Rotem,[1,2,11] Bruce E. Johnson,[1,3,4,11] James M. McFarland,[2]

(Author list continued on next page)

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[2]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[3]Harvard Medical School, Boston, MA 02115, USA
[4]Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA
[5]Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[6]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[7]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[8]Center for Computational Molecular Biology, Brown University, Providence, RI 02912, USA
[9]Division of Applied Mathematics, Brown University, Providence, RI 02912, USA
[10]Department of Informatics and Analytics, Dana-Farber Cancer Institute, Boston, MA 02215, USA

(Affiliations continued on next page)

SUMMARY

Prognostically relevant RNA expression states exist in pancreatic ductal adenocarcinoma (PDAC), but our understanding of their drivers, stability, and relationship to therapeutic response is limited. To examine these attributes systematically, we profiled metastatic biopsies and matched organoid models at single-cell resolution. *In vivo*, we identify a new intermediate PDAC transcriptional cell state and uncover distinct site- and state-specific tumor microenvironments (TMEs). Benchmarking models against this reference map, we reveal strong culture-specific biases in cancer cell transcriptional state representation driven by altered TME signals. We restore expression state heterogeneity by adding back *in vivo*-relevant factors and show plasticity in culture models. Further, we prove that non-genetic modulation of cell state can strongly influence drug responses, uncovering state-specific vulnerabilities. This work provides a broadly applicable framework for aligning cell states across *in vivo* and *ex vivo* settings, identifying drivers of transcriptional plasticity and manipulating cell state to target associated vulnerabilities.

INTRODUCTION

Recent advances in high-throughput genomic sequencing have led to a detailed understanding of the genetic alterations that underlie human tumors (Garraway and Lander, 2013). These reference maps have driven a "mutation-centric" view of cancer that informs our current approach to precision medicine. In this framework, DNA alterations are used as biomarkers to guide therapy selection (Hyman et al., 2017), and *ex vivo* models are used to validate mutational associations and power therapeutic discovery efforts. To maintain translational relevance, the fidelity of *ex vivo* models to *in vivo* attributes is paramount.

Driven by the mutation-centric view of cancer, model fidelity is typically assessed via genomic similarity (Ben-David et al., 2019; Byrne et al., 2017; Drost and Clevers, 2018; Gillet et al., 2013). Assessment of mutational fidelity is feasible and useful for two reasons: first, bulk measurements of DNA alterations have favorable signal-to-noise profiles (i.e., somatic mutations are present or absent); and, second, some cancers respond to therapies in a genotype-directed manner (Garraway and Lander, 2013; Hyman et al., 2017). However, a growing body of evidence indicates that using mutations alone to assign therapies has limitations (Nam et al., 2021; van de Haar et al., 2021). The advent of single-cell genomic technologies has confirmed extensive mutational heterogeneity in human tumors but also revealed that the complexity of cancer extends to variation in cell transcriptional state. The relationship between cell state and therapeutic sensitivity represents a new but poorly understood opportunity for cancer therapeutic development (Hahn et al., 2021).

Ewa T. Sicinska,[3,12] Tyler E. Jacks,[6] Ryan J. Sullivan,[3,13] Geoffrey I. Shapiro,[1,3,4] Thomas E. Clancy,[1,3,14] Kimberly Perez,[1,3,4] Douglas A. Rubinson,[1,3,4] Kimmie Ng,[1,3,4] James M. Cleary,[1,3,4] Lorin Crawford,[8,15,16] Scott R. Manalis,[6,17] Jonathan A. Nowak,[3,11,18] Brian M. Wolpin,[1,3,4,21,22] William C. Hahn,[1,2,3,4,21,22] Andrew J. Aguirre,[1,2,3,4,21,22,*] and Alex K. Shalek[2,3,5,6,7,19,21,22,23,*]

[11]Center for Cancer Genomics, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[12]Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[13]Massachusetts General Hospital Cancer Center, Boston, MA 02114, USA
[14]Department of Surgery, Brigham and Women's Hospital, Boston, MA 02115, USA
[15]Department of Biostatistics, Brown University, Providence, RI 02912, USA
[16]Microsoft Research New England, Cambridge, MA 02142, USA
[17]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
[18]Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA
[19]Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA 02139, USA
[20]These authors contributed equally
[21]These authors contributed equally
[22]Senior author
[23]Lead contact
*Correspondence: pswinter@mit.edu (P.S.W.), andrew_aguirre@dfci.harvard.edu (A.J.A.), shalek@mit.edu (A.K.S.)
https://doi.org/10.1016/j.cell.2021.11.017

Cell state, as measured by RNA expression, is a complex representation of tumor phenotype because it integrates inputs from cell-intrinsic (e.g., mutational background, epigenetic state) and cell-extrinsic (e.g., cell-to-cell interactions, tissue architecture) sources. Although the field has generated high-resolution single-cell RNA sequencing (scRNA-seq) maps of cancer cell transcriptional states across diverse contexts (Filbin et al., 2018; Hovestadt et al., 2019; Kim et al., 2018; Neftel et al., 2019; Patel et al., 2014; Puram et al., 2017; Sade-Feldman et al., 2019; Suvà and Tirosh, 2019; Tirosh et al., 2016a; Tirosh et al., 2016b; van Galen et al., 2019; Venteicher et al., 2017), we have not mapped their stability or the relative influences of cell-intrinsic and cell-extrinsic factors in specifying them. Moreover, we have a limited understanding of the degree to which models accurately recapitulate the distribution of cancer cell states seen in patients.

Despite extensive genomic characterization of pancreatic ductal adenocarcinoma (PDAC), most cancers do not harbor therapeutically tractable alterations (Aguirre et al., 2018; Bailey et al., 2016; Cancer Genome Atlas Research Network, 2017). However, RNA subtypes (states) derived from bulk measurements have emerged as an important clinical biomarker (Aguirre et al., 2018; Aung et al., 2018; Bailey et al., 2016; Cancer Genome Atlas Research Network, 2017; Chan-Seng-Yue et al., 2020; Collisson et al., 2019; Collisson et al., 2011; Connor et al., 2019; Hayashi et al., 2020; Moffitt et al., 2015; O'Kane et al., 2020; Porter et al., 2019; Tiriac et al., 2018). While PDAC subtyping studies have largely recovered common expression features, it remains unclear whether these bulk RNA-seq state measurements mask heterogeneity at the single-cell level, which features derive from malignant versus non-malignant cells, and how well they are recapitulated in laboratory models. Moreover, the PDAC tumor microenvironment (TME) includes numerous non-malignant immune and stromal cell types (Balachandran et al., 2019; Bernard et al., 2019; Elyada et al., 2019; Grünwald et al., 2021; Ligorio et al., 2019), but their variation across different sites of disease and their effects on malignant cell state and therapeutic response is not well characterized. Given the lack of mutational biomarkers for PDAC, understanding how cell state is shaped by the local TME and whether cell state can be used as a tractable biomarker for therapy selection remains of critical importance.

To parse the instructive roles of cell-intrinsic and cell-extrinsic inputs to cancer cell state, we developed and employed an optimized translational workflow to perform both high-resolution profiling of metastatic PDAC patient tissue using scRNA-seq (Gierahn et al., 2017; Hughes et al., 2020) and derivation of matched organoid models (Boj et al., 2015; Tiriac et al., 2018) from the same core needle biopsy. Using this approach, we generated a single-cell map of metastatic PDAC and used it as a reference to benchmark cell states in matched organoid models. We identify a new intermediate co-expressor (IC) PDAC cell state, uncover distinct site- and subtype-specific TMEs, and demonstrate that TME signals are critical regulators of cancer cell state, plasticity, and response to therapy.

RESULTS

Model systems retain genetic fidelity but lose expression state heterogeneity

Current precision medicine pipelines focus on preserving mutational fidelity in cancer models; however, it is unclear how well these same models represent prognostic RNA states (Figure 1A). We compared bulk DNA and RNA-sequencing profiles of primary and metastatic patient tumors with established cell lines and a cohort of newly generated patient-derived organoids to understand how each model system represents the distribution of mutational and RNA phenotypes seen in patients. We observed no significant difference in driver oncogene alteration frequencies among the groups, suggesting that model systems are relatively faithful genetic representations of patient cohorts (Fisher's exact test; Figure 1B, far right). Next, we assessed PDAC subtypes derived from several recent bulk RNA-sequencing studies (Figure 1C) (Bailey et al., 2016; Chan-Seng-Yue et al., 2020; Collisson et al., 2011; Moffitt et al., 2015). Subsets of signatures from each study were highly concordant and separated primarily into classical-like (clade 1) and basal-like (clade 2) groups. Clade 3 signatures were generally lower in expression but tended to associate with clade 1, while clade 4

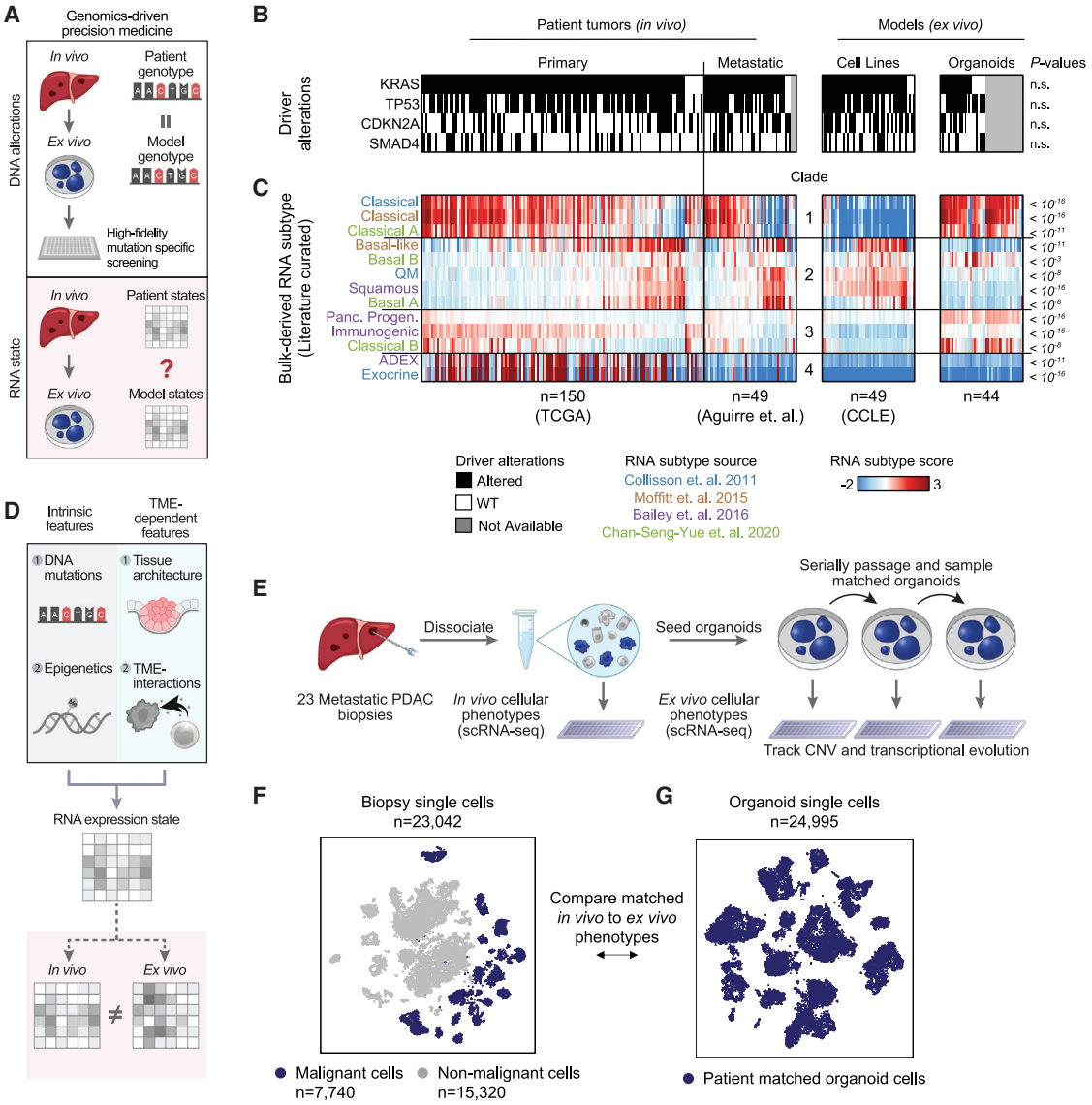


Figure 1. Assessing transcriptional states in patient tumors and cancer models

(A) Precision medicine pipelines assess model fidelity for genetics but typically do not evaluate RNA states.

(B) Alterations in PDAC driver genes across primary resections (TCGA), metastatic biopsies (Panc-Seq), and organoid and cell line (CCLE) models. Grey indicates where genomic data were not available. *P*-values by Fisher's exact test.

(C) Comparison of PDAC expression signatures from bulk RNA-sequencing in primary and metastatic tumors, cell lines, and organoid models in (B). Rows are clustered, columns are sorted by average basal-classical score difference. *P*-values indicate differences between patient tumors, cell lines, and organoids by ANOVA.

(D) Schematic of contributors to RNA state that may lead to differences between *in vivo* and *ex vivo* expression patterns.

(E) Metastatic patient samples were collected via core needle biopsies and dissociated. Biopsy cells were allocated for scRNA-seq, and patient-matched organoids were developed with downstream serial scRNA-seq sampling.

(F and G) *t*-distributed stochastic neighbor embedding (*t*-SNE) for biopsy (F) and matched patient-derived organoid cells (G).

See also Figure S1; Tables S1 and S2.

signatures represented exocrine pancreas expression patterns and tended to be expressed only in primary disease, suggesting that they may be due to contributions from the primary PDAC TME. Relatedly, the documented low malignant cellularity in PDAC may obstruct malignant cell-specific signature identification in bulk RNA-sequencing datasets (Collisson et al., 2019).

In contrast to the mutational data, we observed significant differences for all RNA signatures when comparing tumors to model types (Figure 1C). Whereas primary and metastatic samples include tumors with both classical-like (clade 1) and basal-like (clade 2) signatures, cell lines exhibited predominantly basal-like (clade 2) subtypes while organoids were nearly entirely

classical-like (clade 1) with partially overlapping expression of some clade 3 signatures (Figures 1C and S1A). These observations demonstrate that despite preservation of genomic alterations, neither PDAC cell lines nor organoids represent the full repertoire of expression subtypes seen in patient cohorts.

Single-cell profiling of metastatic PDAC and matched organoid models

These findings highlight a critical need for new approaches to identify the determinants of cancer cell state both *in vivo* and in model systems. Cell state is a complex attribute since it integrates cell-intrinsic as well as TME-dependent features; therefore, multiple mechanisms (e.g., clonal selection and/or plasticity) may contribute to the divergence between *in vivo* and *ex vivo* expression patterns (Figure 1D). We hypothesized that a dataset at single-cell resolution allowing for matched comparisons of *in vivo* and *ex vivo* attributes would enable a better understanding of cell state drivers, stability, and functional relevance. To this end, we established a pipeline to generate matched scRNA-seq profiles and organoid models using core needle biopsies from patients with metastatic PDAC (n = 23) (Figures 1E and S1B; Table S1). Most samples were obtained from metastatic lesions residing in the liver (19/23), and the majority (21/23) were analyzed by targeted DNA-sequencing, yielding the expected mutational patterns (Figure S1B).

Our pipeline generated approximately 1,000 high-quality single cells per biopsy and successful early-passage organoid cultures from 70% of patient samples (16/23 samples reaching at least passage 2) (Figures 1E, 1F, and S1B–S1D). Consistent with other studies, we observed patient-specific and admixed clusters of single cells suggesting the presence of both malignant and non-malignant cells in each biopsy (Figures S1E and S1F; STAR Methods) (Kim et al., 2018; Puram et al., 2017; Sade-Feldman et al., 2019; Tirosh et al., 2016a). Inferred copy number variation (CNV) alteration scores separated putative cancerous and non-cancerous cells in each biopsy and demonstrated high concordance with reference targeted DNA-seq (Figures S1G and S1H) (Patel et al., 2014; Tirosh et al., 2016b). CNV analysis and manual inspection of expression patterns for known markers across single cells supported the identification of malignant cells as well as 11 unique non-malignant cell types (Figures S1I–S1K; Table S2). Thus, despite the documented low malignant cellularity in PDAC (Aguirre et al., 2018; Cancer Genome Atlas Research Network, 2017; Chan-Seng-Yue et al., 2020), we established a robust pipeline that retrieved high-quality malignant (n = 7,740) and non-malignant (n = 15,302) single-cell transcriptomes from metastatic PDAC needle biopsies, as well as those of matched organoids (n = 24,995) (Figures 1E–1G).

Single-cell resolution identifies an IC cancer cell state in metastatic PDAC

After separating the CNV-confirmed PDAC cells from non-malignant cells (excluding one neuroendocrine sample; Figure S2A; see STAR Methods), we first interrogated whether previously described RNA subtypes (Figure 1C) are represented in single metastatic PDAC cells. Subsets of metastatic cells scored highly for either classical-like or basal-like signatures derived from independent bulk studies indicating the general relevance of these expression programs to *in vivo* PDAC biology (Figure 2A). The remaining literature-derived signatures had low expression in these metastatic cells, suggesting either specificity for primary disease or influence from non-malignant expression in bulk RNA profiles. Interestingly, a large subset of cells showed low expression for all previously proposed signatures, highlighting a gap in our understanding of PDAC expression states (Figure 2A; "Low for bulk subtypes").

Given this large subset of cells that score weakly for previously identified signatures, we sought to understand the spectrum of *in vivo* expression states through an unbiased analysis of our single-cell dataset. Principal component analysis (PCA) on all malignant cells revealed that genes enriched for signatures of epithelial/mesenchymal transition (EMT) (PC1) (Gröger et al., 2012), basal-like and classical PDAC states (PC1 and PC2), and cell cycle (PC3 and PC8) (Tirosh et al., 2016a) drove the major axes of variation in the dataset (Figures S2B and S2C). The PC1-PC2 score difference was correlated with the basal-classical signature axis previously identified in bulk studies ($r = 0.8$). We identified 1,909 genes significantly correlated with either end of the basal or classical-enriched continuum within our single-cell cohort (Figure S2D; Table S3; STAR Methods). Inspection of these genes revealed that more basal-like cells are associated with transforming growth factor beta (TGF-β) signaling, WNT signaling, EMT, and cell cycle progression. Epithelial and pancreatic lineage programs are enriched in the cells with more classical-like attributes (Figures S2E and S2F). We term the signatures derived from our single-cell cohort as single-cell basal (scBasal) and single-cell classical (scClassical) (Figure 2B).

Single-cell analysis enabled a dissection of expression states that are confounded by contaminating non-malignant cells in bulk measurements. For example, in bulk RNA-seq studies (Aguirre et al., 2018; Cancer Genome Atlas Research Network, 2017), EMT programs are strongly correlated with markers of fibroblasts ($r = 0.9$), inversely related to purity metrics ($r = -0.5$), and poorly correlated with basal-like expression ($r = -0.08$). At single-cell resolution, we observed a subset of cells that express both scBasal and EMT programs, while other scBasal cells had low expression of EMT programs (Figure 2C, left). Similarly, the origin of cytokine response signatures can be difficult to interpret from bulk studies, as interferon (IFN) response gene signatures are positively correlated with markers of several cell types including macrophages ($r = 0.6$) and T cells ($r = 0.4$), negatively associated with purity metrics ($r = -0.4$), and not associated with either basal-like or classical scores ($r = 0.01$ and -0.04, respectively). In single cells, we observe clear patterns of association between IFN response and a subset of cells with intermediate scBasal expression (Figure 2C, right). These findings suggest not all basal-like cells have the same underlying attributes and highlight the importance of utilizing scRNA-seq to understand how malignant cells sense and respond to their local TME.

Although some studies suggest that basal and classical programs exist only in discrete cell populations (Chan-Seng-Yue et al., 2020), we observed that the scBasal and scClassical programs were not mutually exclusive. Rather, we identified cells that are intermediate for scBasal and scClassical gene expression signatures, co-expressed features of both programs to varying

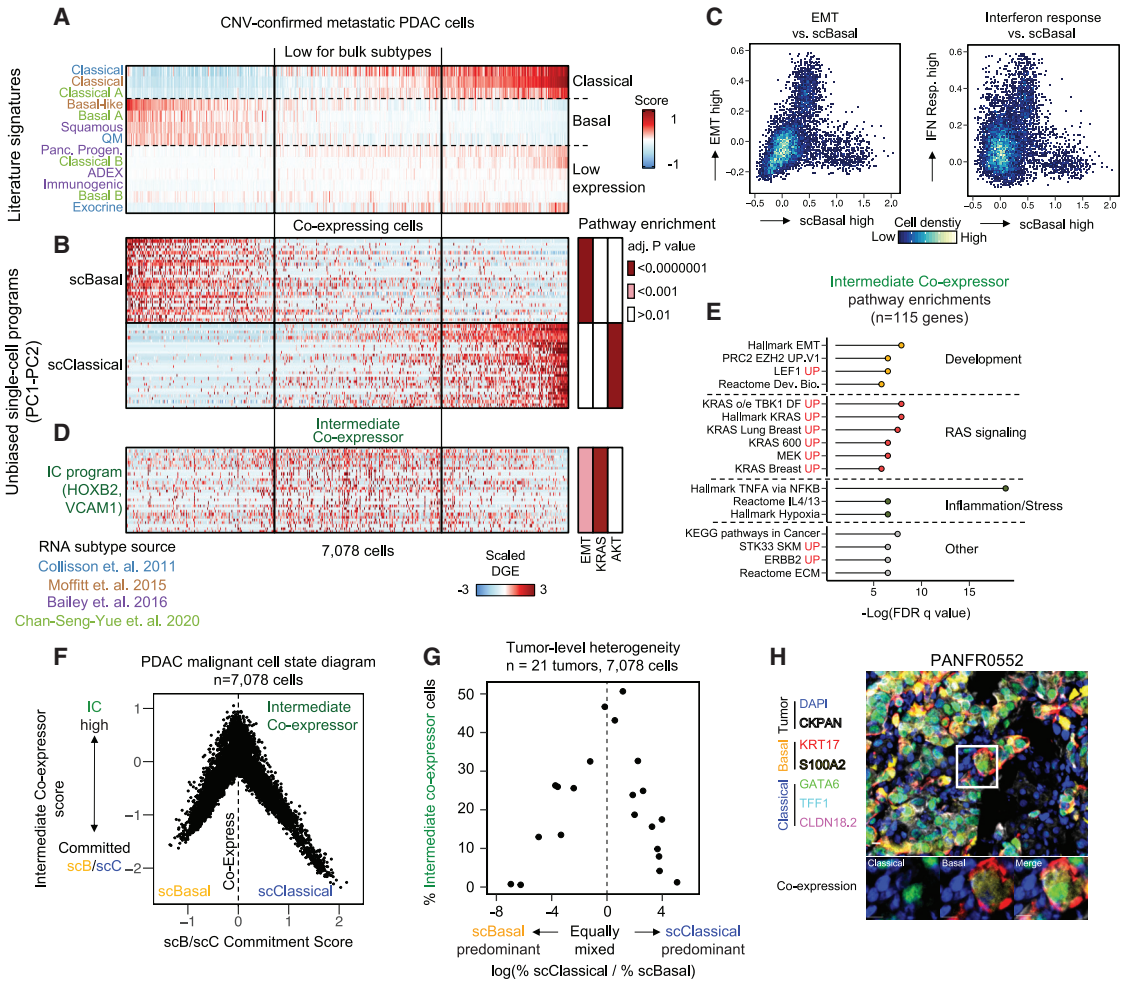


Figure 2. An intermediate co-expressor state bridges basal and classical phenotypes

(A) Signature scores (rows) for bulk derived expression subtypes in malignant cells (columns).

(B) Heatmaps depict the expression of scBasal and scClassical expression programs and highlight a co-expressing cell population.

(C) Variation in EMT and IFN response signature expression within malignant cells that have scBasal expression.

(D) The intermediate co-expressor ("IC") expression program is enriched in co-expressing cells. Enrichment adjusted *P*-values (hypergeometric test) for EMT, *KRAS*, and *AKT* gene sets are indicated at right in (B) and (D).

(E) Gene set enrichment analysis for the 115 genes specific to the intermediate co-expressor program.

(F) Malignant cell state diagram for PDAC. ScBasal-scClassical commitment score (x axis) and IC score (y axis).

(G) Frequency of co-expressing cells is related to increased mixing of scBasal and scClassical cell populations. Log ratio of % scBasal and scClassical cells in each sample (x axis; dotted line at 0 indicates equal percentages of scBasal and scClassical cells) versus their % IC cells (y axis).

(H) Multiplex immunofluorescence analysis identifies co-expressing cells in matched metastatic samples. White box indicates region for co-expression insets at bottom. Scale bar represents 10 μm.

See also Figures S2, S3, and S4; Table S3.

degrees, and were poorly described by previously identified bulk RNA subtypes (Figures 2B ["Co-expressing cells"], S3A, and S3B). We identified 115 genes whose expression was correlated with scBasal-scClassical co-expression and termed this gene set the IC state (Figures 2D, S3C, and S3D; Table S3; STAR Methods).

The IC state is enriched for developmental, RAS signaling, and inflammation/stress response gene sets (Figure 2E). Signatures of RAS signaling are enriched in the IC state even compared with scBasal and scClassical programs, and, by contrast, scClassical states are enriched for AKT-associated gene sets (Figures 2B, 2D, 2E, and S2F). We also assessed whether the IC state overlapped with any phenotypes recently reported in a study of normal pancreas progenitors (Qadir et al., 2020). We found that both scBasal and scClassical gene expression signatures are expressed by pro-ductal progenitor cells, while the IC gene expression program is enriched in an undifferentiated, stress-responsive progenitor population (Figure S3E) (Qadir et al., 2020). Although the IC state may not represent a distinct step along a developmental trajectory, it may represent a similar stress-induced transition state in a cancer context.

We next assessed how this new IC state relates to previously proposed bulk signatures to clarify potential inter-relationships. Pairwise correlation of our PDAC single-cell and established bulk RNA-seq signatures in malignant cells revealed that the IC state is unique and not well described by prior signatures (Figure S3F). Our findings suggest that malignant PDAC cells organize in a tripartite cell state framework that spans committed basal and classical phenotypes, with considerable signature co-expression in single cells (Figure 2F). Similar to the variation in EMT scores observed in scBasal-expressing cancer cells (Figure 2C), we noted heterogeneity among co-expressing cells for the IC program (Figure 2F).

Classification of tumors by their malignant pseudo-bulk signature expression stratified the cohort into those that expressed predominantly scBasal, scClassical, or IC signatures (Figure S3G–S3I). Individual patient specimens still exhibited significant heterogeneity at the cellular level, containing at least two and sometimes all three malignant cell states (Figure S3J). Samples with greater malignant cell state diversity (i.e., higher proportions of both scBasal and scClassical cells) also exhibited a higher proportion of cells expressing the IC state, suggesting the IC state may serve as a transition between scBasal and scClassical poles (Figure 2G).

Multiplex immunofluorescence confirms a tripartite cell state framework in metastatic and primary PDAC

To validate this extensive heterogeneity and the presence of co-expressing cells in our metastatic cohort, we used a subtype-specific multiplex immunofluorescence (mIF) panel to categorize single malignant cells by their patterns of marker detection in 10 matched cases from our single cell study (Figure S4A; Table S3; STAR Methods). We observed overlap of basal and classical markers within single cells at the protein level, corroborating the existence of co-expressing cells (Figures 2H and S4B). Moreover, we observed a significant correlation between malignant phenotypes assessed by mIF protein detection for samples from the same RNA subtype (average $r = 0.52$) compared to those of different subtypes (average $r = 0.06$, $p < 10^{-7}$, Student's t test) (Figure S4C, white dots). Deep sampling of each matched biopsy using mIF (mean = 13,078 cells per sample) also confirmed the increased fraction of co-expressing cells in tumors with equal basal and classical cell state mixing (Figure S4D). Finally, we used mIF to identify co-expressing cells in primary tumor samples, suggesting that the IC state may be a general feature of PDAC (Figure S4E).

Benchmarking model expression state using matched *in vivo* reference maps

With this high-resolution reference map of *in vivo* malignant cell states, we next asked whether matched *ex vivo* models retained the cell state distribution of the tissues from which they were derived. An unbiased comparison of all malignant cells (biopsy and organoid; 32,073 cells) revealed separation of biopsy and organoid profiles, while organoid samples from iterative passages *ex vivo* clustered together (Figures S5A and S5B). After removing low-frequency non-malignant cells (Figures S5B–S5D; STAR Methods), we observed that models attempted from biopsies with high malignant-cell-averaged scBasal or IC states exhibited lower rates of long-term propagation than models derived from scClassical tumors (Figure 3A). When comparing early passage CNV-confirmed organoid cancer cells to their cognates from patient tissues, culture in an *ex vivo* microenvironment caused greater deviation in cell state than CNV-defined genotype (Figure 3A; STAR Methods).

We next examined how *ex vivo* transcriptional states in our organoid cohort differed from their matched patient samples (Figure 3B). We observed a striking loss of scBasal gene expression and to a lesser extent, the IC program. By contrast, aggregate scClassical gene expression remained largely unchanged in organoid conditions. This comparative analysis also nominated a set of upregulated organoid-specific genes that were not present *in vivo*, including markers of epithelial identity, oxidative stress response pathways (e.g., NRF2 target genes), and amino acid metabolism (hereafter collectively referred to as ''organoid-specific'' gene expression) (Figure 3C, bottom; Table S4). These findings suggest that changes to microenvironmental growth conditions significantly alter cellular transcriptional states and induce culture-specific expression programs, highlighting the importance of benchmarking *ex vivo* models using matched *in vivo* states as a reference.

Transcriptional state heterogeneity is shaped by the *ex vivo* microenvironment

Regardless of the cell state distribution in the original biopsy, individual models assumed more scClassical or organoid-specific cell states over time in culture (Figure 3D), mirroring the results in our larger bulk RNA-seq cohort (Figures 1B and 3C [top]). To better understand the drivers of this cell state bias, we first investigated genetic alterations associated with either the basal or classical subtypes. Prior work has suggested that *KRAS* amplifications associate with basal features (Chan-Seng-Yue et al., 2020; Miyabayashi et al., 2020), while amplifications of lineage transcription factors like *GATA6* associate with classical subtypes (Chan-Seng-Yue et al., 2020). In agreement with prior bulk studies, we observed a significant association between single-cell inferred *KRAS* copy number gain and the scBasal state in metastatic cells ($p < 0.03$ Fisher's exact test) (Figure S5E). Four *KRAS*-amplified samples proliferated *ex vivo* and maintained this alteration, but their malignant cell state shifted from scBasal *in vivo* to scClassical in organoid culture (Figures 3D [black dots] and S5F), including one sample where the same *KRAS*-amplified clone exhibited a similar scBasal (biopsy) to scClassical (organoid) state shift (Figure 3E). These findings demonstrate that *KRAS* amplification alone is not sufficient to lock the scBasal expression state, PDAC cells are plastic, and the microenvironment can influence cell state independent of genotype in this context.

We next compared genetic heterogeneity (inferred CNV subclones) and transcriptional states from matched biopsy tissue and organoid samples from iterative passages. ScClassical tumors tended to maintain their genotype and transcriptional state both early in culture and at later passages (e.g., PANFR0631) (Figures 3D and S5G [clone A]). In contrast, most models derived from scBasal or IC tumors exhibited early cell-state deviation and cessation of growth within 100 days of initiation (e.g., PANFR0552) (Figures 3D and S5H). In some cases, we observed

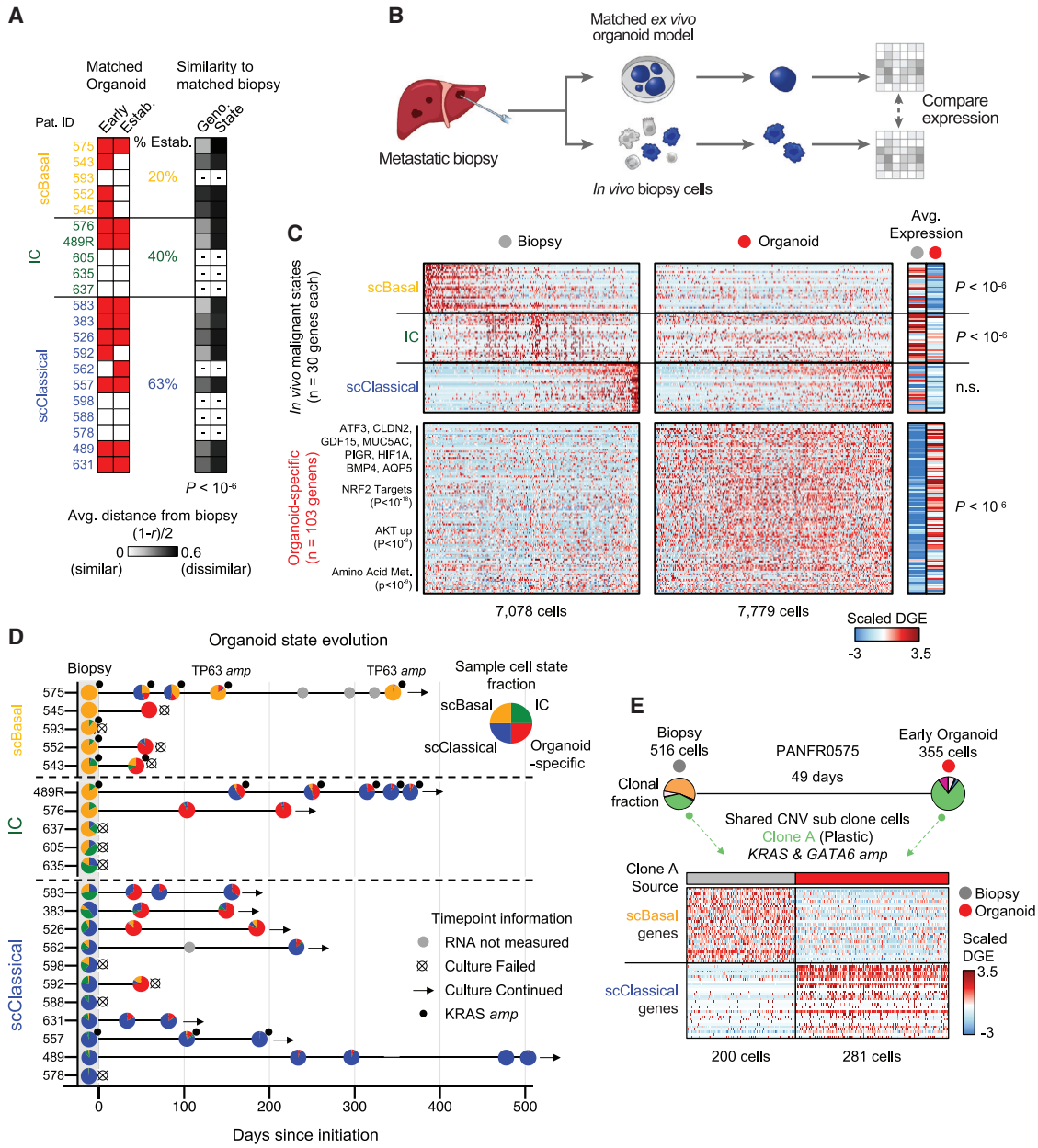


Figure 3. Organoid culture selects against the scBasal state with transcriptional evolution over time

(A) Outgrowth and similarity between organoids and matched biopsy samples. Red fill indicates successful early (Early) and long-term (Estab.) culture. Right gray scale indicates similarity between each biopsy-early organoid pair for inferred CNVs (genotype, Geno.) or cell state (State). *P*-value for Geno. versus State differences determined by Student's t test.

(B) Schematic for matched *in vivo* malignant cell and organoid comparison.

(C) Single-cell and average expression of malignant programs (top) and organoid-specific genes (bottom) in biopsy cells and matched, early passage organoid cells (n = 13 models). *P*-values determined by Student's t test. Parenthetical *P*-values (left) indicate hypergeometric test for pathway enrichments.

(D) Swimmer's plot depicting organoid state evolution in culture. Pie charts indicate the fraction of cell states at each time point.

(E) Clonal fractions from *KRAS*-amplified PANFR0575 biopsy (gray) and organoid (red) cells. Clone A (green) is present in both. Heatmap shows expression of scBasal and scClassical states in clone A from both contexts.

See also Figure S5.

outgrowth of an scClassical sub-clone in culture. For example, the *in vivo* scBasal clones from PANFR0489R rapidly decreased in abundance while clones with scClassical or organoid-specific states emerged at later passages (Figures S5I and S5J). By contrast, *in vivo* scBasal clones from PANFR0575 were maintained initially in culture, but their cell state was highly plastic

and changed to scClassical when measured as organoids at passages 2 and 3 (Figures 3E and S5K). After > 100 additional days in *ex vivo* culture, PANFR0575 regained scBasal expression and clones with inferred *TP63* amplifications, a squamous-specifying transcription factor (Somerville et al., 2018), dominated the culture (Figure S5K [clones D and E]). Our detailed analysis reveals that both plasticity and selection occur in the same model, where certain clones demonstrate plasticity in response to microenvironmental signals, and certain genotypes, though rare, may still exert a strong effect despite opposing cues from the microenvironment. In sum, these findings underscore the need to consider both mutational and transcriptional state to ensure faithful representation of *in vivo* cancer cell phenotypes.

Media formulation influences PDAC transcriptional states *ex vivo*

Given the evidence for selection and plasticity in standard organoid conditions, we tested whether specific aspects of the *ex vivo* culture environment governed cell state determination. We first evaluated the effects of extracellular matrix dimensionality by culturing established 3-dimensional (3D) organoid models as 2-dimensional (2D) cell lines in the same organoid medium. This did not affect transcriptional subtype across the 4 models tested (Figure S5L). Next, we asked whether culturing established organoid models in altered media conditions could rescue expression heterogeneity (Figure 4A). We cultured 4 organoid models in media without any additives for 6 days (“Minimal” medium) (Table S5; STAR Methods) and observed a robust increase in scBasal expression, a decrease in organoid-specific gene expression, and stable CNV profiles (Figures 4B, S5M, and S5N). With greater time in minimal medium, the distribution of cell states shifted even further toward IC and scBasal (Figure 4C). Since minimal medium lacks serum and mitogens to maintain prolonged cell growth, we tested a reduced organoid media formulation (“OWRNA”) (Table S5; STAR Methods) and found that OWRNA supported proliferation, strengthened scClassical, and allowed for scBasal expression (Figure S5O).

Given the divergence in cell state for cell line versus organoid models of PDAC (Figures 1A and 1B), we took an established cell line and an organoid model and cultured them in the reciprocal media condition. Organoid cells grown in standard cancer cell line medium (“Cell line media,” RPMI-1640 with 10% fetal bovine serum) (STAR Methods) gained scBasal expression, while CFPAC1, an established pancreatic cancer cell line, lost scBasal features when grown in complete organoid media (Figures 4D, 4E, S5P, and S5Q). To understand whether these state changes are functionally significant, we exposed CFPAC1 cells grown in basal media conditions (“cell line media”) or classical conditions (complete organoid media) to SN-38 and paclitaxel. Strikingly, in both instances, we observed greater sensitivity among cells grown in organoid media, suggesting that environmental factors can shape therapeutic response through state changes (Figure 4F). These findings demonstrate that even established models are sensitive to changes in culture conditions and that a combination of up-front selective pressures and culture medium-imprinting shapes the cell-state biases in extant model systems, influencing their functional attributes.

Although maintaining organoids in minimal or cell line media conditions resulted in partial recovery of scBasal and IC expression, we failed to observe fully polarized models (Figures S5P and S5Q), suggesting that these conditions lack critical TME components to fully specify cell state. We used differential expression across each biopsy-organoid pair to nominate genes that were expressed *in vivo* but missing *ex vivo* (Figure 3B; STAR Methods). Genes differentially expressed by malignant cells *in vivo* were related to soluble cytokine signaling, cell-cell communication, and microenvironmental interactions (Figure 4G). Hierarchical clustering revealed state-dependent expression patterns for *in vivo*-specific genes (Figure 4H; Table S4). For example, IFN response and EMT genes were significantly upregulated in scBasal and IC malignant cells *in vivo* (clusters 1 and 2), while genes associated with cell-cell interactions and surface glycoproteins were more strongly expressed in IC and scClassical cells (cluster 3). Genes related to biological adhesion were more uniform in their expression across the subtypes (cluster 4) (Figure 4H). The relative absence of these genes in organoid culture and their differences in expression across transcriptional subtypes *in vivo* suggests that TME signals might play a role in fully specifying cancer cell transcriptional state and potentially therapeutic responses.

Composition of the metastatic TME and site-specific differences in mesenchymal populations

The presence of TME-associated expression patterns in cancer cells *in vivo* suggested there may be subtype-dependent structure to, and instructive signaling from, the metastatic TME; however, relatively little is known about the composition of the metastatic TME in PDAC. We analyzed the non-malignant cells (n = 14,811) in the metastatic niche to further subclassify cell types and provide a more complete picture of the immune/stromal composition of metastatic disease (Figure 5A). Sub-clustering of T cells and natural killer (NK) cells revealed 4 cell types (*CD4*+ T, *CD8*+ T, NK, and *CD16*+ [*FCGR3A*+] NK cells) (Figures S6A and S6B; STAR Methods) Similarly, an unbiased analysis within the monocyte/macrophage compartment revealed 3 subsets of tumor associated macrophages (TAMs) (*FCN1*+ “monocyte-like” TAMs, *C1QC*+ TAMs, and *SPP1*+ TAMs) (Figures S6C and S6D; Table S2) (Zhang et al., 2020; Zilionis et al., 2019). Marker expression across all previously described non-malignant cells is summarized in Figure S6E.

Whether the TME differs between primary PDAC and different metastatic sites is not well understood. Although we found equal distribution of immune cells across different metastatic sites, mesenchymal cells clustered by the site of disease (Figures S6F and S6G). Despite uniform expression of a previously described myofibroblast signature (Elyada et al., 2019; Öhlund et al., 2017), an unbiased analysis revealed divergent mesenchymal states favoring expression of either fibroblast-like or pericyte-like genes (Figures S6H–S6K; Table S2) (Bartoschek et al., 2018; Di Carlo and Peduto, 2018; Hosaka et al., 2016; Pelon et al., 2020). While biopsies from each location contained both subsets, we detected a strong association between liver biopsies and the pericyte-like mesenchymal state (Figures 5B, S6L, and S6M). We validated this association in larger cohorts and observed a similar relationship between pericyte-like

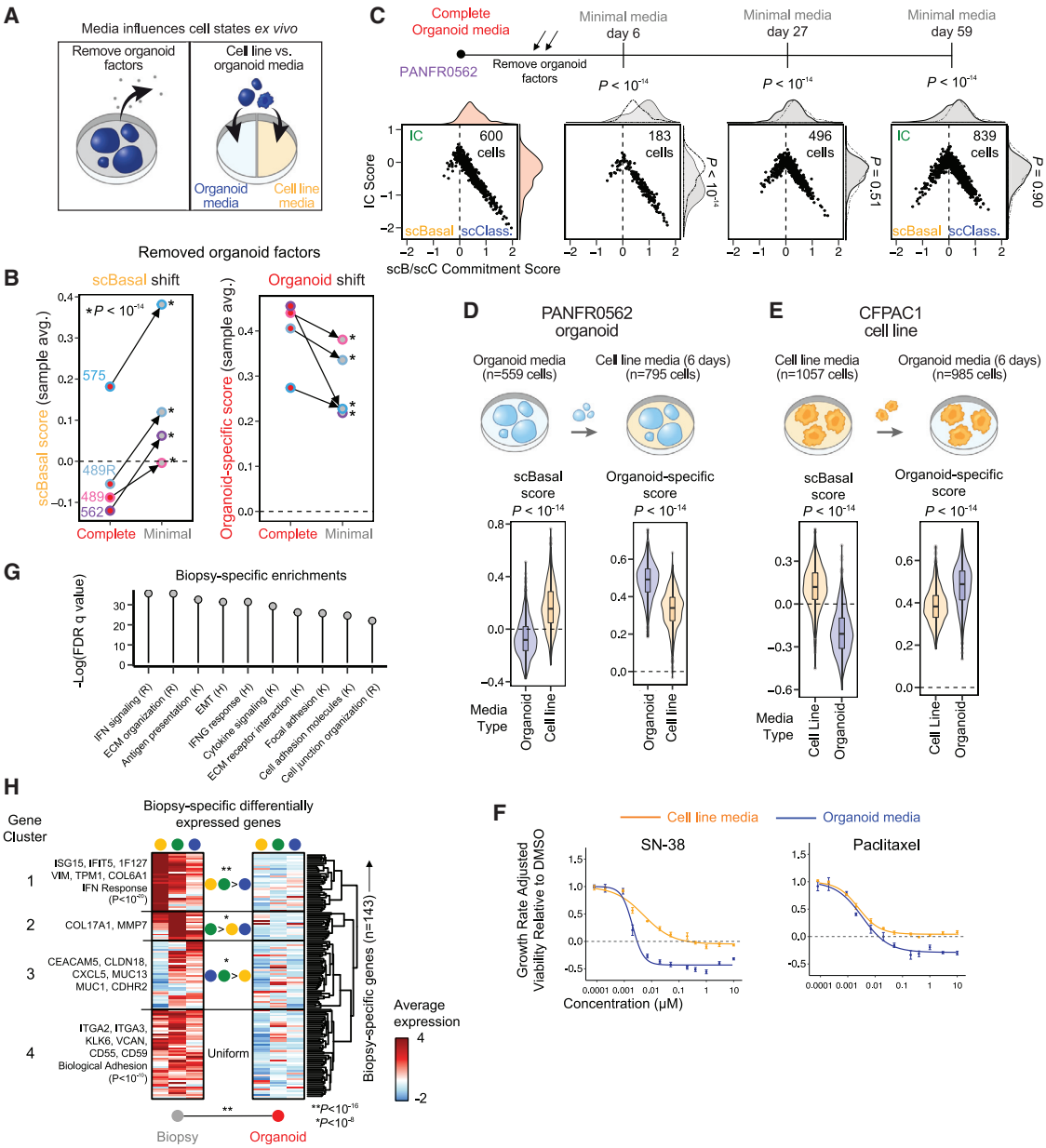


Figure 4. Modulation of the media microenvironment allows recovery of scBasal states

(A) Strategies to recover scBasal expression in different media conditions.

(B) Tied dot plot for sample average scBasal score (left) and organoid-specific score (right) in the indicated conditions. Color outline indicates sample identity. *P*-value compares single cell distributions within models and was calculated by Student's t test.

(C) Cell state diagrams for organoid cells cultured in minimal media. *P*-values are for that time point versus the complete media condition and compare B/C commitment (top) and IC scores (right) by ANOVA followed by Tukey's HSD.

(D and E) Violin plots depict scBasal and organoid-specific expression scores in PANFR0562 organoid cells (D) or CFPAC1 cell line (E) after 6 days in Complete organoid medium or in "Cell line" medium. *P*-values for differences were calculated by Student's t test.

(F) CFPAC1 cell line growth rate-adjusted dose response curves to SN-38 and paclitaxel after culture in standard "Cell line" medium or in Complete organoid medium. Points are the mean ± SD of 3 technical replicates. Curves are representative of 2 independent experiments.

(G) Significant pathway enrichments (*P*-value < 10^{-12}) for top *in vivo* differentially expressed genes (143 genes).

(H) Average expression in biopsy (left) and organoid cells (right) for the 143 top *in vivo* differentially expressed genes (rows), organized by originating tumor's overall transcriptional subtype (colored dots). Parenthetical *P*-values (left) for enrichment of indicated pathways are by hypergeometric test. Overall biopsy versus organoid expression difference is determined by Student's t test (bottom). *P*-values computed by one-way ANOVA followed by Tukey's HSD (center) are for differences in average expression between biopsy transcriptional subtypes (*P-value < 10^{-8}; **P-value < 10^{-16}).

See also Figure S5; Tables S4 and S5.

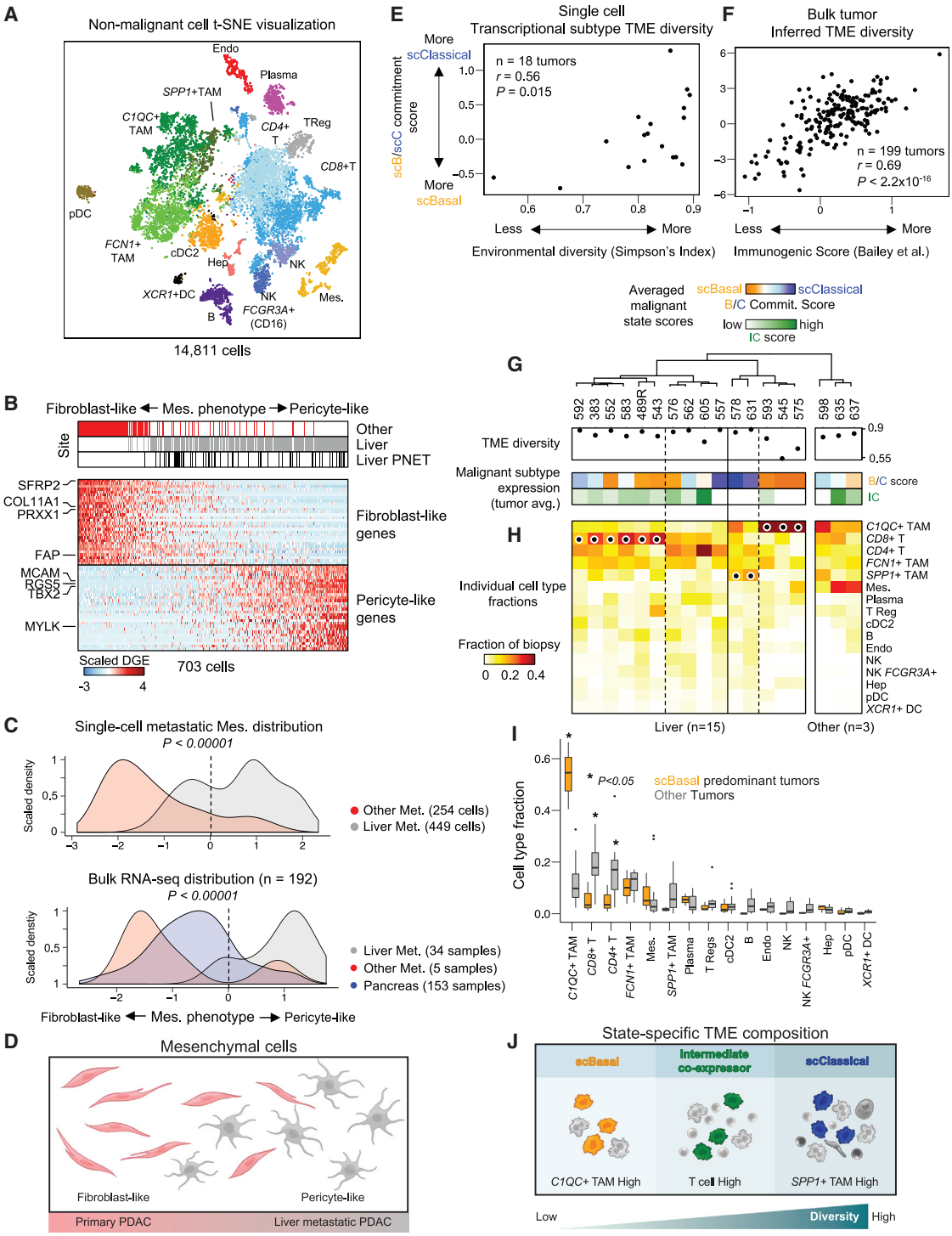


Figure 5. Distinct mesenchymal phenotypes and transcriptional state-specific immune heterogeneity exist in the liver metastatic microenvironment

(A) *t*-SNE visualization of non-malignant cells identified in the metastatic microenvironment, abbreviations as in Figure S1J (TAM, tumor associated macrophage).

(B) Expression of Fibroblast-like (PC2 low) and Pericyte-like (PC2 high) mesenchymal (Mes.) cell programs across different metastatic sites (top).

(C) Density plots for mesenchymal cell phenotype score in single cells from our metastatic cohort (top) or previously published PDAC bulk RNA-seq profiles (bottom), fill indicates anatomic site. *P*-value determined by Student's t test (top) or by ANOVA followed by Tukey's HSD (bottom).

(D) Summary of mesenchymal phenotypes in primary versus liver metastatic PDAC.

(legend continued on next page)

expression and liver metastases, while in contrast, tumors in the pancreas favored fibroblast-like mesenchymal expression (Figure 5C). Thus, we observed diverse immune and stromal cell types in the metastatic TME and identified mesenchymal features unique to the liver niche compared with primary disease (Figure 5D).

Transcriptional subtypes associate with distinct immune microenvironments

We next searched for associations between malignant cell state and the composition of the TME. Five samples were excluded from this analysis on the basis of low cell counts (< 200 cells) or indeterminant transcriptional subtype (Figure S6N). We applied Simpson's diversity index to define each tumor's overall microenvironmental composition (STAR Methods). Tumors with higher average malignant scClassical or IC expression harbored greater TME diversity, while strongly scBasal tumors exhibited more homogeneous TMEs (Figure 5E). We observed a similar pattern in bulk samples where we inferred diversity by using the "immunogenic" signature (Bailey et al., 2016) to indicate greater immune cell infiltration (Figure 5F). Clustering over the cell type fractions in each biopsy revealed the non-malignant cell types driving overall diversity differences (Figures 5G and 5H). C1QC+ TAMs dominated the TME of strongly scBasal tumors, and both CD8+ and CD4+ T cells were significantly depleted in scBasal contexts compared to the rest of the samples in the cohort (Figures 5H and 5I). By contrast, T cells were most often isolated from, and their abundance positively correlated with, higher IC malignant fractions in our cohort (Figures 5I and S6O). We observed similar patterns in bulk RNA-seq data from The Cancer Genome Atlas (TCGA) (Weinstein et al., 2013; Vivian et al., 2017), noting reduced levels of immune-related gene expression in other epithelial tumors with high basal/squamous gene expression (Figure S6P [cluster 4]). Together, these findings suggest coordination between cancer cell states and the local TME, with decreased immune cell diversity in basal contexts (Figure 5J).

State-specific TME signals drive transcriptional heterogeneity and drug response

Based on these observations, we hypothesized that soluble factors specific to the TME of each transcriptional subtype may drive cancer cell states and potentially influence their therapeutic responses (Figure 6A). *In vivo*, the secreted factor milieu surrounding cancer cells originates from at least two sources: malignant cells themselves ("autocrine" factors) and non-malignant cells ("paracrine" factors) (Figure 6A). We first identified secreted factors differentially expressed by cancer cells in each transcriptional state ("autocrine" signals) and applied these to rescue IC and scBasal expression in *ex vivo* models (Figure 6B; Table S6; STAR Methods). *TGFB2* was the top differentially expressed factor shared by malignant cells in both scBasal and IC TMEs (Figure 6B), and organoids cultured with TGF-β ligands exhibited a pronounced shift toward IC and scBasal states (Figure 6C). The reemergence of scBasal transcriptional heterogeneity in both minimal media (Figure 5C) and TGF-β conditions (Figure 6C) suggested that different types of microenvironmental pressure can lead to the basal-like cell state. These experiments also indicate that culture conditions can be tuned to achieve compositional differences spanning scClassical, heterogenous, and scBasal expression, akin to those seen *in vivo*.

We next assessed whether TME signals, like media formulation (Figure 4F), could influence drug sensitivity through altering transcriptional state. In an isogenic organoid system (STAR Methods), models induced to adopt scBasal expression through exposure to TGF-β for 3 weeks were less sensitive to several standard-of-care chemotherapeutic agents including gemcitabine, paclitaxel, and SN-38, the active analog of irinotecan (Figures 6D, S7A, and S7B; Table S7). The duration of exposure to TGF-β corresponded with the degree of state shift, and these states were highly plastic, as withdrawal of TGF-β resulted in a return to scClassical expression (Figure 6E). Drug sensitivity tracked with cancer cell state, as models were re-sensitized to chemotherapeutic agents upon shifting back to the scClassical state (Figures 6E and 6F). These observations agree with recent clinical trial data showing that patients with basal phenotypes tend to have poorer outcomes with combination chemotherapy (Aung et al., 2018; Bailey et al., 2016; Collisson et al., 2011; Connor et al., 2019; Moffitt et al., 2015). Together, these findings highlight the remarkable phenotypic and functional plasticity inherent in tumor models.

To assess for state-specific vulnerabilities more broadly, we tested a 24-drug panel in isogenic organoid systems with the scBasal state induced via two distinct routes, either by culturing in minimal or TGF-β-containing medium (Figure 6G; Table S7). Within the same model, scClassical states were on average more sensitive to both chemotherapy and agents targeting DNA-damage repair pathways while scBasal states were more sensitive to mitogen-activated protein kinase (MAPK) pathway inhibitors (MEK and ERK inhibitors) (Figures 6G, S7A, and S7B). These findings provide direct evidence that transcriptional state can be a major determinant of drug response and that

(E) Correlation between microenvironment diversity (Simpson's index, x axis) and the average malignant scBasal-scClassical commitment score (y axis) for each scRNA-seq sample.
(F) Correlation between TME diversity as inferred by immunogenic signature score (x axis) versus average tumor scBasal-scClassical commitment score (y axis) in primary and metastatic bulk RNA-seq samples.
(G and H) Sample-level (columns) variation in Simpson's index (G, dot plot), average malignant scBasal-scClassical (G, top heat bar) and IC (G, bottom heat bar) expression, and fraction of non-malignant cell types (H). Samples were clustered and ordered within metastatic site (liver versus other) by their fractional representation of cell types. Dots indicate top statistically significant cell type frequency differences calculated using a Kruskal-Wallis test with multiple hypothesis correction.
(I) Boxplots compare cell type fractions between the scBasal predominant tumors with low diversity (PANFR0593, 575, 545) and all others. *P*-value determined by Student's t test.
(J) Summary of associations between microenvironmental diversity, non-malignant cell types, and malignant cell state.
See also Figure S6; Table S2.

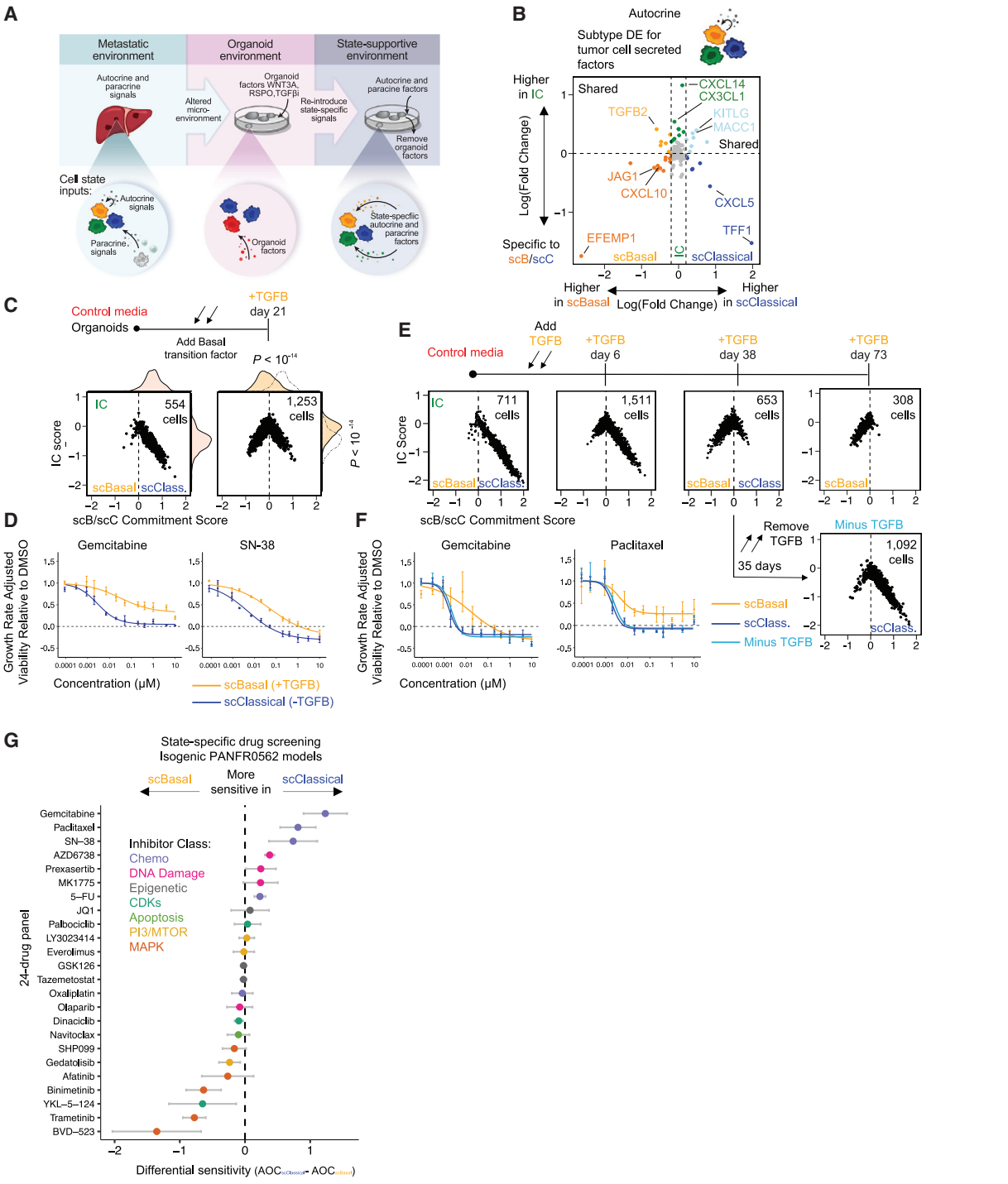


Figure 6. Tumor state-specific factors rescue malignant transcriptional heterogeneity and reveal state-specific drug sensitivities

(A) Schematic describing microenvironmental inputs to tumor cell state *in vivo* (left, "Metastatic environment") versus *ex vivo* (center, "Organoid environment") and a strategy to recover malignant transcriptional heterogeneity *ex vivo* (right, "State-supportive environment").

(B) Differential expression (Wilcoxon rank sum test) of secreted factors between *in vivo* tumor cells scored as scBasal versus scClassical (x axis) and IC malignant cells versus the rest (y axis). State-specific genes that pass significance after multiple hypothesis correction ($p < 0.05$) are colored by their group association.

(C and D) Cell state diagrams with marginal density plots (C) and growth rate-adjusted dose response curves to gemcitabine and SN-38 (D) for organoid model PANFR0562 cultured for 3 weeks in control medium (OWRNA) or in control medium with TGF-β. *P*-values in (C) for group differences between B/C commitment (top) and IC scores (right) were calculated by ANOVA followed by Tukey's HSD. In (D), points are the mean ± SD of 3 technical replicates. Curves are representative of 2 independent experiments.

(legend continued on next page)

differential targeting of cell states represents an actionable therapeutic paradigm.

Paracrine signals from the local TME direct cancer cell transcriptional phenotypes

We next searched for paracrine factors differentially expressed by the non-malignant cells in each subtype. We noted an increasing number of differentially expressed factors in the IC and scBasal contexts and mapped each paracrine factor to its cognate cell type to summarize the overall secreted factor combinations that shape subtype-specific TMEs in metastatic PDAC (Figures 7A and 7B; Table S6; STAR Methods). Interestingly, *IFNG* from *CD8+* T cells was most highly expressed in the IC TME, consistent with a higher T cell fraction in IC tumors (Figures 5G and S6O) and the relative increase in IFN-responsive gene expression in IC and scBasal malignant cells (Figures 4G and 4H). Correspondingly, cells from organoid models exposed to interferon γ (IFNγ) showed increased IFN response gene expression (IFN response score) and a concomitant shift toward the IC state (Figures 7C, S7C, and S7D).

We next examined two *in vivo* scenarios for evidence that the TME can drive cell state variation by analyzing samples from (1) distinct metastatic sites within the same patient and (2) the same metastatic site before and after immunotherapy. In the first case (PANRF0473), a larger fraction of T cells within the liver metastatic niche expressed high levels of *IFNG* (IFNG expression score) relative to the lung metastatic niche. Correspondingly, malignant cells in the liver showed evidence of IFN response and higher IC state expression (Figure 7D, S7E,F). Similarly, in serial samples from the same liver lesion in patient PANFR0489, the post-progression biopsy harbored a higher fraction of T cells with high IFNG expression scores, consistent with changes to the TME stemming from immunotherapy. Post-progression malignant cells again expressed higher IFN response scores, and their transcriptional phenotypes were shifted toward the IC and scBasal states relative to the pre-treatment malignant cells (Figures 7E, 7F, S7G, and S7H). These paired biopsies allow a window into the complex interplay between the TME and cancer cells within the same patient and, consistent with our organoid studies, illustrate the critical role that TME-supplied cytokines and immune activation may play in directing cancer cell state.

DISCUSSION

Single-cell atlases of cancer have revolutionized our understanding of human malignancies and revealed that mutational and transcriptional heterogeneity are common. Critical next steps include understanding what drives cancer cell state and whether it can be targeted therapeutically. Here, we provide a systematic framework for assessing cancer cell states, identifying drivers of transcriptional plasticity, and evaluating their functional significance in model systems. The use of matched *in vivo* tissue as a reference for culture model state fidelity, an approach that has been applied in normal organoid systems (Fujii et al., 2018) but not in cancer, enabled a structured dissection of cell-intrinsic versus TME-induced contributions to malignant cell state. As the community continues to catalog these transcriptional states in clinical samples at single-cell resolution, we anticipate this framework will be broadly applicable for understanding their functional significance across a variety of cancers.

Single-cell resolution enabled us to appreciate the layering of malignant states whereby single cells can be unified in their similarity for an expression subtype (e.g., scBasal or IC) but differ in their expression of TME-influenced programs (e.g., IFN response). Future studies across more patients and different anatomic disease sites at single-cell resolution will be needed to fully parse which invariant "core" genes mark archetypal cell states in PDAC and which expression features are superimposed by the TME. Clinically annotated datasets will also aid in the assessment of these expression states for prognostic value and their utility in nominating cell state-specific therapeutic liabilities.

We also uncovered formerly unappreciated relationships between cancer cell transcriptional states and the local TME. Similar to the relationship between inflammation and tumorigenesis (Alonso-Curbelo et al., 2021; Li et al., 2021), our data support a model wherein as tumors become inflamed and immune-activated, malignant cells display enhanced state plasticity. These relationships may have implications for PDAC immunotherapy strategies given that a productive CD8 T cell response may promote more aggressive basal-like states, as suggested by our paired pre- and post-immunotherapy samples. Coordination between basal/mesenchymal malignant cell states and immune responses may be a broadly relevant phenomenon given our observations in other basal-like tumors and recent work in glioblastoma (Hara et al., 2021). A recent publication further supports the idea that local TME variation can influence clinical outcomes in PDAC (Grünwald et al., 2021), while our work provides direct evidence that TME signals are critical drivers of malignant cell state and that cell state dictates drug sensitivity. Larger cohorts of longitudinal single-cell measurements, both spatially resolved and transcriptome-wide, from individual patients undergoing therapy are needed to fully assess variation in the TME, the kinetics of cell state plasticity, and their consequences for therapeutic response.

While the genetic evolution of *ex vivo* models is an established phenomenon (Ben-David et al., 2019; Ben-David et al., 2018), our study shows similar *ex vivo* evolution along a transcriptional axis that causes discordance between a patient sample and its matched avatar. For example, biopsies taken from PANFR0489R and PANFR0576 establish long-term cultures *ex vivo*, but they do not represent the dominant cell state of the corresponding patient tissue *in vivo* (Figure 3D). Given our demonstrated link between

(E) Cell state diagram time series for PANFR0562 organoids cultured with TGF-β or after TGF-β removal.
(F) Growth rate-adjusted dose response curves to gemcitabine and paclitaxel for models in (E). Points are the mean ± SD of 3 technical replicates.
(G) State-specific drug sensitivities in isogenic organoid model pairs skewed toward scBasal or scClassical states by altering media formulation. Points are the mean ± SEM of 2–6 biologic replicates for the difference in growth rate-corrected Area Over the Curve (AOC) between each scBasal-scClassical model pair.
See also Figure S7; Tables S5, S6, and S7.

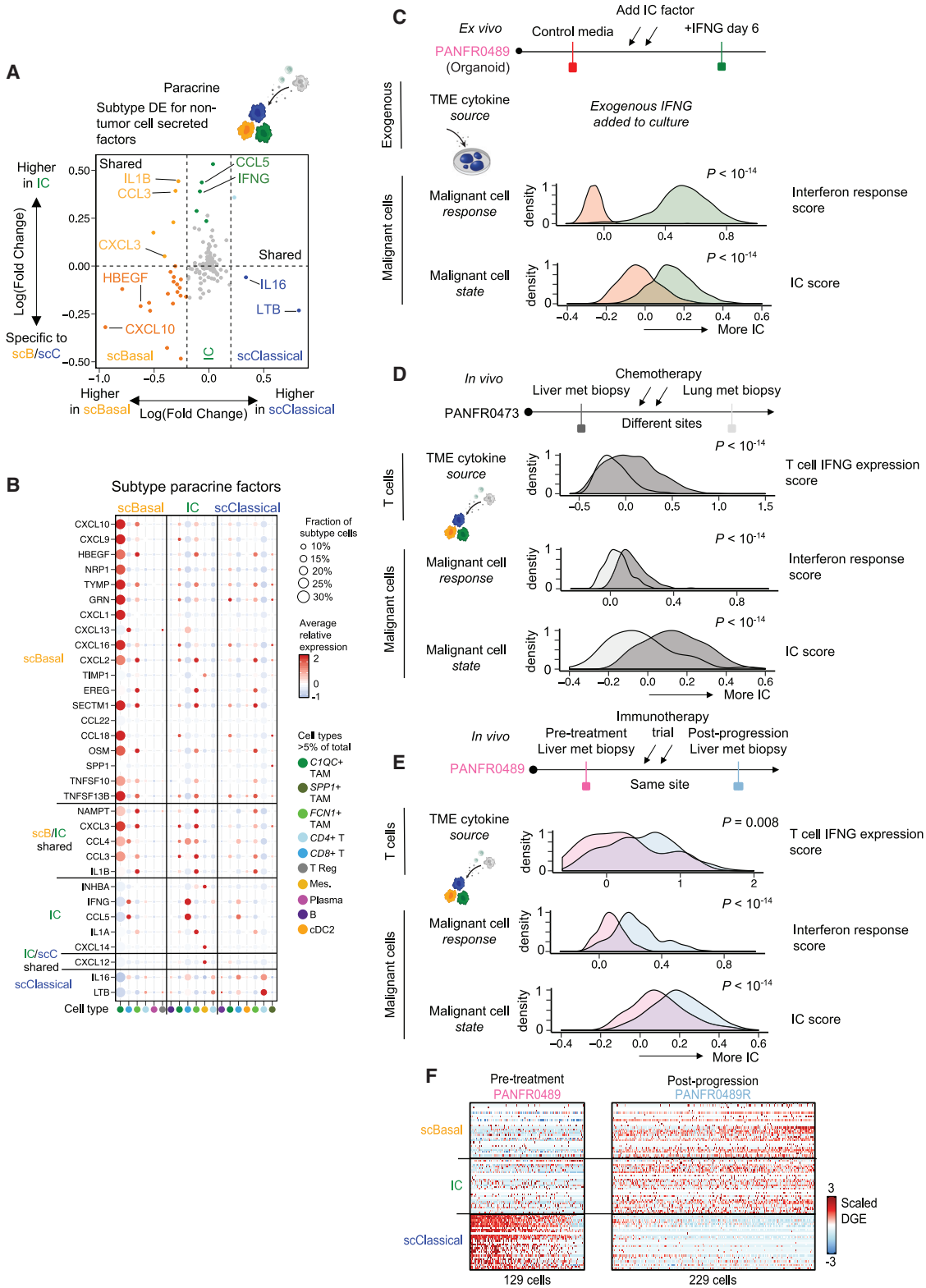


Figure 7. Malignant transcriptional states respond to TME alterations in organoid models and *in vivo*

(A) Differential expression (Wilcoxon rank sum test) of secreted factors by non-malignant cells (paracrine) grouped by their sample-averaged malignant cell expression state in scBasal and scClassical (x axis) tumors and IC biopsies and the rest (y axis). State-specific genes that pass significance after multiple hypothesis correction ($p < 0.05$) are colored by their group association.

(legend continued on next page)

cell state and therapeutic response, these findings establish the necessity for preserving transcriptional fidelity in personalized medicine pipelines (Hahn et al., 2021). Since models from different patients may also harbor differential plasticity and adaptability, future efforts will need to evaluate models in multiple conditions to account for this latent property, define its drivers (e.g., genetics, epigenetics), and accurately map vulnerabilities. Importantly, manipulating the soluble microenvironment may offer a more tractable approach for state-specific high-throughput screening compared to more complex heterotypic co-cultures or patient-derived xenograft systems.

Within the context of PDAC, clinical studies are ongoing to evaluate the efficacy of gemcitabine/nab-paclitaxel and FOLFIRINOX in patients with basal- versus classical-predominant metastatic disease (e.g., PASS01 – NCT04469556). Our observations suggest that basal tumors may exhibit broadly decreased sensitivity to chemotherapy and highlight the need for new strategies to target this transcriptional subtype of PDAC. Importantly, we show that the transition to an scBasal state may render cells sensitive to other classes of inhibitors. Since most PDAC tumors are heterogenous, combination strategies that suppress distinct cell states may be necessary for maximal synergistic effect (Palmer et al., 2019; Palmer and Sorger, 2017). Furthermore, the ability of the TME to drive malignant cell state transitions suggests that next generation therapeutic strategies may also need to target site-specific supportive cells in the TME to control cell-state evolution during therapy.

In sum, we provide a widely applicable framework to benchmark cell states in patient-derived model systems, identify the drivers of malignant transcriptional heterogeneity, and examine the functional significance of cell state. As efforts to characterize cell states across malignancies provide new and increasingly higher resolution maps of patient tissues, an important next step will be to understand better how to control and target cell state. We anticipate that our approach will provide a path toward the systematic evaluation of cell state as a targetable feature in cancer.

Limitations of the study

We focused on understanding how the TME supports cell state, and thus we cannot comment directly on the epigenetic mechanisms through which these cell state transitions occur. Additional studies into the regulatory mechanisms underlying PDAC state transitions will be a critical next step in further delineating the relationships between cell-intrinsic and extrinsic factors and state. In addition, we tracked CNV alterations to define genetic clones, but future studies using barcode-based lineage tracing approaches are needed to confirm whether individual cells transition between scClassical and scBasal states via the IC state as our data suggest (Wagner and Klein, 2020). While we focused predominantly on metastatic biopsies from the liver, it will be important in future studies to more broadly analyze how site-specific cues drive state plasticity. Even with our limited number of samples from outside the liver, we uncovered site-specific differences in mesenchymal populations which revealed important differences between metastatic and primary disease. Given the pivotal role that has been suggested for the fibrotic TME in primary disease (Ho et al., 2020; Sahai et al., 2020), these findings carry important implications for targeting the stromal compartment in primary versus metastatic PDAC. Future studies with larger sample and cell numbers will be needed to make comparisons across cell types in primary and metastatic disease to fully understand how the TME regulates transcriptional phenotypes in these distinct niches. Here, advances in spatial transcriptomics will likely empower a deeper understanding of these relationships across compartments (Longo et al., 2021). Finally, future studies that broadly map the landscape of RNA state-dependent drug response across various cancer models will be needed to fully define the links between malignant cell transcriptional plasticity and therapeutic response.

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
  - Human PDAC patient-derived organoid models
- METHOD DETAILS
  - Sample preparation for single-cell RNA-sequencing of clinical samples, organoids, and cell lines
  - Assessing organoid and cell line transcriptional states under different matrix and media conditions

(B) Dot plot for state-specific significant differentially expressed paracrine factors (rows) by subtype-specific non-malignant cell types (columns). Dot size represents that cell type's fraction within tumors of each subtype, and fill color indicates average expression. Only cell types with a fractional representation > 5% from each subtype are visualized.

(C) Density plots of IFN response score (top) and IC score (bottom) in control organoid cells and after addition of IFNγ for 6 days. *P*-values were calculated by Student's t test.

(D) Biopsy samples from distinct metastatic sites (liver, dark gray; lung, light gray) in the same patient (PANFR0473) demonstrate co-variation in T cell IFNG expression (top), malignant cell IFN response score (middle), and malignant IC score (bottom). *P*-values for density plot differences were calculated by Student's t test.

(E) Biopsy samples from the same lesion pre- and post-immunotherapy (checkpoint inhibitor plus a macrophage-targeting agent; pre-, PANFR0489, pink; post-, PANFR0489R, blue) demonstrate coordinated changes with treatment in T cell IFNG expression (top), malignant cell IFN response score (middle), and malignant IC score (bottom). *P*-values for density plot differences were calculated by Student's t test.

(F) Heatmap for malignant cell state shifts from samples in (E).

See also Figure S7; Tables S5, S6, and S7.

- Compound sensitivity testing in cell lines and organoid models
- Single-cell RNA-seq (scRNA-seq) data library generation, sequencing, and alignment
- Bulk RNA- and DNA-sequencing of organoids
- Multiplex immunofluorescence imaging

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Mutation and CNV identification from bulk DNA-sequencing
  - Bulk RNA-sequencing analysis
  - Single-cell data quality pre-processing and initial cell type discovery
  - Single-cell CNV identification
  - Subclonal analysis with single-cell inferred CNVs
  - Subclustering of malignant and non-malignant cells
  - Generation of expression signatures/scores
  - Analysis of normal pancreas progenitor data
  - Matched organoid clustering and cell-typing
  - Correlation distances for genotype and transcriptional cell state
  - Matched biopsy versus organoid malignant cell comparison
  - TME associations
  - Biopsy paracrine and autocrine subtype-specific factor analysis
  - Tumor phenotyping from mIF data

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.11.017.

ACKNOWLEDGMENTS

This work was funded by the Lustgarten Foundation Dedicated Laboratory (B.M.W. and A.J.A.), Dana-Farber Cancer Institute Hale Center for Pancreatic Cancer Research (A.J.A., B.M.W., and W.C.H.), the Doris Duke Charitable Foundation (A.J.A.), Pancreatic Cancer Action Network (A.J.A. and B.M.W.), NIH-NCI K08 CA218420-02 (A.J.A.), P50 CA127003 (A.J.A., B.M.W., and G.I.S.), U01 CA176058 (W.C.H.), U01 CA224146 (W.C.H. and A.J.A.), U01 CA210171 (B.M.W.), U01 28020510 (W.C.H. and A.K.S.), 1U2CCA23319501 (A.K.S.), U54 CA217377 (W.C.H., S.R.M., and A.K.S.), U01 CA250549 (W.C.H., S.R., and A.J.A.), Stand Up To Cancer (B.M.W.), Noble Effort Fund (B.M.W.), Wexler Family Fund (B.M.W.), Promises for Purple (B.M.W.), Hope Funds for Cancer Research Postdoctoral Fellowship (S.R.), Harvard Catalyst/The Harvard Clinical and Translational Science Center UL1 TR002541 (S.R.), Koch Institute for Integrative Cancer Research at MIT and the Dana-Farber/Harvard Cancer Center Bridge Project (W.C.H., S.R.M., and S.R.), Dana-Farber Cancer Institute Gloria Spivak Faculty Advancement Fund (S.R.), Hopper-Belmont Foundation Inspiration Award (S.R.), Marcotte Center for Cancer Research (B.E.J.), Finnish Cultural Foundation (S.A.V.), Orion Research Foundation sr (S.A.V.), Damon Runyon Cancer Research Foundation (A.M.J.), NIH K99 Award CA241072-01 (A.M.J.), Koch Institute Cancer Core Grant NIH P30 CA14051 (T.E.J. and A.M.J.), Broman Family Fund for Pancreatic Cancer Research (K.N.), Ludwig Center for Molecular Oncology at MIT (A.W.N.), Searle Scholars Program (A.K.S.), Beckman Young Investigator Program (A.K.S.), Sloan Fellowship in Chemistry (A.K.S.), and the Pew-Stewart Scholars Program for Cancer Research (A.K.S.). This work was supported by the National Cancer Institute's Office of Cancer Genomics Cancer Target Discovery and Development (CTD^2) initiative. We also thank SciStories for assistance with schematics and Bryan Bryson for helpful discussions of macrophage phenotypes.

AUTHOR CONTRIBUTIONS

Conceptualization, S.R., P.S.W., B.M.W., W.C.H., A.J.A., A.K.S.; Methodology, S.R., P.S.W., A.W.N., H.L.W., L.C., J.A.N., B.M.W., W.C.H., A.J.A., A.K.S.; Validation, S.R., P.S.W., A.W.N., H.L.W., K.E.L., J.A.N., B.M.W., W.C.H., A.J.A., A.K.S.; Formal analysis, S.R., P.S.W., A.W.N., H.L.W., A.D., K.S.K., A.A.B., N.L., L.F.S., Y.Y.L., A.D. Cherniack, M.A.B., E.F.C., M.Y.T., J.M.M., L.C.; Investigation, S.R., P.S.W., A.W.N., H.L.W., K.E.L., J.G.-R., R.L.K., N.M., M.S.R., R.W.S.N., J.W., I.W.; Resources, S.A.V., A.D. Costa, E.K.H., D.Y.R., L.K.B., A.R., B.E.J., E.T.S., R.J.S., G.I.S, T.E.C., K.P., D.A.R., K.N., J.M.C., J.A.N., B.M.W.; Data Curation, P.S.W., S.R., A.W.N., H.L.W., K.S.K.; Writing – original draft, S.R., P.S.W., A.W.N., H.L.W.; Writing – review & editing, S.R., P.S.W., A.W.N., H.L.W., A.M.J., L.C., J.A.N., S.R.M., B.M.W., W.C.H., A.J.A., A.K.S.; Visualization, S.R., P.S.W., A.W.N., H.L.W.; Supervision, B.E.J., T.E.J., L.C., J.A.N., S.R.M., B.M.W., W.C.H., A.J.A., A.K.S.; Project administration, S.R., P.S.W., B.M.W., W.C.H., A.J.A., A.K.S.; Funding acquisition, S.R., P.S.W., S.R.M., B.M.W., W.C.H., A.J.A., A.K.S.

DECLARATION OF INTERESTS

B.M.W. reports research support from Celgene and Eli Lilly and consulting for BioLineRx, Celgene, G1 Therapeutics, and GRAIL. W.C.H. is a consultant for Thermo Fisher, Solasta Ventures, MPM Capital, Tyra Biosciences, iTeos, Frontier Medicines, Function Oncology, KSQ Therapeutics, Jubilant Therapeutics, RAPPTA Therapeutics, and Paraxel. A.K.S. reports compensation for consulting and/or SAB membership from Merck, Honeycomb Biotechnologies, Cellarity, Repertoire Immune Medicines, Ochre Bio, Third Rock Ventures, Hovione, Relation Therapeutics, FL82, and Dahlia Biosciences. A.J.A. has consulted for Oncorus, Arrakis Therapeutics, and Merck & Co. and has research funding from Mirati Therapeutics, Syros, Deerfield, and Novo Ventures that are unrelated to this project. S.R.M. is a founder of Travera. S.R. has equity in Amgen. J.M.C. has received research funding from Merck, Tesaro, AstraZeneca, Bayer, and Esperas Pharma; has served as a consultant to Bristol Myers Squibb; and has received travel funding from Bristol Myers Squibb. A.R. is an employee of AstraZeneca and an equity holder in Celsius Therapeutics and NucleAI. Y.Y.L. reports equity from g.Root Biomedical Services. A.D. Cherniack reports research support from Bayer. R.J.S reports research support from Merck and consulting and/or SAB membership for Bristol Myers Squibb, Merck, Novartis, and Pfizer. G.I.S. reports sponsored research support from Eli Lilly, Merck KGaA/EMD-Serono, Merck, and Sierra Oncology and has served on advisory boards for Pfizer, Eli Lilly, G1 Therapeutics, Roche, Merck KGaA/EMD-Serono, Sierra Oncology, Bicycle Therapeutics, Fusion Pharmaceuticals, Cybrexa Therapeutics, Astex, Almac, Ipsen, Bayer, Angiex, Daiichi Sankyo, Seattle Genetics, Boehringer Ingelheim, ImmunoMet, Asana, Artios, Atrin, Concarlo Holdings, Syros, Zentalis, CytomX Therapeutics, Blueprint Medicines, and Kymera Therapeutics. T.J. is a member of the Board of Directors of Amgen and Thermo Fisher Scientific; a co-Founder of Dragonfly Therapeutics and T2 Biosystems; serves on the Scientific Advisory Board of Dragonfly Therapeutics, SQZ Biotech, and Skyhawk Therapeutics; and is the President of Break Through Cancer. None of these affiliations represent a conflict of interest with respect to the design or execution of this study or interpretation of data presented in this manuscript. T.J.'s laboratory currently also receives funding from the Johnson & Johnson Lung Cancer Initiative and The Lustgarten Foundation for Pancreatic Cancer Research, but this funding did not support the research described in this manuscript. K.N. reports research support from Revolution Medicines, Evergrande Group, Pharmavite, and Janssen; is a member of SABs for SeaGen, BiomX, and Bicara Therapeutics; and has consulted for X-Biotix Therapeutics and Redesign Health. A.K.S., P.S.W., A.W.N., S.R., W.C.H., A.J.A., B.W.M., and J.G.R. have filed a patent (Publication number WO/2021/081491; International application number PCT/US2020/057342) related to this work.

INCLUSION AND DIVERSITY

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure ethnic or other types of diversity in the recruitment of human

subjects. We worked to ensure diversity in experimental samples through the selection of the cell lines. We worked to ensure diversity in experimental samples through the selection of the genomic datasets. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: March 12, 2021
Revised: September 24, 2021
Accepted: November 11, 2021
Published: December 9, 2021

REFERENCES

Abo, R.P., Ducar, M., Garcia, E.P., Thorner, A.R., Rojas-Rudilla, V., Lin, L., Sholl, L.M., Hahn, W.C., Meyerson, M., Lindeman, N.I., et al. (2015). BreaKmer: detection of structural variation in targeted massively parallel sequencing data using kmers. Nucleic Acids Res. *43*, e19.

Aguirre, A.J., Nowak, J.A., Camarda, N.D., Moffitt, R.A., Ghazani, A.A., Hazar-Rethinam, M., Raghavan, S., Kim, J., Brais, L.K., Ragon, D., et al. (2018). Real-time Genomic Characterization of Advanced Pancreatic Cancer to Enable Precision Medicine. Cancer Discov. *8*, 1096–1111.

Alonso-Curbelo, D., Ho, Y.J., Burdziak, C., Maag, J.L.V., Morris, J.P., 4th, Chandwani, R., Chen, H.A., Tsanov, K.M., Barriga, F.M., Luan, W., et al. (2021). A gene-environment-induced epigenetic program initiates tumorigenesis. Nature *590*, 642–648.

Aung, K.L., Fischer, S.E., Denroche, R.E., Jang, G.H., Dodd, A., Creighton, S., Southwood, B., Liang, S.B., Chadwick, D., Zhang, A., et al. (2018). Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. Clin. Cancer Res. *24*, 1344–1354.

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., Quinn, M.C., et al.; Australian Pancreatic Cancer Genome Initiative (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. Nature *531*, 47–52.

Balachandran, V.P., Beatty, G.L., and Dougan, S.K. (2019). Broadening the Impact of Immunotherapy to Pancreatic Cancer: Challenges and Opportunities. Gastroenterology *156*, 2056–2072.

Bartoschek, M., Oskolkov, N., Bocci, M., Lövrot, J., Larsson, C., Sommarin, M., Madsen, C.D., Lindgren, D., Pekar, G., Karlsson, G., et al. (2018). Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. Nat. Commun. *9*, 5150.

Ben-David, U., Siranosian, B., Ha, G., Tang, H., Oren, Y., Hinohara, K., Strathdee, C.A., Dempster, J., Lyons, N.J., Burns, R., et al. (2018). Genetic and transcriptional evolution alters cancer cell line drug response. Nature *560*, 325–330.

Ben-David, U., Beroukhim, R., and Golub, T.R. (2019). Genomic evolution of cancer models: perils and opportunities. Nat. Rev. Cancer *19*, 97–109.

Bernard, V., Semaan, A., Huang, J., San Lucas, F.A., Mulu, F.C., Stephens, B.M., Guerrero, P.A., Huang, Y., Zhao, J., Kamyabi, N., et al. (2019). Single-Cell Transcriptomics of Pancreatic Cancer Precursors Demonstrates Epithelial and Microenvironmental Heterogeneity as an Early Event in Neoplastic Progression. Clin. Cancer Res. *25*, 2194–2205.

Bi, W.L., Greenwald, N.F., Ramkissoon, S.H., Abedalthagafi, M., Coy, S.M., Ligon, K.L., Mei, Y., MacConaill, L., Ducar, M., Min, L., et al. (2017). Clinical Identification of Oncogenic Drivers and Copy-Number Alterations in Pituitary Tumors. Endocrinology *158*, 2284–2291.

Boj, S.F., Hwang, C.I., Baker, L.A., Chio, I.I., Engle, D.D., Corbo, V., Jager, M., Ponz-Sarvise, M., Tiriac, H., Spector, M.S., et al. (2015). Organoid models of human and mouse ductal pancreatic cancer. Cell *160*, 324–338.

Brunton, H., Caligiuri, G., Cunningham, R., Upstill-Goddard, R., Bailey, U.M., Garner, I.M., Nourse, C., Dreyer, S., Jones, M., Moran-Jones, K., et al.; Glasgow Precision Oncology Laboratory; Australian Pancreatic Cancer Genome Initiative (2020). HNF4A and GATA6 Loss Reveals Therapeutically Actionable Subtypes in Pancreatic Cancer. Cell Rep. *31*, 107625.

Byrne, A.T., Alférez, D.G., Amant, F., Annibali, D., Arribas, J., Biankin, A.V., Bruna, A., Budinská, E., Caldas, C., Chang, D.K., et al. (2017). Interrogating open issues in cancer medicine with patient-derived xenografts. Nat. Rev. Cancer *17*, 632.

Cancer Genome Atlas Research Network (2017). Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. Cancer Cell *32*, 185–203.e13.

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.

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.

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.

Connor, A.A., Denroche, R.E., Jang, G.H., Lemire, M., Zhang, A., Chan-Seng-Yue, M., Wilson, G., Grant, R.C., Merico, D., Lungu, I., et al. (2019). Integration of Genomic and Transcriptional Features in Pancreatic Cancer Reveals Increased Cell Cycle Progression in Metastases. Cancer Cell *35*, 267–282 e267.

Di Carlo, S.E., and Peduto, L. (2018). The perivascular origin of pathological fibroblasts. J. Clin. Invest. *128*, 54–63.

Drost, J., and Clevers, H. (2018). Organoids in cancer research. Nat. Rev. Cancer *18*, 407–418.

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.

Fan, J., Lee, H.O., Lee, S., Ryu, D.E., Lee, S., Xue, C., Kim, S.J., Kim, K., Barkas, N., Park, P.J., et al. (2018). Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. Genome Res. *28*, 1217–1227.

Filbin, M.G., Tirosh, I., Hovestadt, V., Shaw, M.L., Escalante, L.E., Mathewson, N.D., Neftel, C., Frank, N., Pelton, K., Hebert, C.M., et al. (2018). Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. Science *360*, 331–335.

Fujii, M., Matano, M., Toshimitsu, K., Takano, A., Mikami, Y., Nishikori, S., Sugimoto, S., and Sato, T. (2018). Human Intestinal Organoids Maintain Self-Renewal Capacity and Cellular Diversity in Niche-Inspired Culture Condition. Cell Stem Cell *23*, 787–793 e786.

Garcia, E.P., Minkovsky, A., Jia, Y., Ducar, M.D., Shivdasani, P., Gong, X., Ligon, A.H., Sholl, L.M., Kuo, F.C., MacConaill, L.E., et al. (2017). Validation of OncoPanel: A Targeted Next-Generation Sequencing Assay for the Detection of Somatic Variants in Cancer. Arch. Pathol. Lab. Med. *141*, 751–758.

Garraway, L.A., and Lander, E.S. (2013). Lessons from the cancer genome. Cell *153*, 17–37.

Ghandi, M., Huang, F.W., Jané-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., 3rd, Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., et al. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature *569*, 503–508.

Gierahn, T.M., Wadsworth, M.H., 2nd, Hughes, T.K., Bryson, B.D., Butler, A., Satija, R., Fortune, S., Love, J.C., and Shalek, A.K. (2017). Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat. Methods *14*, 395–398.

Gillet, J.P., Varma, S., and Gottesman, M.M. (2013). The clinical relevance of cancer cell lines. J. Natl. Cancer Inst. *105*, 452–458.

Gröger, C.J., Grubinger, M., Waldhör, T., Vierlinger, K., and Mikulits, W. (2012). Meta-analysis of gene expression signatures defining the epithelial to mesenchymal transition during cancer progression. PLoS ONE *7*, e51136.

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.

Hafner, M., Niepel, M., Chung, M., and Sorger, P.K. (2016). Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nat. Methods *13*, 521–527.

Hahn, W.C., Bader, J.S., Braun, T.P., Califano, A., Clemons, P.A., Druker, B.J., Ewald, A.J., Fu, H., Jagu, S., Kemp, C.J., et al.; Cancer Target Discovery and Development Network (2021). An expanded universe of cancer targets. Cell *184*, 1142–1155.

Hara, T., Chanoch-Myers, R., Mathewson, N.D., Myskiw, C., Atta, L., Bussema, L., Eichhorn, S.W., Greenwald, A.C., Kinker, G.S., Rodman, C., et al. (2021). Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. Cancer Cell *39*, 779–792 e711.

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. Nature Cancer *1*, 59–74.

Ho, W.J., Jaffee, E.M., and Zheng, L. (2020). The tumour microenvironment in pancreatic cancer - clinical challenges and opportunities. Nat. Rev. Clin. Oncol. *17*, 527–540.

Hosaka, K., Yang, Y., Seki, T., Fischer, C., Dubey, O., Fredlund, E., Hartman, J., Religa, P., Morikawa, H., Ishii, Y., et al. (2016). Pericyte-fibroblast transition promotes tumor growth and metastasis. Proc. Natl. Acad. Sci. USA *113*, E5618–E5627.

Hovestadt, V., Smith, K.S., Bihannic, L., Filbin, M.G., Shaw, M.L., Baumgartner, A., DeWitt, J.C., Groves, A., Mayr, L., Weisman, H.R., et al. (2019). Resolving medulloblastoma cellular architecture by single-cell genomics. Nature *572*, 74–79.

Hughes, T.K., Wadsworth, M.H., 2nd, Gierahn, T.M., Do, T., Weiss, D., Andrade, P.R., Ma, F., de Andrade Silva, B.J., Shao, S., Tsoi, L.C., et al. (2020). Second-Strand Synthesis-Based Massively Parallel scRNA-Seq Reveals Cellular States and Molecular Features of Human Inflammatory Skin Pathologies. Immunity *53*, 878–894 e877.

Hyman, D.M., Taylor, B.S., and Baselga, J. (2017). Implementing Genome-Driven Oncology. Cell *168*, 584–599.

Kim, J.H., Park, S.Y., Jun, Y., Kim, J.Y., and Nam, J.S. (2017). Roles of Wnt Target Genes in the Journey of Cancer Stem Cells. Int. J. Mol. Sci. *18*, 1604.

Kim, C., Gao, R., Sei, E., Brandt, R., Hartman, J., Hatschek, T., Crosetto, N., Foukakis, T., and Navin, N.E. (2018). Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell *173*, 879–893 e813.

Li, B., Gould, J., Yang, Y., Sarkizova, S., Tabaka, M., Ashenberg, O., Rosen, Y., Slyper, M., Kowalczyk, M.S., Villani, A.C., et al. (2020). Cumulus provides cloud-based data analysis for large-scale single-cell and single-nucleus RNA-seq. Nat. Methods *17*, 793–798.

Li, Y., He, Y., Peng, J., Su, Z., Li, Z., Zhang, B., Ma, J., Zhuo, M., Zou, D., Liu, X., et al. (2021). Mutant Kras co-opts a proto-oncogenic enhancer network in inflammation-induced metaplastic progenitor cells to initiate pancreatic cancer. Nature Cancer *2*, 49–65.

Ligorio, M., Sil, S., Malagon-Lopez, J., Nieman, L.T., Misale, S., Di Pilato, M., Ebright, R.Y., Karabacak, M.N., Kulkarni, A.S., Liu, A., et al. (2019). Stromal Microenvironment Shapes the Intratumoral Architecture of Pancreatic Cancer. Cell *178*, 160–175 e127.

Longo, S.K., Guo, M.G., Ji, A.L., and Khavari, P.A. (2021). Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. Nat. Rev. Genet. *22*, 627–644.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Miyabayashi, K., Baker, L.A., Deschênes, A., Traub, B., Caligiuri, G., Plenker, D., Alagesan, B., Belleau, P., Li, S., Kendall, J., et al. (2020). Intraductal Transplantation Models of Human Pancreatic Ductal Adenocarcinoma Reveal Progressive Transition of Molecular Subtypes. Cancer Discov. *10*, 1566–1589.

Moffitt, R.A., Marayati, R., Flate, E.L., Volmar, K.E., Loeza, S.G., 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.

Nam, A.S., Chaligne, R., and Landau, D.A. (2021). Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. Nat. Rev. Genet. *22*, 3–18.

Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R., Silverbush, D., Shaw, M.L., Hebert, C.M., et al. (2019). An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell *178*, 835–849 e821.

O'Kane, G.M., Grunwald, B.T., Jang, G.H., Masoomian, M., Picardo, S., Grant, R.C., Denroche, R.E., Zhang, A., Wang, Y., Lam, B., et al. (2020). GATA6 Expression Distinguishes Classical and Basal-like Subtypes in Advanced Pancreatic Cancer. Clin Cancer Res. *26*, 4901–4910.

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.

Palmer, A.C., and Sorger, P.K. (2017). Combination Cancer Therapy Can Confer Benefit via Patient-to-Patient Variability without Drug Additivity or Synergy. Cell *171*, 1678–1691 e1613.

Palmer, A.C., Chidley, C., and Sorger, P.K. (2019). A curative combination cancer therapy achieves high fractional cell killing through low cross-resistance and drug additivity. eLife *8*, e50036.

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.

Pelon, F., Bourachot, B., Kieffer, Y., Magagna, I., Mermet-Meillon, F., Bonnet, I., Costa, A., Givel, A.M., Attieh, Y., Barbazan, J., et al. (2020). Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms. Nat. Commun. *11*, 404.

Porter, R.L., Magnus, N.K.C., Thapar, V., Morris, R., Szabolcs, A., Neyaz, A., Kulkarni, A.S., Tai, E., Chougule, A., Hillis, A., et al. (2019). Epithelial to mesenchymal plasticity and differential response to therapies in pancreatic ductal adenocarcinoma. Proc. Natl. Acad. Sci. USA *116*, 26835–26845.

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 e1624.

Qadir, M.M.F., Álvarez-Cubela, S., Klein, D., van Dijk, J., Muñiz-Anquela, R., Moreno-Hernández, Y.B., Lanzoni, G., Sadiq, S., Navarro-Rubio, B., García, M.T., et al. (2020). Single-cell resolution analysis of the human pancreatic ductal progenitor cell niche. Proc. Natl. Acad. Sci. USA *117*, 10876–10887.

Sade-Feldman, M., Yizhak, K., Bjorgaard, S.L., Ray, J.P., de Boer, C.G., Jenkins, R.W., Lieb, D.J., Chen, J.H., Frederick, D.T., Barzily-Rokni, M., et al. (2019). Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. Cell *176*, 404.

Sahai, E., Astsaturov, I., Cukierman, E., DeNardo, D.G., Egeblad, M., Evans, R.M., Fearon, D., Greten, F.R., Hingorani, S.R., Hunter, T., et al. (2020). A framework for advancing our understanding of cancer-associated fibroblasts. Nat. Rev. Cancer *20*, 174–186.

Sholl, L.M., Do, K., Shivdasani, P., Cerami, E., Dubuc, A.M., Kuo, F.C., Garcia, E.P., Jia, Y., Davineni, P., Abo, R.P., et al. (2016). Institutional implementation of clinical tumor profiling on an unselected cancer population. JCI Insight *1*, e87062.

Somerville, T.D.D., Xu, Y., Miyabayashi, K., Tiriac, H., Cleary, C.R., Maia-Silva, D., Milazzo, J.P., Tuveson, D.A., and Vakoc, C.R. (2018). TP63-Mediated

Enhancer Reprogramming Drives the Squamous Subtype of Pancreatic Ductal Adenocarcinoma. Cell Rep. *25*, 1741–1755 e1747.

Spurr, L.F., Touat, M., Taylor, A.M., Dubuc, A.M., Shih, J., Meredith, D.M., Pisano, W.V., Meyerson, M.L., Ligon, K.L., Cherniack, A.D., et al. (2021). Quantification of aneuploidy in targeted sequencing data using ASCETS. Bioinformatics *37*, 2461–2463.

Suvà, M.L., and Tirosh, I. (2019). Single-Cell RNA Sequencing in Cancer: Lessons Learned and Emerging Challenges. Mol. Cell *75*, 7–12.

Tiriac, H., Belleau, P., Engle, D.D., Plenker, D., Deschênes, A., Somerville, T.D.D., Froeling, F.E.M., Burkhart, R.A., Denroche, R.E., Jang, G.H., et al. (2018). Organoid Profiling Identifies Common Responders to Chemotherapy in Pancreatic Cancer. Cancer Discov. *8*, 1112–1129.

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016a). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196.

Tirosh, I., Venteicher, A.S., Hebert, C., Escalante, L.E., Patel, A.P., Yizhak, K., Fisher, J.M., Rodman, C., Mount, C., Filbin, M.G., et al. (2016b). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature *539*, 309–313.

van de Haar, J., Hoes, L.R., Roepman, P., Lolkema, M.P., Verheul, H.M.W., Gelderblom, H., de Langen, A.J., Smit, E.F., Cuppen, E., Wessels, L.F.A., and Voest, E.E. (2021). Limited evolution of the actionable metastatic cancer genome under therapeutic pressure. Nat. Med. *27*, 1553–1563.

van Galen, P., Hovestadt, V., Wadsworth Ii, M.H., Hughes, T.K., Griffin, G.K., Battaglia, S., Verga, J.A., Stephansky, J., Pastika, T.J., Lombardi Story, J., et al. (2019). Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. Cell *176*, 1265–1281 e1224.

Venteicher, A.S., Tirosh, I., Hebert, C., Yizhak, K., Neftel, C., Filbin, M.G., Hovestadt, V., Escalante, L.E., Shaw, M.L., Rodman, C., et al. (2017). Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. Science *355*, eaai8478.

Vivian, J., Rao, A.A., Nothaft, F.A., Ketchum, C., Armstrong, J., Novak, A., Pfeil, J., Narkizian, J., Deran, A.D., Musselman-Brown, A., et al. (2017). Toil enables reproducible, open source, big biomedical data analyses. Nat. Biotechnol. *35*, 314–316.

Wagner, D.E., and Klein, A.M. (2020). Lineage tracing meets single-cell omics: opportunities and challenges. Nat. Rev. Genet. *21*, 410–427.

Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M.; Cancer Genome Atlas Research Network (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet. *45*, 1113–1120.

Wöll, S., Schlitter, A.M., Dhaene, K., Roller, M., Esposito, I., Sahin, U., and Türeci, Ö. (2014). Claudin 18.2 is a target for IMAB362 antibody in pancreatic neoplasms. Int. J. Cancer *134*, 731–739.

Zhang, L., Li, Z., Skrzypczynska, K.M., Fang, Q., Zhang, W., O'Brien, S.A., He, Y., Wang, L., Zhang, Q., Kim, A., et al. (2020). Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. Cell *181*, 442–459 e429.

Zilionis, R., Engblom, C., Pfirschke, C., Savova, V., Zemmour, D., Saatcioglu, H.D., Krishnan, I., Maroni, G., Meyerovitz, C.V., Kerwin, C.M., et al. (2019). Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. Immunity *50*, 1317–1334 e1310.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-cytokeratin 17, clone E3	Thermo Fisher	Cat# MA5-13539; RRID:AB_10980102
Rabbit monoclonal anti-S100A2, clone EPR5392	Abcam	Cat# ab109494; RRID:AB_10859000
Rabbit monoclonal anti-Claudin18.2, clone EPR19202-244	Abcam	Cat# ab241330
Rabbit monoclonal anti-GATA-6 XP, clone D61E4	Cell Signaling Technology	Cat# 5851; RRID:AB_10705521
Rabbit monoclonal anti-TFF1 (estrogen inducible protein pS2), clone EPR3972	Abcam	Cat# ab92377; RRID:AB_10562112
Mouse monoclonal anti-cytokeratin, clone AE1/AE3	Agilent/Dako	Cat# M3515; RRID:AB_2132885
Mouse monoclonal anti-pan-keratin, clone C11	Cell Signaling Technology	Cat# 4545; RRID:AB_490860
Opal polymer HRP anti-mouse and anti-rabbit secondary antibody	Akoya Biosciences	Cat# ARH1001EA; RRID:AB_2890927
Biological samples		
Human PDAC samples	This study	N/A
Chemicals, peptides, and recombinant proteins		
Advanced DMEM/F12	Thermo Fisher	Cat# 12634028
RPMI 1640	Corning	Cat# 10-040-CV
Fetal bovine serum	Sigma	Cat# F4135
Penicillin/streptomycin	Thermo Fisher	Cat# 15140122
Primocin	Invivogen	Cat# ant-pm-1
HEPES	Thermo Fisher	Cat# 15630080
GlutaMAX	Thermo Fisher	Cat# 35050061
A83-01	Tocris	Cat# 2939
Recombinant mouse Noggin	Peprotech	Cat# 250-38E
Recombinant mouse EGF	Peprotech	Cat# 315-09
Recombinant human FGF10	Peprotech	Cat# 100-26
Human [Leu15]-Gastrin I	Sigma	Cat# G9145
N-acetylcysteine	Sigma	Cat# A9165
Nicotinamide	Sigma	Cat# N0636
B-27 supplement	Thermo Fisher	Cat# 17504044
Recombinant human TGF-β1	Peprotech	Cat# 100-21
Recombinant human IFNγ	Peprotech	Cat# 300-02
Growth factor reduced Matrigel	Corning	Cat# 356231
TrypLE Express	Thermo Fisher	Cat# 12604054
Trypsin-EDTA (0.25%)	Thermo Fisher	Cat# 25200056
CellTiterGlo 3D	Promega	Cat# G9683
Collagenase XI	Sigma	Cat# C7657
Dnase	StemCell Technologies	Cat# 07900
Y-27632	Selleck Chemicals	Cat# S1049
ACK lysing biffer	Thermo Fisher	Cat# A1049201
Trypan blue solution, 0.4%	Thermo Fisher	Cat# 15250061

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Spectral DAPI	Akoya Biosciences	Cat# FP1490
Opal 520 reagent pack	Akoya Biosciences	Cat# FP1487001KT
Opal 540 reagent pack	Akoya Biosciences	Cat# FP1494001KT
Opal 570 reagent pack	Akoya Biosciences	Cat# FP1488001KT
Opal 620 reagent pack	Akoya Biosciences	Cat# FP1495001KT
Opal 650 reagent pack	Akoya Biosciences	Cat# FP1496001KT
Opal 690 reagent pack	Akoya Biosciences	Cat# FP1497001KT
Xylenes (histological)	Fisher Scientific, X3P1GAL	Cat# X3P-1GAL
BOND Epitope Retrieval Solution 1	Leica Biosystems	Cat# AR9961
BOND Epitope Retrieval Solution 2	Leica Biosystems	Cat# AR9640
Antibody diluent/block	Akoya Biosciences	Cat# ARD1001EA
1x Plus Automation Amplification Diluent	Akoya Biosciences	Cat# FP1609
ProLong Gold Antifade Mountant	Fisher Scientific/Molecular Probes	Cat# P36930
Gemcitabine	Selleck Chemicals	Cat# S1149
Paclitaxel	Selleck Chemicals	Cat# S1150
SN-38	Selleck Chemicals	Cat# S4908
Trametinib	Selleck Chemicals	Cat# S2673
5-FU	Selleck Chemicals	Cat# S1209
Afatinib	Selleck Chemicals	Cat# S1011
AZD6738 (Ceralasertib)	Selleck Chemicals	Cat# S7693
Binimetinib	Selleck Chemicals	Cat# S7007
BVD-523 (Ulixertinib)	Selleck Chemicals	Cat# S7854
Dinaciclib	Selleck Chemicals	Cat# S2768
Everolimus	Selleck Chemicals	Cat# S1120
Gedatolisib	Selleck Chemicals	Cat# S2628
GSK126	Selleck Chemicals	Cat# S7061
(+)-JQ1	Selleck Chemicals	Cat# S7110
LY3023414 (Samotolisib)	Selleck Chemicals	Cat# S8322
MK-1775 (Adavosertib)	Selleck Chemicals	Cat# S1525
Navitoclax	Selleck Chemicals	Cat# S1001
Olaparib	Selleck Chemicals	Cat# S1060
Oxaliplatin	Selleck Chemicals	Cat# S1224
Palbociclib	Selleck Chemicals	Cat# S4482
Prexasertib	Selleck Chemicals	Cat# S6385
SHP099	Selleck Chemicals	Cat# S8278
Tazemetostat	Selleck Chemicals	Cat# S7128
YKL-5-124	Selleck Chemicals	Cat# S8863
2-mercaptoethanol	Sigma	Cat# M3148
Buffer RLT	QIAGEN	Cat# 79216
Buffer RLT Plus	QIAGEN	Cat# 1053393
Deoxynucleotide (dNTP) solution mix	NewEngland BioLabs	Cat# N0447L
Superase.In RNase Inhibitor	Thermo Fisher	Cat# AM2696
Maxima H minus reverse transcriptase	Fisher Scientific	Cat# EP0753
AMPure XP beads	Beckman Coulter	Cat# A63881
Guanidinium thiocyanate	Thermo Fisher	Cat# AM9422
N-Lauroylsarcosine sodium salt solution (Sarkosyl NL)	Sigma	Cat# L7414
Exonuclease I	New England BioLabs	Cat# M0293S
Klenow Fragment	New England BioLabs	Cat# M0212L
Polycarbonate membrane filters 62x22	Fisher Scientific/Sterlitech Corporation	Cat# NC1421644

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MACOSKO-2011-10 mRNA Capture Beads	Fisher Scientific/ChemGenes	Cat# NC0927472
ERCC RNA spike-in mix	Thermo Fisher	Cat# 4456740
Critical commercial assays		
AllPrep DNA/RNA/miRNA Universal kit	QIAGEN	Cat# 80224
Nextera XT DNA Library Preparation Kit	Illumina	Cat# FC-131-1096
Nextseq 500/550 High output v2.5 kit (75 cycles)	Illumina	Cat# 20024906
NovaSeq 6000 S2 kit (100 cycles)	Illumina	Cat# 20012862
TruSeq Stranded mRNA Library Prep kit	Illumina	Cat# 20020595
Kapa HiFi HotStart ReadyMix	Kapa Biosystems	Cat# KK2602
KAPA HyperPrep kit (PCR-free)	Kapa Biosystems	Cat# KK8505
High Sensitivity D5000 ScreenTape	Agilent	Cat# 5067-5592
Qubit dsDNA High-Sensitivity kit	Thermo Fisher	Cat# Q32854
Quant-iT Ribogreen RNA Assay kit	Thermo Fisher	Cat# R11490
Quant-iT PicoGreen dsDNA Assay kit	Thermo Fisher	Cat# P11496
Deposited data		
Bulk and single-cell transcriptomic data from PDAC patient samples and organoid models	This study	https://singlecell.broadinstitute.org/single_cell/study/SCP1644 dbGaP: phs002712.v1.p1
Primary PDAC genomic and transcriptomic data (TCGA, Pancreatic Ductal Adenocarcinoma)	Cancer Genome Atlas Research Network, 2017	https://portal.gdc.cancer.gov/projects/TCGA-PAAD
Metastatic PDAC genomic and transcriptomic data (Panc-Seq)	Aguirre et al., 2018	dbGaP: phs001652.v1.p1
CCLE transcriptomic data	Ghandi et al., 2019	https://portals.broadinstitute.org/ccle
TCGA transcriptomic data (other malignancies)	Vivian et al., 2017	https://toil.xenahubs.net
Experimental models: Cell lines		
Human PDAC organoids	This study	N/A
CFPAC-1	ATCC	Cat# CRL-1918; RRID:CVCL_1119
L Wnt-3A cells for Wnt-3A conditioned medium	ATCC	Cat# CRL-2647; RRID:CVCL_0635
Cultrex 293T cells for R-spondin1 conditioned medium	Trevigen	Cat# 3710-001-K; RRID:CVCL_RU08
Oligonucleotides		
Seq-Well ISPCR: AAG CAG TGG TAT CAA CGC AGA GT	Integrated DNA Technologies	N/A
Custom Read 1 Primer: GCC TGT CCG CGG AAG CAG TGG TAT CAA CGC AGA GTA C	Integrated DNA Technologies	N/A
Seq-Well 5¢ TSO: AAG CAG TGG TAT CAA CGC AGA GTG AAT rGrGrG	Integrated DNA Technologies	N/A
Seq-Well Custom P5-SMART PCR hybrid oligo: AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG CCT GTC CGC GGA AGC AGT GGT ATC AAC GCA GAG TAC	Integrated DNA Technologies	N/A
Seq-Well dN-SMRT oligo: AAG CAG TGG TAT CAA CGC AGA GTG ANN NGG NNN B	Integrated DNA Technologies	N/A
Software and algorithms		
R project for statistical computing v3.5.1	R Core Team	https://www.r-project.org
R package – Seurat v2.3.4	GitHub	https://github.com/satijalab/seurat

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R package – Circlize v0.4.8	CRAN	https://CRAN.R-project.org/ package = circlize
R package – infercnv v0.99.4	GitHub	https://github.com/broadinstitute/inferCNV
R package – data.table v1.12.0	GitHub	https://github.com/Rdatatable/data.table
R package – ggplot2 v3.2.1	CRAN	https://CRAN.R-project.org/ package = ggplot2
R package – ComplexHeatmap v2.7.3	Bioconductor	https://bioconductor.org/packages/ComplexHeatmap/
R package – dplyr v1.0.7	CRAN	https://cran.r-project.org/web/packages/dplyr/
STAR	GitHub	https://github.com/alexdobin/STAR
Cumulus	Li et al., 2020	https://cumulus.readthedocs.io/
Broad Picard pipeline v1.90	GitHub	https://broadinstitute.github.io/picard/
Genome Analysis Toolkit (GATK) v1.6-5-g557da77 and v.4.1.6.0	Broad Institute	https://gatk.broadinstitute.org/hc/en-us
Python Programming Language v3.7.4	Python	https://www.python.org
Other		
Leica BOND RX Research Stainer	Leica Biosystems	https://www.leicabiosystems.com/ihc-ish-fish/fully-automated-ihc-ish-instruments/bond-rx/
Vectra 3.0 Automated Quantitative Imaging System	PerkinElmer/Akoya Biosciences	https://www.akoyabio.com/phenoptics/mantra-vectra-instruments/vectra-3-0/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be sent to and will be fulfilled by Dr. Alex K. Shalek (shalek@mit.edu).

Materials availability

Organoid models generated in this study are available upon request with a materials transfer agreement.

Data and code availability

De-identified single-cell RNA-seq data are publicly available for download and visualization via the Single Cell Portal: https://singlecell.broadinstitute.org/single_cell/study/SCP1644. Genomic and transcriptomic data will be available at the NCBI Database of Genotypes and Phenotypes (dbGaP). This paper analyzes existing, publicly available data. The links and accession numbers for these datasets are listed in the key resources table.

Code is available from the lead contact upon request.

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

Eligible participants were recruited from outpatient clinics and inpatient units at the Dana-Farber Cancer Institute and the Brigham and Women's Hospital. Investigators obtained written, informed consent from patients at least 18 years old with pancreatic cancer for Dana-Farber/Harvard Cancer Center Institutional Review Board (IRB)-approved protocols 11-104, 17-000, 03-189, and/or 14-408 for tissue collection, molecular analysis, and organoid generation. ScRNA-seq samples were collected from 23 patients between October 2018 and December 2020, both male (n = 13) and female (n = 10). Organoid samples for bulk genomic and transcriptomic analyses were collected between May 2015 and January 2018. Core needle biopsy specimens were collected and the first core was sent for pathologic analysis. One or more additional cores were then allocated for scRNA-seq and/or organoid generation. Clinical features of our patient cohort are included in **Table S1**.

Human PDAC patient-derived organoid models

Tissue samples were minced into small portions using a scalpel and then digested at 37°C for 15 min using digest medium that consisted of human complete organoid medium (see below), 1 mg/mL collagenase XI (Sigma Aldrich), 10 μg/mL DNase (Stem Cell Technologies), and 10 μM Y27632 (Selleck) (Tiriac et al., 2018). After dissociation, a portion of the cells from the fresh tumor specimen were allocated for scRNA-seq, and the remainder were initiated and maintained as patient-derived organoid cultures as previously described (Boj et al., 2015; Tiriac et al., 2018). In brief, digested cells were seeded in 3-dimensional (3D) Growth-factor Reduced Matrigel (Corning), fed with human complete organoid medium containing Advanced DMEM/F12 (GIBCO), 10 mM HEPES (GIBCO), 1x GlutaMAX (GIBCO), 500 nM A83-01 (Tocris), 50 ng/mL mEGF (Peprotech), 100 ng/mL mNoggin (Peprotech), 100 ng/mL hFGF10 (Peprotech), 10 nM hGastrin I (Sigma), 1.25 mM N-acetylcysteine (Sigma), 10 mM Nicotinamide (Sigma), 1x B27 supplement (GIBCO), RSPONDIN-1 conditioned media 10% final, WNT3A conditioned media 50% final, 100 U/mL penicillin/streptomycin (GIBCO), and 1x Primocin (Invivogen) (Table S5), and maintained at 37°C in 5% CO_2. 10 μM Y27632 (Selleck) was included in the culture medium of newly initiated samples until the first media exchange. For propagation, organoids were dissociated with TrypLE Express (GIBCO) before re-seeding into fresh Matrigel and culture medium.

After initial processing of fresh tissue specimens, we monitored samples closely for organoid growth. We did not passage organoids at set time intervals, as there was significant variability in the time needed to establish relatively robust growth of organoids (Figure 3D). Instead, we maintained early passage organoids until they reached relative confluence, and then passaged them at low split ratios (1:1, 1:1.5, or 1:2 dilutions) in complete organoid medium to promote continued growth. In one case, PANFR0489R, cells persisted as individuals and small organoids after initiation in complete organoid medium, but did not grow and expand cell numbers significantly. Approximately 15 weeks after initiation, we switched a portion of the surviving cells to organoid medium without A83-01 or mNoggin, and observed renewed growth of organoids under these media conditions but not of those that remained in complete organoid medium. Consequently, we expanded this sample in media without A83-01 or mNoggin, including performing early passage scRNA-seq. After several additional passages, once the organoids were robustly growing, we were able to transition this model back to complete organoid medium with no apparent change in growth rate, morphology, or transcriptional state. All other serially sampled organoids were maintained and assessed in complete medium except as indicated when specific media alterations or experimental perturbations were performed. The identify of organoid models was authenticated by comparison of their inferred CNV profiles with targeted genomic sequencing and CNV profiles of matched patient tissue and with inferred CNV profiles from patient tissue and earlier passage models in the case of samples serially assessed with scRNA-seq. The identify of cell line models was authenticated by short tandem repeat (STR) analysis. Cell line and organoid cultures were routinely tested for mycoplasma contamination.

METHOD DETAILS

Sample preparation for single-cell RNA-sequencing of clinical samples, organoids, and cell lines

Patient specimens were dissociated for paired scRNA-seq and organoid generation as described above. In our initial process optimization for fresh patient specimens, we found that dissociation times below 30 min, while not always completely digesting all biopsy material and potentially affecting the representation of difficult to dissociate cell types (e.g., fibroblasts), resulted in greater cell viability and improved RNA quality downstream. After tissue dissociation, cells were washed, treated with ACK lysing buffer (Thermo Fisher) to lyse red blood cells, washed again, and counted using a hemocytometer with 0.4% Trypan blue (Thermo Fisher) added at 1:1 dilution for viability assessment. We allowed residual tissue chunks to settle before selecting a predominance of single cells for counting and Seq-Well processing. We allocated between 10,000 and 15,000 viable cells per Seq-Well array, and where possible we prepared two arrays per sample. Most samples were processed and loaded onto Seq-Well arrays within 2-3 h of biopsy acquisition. Remaining cells and tissue chunks were allocated for patient-matched organoid generation.

For scRNA-seq of organoid samples and cell lines (CFPAC1), we passaged models and allowed them to grow for 6 days before dissociating to single cells (organoids – TrypLE Express, Thermo Fisher; cell lines – 0.25% Trypsin-EDTA, Thermo Fisher), counting, and allocating 15,000 viable cells for Seq-Well. By standardizing the collection of organoid scRNA-seq samples at 6 days after passaging, we tried to minimize bias arising from cell cycle differences in samples at different degrees of confluence.

Assessing organoid and cell line transcriptional states under different matrix and media conditions

For adaptation of patient-derived organoids onto 2-dimensional (2D) culture surfaces as patient-derived cell lines, tissue culture plates were pre-coated with 100 μg/mL Matrigel dissolved in basal media for 2 h at 37°C before washing with PBS. Established organoid models were dissociated and seeded onto these Matrigel-coated culture wells in complete organoid media. In parallel, a portion of these passage-matched organoid cells were re-seeded into Matrigel droplets as above. Cells were cultured in both matrix conditions in complete organoid media until they were confluent, approximately 2-3 weeks. Cells were collected and lysed using Trizol before snap freezing. RNA was isolated and purified as described below (“Bulk RNA-sequencing of organoids” section) using chloroform extraction, aqueous phase isolation, and processing using the QIAGEN AllPrep DNA/RNA/miRNA Universal kit before being submitted for sequencing.

For scRNA-seq assessment of organoid cell states when cultured under different media conditions, established organoid models were passaged as above by dissociating and reseeding into Matrigel droplets. A portion of the cells were cultured with complete organoid media (“Complete media”), while a distinct portion of passage-matched cells were cultured in “Minimal” media, which

consisted of Advanced DMEM/F12 (Thermo Fisher), 10 mM HEPES (Thermo Fisher), 1x GlutaMAX (Thermo Fisher), 100 U/mL penicillin/streptomycin (Thermo Fisher), and 1x Primocin (Invivogen) (Table S5). Cells were cultured for 6 days before being collected, dissociated, and aliquoted for scRNA-seq. Images were taken with an Olympus XM10 camera mounted to an Olympus CKX41 microscope 1 day after seeding and again after 11 days in culture to assess organoid growth in both conditions. The portion of cells cultured in minimal media were maintained in the same conditions for a longer duration and harvested again for scRNA-seq at 27 days and 59 days after the initial introduction of minimal media. To mirror the standard scRNA-seq workflow, the cells harvested at the 27- and 59-day time points were collected 6 days after passaging.

In addition to the minimal media experiment, organoid cells were also cultured in standard cell line media (''RP10''), which contains RPMI-1640 (Thermo Fisher) and 100 U/mL penicillin/streptomycin (Thermo Fisher) with 10% fetal bovine serum (Sigma), or in reduced organoid media ''OWRNA,'' which consists of Advanced DMEM/F12 (Thermo Fisher), 10 mM HEPES (Thermo Fisher), 1x GlutaMAX (Thermo Fisher), 50 ng/mL mEGF (Peprotech), 100 ng/mL hFGF10 (Peprotech), 10 nM hGastrin I (Sigma), 1.25 mM N-acetylcysteine (Sigma), 10 mM Nicotinamide (Sigma), 1x B27 supplement (Thermo Fisher), 100 U/mL penicillin/streptomycin (Thermo Fisher), and 1x Primocin (Invivogen) (i.e., complete organoid medium with removal of WNT3A, RSPONDIN-1, NOGGIN, and A-8301; Table S5). Furthermore, OWRNA reduced organoid medium served as the baseline control medium when assessing the effect of specific factors (IFNGγ and TGF-β1) from the TME on malignant cell states. Cells were cultured for 6 days before being collected, dissociated, and aliquoted for scRNA-seq in each of the following conditions: RP10, OWRNA, OWRNA with 50 ng/mL IFNGγ (Peprotech), and OWRNA with 5 ng/mL TGF-β1 (Peprotech) (Table S5). The cells cultured under the IFNGγ and TGF-β1 conditions were maintained in culture and harvested again for scRNA-seq 38 days after being introduced to these new media conditions. For these longer duration time points, cells were again passaged 6 days before collecting for scRNA-seq.

For scRNA-seq assessment of transcriptional states of the established pancreatic cancer cell line CFPAC1 under different media conditions, CFPAC1 cells were cultured in parallel in either standard cell line medium RP10 or complete organoid medium for 6 days before being collected, dissociated, and aliquoted for scRNA-seq.

Compound sensitivity testing in cell lines and organoid models

Organoids were propagated in their respective media formulations for the indicated times (e.g., 3 weeks, 6 weeks) before performing drug sensitivity testing. After dissociation, cells were seeded into ultra-low attachment 384-well plates (Corning) at 1000 viable cells per well with 20 μL of their respective medium containing 10% Matrigel by volume. After 24 h, compounds were added to each well over 12-point dose curves along with DMSO controls using a Tecan D300e digital dispenser. A parallel untreated plate was harvested at this 24 h time point for growth rate correction. Drug- or DMSO-treated cells were cultured for 5 days before assessing for viability. Cell viability was measured by adding 20 uL of CellTiter-Glo 3D (Promega) to each well, incubating for 1 h at room temperature on a shaker, and measuring luminescence using an EnVision plate reader (PerkinElmer). Given plate well randomization, raw luminescence data were deconvoluted with an in-house Python script (Python v3.7.4). Each condition was tested in triplicate, and each dose point was normalized to DMSO controls to estimate relative viability. Growth rate correction was performed as previously described, with growth-rate adjusted dose response curves fit to a 3-parameter sigmoidal curve (Hafner et al., 2016). Each experiment was performed using cells cultured in paired basal (OWRNA+TGF-β1 or Minimal) and classical (OWRNA) media conditions, and independent experiments were performed as summarized in Figure S7A and Table S7. To quantify the relative sensitivity of basal versus classical models to a given compound, the difference in growth-rate adjusted areas over the curve (AOC) for organoids cultured in paired media conditions was calculated for each compound within each experiment, as illustrated in Figure 6G and Figure S7A.

CFPAC1 cells were cultured in standard cell line medium RP10 or in complete organoid medium for 2 or 5 weeks before seeding for drug testing. Cells were dissociated and seeded into 384-well tissue culture plates (Greiner Bio-One) at 400 viable cells per well with 20 μL of their respective medium. Addition of compounds and sample harvesting were as above, except cell viability was measured by adding 20 uL of CellTiter-Glo (Promega) to each well and incubating for 15 min at room temperature on a shaker before measuring luminescence. Data analysis and dose response curve fitting were performed as described above.

Single-cell RNA-seq (scRNA-seq) data library generation, sequencing, and alignment

ScRNA-seq processing followed the Seq-Well protocol, uniquely compatible with low-input samples (Gierahn et al., 2017; Hughes et al., 2020). Briefly, arrays were preloaded with RNA capture beads (Fisher Scientific/ChemGenes) and stored in quenching buffer until used. Prior to cell loading, arrays were resuspended in 5 mL RPMI-1640 medium (Thermo Fisher) with 10% fetal bovine serum (Sigma), hereafter referred to as RP10. After dissociation, single-cell suspensions were manually counted and diluted to 15,000 cells per 200 μL of RP10 when cell numbers allowed. Excess RP10 was aspirated from the array and cells were loaded onto arrays. Excess cells were washed off with PBS (4x5 mL), briefly left in RPMI (5 mL) and cell+bead pairs were sealed for 40 min at 37°C using a polycarbonate membrane (Fisher Scientific). Arrays were rocked in lysis buffer for 20 min and RNA was hybridized onto the beads for 40 min. Beads were removed and reverse transcription was performed overnight using Maxima H Minus Reverse Transcriptase (Fisher Scientific). Prior to sequencing, the beads underwent an exonuclease treatment (New England BioLabs) and second strand synthesis *en masse* followed by whole transcriptome amplification (WTA, Kapa Biosystems) in 1,500 bead reactions (50 μL). cDNA was isolated using Agencourt AMPure XP beads (Beckman Coulter) at 0.6X SPRI (solid-phase reversible immobilization) followed by a 1X SPRI and quantified using a Qubit dsDNA High Sensitivity assay kit (Thermo Fisher). Library preparation was performed using Nextera XT DNA tagmentation (Illumina FC-131-1096) and N700 and N500 indices specific to a given sample. Tagmented and

amplified sequences were purified with a 0.6X SPRI, and cDNA was quantified (Qubit dsDNA High sensitivity assay kit, Thermo Fisher) and the base pair distribution measured (High sensitivity D5000 screen tape, Agilent). cDNA was loaded onto either an Illumina Nextseq (75 Cycle NextSeq 500/550 v2.5 kit) or Novaseq (100 Cycle NovaSeq 6000 S2 kit) at 2.4 pM. Regardless of platform, the paired end read structure was 21 bases (cell barcode and UMI) by 50 bases (transcriptomic information) with an 8 base pair (bp) custom read one primer. The demultiplex and alignment protocol was followed as previously described (Macosko et al., 2015). While Novaseq data were directly output as FASTQs, Nextseq BCL files were converted to FASTQs using bcl2fastq2. The resultant Nextseq and Novaseq FASTQs were demultiplexed by sample based on Nextera N700 and N500 indices. Reads were then aligned to the hg19 transcriptome using the cumulus/dropseq_tools pipeline on Terra maintained by the Broad Institute using standard settings (Li et al., 2020).

Bulk RNA- and DNA-sequencing of organoids

RNA was obtained for bulk RNA-sequencing from established organoids using one of two approaches. Dissociated organoids were resuspended into cold Matrigel, added as droplets to tissue culture plates (Greiner BioOne), and allowed to polymerize for 30 min before addition of media. Organoids were grown for 14-21 days (until confluent) under these conditions with regular media changes. At the time of harvest, cells were washed with cold phosphate buffered saline (PBS) at 4°C, then lysed with Trizol (Invitrogen) before snap-freezing. To isolate RNA, we performed chloroform extraction with isolation of the aqueous phase before processing RNA as per protocols outlined in the QIAGEN AllPrep DNA/RNA/miRNA Universal kit.

In the second approach, used to obtain both RNA and DNA, dissociated organoids were resuspended in a solution of 10% Matrigel in complete organoid media (volume/volume) and cultured in ultra-low-attachment culture flasks (Corning). Organoids were grown for 14-21 days (until confluent) before pelleting, washing with cold PBS at 4°C until most Matrigel was dissipated, and then snap frozen. Cell pellets were homogenized using buffer RLT Plus (QIAGEN) and a Precellys homogenizer. Samples were then processed for both DNA extraction and RNA isolation as per the QIAGEN AllPrep DNA/RNA/miRNA Universal kit. Purified RNA and DNA were then submitted for sequencing by the Broad Institute Genomics Platform.

For bulk RNA-sequencing, total RNA was quantified using the Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher) and normalized to 5 ng/μL. Following plating, 2 μL of a 1:1000 dilution of ERCC RNA controls (Thermo Fisher) were spiked into each sample. An aliquot of 200 ng for each sample was transferred into library preparation which uses an automated variant of the Illumina TruSeq Stranded mRNA Sample Preparation Kit. This method preserves strand orientation of the RNA transcript, and uses oligo dT beads to select mRNA from the total RNA sample followed by heat fragmentation and cDNA synthesis from the RNA template. The resultant 400 bp cDNA then goes through dual-indexed library preparation: 'A' base addition, adaptor ligation using P7 adapters, and PCR enrichment using P5 adapters. After enrichment, the libraries were quantified using Quant-iT PicoGreen (1:200 dilution; Thermo Fisher). After normalizing samples to 5 ng/μL, the set was pooled and quantified using the KAPA Library Quantification Kit for Illumina Sequencing Platforms. The entire process was performed in 96-well format and all pipetting was done by either Agilent Bravo or Hamilton Starlet.

Pooled libraries were normalized to 2 nM and denatured using 0.1 N NaOH prior to sequencing. Flowcell cluster amplification and sequencing were performed according to the manufacturer's protocols using the NovaSeq 6000. Each run was a 101 bp paired-end with an eight-base index barcode read. Data were analyzed using the Broad Picard pipeline which includes de-multiplexing and data aggregation (https://broadinstitute.github.io/picard/). FASTQ files were then processed as described below (see Bulk RNA-sequencing analysis).

For whole genome sequencing, 350 ng of genomic DNA was fragmented using a Covaris focused-ultrasonicator targeting 385bp fragments followed by size selection using SPRI cleanup. Library preparation was performed using a KAPA HyperPrep without amplification kit (KAPA Biosystems) with palindromic forked adapters with unique 8-base index sequences embedded within the adaptor (Roche). Libraries were then quantified using quantitative PCR (kit purchased from KAPA Biosystems) with probes specific to the adaptor ends on an Agilent Bravo liquid handling platform. Libraries were normalized to 2.2 nM, pooled into 24-plexes, combined with NovaSeq Cluster Amp Reagents DPX1, DPX2, and DPX3, and loaded into single lanes of a NovaSeq 6000 S4 flowcell using a Hamilton Starlet Liquid Handling system. Cluster amplification and sequencing occurred utilizing sequencing-by-synthesis kits to produce 151bp paired-end reads. Output from Illumina software was processed by the Broad Picard pipeline (https://broadinstitute.github.io/picard/) to yield BAM files containing demultiplexed, aggregated aligned reads. BAM files were then processed as described below (see Mutation and CNV identification from bulk DNA-sequencing).

Multiplex immunofluorescence imaging

A multi-marker panel was developed to characterize malignant cell subtype in formalin-fixed paraffin-embedded (FFPE) 4 μm tissue sections using multiplex immunofluorescence (mIF). The panel comprises markers associated with either a basal (Keratin-17: Thermo Fisher MA513539 and S100A2: Abcam 109494) or classical (CLDN18.2: Abcam 241330, GATA6: Cell Signaling Technology 5851 and TFF1: Abcam 92377) subtype. Additionally, DAPI (Akoya Biosciences FP1490) was included for identification of nuclei and pan-cytokeratin (AE1/AE3: Dako M3515; C11: Cell Signaling Technology 4545) for identification of epithelial cells. Secondary Opal Polymer HRP anti-mouse and anti-rabbit antibody (Dako ARH1001EA), Tyramide signal amplification, and Opal fluorophores (Akoya Biosciences) were used to detect primary antibodies (Keratin-17, Opal 520; S100A2, Opal 650; GATA6, Opal 540; CLDN18.2, Opal 570; TFF1, Opal 690; panCK, Opal 620).

These specific mIF markers were chosen for several reasons. KRT17, S100A2, and TFF1 are included in the original basal and classical RNA gene signatures (Moffitt et al., 2015). GATA6 is also a classical marker in an extended RNA gene panel and has been reported to be a potential driver of the classical phenotype (Brunton et al., 2020; Moffitt et al., 2015; O'Kane et al., 2020). Single markers of S100A2 and GATA6 have also been used extensively in imaging experiments in past literature to mark cells in the basal or classical state, respectively (Aung et al., 2018; Chan-Seng-Yue et al., 2020; Miyabayashi et al., 2020; O'Kane et al., 2020; Somerville et al., 2018). CLDN18.2 has also been associated with classical phenotypes, and antibody-based therapies targeting CLDN18.2 have been developed and tested in PDAC (Wöll et al., 2014). Furthermore, the markers chosen for the mIF subtyping panel agreed with those selected as optimal markers to differentiate basal-like versus classical by an international panel of experts at a workshop on pancreatic cancer subtyping held at Memorial Sloan Kettering in 2019. Here, we used multiple markers for each state to provide greater confidence in cell state identification and to assess marker heterogeneity across our mIF cohort.

Prior to use in multiplex staining, primary antibodies were first optimized via immunohistochemistry on control tissue to confirm contextual specificity. Monoplex immunofluorescence and iterative multiplex fluorescent staining were then used to optimize staining order, antibody-fluorophore assignments and fluorophore concentrations. Multiplex staining was performed using a Leica BOND RX Research Stainer (Leica Biosystems, Buffalo, IL) with sequential cycles of antigen retrieval, protein blocking, primary antibody incubation, secondary antibody incubation, and fluorescent labeling. Overview images of stained slides were acquired at 10X magnification using a Vectra 3.0 Automated Quantitative Imaging System (Perkin Elmer, Waltham, MA) and regions of interest (ROIs) were selected for multispectral image acquisition at 20X. After unmixing using a spectral library of single-color references, each image was inspected to ensure uniform staining quality and adequate tumor representation.

QUANTIFICATION AND STATISTICAL ANALYSIS

Mutation and CNV identification from bulk DNA-sequencing

For targeted DNA-sequencing of clinical samples, next-generation sequencing using a custom-designed hybrid capture library preparation was performed on an Illumina HiSeq 2500 with 2x100 paired-end reads, as previously described (Garcia et al., 2017; Sholl et al., 2016). Sequence reads were aligned to reference sequence b37 edition from the Human Genome Reference Consortium using bwa, and further processed using Picard (version 1.90, http://broadinstitute.github.io/picard/) to remove duplicates and Genome Analysis Toolkit (GATK, version 1.6-5-g557da77) to perform localized realignment around indel sites. Single nucleotide variants were called using MuTect v1.1.45, insertions and deletions were called using GATK Indelocator. Copy number variants (CNV) and structural variants were called using the internally-developed algorithms RobustCNV and BreaKmer followed by manual review (Abo et al., 2015). RobustCNV calculates copy ratios by performing a robust linear regression against a panel of normal samples. The data were segmented using circular binary segmentation, and event identification was performed based on the observed variance of the data points (Bi et al., 2017).

We computed the cytoband-level copy number calls and weighted (by length) average segment means across the covered regions of each cytoband using ASCETS (Spurr et al., 2021). Briefly, cytobands were considered amplified/deleted if more than 70% of the covered regions had a log2 copy ratio of greater than 0.2/less than −0.2, and were considered neutral if more than 70% of the covered regions had a log2 copy ratio between −0.2 and 0.2.

For comparisons of driver mutation frequencies across patient tumors, cell lines, and organoid models, mutation and CNV calls for *KRAS*, *TP53*, *CDKN2A*, and *SMAD4* were either compiled from prior publications for patient samples and CCLE cell lines (Aguirre et al., 2018; Cancer Genome Atlas Research Network, 2017; Ghandi et al., 2019) or generated from whole genome sequencing of organoid models. Organoid cohort variants were called from tumor/germline pairs using GATK (v.4.1.6.0, Paired tumor-control mode). Potential germline variants were additionally filtered using gnomAD (v2.1). Significance of short-nucleotide variants (SNVs) was determined using MuSiC2 (v0.2, q-value < 0.1). CNVs were initially called using GATK (v.4.1.6.0) followed by analysis with GISTIC (v2.0.23). Genomic alterations in *KRAS*, *TP53*, *CDKN2A*, and *SMAD4* were binarized by counting a gene in the given sample as altered if it contained at least one of the following alterations: missense mutation, nonsense mutation, splice site alteration, frameshift insertion or deletion, in-frame insertion or deletion, a high amplification, or a homozygous deletion. Statistical significance of alteration occurrence per gene across sample cohorts was determined using a Fisher's exact test with multiple test correction using the Benjamini-Hochberg procedure (R v4.0.3).

Bulk RNA-sequencing analysis

FASTQs for bulk RNA expression profiles were downloaded from the relevant repository (TCGA, https://toil.xenahubs.net; PDAC Cell lines, https://portals.broadinstitute.org/ccle), available in-house (Panc-Seq, metastatic PDAC), or generated for this study (organoid cohort) (Aguirre et al., 2018; (Weinstein et al., 2013) Cancer Genome Atlas Research Network, 2017; Ghandi et al., 2019; Vivian et al., 2017). All were processed using the same pipeline. Briefly, each sample's sequences were marked for duplicates and then mapped to hg38 using STAR. After running QC checks using RNaseqQC, gene-level count matrices were generated using RSEM. Instructions to run the pipeline are given in the Broad CCLE github repository https://github.com/broadinstitute/ccle_processing. Length-normalized values (TPM) were then transformed according to log2(TPM+1) for downstream analysis. The entire dataset was scaled and

centered to allow relative comparisons across sample types (e.g., tumors, organoids, and cell lines). Signature scores were computed as below (e.g., basal and classical; see Generation of expression signatures/scores below) (Puram et al., 2017).

Single-cell data quality pre-processing and initial cell type discovery

All single-cell data analysis was performed using the R language for Statistical Computing (v3.5.1). Each biopsy sample's digital gene expression (DGE) matrix (cells x genes) was trimmed to exclude low quality cells (< 400 genes detected; < 1,000 UMIs; > 50% mitochondrial reads) before being merged together (preserving all unique genes) to create the larger biopsy dataset. The merged dataset was further trimmed to remove cells with > 8,000 genes which represent outliers and likely doublet cells. We also removed genes that were not detected in at least 50 cells. The same metrics were applied to the organoid single-cell cohort (see below). On a per cell basis, UMI count data was divided by total transcripts captured and multiplied by a scaling factor of 10,000. These normalized values were then natural log transformed for downstream analysis (i.e., log-normalized cell x gene matrix). Initial exploration of the data was performed using the R package Seurat (v2.3.4) and followed two steps: 1) SNN-guided quality assessment and 2) cell type composition determination. In step 1, we intentionally left cells in the DGE matrix of dubious quality (e.g., % mitochondrial reads > 25% but < 50%), performed PCA over the variable genes (n = 1,070 genes), and input the first 50 PCs (determined by Jackstraw analysis implemented through Seurat) to build an SNN graph and cluster the cells (res = 1; k.param = 40). The inclusion of poor-quality cells essentially acts as a variance "sink" for other poor-quality cells and they cluster together based on their shared patterns in quality-associated gene expression. This method helped to nominate additional low quality (e.g., defined exclusively by mitochondrial genes) or likely doublet cells (e.g., clusters defined by co-expression of divergent lineage markers) which were removed from the dataset (n = 1,678 cells). This led to an overall high-quality dataset of single-cells with a low overall fraction of mitochondrial reads (median = 0.09) for downstream analysis (Figure S1D).

Using the trimmed dataset, we proceeded to step 2 using a very similar workflow as above but with slightly altered input conditions for defining clusters. Here we used PCs 1-45 and their associated statistically significant genes for building the SNN graph and determining cluster membership (resolution = 1.2; k.param = 40). This identified the 36 clusters shown (visualized using *t*-SNE; perplexity, 40; iterations, 2,500) in Figure S1E. The expression of known markers was used to collapse clusters containing shared lineage information. For example, clusters 1, 2, and 4 all express high levels of macrophage markers—*CD14*, *FCGR3A (CD16)*, *CD68*—and were accordingly collapsed for this first pass analysis (Figure S1E,I). To aid our cell type identification, we performed an ROC test implemented in Seurat to confirm the specificity (power > 0.6) of the top marker genes used to discern the cell types. Combined with inferred CNV information (see below), this analysis confirmed the presence of 11 broad non-malignant cell types in our biopsy dataset (Table S2). Variation in the SNN graph parameters above did not strongly affect cell type identification.

Single-cell CNV identification

To confirm the identity of the putative malignant clusters identified in Figure S1F, we estimated single-cell CNVs as previously described by computing the average expression in a sliding window of 100 genes within each chromosome after sorting the detected genes by their chromosomal coordinates (Patel et al., 2014; Tirosh et al., 2016b). We used all T/NK, Fib, Hep, and Endo cells identified above as reference normal populations for this analysis. Complete information on the inferCNV workflow used for this analysis can be found here https://github.com/broadinstitute/inferCNV/wiki. To compare with bulk targeted DNA-sequencing, we collapsed individual probes to cytoband-level information (weighted average of log2 ratios across each cytoband, see above) within each sample. ScRNA-seq-inferred CNVs showed high concordance across samples with the bulk measurements and suggests that, at least by this metric, we are likely sampling the same dominant clones within sequential but distinct cores from each needle biopsy procedure (Figure S1G). For plotting CNV profiles in putative malignant versus normal cells (Figure S1H), we computed the average CNV signal for the top 5% of altered cells in each biopsy and correlated all cells in that biopsy to the averaged profile as has been previously described (Tirosh et al., 2016a). Relation of this correlation coefficient to the CNV score (mean square deviation from diploidy) in the single cells from each biopsy shows consistent separation of malignant from non-malignant cells, and, combined with membership in patient-specific SNN clusters, substantiates the identification of malignant cells in our dataset. One patient sample, PANFR0604, did not contain any malignant cells within the core biopsy used for scRNA-seq analysis.

Subclonal analysis with single-cell inferred CNVs

The inferCNV workflow can be used to call subclonal genomic variation with high sensitivity and is described at https://github.com/broadinstitute/inferCNV/wiki (Fan et al., 2018; Patel et al., 2014; Tirosh et al., 2016b). Briefly, we used a six-state Hidden Markov Model (i6-HMM) to predict relative copy number status (complete loss to > 3x gain) across putative altered regions in each cell. A Bayesian latent mixture model then evaluated the posterior probability that a given copy number alteration is a true positive. We set a relatively stringent cutoff for this step (BayesMaxPNormal = 0.2) to only include high probability alterations for downstream clustering. The results of this filtered i6-HMM output were then used to cluster the single cells using Ward's method. We used inferCNV's "random trees" method to test for statistical significance ($p < 0.05$, 100 random permutations for each split) at each tree bifurcation and only retained subclusters that had statistical evidence underlying the presumed heterogeneity. To track subclonal heterogeneity between biopsy and matched organoid cells in Figure 3E and Figure S5E-K, the above workflow was implemented within each biopsy and the relevant matched organoid samples, essentially treating all cells as the same "tumor" and allowing the CNVs to determine cell sorting agnostic to sample-of-origin. The results of the HMM output

can be used to infer gene-level information based on which genes are in the affected window. This (like the rest of the HMM workflow) is computed over groups of cells (e.g., samples or sub-clones) and used to map *KRAS* and other alterations to samples or sub-clones (Figure 3E, Figure S5E-K).

Subclustering of malignant and non-malignant cells

Detailed phenotyping required splitting the dataset into malignant and non-malignant fractions. After subsetting to only the malignant cells, we re-scaled the data and ran PCA including the first 35 PCs for SNN clustering and *t*-SNE visualization. This PCA was used to determine the PanNET identity for PANFR0580 (Figure S2A). After removing PANFR0580, we repeated the steps above and used this new PCA for the remainder of PDAC malignant cell analysis. Subsequent phenotyping for malignant cells is discussed below (Generation of expression signatures/scores). A similar approach was used for calling the non-malignant subsets in Figure 5A. To determine the specific phenotypes within T/NK, macrophage, and mesenchymal populations, we separately subclustered these groups using PCs 1-20 and a resolution of 0.6 in each case. Of note, subclustering within the macrophages revealed a distinct cluster of cells co-expressing markers of both T/NK cells and macrophages (n = 491 cells). We discarded these cells as likely doublets, as have others, and re-ran the macrophage PCA and clustering (Zhang et al., 2020; Zilionis et al., 2019). These cells are included in the full dataset in case they are of interest to others. Each unbiased analysis helped to define the non-malignant phenotypes summarized in Figure 5 and Figure S6.

Generation of expression signatures/scores

All expression scores were computed as previously described by taking a given input set of genes and comparing their average relative expression to that of a control set (n = 100 genes) randomly sampled to mirror the expression distribution of the genes used for the input (Tirosh et al., 2016b). While all scores were computed in the same way, choosing the genes for input varied. We have outlined the relevant approaches below. Where correlations (Pearson's *r)* are performed over genes, we used the log-transformed UMI count data for each case. Unless otherwise noted, we selected the top 30 statistically significant genes for each signature (> 3 SD above the mean for shuffled data) for visualization and scoring.

Cell cycle

We utilized previously established signatures for G1/S (n = 43 genes) and G2/M (n = 55 genes) to place each cell along this dynamic process (Tirosh et al., 2016a). After inspecting the distribution of scores in the complete dataset, we considered any cell > 1.5 SD above the mean for either the G1/S or the G2/M scores to be cycling (van Galen et al., 2019).

PDAC bulk subtype signatures

We scored malignant cells within our single-cell cohort for expression of previously published signatures derived from bulk RNA-sequencing of primary and metastatic tumors.

scBasal and scClassical programs

We first scored each malignant single cell for the basal-like and classical genes identified by Moffitt et al., 2015 as these were well described by unbiased analysis in our data (PCA, Figure S2B,C, S3B). To derive refined single-cell basal (scBasal) and single-cell classical (scClassical) signatures using our malignant cohort and determine programs associated with these cell states, we correlated the aforementioned basal and classical scores to the entire gene expression matrix containing malignant cells and identified the 1,909 genes significantly associated with either subtype ($r > 0.1$; > 3 SD above the mean for shuffled data, full data in Table S3). Biological pathway correlates for scBasal and scClassical are summarized in Figure S2E,F [WNT signaling (Kim et al., 2017); EMT (Gröger et al., 2012); cell cycle progression (Tirosh et al., 2016a)]. For visualization, we use the scBasal and scClassical genes (top 30 correlated genes for each). In Figure 2C we score single cells for EMT (Gröger et al., 2012) and the union of Hallmark and Reactome interferon response gene sets to show their divergence within cells expressing the scBasal state.

Intermediate co-expressor (IC) program

Ordering the cells by their polarization or ''score difference,'' simply the difference of the two scores, using these basal and classical scores related to PC1 and PC2 revealed a significant fraction of cells co-expressing intermediate levels of both cell states (Figure 2B, Figure S3A,B). Co-expressing cells showed associations with features across several additional PCs, but lacked a single dominant axis. To define a consensus set of genes that are preferentially expressed by coexpressing cells, we computed the Euclidean distance to the line representing equal basal and classical co-expression for each cell. To limit the influence of cell quality on this analysis and to specifically identify genes related to co-expression, we used cells from each group (basal, intermediate, and classical) with fractionally low mitochondrial genes (< 0.2) and non-zero basal or classical expression (basal or classical score > 0) and correlated their Euclidean distance (Figure S3C) to the entire gene expression matrix of malignant cells. Next, for each gene positively associated with this co-expressor state (Pearson's $r > 0$), we subtracted the second highest correlation coefficient for each subtype-associated gene (basal and classical), and then re-ranked the matrix by this corrected value. This enriched for genes more specific to the co-expressor state by excluding those that were also associated with basal or classical programs. We then selected the 115 genes with a corrected correlation value > 0.1 ($p < 0.00001$, shuffled data) as our intermediate co-expressor (IC) signature (Figure S3D, Table S3). Single cells were classified based on Euclidian distance to co-expression, where cells with Euclidian distance < 0.2 are defined as intermediate co-expressor and the remainder (Euclidian distance > 0.2) by their maximal of either scBasal or scClassical scores. We binned each organoid cell (Figure 4C,D) by its maximal expression for one of the 3 *in vivo* scores (scBasal, scClassical, or IC). Here a cell must be within 1 SD of the mean expression for a given subtype *in vivo*, else it was considered ''organoid-specific'' as

this program was superimposed on all organoid cells, regardless of their subtype identity (Figure 4C). We used these classifications to summarize overall sample malignant cell composition and visualize the groups. Tumor heterogeneity measures were not significantly affected by changing these cutoffs.

Non-Malignant programs

TAM signatures were determined similar to above and previous work (van Galen et al., 2019; Zhang et al., 2020; Zilionis et al., 2019). Using PCA as an anchor (Figure S6C), we correlated expression within the TAM compartment to either *FCN1*, *SPP1*, or *C1QC* (top loaded genes on each relevant PC) and merged the resultant correlation coefficients for every detected gene to the 3 subtypes into one matrix (i.e., a 16,920 × 3 matrix). For each TAM type (i.e., vector of correlation coeffects to each marker), we first ranked the matrix by decreasing correlation coefficient, selected only the most significantly associated genes to that type ($r > 0.1$; > 3 SD above the mean for shuffled data), subtracted the second highest correlation coefficient for each subtype-associated gene, and then re-ranked the matrix by this corrected value. We repeated this procedure for each TAM subtype independently. This ensures that the genes selected are specific to a given TAM subset and do not describe general TAM features. The top 30 genes for each were used for scoring and visualization (Table S2; Figure S6D).

Mesenchymal phenotypes were determined using a similar workflow. To examine mesenchymal heterogeneity, we removed a subset of adrenal endocrine cells (cluster 4, 40 cells; Figure 5C) and then performed PCA on the remaining mesenchymal cells. PC1 was driven by spillover genes (likely contributed from ambient RNA) and lacked any coherent biological program and was not considered further. PCs 2 and 3 by contrast were consistent with variable mesenchymal (PC2) and inflammatory CAF (PC3) phenotypes. All these cells scored highly for previous myCAF gene expression programs so this phenotype did not fully explain the heterogeneity in mesenchymal cells. Again, using correlation, we determined the genes driving low PC2 scores (Fibroblast-like), and high PC2 scores (Pericyte-like), as well as those associated with the high PC3 scores (Inflammatory). As before, we used the top 30 genes within each subset for scoring and visualization. These same genes (Fibroblast-like and Pericyte-like) were used to examine bulk RNA-seq profiles and their difference in each sample quantifies which phenotype is favored in the bulk averages (Figure 5C).

Analysis of normal pancreas progenitor data

We obtained the genes by cells matrix of normal pancreas progenitors (Qadir et al., 2020) at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131886. We clustered the original data and excluded a small subset of immune cells (CD45+). We then collapsed cell types from the original paper into broad categories (Pro-Acinar, Pro-ductal, Undifferentiated, and Mesenchymal) based on lineage marker expression. For analysis in Figure S3E we averaged the expression for scBasal, scClassical and IC genes in each group. For use in Figure S3F we generated signatures for each population using differential expression (FindAllMarkers function in Seurat using the ''wilcox'' option) and scored our single cells for these normal-derived signatures as above.

Matched organoid clustering and cell-typing

After applying similar quality metrics as above, we performed PCA, SNN clustering, and *t*-SNE embedding for 32,073 cells including organoid cells and all malignant cells from primary PDAC biopsies (PCs 1-50; resolution = 1.2; k.param = 45; perplexity = 45; max_iter = 2,500), and identified 39 total clusters. Organoids clustered separately from their matched biopsies, suggesting expression and/or CNV related drift in culture. Only two SNN clusters—clusters 4 and 32—were admixed by sample. We determined the specific gene expression programs in these two clusters via differential expression testing by Wilcoxon rank sum test ($p < 0.05$, Bonferroni correction; log(fold change) > 0.5). These comparisons were done in a ''1 versus rest'' fashion, testing for genes defining each cluster (4 or 32) compared to the entire dataset. Their expression profiles were consistent with non-malignant cell types, likely fibroblasts (cluster 32) and epithelial cells (cluster 4; Figure S5B,C).

Correlation distances for genotype and transcriptional cell state

To generate correlation distances for genotype and transcriptional cell state, each single cell in a biopsy-organoid pair was represented by two vectors of information: (i) a cell state vector containing expression values for scBasal and scClassical genes (n = 60 genes) and (ii) a genotype vector containing the average CNV score for each cytoband. The cell state and genotype distances between every single cell within a biopsy/early organoid pair was computed from these vectors using a correlation-based (Pearson's *r*) distance metric of the form $d = (1-r)/2$. This resulted in two distance matrices of n x n dimension where n is the total number of cells from each biopsy/early organoid sample pair. Values in Figure 3A are computed by averaging the values for *d* between only early organoid and matched biopsy cells.

Matched biopsy versus organoid malignant cell comparison

For CNV-confirmed malignant cells from each biopsy and its matched organoid (earliest passage), we used differential expression (Wilcoxon rank sum test; $p < 0.05$, Bonferroni correction; log(fold change) > 0.3) to understand the features lost from malignant cells in the *in vivo* setting and gained when transitioning into growth in organoid culture. We required any gene to be significantly differentially expressed in at least 3 model-biopsy comparisons to summarize the consistent changes. We repeated this same workflow for both organoid- and biopsy-specific genes (Table S4) outlined in Figure 3C and Figure 4G,H, respectively.

TME associations

We determined the transcriptional-subtype-dependent composition of the TME (Figure 5E-I) following two steps. First, we computed the Simpson's Index (measure of ecological diversity) using the count of each non-malignant cell type captured from each sample as input (Figure 5E,G) and correlated each biopsy's diversity score to its scBasal versus scClassical commitment score. Importantly, the number of non-malignant cells captured from each biopsy was not associated with basal versus classical commitment score ($r = 0.09$). Next, to understand which cell types drive these differences, we computed the fractional representation for every non-malignant cell type in each core needle biopsy and determined their pairwise correlation distance (Pearson's *r*) followed by hierarchical clustering using Ward's method (dendrogram in Figure 5G). For both analyses we only used samples with > 200 non-malignant cells captured (Figure S6N).

Biopsy paracrine and autocrine subtype-specific factor analysis

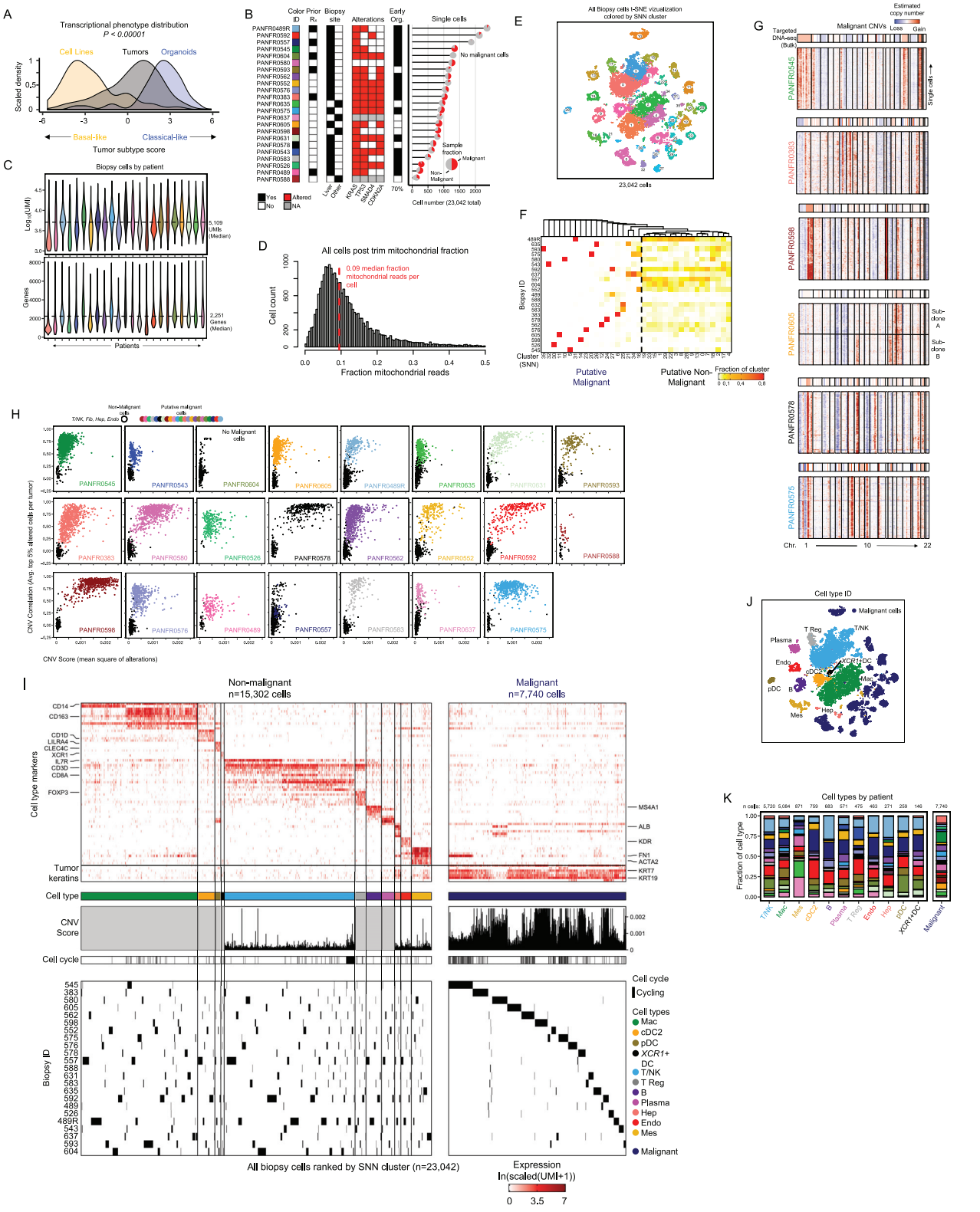
Factors present in the TME but absent from organoid culture could originate from at least two sources, the malignant cells themselves (autocrine) or non-malignant cells in the local microenvironment (paracrine). We examined any gene with gene ontology annotations related to ''cytokines,'' ''chemokines,'' or ''growth factors'' and took the union of these lists, yielding 321 genes, 218 of which were detected in our dataset. For ''autocrine'' factors we performed differential expression between malignant cells binned as scBasal and scClassical, and then IC versus rest. A gene was considered differentially expressed if it passed a $p < 0.05$ with Bonferroni correction and a log(fold change) > 0.2 in one of these comparisons. Genes were then assigned to subtypes based on the log fold change direction (Figure 6B, Table S6). ''Paracrine'' factors were determined in a similar manner with slight modifications. We grouped non-malignant cells into basal, classical or IC based on the average expression and clustering for malignant programs from their respective tumor samples (Figure S3G). We then assessed for differential expression between all cells from a given group and the rest using the same cutoffs as above and sorted factors into subtypes based on their log fold change directionality (Figure 7A, Table S6). We visualized which cell type contributed the highest average expression for each factor among the cell types from each of the respective cell state-specific TMEs (Figure 7B). We note that our use of ''paracrine'' and ''autocrine'' here is somewhat inexact as these secreted factors could act in either manner depending on the context. We merely use this nomenclature to reflect a ''cancer cell centric'' view, i.e., factors secreted by malignant cells are autocrine and those deriving from the TME are paracrine.

Tumor phenotyping from mIF data

Supervised machine learning algorithms were applied for tissue and cell segmentation (inForm 2.4.1, Akoya Biosciences). Single-cell-level imaging data were exported and further processed and analyzed using R (v3.6.2). To assign phenotypes to individual malignant epithelial cells, mean expression intensity in the relevant subcellular compartment was first used to classify cells as positive or negative for each of the 5 markers. Combinatorial expression patterns for the five markers were then used to phenotypically classify cells as basal, classical, co-expressing / IC or marker negative (3 combinations of 2 basal markers, 7 combinations of 3 classical markers, 1 pan-marker negative, 21 combinations of co-expression of basal and classical markers, Figure S4A, Table S3). Tumor subtype composition was assessed by calculating the fraction of total malignant cells positive for each cell phenotype (Figure S4B, excluding pan-marker negative cells).

CellPress
OPEN ACCESS

Supplemental figures



(legend on next page)

Figure S1. Clinical and genomic features, quality metrics, unsupervised cell type identification, and malignant cell confirmation across the biopsy cohort, related to Figure 1

(A) Distribution of PDAC patient tumor and cell model phenotypes along the basal-classical spectrum for bulk RNA-seq samples in Figure 1C. *P*-value for group differences was calculated by ANOVA.

(B) Clinical and molecular features for all patients included in the dataset (R_x = Therapy; Other = Adrenal (PANFR0637), Omentum (PANFR0635, PANFR0598), Peritoneum (PANFR0588); Org. at P2 = Organoid measured at passage 2). Mutations were determined by bulk targeted DNA-seq (Red, Altered; White, wild type; Grey, Data not available). Number of single cells captured per biopsy and their malignant and non-malignant fraction is visualized at the right. The scRNA-seq biopsy for sample PANFR0604 did not contain any malignant cells.

(C) Distribution of unique molecules and genes captured in quality cells per biopsy, median values are indicated for each metric (dotted line) and violin plots are colored by patient (top, Log_{10}(UMIs); bottom, number of genes).

(D) Distribution of fraction mitochondrial reads across the entire trimmed biopsy dataset (n = 23,042 cells). Red dotted line denotes the median.

(E) *t*-SNE visualization of the entire single-cell biopsy dataset colored by the identified SNN clusters (inset numbers).

(F) Distribution of single cells captured per biopsy across the identified and putative malignant and non-malignant SNN clusters.

(G) Heatmaps represent select scRNA-seq-derived copy number profiles where expression across the transcriptome is organized by chromosome (columns) for each single putative malignant cell (rows) from a given biopsy. Top bar indicates reference bulk targeted DNA-seq for the same patient.

(H) CNV correlation (averaged top 5% of altered cells per biopsy) versus CNV score (mean square of modified expression) for each single putative malignant (colored points) and reference normal (empty black circles) cell within a given biopsy. One sample, PANFR0604, did not contain any malignant cells.

(I) Overview of cell-typing for all cells in the biopsy dataset. Cells are ordered by SNN cluster and separated by cell types. Top heatmap represents expression levels for a subset of select markers (n = 73 genes) used to identify cell types. Color bar indicates cell types and binarized cell cycle phenotypes are labeled (black, cycling; white, not). CNV scores (mean square of alterations per cell) used to parse malignant from non-malignant are shown using T/NK, endothelial, fibroblasts, and hepatocytes as reference; gray boxes denote normal cell types where we did not compute reference CNV scores. Bottom panel shows biopsy of origin for each cell. The data are split by non-malignant (n = 15,302) and malignant (7,740) identity.

(J) *t*-SNE visualization as in **S1E** but colored by cell types identified. Endo, Endothelial; Mes, Mesenchymal; B, B cell; Hep, Hepatocyte; DC, Dendritic cell; pDC, Plasmacytoid dendritic cell; Mac, Macrophage; T, T cell; NK, Natural killer cell.

(K) Fraction of each cell type contributed by each biopsy sample (color fill, patient ID as in S1B), cell type totals are noted at the top of each bar.

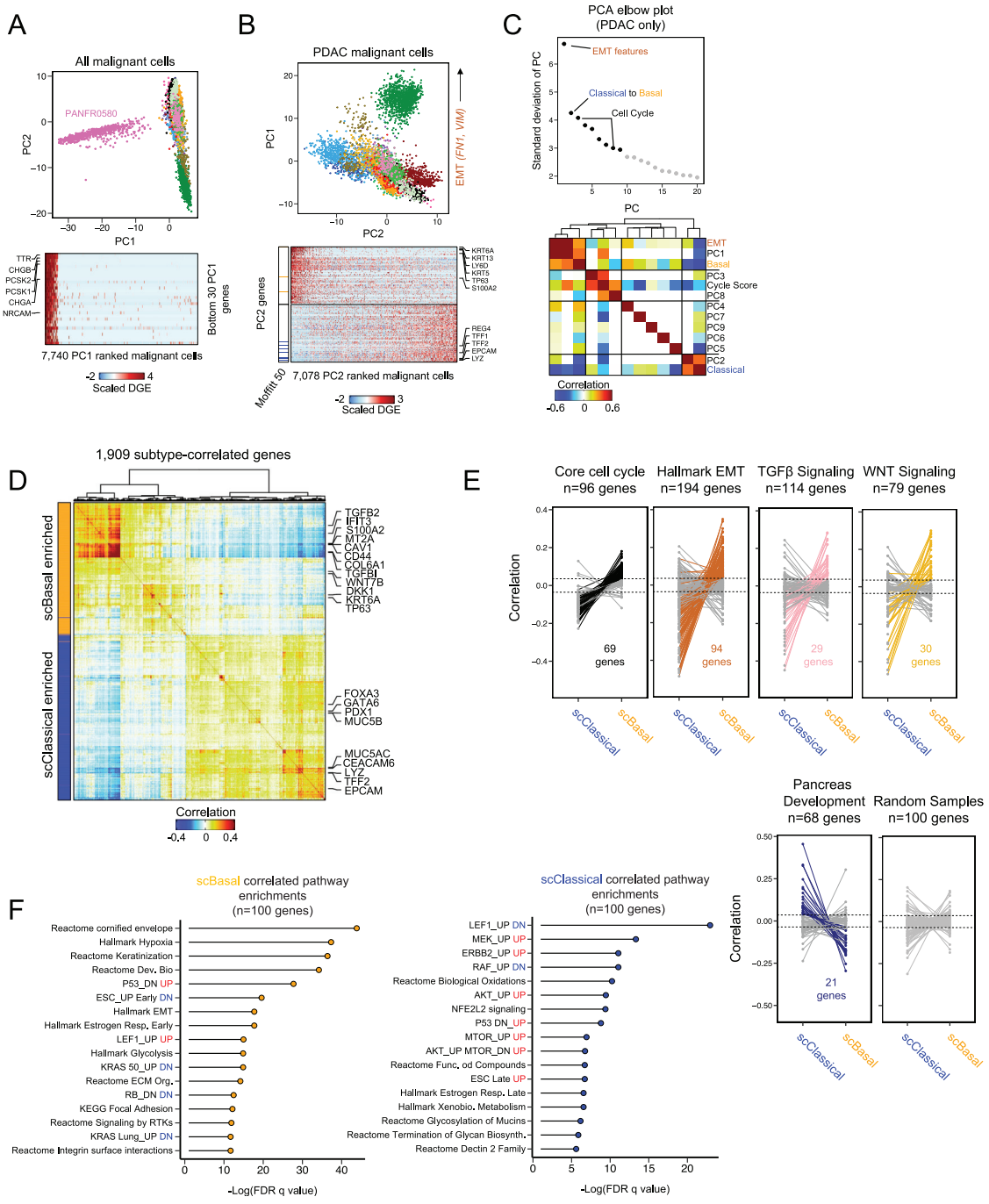


Figure S2. Identifying and contextualizing single-cell scBasal and scClassical associated biology, related to Figure 2

(A) Principal component analysis (PCA) and scatterplot for PC1 and PC2 across all malignant cells (n = 7,740, colored by patient ID as in Figure S1B) separates PANFR0580's malignant cells (n = 662) from the rest. Heatmap for genes with the strongest negative loading on PC1 (n = 30) denotes a neuroendocrine identity (*TTR, CHGB*). This tumor was later classified by histology as a pancreatic neuroendocrine tumor (PanNET).

(B) Principal component analysis (PCA) and scatterplot for PC1 and PC2 across all PDAC malignant cells (n = 7,078, colored by patient ID as in Figure S1B), excluding PANFR0580's PanNET malignant cells. Heatmap for genes with the strongest positive and negative loadings on PC2 (n = 30 each) shows overlap with genes from the Moffitt basal and classical signatures.

(C) Principal component (PC) elbow plot (top) showing the standard deviation for the first 20 components calculated over the verified PDAC malignant cell variable genes. Line is drawn at the putative "elbow" (black versus gray points) as inclusion of additional PCs described overlapping information or quality metrics. Cross-correlational analysis (bottom) for each single-cell's embeddings across first 9 PCs (black points) and scores for literature curated gene sets describing EMT, classical and basal, and cell cycle phenotypes.

(legend continued on next page)

(D) Pairwise correlation of genes significantly enriched in scBasal or scClassical expression states within single-cell cohort. Left bar indicates the subtype association of each gene (orange, scBasal-enriched; blue, scClassical-enriched).
(E) Tied dot plots depict the correlation coefficient for each gene (points) from select literature-derived gene sets, indicated at the top of each plot, to either scBasal or scClassical cell states. Dotted lines represent significance threshold (3 SD above the mean of shuffled data), points and lines are colored if that gene passes the threshold.
(F) GSEA pathway enrichments for the top 100 genes correlated to either scBasal or scClassical expression scores.

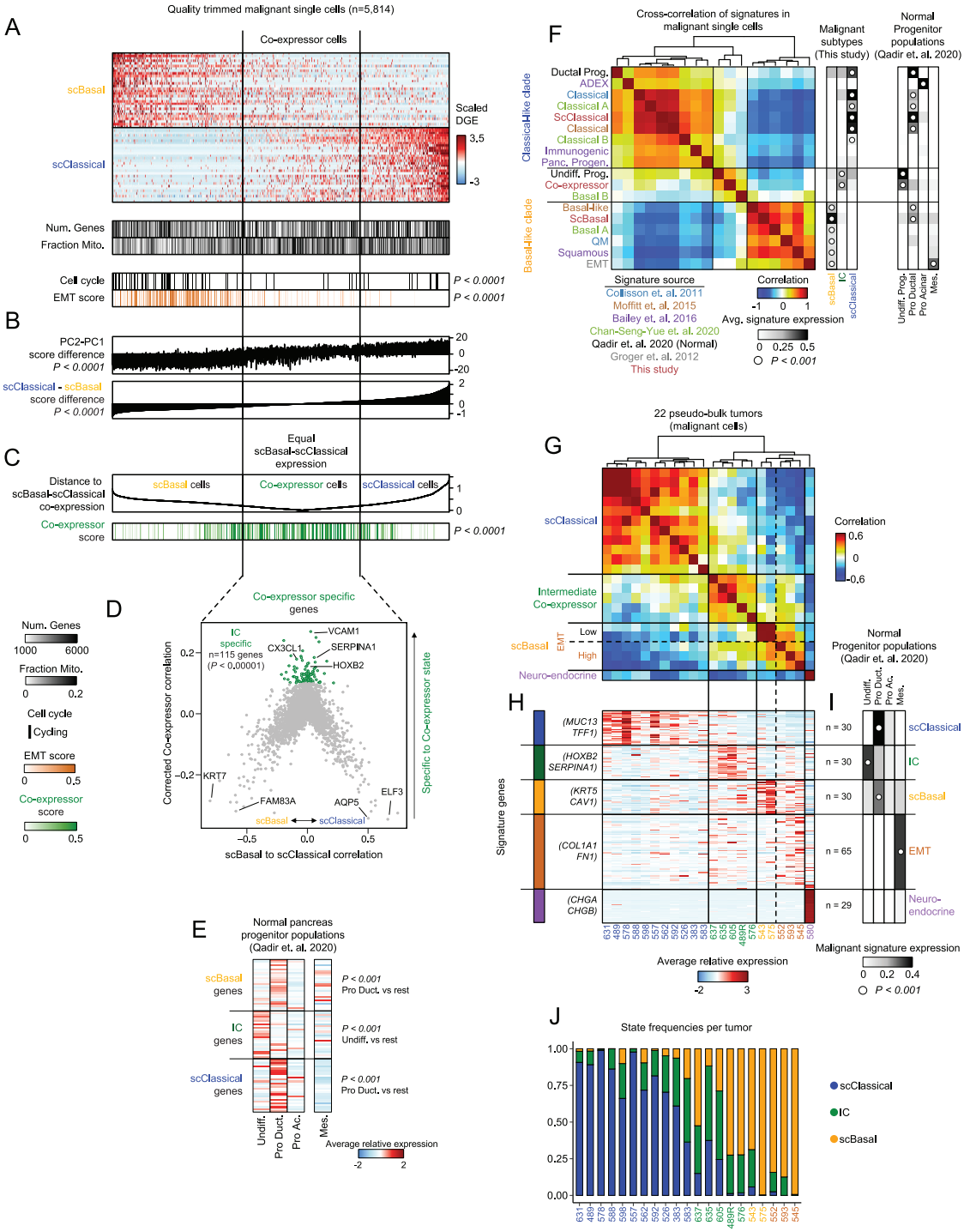


Figure S3. Cells with co-expression of scBasal and scClassical states express a distinct intermediate co-expressor gene program, related to Figure 2

(A) Expression of scBasal and scClassical gene programs, with cells ordered by their scBasal-scClassical score difference. Quality metrics, EMT scores and the binarized cell cycle program are shown for each single cell below the heatmap. *P*-value for binarized cycling group differences was calculated using Fisher's Exact test. *P*-value for EMT score was calculated by Kruskal-Wallis test with multiple hypothesis correction.

(B) Difference between PC1 and PC2 (top) and scClassical–scBasal score difference (bottom) are shown. Cells with equal scBasal and scClassical expression are associated with intermediate PC scores and cells are ordered as in S3A. *P*-values for group differences were calculated by Kruskal-Wallis test with multiple hypothesis correction.

(legend continued on next page)

(C) Euclidean distance for each cell to co-expression (y = x) of scBasal (x) and scClassical (y) expression scores. Bottom track indicates the score derived from the genes specific to the intermediate co-expressor state shown in S3D. *P*-values for group differences were calculated by Kruskal-Wallis test with multiple hypothesis correction.
(D) Gene correlation to either scBasal or scClassical score (x axis) or the corrected intermediate co-expressor correlation (Euclidean distance in S3C). Green highlighted genes have corrected intermediate co-expressor correlation > 0.1 ($p < 0.00001$ above shuffled).
(E) Averaged expression of all three malignant programs in normal pancreatic progenitor niche subsets (Acinar (Pro Ac.), Ductal (Pro Duct.), or Undifferentiated (Undiff.) and mesenchymal (Mes.) cells defined in Qadir et al., 2020. *P*-values for each set of genes are computed by Kruskal-Wallis test with multiple hypothesis correction.
(F) Cross-correlation between new and previously proposed expression signatures in our PDAC single cells. Average expression for each signature (rows) is shown at the right for cells in the malignant subtypes from our cohort and the normal pancreatic progenitor cells from Qadir et al., 2020. White dot indicates the subset with the highest average significant expression for each signature (Kruskal-Wallis test); no white dot indicates no significant expression.
(G) Pairwise correlation for biopsies with malignant cells (n = 22). Data are correlation coefficients for the average expression of all signature genes in the malignant cells from a given biopsy. Clade identities are at left with the one PanNET tumor (PANFR0580) included for comparison.
(H) Average expression for the 184 genes used for clustering in S3G. Clade identity colors match text color in S3G, and individual samples (columns) are ordered as in S3G with sample ID numbers provided below.
(I) Scores for the expression of genes in S3H (gray scale heat) across the 4 main cell types found in the pancreatic progenitor niche (Qadir et al., 2020). White dot indicates the normal subset with the highest average expression for each malignant program (Kruskal-Wallis test), none of the normal subsets significantly express the neuroendocrine gene signature.
(J) Frequencies of scBasal, scClassical, and IC states in individual tumors demonstrates intratumoral transcriptional heterogeneity.

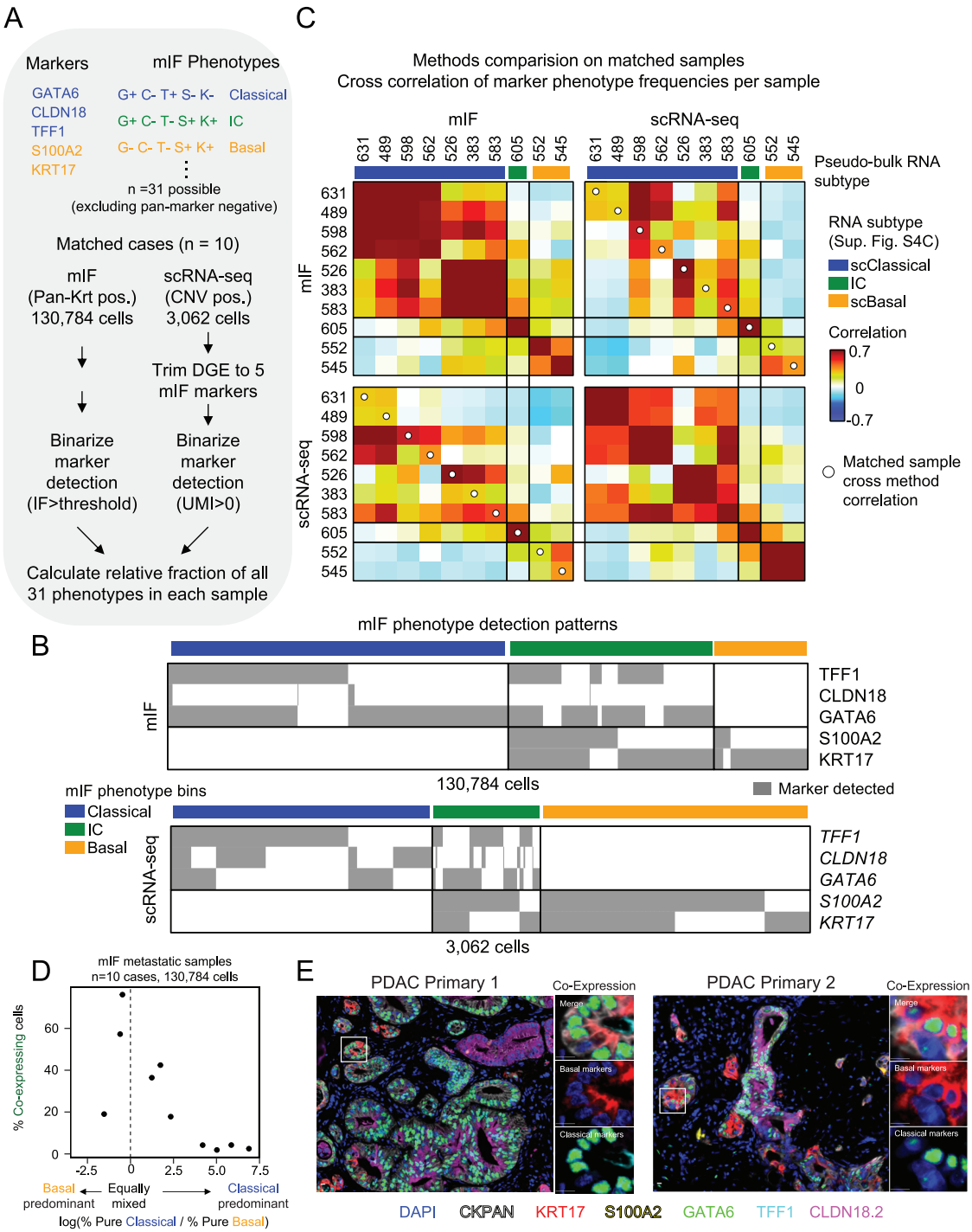


Figure S4. Multiplex immunofluorescence is concordant with scRNA-seq and demonstrates intratumoral heterogeneity with the presence of intermediate co-expressor cells, related to Figure 2

(A) Schematic for comparison of the matched datasets by combinatorial marker phenotypes.

(B) Marker detection in each single cell from the 10 samples in the mIF (top, 130,784 cells) and matched scRNA-seq datasets (bottom, 3,062 cells). Cells are sorted by their combinatorial phenotype outlined in S4A.

(C) Comparison within and between modalities on matched samples. Samples are sorted by their pseudo-bulk RNA subtype identity (color bar, top; dendrogram from Figure S3G). Pearson correlations (red to blue heat) were performed over the fractional representation of each marker pattern in S4A for each biopsy using either modality (scRNA-seq or mIF). The upper left and lower right quadrants cross-compare cell state composition from individual biopsies assessed by mIF and scRNA-seq, respectively, with both showing agreement with averaged RNA subtype (top color bar). The upper right and lower left quadrants cross-compare

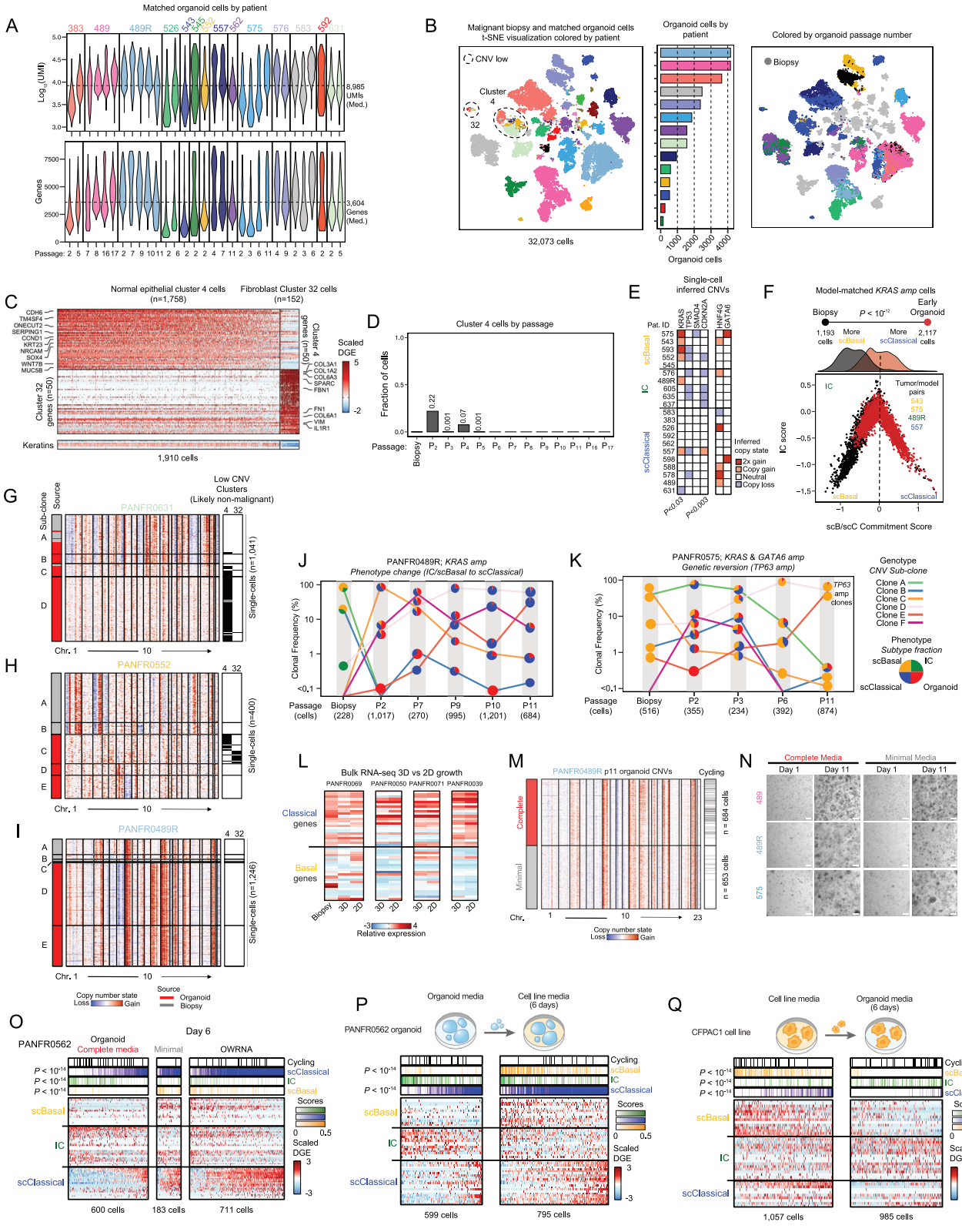
(legend continued on next page)

modalities (scRNA-seq versus mIF) for each sample (white dots), demonstrating that individual marker patterns for each sample are similar despite being measured by different methods in different cores from the same patient.

(D) Frequency of co-expressing cells is related to increased mixing of basal and classical cell populations within patient samples assessed by mIF. Log ratio of % basal and classical cells in each sample (x axis; dotted line at 0 indicates equal percentages of basal and classical cells) versus their % co-expressing cells (y axis).

(E) mIF images and marker detection in single cells from two primary PDAC samples identifies IC cells in primary specimens also. Scale bar represents 10 μm.

CellPress
OPEN ACCESS



(legend on next page)

Figure S5. Quality metrics, cell type identification, serial sampling, and media influences on transcriptional state in patient-matched organoids, related to Figures 3 and 4

(A) Distribution of unique molecules and genes captured in quality cells per organoid sample, median values are indicated for each metric (dotted line) and violin plots are colored by patient ID (top, Log_{10}(UMIs); bottom, number of genes).

(B) *t*-SNE visualization of all biopsy and matched organoid cells from iterative passages, colored by patient ID (left). Dotted circles indicate the only two SNN clusters (4 and 32) with appreciably admixed clusters and low CNV scores, the rest were patient-specific. Bar chart shows number of organoid cells recovered per matched sample (center). On the right *t*-SNE visualization, cells are organized in the same manner but organoid cells are colored by passage number.

(C) Relative expression for genes defining cluster 4 (top) and cluster 32 (bottom). Cluster 4 had an ambiguous epithelial identity while cluster 32 cells were defined by canonical fibroblast genes and low to absent detection of CNVs.

(D) Fraction of cluster 4 cells at each passage, demonstrating that these cells did not survive iterative passaging.

(E) Single-cell inferred copy number alterations for each sample in the biopsy cohort. Tumors are grouped by expression of their dominant subtype based on the clustering in Figure S3G, *P*-values comparing presence of each alteration among the groups (scBasal, scClassical, IC) are determined by Fisher's exact test.

(F) Cell state diagram with marginal density plots for all cells with inferred *KRAS* amplifications (4 biopsy-model pairs) in biopsy (black) and organoid (red) microenvironments. *P*-value compares biopsy versus early passage organoid score distributions (top density) and was determined by Student's t test.

(G-I) Heatmaps show inferred CNV copy number status for every cell in each of three biopsy/early passage organoid pairs. Cells (in rows) are ordered by hierarchical clustering of their CNV profiles and letters on the far left indicate subclones that have significant statistical evidence for tree-splitting (STAR Methods). Each cell's origin is also noted ("Source" column; biopsy tissue, gray; early passage organoid, red). Right metadata bars indicate if that cell came from an admixed, likely non-malignant SNN cluster (4 or 32 in S5B).

(J-K) Matched cell state and genotype evolution at each passage in PANFR0489R (S5J) and PANFR0575 (S5K). Frequencies of individual CNV clones at each time point (y axis) are tied by colored lines. Fill represents the transcriptional cell state fraction for each CNV clone. In sample PANFR0575 (S5K), clones D and E had inferred *TP63* amplifications which expanded over time.

(L) Bulk RNA-seq expression of basal and classical states in 4 organoid models cultured in 3D or adapted to 2D in complete organoid media.

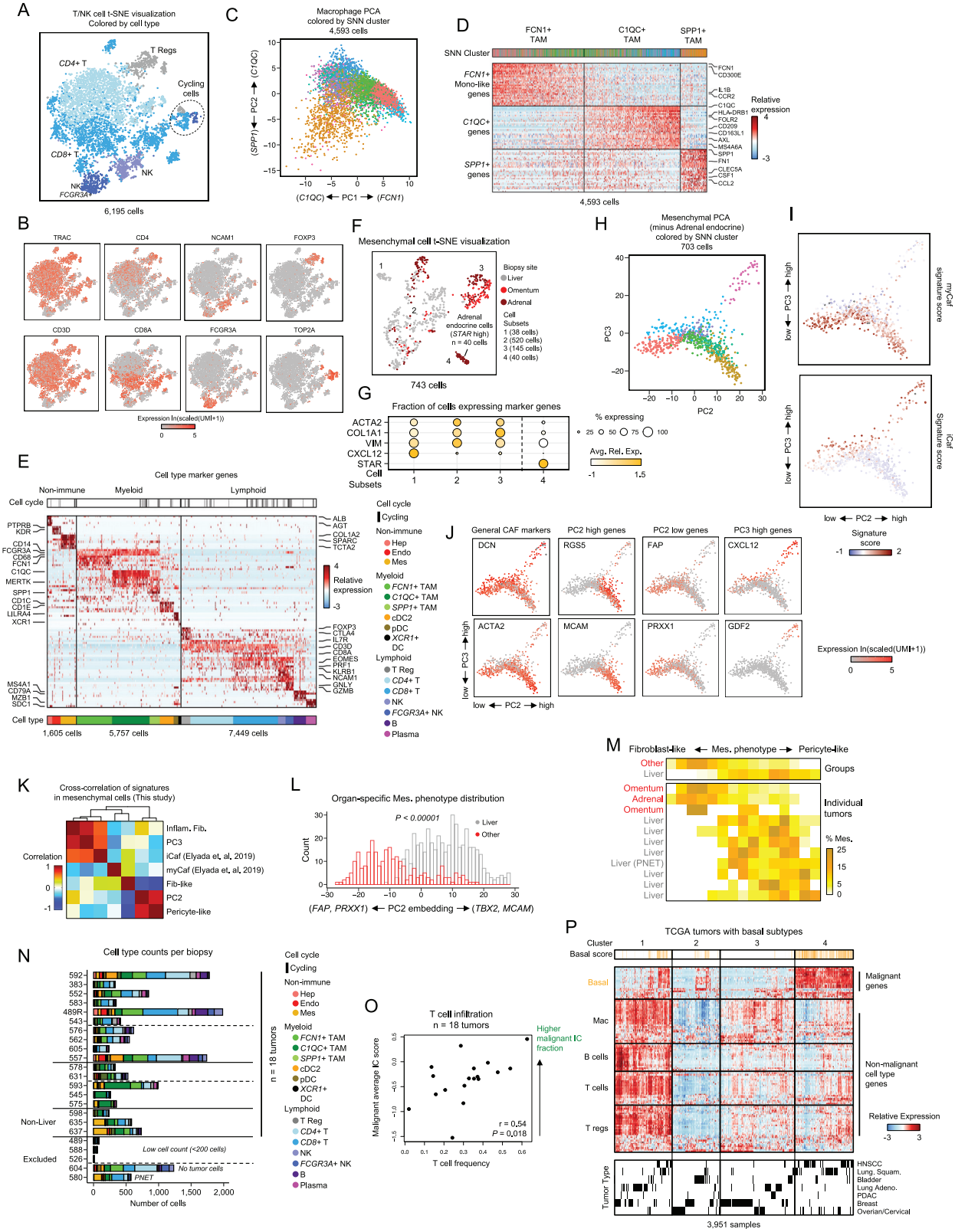
(M) Inferred CNVs for each cell from the PANFR489R samples cultured in either Minimal (gray) or Complete (red) organoid media conditions in Fig. 4B.

(N) Brightfield images of organoids grown in standard organoid media ("Complete") or in media without any growth factors ("Minimal") at days 1 and 11 after seeding. Scale bar represents 200 μm.

(O) Expression of scBasal, IC, and scClassical genes (rows) in organoid cells (columns) from model PANFR0562 cultured for 6 days in Complete medium, Minimal medium, or OWRNA medium. *P*-values for group differences were calculated by ANOVA followed by Tukey's HSD.

(P) Expression of scBasal, IC, and scClassical genes (rows) in organoid cells (columns) from model PANFR0562 cultured for 6 days in Complete organoid medium or in "Cell line" medium. *P*-values for group differences were calculated by ANOVA followed by Tukey's HSD.

(Q) Expression of scBasal, IC, and scClassical genes (rows) in single cells from the PDAC cell line CFPAC1 (columns) cultured for 6 days in standard "Cell line" medium or in Complete organoid medium. *P*-values for group differences were calculated by ANOVA followed by Tukey's HSD.



(legend on next page)

Figure S6. Identification of T cell, natural killer cell, macrophage, and mesenchymal heterogeneity and transcriptional state associations in the metastatic microenvironment, related to Figure 5

(A) *t*-SNE visualization of sub-clustering (SNN) performed on T/NK cells in the metastatic cohort. Cells are colored by their type identity based on shared SNN cluster membership.

(B) Select cell type marker expression overlaid on the *t*-SNE visualization from S6A**.**

(C) PCA identifies 3 major subsets of TAMs in the metastatic niche. PC1 largely separates *FCN1*+ monocyte-like TAMs from more committed macrophage phenotypes. PC2 separates *SPP1*+ from *C1QC*+ macrophage phenotypes.

(D) Heatmap visualization of the gene expression programs specific to each TAM subset identified by the PCA and SNN clusters in S6C.

(E) Relative expression for select cell type markers and binarized cell cycle program (top bar; black, cycling) with cell type colors (bottom color bar) as in Figure 5A.

(F) *t*-SNE visualization of sub-clustering (SNN) performed on mesenchymal cells colored by their anatomical site. Cell subsets (1-4) determined by SNN clustering.

(G) Average expression of the indicated mesenchymal/fibroblast and adrenal endocrine marker genes in each of the cell subsets (1-4) identified in S6F. Dot size indicates fraction of cells expressing a given gene.

(H) PCA over mesenchymal cells in the cohort (excluding Adrenal endocrine cells; subset 4, **S6F**). Scatterplot of PC2 versus PC3 defines 3 states for mesenchymal cells in our cohort.

(I) Same visualization as in S6H, but cells are colored by previously identified myCAF or iCAF signature scores. myCAF is evenly distributed across PC2 and iCAF associates with higher PC3 scores.

(J) Expression for select markers overlaid on the PCA from S6H.

(K) Cross-correlation of mesenchymal signatures in single-cells. Fibroblast-like versus pericyte-like signatures provide non-overlapping information. PC3 inflammatory phenotypes are similar to the previously reported iCaf phenotype and our PC3-derived inflammatory mesenchymal signature.

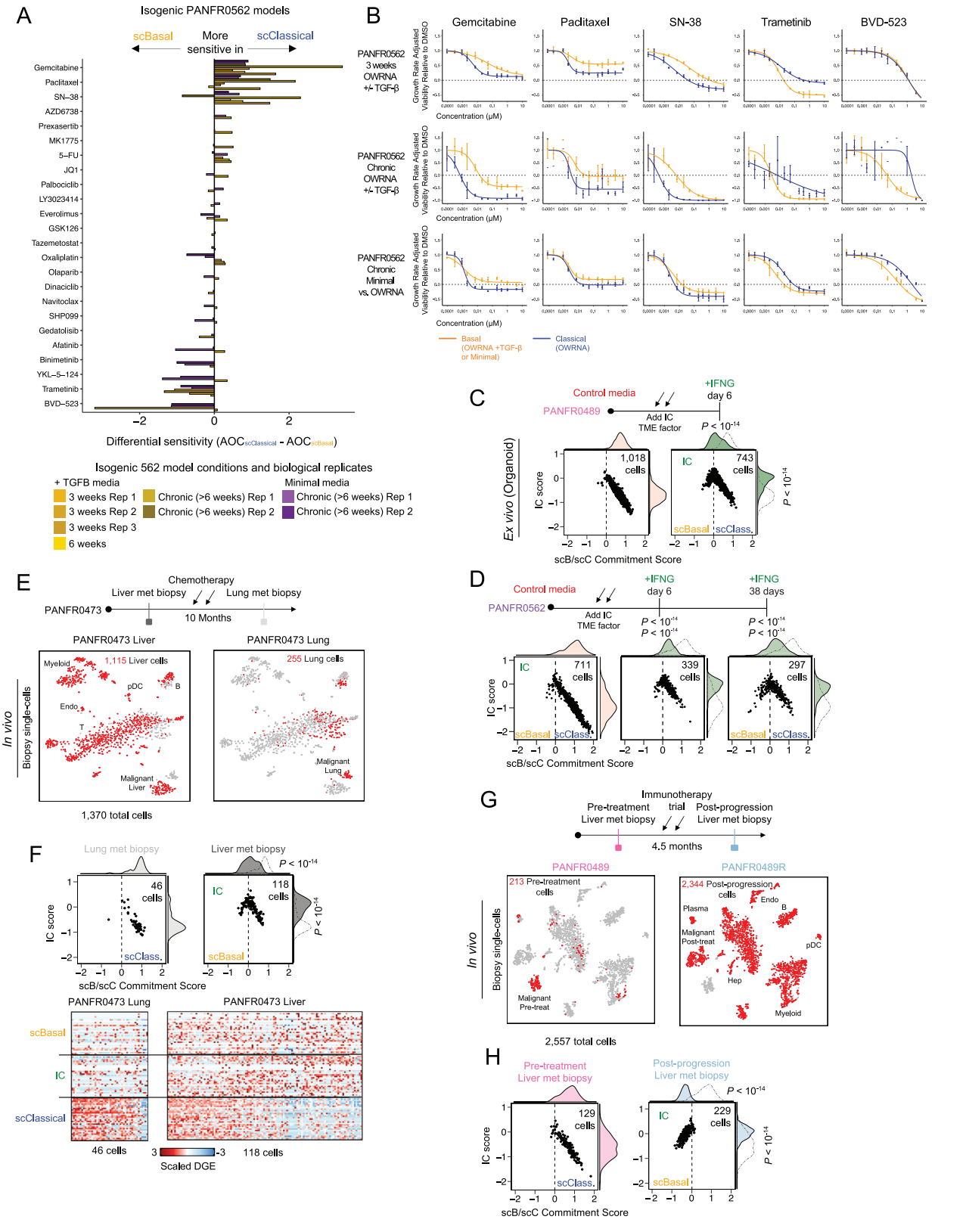
(L) Frequency of mesenchymal cells (y axis, cell count) across PC2 scores, colored by site of biopsy tissue. *P*-value determined by Student's t test.

(M) Comparison of the distribution of mesenchymal phenotypes across different sites as groups (top) or individual tumors (bottom). Heat indicates the fraction of mesenchymal cells in that score bin.

(N) Non-malignant cell types (color fill) and counts per biopsy, included samples organized as in Figure 5G. Five biopsies were excluded from the analysis in Figure 5E,5G-I because they either had low cell capture or were from a tumor with indeterminant malignant transcriptional subtype due to negligible malignant cell capture.

(O) Correlation between T cell fraction and malignant IC score.

(P) Cross-TCGA analysis for basal and immune cell type markers in epithelial tumors with known basal subtypes. Clusters were determined by dendrogram splitting, and disease type for each sample is indicated at bottom.



(legend on next page)

Figure S7. State-specific drug responses and TME-associated state shifts in organoid models and patient samples, related to Figures 6 and 7

(A) Independent experiments demonstrating state-specific drug sensitivities in isogenic PANFR0562 organoid model pairs skewed toward scBasal or scClassical states by altering media formulation (scBasal, TGF-β or Minimal media; scClassical, OWRNA media) for different time periods. Bars represent the difference in growth rate-corrected Area Over the Curve (AOC) between each scBasal-scClassical model pair for each independent experiment depicted in Figure 6G.

(B) Growth rate-adjusted dose response curves for isogenic PANFR0562 organoid model pairs in different media conditions and at different time points. Points are mean ± SD of 3 technical replicates. Orange curves represent responses in more scBasal conditions and blue curves represent responses in more scClassical conditions.

(C) Cell state diagrams with marginal density plots for the organoid model PANFR0489 depicted in Figure 7C in control medium (OWRNA, reduced organoid medium) or control medium with IFNγ. *P*-values for group differences between B/C commitment (top) and IC scores (right) were calculated by ANOVA followed by Tukey's HSD.

(D) Cell state diagrams with marginal density plots for organoid model PANFR0562 at 2 time points with exposure to IFNγ. *P*-values for group differences between B/C commitment (top) and IC scores (right) were calculated by ANOVA followed by Tukey's HSD.

(E) *t*-SNE visualization of single cells from metastatic biopsies collected from the same patient (PANFR0473, Figure 7D) but at two different sites (liver, left; lung, right). All cells from both samples are shown in both *t*-SNE plots, but cells for each site are labeled red in the corresponding plot.

(F) Cell state diagrams with marginal density plots (top) and expression heatmaps (bottom) depict cell state differences between malignant cells from two distinct metastatic sites for the samples (PANFR0473) in Figure 7D and S7E. *P*-values for group differences between B/C commitment (top) and IC scores (right) were calculated by ANOVA followed by Tukey's HSD.

(G) *t*-SNE visualization of single cells from metastatic biopsies collected from the same site in one patient (PANFR0489, Figure 7E) pre- and post-immunotherapy. All cells from both samples are shown in both *t*-SNE plots, but cells for each time point are labeled red in the corresponding plot.

(H) Cell state diagrams with marginal density plots from metastatic samples collected pre- and post-immunotherapy (PANFR0489, Figure 7E,F and S7G). *P*-values for group differences between B/C commitment (top) and IC scores (right) were calculated by ANOVA followed by Tukey's HSD.