



OPEN

Spatially restricted drivers and transitional cell populations cooperate with the microenvironment in untreated and chemo-resistant pancreatic cancer

Daniel Cui Zhou ^{1,2,18}, Reyka G. Jayasinghe^{1,2,18}, Siqi Chen^{1,2}, John M. Herndon ^{3,4}, Michael D. Iglesia ^{1,2}, Pooja Navale^{1,5}, Michael C. Wendt^{2,6,7}, Wagma Caravan ^{1,2}, Kazuhito Sato ^{1,2}, Erik Storrs^{1,2}, Chia-Kuei Mo^{1,2}, Jingxian Liu ^{1,2}, Austin N. Southard-Smith^{1,2}, Yige Wu^{1,2}, Nataly Naser Al Deen^{1,2}, John M. Baer^{1,5}, Robert S. Fulton², Matthew A. Wyczalkowski ^{1,2}, Ruiyang Liu^{1,2}, Catrina C. Fronick², Lucinda A. Fulton², Andrew Shinkle^{1,2}, Lisa Thammavong^{1,2}, Houxiang Zhu^{1,2}, Hua Sun ^{1,2}, Liang-Bo Wang ^{1,2}, Yize Li ^{1,2}, Chong Zuo¹, Joshua F. McMichael ^{1,2}, Sherri R. Davies ³, Elizabeth L. Appelbaum², Keenan J. Robbins ^{3,4}, Sara E. Chasnoff³, Xiaolu Yang¹, Ashley N. Reeb^{1,8}, Clara Oh^{1,2}, Mamatha Serasanambati^{1,2}, Preet Lal^{1,2}, Rajees Varghese^{1,2}, Jay R. Mashl^{1,2}, Jennifer Ponce², Nadezhda V. Terekhanova^{1,2}, Lijun Yao^{1,2}, Fang Wang⁹, Lijun Chen¹⁰, Michael Schnaubelt¹⁰, Rita Jui-Hsien Lu ^{1,2}, Julie K. Schwarz^{4,11,12}, Sidharth V. Puram⁸, Albert H. Kim ^{4,13}, Sheng-Kwei Song^{1,14}, Kooresh I. Shoghi^{1,14}, Ken S. Lau ¹⁵, Tao Ju¹⁶, Ken Chen ⁹, Deyali Chatterjee¹⁷, William G. Hawkins ^{3,4}, Hui Zhang¹⁰, Samuel Achilefu^{1,14}, Milan G. Chheda^{1,4}, Stephen T. Oh^{1,5}, William E. Gillanders^{3,4}, Feng Chen¹, David G. DeNardo ^{1,4,5,19} ✉, Ryan C. Fields ^{3,4,19} ✉ and Li Ding ^{1,2,4,6,19} ✉

Pancreatic ductal adenocarcinoma is a lethal disease with limited treatment options and poor survival. We studied 83 spatial samples from 31 patients (11 treatment-naïve and 20 treated) using single-cell/nucleus RNA sequencing, bulk-proteogenomics, spatial transcriptomics and cellular imaging. Subpopulations of tumor cells exhibited signatures of proliferation, KRAS signaling, cell stress and epithelial-to-mesenchymal transition. Mapping mutations and copy number events distinguished tumor populations from normal and transitional cells, including acinar-to-ductal metaplasia and pancreatic intraepithelial neoplasia. Pathology-assisted deconvolution of spatial transcriptomic data identified tumor and transitional subpopulations with distinct histological features. We showed coordinated expression of TIGIT in exhausted and regulatory T cells and Nectin in tumor cells. Chemo-resistant samples contain a threefold enrichment of inflammatory cancer-associated fibroblasts that upregulate metallothioneins. Our study reveals a deeper understanding of the intricate substructure of pancreatic ductal adenocarcinoma tumors that could help improve therapy for patients with this disease.

Pancreatic ductal adenocarcinoma (PDAC) has an 11% 5-yr survival rate¹ due to late detection, early metastases and therapy resistance^{2–6}. First-line treatment is surgery followed by radiation and/or chemotherapy^{7,8}, with immunotherapy options being limited^{9,10}. Drivers such as *KRAS*, *TP53*, *CDKN2A* and *SMAD4* (HUGO Gene Nomenclature Committee at the European Bioinformatics Institute, <https://www.genenames.org/>) have been identified¹¹ as have transcriptional subtypes of classical and basal-like^{12,13}.

Single-cell technologies enable analysis regardless of tumor content and facilitate dissection of the tumor microenvironment (TME), whose role in PDAC remains largely unknown. For instance, cancer-associated fibroblast (CAF) subtypes have been

identified and cytotoxic natural killer (NK) and CD8⁺ T cells are often numerically and functionally impaired^{14–17}. This creates an immunosuppressed, pro-tumorigenic environment, but how this occurs is poorly understood^{18,19}. There is a growing appreciation surrounding acinar-to-ductal metaplasia (ADM), in which acinar cells start expressing ductal markers. Animal models posit acinar cells as the origin of PDAC when *KRAS*(G12D) is expressed^{20–23}, but this hypothesis is difficult to evaluate in humans due to the paucity of acinar and ADM cells sampled at single-cell resolution^{24–28}. Recent efforts have focused on acinar heterogeneity in chronic pancreatitis²⁹ and healthy human pancreas³⁰, but adequate sampling of ADM cells is still lacking.

A full list of affiliations appears at the end of the paper.

As part of the Human Tumor Atlas Network consortium, we used a spatially distinct, multi-sampling approach to analyze 83 PDAC samples across 31 patients³¹. Samples are physically separate from one another, which allowed interrogating both inter- and intra-tumor heterogeneity via extensive omics, including bulk DNA and RNA sequencing (RNA-seq), bulk proteomics and phosphoproteomics, single-cell and single-nucleus RNA-seq (scRNA-seq and snRNA-seq, respectively), cellular imaging and spatial transcriptomics. We identified and validated transitional populations and their associated molecular signatures along the spectrum from normal pancreas to PDAC that were previously proposed in mouse models. We characterized differential impact of chemotherapy on the abundance and transcriptional programs of tumor and stroma populations using multi-omic approaches. We highlight the necessity of spatial sequencing for polyclonal/heterogeneous PDAC tumor characterization.

Results

Study design and overview of the study cohort. We collected 73 PDAC samples from 21 patients undergoing standard treatment, including four normal adjacent tissue samples. Treatment groups included seven treatment-naïve cases, eight neoadjuvant FOLFIRINOX (a treatment regimen comprising folic acid, 5-fluorouracil, irinotecan and oxaliplatin) cases, four neoadjuvant gemcitabine + nab-paclitaxel cases, one mixed (FOLFIRINOX and gemcitabine + nab-paclitaxel) and one chemoradiation (Chemo-RT) case (Supplementary Table 1). Each tumor was spatially sampled 2–4 times, with sample segments subsequently used to generate histologic, imaging and omics data; hematoxylin and eosin (H&E) slides; scRNA-seq; bulk mass spectrometry-based proteomics and phosphoproteomics; bulk whole-exome sequencing (WES); and bulk RNA-seq (Fig. 1a, Supplementary Table 2 and Methods). We generated scRNA-seq data for all 73 samples, WES for 64 samples and bulk RNA-seq for 65 samples. A subset ($n = 30$) underwent tandem mass tag (TMT) 11 proteomic and phosphoproteomic characterization (Fig. 1b). Following quality control, we clustered 232,764 cells across all samples based on expression profiles and assigned cell types based on marker gene expression (Fig. 1c, Extended Data Fig. 1a–c, Methods and Supplementary Note). Using the fraction of tumor cells as a proxy for tumor purity, estimates ranged from 0.10% to 82.69% across samples, with an average of 16.28%. H&E pathology review revealed that within-patient tumor content differences across samples averaged 24%, with a range of 5% to 64% (Extended Data Fig. 1d, Methods and Supplementary Note), consistent with tumor percentages from scRNA-seq (Pearson $R = 0.40$, $P = 0.001$). Principal component analysis (PCA) of bulk proteomic and phosphoproteomic data confirms that, while most within-tumor regions cluster closely, several specimens from the same tumor have substantial intra-tumor heterogeneity (Extended Data Fig. 1e,f).

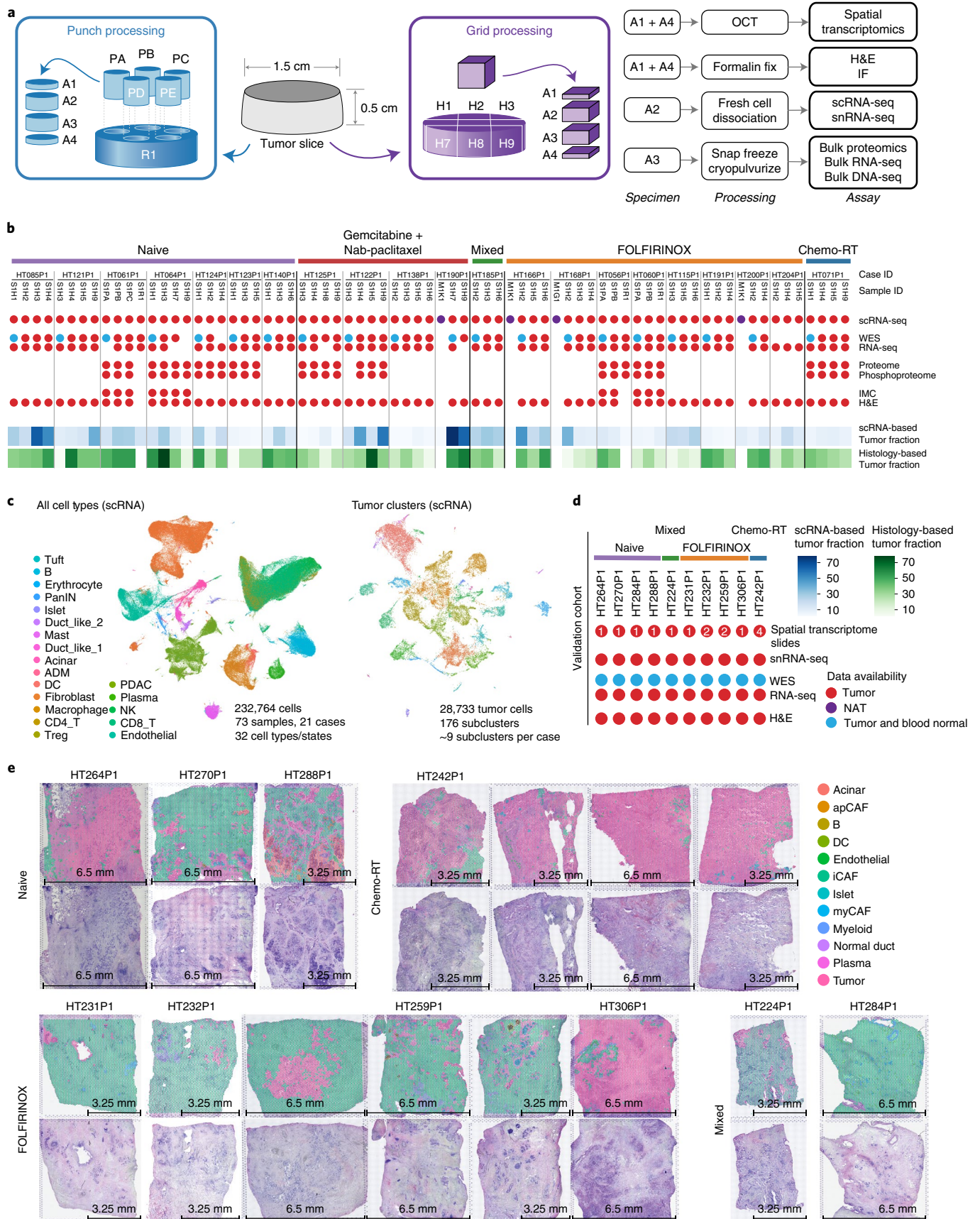
We further generated snRNA-seq with matching spatial transcriptomics, RNA-seq and WES data for an additional 10 cases, bringing the total cohort to 31 patients, 83 sc/snRNA-seq samples and 15 spatial

transcriptomics slides (Fig. 1d). The treatment groups included three treatment-naïve cases, four neoadjuvant FOLFIRINOX cases, one mixed (FOLFIRINOX and gemcitabine + nab-paclitaxel) and one Chemo-RT case. Following quality control, we assigned cell types to 83,860 nuclei based on marker gene expression and used the paired snRNA-seq to label spots in the spatial transcriptomics slides (Fig. 1d,e and Methods).

PDAC tumor subclusters with distinct cellular functions. Pathway enrichment analysis between case-level tumor subpopulations to dissect tumor heterogeneity (Fig. 2a and Methods) identified case-specific subpopulations enriched in pathways including cell proliferation, cell stress response, epithelial-to-mesenchymal transition and immune-related pathways that displayed spatial heterogeneity (Fig. 2a and Extended Data Fig. 1g). Actively proliferating clusters were present in most cases and were characterized by upregulation of genes belonging to the Molecular Signatures Database hallmark gene sets³² for E2F targets, G2M checkpoint, MYC targets and mitotic spindle. These clusters also exhibited increased oxidative phosphorylation, in line with previous reports³³ (Fig. 2a–c). It was common (15 of 21 cases) for tumor subclusters enriched in certain pathways to originate predominantly from only one of the spatially distinct samples from each case, such as S1H3 in HT185P1 and S1H4 in HT200P1 (Fig. 2d,e). Other sets of co-upregulated genes included ‘KRAS signaling up’ and ‘inflammatory response’, which lead to pancreatitis, pancreatic intraepithelial neoplasia (PanIN) and eventually PDAC³⁴. Increased expression of these sets occurred in samples with lower numbers of proliferating tumor cells, as demonstrated by clusters 7 and 0 from sample HT200P1_M1K1 and cluster 10 from HT185P1_S1H2 (Fig. 2b,c). KRAS-associated inflammatory response was expressed in clusters with increased expression of gene sets associated with cell stress (defined by the TP53 pathway, hypoxia and TNFA signaling via NFkB). This could indicate that parts of the tumor with the most actively proliferating cells were least impacted by KRAS-driven inflammation, or that tumor cells modulate their KRAS-driven associated inflammation during proliferation (Fig. 2d,e and Methods).

Using the spatial transcriptomics cohort, we characterized spatial heterogeneity by integrating histology features. Most slides had dense stroma intermingled with tumor populations (Fig. 1e). For HT264P1, six of eight tumor subclusters were mapped from snRNA-seq data to the spatial transcriptomics spots using cell-type label transfer and Robust Cell Type Decomposition (RCTD) to deconvolve the cell-type composition of each spot (Methods and Fig. 2f,g)³⁵. Subpopulation ‘Tumor_2’ clustered separately from other tumor cells and had a distinct ‘lower grade’ morphology (as annotated by pathology) (Fig. 2g–i). Gene and pathway enrichment analysis revealed that the ‘Tumor_2’ population upregulates fucosylation, hydroxylation and HIF pathways, and genes associated with the basal-like tumor subtype (*BTNL8*, *AGR3* and *LYZ*) (Fig. 2j,k). These findings suggest that this population is a spatially distinct cluster with different H&E morphology and likely more

Fig. 1 | Sampling strategy and cohort overview. **a**, Spatial sampling approach. At least two punches or grids were selected from each tumor for comprehensive imaging and omics characterization. Prefixes: ‘P’ denotes tissue punches, ‘H’ denotes tissue grids, ‘R’ denotes remainder tissue and ‘A’ denotes a piece of tissue. Each associated piece of tissue was processed in a systematic fashion and utilized for the listed assays. **b**, Top, data overview of the cohort. Samples are organized by treatment status. M1K1 and M1G1 denote normal adjacent tissue (NAT) samples. Dots for each associated assay indicate data availability: red for tumor tissue, purple for NAT, and blue for tumor tissue and blood normal. Bottom, scRNA-based (blue) and histology-based estimates (green) of tumor purity (Methods). **c**, Overview of all cell types profiled in the scRNA-seq cohort. The left UMAP shows that a total of 232,764 cells were profiled from 73 samples and 21 cases, representing 32 cell types or states (colored by cell types). The right UMAP shows the 28,733 tumor cells subset (colored by sample). **d**, Data overview of the validation cohort. Treatment status is listed on the top followed by sample name and data availability. Dots for each associated assay indicate data availability: red for tumor tissue, and blue for tumor tissue and blood normal. For spatial transcriptome slides, the number indicated in the circle denotes the number of spatial transcriptome slides generated from that sample. **e**, Overview of spatial transcriptomics cohort. Samples are organized according to treatment groups. For each sample, the bottom image shows the H&E and the top image has cell types overlaid. DC, Dendritic Cells; ID, Identification; IF, Immunofluorescence; IMC, Immunohistochemistry.



aggressive, highlighting the heterogeneous composition of tumor cells in PDAC in terms of pathology and transcriptomics. We identified ‘Tumor_4’ as a proliferative population that upregulates cell cycle pathways (Fig. 2j,k). Lastly, we characterized ‘Tumor_3’, which is scattered around the periphery of the large tumor regions in H&E (Fig. 2g,h). This population has high expression of laminin genes and is enriched in integrin and fibril formation pathways, which interact with the extracellular matrix and may be involved in tumor expansion (Fig. 2j,k)³⁶. In summary, we observe spatially and transcriptionally distinct tumor subpopulations from the same patient sample, within the same H&E section.

KRAS signaling and spatial drivers in pancreatic cancer. We observed substantial variation of driver mutation variant allele fractions (VAFs) between samples (Extended Data Fig. 1g and Supplementary Data 1). We detected a *KRAS* hotspot mutation in at least one sample in all cases, except for HT138P1 (HT204P1 lacks WES data). For low VAF (<0.01) mutations, 10 of 20 cases had samples whose mutation profiles differed from one another. We also identified pathogenic germline variants³⁷ (Methods). HT138P1 carried a pathogenic germline *BRCA2* variant. Three cases carried variants in the homology-directed DNA repair pathway (*FANCC* D23*, *BRCA2* I1470* and K607*, and *ATM* Y1124*), and, expectedly, all spatial samples carried the same variant in each case. Using RNA-seq, we classified samples into established subtypes^{12,38,39} and determined immune subtypes and stromal and immune compartment scores using xCell⁴⁰ and ESTIMATE⁴¹, respectively (Extended Data Figs. 1g and 2a,b, Supplementary Note and Methods).

Tumor cells grouped into patient-specific clusters, consistent with the genomic landscapes within each patient (Fig. 3a and Methods). We mapped mutation and copy number alterations to single cells (Fig. 3b and Methods). We tested the impact of *KRAS* hotspot variants by comparing gene expression profiles of each subset of tumor cells with a given *KRAS* mutation (Methods). Interestingly, tumor cells harboring *KRAS* G12V upregulate several genes associated with more aggressive or metastatic tumors, including *COL1A1*, *VIM* and *MUC5B* (Extended Data Fig. 2c)^{42–44}. We identified five cases with multiple *KRAS* hotspot drivers, which we interpret as synchronous primary tumor clones^{11,45,46} (Extended Data Fig. 2d). For case HT061P1, we obtained four subpopulations of clustered tumor cells, three small clusters largely derived from punch A and one large cluster common to all three punches (the remainder, as expected, represented all clusters) (Fig. 3c and Extended Data Fig. 2d,e). *KRAS* G12V cells faithfully map into one cluster from punch A predominant populations, with G12D cells mapping onto the large mixed cluster. Thus, two distinct clones carrying different *KRAS* driver mutations in the same patient are spatially separated, with differing gene expression profiles (Fig. 3c).

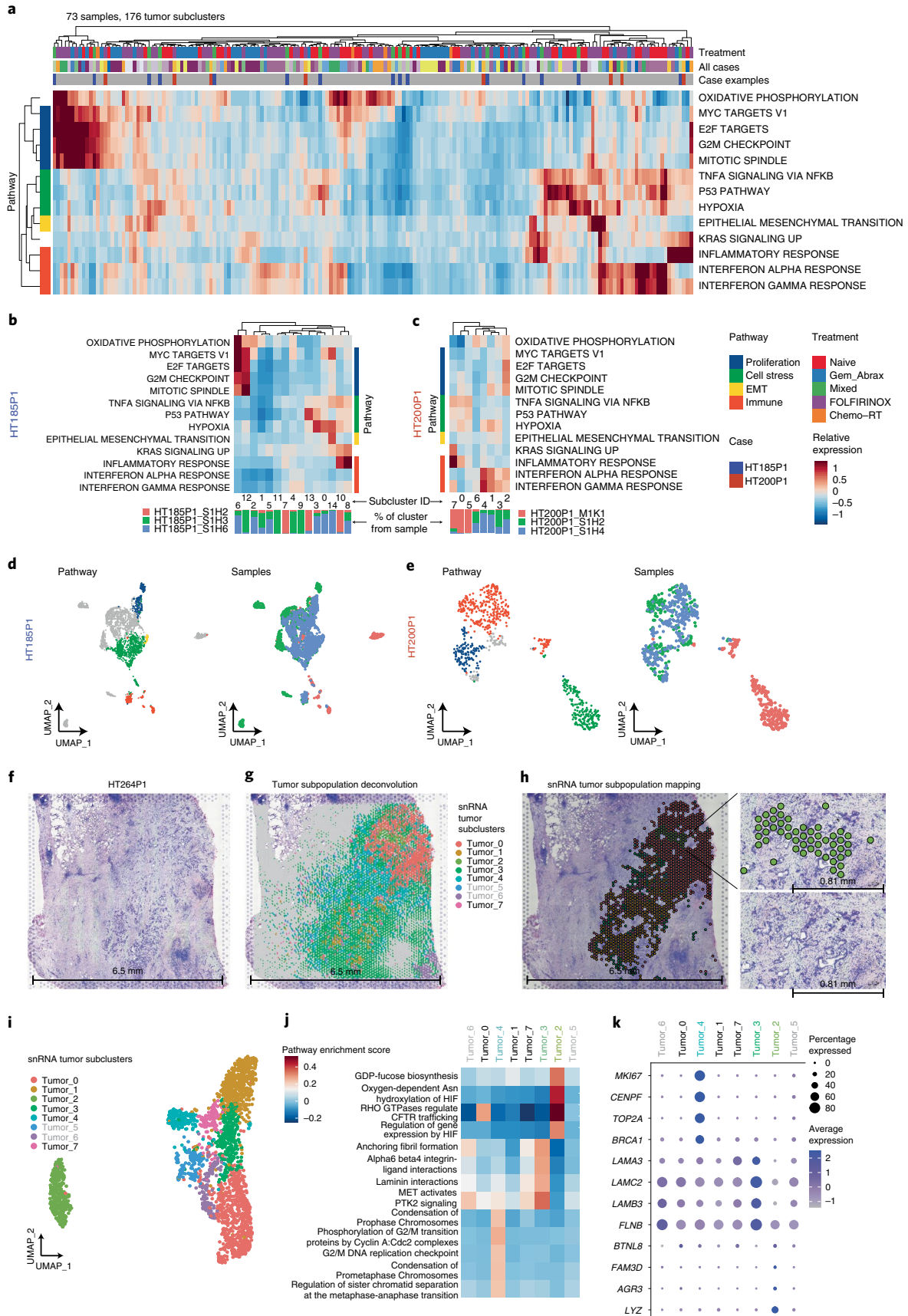
Using inferCNV⁴⁷, we identified copy number variation (CNV) signatures at both focal and arm levels that are unique to the respective *KRAS* subclones in case HT061P1 (Fig. 3c, Extended Data Fig. 2f and Methods). Copy number alterations were further evaluated using WES (Supplementary Data 2 and Methods). The G12D population shows amplifications of *AKT2* and *MYC*, while both G12D and G12V clusters harbor amplifications in *GATA6*, among others (Fig. 3c and Extended Data Fig. 2f). We reconstructed a lineage tree using MEDALT⁴⁸ (Fig. 3d and Methods), separating tumor cells into two major groups, each branching from a normal duct origin, consistent with both gene expression-based clustering and spatial origin of the cells (Fig. 3c,d). Together, we propose a model that integrates the gene expression and CNV data (Fig. 3e and Supplementary Note).

To assess proteomic heterogeneity, we conducted a global pairwise correlation analysis for the $n=30$ samples from 9 tumor cases (Methods and Supplementary Note). We determined the impact of mutations on downstream targets by analyzing the associated changes at protein and phosphorylation levels in several oncogenic pathways (Methods). We found a large degree of differential regulation, between and within tumors, in several phosphosites within the PI3K/PDK1/Akt and Raf/Mek/Erk pathways (Fig. 3f, Extended Data Fig. 2g–i and Supplementary Note).

Transitional populations between acinar and tumor cells. The hypothesis that PDAC arises from acinar cells that undergo ADM^{49–51} has been examined in mice, but not humans⁵². A major hurdle has been the small numbers of acinar and ADM cells that have been sampled from patients at single-cell resolution^{24,25}.

We identified populations of acinar cells expressing acinar markers (*PRSSI*, *CELA3A*) from multiple samples (Fig. 4a–d). The Acinar-REG⁺ cluster exhibits high expression of regenerating proteins^{25,30,53,54} thought to promote ADM and PanIN in PDAC^{55,56}. Two mixed populations of ductal cells^{30,54} lacked genomic alterations and maintained high expression of ductal markers (*CFTR*, *SLC44A4*, *ANXA4*, *SOX9*). Duct-like1 expresses *SPP1* and *CRP*, which have been observed in stressed cells and have progenitor-like features from the pancreatic ductal niche²⁵. Duct-like2 expresses normal ductal genes to a lesser extent and shows increased expression of mucus secretion (*MUC5B*) and trefoil factor genes. Highly expressed markers in Duct-like2 suggest these cells are a major source of malignant PDAC cells²⁴, and transcriptionally resemble cells identified in healthy pancreases³⁰. Finally, this cluster exhibits expression of *ONECUT2*, a transcription factor exclusively expressed in metaplastic cells derived from acinar origin in a mouse model⁵⁴. Peng et al.²⁴ hypothesized that subclusters of Duct-like2 could be PanIN-like, but we find that our PanIN and ADM populations are distinct from Duct-like2 and are only identified in our cohort (Extended Data Fig. 3a,b and Supplementary Note). PDAC exhibits high expression

Fig. 2 | Tumor subclusters with distinct cellular functions. **a**, Differential pathway enrichment case-level tumor subpopulations. Each column is a different tumor subcluster; the top bar indicates treatment and the bottom indicates case ID. Two samples are highlighted: HT185P1 (red) and HT200P1 (blue). The heatmap denotes relative expression of each pathway for each tumor subcluster, grouped by pathway similarity indicated on the left (Methods). **b**, Tumor cluster pathway enrichment for HT185P1. For panels **b** and **c**, each column indicates a tumor subcluster (ID is listed at the bottom of each column), below which is a bar plot indicating the percentage of the cluster that comes from each spatial sample. The heatmap is colored by the relative expression of each pathway. **c**, Tumor cluster pathway enrichment for HT200P1. **d**, UMAP of tumor subclusters for HT185P1. The left UMAP highlights tumor subclusters with high relative expression of associated pathways outlined in panel **b**. The right UMAP indicates tumor subclusters colored by spatial samples. **e**, UMAP of tumor subclusters for HT200P1. UMAPs are equivalent to those shown in panel **d** but for pathways in panel **c**. **f**, H&E of the section used for spatial transcriptomics for HT264P1. **g**, snRNA-seq tumor subcluster mapping using the RCTD deconvolution approach for each spatial transcriptomic spot (Methods). Gray regions in each pie chart denote nontumor cell types and grayed out tumor subpopulations (Tumor_5 and Tumor_6) represent subpopulations with minimal mapping to the H&E section. **h**, Left, simplified version of **g**, where each spot with high confidence was assigned a single tumor subpopulation identity. Right, magnification of the Tumor_2 subpopulation, which has a different morphology from the surrounding tumor. **i**, UMAP of tumor nuclei subclusters from the paired HT264P1 snRNA-seq data. **j**, Pathway enrichment analysis of tumor subclusters focusing on the Tumor_2, Tumor_3 and Tumor_4 subpopulations. **k**, Top DEGs of tumor subclusters focusing on the Tumor_2, Tumor_3 and Tumor_4 subpopulations. The size of each bubble indicates the percentage of cells expressing the gene of interest and color indicates average expression. EMT, Epithelial-mesenchymal transition.



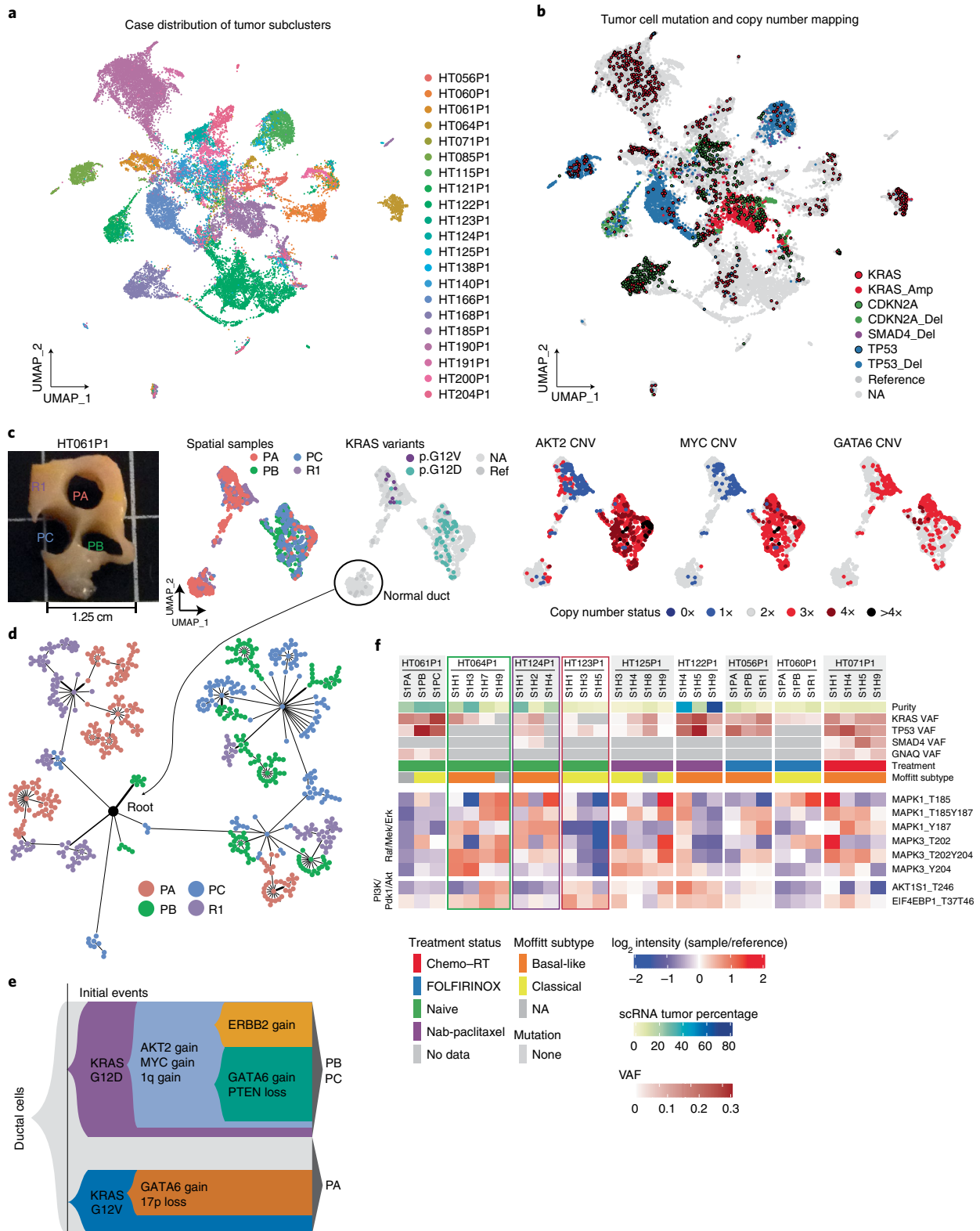


Fig. 3 | Genomic landscape and oncogenic driver heterogeneity. **a**, Tumor cell clusters labeled by case ID. **b**, Tumor cells labeled with genomic alterations including mutations and copy number alterations. Mutations are denoted by colored circles with a black outline, while deep copy number events (gain/loss of more than 1 copy) are denoted by colored circles without an outline. **c**, Ductal cells from case HT061P1. From left to right, tissue sample spatial locations; spatial sample IDs (R1 denotes the remainder tissue); KRAS variant mapping; and AKT2 CNV, MYC CNV and GATA6 CNV mapping. Spatial samples are labeled with a 'P' to denote punches and 'R' to denote remainder tissue. Copy number calls were obtained using inferCNV and CNV status is indicated by color. **d**, CNV-based lineage tree of a subset of ductal cells from HT061P1. **e**, Proposed model of tumor progression for HT061P1. **f**, Bulk phosphosite levels in the PI3K/Pdk1/Akt and Raf/Mek/Erk pathways. Cells filled in gray denote missing data. Samples with proteomics/phosphoproteomics all did not have mutations in *CDKN2A*. NA, No coverage for mutant or reference allele; Ref, Reference allele.

of *FXYD3*, *S100P* and *KRT17* in addition to copy number and driver mutations. PanIN-like cells were derived from 19 patients with a large proportion from sample HT168P1 (>1,700 cells) and, consistent with previous studies, we observe PDAC-initiating mutations in *KRAS* and *CDKN2A* (from HT168P1) within PanIN populations⁵². PanIN exhibits increased expression of extracellular matrix-related genes (*DCN*, *SPARC*, *SPON1*), a diversity of collagens⁵⁷, genes involved in acinar-to-ductal reprogramming (*KLF4*, *MMP7*)^{58,59} and other markers of early-stage malignancy (*CXCL12*, *TIMP3*, *ITGA1*, *MUC5AC*)^{60–63}.

We detected ADM cells in eight samples, including from cases HT122P1 and HT168P1 that represented a large proportion of ADM cells (>700 cells) (Supplementary Data 3). We found PDAC cells harboring several distinct alterations (Fig. 4a,b). Cells expressing acinar and ductal markers clustered separately from acinar cells and tended to be between acinar cells and normal ductal lineages on the Uniform Manifold Approximation and Projection (UMAP) (Duct-like1, Duct-like2), termed ADM_Normal, or between acinar and PanIN, denoted ADM_Tumor (Fig. 4b,c). While tumor and acinar cells express ductal and acinar markers in a mutually exclusive pattern, ADM cells express a combination, suggestive of an intermediate, reversible state (Fig. 4d). While both ADM_Tumor and ADM_Normal have decreased expression of acinar markers, they also have increased expression of PDAC markers and Duct-like1 markers, respectively. Following copy number analysis (CopyKAT, Methods), we found that a majority of predicted aneuploid cells are annotated as PanIN ($n = 524$), while only a handful ($n = 30$) in ADM_Tumor are labeled aneuploid (Fig. 4e,f). By mapping both *KRAS* and *CDKN2A* mutations, we identified several cells in the ADM_tumor population with either a *KRAS* mutation ($n = 1$) or *CDKN2A* mutation ($n = 7$), although this was not as widespread as the predicted PanIN populations (*KRAS*: 23 cells; *CDKN2A*: 163 cells) (Fig. 4g,h and Supplementary Data 3).

We examined whether acinar cells transition to different expression states (tumor or normal) by way of the two distinct ADM cell populations by performing Monocle analysis (Methods). We found two different transition states starting either with acinar cells transitioning towards the normal ductal cell route with ADM_Normal cells in between or with cells transitioning from acinar cells towards PanIN cells with ADM_Tumor cells in between (Fig. 4i). This suggests ADM_Normal is a transition state more similar to normal ductal cells and largely lacking genomic alterations, while ADM_Tumor is more related to PanIN and has a few alterations (for example, *CDKN2A*, aneuploidy). Recent studies in mice suggest acinar-derived tumors are preceded by PanINs, while ductal-derived tumors are PanIN independent⁶⁴.

Validation of ADM using snRNA-seq, immunohistochemistry and mouse models. We orthogonally surveyed two samples by snRNA-seq to see if cells expressing acinar and ductal features

could be identified from frozen tissue (Fig. 5a–c). ADM cells in HT288P1 and HT412P1 snRNA-seq samples have higher expression of Duct-like1 features than tumor cells, suggesting similarity to the ADM_Normal population in scRNA-seq.

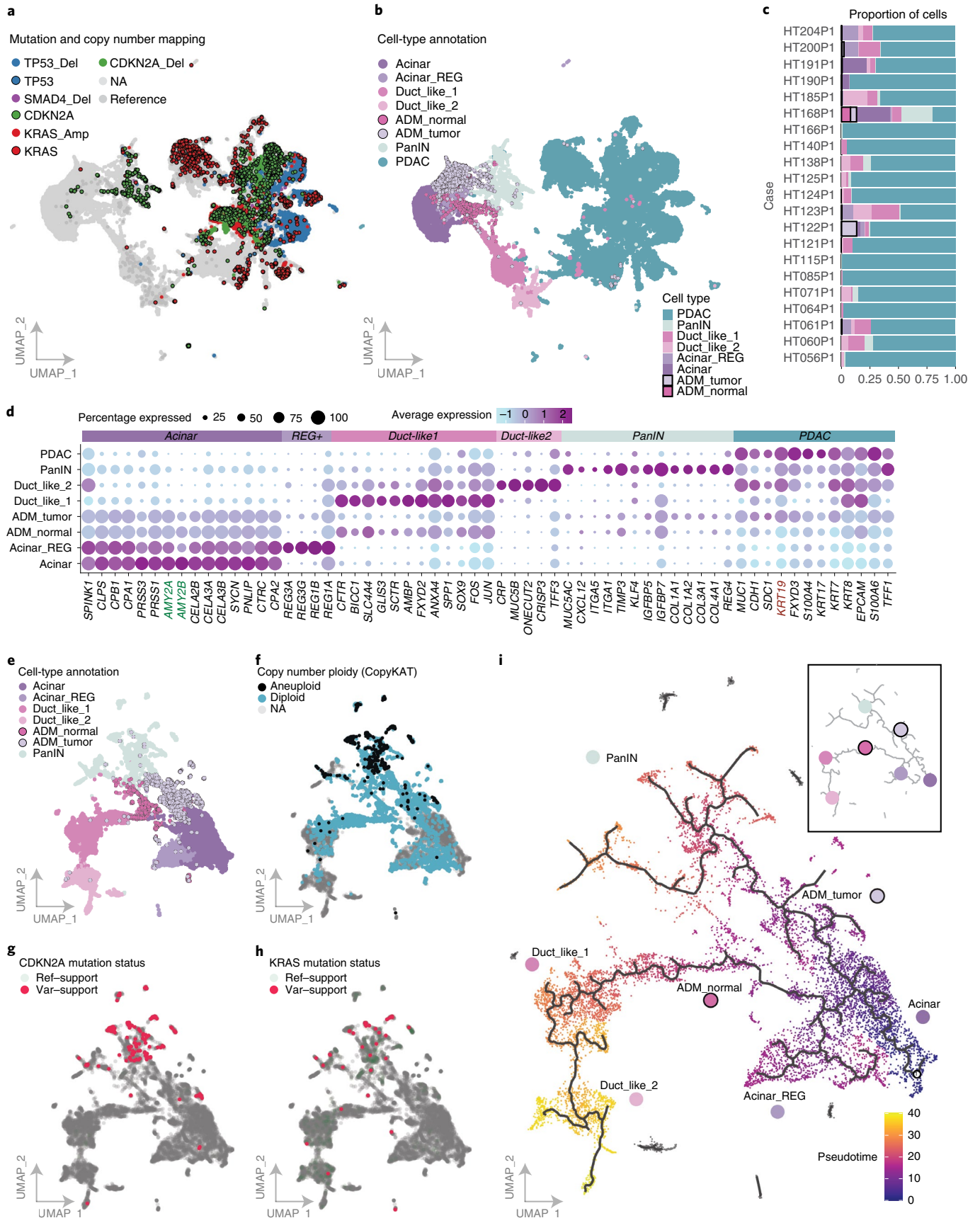
As validation, we performed immunofluorescence staining on tumor and normal formalin-fixed paraffin-embedded (FFPE) sections with amylase (acinar), cytokeratin-19 (ductal), Hoechst (nuclei) and Ki67 (proliferation) to evaluate co-staining patterns within individual cells (Fig. 5d, Extended Data Fig. 4a and Methods). In HT122P1 and HT288P1, we observe co-staining of acinar and ductal markers within multiple individual cells across several tumor regions. As controls, we provide a normal section with intermixed acinar and ductal cells, but lacking a co-staining expression pattern (HT288P1), and a tumor section that is predominantly stained by cytokeratin-19 (HT190P1). The paucity of ADM using immunofluorescence was recently validated in a tamoxifen-induced PDAC mouse model where acinar transformed ductal cells similarly co-stained for amylase and Cytokeratin-19 (ref. 65). Although rare, this co-staining pattern is confirmed at single-cell resolution in the same samples for which we performed immunofluorescence, thanks to our spatial sampling strategy. We identified two additional samples (HT412P1, HT434P1) with high acinar content with the same co-expression patterns (Extended Data Fig. 4b,c).

Finally, we performed scRNA-seq on eight mice from mouse models including induction of pancreatitis, *KRAS*-driven early- and late-stage transformation (KPC-OG GEMM mice) and normal pancreas tissues (Methods)⁶⁶. We identified two pancreatitis-acinar populations, one tumor-acinar population and one normal-acinar (Extended Data Fig. 5a–c). Markers within the pancreatitis-acinar populations overlap differentially expressed genes (DEGs) identified in another study⁶⁷ (Extended Data Fig. 5d). The tumor-acinar population from the KPC-OG model was the only acinar cluster with GFP expression (Extended Data Fig. 5b). Within this tumor model, GFP is associated with early transformation and metaplasia. While the tumor-acinar population expresses *Reg3a*, which is overexpressed in ADM regions⁶⁸, it also maintains high expression of *Sox9*, a ductal lineage marker in normal ductal and cancer cells, suggesting early-stage metaplasia.

These results support the identification of this rare population with acinar and ductal-like features seemingly lacking widespread genomic alterations. Our findings enable us to expand the proposed models of acinar origin to human PDAC development (Fig. 5e).

Transitional populations in histological features by spatial transcriptomics. Spatial transcriptomics enables the identification of histology features (for example, PanIN) to complement findings in sc/snRNA-seq. Each H&E image with associated spatial transcriptomics data was annotated by a pathologist (Extended Data Fig. 6a). Of ten paired spatial transcriptomics/snRNA-seq samples, HT288P1 presented with ADM (88 cells). We compared pathology

Fig. 4 | Acinar, ductal and transitional populations. **a**, UMAP clustering of acinar, ductal, transitional and PDAC tumor cells across the single-cell cohort. Mutations and deep copy number events were mapped to individual cells (Methods). Copy number events are indicated by colored dots with no outline, while an outlined black circle denotes a mutation event. **b**, UMAP clustering of acinar, ductal, transitional and PDAC tumor cells colored by cell-type annotation from single-cell RNA-seq samples. **c**, Proportion of cells identified as acinar, ductal, transitional and PDAC tumor cells by sample. **d**, Highly expressed genes identified in each cell type. The size of the bubble indicates the percentage of cells expressing the gene of interest and color indicates average expression. The bubble plot is ordered by expression across each cell-type group. *AMY2A/2B* and *KRT19* are indicated in bold because they are genes used for staining acinar cells and ductal cells, respectively, in the following immunofluorescence assays. ADM cells show expression of both of these markers in the scRNA data. **e**, Cell-type annotation and genomic alterations mapped across acinar, normal ductal, PanIN and ADM populations. UMAP of cell-type annotation indicates two distinct ADM populations annotated as ADM_Normal and ADM_Tumor. **f**, Copy number was annotated using CopyKAT, which predicts aneuploid cells independent of identifying tumor populations. Cells are colored by ploidy status. **g,h**, *CDKN2A* (**g**) and *KRAS* (**h**) mutation mappings are indicated in the lower two UMAPs and are colored by reference, variant and variant/reference supporting cells. **i**, Monocle pseudotime analysis indicates a cell-state transition from acinar cells to ADM_Tumor and ADM_Normal states independently. Each cell is colored by pseudotime which is a measurement of the change each cell is making through a process (for example, differentiation) and is annotated with a trajectory of change in the solid line overlaying the UMAP. Inset of the trajectory shows a summary of the cell-state transitions, with dots indicating cell type.



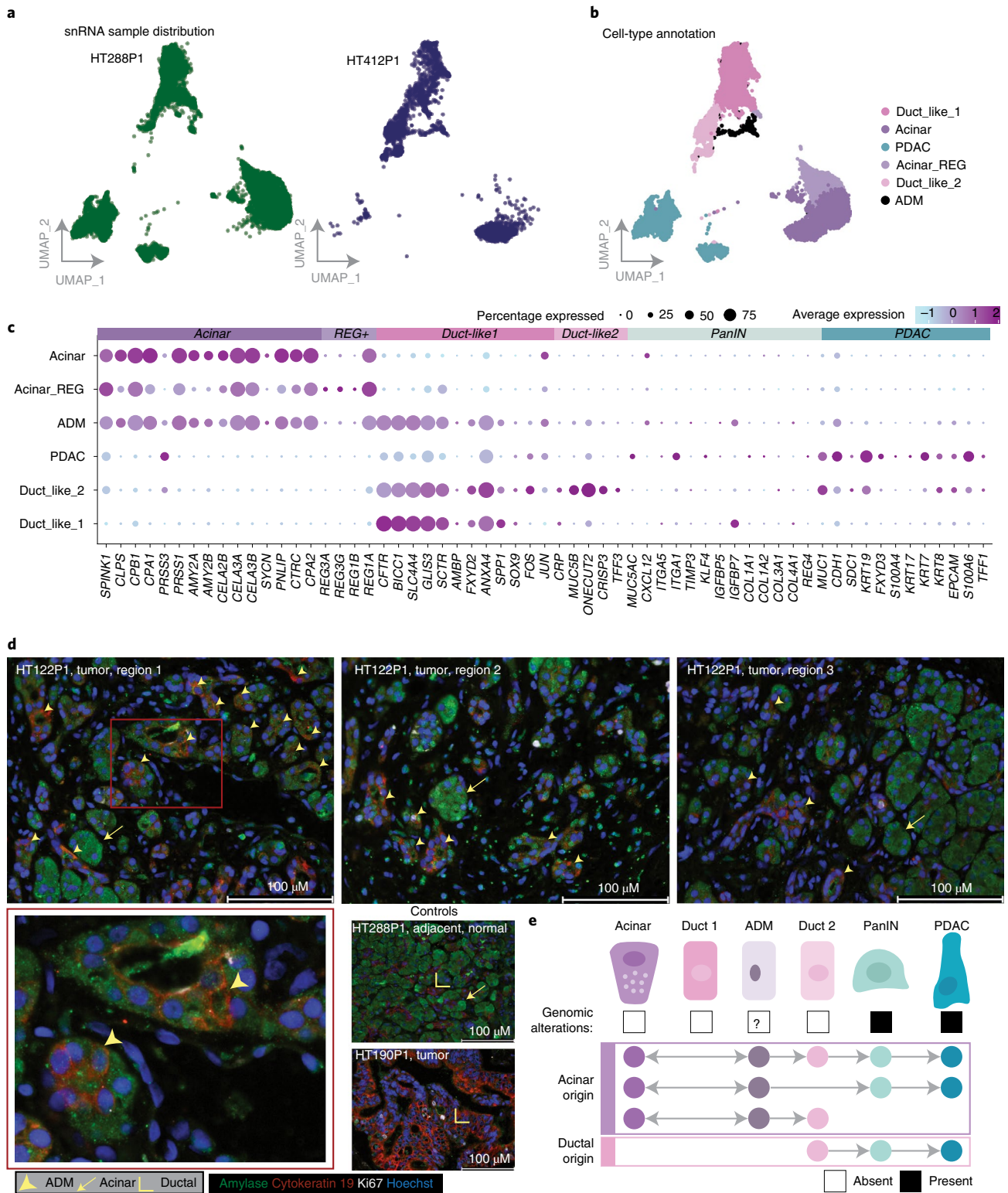


Fig. 5 | Validation of ADM using snRNA and immunofluorescence. a, UMAP plots of acinar and ductal cells from two cases, HT288P1 and HT412P1. Cells are colored by sample. **b**, UMAP plots of acinar and ductal cells colored by cell types. **c**, Gene expression signatures derived from scRNA data of acinar and ductal genes across cell types. Each dot indicates expression of a given gene in an annotated cell cluster. The size indicates the percentage of cells expressing that gene and the color is average expression. ADM cells show expression of both acinar and ductal markers in the snRNA data. **d**, Immunofluorescence staining of tumor and NAT sections. Amylase stains acinar cells (green), cytokeratin-19 stains ductal cells (red), Ki67 stains proliferating cells (white) and Hoechst stains nuclei (blue). For select sections, individual cells expressing both acinar and ductal markers, indicating ADM, are highlighted by the yellow arrowheads. Acinar cells are denoted with a yellow arrow and ductal cells with an outlined yellow angle. **e**, Proposed models of PDAC development. Development of PDAC along the spectrum from normal to PDAC in humans was initially suggested to be derived predominantly from ductal origin, but, with evidence of ADM cells in humans, an additional model of transition from acinar origin to PDAC is proposed.

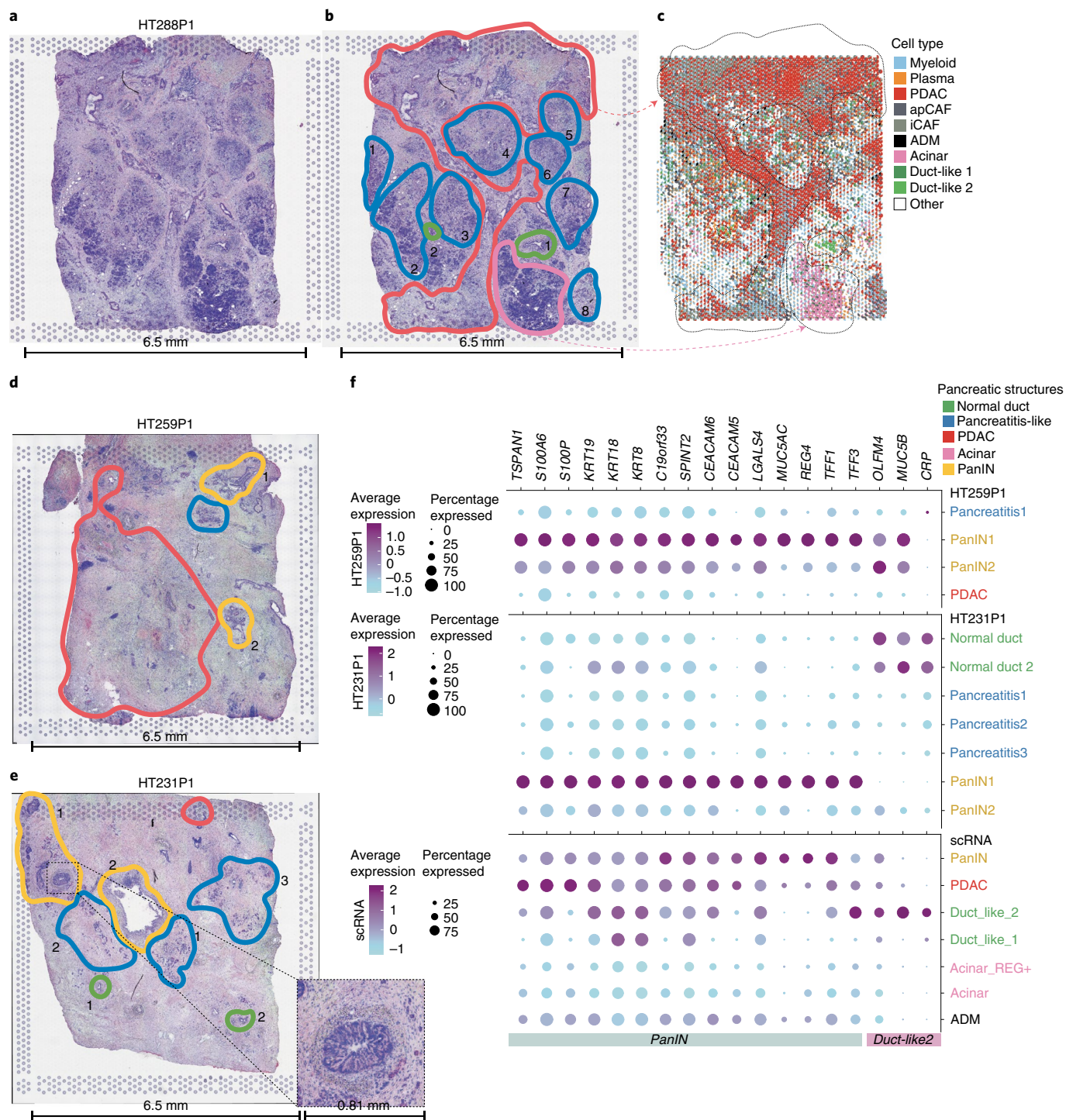


Fig. 6 | Tumor and transitional cell heterogeneity in spatial transcriptomics data. **a**, H&E of the section used for spatial transcriptomics for HT288P1. **b**, Pathologist-annotated regions of HT288P1. Regions highlighted include tumor (red), pancreatitis-like (blue), normal duct (green) and acinar (pink). Numbers are listed next to each annotated area if more than one are reported. **c**, snRNA-seq mapping using the RCTD deconvolution approach for each spatial transcriptomics spot (Methods). Each spot is colored by the expected representation of the cell types discovered from snRNA-seq. Regions of tumor, normal duct and acinar are highlighted with black dashes, showing similarity between pathology-assisted annotations and deconvolution. **d**, Pathologist-annotated regions of HT259P1. Numbers are listed next to each annotated area if more than one is reported. **e**, Pathologist-annotated regions of HT231P1. Numbers are listed next to each annotated area if more than one is reported. **f**, DEGs identified by spatial transcriptomics pathology-assisted regions with a focus on normal duct and PanIN. The size of the bubble indicates the percentage of cells expressing that target gene and color indicates average expression. Top, HT259P1 spatial transcriptome DEG analysis. Middle, HT231P1 spatial transcriptome DEG analysis. Bottom, spatial transcriptome-derived DEGs mapped to scRNA-seq acinar, ductal and transition populations.

annotations with spots defined via integration with snRNA-seq and with RCTD (Fig. 6a–c)³⁵. We observe strong concordance between pathology-defined and molecular regions for tumor, normal duct and acinar cells. Myeloid and plasma cells mapped throughout annotated pancreatitis regions, Duct-like2 mapped to annotated normal ductal structures and ADM mapped sparsely throughout the section. It remains difficult to validate ADM in spatial transcriptomics data since it has not yet reached single-cell resolution and ADM cells are rare. Under these circumstances, pancreatitis regions may appear to have both acinar and ductal features in agreement with Tosti et al.³⁰.

We annotated spatial transcriptomics spots in select samples with PanIN and a ductal structure within the capture area (Fig. 6a–e). Extracted spots were then subsetted from the full object and DEGs were analyzed. Isolated PanIN regions exhibited distinct DEGs that spanned multiple samples (Fig. 6f and Supplementary Data 4). DEGs identified with spatial transcriptomics data for normal duct, tumor and PanIN were compared against annotated scRNA-seq data, corroborating our initial annotations of PanIN, normal duct and tumor. Interestingly, the two cases having multiple PanIN regions identified by spatial transcriptomics exhibited distinct DEGs that differed by each uniquely annotated region. Our combined analysis strongly supports the presence of PanIN and existence of ADM in human samples.

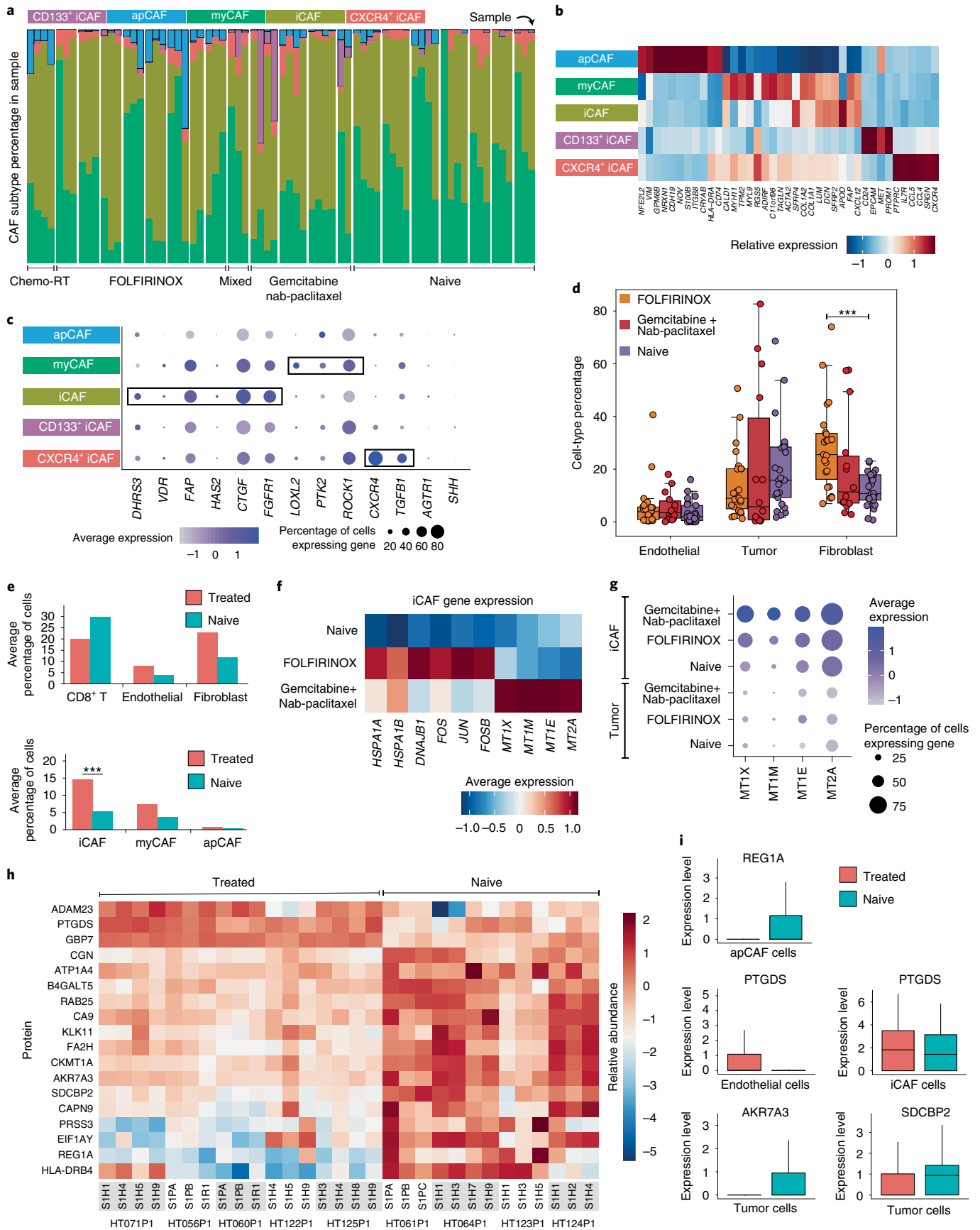
CAF subpopulations in PDAC TME. We identified three subtypes of CAFs: myofibroblastic CAFs (myCAFs), inflammatory CAFs (iCAFs) and antigen-presenting CAFs (apCAFs)^{14,15,69} (Fig. 7a and Extended Data Fig. 7a). We also observed subpopulations of iCAFs, denoted as CXCR4⁺ iCAFs and CD133⁺ iCAFs. Several CAF markers, including *ACTA2* and *FAP*, used to identify CAF subtypes are not definitive, being often expressed in both iCAFs and myCAFs⁷⁰. We identified the top DEGs between each CAF subtype; *TAGLN* and *ACTA2* discern myCAFs, *FAP* and *CXCL12* distinguish iCAFs, and apCAFs express *HLA-DRA* and *CD74* (refs.^{14,71}) (Fig. 7b and Extended Data Fig. 7b). CXCR4⁺ iCAFs and CD133⁺ iCAFs are defined by *CXCR4* and *CD133* (*PROM1*), respectively, although they also weakly express myCAF and apCAF marker genes. While most CAFs in tumors are iCAFs or myCAFs, the other CAF subtypes are present at low numbers throughout. These CD133⁺ iCAFs carry no genomic alterations, but express cancer stem cell markers, including *CD133*, *MET*, *EPCAM*, *CD24* and *CD44*. We observed high *CD44* expression in apCAFs and CXCR4⁺ iCAFs. *VIM* and *NFE2L2* were highly expressed in apCAFs, which were more abundant in treated samples ($P < 10^{-5}$) (Fig. 7b). These results suggest that small unique CAF subpopulations that express cancer-driving programs exist within standard CAF subtypes. We examined expression of CAF genes currently targeted by clinical trials registered since January 2020 (ref.¹⁵) (Fig. 7c). As treated samples have a depletion of myCAFs and enrichment of iCAFs, the effectiveness of additional therapies targeting CAFs may differ across treatment groups. Further, tumor-specific CAF clusters (relative to normal adjacent tissue) were enriched for TME-remodeling pathways (Extended Data Fig. 7c–e and Supplementary Note).

By assessing cell-type enrichment across treatment groups, we detected modest changes in endothelial and tumor cells and the largest difference within fibroblasts, where both treated groups had higher numbers than the treatment-naïve group (Fig. 7d). This is largely driven by a threefold higher level of iCAFs in FOLFIRINOX and Gemcitabine + nab-paclitaxel samples ($P < 10^{-3}$), with comparable myCAF abundance between treatment groups (Fig. 7e). As iCAFs are considered to be pro-tumorigenic⁷², this large increase of iCAFs after treatment may be associated with treatment resistance. We observed upregulation in heat shock genes, AP-1 pathway genes and metallothionein genes in treated iCAFs (Fig. 7f). As we only observe substantial expression of metallothionein genes in iCAFs, their prognostic value for predicting chemotherapy resistance originates from the stroma, rather than tumor cells⁷³ (Fig. 7g). Heat shock and AP-1 genes were more highly expressed in FOLFIRINOX samples, while metallothioneins were more highly expressed in Gemcitabine + nab-paclitaxel samples, suggesting iCAF heterogeneity based on treatment regimen (Fig. 7f). These observations suggest that treated tumors have much higher levels of iCAFs, which are potential targets for chemo-resistant tumors.

We identified 18 proteins having at least a twofold or greater change between treated and untreated samples (Methods and Supplementary Note). GBP6, PTGDS and ADAM23 were elevated in treated samples, while REG1A, EIF1AY, PRSS3 and HLA-DRB4 were elevated in naïve samples (Fig. 7h). While these proteins display overall differences between treated and naïve samples, we observe modest heterogeneity between spatial samples and between a subset of tumor cases (Methods). To assess whether these proteins are signals originating from tumor cells or the TME, we compared their expression profiles in scRNA data within each cell type among treatment groups. *REG1A* is upregulated in naïve apCAFs, while *PTGDS* is upregulated in treated endothelial and iCAF cells; neither were observed in tumor cells. (Fig. 7i). Only *AKR7A3* and *SDCBP2* were consistently upregulated strictly in tumor cells. These results suggest that several of these differentially abundant proteins may originate from the TME.

Immunosuppressive PDAC TME and treatment. To examine the immunosuppressed TME characteristics of PDAC¹⁸, we identified and reclustered immune cells into lymphocytes or myeloid/dendritic cells. In the latter group, we further distinguished between type I and II classical dendritic cells (cDC1, cDC2), macrophages, monocytes and neutrophils. Myeloid cells and classical dendritic cells strongly express TME-remodeling pathway genes, such as angiogenesis and hypoxia pathways, at higher levels than tumor cells (Extended Data Fig. 8a–c). While tumor cells do not have high expression of *NFE2L2* relative to myeloid cells, elevated expression occurs downstream of the Nrf2 pathway (*NQO1*, *GPX2*), which regulates oxidative damage repair (Fig. 8a). Such activation may be triggered via paracrine interactions with TME cells and would indicate that myeloid and dendritic cells contribute towards a pro-tumor TME. Within lymphocytes, we observed slight enrichment of CD4/CD8⁺ T cell subsets in treated samples and expression

Fig. 7 | CAF subpopulations across treatment groups. **a**, CAF subtype distribution across the cohort and across treatment groups. CAF subtype colors are indicated at the top and are consistent throughout panels **a–c**. **b**, Key markers and DEGs in each CAF subtype. **c**, Expression of genes currently targeted by clinical trials across CAF subtypes. The size of the bubble indicates the percentage of cells expressing the gene of interest and color indicates average expression. **d**, Cell-type percentage differences in tumor, endothelial and fibroblast cells among treatment groups. The mixed and Chemo-RT singleton cases were excluded in these analyses ($n = 41$ treated, $n = 25$ untreated). *** $P < 10^{-3}$ (exact $P = 0.0026360$), using a two-sided Wilcoxon rank sum test. The boxplots show the median with $1.5 \times$ interquartile range whiskers. **e**, Top, average cell-type percentages split into treated versus untreated groups. Bottom, average cell-type distributions of the main CAF subtypes (iCAF, myCAF and apCAF) split into treated versus untreated groups ($n = 41$ treated, $n = 25$ untreated). *** $P < 10^{-3}$ (exact $P = 0.0065984$), using a two-sided Wilcoxon rank sum test. **f**, DEGs between treated and untreated iCAFs. **g**, Expression of metallothionein genes across treatment groups in iCAFs and tumor cells. **h**, Top differentially expressed proteins across treated and untreated samples. **i**, Differential gene expression in specific cell types that match the proteins in panel **h** ($n = 41$ treated samples, $n = 25$ untreated samples). The boxplots show the median with $1.5 \times$ interquartile range whiskers.



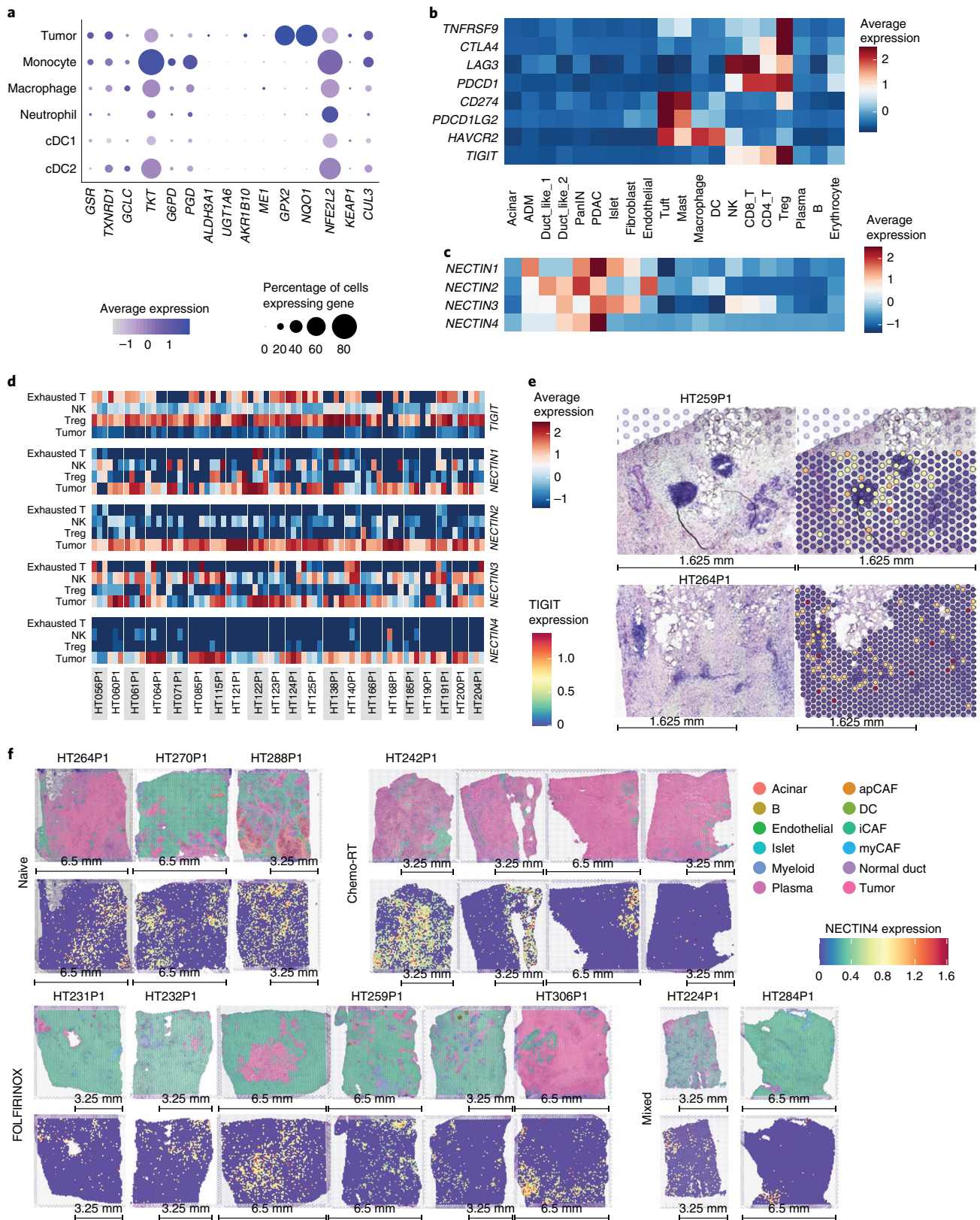


Fig. 8 | Myeloid and lymphocyte populations in the TME. a, Expression of Nrf2 pathway genes in myeloid and tumor cells. The size of the bubble indicates the percentage of cells expressing the gene of interest and color indicates average expression. **b**, Expression of immune checkpoint receptor and ligand genes across cell types. **c**, Expression of the four nectin receptors across all cell types. **d**, Average expression of *TIGIT*, *NECTIN1*, *NECTIN2*, *NECTIN3* and *NECTIN4* in exhausted T cells, NK cells, Tregs and tumor cells. Each column denotes a spatial sample and columns are grouped by case ID. **e**, *TIGIT* expression in lymphocyte-infiltrated regions from two spatial transcriptomics cases. **f**, *NECTIN4* expression colocalization with tumor spots across the spatial transcriptomics cohort. For each slide, the bottom image shows *TIGIT* expression and the top shows cell types. Tumor cells are colored bright pink.

of heat shock genes in FOLFIRINOX samples (Extended Data Fig. 8d,e,h and Supplementary Note).

We analyzed the expression of common immune checkpoint receptors and ligands, including PD-1 (*PDCD1*), *CTLA4* and *TIGIT* (Fig. 8b). We observed strong expression of exhaustion markers in NK, CD4⁺ and CD8⁺ T, and regulatory T (Treg) cells. We do not detect any significant expression of immune checkpoint genes in tumor cells or transitioning populations, including *PD-L1* and *PD-L2*, consistent with the poor response of PDAC to anti-*PD-L1*/*PD-L2* immunotherapy^{74,75}. Receptor–ligand analyses reveal interactions between the *TIGIT* receptor in lymphocytes and NECTIN ligands across all samples, which we found highly expressed in tumor cells, but somewhat less in other transitioning ductal populations (Fig. 8c and Methods). This is consistent with reports that *NECTIN4* is a potential target for immune checkpoint blockade^{76,77}. *TIGIT* interaction with NECTIN inactivates T and NK effector function, which the tumor could exploit for immune evasion^{76,77}. *NECTIN1* and *NECTIN4* are most strongly expressed in tumor cells, but *NECTIN2* and *NECTIN3* are also expressed in some lymphocyte cell types, while *TIGIT* is largely expressed in Tregs and exhausted CD4⁺ T cells (Fig. 8c and Extended Data Fig. 8i).

NECTIN1/2/3 are expressed in fibroblasts, endothelial cells and lymphocytes, respectively, while *NECTIN4* is the most tumor cell-specific NECTIN (Fig. 8c), consistent with previous reports⁷⁶. We analyzed *TIGIT* and NECTIN expression of individual samples and observed high expression of all NECTIN receptors in tumor cells and *TIGIT* in Tregs and exhausted T cells, and noted substantial heterogeneity across cases, particularly in *TIGIT* expression in exhausted T cells and in *NECTIN1* and *NECTIN3* expression in tumor cells (Fig. 8d). Finally, in the snRNA-seq cohort, we observe the same expression pattern of *TIGIT* and *NECTIN* across cell types (Extended Data Fig. 8j). Using spatial transcriptomics data, we focused on two regions in HT259P1 and HT264P1 to show the expression of *TIGIT* in spots proximal to the infiltrating lymphocyte regions (Fig. 8e). We find colocalization of tumor regions in the H&E with expression of *NECTIN4* across most H&E slides, regardless of treatment status (Fig. 8f). These results provide a rationale for targeting the *TIGIT*–*NECTIN* axis to improve anti-tumor T cell activity.

Discussion

Using bulk sequencing and proteomics/phosphoproteomics, single-cell sequencing, spatial transcriptomics and high-resolution cellular imaging on 83 PDAC samples, we identified transitional populations, including ADM and PanIN, and populations of nontransformed acinar and duct cells and PDAC cells. ADM populations express both oncogenes and tumor suppressor genes, significantly upregulating epithelial-to-mesenchymal transition and stem cell genes, compared with tumor cells⁷⁸. The unique expression pattern of ADM as an intermediate state suggests a dynamic transition between tumor and acinar fates and progression towards PDAC via acquisition of a driver *KRAS* event. This is consistent with acinar sensitivity to *KRAS* mutations as a catalyst of ADM and inclination toward PDAC⁷⁸. Driver mutations mapped to PanIN-like cells and tumor cells, consistent with their role as a precursor lesion. We used pathology-assisted spatial transcriptomics to identify distinct tumor subpopulations, as well as PanIN and PDAC-associated chronic pancreatitis. PanIN profiling by spatial transcriptomics provides direct confirmation of scRNA-seq findings.

CAFs are poorly understood in PDAC. Historically presumed to be cancer drivers, they are now known to have dual behavior as either drivers or suppressors of cancer, depending upon numerous factors^{79,80}. We identified iCAFs, myCAFs and apCAFs, further classifying two iCAF subsets as CD133⁺ and CXCR4⁺. We noted that several markers and activated pathways found to be differentially expressed between subtypes are being explored in current

clinical trials¹⁵. We observed higher iCAF abundance in treated samples. This is important, as IL-1-mediated and JAK-STAT signaling in iCAFs have motivated trials of adding IL-1R blockade to standard-of-care (FOLFIRINOX-based) chemotherapy⁷² (ClinicalTrials.gov: [NCT02021422](https://clinicaltrials.gov/ct2/show/study/NCT02021422)) and treating KPC mouse models with a JAK inhibitor decreases tumor size⁸¹. In patients treated with gemcitabine and nab-paclitaxel (either alone or in succession with other therapies), we observed upregulation of metallothionein genes in iCAFs. Metallothionein proteins are associated with resistance to a variety of chemotherapeutics, and may signal a chemoresistance mechanism⁸².

Immunotherapy has revolutionized treatment of many tumors, but is not yet effective for PDAC^{83–85}. Lack of immune checkpoint blockade activity in PDAC is multifaceted, including the shortage of naturally occurring T cell responses, partially due to the inhibition of effective T cell priming and/or T cell exclusion^{86,87}. Single-cell analysis revealed that NECTIN family members, especially *NECTIN4*, are tumor-specific, uncovering a potentially targetable interaction with *TIGIT* in Tregs and exhausted T cells. We observed high expression of all NECTIN genes in tumor cells and *TIGIT* in Tregs and exhausted T cells, but noted substantial heterogeneity across cases. Clarifying the key elements of the immunosuppressive PDAC microenvironment may pave the way for effective immunotherapy in PDAC.

Our study provides a comprehensive analysis of PDAC spatial heterogeneity and treatment effects. We found substantial heterogeneity in PDAC, including spatially separated driver clones, subtype heterogeneity within the same patients and multiple transitional cell populations, including duct-like, ADM and PanIN. Our work provides a resource to identify new targets of clinical relevance. We acknowledge the heterogeneous treatments in the patient population included in the study. Future work using clinical trials, specimens with uniform treatment regimens and comprehensive clinical response data will identify treatment-associated resistance signatures.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-022-01157-1>.

Received: 24 September 2021; Accepted: 13 July 2022;

Published online: 22 August 2022

References

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **72**, 7–33 (2022).
2. McGuigan, A. et al. Pancreatic cancer: a review of clinical diagnosis, epidemiology, treatment and outcomes. *World J. Gastroenterol.* **24**, 4846–4861 (2018).
3. Saad, A. M., Turk, T., Al-Husseini, M. J. & Abdel-Rahman, O. Trends in pancreatic adenocarcinoma incidence and mortality in the United States in the last four decades; a SEER-based study. *BMC Cancer* **18**, 688 (2018).
4. Rawla, P., Sunkara, T. & Gaduputi, V. Epidemiology of pancreatic cancer: global trends, etiology and risk factors. *World J. Oncol.* **10**, 10–27 (2019).
5. Ilic, M. & Ilic, I. Epidemiology of pancreatic cancer. *World J. Gastroenterol.* **22**, 9694–9705 (2016).
6. Viale, P. H. The American Cancer Society's facts & figures: 2020 edition. *J. Adv. Pract. Oncol.* **11**, 135–136 (2020).
7. Conroy, T. et al. FOLFIRINOX or gemcitabine as adjuvant therapy for pancreatic cancer. *N. Engl. J. Med.* **379**, 2395–2406 (2018).
8. Kang, J. et al. Nab-paclitaxel plus gemcitabine versus FOLFIRINOX as the first-line chemotherapy for patients with metastatic pancreatic cancer: retrospective analysis. *Invest. New Drugs* **36**, 732–741 (2018).
9. Morrison, A. H., Byrne, K. T. & Vonderheide, R. H. Immunotherapy and prevention of pancreatic cancer. *Trends Cancer* **4**, 418–428 (2018).

10. Balachandran, V. P., Beatty, G. L. & Dougan, S. K. Broadening the impact of immunotherapy to pancreatic cancer: challenges and opportunities. *Gastroenterology* **156**, 2056–2072 (2019).
11. Raphael, B. J. et al. Integrated genomic characterization of pancreatic ductal adenocarcinoma. *Cancer Cell* **32**, 185–203 (2017).
12. Moffitt, R. A. et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**, 1168–1178 (2015).
13. Maurer, C. et al. Experimental microdissection enables functional harmonisation of pancreatic cancer subtypes. *Gut* **68**, 1034–1043 (2019).
14. Elyada, E. et al. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discov.* **9**, 1102–1123 (2019).
15. Sahai, E. et al. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat. Rev. Cancer* **20**, 174–186 (2020).
16. Schnurr, M. et al. Strategies to relieve immunosuppression in pancreatic cancer. *Immunotherapy* **7**, 363–376 (2015).
17. Looi, C.-K. et al. Therapeutic challenges and current immunomodulatory strategies in targeting the immunosuppressive pancreatic tumor microenvironment. *J. Exp. Clin. Cancer Res.* **38**, 162 (2019).
18. Uzunparmak, B. & Sahin, I. H. Pancreatic cancer microenvironment: a current dilemma. *Clin. Transl. Med.* **8**, 2 (2019).
19. Ren, B. et al. Tumor microenvironment participates in metastasis of pancreatic cancer. *Mol. Cancer* **17**, 108 (2018).
20. De La O, J.-P. et al. Notch and Kras reprogram pancreatic acinar cells to ductal intraepithelial neoplasia. *Proc. Natl Acad. Sci. USA* **105**, 18907–18912 (2008).
21. Gidekel Friedlander, S. Y. et al. Context-dependent transformation of adult pancreatic cells by oncogenic K-Ras. *Cancer Cell* **16**, 379–389 (2009).
22. Habbe, N. et al. Spontaneous induction of murine pancreatic intraepithelial neoplasia (mPanIN) by acinar cell targeting of oncogenic Kras in adult mice. *Proc. Natl Acad. Sci. USA* **105**, 18913–18918 (2008).
23. Tuveson, D. A. et al. Mist1-KrasG12D knock-in mice develop mixed differentiation metastatic exocrine pancreatic carcinoma and hepatocellular carcinoma. *Cancer Res.* **66**, 242–247 (2006).
24. Peng, J. et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* **29**, 725–738 (2019).
25. Qadir, M. M. F. et al. Single-cell resolution analysis of the human pancreatic ductal progenitor cell niche. *Proc. Natl Acad. Sci. USA* **117**, 10876–10887 (2020).
26. Bernard, V. et al. Single-cell transcriptomics of pancreatic cancer precursors demonstrates epithelial and microenvironmental heterogeneity as an early event in neoplastic progression. *Clin. Cancer Res.* **25**, 2194–2205 (2019).
27. Lin, W. et al. Single-cell transcriptome analysis of tumor and stromal compartments of pancreatic ductal adenocarcinoma primary tumors and metastatic lesions. *Genome Med.* **12**, 80 (2020).
28. Raghavan, S. et al. Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. *Cell* **184**, 6119–6137.e26 (2021).
29. Blobner, B. M. et al. Single-cell analyses of human pancreas: characteristics of two populations of acinar cells in chronic pancreatitis. *Am. J. Physiol. Gastrointest. Liver Physiol.* **321**, G449–G460 (2021).
30. Tosti, L. et al. Single-nucleus and in situ RNA-sequencing reveal cell topographies in the human pancreas. *Gastroenterology* **160**, 1330–1344.e11 (2021).
31. Rozenblatt-Rosen, O. et al. The Human Tumor Atlas Network: charting tumor transitions across space and time at single-cell resolution. *Cell* **181**, 236–249 (2020).
32. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
33. Yao, C.-H. et al. Mitochondrial fusion supports increased oxidative phosphorylation during cell proliferation. *eLife* **8**, e41351 (2019).
34. Kitajima, S., Thummalapalli, R. & Barbic, D. A. Inflammation as a driver and vulnerability of KRAS mediated oncogenesis. *Semin. Cell Dev. Biol.* **58**, 127–135 (2016).
35. Cable, D. M. et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* **40**, 517–526 (2022).
36. Walker, C., Mojares, E. & Del Río Hernández, A. Role of extracellular matrix in development and cancer progression. *Int. J. Mol. Sci.* **19**, 3028 (2018).
37. Scott, A. D. et al. CharGer: clinical Characterization of Germline variants. *Bioinformatics* **35**, 865–867 (2019).
38. Collisson, E. A. et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat. Med.* **17**, 500–503 (2011).
39. Bailey, P. et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52 (2016).
40. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
41. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
42. Valque, H., Gouyer, V., Gottrand, F. & Desseyn, J.-L. MUC5B leads to aggressive behavior of breast cancer MCF7 cells. *PLoS ONE* **7**, e46699 (2012).
43. Niknami, Z., Eslamifar, A., Emamirazavi, A., Ebrahimi, A. & Shirkoobi, R. The association of vimentin and fibronectin gene expression with epithelial-mesenchymal transition and tumor malignancy in colorectal carcinoma. *EXCLI J.* **16**, 1009–1017 (2017).
44. Zhang, Z., Wang, Y., Zhang, J., Zhong, J. & Yang, R. COL1A1 promotes metastasis in colorectal cancer by regulating the WNT/PCP pathway. *Mol. Med. Rep.* **17**, 5037–5042 (2018).
45. Kulemann, B. et al. Pancreatic cancer: circulating tumor cells and primary tumors show heterogeneous KRAS mutations. *Sci. Rep.* **7**, 4510 (2017).
46. Hashimoto, D. et al. Heterogeneity of KRAS mutations in pancreatic ductal adenocarcinoma. *Pancreas* **45**, 1111–1114 (2016).
47. *infercnv*: Inferring CNV from Single-Cell RNA-Seq. v0.8.2 (Trinity CTAT Project, 2020).
48. Wang, F. et al. MEDALT: single-cell copy number lineage tracing enabling gene discovery. *Genome Biol.* **22**, 70 (2021).
49. Makohon-Moore, A. P. et al. Precancerous neoplastic cells can move through the pancreatic ductal system. *Nature* **561**, 201–205 (2018).
50. Murphy, S. J. et al. Genetic alterations associated with progression from pancreatic intraepithelial neoplasia to invasive pancreatic tumor. *Gastroenterology* **145**, 1098–1109.e1 (2013).
51. Kopp, J. L. et al. Identification of Sox9-dependent acinar-to-ductal reprogramming as the principal mechanism for initiation of pancreatic ductal adenocarcinoma. *Cancer Cell* **22**, 737–750 (2012).
52. Storz, P. Acinar cell plasticity and development of pancreatic ductal adenocarcinoma. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 296–304 (2017).
53. Muraro, M. J. et al. A single-cell transcriptome atlas of the human pancreas. *Cell Syst.* **3**, 385–394.e3 (2016).
54. Schlesinger, Y. et al. Single-cell transcriptomes of pancreatic preinvasive lesions and cancer reveal acinar metaplastic cells' heterogeneity. *Nat. Commun.* **11**, 4516 (2020).
55. Liu, X. et al. REG3A accelerates pancreatic cancer cell growth under IL-6-associated inflammatory condition: involvement of a REG3A-JAK2/STAT3 positive feedback loop. *Cancer Lett.* **362**, 45–60 (2015).
56. Li, Q. et al. Reg proteins promote acinar-to-ductal metaplasia and act as novel diagnostic and prognostic markers in pancreatic ductal adenocarcinoma. *Oncotarget* **7**, 77838–77853 (2016).
57. Crnogorac-Jurcic, T. et al. Molecular analysis of precursor lesions in familial pancreatic cancer. *PLoS ONE* **8**, e54830 (2013).
58. Crawford, H. C., Scoggins, C. R., Washington, M. K., Matrisian, L. M. & Leach, S. D. Matrix metalloproteinase-7 is expressed by pancreatic cancer precursors and regulates acinar-to-ductal metaplasia in exocrine pancreas. *J. Clin. Invest.* **109**, 1437–1444 (2002).
59. Wei, D. et al. KLF4 is essential for induction of cellular identity change and acinar-to-ductal reprogramming during early pancreatic carcinogenesis. *Cancer Cell* **29**, 324–338 (2016).
60. Demir, I. E. et al. Early pancreatic cancer lesions suppress pain through CXCL12-mediated chemoattraction of Schwann cells. *Proc. Natl Acad. Sci. USA* **114**, E85–E94 (2017).
61. Gharibi, A. et al. ITGA1 is a pre-malignant biomarker that promotes therapy resistance and metastatic potential in pancreatic cancer. *Sci. Rep.* **7**, 10060 (2017).
62. Thomas, R. M. et al. The chemokine receptor CXCR4 is expressed in pancreatic intraepithelial neoplasia. *Gut* **57**, 1555–1560 (2008).
63. Tian, C. et al. Proteomic analyses of ECM during pancreatic ductal adenocarcinoma progression reveal different contributions by tumor and stromal cells. *Proc. Natl Acad. Sci. USA* **116**, 19609–19618 (2019).
64. Ferreira, R. M. M. et al. Duct- and acinar-derived pancreatic ductal adenocarcinomas show distinct tumor progression and marker expression. *Cell Rep.* **21**, 966–978 (2017).
65. Mallya, K. et al. Acinar transformed ductal cells exhibit differential mucin expression in a tamoxifen-induced pancreatic ductal adenocarcinoma mouse model. *Biol. Open* **9**, bio052878 (2020).
66. Hegde, S. et al. Dendritic cell paucity leads to dysfunctional immune surveillance in pancreatic cancer. *Cancer Cell* **37**, 289–307.e9 (2020).
67. Boggs, K. et al. Pancreatic gene expression during recovery after pancreatitis reveals unique transcriptome profiles. *Sci. Rep.* **8**, 1406 (2018).
68. Zhang, H. et al. REG3A/REG3B promotes acinar to ductal metaplasia through binding to EXTL3 and activating the RAS-RAF-MEK-ERK signaling pathway. *Commun. Biol.* **4**, 688 (2021).
69. Helms, E., Kathrina Onate, M. & Sherman, M. H. Fibroblast heterogeneity in the pancreatic tumor microenvironment. *Cancer Discov.* **10**, 648–656 (2020).
70. Kraman, M. et al. Suppression of antitumor immunity by stromal cells expressing fibroblast activation protein- α . *Science* **330**, 827–830 (2010).

71. Öhlund, D. et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J. Exp. Med.* **214**, 579–596 (2017).
72. Hosein, A. N., Brekken, R. A. & Maitra, A. Pancreatic cancer stroma: an update on therapeutic targeting strategies. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 487–505 (2020).
73. Doz, F., Roosen, N. & Rosenblum, M. L. Metallothionein and anticancer agents: the role of metallothionein in cancer chemotherapy. *J. Neurooncol.* **17**, 123–129 (1993).
74. Feng, M. et al. PD-1/PD-L1 and immunotherapy for pancreatic cancer. *Cancer Lett.* **407**, 57–65 (2017).
75. Birnbaum, D. J. et al. Prognostic value of PDL1 expression in pancreatic cancer. *Oncotarget* **7**, 71198–71210 (2016).
76. Reches, A. et al. Nectin4 is a novel TIGIT ligand which combines checkpoint inhibition and tumor specificity. *J. Immunother. Cancer* **8**, e000266 (2020).
77. Gorvel, L. & Olive, D. Targeting the ‘PVR–TIGIT axis’ with immune checkpoint therapies. *F1000Res.* **9**, 354 (2020).
78. Xu, Y., Liu, J., Nipper, M. & Wang, P. Ductal vs. acinar? Recent insights into identifying cell lineage of pancreatic ductal adenocarcinoma. *Ann. Pancreat. Cancer* **2**, 11 (2019).
79. Gieniec, K. A., Butler, L. M., Worthley, D. L. & Woods, S. L. Cancer-associated fibroblasts—heroes or villains? *Br. J. Cancer* **121**, 293–302 (2019).
80. Pereira, B. A. et al. CAF subpopulations: a new reservoir of stromal targets in pancreatic cancer. *Trends Cancer Res.* **5**, 724–741 (2019).
81. Biffi, G. et al. IL1-induced JAK/STAT signaling is antagonized by TGF β to shape CAF heterogeneity in pancreatic ductal adenocarcinoma. *Cancer Discov.* **9**, 282–301 (2019).
82. Rodrigo, M. A. M. et al. Metallothionein isoforms as double agents – their roles in carcinogenesis, cancer progression and chemoresistance. *Drug Resist. Updat.* **52**, 100691 (2020).
83. Brahmer, J. R. et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N. Engl. J. Med.* **366**, 2455–2465 (2012).
84. O’Reilly, E. M. et al. Durvalumab with or without tremelimumab for patients with metastatic pancreatic ductal adenocarcinoma: a phase 2 randomized clinical trial. *JAMA Oncol.* **5**, 1431–1438 (2019).
85. Royal, R. E. et al. Phase 2 trial of single agent Ipilimumab (anti-CTLA-4) for locally advanced or metastatic pancreatic adenocarcinoma. *J. Immunother.* **33**, 828–833 (2010).
86. Bear, A. S., Vonderheide, R. H. & O’Hara, M. H. Challenges and opportunities for pancreatic cancer immunotherapy. *Cancer Cell* **38**, 788–802 (2020).
87. Stromnes, I. M., Hulbert, A., Pierce, R. H., Greenberg, P. D. & Hingorani, S. R. T-cell localization, activation, and clonal expansion in human pancreatic ductal adenocarcinoma. *Cancer Immunol. Res.* **5**, 978–991 (2017).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

¹Department of Medicine, Washington University in St Louis, St Louis, MO, USA. ²McDonnell Genome Institute, Washington University in St Louis, St Louis, MO, USA. ³Department of Surgery, Washington University in St Louis, St Louis, MO, USA. ⁴Siteman Cancer Center, Washington University in St Louis, St Louis, MO, USA. ⁵Department of Pathology and Immunology, Washington University in St Louis, St Louis, MO, USA. ⁶Department of Genetics, Washington University in St Louis, St Louis, MO, USA. ⁷Department of Mathematics, Washington University in St Louis, St Louis, MO, USA. ⁸Department of Otolaryngology–Head & Neck Surgery, Washington University in St Louis, St Louis, MO, USA. ⁹Department of Bioinformatics and Computational Biology, University of Texas MD Anderson Cancer Center, Houston, TX, USA. ¹⁰Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ¹¹Department of Radiation Oncology, Washington University in St Louis, St Louis, MO, USA. ¹²Department of Cell Biology and Physiology, Washington University in St Louis, St Louis, MO, USA. ¹³Department of Neurological Surgery, Washington University in St Louis, St Louis, MO, USA. ¹⁴Department of Radiology, Washington University in St Louis, St Louis, MO, USA. ¹⁵Department of Cell and Developmental Biology and Epithelial Biology Center, Vanderbilt University School of Medicine, Vanderbilt, TN, USA. ¹⁶Department of Computer Science and Engineering, Washington University in St Louis, St Louis, MO, USA. ¹⁷Department of Pathology, University of Texas MD Anderson Cancer Center, Houston, TX, USA. ¹⁸These authors contributed equally: Daniel Cui Zhou, Reyka G. Jayasinghe. ¹⁹These authors jointly supervised this work: David G. DeNardo, Ryan C. Fields, Li Ding. ✉e-mail: ddenardo@wustl.edu; rcfields@wustl.edu; lding@wustl.edu

Methods

Specimens and clinical data. All samples were collected with informed consent in concordance with the Washington University Institutional Review Board (IRB) at the Washington University School of Medicine in St Louis (St Louis, MO). Primary pancreatic adenocarcinoma samples were collected during surgical resection and verified by standard pathology (IRB protocol 201108117). Blood was collected at the time of surgery into vacuum tubes containing heparin or EDTA (BD Bioscience). Cells were isolated by Ficoll-density centrifugation and frozen in FBS with 5% dimethyl sulfoxide. Clinical data were captured in accordance with IRB protocol 20108117, at the time of informed consent, and entered into the REDCap database.

Sample processing. After verification by an attending pathologist, a $1.5 \times 1.5 \times 0.5$ -cm³ portion of the tumor was removed, photographed, weighed and measured. Each piece was then subdivided into 6–9 pieces (depending on the original size) and then further subdivided into four transverse cut pieces. Pieces were each then separately placed into formalin, snap-frozen in liquid nitrogen, DMEM and formalin, respectively. The purpose of the switch from punch sampling to the grid processing method was utility-based, as it minimized remainder tissue.

Genomic DNA and RNA extraction. Tumor tissues and corresponding normal mucosae were obtained from surgically resected specimens, and after a piece was removed for fresh single-cell prep the remaining sample was snap-frozen in liquid nitrogen and stored at -80°C . Before bulk RNA/DNA extraction, samples were cryo-pulverized (Covaris) and aliquoted for bulk extraction methods. Genomic DNA was extracted from tissue samples with either the DNeasy Blood and Tissue Kit (Qiagen, 69504) or the QIAamp DNA Mini Kit (Qiagen, 51304). Total RNA was extracted with TRI reagent (Millipore Sigma, T9424) and treated with DNase I (Qiagen, 79254) using an RNeasy MinElute Cleanup Kit (Qiagen, 74204). RNA integrity was evaluated using either a Bioanalyzer (Agilent Technologies) or TapeStation (Agilent Technologies). Genomic germline DNA was purified from cryopreserved peripheral blood mononuclear cells using the QIAAMP DNA Mini Kit (Qiagen, 51304) according to the manufacturer's instructions (Qiagen). The DNA quantity was assessed by fluorometry using the Qubit dsDNA HS Assay (Q32854) according to manufacturer's instructions (Thermo Fisher Scientific).

WES. First, 100–250 ng of genomic DNA was fragmented on the Covaris LE220 instrument targeting 250-base pair (bp) inserts. Automated dual-indexed libraries were constructed with the KAPA Hyper library prep kit (Roche) on the SciClone NGS platform (Perkin Elmer). Up to ten libraries were pooled at an equimolar ratio by mass before the hybrid capture targeting a 5- μg library pool. The library pools were hybridized with the xGen Exome Research Panel v1.0 reagent (IDT Technologies) that spans a 39-megabase (Mb) target region (19,396 genes) of the human genome. The libraries were hybridized for 16–18 h at 65°C followed by stringent washing to remove spuriously hybridized library fragments. Enriched library fragments were eluted and PCR cycle optimization was performed to prevent over amplification. The enriched libraries were amplified with KAPA HiFi master mix (Roche) before sequencing. The concentration of each captured library pool was accurately determined through quantitative PCR (qPCR) utilizing the KAPA Library Quantification Kit according to the manufacturer's protocol (Roche) to produce cluster counts appropriate for the Illumina NovaSeq-6000 instrument. Then, 2×150 paired-end reads were generated targeting 12 gigabases of sequence to achieve $\sim 100\times$ coverage per library.

RNA-seq. Total RNA integrity was determined using Agilent Bioanalyzer or 4200 TapeStation. Library preparation was performed with 500 ng to 1 μg of total RNA. Ribosomal RNA was blocked using FastSelect reagents (Qiagen) during complementary DNA synthesis. RNA was fragmented in reverse transcriptase buffer with FastSelect reagent and heated to 94°C for 5 min, 75°C for 2 min, 70°C for 2 min, 65°C for 2 min, 60°C for 2 min, 55°C for 2 min, 37°C for 5 min and 25°C for 5 min. mRNA was reverse transcribed to yield cDNA using SuperScript III RT enzyme (Life Technologies, per manufacturer's instructions) and random hexamers. A second strand reaction was performed to yield double-stranded cDNA (ds-cDNA). cDNA was blunt ended, had an A base added to the 3' ends and then had Illumina sequencing adapters ligated to the ends. Ligated fragments were then amplified for 15 cycles using primers incorporating unique dual index tags. Fragments were sequenced on an Illumina NovaSeq-6000 S4 instrument, generating approximately 30 million paired-end 2×150 reads per library.

Single-cell suspension preparation. For each tumor, approximately 15–100 mg of 2–4 sections of each tumor and/or normal piece of tissue were cut into small pieces using a blade and processed separately. Enzymes and reagents from the Human Tumor Dissociation Kit (Miltenyi Biotec, 130-095-929) were added to the tumor tissue along with 1.75 ml of DMEM. The resulting suspension was loaded into a gentleMACS C-tube (Miltenyi Biotec, 130-093-237) and subjected to the gentleMACS Octo Dissociator with Heaters (Miltenyi Biotec, 130-096-427). After 30–60 min on the heated dissociation program (37h_TDK_1), samples were removed from the dissociator and filtered through a 40- μm Mini-Strainer (PluriSelect no. 43-10040-60) or 40- μm Nylon mesh (Fisher Scientific, 22-363-547) into a 15-ml conical tube on ice. The sample was then spun down at 400g for 5 min at

4°C . After removing the supernatant, when a red pellet was visible, the cell pellet was resuspended using 200 μl to 3 ml of ACK Lysis Solution (Thermo Fisher, A1049201) for 1–5 min. To quench the reaction, 10 ml of PBS (Corning, 21-040-CM) with 0.5% BSA (Miltenyi Biotec, 130-091-376) was added and spun down at 400g for 5 min at 4°C . After removing supernatant, the cells were resuspended in 1 ml of PBS with 0.5% BSA, and live and dead cells were visualized using Trypan Blue. If over 40% of dead cells were present, the sample was spun down at 400g for 5 min at 4°C and subjected to the dead cell removal kit (Miltenyi Biotec, 130-090-101). Finally, the sample was spun down at 400g for 5 min at 4°C and resuspended in 500 μl to 1 ml of PBS with 0.5% BSA to a final concentration of 700 to 1,500 cells per μl .

Single-nuclei suspension preparation. First, 15–25 mg of pulverized tissue was placed in a 5-ml Eppendorf tube on ice. Using a wide-bore pipette tip (Rainin), a lysis buffer prepared from the Nuclei Isolation protocol (10x Genomics) and SuperRNase inhibitor (Invitrogen) was added to the tube. The tissue solution was gently pipetted until the lysis liquid turned a slightly cloudy color. (The number of pipetting iterations depended on the specific tissue.) The tissue homogenate was then filtered through a 40- μm strainer (pluriSelect) and washed with a BSA wash buffer (2% BSA + $1 \times$ PBS + RNase inhibitor). The filtrate was collected, centrifuged at 500g for 6 min at 4°C and resuspended with a BSA wash buffer. Then, 100 μl of cell lysis solution was set aside for unstained reference, while the rest was stained with 1 μl of 7AAD per 500 μl of the sample. Nuclei underwent FACS and sorting gates were based on size, granularity and dye staining signal. The final suspension was spun down at 500g for 6 min at 4°C , and resuspended with a BSA wash buffer.

Single-cell/nuclei library prep and sequencing. Utilizing the Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v.3.1 and Chromium instrument, approximately 17,500 to 25,000 cells were partitioned into nanoliter droplets to achieve single-cell resolution for a maximum of 10,000 to 15,000 individual cells per sample (10x Genomics, 1000269). The resulting cDNA was tagged with a common 16-nucleotide (nt) cell barcode and 10-nt Unique Molecular Identifier during the reverse transcriptase (RT) reaction. Full-length cDNA from poly-A mRNA transcripts was enzymatically fragmented and size-selected to optimize the cDNA amplicon size (approximately 400 bp) for library construction (10x Genomics). The concentration of the 10x single-cell library was accurately determined through qPCR (Kapa Biosystems) to produce cluster counts appropriate for the HiSeq 4000 or NovaSeq-6000 platform (Illumina). Then, 26×98 -bp sequence data were generated targeting 50,000 read pairs per cell, which provided digital gene expression profiles for each individual cell.

Spatial transcriptomics prep and sequencing. Optimal cutting temperature (OCT)-embedded tissues were cryosectioned and placed on a Visium Spatial Gene Expression Slide following Visium Spatial Protocols-Tissue Preparation Guide (10x Genomics, CG000240 Rev A). Briefly, fresh tissues were coated carefully and thoroughly with room temperature OCT without any bubbles. OCT-coated tissues were then placed on a metal block chilled in dry ice until the OCT turned solidified and white. After RNA quality check using TapeStation and morphology check using H&E staining for the OCT-embedded tissues, blocks were scored into a proper size that fit the Capture Areas and then sectioned into 10- μm sections. After the tissue placement into the Capture Area, sections were fixed in methanol, stained with H&E and imaged at $\times 20$ magnification using the brightfield imaging setting on a Leica DMI8 microscope. Tissues were then permeabilized for 18 min and Spatial Transcriptomics libraries were constructed following Visium Spatial Gene Expression Reagent Kits User Guide CG000239 Rev A (10x Genomics). Briefly, cDNA was reverse transcribed from the poly-adenylated messenger RNA which was captured by the primers on the slides. Next, the second strand was synthesized and denatured from the first strand. Free cDNA was then transferred from slides to tubes for further amplification and library construction. Libraries were sequenced on the S4 flow cell of the Illumina NovaSeq-6000 system.

KPC-OG GEMM mouse model. Three KPC-OG GEMM mice were killed at 3–5 months old, at a time when pathologically these mice have early metaplasia and PanIN throughout the pancreas, with only microscopic PDAC detectable^{66,88}. Age-matched KPC-OG negative littermates (CRE and OG negative) were treated with caerulein to induce acute pancreatitis by administering 6 hourly intraperitoneal injections (that is, once per hour for 6 h) at a dose of $100 \mu\text{g kg}^{-1}$ given every other day for 1 week. For normal, we extracted tissue from KPC-OG breeders negative for cre that underwent no treatment. Cell types were annotated from previous publications^{14,89}. All mice were bred and maintained under specific pathogen-free conditions, 12-h light/dark cycle, in accordance with the National Institute of Health and American Association for Accreditation of Laboratory Animal Care (NIH-AALAC) standards and consistent with Washington University School of Medicine Institutional Animal Care and Use Committee (IACUC) regulations (protocol no. 19-0856). Ethical approval for all mouse work was given by Washington University School of Medicine IACUC under protocol no. 19-0856.

Somatic variant calling. Somatic variants were called from whole-exome tumor-normal paired BAMs using somaticwrapper v.1.5, a pipeline designed for detection of somatic variants from tumor and normal WES data. The pipeline

merges and filters variant calls from four callers: Strelka v.2.9.2 (ref. ⁹⁰), VarScan v.2.3.8 (ref. ⁹¹), Pindel v.0.2.5 (ref. ⁹²) and MuTect v.1.1.7 (ref. ⁹³). SNV calls were obtained from Strelka, VarScan and MuTect. Indel calls were obtained from Strelka, VarScan and Pindel. The following filters were applied to obtain variant calls of high confidence: normal VAF ≤ 0.02 and tumor VAF ≥ 0.05 , read depth in tumor ≥ 14 and normal ≥ 8 , indel length < 100 bp, all variants must be called by 2 or more callers, all variants must be exonic and variants in dbSNP but not in COSMIC excluded.

KRAS hotspot and within-case genotyping. To verify manually and/or determine the KRAS mutation status at KRAS hotspots G12, G13 and Q61, we used bam-readcount. For each case, we first applied bam-readcount to generate readcounts for each of the nine bases in these loci and then calculated VAF values of all the KRAS hotspots based on reference and alternative base read counts at each position. Additionally, we manually verified every variant present in a sample in a pairwise fashion against other samples within the same case.

Germline variant calling and annotation. Germline variant calling was performed using an in-house pipeline, germlinewrapper v.1.1 (<https://github.com/ding-lab/germlinewrapper>), which implements multiple tools for the detection of germline INDELS and SNVs. Germline SNVs were identified using VarScan v.2.3.8 (with parameters: --min-var-freq 0.10 --p-value 0.10 --min-coverage 3 --strand-filter 1) operating on an mpileup stream produced by samtools v.1.2 (with parameters: -q 1 -Q 13) and GATK v.4.0.0.0 (ref. ⁹⁴) using its haplotype caller in single-sample mode with duplicate and unmapped reads removed and retaining calls with a minimum quality threshold of 10. All resulting variants were limited to the coding regions of the full-length transcripts obtained from Ensembl release 95 plus an additional 2 bp flanking each exon to cover splice donor/acceptor sites. We required variants to have allelic depth ≥ 5 reads and alternative allele frequencies $\geq 20\%$ in both the tumor and normal samples. We used bam-readcount v.0.8 for reference and alternative alleles quantification (with parameters: -q 10 -b 15) in both normal and tumor samples. Additionally, we filtered all variants with $\geq 0.05\%$ frequency in gnomAD v.2.1 (ref. ⁹⁵) and the 1000 Genomes Project⁹⁶.

Germline variant pathogenic classification. For annotation and prioritization of the filtered germline variants, we used our automatic variant classification tool CharGer v.0.5.4 (ref. ⁹⁷), which computes a classification score based on American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) guidelines. CharGer automatically marks as pathogenic those input variants that are marked as known pathogenic in ClinVar's curated database and marks as likely pathogenic those variants with a CharGer score > 8 . All pathogenic or likely pathogenic variants had both their normal and tumor samples reviewed manually by us using the Integrative Genomics Viewer software.

sc/snRNA-seq data preprocessing. For each sample, we obtained the unfiltered feature-barcode matrix per sample by passing the demultiplexed FASTQs to Cell Ranger v.3.1.0 'count' command using default parameters and the prebuilt GRCh38 genome reference v.3.0.0 (GRCh38 and Ensembl 93) for scRNA or the pre-mRNA version for snRNA. Seurat v.3.1.2 (refs. ^{97,98}) was used for all subsequent analyses. First, a series of quality filters was applied to the data to remove those barcodes that fell into any one of these categories recommended by Seurat: too few total transcript counts (< 300); possible debris with too few genes expressed (< 200) and too few unique molecular identifiers (UMIs) ($< 1,000$); possible more than one cell with too many genes expressed ($> 10,000$) and too many unique molecular identifiers ($> 10,000$); possible dead cell or a sign of cellular stress and apoptosis with too high proportion of mitochondrial gene expression over the total transcript counts ($> 10\%$). We constructed a Seurat object using the unfiltered feature-barcode matrix for each sample. Each sample was scaled and normalized using Seurat's 'SCTransform' function to correct for batch effects (with parameters: vars.to.regress = c('nCount_RNA', 'percent.mito'), variable.features = 2,000). Any merged analysis or subsequent subsetting of cells/samples underwent the same scaling and normalization method. Cells were clustered using the original Louvain algorithm⁹⁹ and top 30 PCA dimensions via 'FindNeighbors' and 'FindClusters' (with parameters: resolution = 0.5) functions. The resulting merged and normalized matrix was used for the subsequent analysis. Mouse data were aligned to refdata-gex-mm10-2020-A and GFP was added to the reference using the cellranger mkref function (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/tutorial_mr).

sc/snRNA-seq cell-type annotation. Main cell types were assigned to each cluster by manually reviewing the expression of a comprehensive set of marker genes (Supplementary Note). These assignments were all done by one person to maximize consistency.

Spatially distinct tumor cluster assignment. We used sample provenance of tumor cells as well as a requirement of 95% of cells in a tumor cluster to originate from a sample that is physically 6 mm from another sample to conclude that a subcluster is spatially distinct between samples (Supplementary Table 3).

scVarScan mutation mapping. We applied our in-house tool scVarScan that can identify reads supporting the reference and variant alleles covering the variant site in each individual cell by tracing cell and molecular barcode information in an scRNA bam file. For mapping, we used high-confidence somatic mutations from WES data. Additionally, we use cancerhotspots.org¹⁰⁰ to obtain the most common KRAS hotspot mutations at G12, G13 and Q61, and use scVarScan to detect potential minority KRAS mutations in each sample.

scVarScan statistics. To assess the degree of certainty that mutations were preferentially mapped to tumor cells versus nontumor cells (for which mappings can be reasonably assumed to be noise), we devised the following analysis based on the standard binomial difference of proportions test. Let X_T be the read count for mapped tumor mutations and let N_T be the total read count (mutation plus reference) for the tumor. Similarly, let X_N and N_N be the respective counts for the normal sample. The respective proportions of mapped reads for tumor and normal are clearly $P_T = X_T/N_T$ and $P_N = X_N/N_N$. Also, define the average joint fraction as $P_{\text{avg}} = (X_T + X_N)/(N_T + N_N)$ and its complement as $Q_{\text{avg}} = 1 - P_{\text{avg}}$. The large counts we are working with suggest the binomial distribution is well-approximated by the normal (Gaussian) distribution, as assessed by traditional heuristics $N_T P_{\text{avg}} Q_{\text{avg}} \geq 5$ and $N_N P_{\text{avg}} Q_{\text{avg}} \geq 5$. Adding the standard continuity correction (the normal distribution is continuous, whereas the binomial is discrete), we can then construct the following Z score for the difference of proportions:

$$Z_{\text{score}} = \frac{|P_T - P_N| - (1/N_T + 1/N_N)/2}{P_{\text{avg}} Q_{\text{avg}} \sqrt{1/N_T + 1/N_N}} \quad P = \Phi(Z_{\text{score}}),$$

which is normally distributed with mean 0 and variance 1. The P value for the one-sided test of whether the tumor proportion is statistically greater than the normal proportion is $\Phi(Z_{\text{score}})$, that is, the area under the standard Gaussian curve within the range $Z_{\text{score}} \leq Z < \infty$. We restrict performance of the test only to those cases where $P_T > P_N$ is actually observed, skipping cases of $P_T \leq P_N$, to avoid over-correcting in the calculation of false discovery rate (FDR). Using this method, we determined that the rate of mutations mapping is significant in the following comparison: tumor cells versus nontumor cells ($P \approx 0$), PanIN cells versus nontumor cells ($P \approx 0$), tumor cells versus PanIN cells ($P = 1.04 \times 10^{-11}$).

Single-cell RNA CNV detection. To detect large-scale chromosomal CNVs using single-cell RNA-seq data, inferCNV (v.0.8.2) was used with default parameters recommended for 10x Genomics data. All cells that are not tumor cells were pooled together for the reference normal set. InferCNV was run at a sample level and only with post-quality controlled filtered data. To calculate arm-level CNV events, we used an in-house script to match the gene-level inferCNV output to chromosome bands and take the mean value for each arm.

Single-cell mutation and CNV plotting. For clarity, we assigned each cell, represented by a single dot in a UMAP plot, with only one genetic alteration, in a hierarchical fashion. For clarity and to not overcomplicate plotting due to too many comparison groups, if a mutation and a copy number event are detected in the same cell, the cell is labeled with the mutation. Additionally, when multiple mutations or copy number events are detected in the same cell, we plot them hierarchically as follows: *KRAS* > *CDKN2A* > *SMAD4* > *TP53*.

Differential sc/snRNA expression analyses. For cell-level and cluster-level differential expression, we used the 'FindMarkers' or 'FindAllMarkers' Seurat function as appropriate, using a minimum percentage of 0.25 (parameter min.pct = 0.25) and looking only in the positive direction, as lack of expression is harder to interpret due to the sparsity of the data. The resulting DEGs were then filtered for adjusted $P < 0.05$ and sorted by fold change. All differential expression analyses were carried out using the 'SCT' assay.

Tumor subcluster pathway analysis. To demonstrate tumor heterogeneity in merged scRNA/snRNA data, we first took subsets of tumor cells from each individual case and renormalized with Seurat function 'SCTransform'. We then found case-level clusters with Seurat functions 'FindNeighbors' and 'FindClusters' (top 20 PCA dimensions, resolution = 0.8). Clusters with fewer than 0.1% of total tumor cells across cases were excluded. For each case-level cluster, we found DEGs with function 'FindAllMarkers' with a minimum percentage (min.pct) of 0.1, a minimum percentage difference (min.diff.pct) of 0.1, positive log fold change and adjusted $P < 0.05$. For each DEG list, we ran an enrichment analysis using the function 'enricher' against the 50 MSigDB hallmark gene sets¹⁰¹. The universe background for enrichment analysis was composed of genes detected in more than 0.1% of total tumor cells across cases. For each pathway, genes shown as enriched with adjusted $P < 0.05$ in any cluster were used to calculate the pathway score. Finally, in the merged tumor cell object, we calculated the average expression of genes identified in each pathway, centered and scaled across all clusters as the final score. To present pathways that distinguish tumor clusters the most, we ranked tumor cell-related pathways by their occurrence shown as enriched significantly (adjusted $P < 0.05$) in the enrichment analysis and plotted the most common pathways. The final heatmap was generated with the pheatmap package using the Optimal Leaf Ordering clustering method from the seriation package.

Receptor–ligand interactions. We used the CellPhoneDB tool¹⁰² to detect significant pairs of receptor–ligand interactions between cell types. This comparison was done at the sample level using default parameters between tumor and lymphocyte cell types.

Monocle trajectory analysis. We used the Monocle3 tool (<https://cole-trapnell-lab.github.io/monocle3/>) to infer cell-type transition states among acinar, transitional, PanIN and normal ductal populations. Objects and trajectory mapping were obtained by following tutorials outlined by developers (<https://cole-trapnell-lab.github.io/monocle3/>).

CopyKAT. To predict copy number ploidy without tumor annotation we utilized CopyKAT (<https://github.com/navinlabcode/copykat>) and followed the standard tutorial to define populations of aneuploid tumor cells.

Spatial transcriptomics data preprocessing. For each sample, we obtained the unfiltered feature-barcode matrix per sample by passing the demultiplexed FASTQs and associated H&E image to Space Ranger v.1.1.0 ‘count’ command using default parameters with reorient-images enabled, and the prebuilt GRCh38 genome reference 2020-A (GRCh38 and Ensembl 98). Seurat v.4.0.3 was used for all subsequent analyses. We constructed a Seurat object using the ‘Load10X_Spatial’ function for every slide. Each slide was then scaled and normalized with the ‘SCTransform’ function to correct for batch effects (with parameters: vars.to.regress = c(‘nCount_Spatial’)). Any merged analysis or subsequent subsetting of cells/samples for a sample with several slides underwent the same scaling and normalization method. Spots were clustered using the original Louvain algorithm⁹⁹ and top 20 PCA dimensions via ‘FindNeighbors’ and ‘FindClusters’ functions.

sc/snRNA-seq cell-type annotation. For spot-level cell-type assignment, we used the Seurat functions ‘FindTransferAnchors’ and ‘TransferData’ to perform a cell-type label transfer from the paired snRNA-seq annotations to the spatial transcriptomics spots. For further resolution, we used RCTD to deconvolve cell types within a given spot⁴⁵. We used the default parameters in RCTD using the ‘multi’ mode and a minimum of 25 nuclei for each cell-type identity to deconvolve; <https://github.com/dmccable/spacexr>.

Manual spot selection. In select samples with PanIN identified and a tumor or normal ductal structure within the capture area, we annotated the spatial transcriptomics spots using the Loupe Browser 5.0 and the lasso tool to manually select and annotate groups of spots. Annotated spots were then used to annotate the UMAP object; then, annotated spots were subsetted from the full object and DEGs were calculated using Seurat (FindAllMarkers).

DNA and RNA sample quality control. Bulk sequencing data quality metrics (adaptor content, mapping quality, coverage and swaps/mislabeling) were determined for DNA and RNA bams using our in-house pipeline SeqQEst. The inclusion criteria for paired DNA and RNA bams with sufficient coverage was >50× coding region coverage in WES or >50 Mb mapped depth in RNA-seq data.

RNA quantification. We used our in-house bulk RNA-seq expression analysis pipeline for quantification. Briefly, for each sample, the raw sequence reads were aligned into BAM files using STAR¹⁰³ (v.2.7.4a) two-pass alignment with GRCh38 as the reference. The resulting BAM files were then quantified as a raw count-matrix using read feature counts using Subread¹⁰⁴ (v.2.0.1). For both alignment and quantification, gene annotations were based on Gencode v.34. The raw counts were then converted to FPKM-UQ based on GDC’s formula (https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/#upper-quartile-fpkm) and then log₂ transformed with 1 pseudocount.

Proteomic and phosphoproteomics quantification. Proteomic data processing followed the methods detailed by Clark et al.¹⁰⁵. Briefly, raw mass spectrometry files were converted into open mzML format, then searched using the MSFragger database against a RefSeq protein sequence database appended with an equal number of decoy sequences. The specific parameters and software are detailed in the Clark et al. 2020 study. We then used the ComBat function from the R sva package to correct for TMT batch effects¹⁰⁶.

Pathway analysis. For each comparison, we obtained the top 30 genes ranked by highest fold change that are significantly different between the comparison groups (FDR < 0.05). We used ConsensusPathDB-human for gene set over-representation analysis¹⁰⁷.

Statistics and reproducibility. Relevant statistics are referred to in each of the associated methods sections. We did not use statistical methods to predetermine a sample size and patients were not randomly selected, as they were enrolled as they passed through the clinic. We excluded samples that did not pass sample prep quality control. For all immunofluorescence imaging, at least three regions of each sample were assayed, but immunofluorescence staining was not repeated for the same sample sections.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All human sequencing and imaging data have been deposited via the Human Tumor Atlas Network (HTAN) dbGaP Study Accession: phs002371.v1.p1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002371.v1.p1). In addition, all data have been deposited to the HTAN Data Coordinating Center Data Portal at the National Cancer Institute: <https://data.humantumoratlas.org/> (under the HTAN WUSTL Atlas). References (GRCh38 genome reference v3.0.0 and refdata-gex-mm10-2020-A) used for single-cell analysis of the human and mouse genomes, respectively, are available from public sources, as described in access scripts freely furnished by 10x Genomics: <https://support.10xgenomics.com/single-cell-gene-expression/software/release-notes/build>. Mouse single-cell RNA-seq data are freely available from the National Library of Medicine BioProject (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession: PRJNA835747. Data for single-cell integration from Peng et al.²⁴ were downloaded from the Genome Sequence Archive (PRJCA001063).

Code availability

The code for the mutation mapping from bulk to single cells can be found at: <https://github.com/ding-lab/10Xmapping>. Code for inferCNV post processing can be found at: https://github.com/ding-lab/infer_cnv_postprocessing.git. Code for germline wrapper variant calling can be found at: <https://github.com/ding-lab/germlinewrapper>.

References

- Hingorani, S. R. et al. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).
- Hosein, A. N. et al. Cellular heterogeneity during mouse pancreatic ductal adenocarcinoma progression at single-cell resolution. *JCI Insight* **5**, e129212 (2019).
- Kim, S. et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594 (2018).
- Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
- Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
- Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
- McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
- Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
- 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
- Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
- Chang, M. T. et al. Accelerating discovery of functional mutant alleles in cancer. *Cancer Discov.* **8**, 174–183 (2018).
- Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
- Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
- Clark, D. J. et al. Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell* **180**, 207 (2020).
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

107. Kamburov, A., Stelzl, U., Lehrach, H. & Herwig, R. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.* **41**, D793–D800 (2013).

Acknowledgements

We thank the patients, staff and scientists who contributed to this study as well as the NCI and the Human Tumor Atlas Network (HTAN) consortium. All HTAN consortium members are named at humantumoratlas.org. We also thank the Siteman Cancer Center and the McDonnell Genome Institute for their support. The following grants supported this work: grant no. U2CCA233303 to L.D., R.C.F., W.E.G. and S.T.O.; grant no. U24CA211006 to L.D.; grant no. U24CA209837 to K.I.S.; grant no. R01HG009711 to L.D. and F.C.

Author contributions

L.D., R.C.F., D.G.D., W.E.G., F.C. and S.T.O. conceived and designed the study. R.G.J., J.M.H., C.C.F., L.T., E.L.A., S.E.C., X.Y., K.S., A.N.R., J.P., L.C., F.C., R.S.F., C.C.F., R.V., M. Serasanambati, K.J.R., C.O., L.C., S.V.P., D.C., W.G.H., S.C., N.N., A.S., S.C., W.C., J.M.B. and P.L. developed and performed experiments or data collection. D.C.Z., R.G.J., E.S., C.M., Y.W., N.N., M.A.W., L.W., Y.L., C.Z., R.L., N.V.T., H. Zhu, H.S., F.W., M. Schnaubelt, H. Zhang, J.L., A.N.S., M.C.W. and R.J.L. performed computation and statistical analyses. D.C.Z., R.G.J., E.S., J.M.H., C.Z., K.J.R., D.C., M.G.C., D.G.D.,

W.E.G., R.C.F., L.D., J.F.M., K.C., H. Zhang, L.Y., P.N., J.K.S., M.D.I., J.L., A.N.S. and A.S. performed data interpretation and biological analyses. D.C.Z., R.S.F., L.A.F., J.F.M., M.C.W., R.J.M., S.V.P., A.H.K., S.S., K.I.S., T.J., W.G.H., K.C., H. Zhu, D.C., M.G.C., S.A., D.G.D., S.T.O., F.C., W.E.G., R.C.F., L.D., N.V.T., N.N., M.D.I. and K.S.L. wrote, reviewed and edited the manuscript. L.D., R.C.F., W.E.G., F.C., S.T.O., S.R.D., E.L.A., L.A.F., C.C.F., R.S.F., J.P., A.H.K., K.I.S., T.J. and S.S. were responsible for project administration.

Competing interests

The authors declare no competing interests.

Additional information

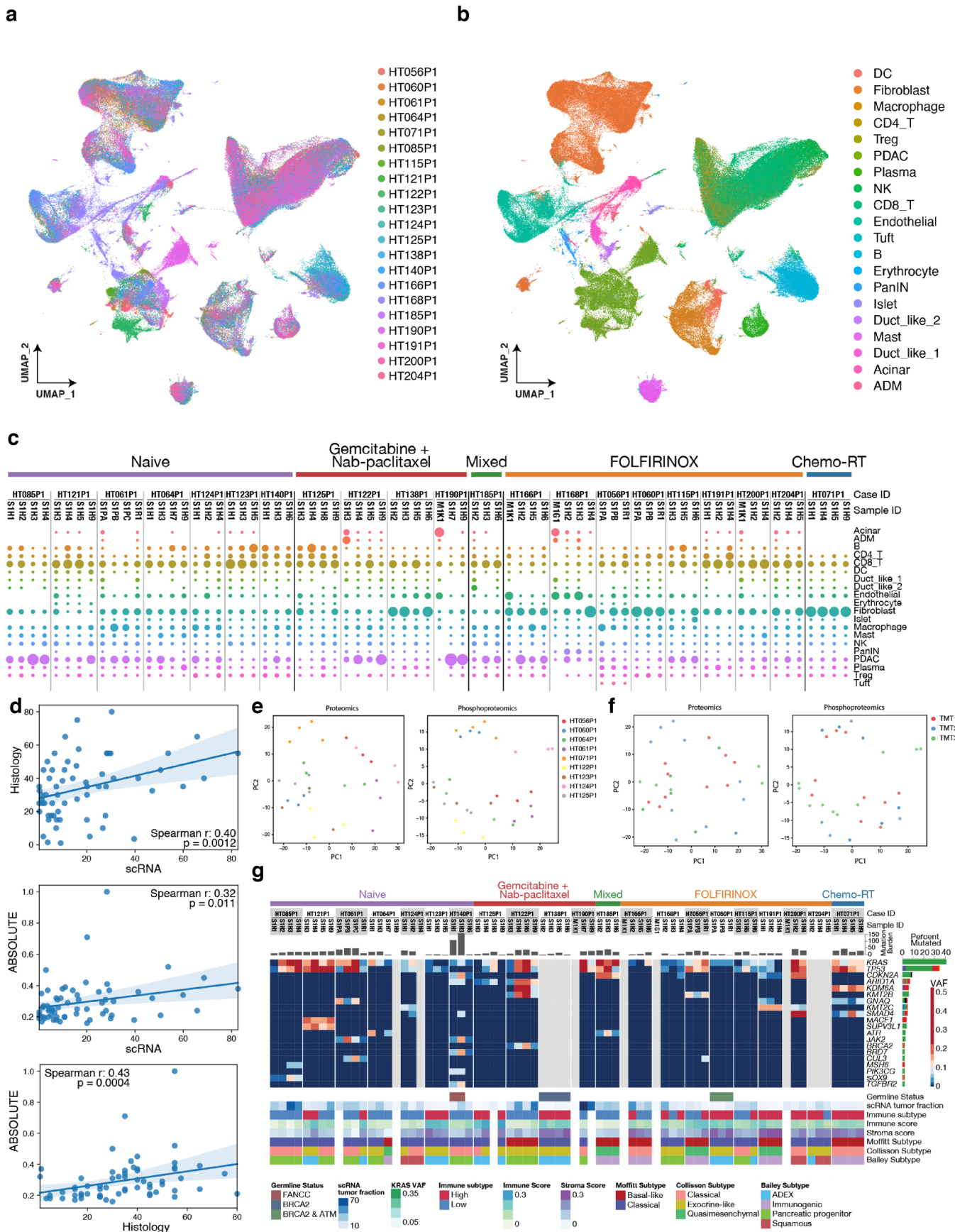
Extended data is available for this paper at <https://doi.org/10.1038/s41588-022-01157-1>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-022-01157-1>.

Correspondence and requests for materials should be addressed to David G. DeNardo, Ryan C. Fields or Li Ding.

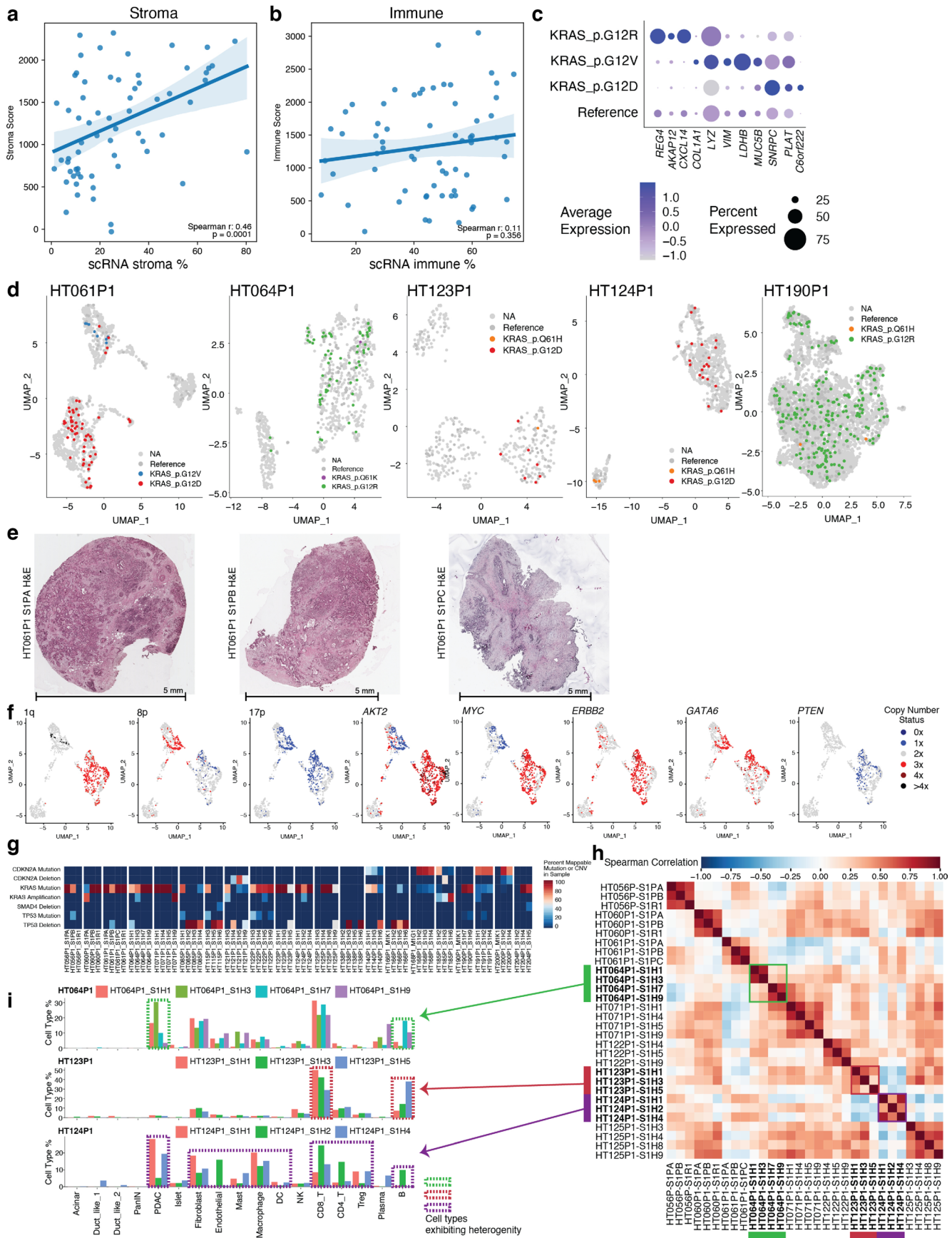
Peer review information *Nature Genetics* thanks Itai Yanai and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



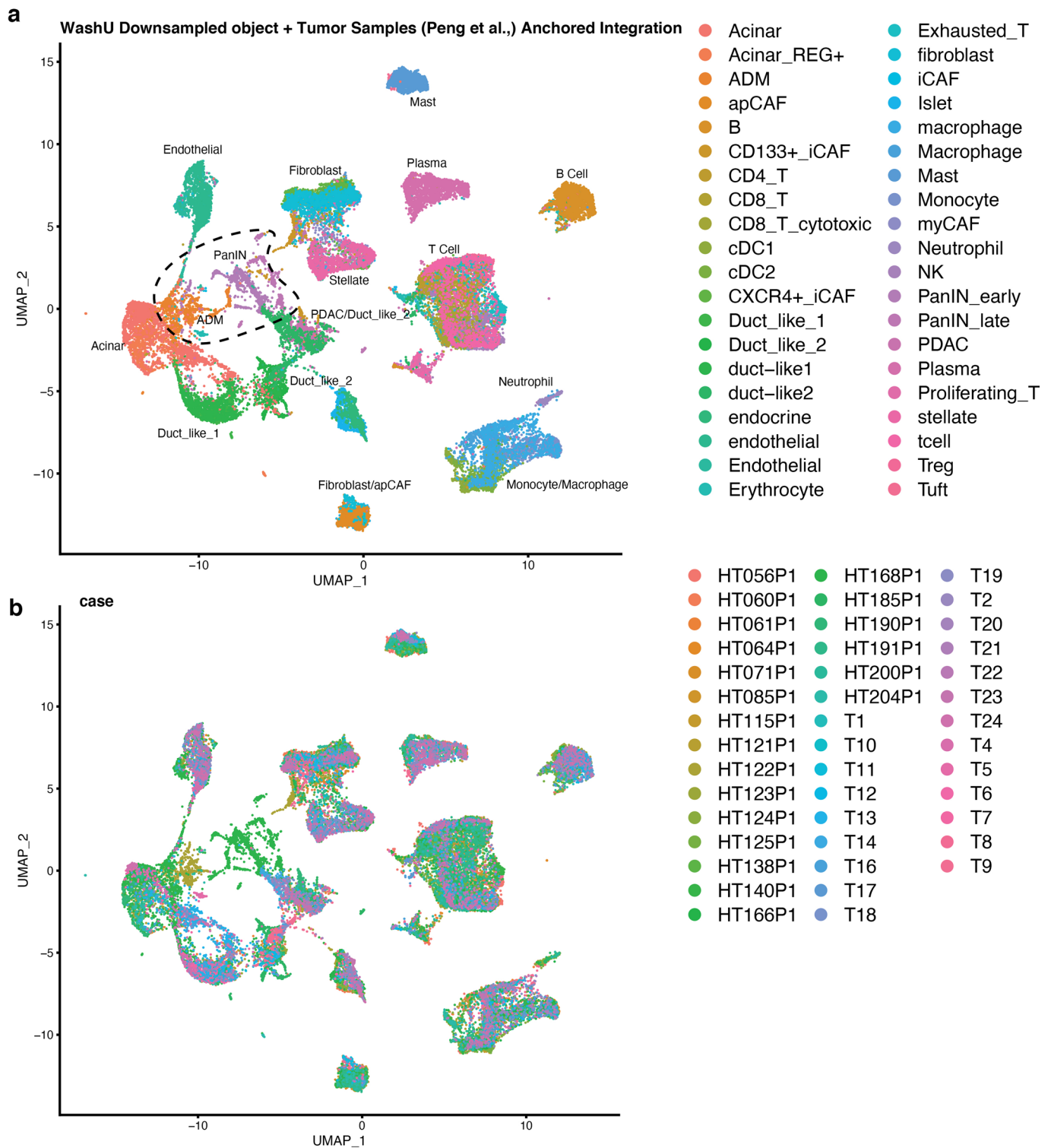
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Details of Data Cohort Overview. **a)** All 232,764 cells labeled by case ID. **b)** Same as A, but cells are labeled by cell type. **c)** scRNA cell type proportions across samples in the cohort. The bigger the circle, the higher the proportion. **d)** Spearman correlations of tumor estimates. The 95% confidence interval is shown. Top: histology vs scRNA, Middle: ABSOLUTE vs scRNA, Bottom: ABSOLUTE vs histology. **e)** Proteomics and phosphoproteomics PCAs labeled by case ID. **f)** Proteomics and phosphoproteomics PCAs labeled by TMT plex. **g)** Top: Genomic landscape of the cohort showing the top significantly mutated genes. The color scale denotes variant allele fraction (VAF) for each gene. The top bar plot indicates mutation burden for each sample. Bottom: Bulk omics overview of the cohort. The first row indicates germline mutation status followed by scRNA tumor fraction, immune subtype, relative scores of immune and stroma, and different PDAC subtypes (Moffitt, Collisson, Bailey) by piece.

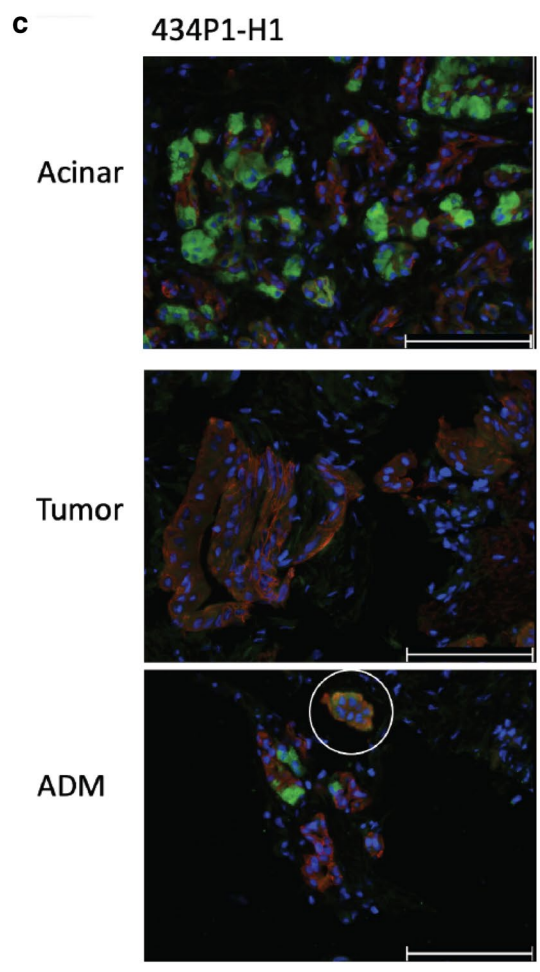
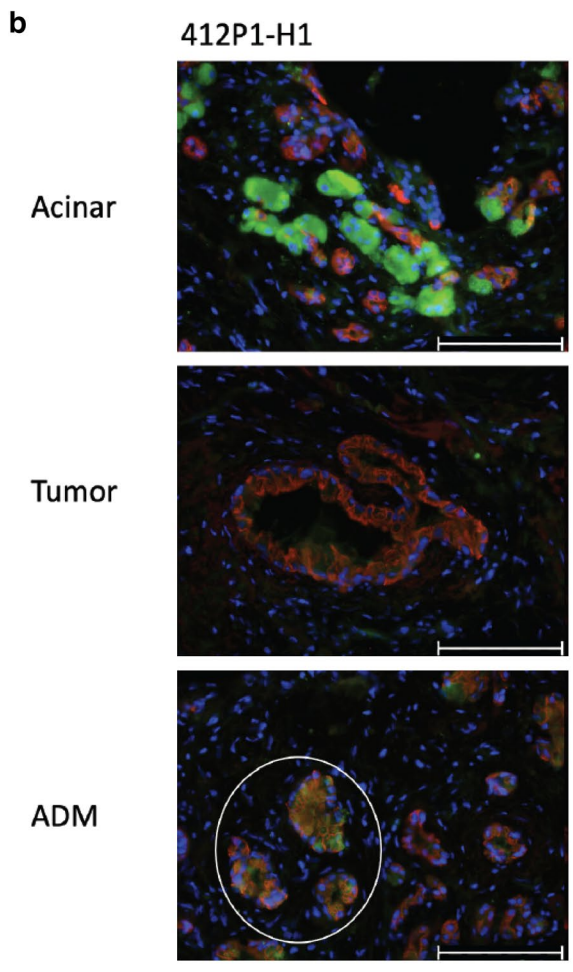
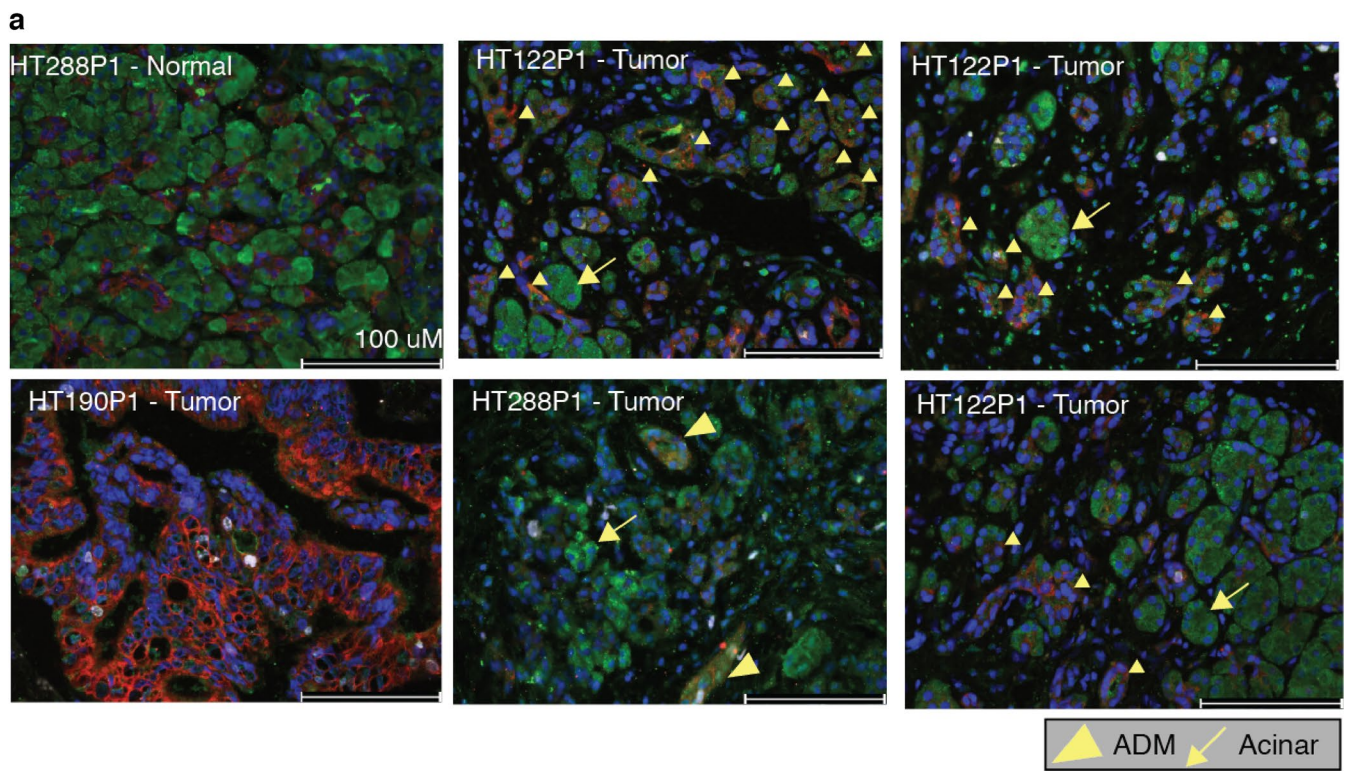


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Genomic Features in Heterogeneous KRAS Subpopulations. **a)** Spearman correlation of scRNA estimates and ESTIMATE stroma score. **b)** Spearman correlation of scRNA estimates and ESTIMATE immune score. For panels A and B, the 95% confidence interval is shown. **c)** Top significant DEGs between specific *KRAS* hotspot mutations. Only cells with a mappable *KRAS* mutation were included in this analysis. **d)** *KRAS* mutations in tumor cells of 5 cases with multiple *KRAS* variants mapped. **e)** H&E images of the spatial samples in HT061P1. **f)** Arm and gene-level CNV events in HT061P1 mapped to different tumor clusters. **g)** Percent of mappable mutations or deep copy number amplifications and/or deletions in each sample for *KRAS*, *TP53*, *CDKN2A*, and *SMAD4*. Samples are grouped together by case and cases are separated by white lines. **h)** Protein-level pairwise spearman correlation between all 30 samples that underwent bulk proteomics. Boxed cases represent cases with high heterogeneity: HT064P1 in green, HT123P1 in red, and HT124P1 in purple. **i)** Cell type proportion distribution of the three heterogeneous cases from panel H. Cell types in dotted boxes represent substantial differences in cellular composition that likely underlie the observed heterogeneity.

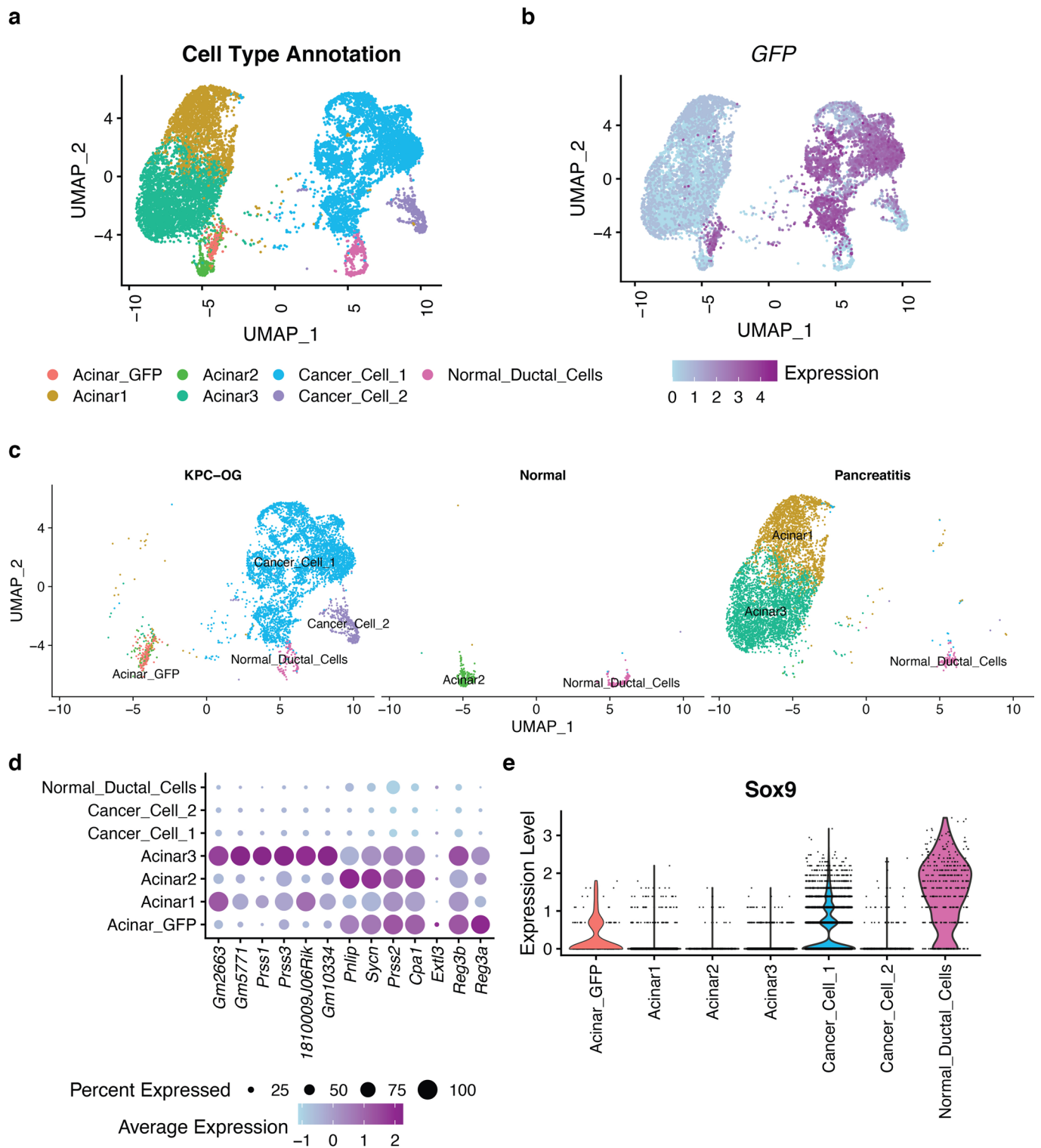


Extended Data Fig. 3 | Evaluating Transitional Populations in Published Studies. **a)** Integration of downsampled cells from tumor samples from Peng et al.²⁴, and WashU samples. UMAP shows integrated single cells colored by cell type. Circled region indicates cells that are specific to the WashU Cohort predominantly made of up PanIN and ADM identified cells. **b)** Integration of downsampled cells from tumor samples from Peng et al., and WashU samples. UMAP shows integrated single cells colored by case. Cases indicated with 'HT#P1' are WashU samples while samples with 'T#' are from Peng et al.

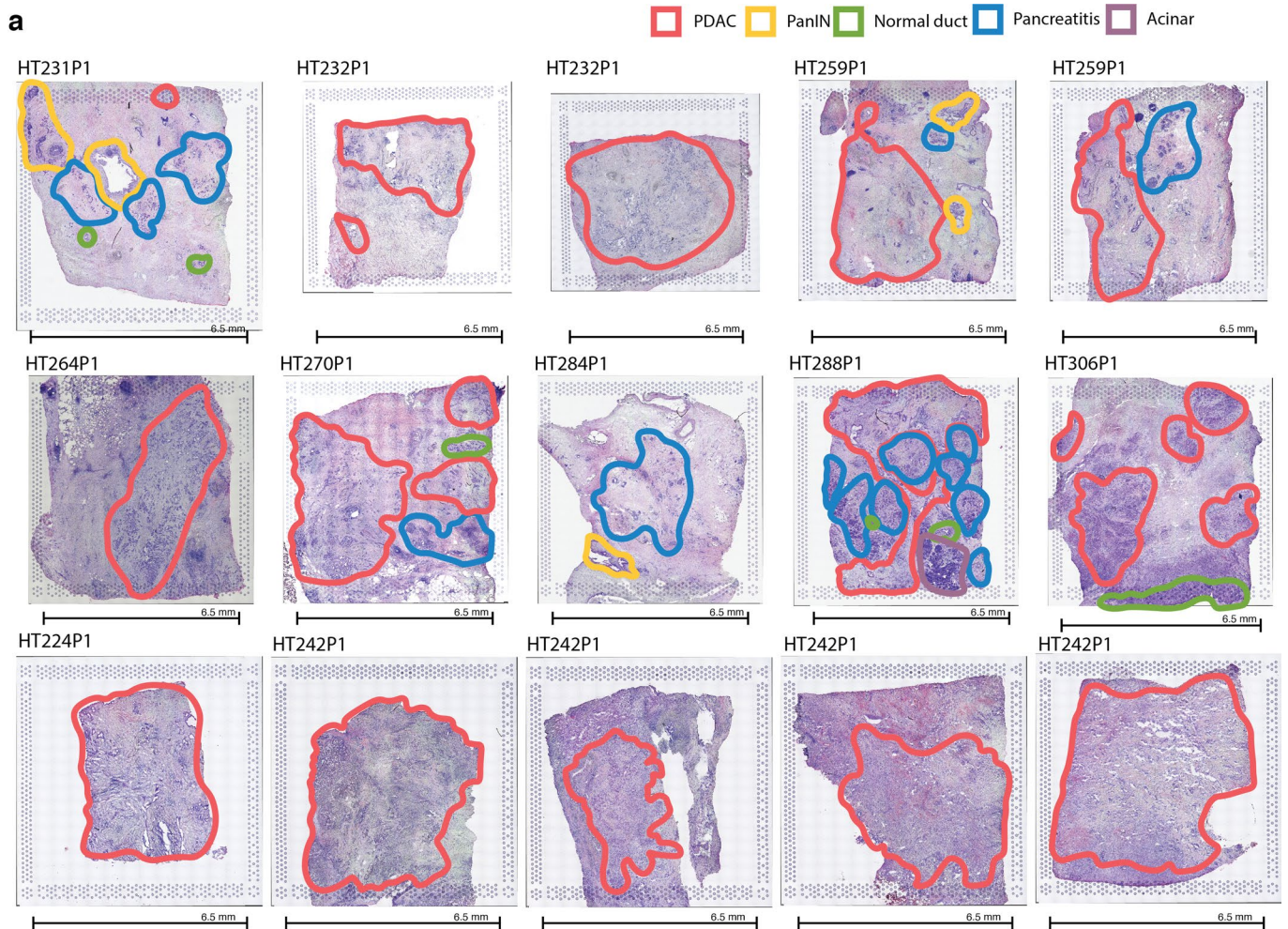


Extended Data Fig. 4 | See next page for caption.

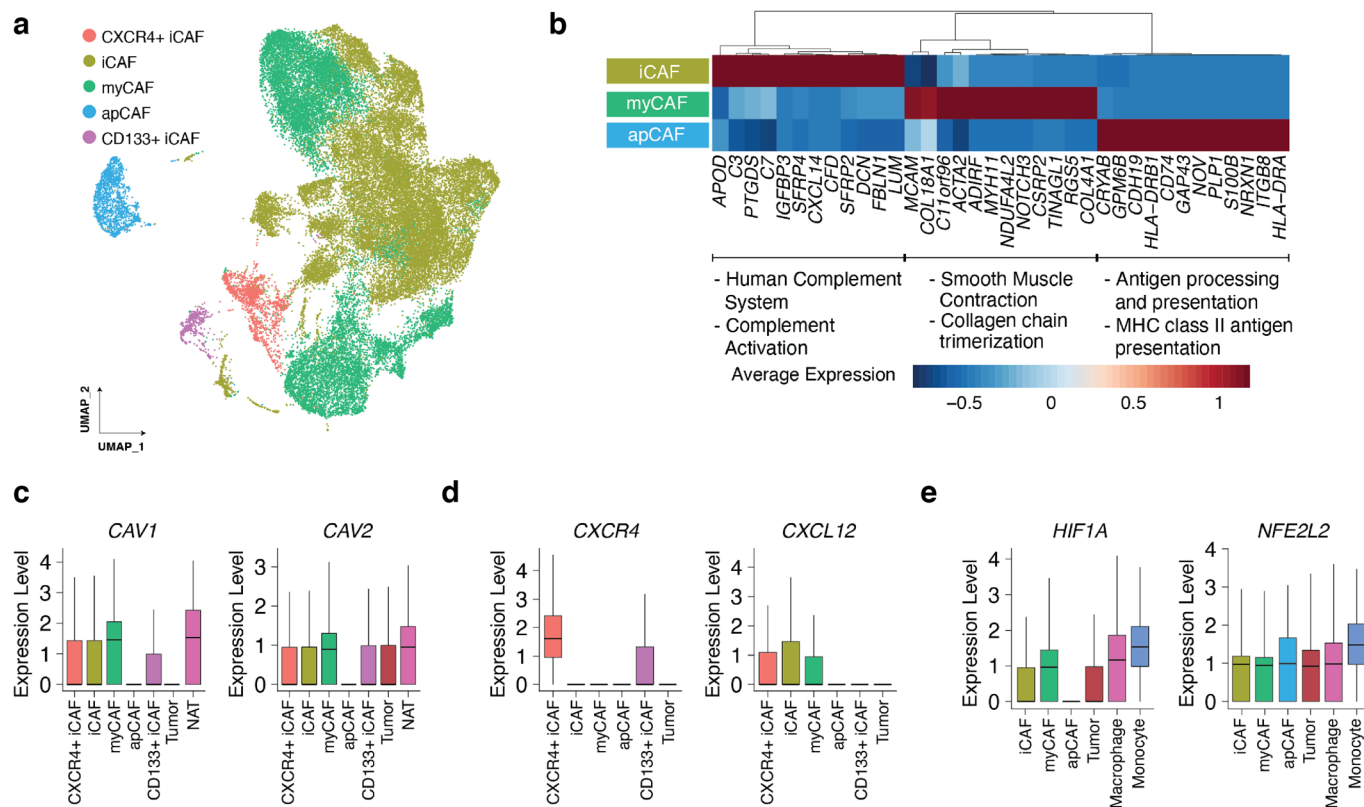
Extended Data Fig. 4 | Combined Channel Immunofluorescence Images of ADM Samples. **a)** Combined channel immunofluorescence staining across four samples HT288P1 (Adjacent Normal), HT190P1 (Tumor), HT122P1 (Tumor) and HT288P1 (Tumor). Amylase stains acinar cells (green), cytokeratin 19 stains ductal cells (red), Ki67 stains proliferating cells (white), and Hoechst stains nuclei (blue). For select sections, individual cells expressing both acinar and ductal markers indicating acinar to ductal metaplasia (ADM) are highlighted by the yellow triangle. Acinar cells are denoted with a yellow arrow. **b)** Combined channel immunofluorescence staining across two samples HT412P1 (Tumor) and **c)** HT434P1 (Tumor). Amylase stains acinar cells (green), cytokeratin 19 stains ductal cells (red), Ki67 stains proliferating cells (white), and Hoechst stains nuclei (blue). Cells exhibiting co-expression of Amylase and cytokeratin 19 are circled in white. Regions of the section with high acinar content, tumor content and ADM content are shown from top to bottom for each sample.



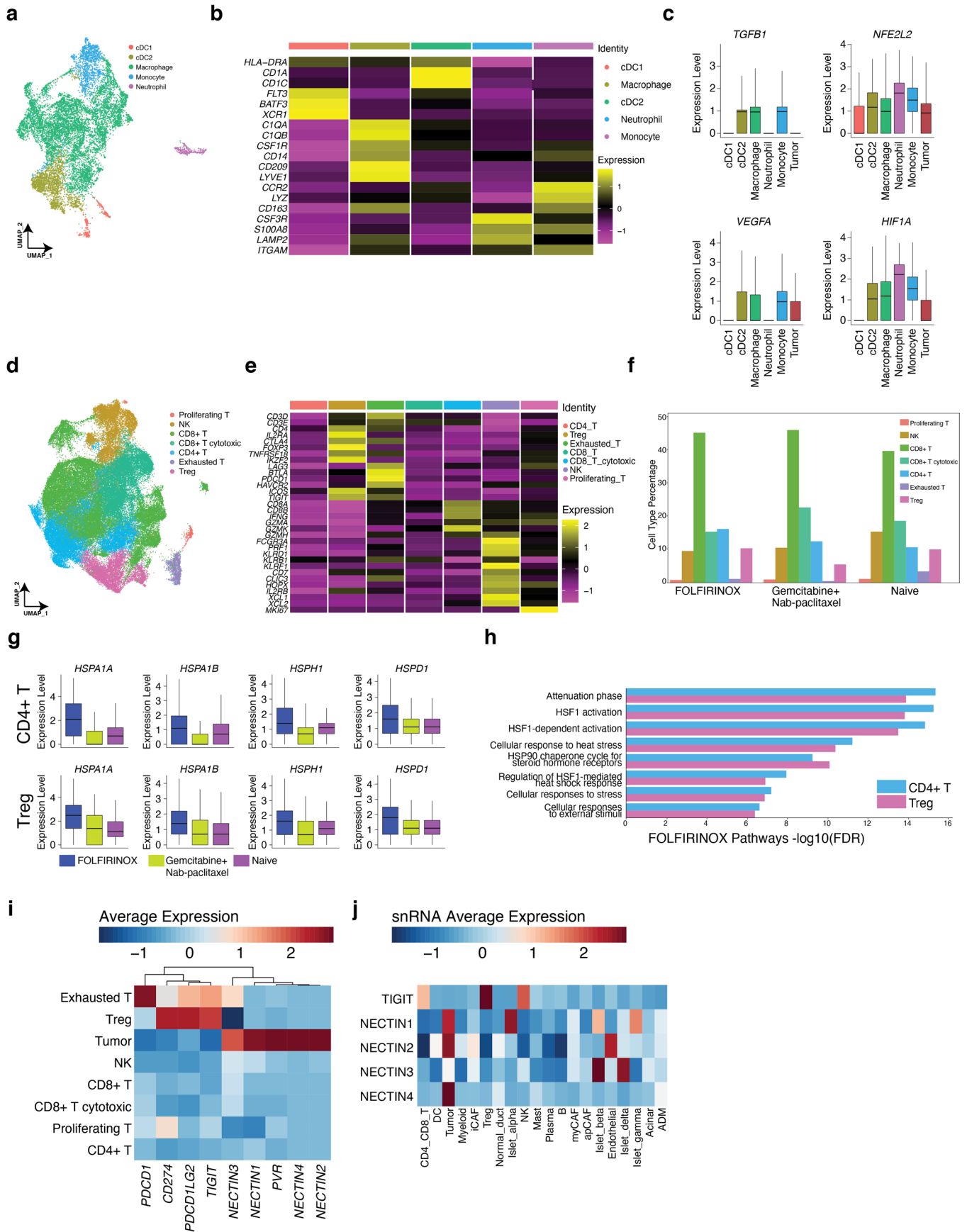
Extended Data Fig. 5 | Single Cell Analysis of Mouse Model Validating Transitional Acinar Populations. **a)** UMAP of acinar and ductal single cells from mouse model. Cells are colored by cell type. **b)** UMAP of GFP expression. Cells are colored by expression value. **c)** UMAP of acinar and ductal cells separated by mouse model from which cells are derived from. **d)** Selected gene expression markers of acinar and ductal genes across cell types. Each dot indicates expression of a given gene in an annotated cell cluster. The size indicates the percent of cells expressing that gene and the color is average expression. **e)** Violin plot showing the distribution of expression levels of Sox9 across each annotated cell type.



Extended Data Fig. 6 | Evaluating Transitional Populations with Spatial Transcriptomics and Published Studies. a) H&E images associated with each piece of tumor that underwent spatial transcriptomics processing. Regions on slides are highlighted based on pathology assisted review. Regions are indicated as tumor (Red), PanIN (Yellow), Normal Duct (Green), Pancreatitis (Blue) and Acinar (Purple).



Extended Data Fig. 7 | CAF Subtypes. **a)** UMAP of all fibroblast cells labeled by CAF subtype. **b)** Top DEGs and pathways across iCAFs, myCAFs, and apCAFs. **c)** *CAV1* and *CAV2* expression in CAF subtypes, tumor cells, and fibroblasts from NAT samples. **d)** *CXCR4* and *CXCL12* expression in CAF subtypes and tumor cells. **e)** *HIF1A* and *NFE2L2* expression in CAF subtypes, tumor cells, macrophages, and monocytes. FDR < 0.0001 for macrophage and monocyte upregulation of *NFE2L2* and *HIF1A*. Panels D-F include expression from all cells in the study of the given cell type and the boxplots show the median with 1.5x IQR whiskers.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Immune Cells in PDAC. **a)** UMAP of myeloid and dendritic cells (DC) labeled by cell type. **b)** Myeloid and DC cell type marker expression. **c)** Expression of the Keap1-Nrf2 (NFE2L2) pathways genes in all myeloid, DC, and tumor cells in the study. The boxplots show the median with 1.5x IQR whiskers. **d)** UMAP of lymphocyte and NK cells labeled by cell type. **e)** Lymphocyte cell type marker expression. **f)** Cell type percentages of lymphocytes across treatment groups. **g)** Expression of heat shock genes across treatment groups in Treg and CD4+ T cells (cells from n = 26 FOLFIRINOX samples, n = 15 Gemcitabine + Nab-paclitaxel samples, n = 25 untreated samples). The boxplots show the median with 1.5x IQR whiskers. **h)** Pathway enrichment of FOLFIRINOX vs treatment-naïve samples in Treg and CD4+ T cells using gene set overrepresentation analysis. **i)** Average expression of genes in lymphocytes and tumor cells in the scRNA data. **j)** Average expression of *TIGIT* and nectin genes across cell types in the snRNA data.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

March 2021

inferCNV v0.8.2, (Tickle et al., 2019), <https://github.com/broadinstitute/infercnv>
 Integrative Genomics Viewer, (Robinson et al., 2011), <https://igv.org>
 Loupe Browser v.5.0, 10X genomics, <https://www.10xgenomics.com/products/loupe-browser>
 magrittr_1.5, R Development Core Team, <https://cran.r-project.org/package=magrittr>
 Matrix_1.2-17, R Development Core Team, <https://CRAN.R-project.org/package=Matrix>
 MEDALT, (Wang et al., 2020), <https://github.com/KChen-lab/MEDALT>
 MuTect v1.1.7, (Cibulskis et al., 2013), <https://github.com/broadinstitute/mutect>
 pheatmap_1.0.12, R Development Core Team, <https://cran.r-project.org/package=pheatmap>
 Pindel v0.2.5, (Ye et al., 2009), <https://github.com/genome/pindel>
 Python v3.7, Python Software Foundation, <https://www.python.org/>
 R v3.6, R Development Core Team, <https://www.r-project.org/>
 RColorBrewer_1.1-2, R Development Core Team, <https://CRAN.R-project.org/package=RColorBrewer>
 RCTD v1.1.2.0, (Cable et al., 2021), <https://github.com/dmccable/RCTD>
 reshape2_1.4.3, R Development Core Team, <https://cran.r-project.org/package=reshape2>
 Samtools v1.2, (Li et al., 2009), <https://www.htslib.org/>
 scVarScan, Li Ding Lab, <https://github.com/ding-lab/10Xmapping>
 SeqQEst, Li Ding Lab, <https://github.com/ding-lab/SeqQEst>
 Seurat v3.1.2 and v4.0.3, (Butler et al., 2018), <https://cran.r-project.org/web/packages/Seurat>
 somaticwrapper v1.5, Li Ding Lab, <https://github.com/ding-lab/somaticwrapper>
 sva v3.40.0, (Huber et al., 2015), <https://bioconductor.org/packages/release/bioc/html/sva.html>
 STAR v2.7.4a, (Dobin et al., 2013), <https://github.com/alexdobin/STAR>
 Strelka v2.9.2, (Kim et al., 2018), <https://github.com/Illumina/strelka>
 stringr_1.4.0, R Development Core Team, <https://cran.r-project.org/package=stringr>
 Subread v2.0.1, (Liao et al., 2013), <https://sourceforge.net/projects/subread/>
 Tidyverse, (Wickham et al., 2019), <https://www.tidyverse.org/>
 VarScan v2.3.8, (Koboldt et al., 2012), <https://dkoboldt.github.io/varscan/>
 viridis_0.5.1, R Development Core Team, <https://github.com/sjmgarnier/viridis>
 viridisLite_0.3.0, R Development Core Team, <https://github.com/sjmgarnier/viridis>
 xCell v1.2, (Aran et al., 2017), <http://xCell.ucsf.edu/>
 Monocle3 v3.10, <https://cole-trapnell-lab.github.io/monocle3/docs/installation/>
 CopyKat 1.0.4 (Gao et al., 2021), <https://github.com/navinlabcode/copykat>
 inferCNV post processing, https://github.com/ding-lab/infer_cnv_postprocessing.git

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All raw image and sequencing data will be deposited into the publicly available HTAN DCC at <https://data.humantumoratlas.org/>. References (GRCh38 genome reference v3.0.0 and refdata-gex-mm10-2020-A) used for single-cell for the human and mouse genomes, respectively are available <https://support.10xgenomics.com/single-cell-gene-expression/software/release-notes/build>. Mouse single-cell RNA-seq data is available using the following BioProject PRJNA835747. Data for single-cell integration from Peng et al. was downloaded from the Genome Sequence Archive under project PRJCA001063.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Tumor samples were collected as patients underwent surgery during the collection time period on a rolling basis. The sample size in the manuscript is comparable or greater than most studies published to date. Tumor samples were collected as patients underwent surgery during the collection time period on a rolling basis, and thus no sample size was calculated. As the goal was to build an atlas, rather than to obtain sufficient statistical power, we included as many samples as we could up until manuscript submission. Nevertheless, our sample sizes (n > 80 patients; n > 230k cells) are comparable or much greater than most single cell studies published to date.

Data exclusions

We excluded two bulk RNA-Seq samples that did not pass quality control (QC) based on FASTQC and HLA-typing checks.

Replication

We analyzed previously published data on single cell pancreatic cancer studies to ensure that our cell typing is consistent. We included a cohort of 10 tumors that underwent single nucleus RNA sequencing and used these results to replicate/validate our single cell findings.

Additionally, we validated results using orthogonal technologies such as immunofluorescence staining and spatial transcriptomics. Validated results included cell type assignment, differentially expressed genes in cell types of interest, and staining for the presence of key markers in sample-matched tissue specimens.

Randomization Due to our patient accrual on a rolling basis, we did not 'allocate' patients into specific treatment groups, thus the imbalance in the number of patients in each treatment group (e.g. we have a high number of samples in the FOLFIRINOX-treated group). Moreover, the study was designed around discovery and characterization, rather than clinical comparison of treatment groups, e.g. in terms of survival. Consequently, there was no issue of controlling for covariates to reduce statistical noise for improving regression. Sequencing was done in randomized batches and we corrected proteomic batch effects (due to TMT plexes) using ComBat. Mice from our KPC-OG breeding colony were genotyped for all alleles. Three mice with the with full genotype (p48-Cre+;LSL-KRASG12D+;p53flox;OVA-GFP+) were allocated into the KPC-OG tumor group. p48-Cre(-) littermate controls were used for non-cancer control groups, and randomly allocated into either the normal pancreas (no treatment) or pancreatitis (cerulein treated) group.

Blinding Blinding was not relevant in our study as samples were collected from patients as they came into the clinic. As every sample was included in the study and processed in the same standardized way, knowledge of clinical phenotypes would not affect the study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a Involved in the study

Antibodies

Eukaryotic cell lines

Palaeontology and archaeology

Animals and other organisms

Human research participants

Clinical data

Dual use research of concern

Methods

n/a Involved in the study

ChIP-seq

Flow cytometry

MRI-based neuroimaging

Antibodies

Antibodies used

α -Amylase, Sigma, Catalog: A8273
 Cytokeratin 19, Santa Cruz, Catalog: sc-376126
 Ki67, Thermo, Catalog: 14-5698-82
 Alexa Fluor 488 AffiniPure F(ab')₂ Fragment D α Rabbit, Jackson Immuno Research, Catalog: 711-546-152
 Alexa Fluor 555 AffiniPure F(ab')₂ Fragment D α Mouse, Thermo, Catalog: A31570
 Alexa Fluor 647 AffiniPure F(ab')₂ Fragment D α Rat, Jackson Immuno Research, Catalog: 712-606-153
 Hoechst, Life Technologies, Catalog: H3570

Validation

α -Amylase, manufactured by Sigma (A8273) - validation data in the following PMID: 34358441
 Cytokeratin 19, manufactured by Santa Cruz - validation data in the following PMID: 33951479
 Ki67, manufactured by Thermo - validation data in the following PMID: 32750316

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals The breed of mice are C57BL/6, all male and pancreatic tissue was extracted at 4 months.

Wild animals N/A

Field-collected samples N/A

Ethics oversight Washington University School of Medicine IACUC under protocol #19-0856

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics All patients involved in this study were between the ages of 38-82 with pancreatic ductal adenocarcinoma. Patients were either treatment naive or had neoadjuvant therapy prior to sample collection. Detailed clinical information can be found in Supplementary Table 1.

Recruitment

Participants for this study were recruited from all patients undergoing surgical resection of primary pancreatic ductal adenocarcinoma, and were screened only on the basis of resectable disease identified by the schedule contained within the electronic health record system (EPIC). Clinical research staff conducted a manual review to ensure each participant was assigned a number and consented either through the Hepato-Biliary clinic at the Washington University in St. Louis, Department of Surgery and Barnes Jewish Hospital or on the day of surgery. All patients provided informed consent approved by the Washington University in St. Louis Institutional Review Board.

Ethics oversight

Primary pancreatic adenocarcinomas either treatment naïve or neoadjuvant treated (Supplementary Table S1) were collected with written consent from all patients under the protocol 11-08117 with approval from the Washington University in St. Louis Institutional Review Board. Consent included the use of all de-identified patient data for publication. Participants were not compensated.

Note that full information on the approval of the study protocol must also be provided in the manuscript.