Alternative polyadenylation characterizes epithelial and fibroblast phenotypic heterogeneity in pancreatic ductal adenocarcinoma

Swati Venkat and Michael E. Feigin*

Department of Pharmacology and Therapeutics, Roswell Park Comprehensive Cancer Center, Buffalo, NY

* Corresponding author: MEF

Email: michael.feigin@roswellpark.org

Twitter: @TheFeiginLab

ORCID:

SV: 0000-0001-7551-3888

MEF: 0000-0002-8189-5568

Author Contributions:

Performed data analysis: SV

Wrote the manuscript: SV, MEF

Supervised the study: MEF

Competing Interest Statement: None

Abstract

Human tumors are characterized by extensive intratumoral transcriptional variability within the cancer cell and stromal compartments. This variation drives phenotypic heterogeneity, producing cell states with differential pro- and anti-tumorigenic properties. While bulk RNA sequencing cannot achieve cell type specific transcriptional granularity, single cell sequencing has permitted an unprecedented view of these cell states. Despite this knowledge, we lack an understanding of the mechanistic drivers of this transcriptional and phenotypic heterogeneity. 3' untranslated region alternative polyadenylation (3' UTR-APA) drives gene expression alterations through regulation of 3' UTR length. These 3' UTR alterations modulate mRNA stability, protein expression and protein localization, resulting in cellular phenotypes including differentiation, cell proliferation, and migration. Therefore, we sought to determine whether 3' UTR-APA events could characterize phenotypic heterogeneity of tumor cell states. Here we analyze the largest single cell human pancreatic ductal adenocarcinoma (PDAC) dataset and resolve 3' UTR-APA patterns across PDAC cell states. We find that increased proximal 3' UTR-APA is associated with PDAC progression and characterizes a metastatic ductal epithelial subpopulation and an inflammatory fibroblast population. Furthermore, we find significant 3' UTR shortening events in cell state-specific marker genes associated with increased expression. Therefore, we propose that 3' UTR-APA drives phenotypic heterogeneity in cancer.

Background

Pancreatic ductal adenocarcinoma (PDAC) is a lethal disease with a 5-year survival rate of 10% [1]. PDAC tumors are characterized by a dense stroma and a high degree of cell type specific phenotypic variation that is integral to disease progression and drug resistance [2–4]. Over the past decade, bulk and single cell RNA sequencing (scRNA-seq) analyses uncovered substantial inter- and intratumoral transcriptional heterogeneity [5–7]. These studies have formed the basis for patient stratification and delineation of phenotypically distinct epithelial and stromal subpopulations. For example, tumor epithelial cells have been found to exist in subpopulations that exhibit differing proliferative and metastatic potential [6,8,9]. Similarly, phenotypically distinct subsets of cancer associated fibroblasts (CAFs) characterized by unique transcriptional profiles have been identified within the tumor microenvironment [10,11]. Two major CAF subclasses, inflammatory CAFs (iCAFs) and myofibroblastic CAFs (myCAFs), have distinct but crucial roles in tumor progression and therapeutic resistance [12,13]. However, mechanistic drivers of such transcriptional and phenotypic heterogeneity in PDAC remain unclear. Recently, we performed an in-depth analysis of sequencing data on PDAC tumors that established 3' UTR alternative polyadenylation (APA) as a mechanistic driver of oncogene expression [14–16]. Specific PDAC oncogenes were found to undergo proximal 3' UTR-APA (usage of proximal 3' UTR polyadenylation site) resulting in shorter 3' UTRs, driving increased expression. However, as this analysis made use of bulk RNA-seq data, it was impossible to determine the contribution of APA to cell type specific transcriptional heterogeneity. To determine if APA could be a mechanistic driver of phenotypic variation in cancer we now leverage the largest scRNA-seq human PDAC dataset recently published by Peng *et al*. [17]. Unlike bulk sequencing data, the majority of single cell sequencing protocols are 3' biased, allowing robust detection of 3'-UTR-APA changes and the associated transcriptional heterogeneity in a high-resolution dataset

[7,18–20]. To our knowledge, this is the first investigation of APA events associated with intratumoral heterogeneity.

Results and Discussion

To understand whether APA is associated with cell type specific phenotypic variation, we sought to identify cell types that exhibit substantial 3' UTR-APA events. To achieve this, we reanalyzed the scRNA-seq PDAC dataset (Additional file Fig. S1a) comprised of 11 normal pancreata and 24 tumor samples. We focused on cell types that form a significant proportion of the tumor, including acinar and ductal epithelial cells, and stromal fibroblasts and stellate cells. After quality control (see Methods), we processed a total of 22053 tumor cells across 21 tumor samples and 10345 normal cells across 11 pancreata for downstream analyses. We adapted a recently published algorithm to detect 3' UTR-APA events from scRNA-seq data [18] (Additional file Fig. S1b). In concordance with previous findings, tumor tissues exhibited significantly higher proximal 3' UTR-APA gene events (3' UTR shortening) as compared to normal tissues [14,16]. In particular, tumor ductal cells showed significantly higher numbers of proximal 3' UTR-APA events (1177 genes expressed shorter 3' UTRs and 250 genes expressed longer 3' UTRs) compared to normal ductal cells (Fig. 1a). While fibroblasts, acinar cells and stellate cells in PDAC tumors exhibited a higher number of proximal 3' UTR-APA events compared to their normal counterparts, PDAC ductal cells exhibited the highest ratio of proximal to distal 3' UTR-APA events (~5:1) compared to other cell types. While a bulk PDAC RNA-seq study would reveal significant 3' UTR-APA events occurring across a mixture of these cell types, it would fail to resolve cell type specific 3' UTR-APA events. The extent of proximal 3' UTR-APA in PDAC ductal cells motivated us to probe APA events within this transcriptionally diverse cell population. Peng and colleagues identified two subsets of PDAC ductal cells, namely ductal cell

type 1 and ductal cell type 2. Ductal cell type 2 constituted the majority of the PDAC ductal cells and exhibited a malignant gene expression profile. Ductal cell type 1 expressed an abnormal gene expression profile that was distinct from the normal cells, representing a transcriptional state between normal and tumor ductal cells [17]. We performed dimensionality reduction and clustering to delineate these transcriptionally distinct subsets of normal and tumor ductal cells (Fig. 1b). Clustering revealed 6 transcriptionally distinct subclusters: normal ductal cells (dA), tumor ductal cell type 1 (dB) and tumor ductal cell type 2 (composed of subclusters dC, dD, dE, dF). Interpatient as well as intrapatient heterogeneity was detected in ductal cell type 2 with the majority of the patients represented in subcluster dC and a minority in subclusters dD, dE and dF (Additional file, Fig. S2a). Subcluster dE specific genes were enriched for metastatic markers (*HMGA1, ENO1, GABRP, IGFBP2, SDC1, LGALS1*) (Additional file, Fig. S2b) and pathway enrichment analysis of dE overexpressed genes showed epithelial to mesenchymal transition (EMT) as a top hit supporting its metastatic phenotype (Additional file, Fig. S2c) [21–26]. In contrast, gene expression and pathway analysis of subcluster dD specific genes showed enrichment for well-differentiated PDAC markers (*REG4, TFF1, TFF2, TFF3, VSIG2, LGALS4*), highlighting the extensive phenotypic heterogeneity exhibited by PDAC ductal cells (Additional file, Fig. S2d) [6].

We first sought to characterize APA patterns across the broad ductal cell states (normal ductal cells, tumor ductal cell type 1, tumor ductal cell type 2) to determine the relationship between APA and tumor progression. We determined the mean proximal polyA site usage index (mean proximal PUI), the extent of 3' UTR proximal site usage for each cell, averaged over all genes (see Methods, [18]). A higher mean proximal PUI indicates enhanced cleavage at proximal polyadenylation sites in the cell (resulting in shorter 3' UTRs). We plotted the mean proximal PUI for every ductal cell associated with each cell state (Fig. 1c). Pseudotime analysis

116 confirmed progression from a normal state (normal ductal cells) to an abnormal intermediate
117 state (tumor ductal type 1) to a malignant ductal state (tumor ductal type 2) (Additional file, Fig.
118 S2e). This malignant progression was associated with a progressive and significant increase in
119 mean proximal PUI (Fig 1d). Therefore increased proximal 3'-UTR-APA is associated with
120 malignant progression in PDAC.
121
122 We noted substantial APA heterogeneity within the subclusters comprising tumor ductal cell
123 type 2 (dC, dD, dE) and therefore quantified proximal 3' UTR-APA patterns across these cells.
124 The cells in the metastatic subcluster dE showed a significant increase in mean proximal PUI
125 compared to dC, indicating increased 3' UTR shortening events in dE (Fig. 1e). In contrast, the
126 cells in the well-differentiated PDAC subcluster dD showed a significant decrease in mean
127 proximal PUI compared to dC, indicating decreased 3' UTR shortening events (Additional file,
128 Fig. S2f). To determine if these APA events are associated with known metastatic driver genes,
129 we performed pathway enrichment analysis of the 3' UTR altered genes in dE, which revealed
130 EMT as a top hit (Fig. 1f). Furthermore, we found significantly increased proximal APA of
131 metastasis-promoting genes preferentially expressed in dE, including *GABRP* and *SDC1* (Fig.
132 1g, 1h, Additional file, Fig. S2b). This suggests a novel role of proximal 3' UTR-APA in
133 orchestrating the metastatic PDAC phenotype.
134
135 CAFs are a transcriptionally and phenotypically heterogeneous population in the tumor
136 microenvironment that make fundamental contributions to both progression and therapy
137 response [11,12,27–29]. How this transcriptional heterogeneity is developed and maintained
138 during tumorigenesis is integral to the advancement of more effective therapeutic strategies. In
139 PDAC, two major CAF subtypes have been discovered and functionally characterized –
140 myCAFs, responsible for secreting the extracellular matrix components that promote a dense

desmoplastic stroma, and iCAFs, responsible for secreting IL-6 and other inflammatory mediators. To investigate the role of APA in CAF biology, we clustered normal fibroblasts and CAFs and identified transcriptionally differing subclusters within the CAF population (Fig. 2a). Clustering revealed transcriptionally distinct subclusters including normal fibroblast cells (fA) and tumor fibroblast cells (composed of subclusters fB-fE). Pathway analysis and cluster specific gene markers revealed fC as a myCAF population (*ACTA2, POSTN, MMP11, IGFBP3, COL12A1, THBS2*), (Additional file, Fig. S3a, S3c) and fD as an iCAF population (*HAS1, HAS2, CCL2, UGDH, SOD2, LMNA*), (Additional file, Fig. S3b, S3d) [12,13]. To characterize 3' UTR-APA patterns, we determined the mean proximal PUI for every normal and tumor fibroblast cell (Fig. 2b). In contrast to normal fibroblasts, the tumor fibroblast population showed a small but significant increase in proximal 3' UTR-APA (Additional file, Fig. S3e), indicative of more 3' UTR shortening events in CAFs. We next examined 3' UTR-APA underlying CAF heterogeneity and found no significant difference between normal fibroblasts and the myCAF population (Fig. 2c). In contrast, there was a significant increase in 3' UTR shortening in the iCAF population (Fig. 2d, Additional file, Fig. S3f), revealing that increased proximal APA characterizes the inflammatory CAF phenotype. Importantly, we found significant increased proximal APA of critical iCAF markers such as *SOD2* and *UGDH* associated with their increased expression in iCAFs (Fig. 2e, 2f, Additional file, Fig. S3b). This suggests a novel role of 3' UTR-APA in orchestrating the inflammatory CAF phenotype.

Conclusions

3' UTR-APA is an underappreciated driver of gene dysregulation in cancer. Single cell sequencing has revealed that tumors have high degrees of transcriptional and phenotypic heterogeneity, both within the cancer cell and stromal compartments. However, drivers of such complex phenotypic heterogeneity remain unclear. In this study, we investigated 3'UTR-APA

associated phenotypic heterogeneity using single cell data. To our knowledge, this is the first investigation of APA events associated with intratumoral heterogeneity. We demonstrate that 3' UTR shortening increases progressively during PDAC progression. Furthermore, 3' UTR shortening of critical metastatic and iCAF marker genes is associated with increased expression, thereby defining cell identity. Increased proximal 3' UTR-APA characterizes a metastatic ductal subpopulation in tumor epithelial cells as well as an inflammatory CAF population in the PDAC stroma. We propose that 3' UTR-APA drives phenotypic heterogeneity both in the tumor epithelium and within the tumor microenvironment.

Methods

Bioinformatic processing of human scRNA-seq data. scRNA-seq FASTQ files of 24 PDAC patients and 11 normal pancreata were downloaded from Genome Sequence Archive (GSA) (Accession: CRA001160, Bioproject: PRJCA001063). Cell Ranger 3.1.0 using standard parameters was used to align each file to the hg19 genome [30]. Appropriate chemistry and alignment by Cell Ranger was detected for 21 patients and 11 normal tissues and these data were used for downstream analyses. We focused on annotated cells (Peng *et al*. [17]) with at most 6000 genes/cell (to eliminate doublets) and with at least 200 genes/cells. Cells with >10% mitochondrial counts and genes occurring in <3 cells were excluded from the analysis. This yielded 10345 normal cells and 22053 tumor cells for the analysis of 3' UTR-APA events.

Analysis of 3'UTR-APA events. Analysis of 3' UTR-APA events was performed by manual implementation of the scRNA-seq algorithm proposed in [18] (Fig. S1b). Briefly, PCR duplicates were discarded from aligned BAM files using UMI tools [31]. These files were used to detect peaks in 3' UTR read density using Homer findPeaks function [32,33]. Additionally, cell type identity was obtained from Peng *et al*. [17] and we used this information to annotate major cell types and generate cell type specific BAM files. Reads in cell type specific BAMs that mapped to Homer-determined peak positions were measured using Feature counts (Rsubread package) [34]. Low count peaks (<10 CPMs over all cell clusters) and peaks with A-rich sequences [18] were filtered out, allowing identification of statistically significant 3'UTR-APA events and mean proximal PUI at a single cell level exactly as described in [18]. IGV plots were used to visualize the read density changes for the 3' UTR altered genes. Frequency density plots were used to visualize distribution of mean proximal PUI across single cells in a subcluster and significant differences between subclusters were assessed using the Wilcoxon ranked sum test with continuity correction.

Bioinformatics analyses and statistical methods. Subsequent analyses were carried out in R 4.0.4. Monocle3 was used to analyze single cell trajectories to determine cell state transitions [35]. Top 200 differentially expressed genes between normal and tumor type 2 ductal cells were used for dimensionality reduction via UMAP and clustering and the mean proximal PUI for each cell was overlayed. The top 25 cluster-specific marker genes were identified using the top_markers function in Monocle3. Differentially expressed genes between the subclusters were identified using the FindMarkers function in Seurat4 [36]. Gene Set Enrichment Analysis (GSEA) and Enrichr were used to perform pathway analysis using the MSigDB hallmark, KEGG and Reactome gene sets [37,38]. Enrichment of the input genes (3'-UTR-APA altered genes/differentially expressed genes) in Enrichr was computed using the Fisher's exact test and p-values were adjusted using the Benjamini-Hochberg correction (FDR < 0.01). A similar approach was implemented for analysis of fibroblasts.

Availability of data and materials. The R code written for this analysis is available on GitHub (https://github.com/feiginlab/APA_PDA).

Acknowledgments

This work was supported by NCI grant P30 CA016056, and an award from the Roswell Park Alliance Foundation to MEF. We thank the members of the Feigin laboratory and Dr. Ethan Abel for their helpful comments on the manuscript.

References

1 Siegel, R.L. *et al.* (2021) Cancer Statistics, 2021. *CA. Cancer J. Clin.* 71, 7–33

2 Venkat, S. *et al.* (2021) Drivers of Gene Expression Dysregulation in Pancreatic Cancer. *Trends in Cancer* DOI: 10.1016/j.trecan.2021.01.008

3 Crawford, H.C. *et al.* (2019) Signaling Networks That Control Cellular Plasticity in Pancreatic Tumorigenesis, Progression, and Metastasis. *Gastroenterology* 156, 2073–2084

4 Oldfield, L.E. *et al.* (2017) Molecular Events in the Natural History of Pancreatic Cancer. *Trends in Cancer* 3, 336–346

5 Hayashi, A. *et al.* (2020) A unifying paradigm for transcriptional heterogeneity and squamous features in pancreatic ductal adenocarcinoma. *Nat. Cancer* 1, 59–74

6 Moffitt, R.A. *et al.* (2015) Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* 47, 1168–1178

7 Bernard, V. *et al.* (2018) 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

8 Bailey, P. *et al.* (2016) Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* 531, 47–52

9 Collisson, E.A. *et al.* (2019) Molecular subtypes of pancreatic cancer. *Nat. Rev. Gastroenterol. Hepatol.* 16, 207–220

10 Helms, E. *et al.* (2020) Fibroblast Heterogeneity in the Pancreatic Tumor Microenvironment. *Cancer Discov.* 10, 648–656

11 Pereira, B.A. *et al.* (2019) CAF Subpopulations: A New Reservoir of Stromal Targets in Pancreatic Cancer. *Trends in Cancer* 5, 724–741

12 Öhlund, D. *et al.* (2017) Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J. Exp. Med.* 214, 579–596

13 Elyada, E. *et al.* (2019) Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discov.* 9, 1102–1123

14 Mayr, C. and Bartel, D.P. (2009) Widespread Shortening of 3′UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell* 138, 673–684

15 Elkon, R. *et al.* (2013) Alternative cleavage and polyadenylation: Extent, regulation and function. *Nat Rev Genet.* 14, 496–506

16 Venkat, S. *et al.* (2020) Alternative polyadenylation drives oncogenic gene expression in pancreatic ductal adenocarcinoma. *Genome Res.* 30, 347–360

17 Peng, J. *et al.* (2019) Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* 29, 725–738

18 Shulman, E.D. and Elkon, R. (2019) Cell-type-specific analysis of alternative polyadenylation using single-cell transcriptomics data. *Nucleic Acids Res.* 47, 10027–10039

19 Dominguez, C.X. *et al.* (2020) Single-cell RNA sequencing reveals stromal evolution into LRRC15+ myofibroblasts as a determinant of patient response to cancer immunotherapy. *Cancer Discov.* 10, 232–253

20 Ziegenhain, C. *et al.* (2017) Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* 65, 631-643.e4

21 Jiang, S.H. *et al.* (2019) GABRP regulates chemokine signalling, macrophage recruitment and tumour progression in pancreatic cancer through tuning KCNN4-mediated Ca2+ signalling in a GABA-independent manner. *Gut* 68, 1994–2006

22 Chen, X. *et al.* (2020) The HPA/SDC1 axis promotes invasion and metastasis of pancreatic cancer cells by activating EMT via FGF2 upregulation. *Oncol. Lett.* 19, 211–220

23 Liau, S.S. *et al.* (2006) HMGA1 is a determinant of cellular invasiveness and in vivo metastatic potential in pancreatic adenocarcinoma. *Cancer Res.* 66, 11613–11622

24 Liu, H. *et al.* (2017) Silencing IGFBP-2 decreases pancreatic cancer metastasis and enhances chemotherapeutic sensitivity. *Oncotarget* 8, 61674–61686

25 Principe, M. *et al.* (2015) Targeting of surface alpha-enolase inhibits the invasiveness of pancreatic cancer cells. *Oncotarget* 6, 11098–11113

26 Orozco, C.A. *et al.* (2018) Targeting galectin-1 inhibits pancreatic cancer progression by modulating tumor–stroma crosstalk. *Proc. Natl. Acad. Sci. U. S. A.* 115, E3769–E3778

27 Biffi, G. and Tuveson, D.A. (2020) Diversity and biology of cancer-associated fibroblasts. *Physiol. Rev.* DOI: 10.1152/physrev.00048.2019

28 Neuzillet, C. *et al.* (2019) Inter- and intra-tumoural heterogeneity in cancer-associated fibroblasts of human pancreatic ductal adenocarcinoma. *J. Pathol.* 248, 51–65

29 Biffi, G. and Tuveson, D.A. (2018) Deciphering cancer fibroblasts. *J. Exp. Med.* 215, 2967

30 Zheng, G.X.Y. *et al.* (2017) Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049

31 Smith, T. *et al.* (2017) UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* 27, 491–499

32 Heinz, S. *et al.* (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* 38, 576–589

33 Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842

34 Liao, Y. *et al.* (2019) The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 47(8):e47

35 Trapnell, C. *et al.* (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386

36 Hao, Y. *et al.* (2020) Integrated analysis of multimodal single-cell data. *Cell* 184(13), 3573-3587

37 Edward, C.Y. *et al.* (2015) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14, 128

38 Subramanian, A. *et al.* (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15545–15550

Figure legends

Figure 1. Proximal APA in tumor epithelium is associated with PDAC progression and malignant phenotypes.

1a. A plot of number of shortened (red) and lengthened (blue) 3' UTR-APA events across four PDAC cell types compared to their counterparts in normal pancreas.

1b. UMAP embedding of ductal cells (dots) from normal pancreata and tumor patients. Color indicates the ductal cell type membership. Notations dA-dF denote the subclusters.

1c. UMAP embedding of ductal cells from normal pancreata and tumor patients. Color indicates degree of mean proximal PUI in each cell (blue, low; green, high).

1d. Distribution of mean proximal PUI of single cells in normal ductal cells (orange), tumor ductal cell type 1 (green) and ductal cell type 2 (blue) (every pairwise comparison yielded $p<10^{-7}$).

1e. Distribution of mean proximal PUI of single cells in subcluster dE (green) compared to subcluster dC (brown) ($p<10^{-16}$).

1f. Significantly enriched pathways (FDR < 0.01) associated with 3' UTR altered genes between subclusters dE and dC.

1g. IGV plot highlighting the 3' UTR density profile differences of the metastatic gene *GABRP* between subclusters dC (brown) and dE (green).

1h. IGV plot highlighting the 3' UTR density profile differences of the metastatic gene *SDC1* between subclusters dC (brown) and dE (green).

Figure 2. Increased proximal APA characterizes the inflammatory CAF phenotype.

1a. UMAP embedding of fibroblast cells (dots) from normal pancreata and tumor patients. Color indicates the fibroblast cell type membership. Notations fA-fE denote the subclusters.

1b. UMAP embedding of fibroblast cells from normal pancreata and tumor patients. Color indicates degree of mean proximal PUI in each cell (blue, low; green, high).

1c. Distribution of mean proximal PUI of single cells ($p=0.6$) in normal fibroblast cells (orange) and myCAFs (purple).

1d. Distribution of mean proximal PUI of single cells ($p<10^{-16}$) in iCAFs (green) compared to normal fibroblast cells (orange).

1e. IGV plot highlighting the 3'-UTR density profile differences of the iCAF activated transcription factor *SOD2* between iCAFs (green) and myCAFs (purple).

1f. IGV plot highlighting the 3'-UTR density profile differences of the iCAF marker *UGDH* between iCAFs (green) and myCAFs (purple).

Figure S1. Description of the scRNA-seq dataset and the workflow used to quantify 3' UTR-APA in PDAC.

S1a. A pie graph representing the single cell dataset that was used for downstream analyses of 3' UTR-APA patterns. Proportion of ductal and acinar cells in the epithelium, and fibroblast and stellate cells in the stroma are highlighted.

S1b. The workflow implemented to detect and quantify 3' UTR-APA events from single cell sequencing data (adapted from [18]).

Figure S2. Proximal APA in tumor epithelium is associated with PDAC progression and malignant phenotypes.

S2a. Barplot showing contribution of different ductal subclusters to each PDAC patient.

S2b. Violin plots of select metastatic markers across ductal cell type 2 subclusters ($p<0.001$).

S2c. Significant enriched pathways (FDR < 0.01) associated with genes overexpressed in dE compared to dC.

S2d. Violin plots of select well-differentiated PDAC markers across ductal cell type 2 subclusters ($p<0.001$).

S2e. Pseudo-time analysis depicting progression of ductal cell states (purple, early; yellow, late) based on their gene expression profiles.

S2f. Distribution of mean proximal PUI of single cells in subcluster dD (purple) compared to subcluster dC (brown) ($p<10^{-16}$).

Figure S3. Increased proximal APA characterizes the inflammatory CAF phenotype.

S3a. Violin plots of myCAF markers across normal fibroblasts (fA, orange) and specific tumor fibroblast subclusters (fC, purple; fD, green).

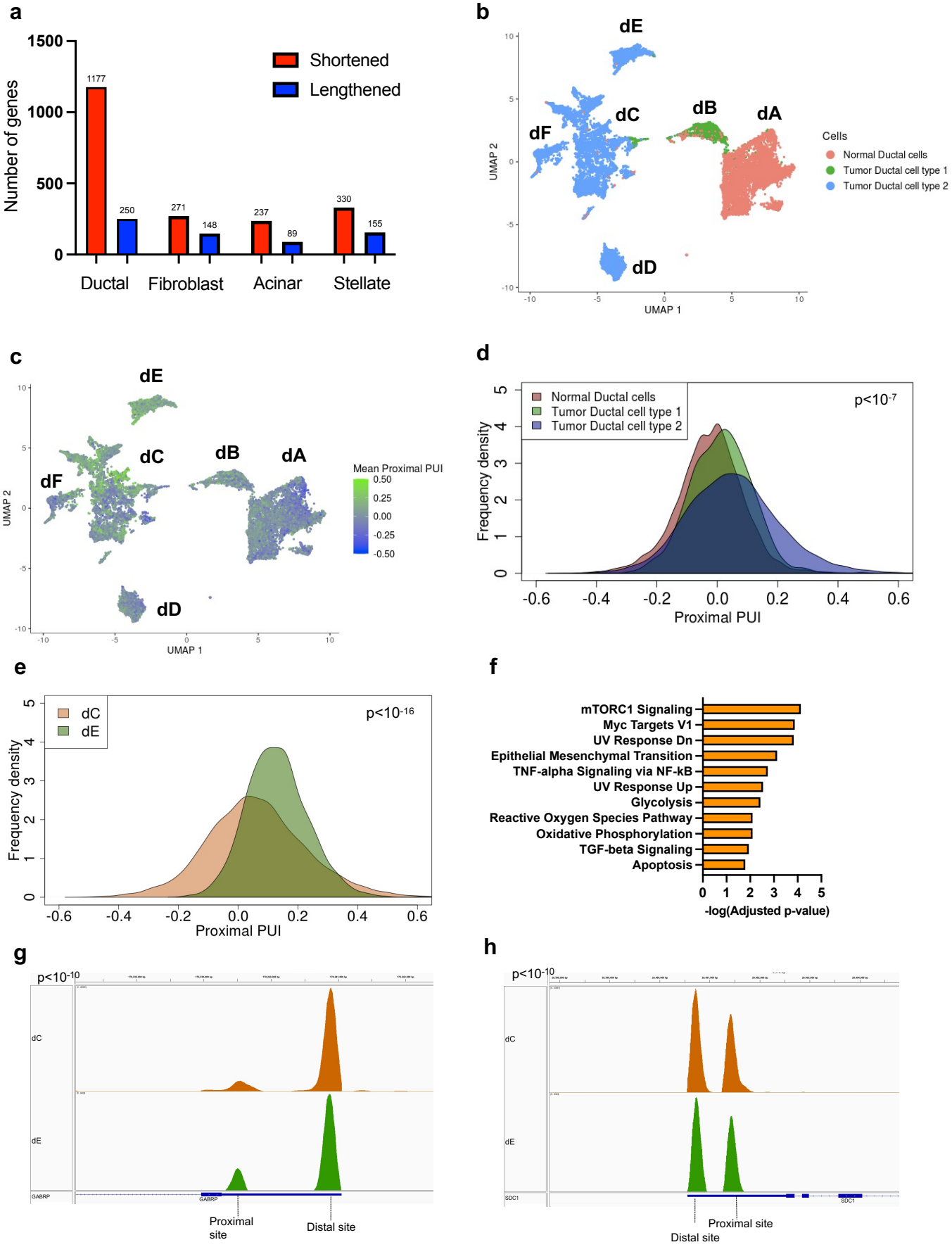
S3b. Violin plots of iCAF markers across normal fibroblast (fA) and specific tumor fibroblast subclusters (fC, fD).

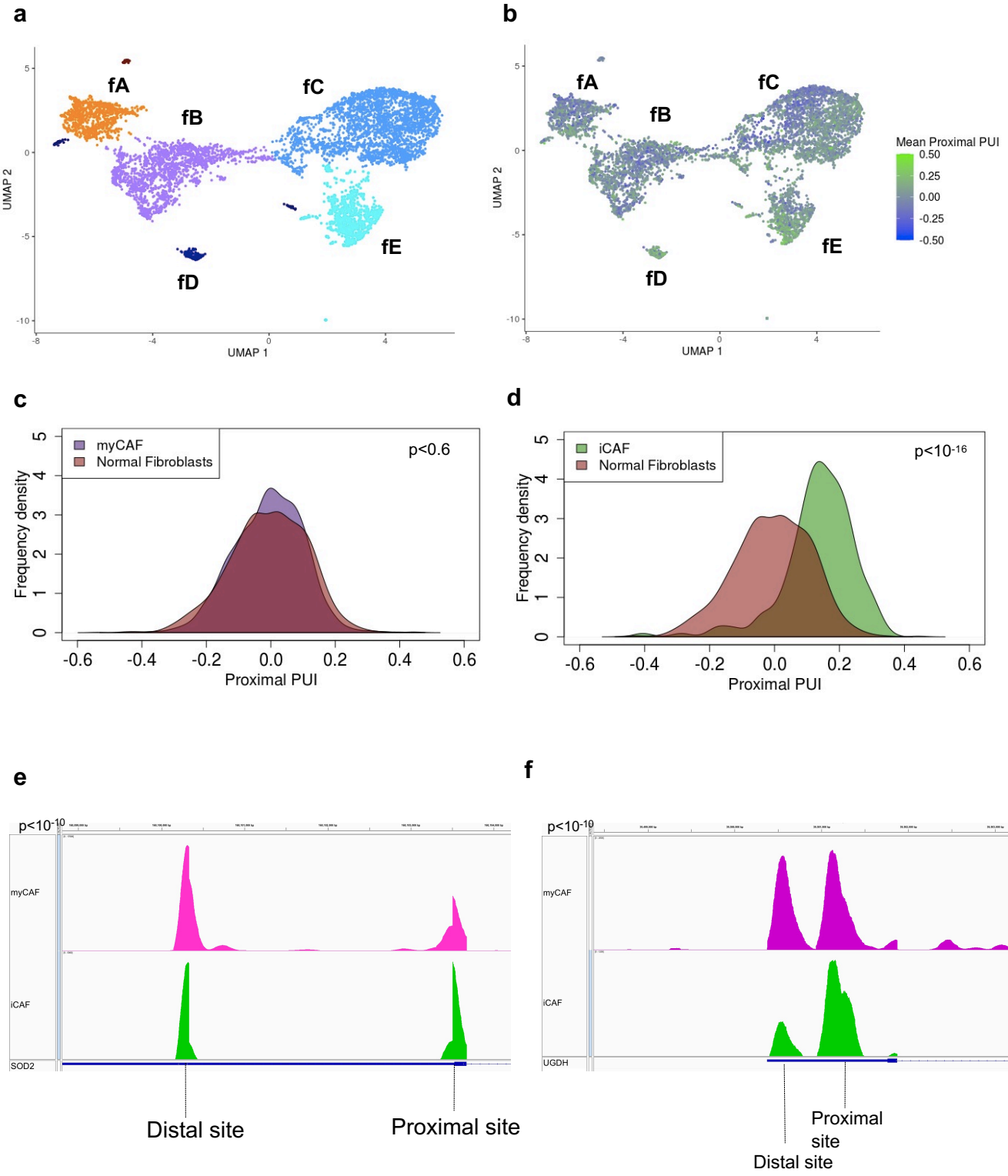
S3c. GSEA of significantly downregulated pathways in iCAFs compared to my CAFs.

S3d. GSEA of significantly upregulated pathways in iCAFs compared to my CAFs.

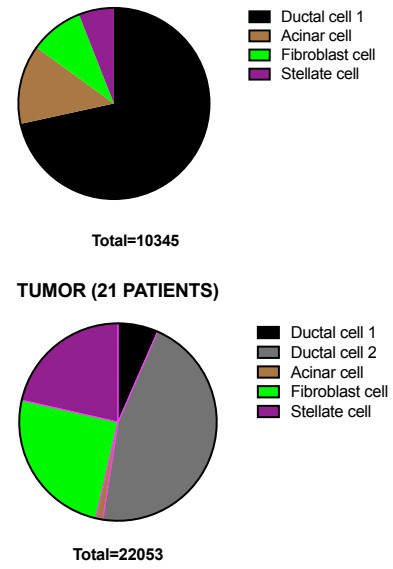
S3e. Distribution of mean proximal PUI of single cells of normal fibroblasts (orange) compared to tumor fibroblasts (blue) ($p<0.005$).

S3f. Distribution of mean proximal PUI of single cells of iCAFs (green) compared to myCAFs (purple) ($p<10^{-16}$).

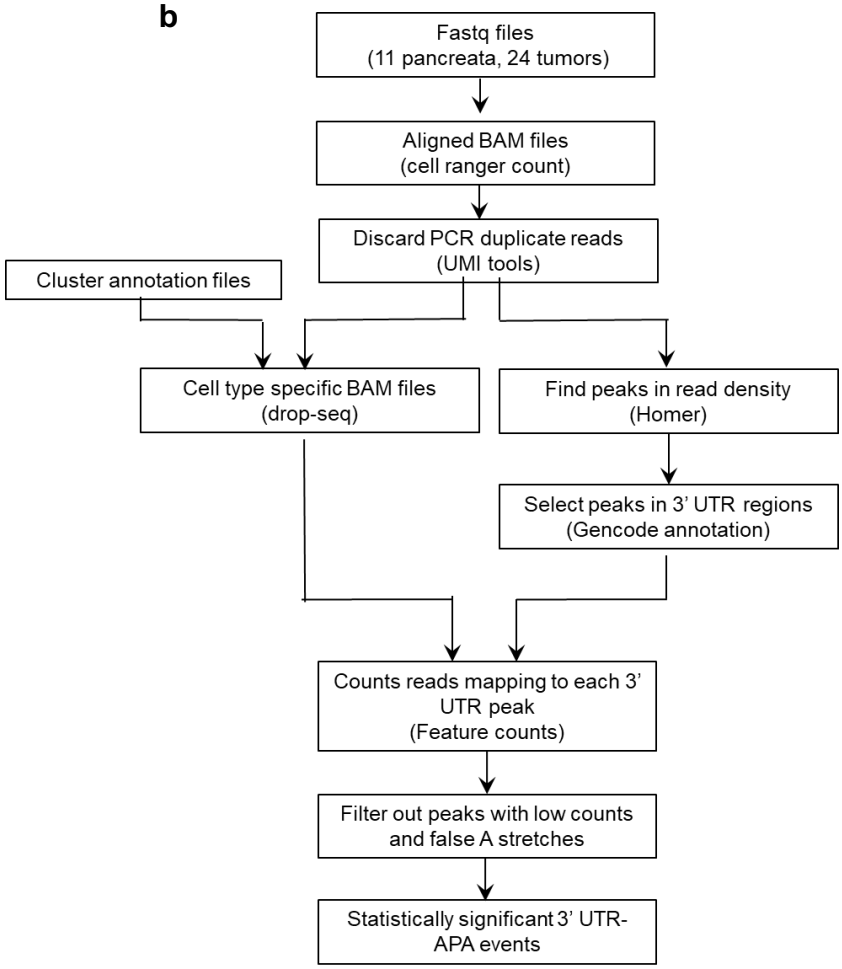

a
Number of genes
1500
1000
500
0
Shortened
Lengthened
1177
250
271
148
237
89
330
155
Ductal
Fibroblast
Acinar
Stellate
b
dE
dC
dB
dA
dF
dD
UMAP 2
UMAP 1
Cells
Normal Ductal cells
Tumor Ductal cell type 1
Tumor Ductal cell type 2
c
dE
dC
dB
dA
dF
dD
Mean Proximal PUI
0.50
0.25
0.00
-0.25
-0.50
UMAP 2
UMAP 1
d
Normal Ductal cells
Tumor Ductal cell type 1
Tumor Ductal cell type 2
p<10-7
Frequency density
Proximal PUI
e
dC
dE
p<10-16
Frequency density
Proximal PUI
f
mTORC1 Signaling
Myc Targets V1
UV Response Dn
Epithelial Mesenchymal Transition
TNF-alpha Signaling via NF-kB
UV Response Up
Glycolysis
Reactive Oxygen Species Pathway
Oxidative Phosphorylation
TGF-beta Signaling
Apoptosis
-log(Adjusted p-value)
g
p<10-10
dC
dE
GABRP
Proximal site
Distal site
h
p<10-10
dC
dE
SDC1
Proximal site
Distal site

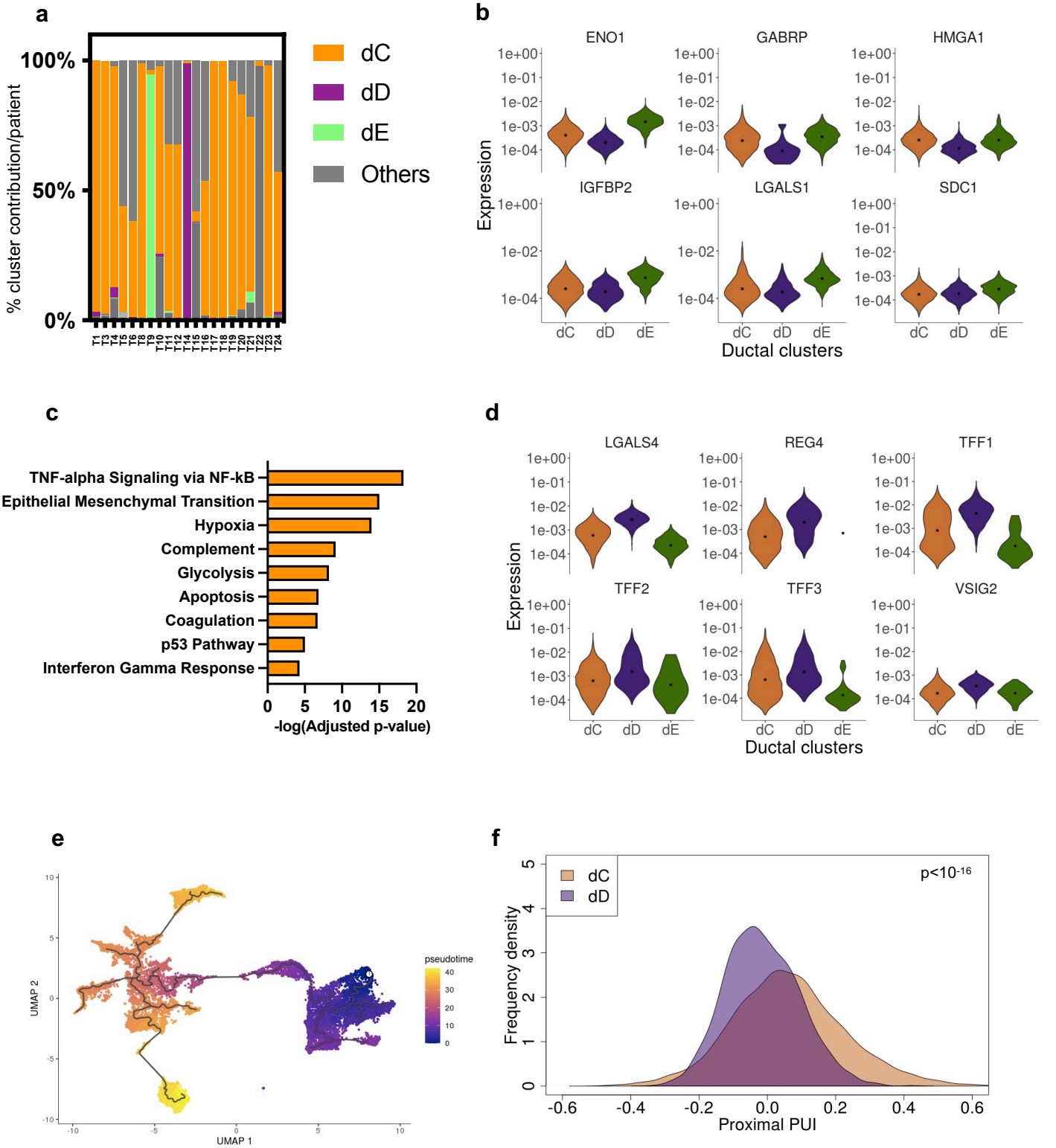

a
fA
fB
fC
fD
fE
UMAP 1
UMAP 2
b
Mean Proximal PUI
0.50
0.25
0.00
-0.25
-0.50
c
myCAF
Normal Fibroblasts
p<0.6
Frequency density
Proximal PUI
d
iCAF
Normal Fibroblasts
p<10-16
e
p<10-10
myCAF
iCAF
SOD2
Distal site
Proximal site
f
p<10-10
UGDH
Proximal site
Distal site

a

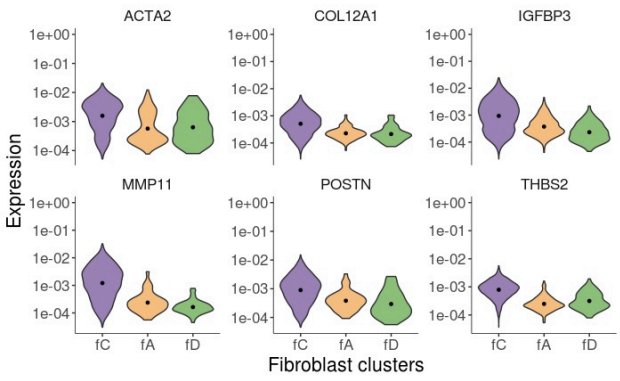

NORMAL (11 Pancreata)
Ductal cell 1
Acinar cell
Fibroblast cell
Stellate cell
Total=10345
TUMOR (21 PATIENTS)
Ductal cell 1
Ductal cell 2
Acinar cell
Fibroblast cell
Stellate cell
Total=22053

b

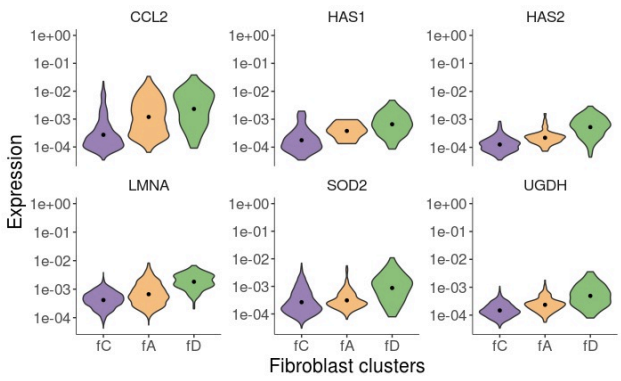

Fastq files
(11 pancreata, 24 tumors)
Aligned BAM files
(cell ranger count)
Discard PCR duplicate reads
(UMI tools)
Cluster annotation files
Cell type specific BAM files
(drop-seq)
Find peaks in read density
(Homer)
Select peaks in 3' UTR regions
(Gencode annotation)
Counts reads mapping to each 3'
UTR peak
(Feature counts)
Filter out peaks with low counts
and false A stretches
Statistically significant 3' UTR-
APA events


a
100%
50%
0%
% cluster contribution/patient
dC
dD
dE
Others
T1 T3 T4 T5 T6 T8 T9 T10 T11 T12 T14 T15 T16 T17 T18 T19 T20 T21 T22 T23 T24
b
ENO1
GABRP
HMGA1
IGFBP2
LGALS1
SDC1
Expression
1e+00
1e-01
1e-02
1e-03
1e-04
dC dD dE
Ductal clusters
c
TNF-alpha Signaling via NF-kB
Epithelial Mesenchymal Transition
Hypoxia
Complement
Glycolysis
Apoptosis
Coagulation
p53 Pathway
Interferon Gamma Response
0 5 10 15 20
-log(Adjusted p-value)
d
LGALS4
REG4
TFF1
TFF2
TFF3
VSIG2
Expression
Ductal clusters
e
UMAP 1
UMAP 2
pseudotime
40
30
20
10
0
f
dC
dD
p<10-16
Frequency density
Proximal PUI
-0.6 -0.4 -0.2 0.0 0.2 0.4 0.6

a

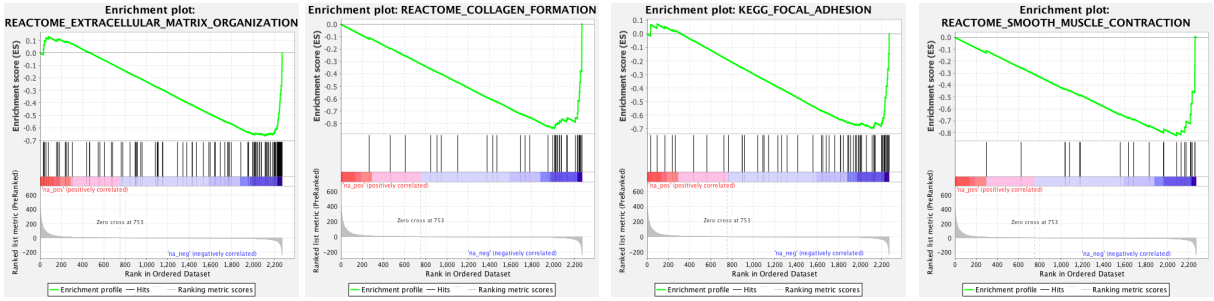


b



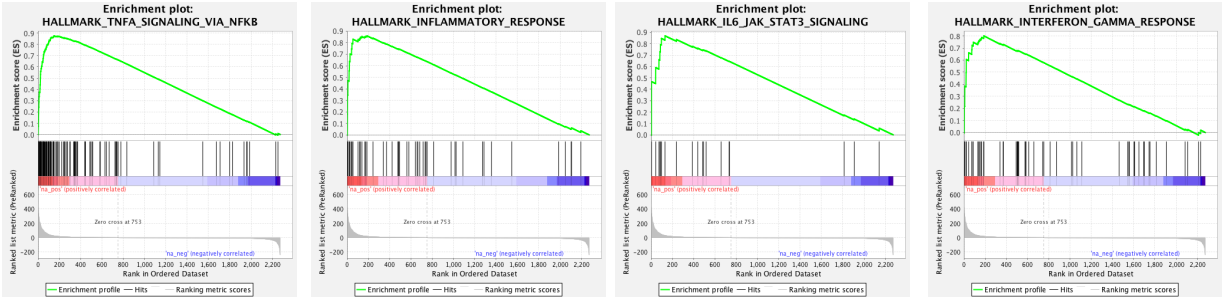
c

Pathways downregulated in iCAFs vs myCAFs

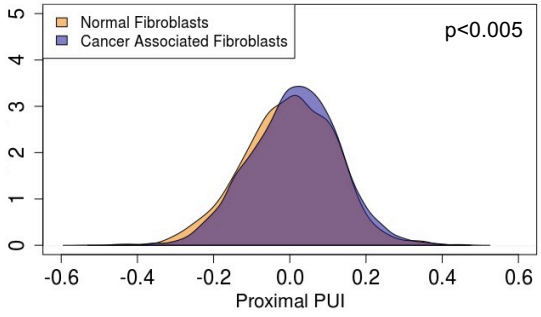


d

Pathways upregulated in iCAFs vs myCAFs



e



f