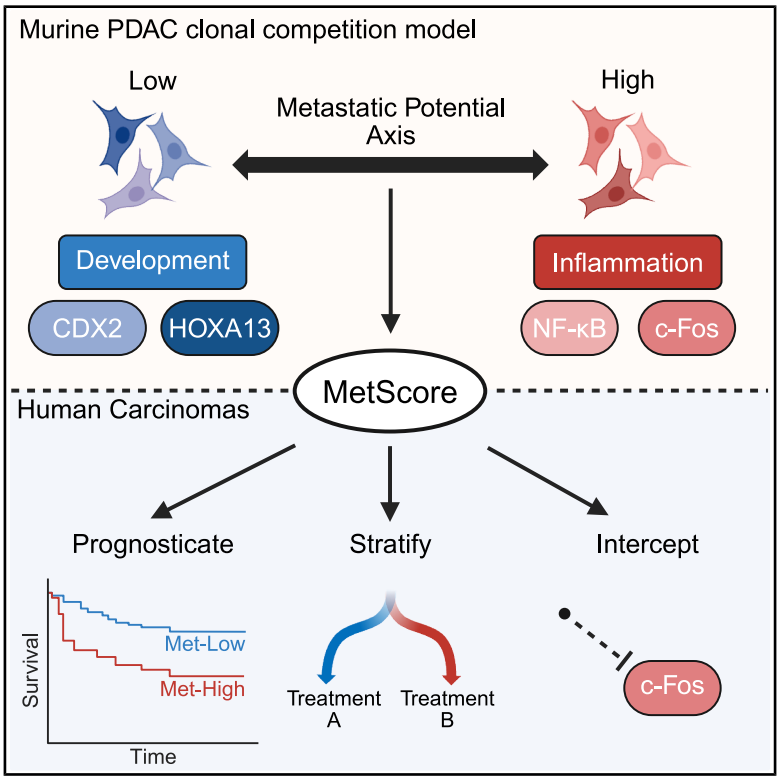
Cell Reports

Article

A cell-state axis underlying colonization in carcinomas with implications for metastasis risk prediction and interception

Graphical abstract



Highlights

- Tracking barcoded cells in immunocompetent mice isolates highly metastatic clones
- Multiomic analysis identifies novel metastatic-potential axis captured in MetScore
- MetScore is orthogonal to previous stratifications and implicates immune remodeling
- MetScore offers prognostic and predictive value across human carcinomas

Authors

Jesse S. Handler, Zijie Li, Rachel K. Dveirin, ..., Hani Goodarzi, Elana J. Fertig, Reza Kalhor

Correspondence

jesse.stone.handler@emory.edu (J.S.H.), kalhor@jhu.edu (R.K.)

In brief

Handler et al. track PDAC subclones in metastatic competition under immunosurveillance using barcodes, identifying a metastatic-potential cell-state axis orthogonal to normal-to-PDAC and classical-basal axes. Scoring samples on this axis using MetScore, they show it is conserved across human carcinomas and offers prognostic and predictive value in patients.

Handler et al., 2025, Cell Reports *44*, 116701
December 23, 2025
https://doi.org/10.1016/j.celrep.2025.116701

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

A cell-state axis underlying colonization in carcinomas with implications for metastasis risk prediction and interception

Jesse S. Handler,[1,2,3,11,*] Zijie Li,[1,2,4] Rachel K. Dveirin,[1,2] Jessica D. Lin,[1,2,5] Weixiang Fang,[1,2] Shihan Wu,[1,2] James E. Forsmo,[1,2] Hani Goodarzi,[6,7] Elana J. Fertig,[1,3,8,9] and Reza Kalhor[1,2,5,10,12,*]

[1]Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[2]Center for Epigenetics, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[3]Sidney Kimmel Comprehensive Cancer Center, Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[4]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, USA
[5]Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[6]Department of Biochemistry & Biophysics and Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, USA
[7]Arc Institute, Palo Alto, CA, USA
[8]Convergence Institute, Johns Hopkins Data Science and AI Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA
[9]Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD, USA
[10]Department of Molecular Biology and Genetics, Department of Medicine, and Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[11]Present address: Department of Hematology and Medical Oncology, Winship Cancer Institute at Emory University School of Medicine, Atlanta, GA, USA
[12]Lead contact
*Correspondence: jesse.stone.handler@emory.edu (J.S.H.), kalhor@jhu.edu (R.K.)
https://doi.org/10.1016/j.celrep.2025.116701

SUMMARY

Metastasis to the liver drives mortality in pancreatic ductal adenocarcinoma (PDAC), yet mechanisms of colonization remain unclear. Using genomic barcoding, we developed a clonal competition model under immune surveillance, isolating murine PDAC subclones with high or low liver-colonization potential. Combined transcriptome and chromatin-accessibility analyses revealed a distinct "metastatic-potential axis," separate from the normal-to-PDAC and classical-basal axes. We established "MetScore" as a biomarker of this axis. MetScore distinguishes metastases from primary PDAC tumors in patients, predicts outcomes beyond classical-basal classifications, and generalizes across carcinoma subtypes, suggesting conserved colonization mechanisms. High-MetScore PDAC cells preferentially occupy immune cell-enriched niches, suggesting they remodel the metastatic microenvironment. Functional screening identified c-Fos as a positive mediator of colonization and a candidate anti-metastatic target. Collectively, we identify a cell-state axis underpinning PDAC liver colonization, introduce MetScore as a broadly applicable biomarker, and nominate actionable targets for peri-operative therapeutic intervention.

INTRODUCTION

The majority of patients with localized pancreatic ductal adenocarcinoma (PDAC) experience distant recurrence following surgical resection,[1] a uniformly fatal event. Liver is the most common metastatic site, and hepatic spread is associated with a particularly poor prognosis.[2] Current adjuvant therapies provide only modest benefit[3] and fail to prevent recurrence in many patients. These challenges highlight a critical unmet need: understanding the molecular mechanisms that enable PDAC liver metastasis to guide the development of novel therapeutic strategies aimed at its interception.

Addressing these challenges requires a more nuanced understanding of PDAC's molecular heterogeneity. PDAC has been divided into two major transcriptional subtypes,[4–6] termed "classical" and "basal," which subsequent single-cell RNA-sequencing (scRNA-seq) analyses[7,8] showed to exist as a spectrum—or axis—within individual tumors. This framework has proved clinically useful: a classical-basal scoring system called purity independent subtyping of tumors (PurIST)[9] is predictive of response to chemotherapy, and a clinical trial is underway to evaluate whether PurIST-guided adjuvant therapy choice improves patient outcomes (ClinicalTrials.gov: NCT06483555). While there is increased prevalence of liver metastases in patients with the basal subtype,[10] liver

metastasis is frequent in both subtypes, and the specific transcriptomic features determining metastatic potential are poorly understood.

Metastasis encompasses multiple stages, including local invasion, extravasation into the circulation, survival as circulating cells, extravasation at distant sites, navigating initial interactions with the local microenvironment (including resident immune cells), and proliferation to form an overt macrometastasis. In the adjuvant setting (where the primary tumor has been removed), only the post-extravasation stages, collectively referred to here as "colonization," remain therapeutically actionable. However, dissecting the molecular features that endow tumor cells with liver-colonization potential is challenging, as direct comparison of metastases with matched primary tumors would fail to deconvolve features that allowed the seeding subclone(s) to colonize from those that evolved during adaptation to the liver microenvironment.

To address this gap, we leveraged genomic barcoding to develop a clonal competition model of PDAC liver colonization under immune surveillance in mice. Using this model, we identified subclones with high or low colonization. Multi-omic comparison of these subclones revealed a set of genes with positive or negative association with colonization, establishing a novel metastatic-potential axis orthogonal to both the normal-to-PDAC and classical-basal axes. We subsequently demonstrated that this metastatic-potential axis, encapsulated in a single-sample scoring metric called "MetScore," is conserved between mouse and human PDAC and across multiple other human carcinomas. We characterized the microenvironments of highly and poorly metastatic PDAC cells in human tumors and found that the former is enriched for immune cells and depleted for non-immune stromal cells. Finally, we performed a functional screen of MetScore genes associated with immune response, revealing c-Fos as a positive mediator of liver colonization. Overall, our study identifies a cell-state axis underlying metastatic potential in multiple carcinomas and sets the stage for developing both diagnostic tools for assessing metastasis risk and novel therapies targeting the colonization steps of the metastatic cascade.

RESULTS

Isolation of primary PDAC subclones with high and low liver-colonization potential

To characterize the molecular determinants of liver-colonization potential within PDAC tumors, we obtained cells isolated from the primary tumors of two independent KPC mice[11–13] (i.e., *Pdx1-Cre*;*LSL-Kras*$^{G12D/+}$;*Trp53*$^{R172H/+}$), which we denote here as KPC-1 and KPC-2. We barcoded each line by transposition of randomized DNA sequences[14,15] (Figure 1A). This strategy results in each cell receiving a unique combination of sequences that can act as a barcode for tracking it. From barcoded mixtures, we sorted single cells and expanded them to generate monoclonal lines. We obtained 10 barcoded subclones from KPC-1 and seven from KPC-2 and identified the barcode in each subclone using high-throughput sequencing.

We next quantified the liver-colonization potential of the isolated subclones. Equal mixtures of all KPC-1 or KPC-2 subclones were transplanted into immunocompetent syngeneic C57BL/6 mice via splenic injection, a standard approach for modeling metastasis to the liver that includes only the post-extravasation steps[12,13] (Figure 1B). After 4 weeks, liver metastases were harvested and barcode sequencing was performed to identify contributing subclones. The majority (80%) of metastases were polyclonal, with a plurality being biclonal (Figure 1C), consistent with prior reports.[16,17] To quantify colonization potential while minimizing confounding by post-seeding proliferation, we measured the frequency with which each subclone appeared across metastases, rather than its relative abundance. This analysis revealed a broad spectrum of colonization capacities within both KPC-1 and KPC-2 cohorts, with some subclones detected in nearly all metastases and others observed rarely (Figure 1D). Subclones present in >50% of metastases were classified as met-high, and the remainder as met-low. To evaluate whether this phenotype was liver specific, we repeated the assay using intraperitoneal injection, which models post-dissemination colonization of the peritoneum[18] (Figures 1B–1D). Subclonal colonization potential was consistent across both anatomic sites, suggesting that the met-high subclones possess a general advantage in metastatic colonization rather than liver-specific tropism.

To identify interactions between subclones, we systematically tested whether each pair of subclones was co-represented more or less than would be expected by chance. There were no significant interactions in the KPC-1 experiments (Figure S1A). However, in the KPC-2 liver-colonization experiment, there were several instances of met-low subclones being significantly over-represented in the presence of a met-high subclone or another met-low subclone (Fisher's exact test, false discovery rate [FDR] <0.05). These results suggest that met-low subclones can successfully colonize by "piggybacking" on met-high subclones or cooperating with other met-low subclones. In contrast, only met-high subclones could form metastases independently (Figure S1B).

To exclude differences in proliferative capacity as a confounding factor, we conducted *in vitro* proliferation and competition assays, neither of which demonstrated a growth advantage for met-high subclones (Figures 1E and 1F). All subclones exhibited recombination of the *LSL-Kras*$^{G12D/+}$ and *LSL-Trp53*$^{R172H/+}$ alleles (Figure S1C). Additionally, all but one subclone (KPC-2_LoA) had undergone loss-of-heterozygosity at the wild-type (WT) *Trp53* locus; KPC-2_LoA was excluded from further analyses to limit our cohort to fully transformed subclones.[19]

Collectively, these results establish that our met-high PDAC subclones have an advantage in metastatic colonization compared to the met-low.

Identification of genes defining met-high and met-low states

We reasoned that successful cells in our clonal-competition assay must realize their advantage in completing the latter half of the metastatic cascade immediately after injection; therefore, these cells are likely to be expressing the genes that confer them this advantage at the time of injection. We further reasoned that genes responsible for conferring this stable metastatic advantage are likely to be under regulation required for mitotic transmission. In this work, we define "epigenetics" as the mitotically heritable

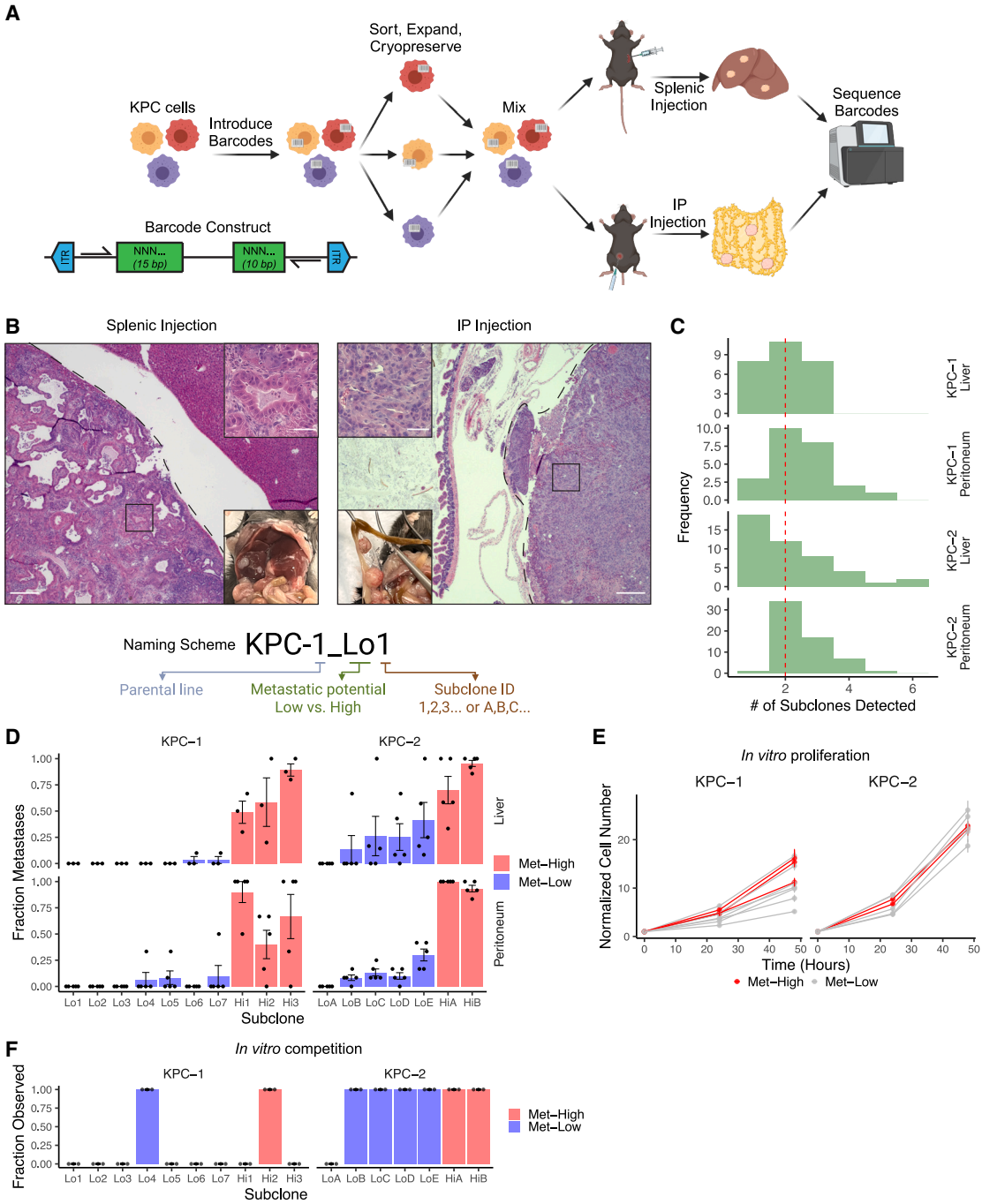


Figure 1. Isolation of primary PDAC subclones with high and low liver-colonization potential

(A) Schematic overview.

(B) Representative gross pathology and H&E-stained formalin-fixed paraffin-embedded (FFPE) sections. Scale bars: 500 μm and 50 μm (insets).

(C and D) (C) Histograms of tumor clonalities—the number of unique subclones per tumor. (D) Detection frequency of each subclone. Points, individual mice; bars, group means ± SEM. Sample sizes: KPC-1 liver, 27 metastases (three mice); KPC-1 peritoneum, 24 metastases (five mice); KPC-2 liver, 46 metastases (five mice); KPC-2 peritoneum, 60 metastases (five mice).

(E) *In vitro* growth curves. Points represent the mean average of four technical replicates, with error bars representing the SEM.

(F) Fraction of three 10-cm dishes in which each subclone was observed after 28 days of passaging.

regulation of gene expression that maintains distinct cell states across cell divisions.[20] Chromatin openness, measurable genome-wide with assay for transposase-accessible chromatin using sequencing (ATAC-seq), integrates diverse epigenetic mechanisms, including DNA methylation, histone modifications, and transcription factor (TF)-driven feedback loops.[21] Therefore, we performed both RNA-seq and ATAC-seq on the met-high ($n = 5$) and met-low ($n = 11$) subclones to characterize their pre-injection transcriptomes and chromatin landscapes, respectively. There were 498 genes with increased expression in met-high subclones and 434 with increased expression in met-low (Table S1), and, out of the 176,964 total shared peaks, there were 2,725 (1.5%) with increased accessibility in met-high subclones and 5,720 (3.2%) with increased accessibility in met-low (Table S2), using an FDR cutoff of 0.05 in both cases. Reassuringly, differentially accessible peaks had the expected signal shape across all the tested subclones (Figure 2A) and a large majority (76%) overlapped candidate *cis*-regulatory elements identified by the Encyclopedia of DNA Elements (ENCODE) project in mouse tissues (Figure S2). Furthermore, unsupervised hierarchical clustering based on either differentially accessible peaks (Figure 2B) or differentially expressed genes (Figure 2C) perfectly segregated met-high subclones from met-low.

Next, we assigned ATAC-seq peaks to their nearest gene and observed correlation of differentially accessible peaks with differential gene expression (Figure 2D). 84% of differentially accessible peaks were assigned to genes with unchanged expression. This result is expected as some peaks may not regulate any genes, their target gene may be in a poised state, or their target may not be the nearest gene. For differentially accessible peaks linked to differentially expressed genes, the genes tended to follow the expected pattern: 97% of significant genes linked to peaks with increased accessibility in met-high subclones had increased expression in met-high subclones, and 91% of significant genes linked to peaks with increased accessibility in met-low subclones had increased expression in met-low subclones. We refer to the genes with differential expression concordant with the differential accessibility of at least one assigned peak as "met-high genes" ($n = 207$; Table S3) or "met-low genes" ($n = 182$; Table S3) based on their increased expression in met-high or met-low subclones, respectively. Met-high and met-low genes define a cell-state axis separating PDAC subclones with high and low metastatic-colonization potential.

To gain insights into the biological functions enabled by these genes, we applied gene set enrichment analysis using the Gene Ontology (GO)[22,23] Biological Process and Kyoto Encyclopedia of Genes and Genomes[24] (KEGG) Pathway databases. The top three out of the five total KEGG gene sets significantly enriched among the met-high genes using an FDR cutoff of 0.05 were related to infection or inflammation (e.g., "TNF signaling pathway"; Figure 2E; Table S4). No GO terms were enriched among the met-high genes. Among met-low genes, there were 50 GO pathways significantly enriched using an FDR cutoff of 0.05 (Figure 2F; Table S4). These pathways were divided by manual curation into three classes: development (19/50; e.g., "muscle cell differentiation"), motility (16/50; e.g., "ameboidal-type cell migration"), and Wnt (7/50; e.g., "Wnt signaling pathway"). The remaining eight pathways could not be confidently assigned to one of these groups. No KEGG terms were enriched among the met-low genes. Review of genes driving enrichment of these pathways (Figure 2G) paints a picture of complex differences between the met-high and met-low subclones, with a clear signal toward activation of inflammation-related genes in met-high and activation of genes specific to adjacent developmental lineages, particularly pancreas neuroendocrine progenitors, in met-low. Interpretation of motility and Wnt pathway enrichment in met-low subclones is more challenging, as the individual genes involved do not point toward a coherent phenotype. In line with the pathway-enrichment results, TF footprinting of ATAC-seq peaks[25] revealed increased differential binding scores for nuclear factor (NF)-κB, a master regulator of inflammation, in met-high subclones (two-sided *t* test, adjusted $p < 2.2E{-}16$ for NFKB2, NFKB1, REL, RELA, and RELB; Figure 2H) and CDX2 and HOXA13, TFs involved in anterior-posterior patterning during embryonic development, in met-low (adjusted $p < 2.2E{-}16$ for both CDX2 and HOXA13; Figure 2H).

The metastatic-potential axis is independent of normal-to-PDAC and classical-basal axes

To place the met-high and met-low subclones within the natural history of PDAC development, we obtained ATAC-seq data from normal pancreata, pre-neoplasia from KC mice (i.e., *Pdx1-Cre*;*LSL-Kras*$^{G12D/+}$), pancreatitis, pre-neoplasia with pancreatitis, and primary PDAC from KPC mice (KPC-0) that Alonso-Curbelo et al.[26] generated in the same mouse strain as ours. All raw data, including our met-high and met-low subclones, were processed through a unified pipeline, followed by batch correction of KPC-1 and KPC-2 samples relative to KPC-0 using ComBat.[27] Principal-component analysis (PCA) recapitulated the expected normal-to-PDAC trajectory along the first principal component (Figure 3A). Met-high and met-low subclones did not separate along the principal components (Figure 3A, inset). To exclude batch correction artifacts, we projected uncorrected subclone data onto PCA axes derived from the Alonso-Curbelo dataset alone, which confirmed that met-high and met-low subclones were indistinguishable along the progression trajectory (Figure S3). These results suggest that the metastatic-potential axis is orthogonal to the normal-to-PDAC axis.

To identify the molecular subtype of our captured subclones, we classified each subclone as classical or basal utilizing PurIST,[9] which generates a single-sample score representing the probability of basal-state occupancy from bulk transcriptomic data. PurIST categorized all met-high and met-low subclones as classical (Figure 3B). Moreover, there was no difference in expression of classical or basal marker genes[5] between met-high and met-low subclones based on single-sample gene set enrichment analysis (ssGSEA) scores or average expression (two-sided Wilcoxon rank-sum tests; Figures 3C and 3D). These results indicate that the metastatic-potential axis is also orthogonal to the classical-basal axis.

The metastatic-potential axis is conserved between mouse and human PDAC

Capture and characterization of successful subclones in their mid-metastatic state is not feasible in patients with PDAC; however, there is a wealth of data from primary and established

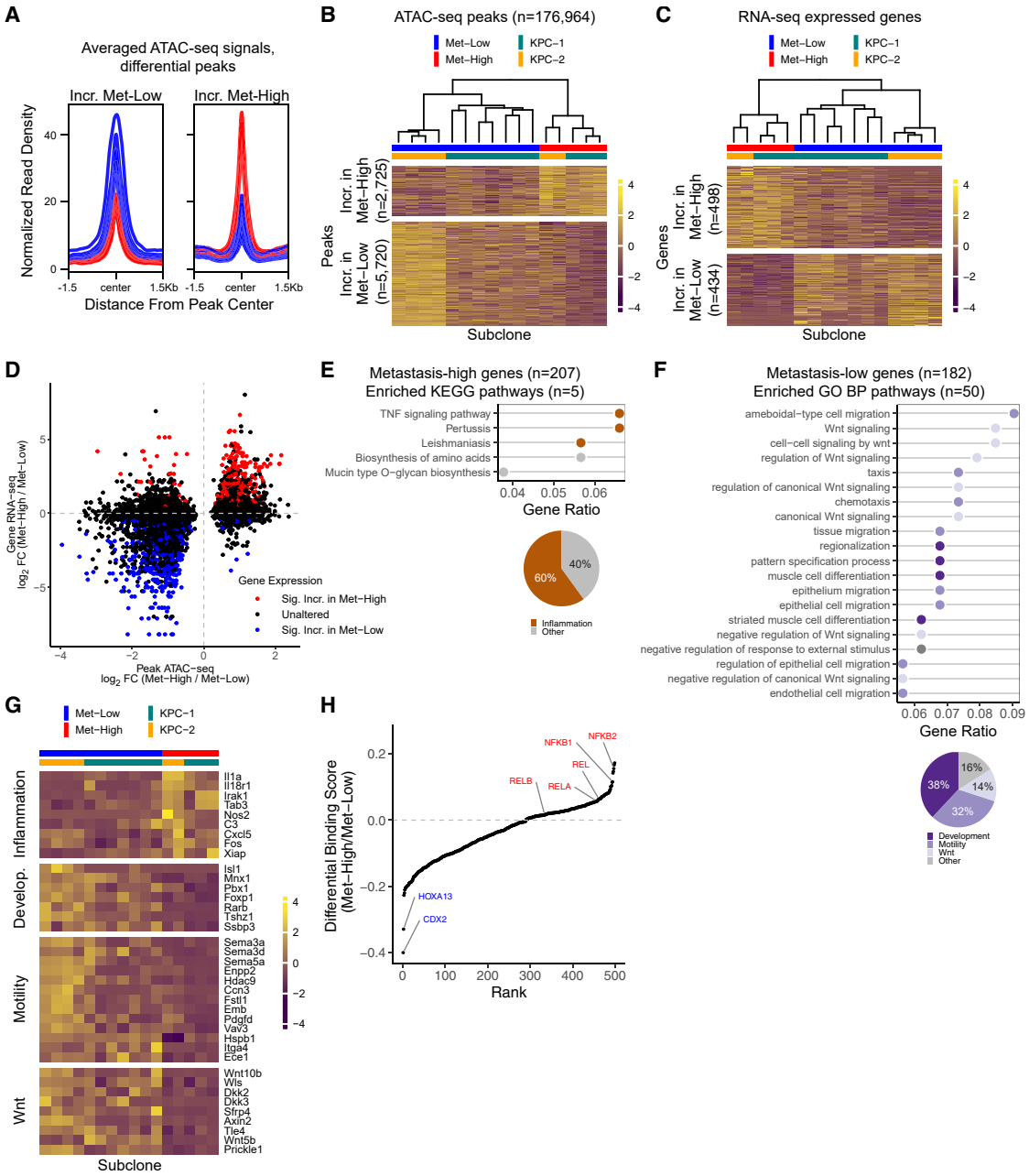


Figure 2. Identification of genes defining met-high and met-low states

(A) Averaged ATAC-seq signal for peaks with significantly greater accessibility in met-low (left) or met-high (right) subclones (FDR <0.05). Each line represents a subclone, colored by metastatic potential (red = high, blue = low).

(B and C) Heatmaps showing scaled, log-transformed normalized (B) accessibility or (C) expression for differentially accessible peaks or differentially expressed genes, respectively (FDR < 0.05). Subclones clustered by Pearson correlation.

(D) Scatterplot showing the relationship between differential chromatin accessibility (*x* axis) and differential gene expression (*y* axis) for each significant peak. Points are colored based on whether the nearest gene is differentially expressed.

(E and F) Dot plots of enriched pathways among met-low (E) or met-high (F) genes (FDR < 0.05).

(G) Heatmap of scaled, log-transformed, normalized expression for module genes across subclones.

(H) Rank-ordered plot of differential binding scores (DBS) for significant TF motifs (n = 499; two-sided t test, Bonferroni-adjusted $p < 0.05$), with positive values indicating increased binding in met-high.

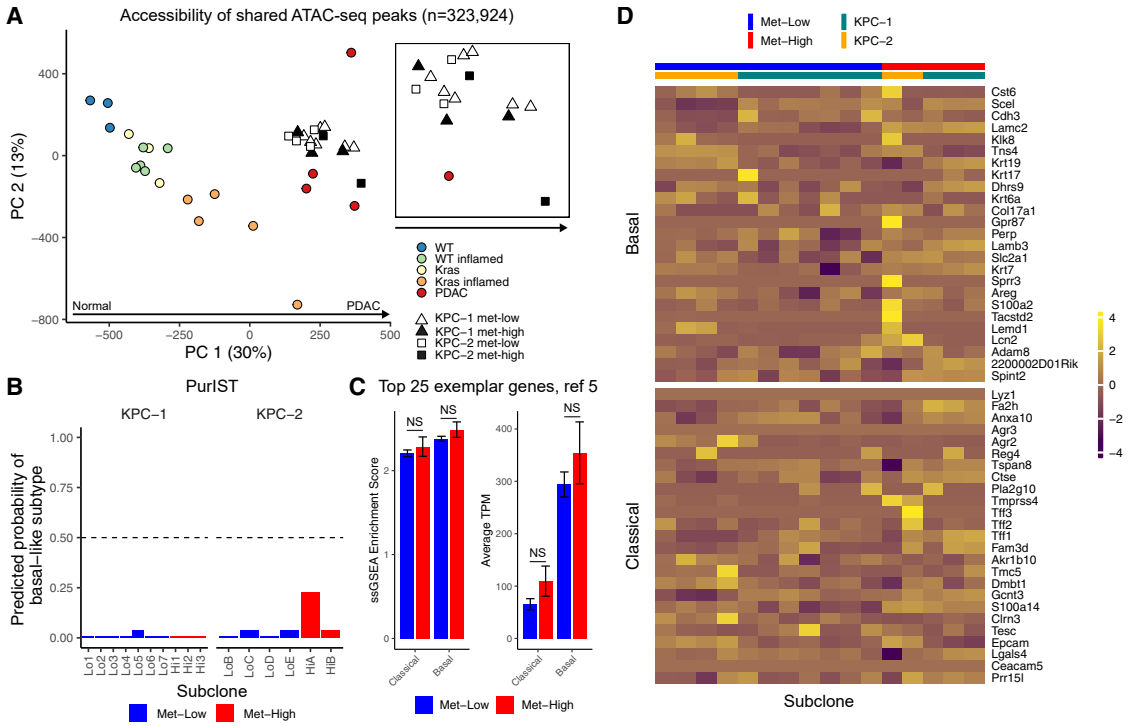


Figure 3. The metastatic-potential axis is independent of normal-to-PDAC and classical-basal axes

(A) PCA of normalized ATAC-seq signal across a consensus peak set.

(B) Predicted basal-like subtype probabilities based on PurIST, with values <0.5 indicating likely classical subtype.

(C) Bar plots showing ssGSEA scores (left) and average transcripts per million (TPM) (right) for classical and basal marker genes across met-high and met-low subclones. Bars, means ± SEM. NS, not significant ($p > 0.05$; two-sided Wilcoxon rank-sum tests).

(D) Heatmap of scaled, log-transformed, normalized expression for the marker genes used in (C) across subclones.

metastatic tumors. To determine whether we can leverage these data to assess our metastatic-potential axis in human PDAC, we tested whether established metastases in the mouse model display the met-high signature. We performed ATAC-seq before and after liver colonization on a monoclonal met-high KPC cell line (Figure S4A). The line was engineered to express GFP, which enabled the purification of cancer cells from stromal cells in the post-metastatic tumor via cell sorting. The ratio of mutant to WT reads at the *Kras* and *Trp53* loci in the sorted tumors showed that they contained less than 5% stromal cell contamination (Figure S4B). The majority of both the met-high (72.7%) and met-low (74.2%) peaks were either unchanged or underwent phenotype-concordant changes in accessibility during liver colonization (e.g., met-high peaks becoming more open; Figure S4C), suggesting that the open chromatin regions that allow a PDAC subclone to successfully colonize the liver are largely maintained during metastatic outgrowth.

With this information, we tested whether our gene-expression signature is enhanced in human PDAC metastases compared to primary tumors. To do so, we established MetScore, which uses rank-based statistics[28] to score a sample's gene-expression profile with respect to enrichment of met-high genes and depletion of met-low genes (n = 202 and 174 human orthologs, respectively; STAR Methods). We first focused on two cohorts of human PDAC with gene expression from primary and metastatic tumors available: a cohort denoted as "PACA-US," which includes microarray data from primary and metastatic tumors harvested from patients at time of autopsy as well as resected primary tumors from living patients,[5] and the International Cancer Genome Consortium PACA-CA cohort, which includes RNA-seq data from tumor biopsies harvested from patients with untreated locally advanced or metastatic PDAC enrolled in the COMPASS[29] and PanGen[30] trials. MetScores were significantly higher in metastases compared to primary tumors in both PACA-US and PACA-CA cohorts (two-sided Wilcoxon rank-sum test, p = 9.1E−6 for PACA-US and 0.040 for PACA-CA; Figure 4A). To control for inter-patient variability, we performed a paired intra-patient comparison in the PACA-US patients with both primary tumors and metastases at the time of sampling, which showed increased MetScores in metastases (linear mixed-effects model, p = 4.5E−4; Figure 4B). Moreover, metastases in PACA-US exhibited higher MetScores than primary tumors in both classical and basal groups based on that study's original labels (two-sided Wilcoxon rank-sum test p = 4.1E−4 for classical and 0.048 for basal; Figure 4C), suggesting that shared mechanisms underlie metastatic colonization across subtypes.

Importantly, met-high and met-low genes accounted for only a small fraction of the thousands of genes differentially expressed between primary tumors and metastases (e.g., 1.1% and 1.2% of metastasis- and primary-upregulated genes in PACA-US,

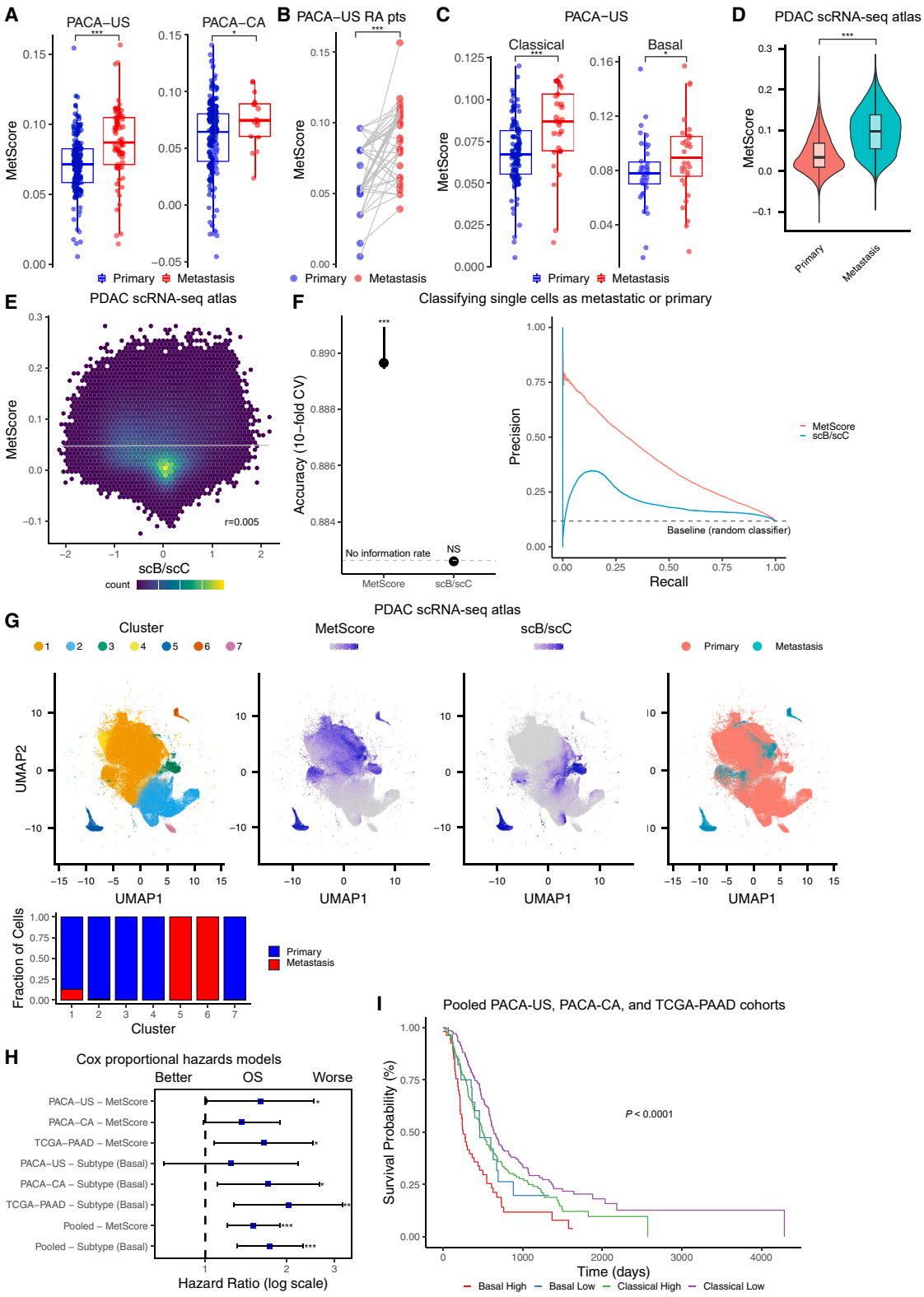


Figure 4. The metastatic-potential axis is conserved between mouse and human PDAC

(A) Boxplots showing MetScore distributions in primary and metastatic tumors across PDAC cohorts. Points represent individual tumors. Two-sided Wilcoxon rank-sum tests; *$p < 0.05$, ***$p < 0.001$.

(legend continued on next page)

respectively; Figure S5A), indicating that a direct primary versus metastasis comparison would not have recovered the metastatic-potential axis. Moreover, a MetScore derived solely from transcriptomic differences between met-high and met-low subclones (MetScore^{RNAonly}) was less effective at distinguishing metastases and primary tumors, demonstrating the added value of integrating chromatin accessibility via RNA-seq and ATAC-seq (Figure S5B). Together, these observations underscore that multi-modal information from the mouse model was critical for distilling the metastatic-potential gene-expression signal from noise in human data.

To assess the relationship between cell-state heterogeneity and metastatic potential within human PDAC at the single-cell level, we leveraged a scRNA-seq atlas[31] comprising 172 primary tumors and 25 metastases, the majority of which (21/25, 84%) were sampled from the liver. We applied MetScore to each tumor cell (n = 214,629 from primary tumors; n = 28,542 from metastases; STAR Methods) and found that cells derived from metastases showed higher scores (Hedges' g = 1.25; generalized linear model, p = 2.4E−8; Figure 4D). These results demonstrate that the MetScore signature reflects tumor cell-intrinsic features and not stromal contamination. We also calculated the single-cell basal versus classical commitment score (scB/scC)[8] for each tumor cell in the atlas and found that MetScore and scB/scC were strikingly uncorrelated (Pearson's r = 0.005; Figure 4E), reaffirming that they are orthogonal biological axes. Accordingly, while a logistic-regression model based on MetScore accurately discriminated metastasis-derived cells from primary tumor-derived cells upon 10-fold cross-validation (one-sided binomial test, p <2.2E−16; precision-recall area under the curve [PRAUC] = 0.394), a model based on scB/scC did not perform any better than the no-information rate (PRAUC = 0.205; Figure 4F).

To further clarify the sources of heterogeneity within and across tumors, we used uniform manifold approximation and projection (UMAP) dimensionality reduction and Louvain clustering, identifying seven distinct states within PDAC cells. Clusters 5 and 6—composed entirely of metastasis-derived cells—exhibited the highest MetScores but were divergent with respect to scB/scC scores, with cluster 5 being more basal and cluster 6 more classical (Figure 4G), demonstrating that elevated metastatic potential can occur independently of basal-state commitment. Most primary tumor-derived clusters (2, 4, and 7) were composed predominantly of cells with a low MetScore. Cluster 1 was more heterogeneous, comprising both primary and metastatic cells and a mixture of high- and low-MetScore states. However, cluster 3—exclusively derived from primary tumors—contained cells with high MetScores, raising the possibility that it represents a subset of primary tumor cells poised for metastasis. Collectively, these results demonstrate that MetScore captures a tumor cell-intrinsic transcriptomic signature underlying metastatic potential that is conserved between mouse and human PDAC and independent of classical-basal subtype.

MetScore is prognostic for survival in human PDAC

Because metastasis is a major driver of mortality in localized and locally advanced PDAC, we explored the relationship between primary tumor MetScore and overall survival in three cohorts with matched survival and transcriptomic data: patients with localized PDAC isolated from the previously analyzed PACA-US cohort; patients with locally advanced PDAC isolated from the previously analyzed PACA-CA cohort; and The Cancer Genome Atlas's (TCGA's) Pancreatic Adenocarcinoma (TCGA-PAAD) cohort,[32] which includes only patients with localized PDAC. In each cohort, we compared the overall survival of patients with tumors in the top half of MetScores to those in the bottom half (Figure S5C). High-MetScore patients had a significantly worse overall survival in all three cohorts (hazard ratio [HR] for high MetScore = 1.6, 1.4, and 1.7; log rank test, p = 0.02, 0.04, and 0.01; for PACA-US, PACA-CA, and TCGA-PAAD, respectively). In contrast to MetScore, MetScore^{RNAonly} was only prognostic in TCGA-PAAD (Figure S5D), further highlighting the importance of incorporating chromatin accessibility to improve classifier robustness and generalizability. Multivariable Cox regression across the three cohorts showed that both MetScore and classical-basal subtype retained independent prognostic significance when included in the same model (Figures 4H and 4I). Meta-analysis confirmed that high MetScore (pooled HR = 1.51, 95% confidence interval [CI] = 1.20–1.89, p = 0.0005) and basal subtype (pooled HR = 1.73, 95% CI = 1.31–2.29, p = 0.0002) were each associated with significantly worse overall survival, with minimal between-study heterogeneity. Individually, MetScore and classical-basal subtype each showed comparable prognostic discrimination (C-index 0.567 for MetScore and 0.566 for classical-basal subtype). A combined model improved performance (C-index 0.590), supporting that these features capture complementary, non-redundant aspects of tumor biology. Together, these results

(B) Dot plot of MetScores for matched primary and metastatic tumors from rapid autopsy patients in the PACA-US cohort. Lines connect samples from the same patient. Linear mixed-effects model; ***p < 0.001.

(C) As in (A) but stratified by molecular subtype as annotated in Moffitt et al.

(D) Violin plot showing MetScore distributions across tumor cells derived from primary tumors and metastases. Generalized linear model; ***p < 0.001.

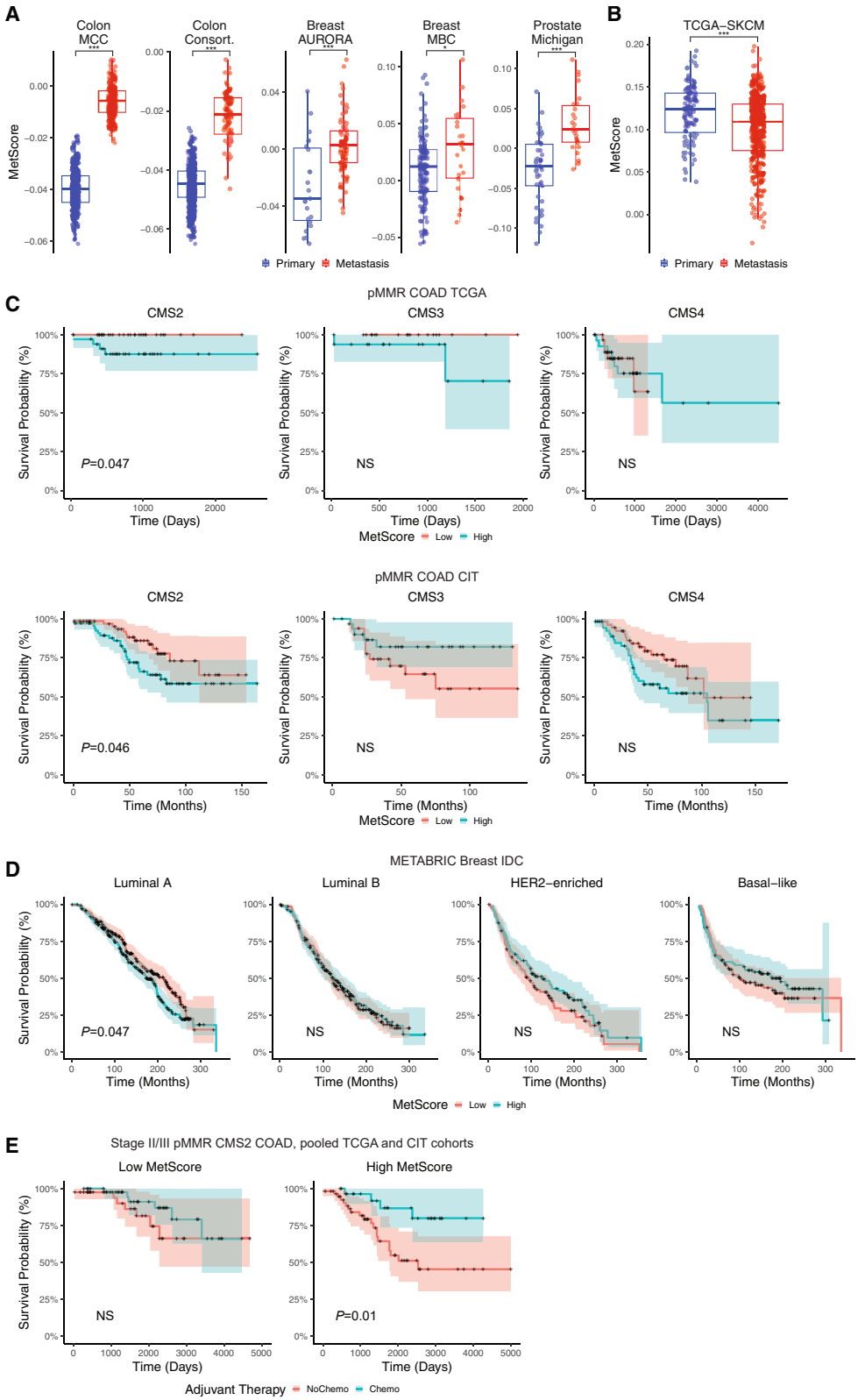
(E) Scatterplot showing the relationship between MetScore and scB/scC across tumor cells. Gray line, linear regression fit. Pearson's r indicated.

(F) Left: dot plot showing accuracy (median ± interquartile range) across 10-fold cross-validation for logistic-regression models predicting donor type (primary vs. metastasis) using either MetScore or scB/scC. One-sided one-sample proportion tests comparing model accuracy to the no-information rate; ***p < 0.001. NS, not significant. Right: PR curves for models trained on MetScore or scB/scC.

(G) Top: UMAP plots of tumor cells colored as indicated. For MetScore and scB/scC, values above the 90th percentile and below the 10th percentile were capped. Bottom: bar plot showing the proportion of primary and metastasis-derived cells within each cluster.

(H) Forest plot showing hazard ratios and 95% confidence intervals for high MetScore (top 50%) and basal subtype (PurIST predicted probability >0.5) with respect to overall survival in Cox proportional hazards models adjusted for both variables. *p < 0.05, **p < 0.01, ***p < 0.001.

(I) Kaplan-Meier curve showing overall survival of patients pooled from cohorts in (H) stratified by MetScore and classical-basal subtype. p value from log rank test.



(legend on next page)

demonstrate that primary tumor MetScore is associated with worse overall survival and, when combined with classical-basal subtype, allows for stratification of patients with PDAC with significantly greater prognostic accuracy.

A shared cell-state axis underlies metastatic potential across multiple human carcinoma subtypes

The steps of the metastatic cascade are similar for different cancer subtypes; thus, we wondered whether the factors that determine metastatic potential are conserved across subtypes. MetScores were significantly higher in metastases compared to primary tumors in all carcinoma cohorts we tested, including two colon adenocarcinoma (COAD) cohorts (two-sided Wilcoxon rank-sum test, $p < 2.2E{-}16$ for Colon-MCC[33] and Colon-Consortium[33]), two breast invasive ductal carcinoma (BRCA) cohorts ($p = 6.0E{-}5$ for Breast-AURORA-US[34] and 0.018 for Breast-MBC[35]), and one prostate adenocarcinoma (PRAD) cohort[36] ($p = 4.4E{-}8$; Figure 5A). In contrast, we observed the opposite trend in a melanoma cohort,[37] where primary tumors had significantly higher MetScores than metastases ($p = 7.9E{-}5$; Figure 5B), suggesting that the metastatic-potential axis defined by MetScore is not shared with this non-epithelial malignancy.[34]

To test whether MetScore is relevant beyond liver and peritoneal metastases, we stratified metastases by anatomic site in cohorts with sufficient sample sizes. In both Colon-MCC and Colon-Consortium, liver and lung metastases each individually had significantly higher MetScores than primary tumors (MCC, FDR < 2.2E−16 for both sites; Consortium, FDR < 2.2E−16 for liver, FDR = 2.5E−12 for lung; Figure S6). In Breast-AURORA, liver, lung, soft tissue, CNS, and bone metastases all individually showed elevated MetScores (FDR = 2.2E−3, 6.4E−3, 6.4E−3, 8.9E−3, and 0.047, respectively), although adrenal and lymph node metastases were not significant. In the Prostate-Michigan cohort, liver, lymph node, and soft tissue metastases each had higher MetScores than primary tumors (FDR = 3.1E−4, 3.1E−4, and 3.2E−4, respectively). Collectively, these findings indicate that the metastatic-potential axis reflects a generalized advantage in metastatic colonization across diverse anatomic sites, not restricted to the liver or peritoneum.

While all carcinomas are epithelial in origin, they can exhibit molecular subtypes with differing degrees of epithelial versus mesenchymal features. To assess whether MetScore's prognostic value depends on epithelial identity, we stratified patients by subtype. In mismatch repair proficient (pMMR) COAD, high primary tumor MetScore was associated with worse overall survival in the strongly epithelial CMS2 ("canonical") subtype across both the TCGA and Cartes d'Identité des Tumeurs COAD (CIT-COAD)[38] cohorts (log rank test, $p = 0.047$ and 0.046, respectively; Figure 5C), but not in the mesenchymal CMS4 subtype. The "metabolic" CMS3 subtype, while epithelial, was the least represented in both datasets, limiting statistical power to detect survival differences. The CMS1 subtype was excluded due to its enrichment for MMR-deficient tumors. Similarly, in the METABRIC[39] breast BRCA cohort, MetScore was prognostic in the strongly epithelial luminal A subtype ($p = 0.047$; Figure 5D), but not in luminal B (less differentiated), basal-like (mesenchymal), or HER2-enriched (variable epithelial-mesenchymal transition [EMT] status) subtypes. Combined, these results indicate that MetScore's prognostic value is not uniform across all carcinoma subtypes but is most apparent in those that retain strong epithelial identity. Moreover, they reinforce that MetScore captures prognostic information that is orthogonal to existing subtype-based classification schemes.

Given MetScore's ability to distinguish primary from metastatic tumors and its prognostic relevance in COAD, we also asked whether it could predict benefit from adjuvant therapy. We pooled patients with stage II/III pMMR CMS2 COAD from the TCGA and CIT cohorts and stratified them by MetScore and peri-operative therapy status. Among patients with high MetScores, adjuvant therapy was associated with significantly improved overall survival (log rank test, $p = 0.014$; Figure 5E), while no significant benefit was observed in patients with low MetScores. Although a formal interaction term in a Cox proportional hazards model was not statistically significant (HR = 0.41, $p = 0.27$), this likely reflects limited power due to the small sample size ($n = 170$ patients) and number of events ($n = 37$ deaths).

Together, these findings indicate that the cell-state axis captured by MetScore is a recurrent feature of metastatic progression in several carcinoma subtypes—particularly those retaining epithelial identity—and may have clinical utility in stratifying patients for adjuvant therapy, especially in COAD.

Met-high PDAC cells enrich immune cells in their microenvironments

Given the enrichment of inflammation-related genes in met-high PDAC subclones, we hypothesized that one mechanism through which met-high cells realize their advantage is by remodeling the metastatic microenvironment. To test this, we calculated the average neoplastic cell MetScore for every tumor in the human PDAC scRNA-seq atlas and assessed correlation with the abundance of stromal cell populations, as originally defined by the study authors. Strikingly, all annotated immune cell abundances were positively correlated with average tumor cell MetScore (Spearman's $\rho = 0.59$, 0.54, 0.53, 0.51, 0.41, 0.38, and 0.34; FDR = 9.2E−16, 4.1E−13, 8.2E−13, 1.1E−11, 7.2E−8, 1.2E−6, and 1.1E−5 for T cell and natural killer cell (T&NK cell), cycling T&NK, myeloid, B cell, mast cell, plasma cell, and cycling myeloid, respectively; Figure 6A; Table S5). Meanwhile, multiple non-immune stromal cell abundances were negatively correlated ($\rho = -0.41$, −0.30, and −0.28; FDR = 8.6E−8, 1.3E−4, and 3.3E−4; for endocrine, cancer-associated fibroblast, and

Figure 5. A shared cell-state axis underlies metastatic potential across multiple human carcinoma subtypes

(A and B) Boxplots showing MetScore distributions in primary and metastatic tumors across the indicated carcinoma cohorts. Points represent individual tumors. Two-sided Wilcoxon rank-sum tests; $^{*}p < 0.05$, $^{***}p < 0.001$.

(C and D) Kaplan-Meier curves showing overall survival stratified by MetScore (high = top 50%, low = bottom 50%) in each cohort. p values from log rank tests.

(E) Kaplan-Meier curves for overall survival in patients with stage II/III pMMR CMS2 COAD in the TCGA and CIT cohorts, stratified by receipt of post-operative therapy (yes vs. no) within low-MetScore (bottom 50%; left) and high-MetScore (top 50%; right) groups. p values from log rank tests.

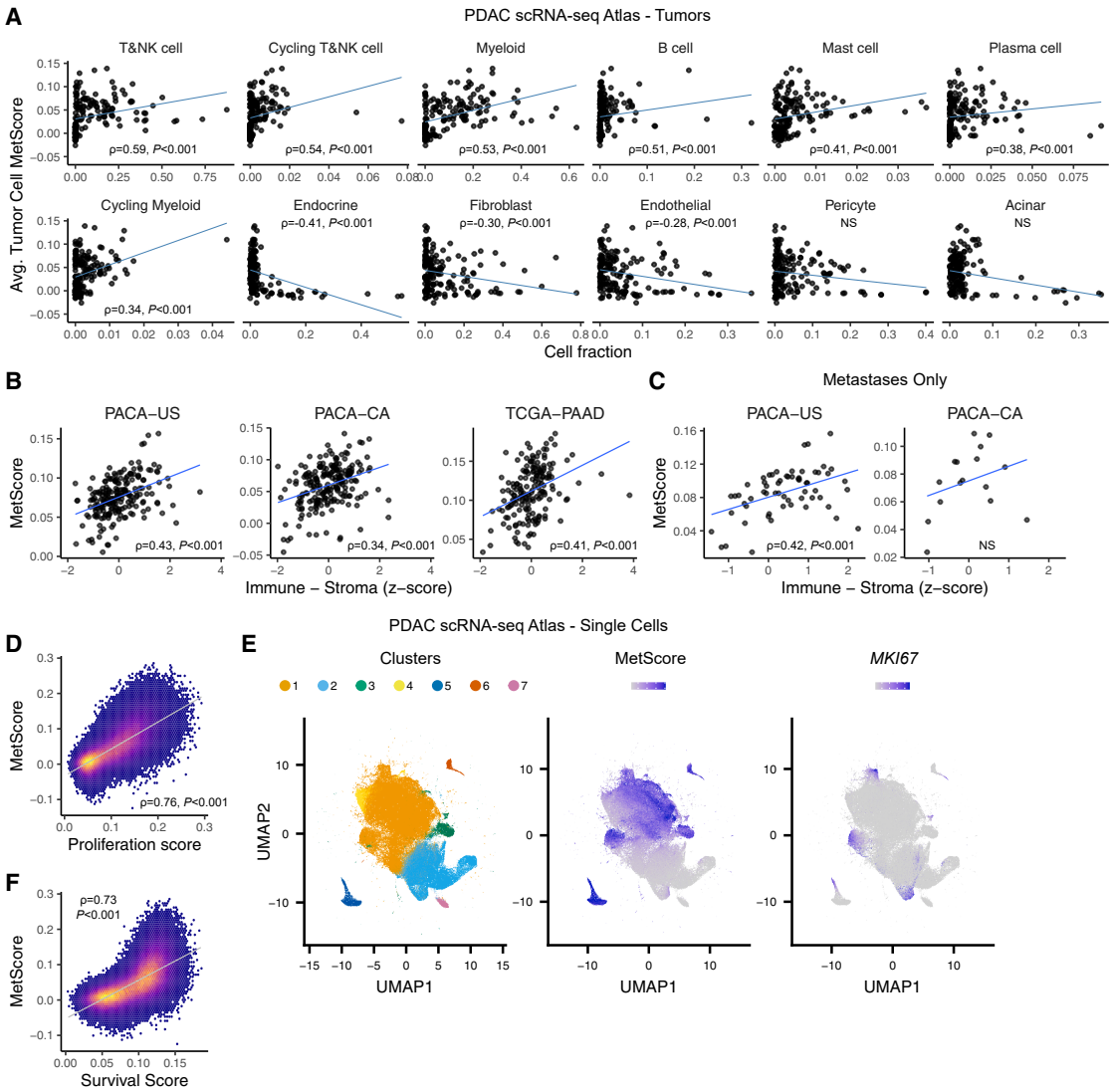


Figure 6. Met-high PDAC cells enrich immune cells in their microenvironments

(A) Scatterplots showing the relationship between average neoplastic cell MetScore and the abundance of stromal populations across tumors in the human PDAC scRNA-seq atlas.

(B) Correlation between bulk tumor MetScore and the immune-stromal score, calculated as the difference between scaled immune and stromal scores from the ESTIMATE algorithm.

(C) The same analysis restricted to metastases.

(D) Scatterplot showing correlation between single-cell MetScore in tumor-enriched epithelial cells in the PDAC scRNA-seq atlas and a proliferation score.

(E) UMAP plots of tumor cells colored as indicated. For MetScore and *MKI67*, values above the 90th percentile and below the 10th percentile were capped.

(F) Scatterplot showing correlation between single-cell MetScore in tumor-enriched epithelial cells in the PDAC scRNA-seq atlas and a survival score.

In (A–D) and (F), linear regression lines are shown, with Spearman's rho and *p* values (FDR adjusted for A–C).

endothelial, respectively; no significant correlation for acinar cells or pericytes).

To validate these findings in additional patient cohorts, we deconvolved bulk transcriptomic data in the PACA-US, PACA-CA, and TCGA-PAAD cohorts into tumor, immune, and non-immune stromal contributions using estimation of stromal and immune cells in malignant tumor tissues using expression data (ESTIMATE).[40] In all three datasets, the difference between immune and non-immune stromal *Z* scores was positively correlated with MetScore (Spearman's ρ = 0.43, 0.34, 0.41; p = 2.0E−10, 2.4E−8, and 4.3E−7; for PACA-US, PACA-CA, and TCGA-PAAD, respectively; Figure 6B). Importantly, this trend was observed in metastases, significantly for PACA-US (ρ = 0.42, p = 9.0E−4; Figure 6C) and exhibiting a strong trend for PACA-CA (ρ = 0.29, p = 0.27), likely due to the limited number of metastases (n = 16) in this cohort versus PACA-US (n = 61).

Together, these findings demonstrate a consistent link between met-high PDAC cells and immune cell-enriched, non-immune

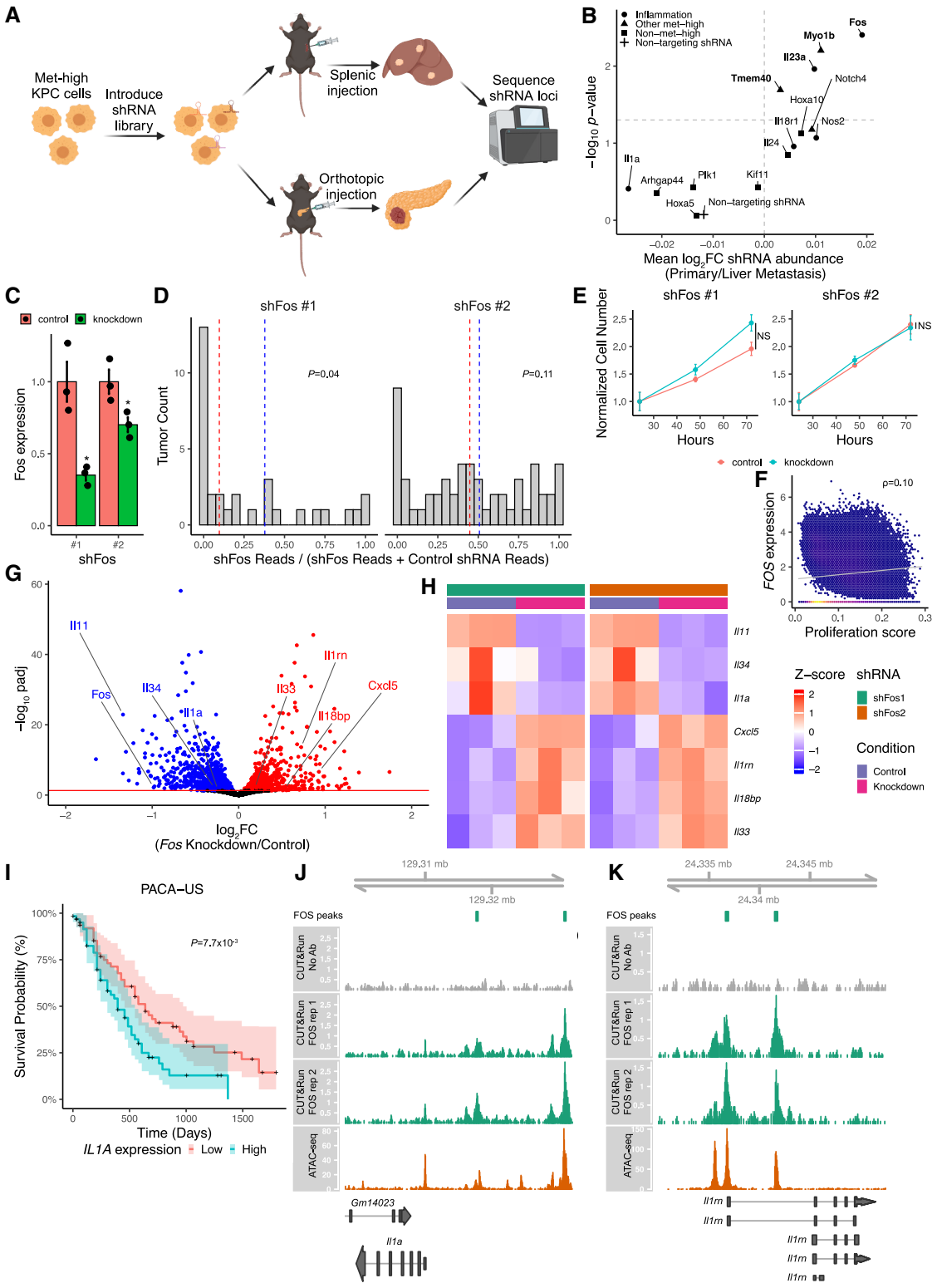


Figure 7. c-Fos is a positive functional mediator of PDAC liver colonization

(A) Experimental design for (B).

(B) Volcano plot showing mean $\log_2$ fold change in shRNA abundance between primary tumor and liver metastasis conditions (averaged across all shRNAs per gene) and $-\log_{10}$ weighted combined *p* value from a linear model. Sample sizes: primary (n = 6 tumors, 6 mice); liver metastasis (n = 36 tumors, 4 mice).

(legend continued on next page)

stromal cell-depleted microenvironments. This association could reflect two possibilities: either met-high cells remodel their surroundings to favor immune accumulation or they arise within pre-existing immune-enriched niches. The recurrence of this pattern in metastases, where new microenvironments must be established, supports remodeling.

While our results point to microenvironmental remodeling as a key cancer cell-extrinsic mechanism through which met-high cells achieve their metastatic advantage, we also found evidence for complementary cancer cell-intrinsic mechanisms. Specifically, MetScore was strongly correlated with a conserved gene-expression signature that predicts cell and lineage-specific differences in growth rate[41] (Spearman's $\rho = 0.76$, $p < 2.2E{-}16$; Figure 6D). Mapping *MKI67* expression across the single-cell UMAP (Figure 6E) revealed groups of *MKI67*$^{+}$ cells within MetScore-high regions—particularly clusters 5 and 6 (exclusively metastasis-derived) and cluster 1 (mixed primary and metastatic). In contrast, MetScore-high cells in cluster 3 (exclusively primary tumor derived) were *MKI67*$^{-}$, indicating that high MetScore is not synonymous with proliferation. A distinct group of *MKI67*$^{+}$ cells was also present within a MetScore-low, basal-differentiated region in cluster 2. These results suggest that met-high PDAC cells—despite comparable intrinsic proliferative potential *in vitro*—are more likely to be in an active proliferative state *in vivo*, particularly following colonization, but that elevated MetScore is not required for robust proliferation. In addition, MetScore was correlated with a gene signature associated with cellular survival ("GOBP_NEGATIVE_REGULATION_OF_APOPTOTIC_SIGNALING_PATHWAY"; ρ 0.73, $p < 2.2E{-}16$; Figure 6F).

Taken together, these findings support a model in which met-high cells realize their advantage through a combination of cancer cell-intrinsic traits, including enhanced proliferation and survival, and cancer cell-extrinsic processes such as remodeling of the target-site microenvironment.

c-Fos is a positive functional mediator of PDAC liver colonization

To identify features of the met-high state that may underlie remodeling of the metastatic microenvironment observed in human PDAC, we performed a short hairpin RNA (shRNA) screen in the liver-colonization context (Figure 7A), targeting inflammation-related met-high genes (*Il1a*, *Il18r1*, *Nos2*, *Fos*, and *Il23a*), other met-high genes (*Tmem40*, *Myo1b*, and *Notch4*), and a few non-met-high genes from the mouse genome to serve as controls (STAR Methods). The results showed that cells carrying shRNAs targeting *Fos* and *Il23a* were significantly depleted in liver metastases relative to primary tumors (linear model, weighted combined $p = 3.9E{-}3$ for *Fos* and 0.011 for *Il23a*; Figure 7B), identifying these inflammation-related genes as novel functional enablers of liver colonization. In addition, shRNAs targeting the met-high genes *Myo1b*, which encodes an atypical myosin, and *Tmem40*, which encodes a transmembrane protein with poorly characterized function, were similarly significantly depleted in liver metastases relative to primary tumors (weighted combined $p = 6.3E{-}3$ for *Myo1b* and 0.020 for *Tmem40*; Figure 7B), suggesting that diverse pathways active within the met-high state contribute to metastatic competence. These findings confirm that met-high genes are enriched for functional regulators of PDAC liver colonization and nominate multiple avenues for therapeutic intervention.

We focused on our top hit, *Fos*, encoding the TF c-Fos, for further validation. Using the KPC-2_HiB subclone, which originates from a different tumor than KPC-1_Hi2 used in the screen, we generated cell lines harboring shRNAs targeting *Fos* or a control non-targeting shRNA and validated knockdown using RT-qPCR (Figure 7C). We then performed splenic injections using equal mixtures of *Fos*-targeting/non-targeting shRNA-harboring cells, harvested the resulting liver metastases, and sequenced the shRNA amplicon. Cells carrying shRNAs targeting *Fos* were depleted in liver metastases relative to the starting population in a manner correlated to the degree of knockdown (Figure 7D). This depletion was significant for shFos #1 (one-sided Wilcoxon signed-rank test, $p = 0.04$) and exhibited a strong trend for shFos #2 that did not reach statistical significance ($p = 0.11$), likely due to the modest degree of knockdown (30%) with this shRNA. Importantly, *Fos* knockdown did not affect *in vitro* proliferation (Figure 7E), suggesting its role in liver metastasis is mediated through mechanisms specifically promoting colonization rather than a generic effect on cellular fitness. Supporting this, *FOS* expression was uncorrelated with the proliferation gene signature among tumor-enriched epithelial cells in the human PDAC scRNA-seq atlas (Spearman's $\rho = 0.10$; Figure 7F).

To identify the downstream programs influenced by c-Fos, we performed RNA-seq on the two *Fos*-knockdown and control

(C) RT-qPCR quantification of *Fos* mRNA in KPC-2_HiB cells expressing shRNAs targeting *Fos* or a non-targeting control. *Rps29*, housekeeping gene. Bars, mean ± SEM from three technical replicates. One-sided Wilcoxon rank-sum test; $^{*}p < 0.05$.
(D) Histograms showing the distribution of shFos representation (shFos/(shFos + shControl)) in liver metastases following intrasplenic injection of mixed populations. Dashed lines indicate pre-injection (blue) and post-injection (red) medians. Left, shFos #1 ($n = 32$ tumors, 3 mice); right, shFos #2 ($n = 56$ tumors, 5 mice). *p* values from one-sided one-sample Wilcoxon signed-rank tests.
(E) *In vitro* proliferation of KPC-2_HiB cells expressing *Fos*-targeting or control shRNAs. Points represent mean ± SEM from technical replicates ($n = 4$). Linear mixed-effects model; NS, not significant.
(F) Scatterplot depicting the relationship between *FOS* expression and a proliferation gene signature across tumor cells in the human PDAC scRNA-seq atlas. A regression line and Spearman's rho (0.10) are shown, indicating negligible correlation.
(G) Volcano plot of differential expression between *Fos*-knockdown and control conditions in KPC-2_HiB cells. Two independent *Fos*-targeting shRNAs and one non-targeting control shRNA were used, each with three technical replicates. Genes colored by significance (red, upregulated; blue, downregulated; black, unchanged) with FDR < 0.05.
(H) Heatmap of scaled, log-transformed, normalized expression for genes highlighted in (G) across technical replicates.
(I) Kaplan-Meier curve for overall survival stratified by *IL1A* expression (high = top 50%; low = bottom 50%) in the PACA-US cohort. *p* value from log rank test.
(J and K) CUT&RUN and ATAC-seq tracks in KPC-2_HiB cells at the *Il1a* (J) and *Il1rn* (K) loci. c-Fos CUT&RUN had two technical replicates; no-antibody CUT&RUN and ATAC-seq had one.

KPC-2_HiB cell lines. Consistent with its role as a pleiotropic TF, c-Fos depletion altered the expression of numerous genes: 802 were significantly upregulated, while 1,116 were downregulated (FDR cutoff <0.05; Table S6), spanning diverse biological processes. This outcome suggests that c-Fos may facilitate metastatic colonization through multiple parallel mechanisms. Regardless, its potential effect on the immune composition of the metastatic niche can be clearly identified: *Il11*, *Il34*, and *Il1a* were consistently downregulated across independent shRNAs (log2FC = −1.3, −0.28, and −0.24; FDR < 2.2E−16, 0.015, and 0.047, respectively; Figures 7G and 7H), whereas *Cxcl5*, *Il1rn*, *Il18bp*, and *Il33* were upregulated (log2FC = 0.93, 0.72, 0.52, and 0.18; FDR = 2.4E−7, 2.4E−14, 0.026, and 0.013, respectively). While the fold changes of some individual secreted factors are modest, the collective perturbation of the secretome is substantial and likely to exert strong effects on the tumor microenvironment. For example, upregulation of *Il1a*, encoding IL-1α, together with downregulation of *Il1rn*, encoding the IL-1 receptor antagonist (IL-1RA), would be expected to cooperate in enhancing IL-1 signaling. Consistent with this conclusion, we found that high *IL1A* expression in PDAC primary tumors was associated with worse overall survival in the PACA-US cohort (Figure 7I), in agreement with prior reports.[42]

We next used CUT&RUN (Cleavage Under Targets & Release Using Nuclease) profiling to map c-Fos binding genome-wide in KPC-2_HiB cells, identifying 18,477 binding sites (Table S7). Assigning each site to its nearest gene nominated 9,492 putative direct transcriptional targets. c-Fos can function as either a transcriptional activator or repressor.[43] By integrating CUT&RUN with RNA-seq from *Fos*-knockdown cells, we identified candidate genes under direct c-Fos transcriptional activation (n = 803) or repression (n = 526), enabling mechanistic assignment of these changes. Among the secreted factors discussed above, *Il34* and *Il1a* appeared to be directly activated, whereas *Cxcl5*, *Il1rn*, and *Il33* were directly repressed. For instance, strong c-Fos occupancy was detected at two open chromatin regions located ∼7.7 and ∼20.9 kb upstream of the *Il1a* transcription start site (TSS; Figure 7J), consistent with distal enhancers and supporting a direct regulatory role. At the *Il1rn* locus, two prominent binding sites were observed (Figure 7K), one at the promoter and the other at a distal enhancer ∼3.5 kb upstream of the TSS, consistent with direct transcriptional repression. These results suggest that much of the c-Fos effects on the secreted factors that would reprogram the metastatic niche are mediated by its direct binding as opposed to its indirect effects on other factors, positioning it as a critical upstream node for therapeutic targeting.

As a whole, these findings establish c-Fos as a reproducible functional driver of PDAC liver colonization, with consistent effects across independent subclones. They nominate c-Fos as a candidate therapeutic target and support a model in which c-Fos promotes colonization through a complex network of downstream targets, including multiple cytokines and cytokine antagonists.

DISCUSSION

In this study, we used DNA barcoding to uniquely label heterogeneous subclones in mouse primary PDAC tumors and measure their performance in liver-colonization assays, identifying a set of genes enriched in subclones with high metastatic potential and another set depleted in these subclones, defining a novel metastatic-potential axis. A prevailing theory of PDAC development positions highly metastatic subclones as being the most advanced along a normal-to-PDAC trajectory.[44] Our results do not support this model. Instead, we propose that the met-high and met-low states reflect a divergence at the endpoint of PDAC's developmental trajectory, with each subclone type achieving high fitness in the primary tumor but with met-high subclones uniquely equipped for navigating the post-extravasation steps of the metastasis cascade.

To assess metastatic potential in humans, we developed MetScore, a rank-based single-sample scoring metric derived from enrichment and depletion patterns of the met-high and met-low gene sets in transcriptomic data. Human PDAC metastases consistently exhibited higher MetScores than primary tumors, and high primary tumor MetScores were associated with worse overall survival independently of classical-basal subtype, suggesting the metastatic-potential axis is conserved between mouse and human PDAC. An unexpected finding was the conservation of the metastatic-potential axis among several diverse human carcinoma subtypes, particularly those that retain epithelial identity, such as CMS2 COAD. This may present an opportunity to more accurately stratify patients with stage II and III CMS2 COAD. Currently, decisions regarding whether to treat with adjuvant therapy and its duration in these patients are guided by clinical risk factors alone. Encouragingly, we observed a strong signal toward MetScore being predictive of adjuvant therapy benefit in stage II/III pMMR CMS2 COAD. Larger retrospective studies are required to validate the relationship between MetScore and adjuvant therapy benefit, explore interactions with clinical risk factors, and choose cutoffs for risk categories.

In addition to prognostic value, MetScore provided an opening to assess the mechanistic drivers of PDAC liver colonization. Across multiple patient cohorts, PDAC cells with a high MetScore occupied microenvironments enriched for immune cells and depleted for non-immune stromal cells. As this pattern was observed in metastases in addition to primary tumors, it suggests active remodeling by met-high cells rather than passive adaptation to pre-existing niches. This remodeling is likely mediated, at least in part, by tumor cell-derived cytokines and cytokine antagonists transcriptionally tuned by c-Fos. This positions c-Fos as a potential therapeutic target, acting as a critical upstream node that coordinates multiple effectors. c-Fos inhibitors are already in clinical use for other indications[45] but have not been evaluated in PDAC. While prior studies have implicated c-Fos in PDAC pathogenesis,[46,47] our work uniquely delineates its specific contribution to liver colonization. Future studies should define the downstream pathways by which c-Fos drives metastatic outgrowth and assess the therapeutic efficacy of c-Fos inhibition in preclinical PDAC models.

Overall, this work substantially expands the understanding of transcriptional heterogeneity in PDAC and identifies a metastatic-potential axis conserved across several carcinoma subtypes. It establishes MetScore as a robust biomarker capable

of prognostication and potentially predicting adjuvant treatment benefit. Moreover, it nominates c-Fos and other novel targets as actionable mediators of metastatic colonization, laying the groundwork for future therapeutic development aimed at intercepting metastasis.

Limitations of the study

While this study provides a framework for understanding metastatic-colonization potential in carcinomas, several technical and conceptual limitations should be acknowledged. First, all *in vivo* experiments were conducted in a single PDAC model (KPC) and exclusively in female mice, and our clonal analyses were based on a limited number of subclones. Nonetheless, the robust performance of MetScore across multiple patient datasets supports the broader relevance of transcriptional and chromatin-accessibility programs identified here. Second, while RNA-seq and ATAC-seq enabled characterization of transcriptomic and chromatin-accessibility differences between met-high and met-low subclones, these methods do not capture all layers of gene regulation. We did not assess DNA methylation, chromatin topology, or post-transcriptional mechanisms that may contribute to stable metastatic phenotypes. Third, our analysis of metastatic potential across human carcinoma subtypes was not exhaustive. While conservation of the metastatic-potential axis was clear across tested carcinomas, other carcinoma subtypes may deviate from this pattern. Similarly, we examined only one non-epithelial malignancy (melanoma), and future studies should investigate metastatic determinants in other non-epithelial solid tumors such as sarcomas. Fourth, the stromal populations we analyzed were defined broadly, yet each comprises multiple heterogeneous sub-states. Further work is needed to delineate the specific stromal subpopulations that define the met-high niche with greater granularity. Finally, our functional studies identifying c-Fos as a positive mediator of liver colonization utilized shRNAs, which may have off-target effects. However, the use of multiple independent shRNAs targeting each gene mitigated the risk of off-target effects generating misleading conclusions. While our data implicate c-Fos-driven cytokines and cytokine antagonists in microenvironmental remodeling, a direct causal link between these secreted factors and the establishment of a pro-metastatic niche remains to be established. In addition, genes enriched in met-low subclones, including neuroendocrine, Wnt signaling, and motility-associated programs, were not functionally evaluated here; whether they actively suppress colonization or are passive markers remains unknown.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Reza Kalhor (kalhor@jhu.edu).

Materials availability

Plasmids generated for stable expression of shRNAs are available on Addgene (shFos #1: 239618; shFos #2 239619; shLuciferase #3 239617). Cell lines generated in this study will be made available on request, but we may require a payment and a completed materials transfer agreement.

Data and code availability

- All ATAC-seq, RNA-seq, and CUT&RUN data are available from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under project ID PRJNA960830 and are publicly available as of the date of publication.
- No original code was used in this study.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Andrew Ewald for comments on the manuscript and Dr. WanYing "Jenny" Lin for technical advice. The authors would also like to acknowledge Drs. Elizabeth Jaffee, Dan Laheru, Andrew Ewald, Ralph Hruban, and Laura Wood for feedback on the project. Protocols and training for mouse surgeries were provided by Ms. Peyton Fair, Dr. Todd Armstrong, Mr. James Leatherman, and Dr. Daniel Zabransky in Dr. Elizabeth Jaffee's group and by Dr. Lei Zheng. The KPC-1/2 lines were generously provided by Dr. Lei Zheng. This work was supported by two Sol Goldman Pancreatic Cancer Pilot Project Grants (J.S.H. and R.K.), the National Institutes of Health (NIH) (R01HG012357 and U01HL156056, R.K.), and the David & Lucile Packard Foundation (2020–71380, R.K.). J.S.H. was supported by a McGlothlin Fellow-to-Faculty Award, a Conquer Cancer Foundation Young Investigator Award (2024YIA-7422622729), and a T32 Training Grant (5T32CA009071) from the NIH. R.K.D. and J.E.F. were supported by National Science Foundation (NSF) Graduate Research Fellowships. Parts of this work were carried out at the Advanced Research Computing at Hopkins (ARCH) core facility, which is supported by NSF grant number OAC 1920103. FFPE processing, sectioning, and H&E staining were carried out by the Johns Hopkins Oncology Tissue Services (OTS) core facility. ATAC-seq libraries for met-high and met-low subclones were prepared and sequenced in the Johns Hopkins Single Cell and Transcriptomics core facility. FACS was carried out at the Wilmer Eye Institute (EY001765) and at the Center for Cell Dynamics at Johns Hopkins.

AUTHOR CONTRIBUTIONS

J.S.H. and R.K. conceived the project. J.S.H. designed and carried out experiments and performed computational analyses. Z.L. helped generate barcoded cell lines. R.K.D. assisted with mouse surgeries. J.D.L. assisted with cell sorting. S.W. assisted with library preparation and RT-qPCR. J.E.F. assisted with generating *Fos*-knockdown cell lines. W.F., H.G., and E.J.F. provided critical feedback on bioinformatic analyses. J.S.H. and R.K. interpreted the data and wrote the manuscript. All authors commented on the manuscript. R.K. supervised the project.

DECLARATION OF INTERESTS

J.S.H. and R.K. are listed as co-inventors on a patent application on the use of this study's findings for clinical metastasis risk prediction. E.J.F. serves on the Scientific Advisory Board of Resistance Bio, as a consultant for Mestag Therapeutics and Merck, and receives research funding from AbbVie Inc. and Roche/Genentech outside the scope of this work. H.G. has outside interest as a co-founder of Exai Bio, Tahoe Therapeutics, and Therna Therapeutics; serves on the board of directors at Exai Bio; and is a scientific advisory board member for Verge Genomics and Deep Forest Biosciences, which are outside the scope of this work.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

The authors used ChatGPT-4o and ChatGPT-5 to edit text for improving clarity and flow. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANTS
  - Cell lines
  - Mouse models
- METHOD DETAILS
  - Barcoding
  - Barcode sequencing – Library preparation and sequencing
  - Barcode sequencing – data processing and analysis
  - Gross pathology and histology
  - Proliferation assays
  - Genotyping
  - Bulk RNA-seq – Library preparation
  - Bulk RNA-seq – data processing and analysis
  - ATAC-seq – Library preparation
  - ATAC-seq – data processing and analysis
  - Transcription factor footprinting
  - Engineering KPC-2_HiA cells to express GFP
  - Isolation of neoplastic cells from liver metastases
  - Scoring human tumor samples for MetScore and assigning molecular subtype
  - Dataset-specific filtering was applied where necessary
  - Single-cell RNA-seq analysis
  - Deconvolution of human tumor bulk transcriptomic profiles using ESTIMATE
  - Pooled shRNA screen and targeted *Fos* knockdown
  - Quantification of *Fos* knockdown by qRT-PCR
  - Characterization of genome-wide c-Fos binding sites using CUT&RUN
  - Data visualization and schematics
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2025.116701.

Received: May 6, 2025
Revised: August 24, 2025
Accepted: November 19, 2025

REFERENCES

1. Groot, V.P., Rezaee, N., Wu, W., Cameron, J.L., Fishman, E.K., Hruban, R.H., Weiss, M.J., Zheng, L., Wolfgang, C.L., and He, J. (2018). Patterns, Timing, and Predictors of Recurrence Following Pancreatectomy for Pancreatic Ductal Adenocarcinoma. Ann. Surg. *267*, 936–945. https://doi.org/10.1097/SLA.0000000000002234.
2. Liu, K.H., Hung, C.Y., Hsueh, S.W., Chang, P.H., Chen, Y.Y., Lu, C.H., Chen, P.T., Yeh, K.Y., Huang, P.W., Tsang, N.M., et al. (2019). Lung Metastases in Patients with Stage IV Pancreatic Cancer: Prevalence, Risk Factors, and Survival Impact. J. Clin. Med. *8*, 1402. https://doi.org/10.3390/jcm8091402.
3. Conroy, T., Hammel, P., Hebbar, M., Ben Abdelghani, M., Wei, A.C., Raoul, J.L., Choné, L., Francois, E., Artru, P., Biagi, J.J., et al. (2018). FOLFIRINOX or Gemcitabine as Adjuvant Therapy for Pancreatic Cancer. N. Engl. J. Med. *379*, 2395–2406. https://doi.org/10.1056/NEJMoa1809775.
4. Bailey, P., Chang, D.K., Nones, K., Johns, A.L., Patch, A.M., Gingras, M.C., Miller, D.K., Christ, A.N., Bruxner, T.J.C., Quinn, M.C., et al. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. Nature *531*, 47–52. https://doi.org/10.1038/nature16965.
5. Moffitt, R.A., Marayati, R., Flate, E.L., Volmar, K.E., Loeza, S.G.H., Hoadley, K.A., Rashid, N.U., Williams, L.A., Eaton, S.C., Chung, A.H., et al. (2015). Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat. Genet. *47*, 1168–1178. https://doi.org/10.1038/ng.3398.
6. Collisson, E.A., Sadanandam, A., Olson, P., Gibb, W.J., Truitt, M., Gu, S., Cooc, J., Weinkle, J., Kim, G.E., Jakkula, L., et al. (2011). Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat. Med. *17*, 500–503. https://doi.org/10.1038/nm.2344.
7. Oh, K., Yoo, Y.J., Torre-Healy, L.A., Rao, M., Fassler, D., Wang, P., Caponegro, M., Gao, M., Kim, J., Sasson, A., et al. (2023). Coordinated single-cell tumor microenvironment dynamics reinforce pancreatic cancer subtype. Nat. Commun. *14*, 5226. https://doi.org/10.1038/s41467-023-40895-6.
8. Raghavan, S., Winter, P.S., Navia, A.W., Williams, H.L., DenAdel, A., Lowder, K.E., Galvez-Reyes, J., Kalekar, R.L., Mulugeta, N., Kapner, K.S., et al. (2021). Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. Cell *184*, 6119–6137.e26. https://doi.org/10.1016/j.cell.2021.11.017.
9. Rashid, N.U., Peng, X.L., Jin, C., Moffitt, R.A., Volmar, K.E., Belt, B.A., Panni, R.Z., Nywening, T.M., Herrera, S.G., Moore, K.J., et al. (2020). Purity Independent Subtyping of Tumors (PurIST), A Clinically Robust, Single-sample Classifier for Tumor Subtyping in Pancreatic Cancer. Clin. Cancer Res. *26*, 82–92. https://doi.org/10.1158/1078-0432.CCR-19-1467.
10. O'Kane, G.M., Grünwald, B.T., Jang, G.H., Masoomian, M., Picardo, S., Grant, R.C., Denroche, R.E., Zhang, A., Wang, Y., Lam, B., et al. (2020). GATA6 Expression Distinguishes Classical and Basal-like Subtypes in Advanced Pancreatic Cancer. Clin. Cancer Res. *26*, 4901–4910. https://doi.org/10.1158/1078-0432.CCR-19-3724.
11. Hingorani, S.R., Wang, L., Multani, A.S., Combs, C., Deramaudt, T.B., Hruban, R.H., Rustgi, A.K., Chang, S., and Tuveson, D.A. (2005). Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. Cancer Cell *7*, 469–483. https://doi.org/10.1016/j.ccr.2005.04.023.
12. Foley, K., Rucki, A.A., Xiao, Q., Zhou, D., Leubner, A., Mo, G., Kleponis, J., Wu, A.A., Sharma, R., Jiang, Q., et al. (2015). Semaphorin 3D autocrine signaling mediates the metastatic role of annexin A2 in pancreatic cancer. Sci. Signal. *8*, ra77. https://doi.org/10.1126/scisignal.aaa5823.
13. Pan, X., Zhou, J., Xiao, Q., Fujiwara, K., Zhang, M., Mo, G., Gong, W., and Zheng, L. (2021). Cancer-associated fibroblast heterogeneity is associated with organ-specific metastasis in pancreatic ductal adenocarcinoma. J. Hematol. Oncol. *14*, 184. https://doi.org/10.1186/s13045-021-01203-1.
14. Kalhor, R., Kalhor, K., Mejia, L., Leeper, K., Graveline, A., Mali, P., and Church, G.M. (2018). Developmental barcoding of whole mouse via homing CRISPR. Science *361*, eaat9804. https://doi.org/10.1126/science.aat9804.
15. Leeper, K., Kalhor, K., Vernet, A., Graveline, A., Church, G.M., Mali, P., and Kalhor, R. (2021). Lineage barcoding in mice with homing CRISPR. Nat. Protoc. *16*, 2088–2108. https://doi.org/10.1038/s41596-020-00485-y.
16. Cheung, K.J., Padmanaban, V., Silvestri, V., Schipper, K., Cohen, J.D., Fairchild, A.N., Gorin, M.A., Verdone, J.E., Pienta, K.J., Bader, J.S., and Ewald, A.J. (2016). Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. Proc. Natl. Acad. Sci. USA *113*, E854–E863. https://doi.org/10.1073/pnas.1508541113.
17. Maddipati, R., and Stanger, B.Z. (2015). Pancreatic Cancer Metastases Harbor Evidence of Polyclonality. Cancer Discov. *5*, 1086–1097. https://doi.org/10.1158/2159-8290.CD-15-0120.
18. Bastiaenen, V.P., Klaver, C.E.L., van der Heijden, M.C.S., Nijman, L.E., Lecca, M.C., Tanis, P.J., Lenos, K.J., and Vermeulen, L. (2020). A mouse model for peritoneal metastases of colorectal origin recapitulates patient heterogeneity. Lab. Invest. *100*, 1465–1474. https://doi.org/10.1038/s41374-020-0448-x.
19. Baslan, T., Morris, J.P., 4th, Zhao, Z., Reyes, J., Ho, Y.J., Tsanov, K.M., Bermeo, J., Tian, S., Zhang, S., Askan, G., et al. (2022). Ordered and

deterministic cancer genome evolution after p53 loss. Nature *608*, 795–802. https://doi.org/10.1038/s41586-022-05082-5.

20. Struhl, K. (2024). The distinction between epigenetics and epigenomics. Trends Genet. *40*, 995–997. https://doi.org/10.1016/j.tig.2024.10.002.
21. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218. https://doi.org/10.1038/nmeth.2688.
22. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29. https://doi.org/10.1038/75556.
23. Gene Ontology Consortium; Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., Drabkin, H.J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N.L., and et al.. (2023). The Gene Ontology knowledgebase in 2023. Genetics *224*, iyad031. https://doi.org/10.1093/genetics/iyad031.
24. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30. https://doi.org/10.1093/nar/28.1.27.
25. Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust, A., Preussner, J., Kuenne, C., Braun, T., et al. (2020). ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat. Commun. *11*, 4267. https://doi.org/10.1038/s41467-020-18035-1.
26. Alonso-Curbelo, D., Ho, Y.J., Burdziak, C., Maag, J.L.V., Morris, J.P., 4th, Chandwani, R., Chen, H.A., Tsanov, K.M., Barriga, F.M., Luan, W., et al. (2021). A gene-environment-induced epigenetic program initiates tumorigenesis. Nature *590*, 642–648. https://doi.org/10.1038/s41586-020-03147-x.
27. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127. https://doi.org/10.1093/biostatistics/kxj037.
28. Bhuva, D.D., Foroutan, M., Xie, Y., Lyu, R., Cursons, J., and Davis, M.J. (2019). Using singscore to predict mutation status in acute myeloid leukemia from transcriptomic signatures. F1000Res. *8*, 776. https://doi.org/10.12688/f1000research.19236.3.
29. Aung, K.L., Fischer, S.E., Denroche, R.E., Jang, G.H., Dodd, A., Creighton, S., Southwood, B., Liang, S.B., Chadwick, D., Zhang, A., et al. (2018). Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. Clin. Cancer Res. *24*, 1344–1354. https://doi.org/10.1158/1078-0432.CCR-17-2994.
30. Jones, M.R., Williamson, L.M., Topham, J.T., Lee, M.K.C., Goytain, A., Ho, J., Denroche, R.E., Jang, G., Pleasance, E., Shen, Y., et al. (2019). NRG1 Gene Fusions Are Recurrent, Clinically Actionable Gene Rearrangements in KRAS Wild-Type Pancreatic Ductal Adenocarcinoma. Clin. Cancer Res. *25*, 4674–4681. https://doi.org/10.1158/1078-0432.CCR-19-0191.
31. Loveless, I.M., Kemp, S.B., Hartway, K.M., Mitchell, J.T., Wu, Y., Zwernik, S.D., Salas-Escabillas, D.J., Brender, S., George, M., Makinwa, Y., et al. (2025). Human Pancreatic Cancer Single-Cell Atlas Reveals Association of CXCL10+ Fibroblasts and Basal Subtype Tumor Cells. Clin. Cancer Res. *31*, 756–772. https://doi.org/10.1158/1078-0432.CCR-24-2183.
32. Cancer Genome Atlas Research Network Electronic address andrew_aguirre@dfci harvard edu; Cancer Genome Atlas Research Network (2017). Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. Cancer Cell *32*, 185–203.e13. https://doi.org/10.1016/j.ccell.2017.07.007.
33. Kamal, Y., Schmit, S.L., Hoehn, H.J., Amos, C.I., and Frost, H.R. (2019). Transcriptomic Differences between Primary Colorectal Adenocarcinomas and Distant Metastases Reveal Metastatic Colorectal Cancer Subtypes. Cancer Res. *79*, 4227–4241. https://doi.org/10.1158/0008-5472.CAN-18-3945.
34. Garcia-Recio, S., Hinoue, T., Wheeler, G.L., Kelly, B.J., Garrido-Castro, A.C., Pascual, T., De Cubas, A.A., Xia, Y., Felsheim, B.M., McClure, M.B., et al. (2023). Multiomics in primary and metastatic breast tumors from the AURORA US network finds microenvironment and epigenetic drivers of metastasis. Nat. Cancer *4*, 128–147. https://doi.org/10.1038/s43018-022-00491-x.
35. Institute, B. The Metastatic Breast Cancer Project. https://mbcproject.org/data-release.
36. Grasso, C.S., Wu, Y.M., Robinson, D.R., Cao, X., Dhanasekaran, S.M., Khan, A.P., Quist, M.J., Jing, X., Lonigro, R.J., Brenner, J.C., et al. (2012). The mutational landscape of lethal castration-resistant prostate cancer. Nature *487*, 239–243. https://doi.org/10.1038/nature11125.
37. Cancer Genome Atlas Network (2015). Genomic Classification of Cutaneous Melanoma. Cell *161*, 1681–1696. https://doi.org/10.1016/j.cell.2015.05.044.
38. Marisa, L., de Reyniès, A., Duval, A., Selves, J., Gaub, M.P., Vescovo, L., Etienne-Grimaldi, M.C., Schiappa, R., Guenot, D., Ayadi, M., et al. (2013). Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLoS Med. *10*, e1001453. https://doi.org/10.1371/journal.pmed.1001453.
39. Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352. https://doi.org/10.1038/nature10983.
40. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. *4*, 2612. https://doi.org/10.1038/ncomms3612.
41. Jiang, Z., Generoso, S.F., Badia, M., Payer, B., and Carey, L.B. (2021). A conserved expression signature predicts growth rate and reveals cell & lineage-specific differences. PLoS Comput. Biol. *17*, e1009582. https://doi.org/10.1371/journal.pcbi.1009582.
42. Gigante, L., Gaudilliere-Le Dain, G., Bertaut, A., Truntzer, C., and Ghiringhelli, F. (2024). Interleukin-1alpha as a Potential Prognostic Biomarker in Pancreatic Cancer. Biomedicines *12*, 1216. https://doi.org/10.3390/biomedicines12061216.
43. Gius, D., Cao, X.M., Rauscher, F.J., 3rd, Cohen, D.R., Curran, T., and Sukhatme, V.P. (1990). Transcriptional activation and repression by Fos are independent functions: the C terminus represses immediate-early gene expression via CArG elements. Mol. Cell Biol. *10*, 4243–4255. https://doi.org/10.1128/mcb.10.8.4243-4255.1990.
44. Burdziak, C., Alonso-Curbelo, D., Walle, T., Reyes, J., Barriga, F.M., Haviv, D., Xie, Y., Zhao, Z., Zhao, C.J., Chen, H.A., et al. (2023). Epigenetic plasticity cooperates with cell-cell interactions to direct pancreatic tumorigenesis. Science *380*, eadd5327. https://doi.org/10.1126/science.add5327.
45. Makino, H., Seki, S., Yahara, Y., Shiozawa, S., Aikawa, Y., Motomura, H., Nogami, M., Watanabe, K., Sainoh, T., Ito, H., et al. (2017). A selective inhibition of c-Fos/activator protein-1 as a potential therapeutic target for intervertebral disc degeneration and associated pain. Sci. Rep. *7*, 16983. https://doi.org/10.1038/s41598-017-17289-y.
46. Wakita, K., Ohyanagi, H., Yamamoto, K., Tokuhisa, T., and Saitoh, Y. (1992). Overexpression of c-Ki-ras and c-fos in human pancreatic carcinomas. Int. J. Pancreatol. *11*, 43–47. https://doi.org/10.1007/BF02925993.
47. You, L., Ren, X., Du, Y., Zhao, W., Cui, M., Chen, G., and Zhao, Y. (2016). c-Fos/ERK promotes the progression from pancreatic intraepithelial neoplasia to pancreatic ductal adenocarcinoma. Oncol. Rep. *36*, 3413–3420. https://doi.org/10.3892/or.2016.5169.
48. Torre-Healy, L.A., Kawalerski, R.R., Oh, K., Chrastecka, L., Peng, X.L., Aguirre, A.J., Rashid, N.U., Yeh, J.J., and Moffitt, R.A. (2023). Open-source curation of a pancreatic ductal adenocarcinoma gene expression analysis platform (pdacR) supports a two-subtype model. Commun. Biol. *6*, 163. https://doi.org/10.1038/s42003-023-04461-6.

49. Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat. Biotechnol. *38*, 675–678. https://doi.org/10.1038/s41587-020-0546-8.

50. Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210. https://doi.org/10.1093/nar/30.1.207.

51. Jain, E., Zañudo, J.G.T., McGillicuddy, M., Abravanel, D.L., Thomas, B.S., Kim, D., Balch, S., Navarro, J., Weiss, J.H., Hernandez, T.G., et al. (2023). The Metastatic Breast Cancer Project: leveraging patient-partnered research to expand the clinical and genomic landscape of metastatic breast cancer and accelerate discoveries. Preprint at medRxiv. https://doi.org/10.1101/2023.06.07.23291117.

52. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095.

53. de Bruijn, I., Kundra, R., Mastrogiacomo, B., Tran, T.N., Sikina, L., Mazor, T., Li, X., Ochoa, A., Zhao, G., Lai, B., et al. (2023). Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. Cancer Res. *83*, 3861–3867. https://doi.org/10.1158/0008-5472.CAN-23-0816.

54. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, pl1. https://doi.org/10.1126/scisignal.2004088.

55. Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330–337. https://doi.org/10.1038/nature11252.

56. Handler, J., Cullis, J., Avanzi, A., Vucic, E.A., and Bar-Sagi, D. (2018). Preneoplastic pancreas cells enter a partially mesenchymal state following transient TGF-beta exposure. Oncogene *37*, 4334–4342. https://doi.org/10.1038/s41388-018-0264-6.

57. Laboratory, T.J. Protocol 21298: Standard PCR Assay - Tg(Pdx1-cre)6Tuv. https://www.jax.org/Protocol?stockNumber=014647&protocolID=21298.

58. Lab, T.J. Kras G12D Conditional PCR. https://jacks-lab.mit.edu/protocols/genotyping/kras_cond.

59. Lab, T.J. p53 R172H, R270H, Ser23Ala. https://jacks-lab.mit.edu/protocols/genotyping/p53_recomb.html

60. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

61. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419. https://doi.org/10.1038/nmeth.4197.

62. Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. *47*, D766–D773. https://doi.org/10.1093/nar/gky955.

63. Love, M.I., Soneson, C., Hickey, P.F., Johnson, L.K., Pierce, N.T., Shepherd, L., Morgan, M., and Patro, R. (2020). Tximeta: Reference sequence checksums for provenance identification in RNA-seq. PLoS Comput. Biol. *16*, e1007664. https://doi.org/10.1371/journal.pcbi.1007664.

64. Martin, F.J., Amode, M.R., Aneja, A., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., et al. (2023). Ensembl 2023. Nucleic Acids Res. *51*, D933–D941. https://doi.org/10.1093/nar/gkac958.

65. Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinf. *14*, 7. https://doi.org/10.1186/1471-2105-14-7.

66. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

67. Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. *109*, 21.29.1–21.29.9. https://doi.org/10.1002/0471142727.mb2129s109.

68. Project, T.E. ENCODE ATAC-seq pipeline. https://github.com/ENCODE-DCC/atac-seq-pipeline.

69. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

70. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

71. Institute, B. Picard. https://broadinstitute.github.io/picard/.

72. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

73. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389–393. https://doi.org/10.1038/nature10730.

74. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

75. Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.

76. Wang, Q., Li, M., Wu, T., Zhan, L., Li, L., Chen, M., Xie, W., Xie, Z., Hu, E., Xu, S., and Yu, G. (2022). Exploring Epigenomic Datasets by ChIPseeker. Curr. Protoc. *2*, e585. https://doi.org/10.1002/cpz1.585.

77. ENCODE Project Consortium; Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., and et al.. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature *583*, 699–710. https://doi.org/10.1038/s41586-020-2493-4.

78. Rauluseviciute, I., Riudavets-Puig, R., Blanc-Mathieu, R., Castro-Mondragon, J.A., Ferenc, K., Kumar, V., Lemma, R.B., Lucas, J., Chèneby, J., Baranasic, D., et al. (2024). JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *52*, D174–D182. https://doi.org/10.1093/nar/gkad1059.

79. Korhonen, J., Martinmäki, P., Pizzi, C., Rastas, P., and Ukkonen, E. (2009). MOODS: fast search for position weight matrix matches in DNA sequences. Bioinformatics *25*, 3181–3182. https://doi.org/10.1093/bioinformatics/btp554.

80. Eide, P.W., Bruun, J., Lothe, R.A., and Sveen, A. (2017). CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. Sci. Rep. *7*, 16618. https://doi.org/10.1038/s41598-017-16747-x.

81. Pelossof, R., Fairchild, L., Huang, C.H., Widmer, C., Sreedharan, V.T., Sinha, N., Lai, D.Y., Guan, Y., Premsrirut, P.K., Tschaharganeh, D.F., et al. (2017). Prediction of potent shRNAs with a sequential classification algorithm. Nat. Biotechnol. *35*, 350–353. https://doi.org/10.1038/nbt.3807.

82. Fellmann, C., Hoffmann, T., Sridhar, V., Hopfgartner, B., Muhar, M., Roth, M., Lai, D.Y., Barbosa, I.A.M., Kwon, J.S., Guan, Y., et al. (2013). An optimized microRNA backbone for effective single-copy

RNAi. Cell Rep. *5*, 1704–1713. https://doi.org/10.1016/j.celrep.2013.11.020.

83. Papadopoulos, D., Ade, C.P., and Eilers, M. (2022). Generation of a pooled shRNA library for functional genomics screens. STAR Protoc. *3*, 101183. https://doi.org/10.1016/j.xpro.2022.101183.

84. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

85. Woo, B.J., Moussavi-Baygi, R., Karner, H., Karimzadeh, M., Garcia, K., Joshi, T., Yin, K., Navickas, A., Gilbert, L.A., Wang, B., et al. (2023). Integrative identification of non-coding regulatory regions driving metastatic prostate cancer. Preprint at bioRxiv. https://doi.org/10.1101/2023.04.14.535921.

86. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313.

87. Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. Use R (Springer), pp. 1–212. https://doi.org/10.1007/978-0-387-98141-3.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
c-Fos	Active Motif	61421; RRID: AB_2793628
Chemicals, peptides, and recombinant proteins		
RPMI 1640 Medium	Gibco	11875093
Fetal Bovine Serum	Thermo Fisher Scientific	16140071
Penicillin-Streptomycin	Gibco	15140122
Non-essential Amino Acids	Gibco	11140050
Sodium Pyruvate	Gibco	11360070
L-glutamine	Gibco	25030081
DMEM	Gibco	11995065
Basic Nucleofector® Kit for Primary Mammalian Epithelial Cells	Lonza	VPI-1005
Puromycin	InvivoGen	ant-pr-1
DMSO	Sigma-Aldrich	D2650
Propidium iodide	Invitrogen	00-6990-50
NEBNext Library Quant Kit for Illumina	New England Biolabs	E7630
MiSeq Reagent Kit v2	Illumina	MS-102-2002
SYBR Gold	Thermo Fisher Scientific	S11494
TRIzol	Thermo Fisher Scientific	15596026
Chloroform	Sigma-Aldrich	C2432
NEBNext Poly(A) mRNA Magnetic Isolation Module	New England Biolabs	E7490
xGen RNA Library Prep Kit	IDT	10009814
Lipofectamine 2000	Thermo Fisher Scientific	11668027
Neomycin	Corning	61-234-RF
Tumor Dissociation Kit, mouse	Miltenyi	130-096-730
ProtoScript II First Strand cDNA Synthesis Kit	New England Biolabs	E6560S
Forget-Me-Not EvaGreen qPCR Master Mix	Biotium	31045
Critical commercial assays		
e-Myco Mycoplasma PCR Detection Kit	Bulldog Bio	2523348
DNeasy Blood and Tissue Kit	Qiagen	69504
DNA Clean & Concentrator-5 Kit	Zymo	D4004
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Q32851
CyQUANT Cell Proliferation Assay	Thermo Fisher Scientific	C7026
RNA Clean & Concentrator-5 kit	Zymo	R1013
Deposited data		
All RNA-seq, ATAC-seq, and CUT&RUN data generated in this study	NCBI SRA	BioProject #PRJNA960830
ATAC-seq data from mouse pancreas pre-neoplasia and PDAC	NCBI SRA	BioProject #PRJNA548087
Mouse cCREs	SCREEN: Search Candidate cis-Regulatory Elements by ENCODE Registry of cCREs (screen.encodeproject.org)	Version 3
Non-redundant vertebrate transcription factor binding profiles	JASPAR (https://jaspar.elixir.no/downloads/)	2022 Version
PACA-CA[29,30]	pdacR[48]	ICGC PACA-CA, 2016—RNAseq
PACA-US[5]	pdacR	Moffit, 2015—MA

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TCGA-PAAD[32]	UCSC Xena Browser[49]	TCGA-PAAD
Colon-MCC[33]	Gene Expression Omnibus[50]	GSE131418
Colon-Consortium[33]	Gene Expression Omnibus	GSE131418
Breast-AURORA[34]	Gene Expression Omnibus	GSE193103
Breast-MBC[51]	cBioPortal[52–54]	The Metastatic Breast Cancer Project
Prostate-Michigan[36]	Gene Expression Omnibus	GSE35988
TCGA-SKCM[37]	UCSC Xena Browser	TCGA-SKCM
TCGA-COAD[55]	UCSC Xena Browser	TCGA-COAD
CIT-COAD[38]	Gene Expression Omnibus	GSE39582
METABRIC[39]	cBioPortal	METABRIC
PDAC scRNA-seq atlas[31]	Zenodo	14199536
Experimental models: Cell lines		
KPC-1	Dr. Lei Zhang	KPC-O
KPC-2	Dr. Lei Zhang	KPC-508F
HEK293T	ATCC	CRL-3216
Experimental models: Organisms/strains		
C57BL/6J	The Jackson Laboratory	000664
Oligonucleotides		
PCR primers	IDT	See Table S8
shRNA ultramers	IDT	See Table S9
xGen UDI indexing primers	IDT	10005922
Recombinant DNA		
PB-U6insert	Addgene	104536
PB-U6insert-EF1puro	Addgene	104537
Super PiggyBac Transposase Expression Vector	System Biosciences	PB210PA-1
pLenti CMV GFP *Neo*	Addgene	17447
pMD2.G	Addgene	12259
pCMV delta R8.2	Addgene	12263
SGEP	Addgene	111170
Software and algorithms		
MARC1 Pipeline	Kalhor Laboratory	https://github.com/Kalhor-Lab/MARC1-Pipeline
ATAC-seq Pipeline	ENCODE	https://github.com/ENCODE-DCC/atac-seq-pipeline
Cutadapt	Bioconda	Version 1.9.1
Bowtie2	Bioconda	Version 2.2.6 or 2.4.4 as noted in method details
Samtools	htslib.org	Version 1.7 or 1.15.1 as noted in method details
Picard	Broad Institute (https://broadinstitute.github.io/picard/)	Version 1.126
MACS2	PyPI (https://pypi.org/project/MACS2/)	Version 2.1.0 or 2.2.7.1 as noted in method details
DiffBind	Bioconductor	Version 3.14.1
ComBat	sva package (Bioconductor)	Version 3.52.0
R programming language	The R Project for Statistical Computing	Version 4.4.3
RStudio	posit	Version 2025.05.1 + 513
deepTools	Bioconda	Version 3.5.5

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ChIPseeker	Bioconductor	Version 1.40.1 or 1.42.1 as noted in method details
TOBIAS	Bioconda	Version 0.15.1
Trimmomatic	The Usadel Lab (http://www.usadellab.org/cms/?page=trimmomatic)	Version 0.39
Salmon	COMBINE lab (https://github.com/COMBINE-lab/salmon/releases)	Version 1.10.1
Tximeta	Bioconductor	Version 1.22.1
DESeq2	Bioconductor	Version 1.44.0
apeglm	Bioconductor	Version 1.28.0
PurIST	Rashid Lab (https://github.com/naimurashid/PurIST)	N/A
GSVA	Bioconductor	Version 2.0.7
singscore	Bioconductor	Version 1.24.0
limma	Bioconductor	Version 3.62.2
CMScaller	Oncosyne (https://github.com/peterawe/CMScaller)	Version 2.0.1
survival	CRAN	Version 3.5–8 or 3.8–3 as noted in method details
ggforitfy	CRAN	Version 0.4.17
meta	CRAN	Version 8.0–2
Seurat	CRAN	Version 5.2.0
caret	CRAN	Version 6.0–94
yardstick	CRAN	Version 1.3.2
ESTIMATE	MD Anderson Cancer Center (https://bioinformatics.mdanderson.org/estimate/rpackage.html)	Version 1.0.13
Bowtie	Bioconda	Version 1.3.0
bestNormalize	CRAN	Version 1.9.1
FastQC	Babraham Institute (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)	Version 0.12.1
Trim Galore	Babraham Institute (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)	Version 0.6.6
Gviz	Bioconductor	Version 1.50.0
ComplexHeatmap	Bioconductor	Version 2.20.0
ggplot2	CRAN	Version 3.5.1

EXPERIMENTAL MODEL AND STUDY PARTICIPANTS

Cell lines

The KPC-1 and KPC-2 cell lines were a gift from Dr. Lei Zheng. They were generated from primary tumors of KPC mice[11] (i.e., *Pdx1-Cre*;*LSL-Kras*$^{G12D/+}$;*Trp53*$^{R172H/+}$) as described previously.[12,13] Cell lines were tested for mycoplasma contamination using a PCR-based kit (Bulldog Bio) and were found to be mycoplasma negative. Their identity was confirmed by genotyping the *Pdx1-Cre*, *LSL-Kras*$^{G12D/+}$, and *Trp53*$^{R172H/+}$ loci (Figure S1). Both KPC-1 and KPC-2 were determined to be female via PCR amplification of the *Smcx* and *Smcy* genomic loci (data not shown). KPC cells were cultured in RPMI 1640 medium supplemented with FBS (10%), penicillin-streptomycin (100 U/mL), non-essential amino acids (1X), sodium pyruvate (1 mM), and L-glutamine (2 mM). HEK293T cells were purchased from ATCC and were cultured in DMEM supplemented with FBS (10%) and penicillin-streptomycin (100 U/mL). HEK293T cells are female.

Mouse models

All animal procedures were approved by Johns Hopkins University's Animal Care and Use Committee (ACUC) and conformed to the relevant local and national regulatory standards. C57BL/6J mice were obtained from The Jackson Laboratory (strain #000664). Eight-week-old female mice were used. Splenic, intraperitoneal, and orthotopic injections were performed as described previously.[12,13,18,56] For splenic and intraperitoneal injections, 500,000 cells were injected. For orthotopic injections, 50,000 cells were injected. In all cases, the injected animals were allowed to incubate for 4 weeks prior to sacrifice. The one exception was the pre- and post-liver colonization ATAC-seq experiment, where we performed splenic injections using 1M cells and sacrificed the mice after 2 weeks. Tumors were harvested in all cases with microdissection using a dissection microscope (ZEISS SteREO Discovery.V8).

METHOD DETAILS

Barcoding

The DNA barcodes utilized in this study were based on a previously published homing guide RNA (hgRNA) library.[14,15] Two transposable plasmid libraries with random bases were mixed. The first library (*Ins21*) contains two stretches of degenerate bases, one 15 bases in length ("LeftBarcode") and the other 10 bases in length ("RightBarcode"). These two stretches are separated by a constant 94 base region. It was constructed from the Addgene #104536 plasmid as described previously.[14] The second library (*Ins25-puro*) is very similar. It contains two stretches of degenerate bases, one 19 bases in length ("LeftBarcode") and the other 10 bases in length ("RightBarcode"), that are separated by a constant 94 base region; however, it also contains a puromycin resistance marker expressed by the EF-1α promoter. It was constructed from the Addgene #104537 plasmid as described previously.[14] The inserts in both plasmid libraries are flanked by PiggyBac inverted repeats, which enable their integration into the genome using the PiggyBac transposase. The constructs express the barcodes in small RNA form from a U6 promoter. The universal amplification primers for these barcodes allow reliable identification in sequencing based upon either the forward or reverse reads.

Nucleofection was used to introduce the barcode libraries into KPC cells (Lonza Basic Nucleofector Kit for Primary Mammalian Epithelial Cells and Nucleofector II Device using program T-020). Barcode libraries were co-transfected with Super PiggyBac Transposase Expression Vector to facilitate integration of the barcode construct into the genome. Two strategies were utilized to encourage a large number of barcode insertions per cell. The first strategy was co-transfection of the *Ins21* and *Ins25-puro* libraries in a 19:1 ratio. Under these conditions, only cells with a large number of integrations would be likely to have integrated a puromycin resistance gene-containing construct, allowing us to eliminate cells with few integrations during antibiotic selection. The second strategy was using a transposase:transposon ratio of 1:10 rather than the more typically used 1:3. Because PiggyBac can both integrate and excise transposons, having less PiggyBac in the cells reduces the likelihood that barcode constructs integrated into the genome will be removed during the initial transposition process.

Following barcode integration, the cells underwent puromycin selection for seven days. Puromycin-resistant cells were then sorted as single cells into wells of a 96-well plate using a Sony Sorter SH800. Propidium iodide was used to exclude dead cells. The resulting colonies were expanded over the course of several weeks while remaining under puromycin selection and then cryopreserved in fetal bovine serum with 10% DMSO.

Barcode sequencing – Library preparation and sequencing

Genomic DNA was isolated from cells or mouse tumors using a DNeasy Blood and Tissue kit as per the manufacturer's instructions. The hgRNA locus was amplified and sequenced using next-generation sequencing as described previously.[15] Briefly, the hgRNA locus was amplified using primers with overhangs containing primer binding sites for Illumina sequencing by synthesis (i.e., PCR1). Then, a second PCR amplification was performed using primers with overhangs containing either P5 or P7 to facilitate binding to the Illumina flow cell and a random DNA sequence (i.e., the i5 or i7 index sequence) to facilitate pooling and deconvolution of multiple samples in the same sequencing run (i.e., PCR2). Deviating from the published protocol, a set of four forward and four reverse degenerate PCR1 primers was used to increase library diversity, and custom PCR2 indexing primers were used to allow for pooling of a large number of samples (Table S8). The resulting libraries were then pooled, purified using a DNA Clean and Concentrator-5 kit, and quantified using either a Qubit dsDNA HS Assay Kit or an NEBNext Library Quant Kit for Illumina. The final libraries were sequenced using an Illumina MiSeq device and MiSeq Reagent Kit v2.

Barcode sequencing – data processing and analysis

Raw sequencing data were processed on a high-performance computing cluster (The Advanced Research Computing at Hopkins [ARCH] "Rockfish" cluster [https://www.arch.jhu.edu/about-arch/]) using the previously published MARC1 pipeline.[15] Briefly, this pipeline decompresses the raw sequencing data, compiles Read 1 and Read 2 sequences from each sample to a list of paired LeftBarcodes and RightBarcodes, sequentially corrects for sequencing errors in the LeftBarcode and RightBarcode regions, and then compiles complete lists of LeftBarcode-RightBarcode counts for each sample.

To characterize the barcodes specific to each subclone, local Python scripts were used to first filter out unique LeftBarcode-RightBarcode pairs with fewer than 3–4 reads, depending on sequencing depth. Then the RightBarcode sequences specific to that subclone were defined by filtering out RightBarcodes whose reads made up less than 1% of the total reads. Finally, each RightBarcode's LeftBarcode mate was defined as the most abundant LeftBarcode out of those paired with that RightBarcode.

To characterize the subclones present in a tumor or *in vitro* culture, local Python scripts were used to first filter out unique LeftBarcode-RightBarcode pairs with fewer than three reads. Then, the identifier-spacer pairs found in that tumor were cross-referenced against LeftBarcode-RightBarcode pairs specific to each subclone included in the experiment (i.e., either all KPC-1-derived subclones or all KPC-2-derived subclones). Since all LeftBarcode-RightBarcode pairs were unique to their assigned subclone, the presence of a single LeftBarcode-RightBarcode pair was sufficient to indicate the presence of its assigned subclone in the tumor. Using this strategy, each tumor was noted for the presence or absence of each subclone included in the experiment.

Gross pathology and histology

Photographs of mouse tumors were taken using an iPhone 13.

Mouse tumors were fixed in 10% neutral buffered formalin for 48 h, after which they were processed into paraffin-embedded tissue blocks as described previously.[56] They were then sectioned and stained with hematoxylin and eosin (H&E) as described previously.[56] Light micrographs of the H&E-stained sections were captured using a Zeiss Axio Vert.A1 microscope.

Proliferation assays

Cells were seeded into wells of a 96-well plate (5,000 or 10,000 cells, depending on the experiment). Relative cell number at each time point (24, 48, and 72 h) was quantified using a CyQUANT Cell Proliferation Assay, for cells in culture, as per the manufacturer's instructions. The resulting cell counts were normalized to the 24-h values. For the *Fos*-knockdown experiments, *pP*-values were calculated using a linear mixed-effects model testing the interaction between time and shRNA (*Fos*-targeting vs. control) with respect to cell number.

Genotyping

Genomic DNA was isolated from each KPC-1 and KPC-2 derived monoclonal line using a DNeasy Blood and Tissue kit as per the manufacturer's instructions. Genotyping PCR reactions were performed according to The Jackson Laboratory for *Pdx1-Cre*,[57] and the Tyler Jacks Laboratory for *Lsl-Kras*^{G12D}[58] and for *Lsl-Trp53*^{R17H}.[59] PCR products were separated on a 1% agarose gel run at 8V/cm for one hour and visualized with SYBR Gold reagent as per the manufacturer's recommendations.

Bulk RNA-seq – Library preparation

For transcriptomic profiling of met-high and met-low monoclonal KPC cell lines and for characterization of gene expression changes downstream of *Fos* knockdown, cells grown *in vitro* were lysed in TRIzol reagent, followed by chloroform extraction and cleanup using the RNA Clean and Concentrator-5 Kit according to the manufacturer's instructions.

For the met-high vs. met-low comparison, mRNA was enriched with the NEBNext Poly(A) mRNA Magnetic Isolation Module. Libraries were prepared using the xGen RNA Library Prep Kit and xGen UDI indexing primers.

For *Fos* knockdown samples, total RNA was shipped to Azenta Life Sciences, where libraries were prepared using their standard poly(A)-selected RNA-seq protocol.

All libraries were sequenced on an Illumina NovaSeq 6000 platform with 150 bp paired-end reads.

Bulk RNA-seq – data processing and analysis

Raw RNA-seq data were processed on a high-performance computing cluster (the Rockfish cluster described in **Barcode sequencing – data processing and analysis**). Adapter sequences were trimmed using Trimmomatic[60] 0.39. Transcripts were then quantified using Salmon[61] 1.10.1 in the mapping-based mode with GC bias correction. The M33 (GRCm39) transcript sequences from GENCODE[62] were used as the reference transcriptome.

Then, locally in R, transcript quantifications were imported and summarized to the gene level using Tximeta[63] 1.22.1.

For the metastasis-high vs. metastasis-low comparison, genes with fewer than 50 reads across all KPC samples were first excluded. Then, the generalized linear model functionality of DESeq2 1.44.0 was utilized to identify differentially expressed genes between metastasis-high and metastasis-low subclones while controlling for parental group status by modeling the metastatic potential (i.e., high vs. low) as a fixed effect and parental group (i.e., KPC-1 vs. KPC-2) as a random effect (design: ~parental_group + metastatic_potential). An FDR cutoff of 0.05 was used.

For the *Fos*-knockdown versus control comparison, genes with fewer than 50 reads across all samples were excluded. Differential expression analysis was performed using DESeq2 (v1.44.0) with a design formula ~ `Knockdown`, pooling replicates from both shRNAs. Log_2 fold changes were shrunken using apeglm (v1.28.0), and significance was determined at an FDR <0.05. For secreted factors highlighted in the main text, we additionally verified that expression changes were consistent across both independent shRNAs (Figure 7H).

To classify the KPC subclones as being either classical or basal subtype, the PurIST classifier[9] was used to calculate the probability of basal subtype classification for each sample. This was performed in R by following the publisher's instructions. The human classical and basal subtype-defining genes used by the classifier were converted to their mouse orthologs using the Ensembl[64] BioMart web portal. Bulk RNA-seq data from each subclone were used as input to the classifier, specifically gene TPM.

To compare expression of classical and basal marker genes between met-high and met-low subclones, single sample gene set enrichment analysis (ssGSEA) was performed using the GSVA package[65] version 2.0.7 using $log_2(x+1)$ TPM as input. Classical

and basal marker genes were defined as the top 25 gene exemplars for the subtype factors from Moffitt et al., 2015[5] restricted to those with well-characterized mouse orthologs. The final gene sets used are listed in Figure 3D. Differences in ssGSEA enrichment score and average TPM between met-high and met-low subclones were assessed using two-sided Wilcoxon rank-sum tests.

Enrichment of GO[22,23] and KEGG[24] (Kyoto Encyclopedia of Genes and Genomes) pathways amongst metastasis-high and metastasis-low genes was assessed in R using the `enrichGO()` and `enrichKEGG()` functions in clusterProfiler[66] 4.12.0. An FDR cutoff of 0.05 was used.

ATAC-seq – Library preparation

ATAC-seq libraries were generated as described previously.[67] Libraries were then pooled and sequenced on a NovaSeq 6000 (met-high and met-low subclones) or NovaSeq X Plus (pre- and post-colonization experiment) using 100 bp paired-end reads.

ATAC-seq – data processing and analysis

Raw sequencing data were processed on a high-performance computing cluster (the Rockfish cluster described in **Barcode sequencing – data processing and analysis**) using the ENCODE project's publicly available ATAC-seq pipeline[68] with default settings. Briefly, adapters were trimmed using cutadapt 1.9.1, trimmed reads were aligned to the mm10 genome using Bowtie2[69] 2.2.6, low-quality, mitochondrial, and duplicate reads were filtered out using Samtools[70] 1.7 with Picard[71] 1.126 to mark duplicates, and peaks were called using Macs2[72] 2.1.0. For two samples, KPC-2_LoC and KPC-2_HiA, two independent libraries were prepared from separate aliquots of cryopreserved cells (i.e., technical replicates). Raw sequencing data from technical replicates were pooled before processing.

ATAC-seq data from normal, pre-neoplasia, pancreatitis, pre-neoplasia+pancreatitis, and primary PDAC generated by Alonso-Curbelo and colleagues[26] were downloaded from NCBI (BioProject #PRJNA548087) and processed in the same manner.

Locally in R, a consensus peak set was generated by merging overlapping peaks from the individual samples' peak sets using DiffBind[73] 3.14.1. Then, a normalized count matrix, i.e., the number of reads aligning to each peak, normalized by the total reads in peaks for that sample, across all samples, was generated using DiffBind.

For analyses that included both the metastasis-high and metastasis-low subclones generated in this study, as well as the normal, pre-neoplasia, pancreatitis, pre-neoplasia+pancreatitis, and primary PDAC generated by Alonso-Curbelo and colleagues,[26] the normalized count matrix was batch corrected using ComBat[27] from the sva package (version 3.52.0). A parametric batch correction was performed with sample type (i.e., normal, pancreatitis, pre-neoplasia, pre-neoplasia+pancreatitis, PDAC) included as a covariate and with the Alonso-Curbelo et al. samples defined as the reference batch.

Principal component analysis was performed using the $\log_2(x+1)$ transformed and scaled normalized count matrix using base R's `prcomp()` function.

The generalized linear model functionality of DESeq2[74] was utilized to identify differentially accessible peaks between metastasis-high and metastasis-low subclones while controlling for parental group status by modeling the metastatic potential (i.e., high vs. low) as a fixed effect and parental group (i.e., KPC-1 vs. KPC-2) as a random effect (design: ~parental_group + metastatic_potential). In the pre-versus post-colonization experiment, DESeq2 was used to perform a straightforward comparison between the two conditions (design: ~condition). An FDR cutoff of 0.05 was used in both cases.

To visualize ATAC-seq signal profiles for peaks with increased accessibility in metastasis-high subclones and separately for peaks with increased accessibility in metastasis-low subclones across all of the tested subclones, a matrix containing scores for the genomic regions of interest was first generated using the `plotProfile()` function of DiffBind 3.14.1, and then plots were generated using the `plotProfile` command of deepTools[75] 3.5.5.

Significantly differentially accessible peaks were assigned to their nearest genes using the `annotatePeak()` function from ChIPseeker[76] 1.40.0 with the mm10 genome as the reference genome. We excluded 124 peaks assigned to predicted genes lacking an Ensembl ID and 2,640 peaks linked to genes that were lowly expressed in both met-high and met-low subclones.

Peaks were assessed for overlap with candidate *cis*-regulatory elements (cCREs). All mouse cCREs (mm10 genome) identified by the ENCODE project were downloaded from SCREEN: Search Candidate cis-Regulatory Elements by ENCODE Registry of cCREs V3.[77] Overlaps were broken down by category. For visualization purposes, "pELS" and "dELS" categories were merged into "Enhancer"; "PLS" and "CA-H3K4me3" categories were merged into "Promoter"; and "CA", "CA-TF", and "TF" were merged into "Candidate RE, NOS".

Transcription factor footprinting

Non-redundant vertebrate transcription factor binding profiles, i.e., motifs, were downloaded from the JASPAR database.[78] Locally in R, this set of motifs was then filtered to exclude motifs corresponding to non-expressed or lowly expressed transcription factors in the KPC cells. Non-expressed genes were defined as those having fewer than 50 total reads across all samples in our RNA-seq dataset, and lowly expressed genes were defined as those whose statistical significance was not calculated by DESeq2 due to low expression in our differential expression analysis.

Then, on a high-performance computing cluster (the Rockfish cluster described in **Barcode sequencing – data processing and analysis**), aligned ATAC-seq reads for every KPC sample were downsampled to the coverage of the least covered sample using the

`view` command in samtools 1.15.1. Then, aligned ATAC-seq reads were merged amongst the metastasis-high samples and separately amongst the metastasis-low samples using the `merge` command in samtools 1.15.1.

Finally, on the Rockfish cluster, transcription factor footprinting was performed by inputting the filtered motifs (n = 511), the metastasis-high and metastasis-low merged aligned ATAC-seq reads, and the complete consensus peak set for all KPC samples (n = 176,964 peaks) to TOBIAS[25] 0.15.1. The `ATACorrect` command was used to correct the ATAC-seq signal in the merged samples for Tn5 insertion bias. Then, the `ScoreBigwig` command was used to calculate a continuous footprint score across peaks in the consensus peak set based on the depletion of signal and the general accessibility of the nearby region. Finally, the `BINDetect` command was used to (1) identify putative TF binding sites within peak regions by matching the known motifs to the peak region DNA sequences using MOODS[79] (MOtif Occurrence Detection Suite); (2) classify each putative TF binding site as being either bound or unbound in each condition (i.e., metastasis-high and metastasis-low) based on a footprint score cutoff; (3) calculate the log_2 fold change in footprint score between the two conditions for each binding site; (4) calculate a differential binding score (DBS) for each motif representing the global distribution of log_2 fold changes across binding sites for that motif; and (5) calculate a *p-value* for each motif by comparing its DBS to 100 DBSs generated from randomly sampled log_2 fold changes from the background distribution.

Locally in R, differential binding score *p-values* were adjusted for multiple hypothesis testing using a Bonferroni adjustment. An adjusted *p-value* cutoff of 0.05 was used.

Engineering KPC-2_HiA cells to express GFP

pLenti CMV GFP *Neo* was packaged into lentiviral particles via co-transfection with VSV.G and gag/pol plasmids into HEK293T cells using Lipofectamine 2000 Transfection Reagent. Lentiviral particles were concentrated via precipitation with lentiviral concentration solution (4X stock is 40% [W/V] PEG-8000 and 1.2M NaCl in PBS [pH 7]). KPC-2_HiA cells were transduced in the presence of polybrene using a low multiplicity of infection. Transduced cells were selected via exposure to medium containing neomycin for 7 days.

Isolation of neoplastic cells from liver metastases

Tumor-laden mouse livers were dissociated into single cells using a Tumor Dissociation Kit, mouse as per the manufacturer's instructions using the 37C_m_TDK_2 program. GFP-positive cells were isolated using a Sony Sorter SH800S. Dead cells were excluded using propidium iodide.

Scoring human tumor samples for MetScore and assigning molecular subtype

Bulk transcriptomic data (normalized microarray signals or RNA-seq counts) from human primary tumors and metastases were obtained (see key resources table).

For microarray datasets with multiple probes per gene (Colon-MCC, Colon-Consortium, Prostate-Michigan, CIT-COAD, METABRIC), the probe with the highest average signal was retained. In RNA-seq datasets containing multiple entries per gene (e.g., Breast-MBC), counts were summed across entries to yield a single value per gene.

Dataset-specific filtering was applied where necessary

- TCGA-PAAD: Neuroendocrine tumors excluded.
- TCGA-COAD and CIT-COAD: Excluded mismatch repair-deficient or unannotated tumors.
- METABRIC: Restricted to patients with invasive ductal carcinoma.

Lowly expressed genes were excluded (average FPKM <1 for TCGA datasets; average TPM <1 for PACA-CA, Breast-AURORA, and Breast-MBC).

Mouse met-high (n = 207) and met-low (n = 182) genes were converted to human orthologs using Ensembl[64] BioMart, yielding 202 met-high and 174 met-low genes. The same procedure was used to generate the broader MetScore^{RNAonly} gene sets (433 high, 403 low).

MetScore or MetScore^{RNAonly} were calculated using the singscore[28] R package (v1.24.0). Genes were ranked with `rankGenes()`, followed by scoring via `singscore()` using human met-high genes as the up-set and met-low genes as the down-set. Only genes detected after dataset-specific filtering were included in scoring.

Comparisons of MetScore or MetScore^{RNAonly} between primary tumors and metastases were performed using two-sided Wilcoxon rank-sum tests. For PACA-US rapid autopsy cases with matched primary and metastatic samples, a linear mixed-effects model (anatomic site as fixed effect, patient as random effect) was applied.

Differential expression analyses in PACA-US and PACA-CA were conducted using the limma R package (v3.62.2) with Benjamini–Hochberg correction (FDR <0.05).

For subtype analyses, PACA-US patients were stratified by classical–basal subtype per original annotations, and MetScore comparisons were performed within subtypes using Wilcoxon rank-sum tests.

For survival analyses, localized or locally advanced primary tumors from PACA-US, PACA-CA, TCGA-PAAD, TCGA-COAD, CIT-COAD, and METABRIC were stratified into high and low MetScore groups (top vs. bottom 50%). For TCGA-COAD and CIT-COAD, patients were further stratified by CMS subtype using the CMScaller package[80] (version 2.0.1). For METABRIC, patients were further

stratified by PAM50 subtype using the authors' original annotations. Kaplan–Meier curves were generated using the survival (version 3.5–8) and ggfortify (version 0.4.17) packages; significance was assessed via log rank tests (*p-value* <0.05).

To assess independent prognostic value, PurIST was applied for classical–basal subtype assignment in PDAC cohorts. Multivariable Cox regression was performed using `coxph()` (survival version 3.8–3). Pooled hazard ratios were calculated using a fixed-effects meta-analysis model based on inverse-variance weighting of log-transformed hazard ratios using the meta R package (version 8.0–2).

To assess predictive power for adjuvant chemotherapy benefit, patients with stage II and III pMMR CMS2 were pooled from the TCGA-COAD and CIT-COAD cohorts. Multivariable Cox regression was performed using `coxph()` (survival version 3.8–3) with MetScore, overall survival, and adjuvant therapy (yes vs. no) as covariates.

Single-cell RNA-seq analysis

Pre-normalised UMI counts together with cell-level metadata for the human PDAC single-cell atlas[31] were downloaded. All analyses were carried out in R 4.4.3 on the Rockfish high-performance computing cluster (described under **Barcode sequencing – data processing and analysis**). The expression matrix was loaded as a Seurat v5.2.0 object.

Tumor-enriched epithelial cells were identified as follows. Cells labelled by the original authors as "DUCTAL" or "CYCLING DUCTAL" were isolated. Variable features were re-identified (`vst`, 2,000 genes), the data were re-scaled, principal-component analysis (`RunPCA`), neighbor graph construction (`FindNeighbors`, dims = 1–10), Louvain clustering (`FindClusters`, resolution = 0.011), and UMAP visualization (`RunUMAP`, dims = 1–10) were repeated. This yielded four clusters, hereafter termed Ductal 1–4.

Cluster Ductal 2 was highly enriched for cells derived from healthy donors or adjacent normal tissue and was inferred to represent non-malignant epithelium. We therefore removed (i) all Ductal 2 cells and (ii) any remaining healthy-donor/adjacent-normal cells, leaving 243,171 tumor-enriched epithelial cells. These were reclustered with the same pipeline (resolution = 0.05), producing the seven tumor clusters discussed in the manuscript.

For each tumor-enriched epithelial cell, we calculated (i) MetScore (method described in **Scoring human tumor samples for MetScore and assigning molecular subtype**, above), (ii) the previously published scBasal/scClassical commitment score (scB/scC),[8] and (iii) a gene signature that predicts cell and lineage-specific differences in growth rate[41] using Seurat's `AddModuleScore`. Violins and UMAP overlays were generated with `FeaturePlot`/`VlnPlot`; expression values were winsorised at the 10th and 90th percentiles for display only. Differences between primary tumor- and metastasis-derived tumor cells were calculated using a generalized linear model with donor type as a fixed effect and donor as a random effect. Cell-wise MetScore × scB/scC correlation and *FOS* × proliferation used Pearson's r.

Binary logistic-regression models were fitted with caret v6.0-94 (method = "glm", family = binomial, classProbs = TRUE, savePredictions = "final"). Cells were stratified by donor type (primary tumor or metastasis); 10-fold cross-validation (CV) was applied. The no-information rate (NIR, 88.3% primary tumor-derived) served as a naive baseline. For each model, the proportion of correct CV predictions was compared with the NIR by a one-sided binomial test (`confusionMatrix`, caret). Precision-recall curves and AUPRC were calculated with yardstick v1.3.2 (event_level = "second"), using the out-of-fold predicted probabilities; baseline precision (random classifier) is shown as a dashed line at 0.117 (metastatic prevalence).

Deconvolution of human tumor bulk transcriptomic profiles using ESTIMATE

To assess stromal composition across tumors, we applied the ESTIMATE algorithm[40] (R package estimate version 1.0.13) to normalized expression matrices. For each dataset, we reformatted the expression matrix to include HGNC gene symbols as the first column and sample identifiers as column headers, and exported the matrix as a tab-delimited text file. Common genes were filtered using `filterCommonGenes()`, and stromal, immune, and ESTIMATE scores were computed using `estimateScore()` with the appropriate platform specification (e.g., `''affymetrix''`).

To quantify the relative balance between immune and non-immune stromal content, we computed z-scores for both the ImmuneScore and StromalScore across all tumors in a given cohort, then subtracted the StromalScore *Z* score from the ImmuneScore *Z* score to generate an `Immune – Stroma (z-score)`. This composite metric captures the relative enrichment of immune versus fibrovascular stromal components in a normalized, cohort-wide manner. Higher values indicate a microenvironment skewed toward immune cell content relative to non-immune stroma (e.g., cancer-associated fibroblasts and endothelial cells).

Pooled shRNA screen and targeted *Fos* knockdown

We designed three miR-30-based shRNAs targeting each gene included in the screen, as well as three negative control shRNAs targeting luciferase, which the KPC cells do not express. The majority of the shRNA sequences used were designed using the SplashRNA algorithm.[81] For genes for which the SplashRNA algorithm did not produce at least three shRNA sequences with SplashRNA scores greater than one, the remaining sequences were selected from the shRNAs in Table 3 in Fellmann et al., 2013.[82] All shRNA sequences used in this study can be found in Table S9.

shRNAs were cloned in a pooled fashion or individually into the SGEP lentiviral expression vector as described previously.[83] The resulting shRNA library or single shRNA-containing plasmids were packaged into lentiviral particles as described above

(see engineering KPC-2_HiA cells to express GFP). KPC-1_Hi2 or KPC-2_HiB cells were transduced in the presence of polybrene using a low multiplicity of infection (< 0.3). GFP-positive cells were isolated using a Sony Sorter SH800S. Dead cells were excluded using propidium iodide.

Genomic DNA was isolated from a day zero pre-injection sample and mouse tumors using a DNeasy Blood and Tissue Kit according to the manufacturer's instructions. When tumors were too large to be digested and loaded onto a single spin column, they were subdivided into smaller chunks, each of which was digested and library prepped separately. Independent draws of the day zero pre-injection sample were library prepped separately (n = 3 for the pooled screen and n = 4 for the *Fos* knockdown experiments). The shRNA locus was amplified using primers flanking the entire shRNA sequence with overhangs containing primer binding sites for Illumina sequencing by synthesis (i.e., PCR1; Table S8). Then, a second PCR amplification was performed using primers with overhangs containing either P5 or P7 to facilitate binding to the Illumina flow cell and a random DNA sequence (i.e., the i5 or i7 index sequence) to facilitate pooling and deconvolution of multiple samples in the same sequencing run (i.e., PCR2; Table S8). The resulting libraries were then pooled, purified using a DNA Clean and Concentrator-5 Kit, and quantified using a NEBNext Library Quant Kit for Illumina. The final libraries were sequenced using an Illumina MiSeq device and MiSeq Reagent Kit v2.

First, reads from each sample were mapped to the expected shRNA sequences present in the library. To accomplish this, Bowtie[84] 1.3.0 was used to align the region of Read1 expected to correspond to the variable region of the shRNA-guide stem against a reference composed of the variable regions present in the library. One base pair mismatches were allowed to account for sequencing error. For the single knockdown experiments, gaps one nucleotide in length were also allowed. This produced a list of counts for each shRNA in the experiment for each sample. This analysis was performed on a high-performance computing cluster (the Rockfish cluster described in **Barcode sequencing – data processing and analysis**).

For the *Fos* knockdown experiments, samples with fewer than 100 total reads aligning to expected shRNA sequences were excluded. Then, for each sample, the Fos-targeting fraction was calculated as shFos reads/(shFos reads + shControl reads). The mean fraction in day zero replicates served as the expected value. A one-sided Wilcoxon signed-rank test evaluated whether post-injection fractions were significantly lower (depletion of shFos).

For the pooled screen, shRNA counts for the three-day zero pre-injection samples were combined. shRNAs whose abundance in the pooled day zero pre-injection sample was less than 0.25% were excluded from further analysis steps. Tumors with fewer than 100 shRNA counts were also excluded from further analysis steps.

The tumor samples were then normalized by dividing counts for each shRNA by the total shRNA counts for that respective sample. Normalized shRNA counts in samples derived from the same tumor (i.e., cases in which the tumor was too large to be digested and loaded onto a single spin column) were combined, and the resulting pooled samples were re-normalized by dividing the normalized counts for each shRNA by the total normalized counts for that pooled sample.

Enrichment/depletion of shRNAs between liver metastases and primary tumors was evaluated as described previously.[85] Briefly, normalized shRNA counts were subjected to linear regression modeling with the sample type (i.e., primary tumor vs. liver metastasis) as a covariate using the `lm()` function in R. The resulting coefficients, standard errors (*SE*), *t*-values, and *pp-value*s were extracted for each shRNA. To account for variability and normalize the data, log_2 fold changes of the non-targeting control shRNAs were normalized using the bestNormalize package (version 1.9.1) to create a transformation object. This transformation was then applied to the log_2 fold changes of targeting shRNAs to calculate *z*-scores.

For weighted combined *p-value* calculation, *z*-scores were divided by the *SE*, and weights were defined as the inverse of the *SE* squared. Subsequently, shRNAs targeting the same gene were aggregated, and the adjusted *z*-scores were summed and normalized by the square root of the sum of the weights to generate combined *z*-scores for each gene. The combined *z*-scores were then used to calculate *p-value*s by applying a two-tailed normal distribution test (`2*pnorm(-abs(combined.z))`). A *p-value* cutoff of 0.05 was used.

Quantification of *Fos* knockdown by qRT-PCR

RNA was extracted using a TRizol–chloroform extraction followed by column cleanup using an RNA Clean and Concentrator-5 Kit as per the manufacturer's instructions. RNA was then reverse transcribed to cDNA using a ProtoScript II First Strand cDNA Synthesis Kit as per the manufacturer's instructions. Quantitative PCR was performed using Forget-Me-Not EvaGreen qPCR Master Mix on a QuantStudio 3 device (Thermo Fisher). See Table S8 for primer sequences. Data was analyzed using the standard curve method. *Rps29* was used as a housekeeping gene. Differences between experimental and control groups were calculated using one-sided Wilcoxon rank-sum tests.

Characterization of genome-wide c-Fos binding sites using CUT&RUN

Cryopreserved cells were shipped to Active Motif (Carlsbad, CA) for CUT&RUN. Briefly, nuclei were immobilized on Concanavalin A-coated magnetic beads and incubated overnight at 4°C with 1 μL (~1 μg) of c-Fos antibody. After washing to remove excess antibody, pMNase was added and activated with $CaCl_2$ for 2 h at 4°C to digest DNA fragments at antibody-bound sites. Released fragments were purified using DNA columns, and sequencing libraries were prepared with the NEBNext DNA Library Prep Kit (New England BioLabs) according to the manufacturer's instructions. Libraries were sequenced on an Illumina NextSeq 2000 platform (38 bp paired-end reads).

Basic data processing steps were performed by Active Motif. Sequencing reads were processed using the nf-core/cutandrun pipeline. Quality metrics were assessed with FastQC (version 0.12.1). Adapter trimming was performed with Trim Galore (version 0.6.6). Trimmed reads were aligned to the mm10 reference genome with Bowtie2 (version 2.4.4), and SAM files were converted to BAM format using Samtools. Duplicate reads were removed with Picard `MarkDuplicates`, mitochondrial reads were excluded, and alignments overlapping ENCODE blacklisted regions were discarded. Peaks were called with MACS2 (version 2.2.7.1). The fraction of reads in peaks (FRiP) was calculated as a measure of dataset quality. BigWig files were generated with deepTools using counts per million (CPM) normalization for visualization.

Downstream analysis was performed locally in R. Consensus peaks were generated from technical replicates using the `dba.count` function of DiffBind (version 3.16.0) with summits extended to 150 bp. Peaks were assigned to their nearest gene using the `annotatePeaks()` function in ChIPseeker (version 1.42.1).

Signal tracks were generated using the Gviz package (version 1.50.0) in R.

Data visualization and schematics

Heatmaps were generated in R using ComplexHeatmap[86] 2.20.0. Unless otherwise noted, all other plots were generated in R using ggplot2[87] 3.5.1. Schematics were created using BioRender.com.

QUANTIFICATION AND STATISTICAL ANALYSIS

The nature and number of experimental replicates utilized are described in the respective figure legends. The quantitative and statistical methods utilized, including filtering criteria, statistical tests used, and significance cutoffs, can be found in method details (in full) and the figure legends (in brief).