nature genetics

Article

https://doi.org/10.1038/s41588-026-02714-8

Developmental and MAPK-responsive transcription factors regulate distinct malignant cell states and associated genetic dependencies in pancreatic cancer

Received: 8 March 2024

Accepted: 14 July 2026

Published online: 26 August 2026

Check for updates

A list of authors and their affiliations appears at the end of the paper

There is broad consensus that the malignant epithelial cells of human pancreatic ductal adenocarcinoma (PDA) comprise multiple, molecularly distinct states. Yet precise characterization of how these are regulated—including their mechanistic determinants, dependencies, plasticity and functional properties—remains elusive. Single-cell master regulator (MR) analysis of multiple PDA cohorts identified malignant cells in three co-existing, molecularly distinct developmental lineage states, with distinct histopathological morphologies and spatial architecture. These include a poorly differentiated lineage driven by epithelial–mesenchymal-transition-related MRs and two well-differentiated states driven by gastrointestinal epithelial development and pancreatic development MRs, respectively. Furthermore, each state comprises two epigenetically distinct substates with low versus high MAPK signaling activity. Barcode-based lineage tracing confirmed both spontaneous and treatment-dependent cross-state plasticity. Furthermore, loss-of-function studies confirmed state-specific MR essentiality, while their ectopic expression effectively reprogrammed cell state, in vitro and in vivo, thus providing a mechanism-based foundation for PDA heterogeneity and a roadmap for pharmacological targeting.

Pancreatic ductal adenocarcinoma (PDA) is the third-leading cause of cancer-related mortality and is highly resistant to cytotoxic, targeted and immune therapies[1]. Compared to many cancers, PDA is remarkable for its relatively uniform DNA alteration repertoire, comprising frequent mutations in *KRAS*, *CDKN2A*, *TP53* and *SMAD4*. These hallmark events are not yet broadly targeted by approved therapies, and there is a paucity of additional actionable alterations. Consequently, despite substantial progress in developing mutant KRAS inhibitors, cytotoxic combinations remain the standard of care, with most patients quickly developing chemoresistance.

Cellular heterogeneity has emerged as a major contributor to cancer chemoresistance, potentially owing to coexistence of malignant subpopulations with distinct transcriptional–epigenetic states and drug sensitivity[2], as well as to the contribution of diverse stromal subpopulations[3]. Specifically, cell states with distinct drug sensitivity profiles may serve as reservoirs of resistance, whose plasticity may eventually regenerate the full tumor heterogeneity[2], undermining bulk tumor profile-based drug sensitivity predictions.

Bulk-based studies of large PDA cohorts identified at least two transcriptional PDA subtypes, with more differentiated tumors (corresponding to classical or progenitor subtypes in prior studies) showing association with better outcome relative to poorly differentiated tumors (called quasi-mesenchymal, basal-like or squamous[4–9]). Although these bulk-level classifiers can be identified through expression of associated marker genes[10], aside from a few studies examining the functional roles of individual genes[11,12], their relevance at the single-cell level is still controversial, and their mechanistic drivers, non-oncogene dependencies and potential plasticity remain elusive.

e-mail: kenolive@columbia.edu; ac2248@cumc.columbia.edu

Single-cell PDA subtype dissection faces two challenges. First, single-cell RNA-seq (scRNA-seq) profiles are very sparse, with >80% of the genes undetectable in each given cell ('gene dropout') even in high-quality datasets, including most of the genes comprising bulk PDA classification signatures. Second, bulk-derived classifiers average signals from dozens of malignant and stromal cell subtypes. Although removing stromal contributions from bulk gene expression signatures helps harmonize discrepancies[9], the application of bulk-tissue-derived signatures to single cells[13,14] precludes novel cell state discovery. Despite these challenges, some studies have taken a more unbiased approach to analyzing single-cell PDA expression profiles, focusing on tumor epithelial cells[15], stromal cells[3,16], spatial patterns[17] and clinical parameters[13,18]. However, how these malignant cell states are regulated, as well as their function, plasticity and non-oncogene dependencies, is still poorly understood.

We addressed these challenges by using gene regulatory network (GRN) analysis to characterize malignant PDA cells based on the activity of over 1,800 regulatory proteins, including transcription factors (TFs) and co-factors (co-TFs) that regulate their transcriptional state. This approach has outperformed gene expression-based analyses in the stratification of cell states[3,19,20] and helped discover experimentally validated mechanisms underlying tumor initiation[21], progression[22], heterogeneity[23] and drug sensitivity[24].

GRN analysis quantifies the activity of each TF and co-TF protein based on the differential expression of its transcriptional targets (regulon), as inferred by the extensively validated ARACNe algorithm[25]. These analyses, as implemented by the VIPER algorithm[26], overcome the gene dropout challenge, enabling protein activity quantification for all TFs and co-TFs—including those whose encoding genes are undetectable in scRNA-seq profiles—comparing favorably even with flow cytometry[3,19,27]. These analyses have proven effective in identifying MR proteins that govern many malignant and stromal cell states[21–23,26,28].

Using metaVIPER[29]—the single-cell extension of VIPER—we performed GRN-based analysis of individual malignant cells from 110 PDA samples from six publicly available scRNA-seq datasets. The analysis identified six co-existing, molecularly distinct cell states, conserved across virtually all samples. These comprise three pancreatic developmental lineages (PDLs), which we termed gastrointestinal lineage state (GLS), primitive lineage state (PLS) and morphogenic state (MOS), based on functional associations of their VIPER-inferred MR proteins; each PDL further stratified into M^+ (MAPK-high) and M^- (MAPK-low) substates based on MAPK pathway activity.

PDL signature analysis identified MOS state associations with poorly differentiated tumors and worse prognosis. Spatial transcriptomics, using 10x Xenium technology, and multiome assay for transposase-accessible chromatin using sequencing (ATAC-seq) confirmed functional and pathological associations and the distinct epigenetic nature of these states. Barcode-based lineage tracing demonstrated widespread spontaneous plasticity across lineages and M^+ or M^- substates, as well as rapid MEK inhibitor-mediated $M^+ \rightarrow M^-$ transition, confirming drug-induced plasticity. Finally, loss-of-function and overexpression assays validated the mechanistic role of MRs in cell state implementation and as PDL-specific non-oncogene dependencies. Taken together, these findings present a mechanism-based rationale for PDA malignant compartment heterogeneity and plasticity, which underlie its therapeutic resistance.

Results

The transcriptional states of PDA

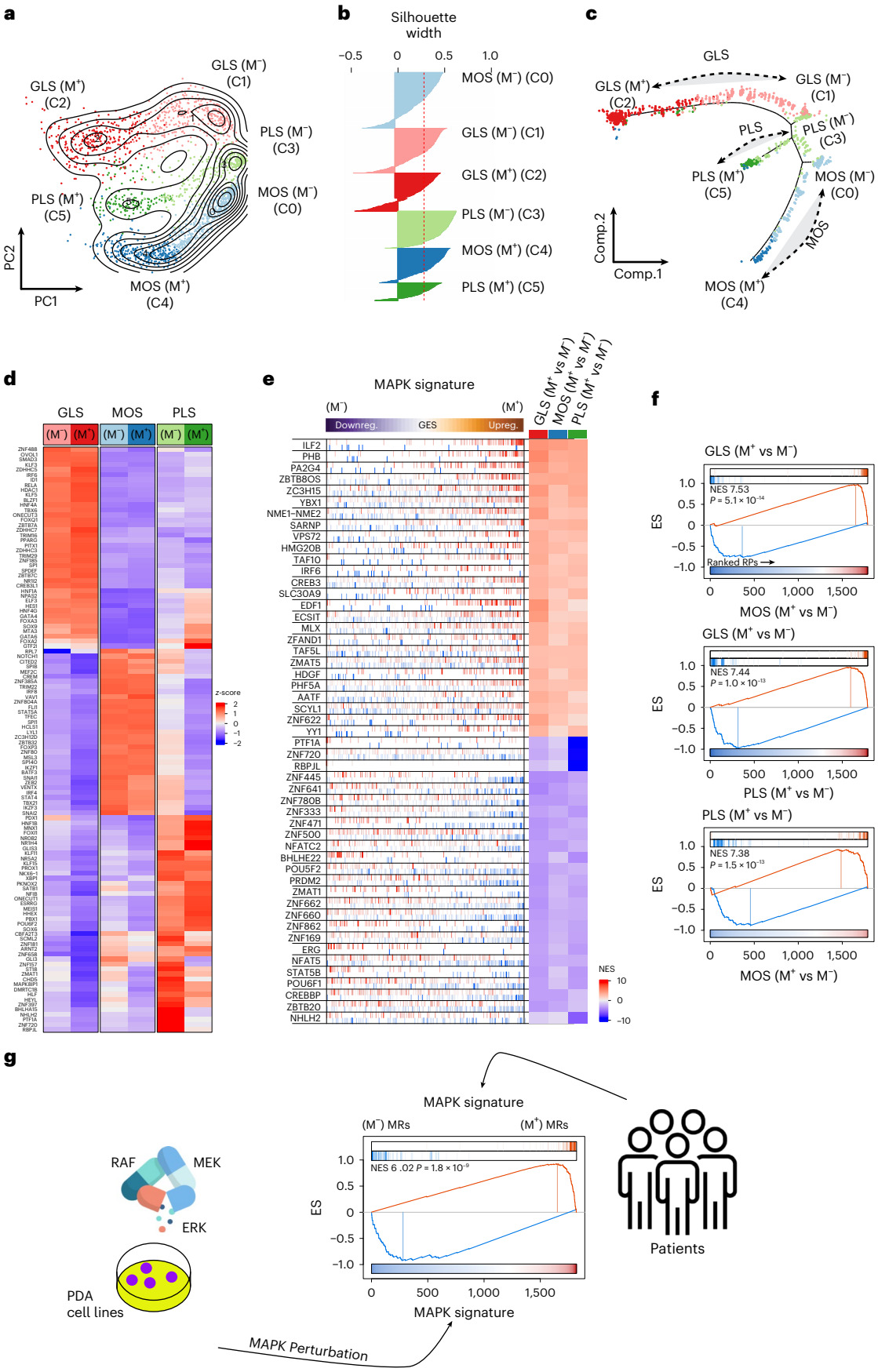
We performed GRN analysis of scRNA-seq profiles from human PDA tumors, using metaVIPER to integrate multiple GRNs generated by ARACNe analysis from four high-quality PDA cohorts, including the International Cancer Genome Consortium (ICGC)[6], The Cancer Genome Atlas (TCGA)[7], the University of North Carolina[5] and Columbia University (CUMC-E)[30]. The latter comprises profiles obtained from laser capture microdissection (LCM) of malignant epithelial samples from 200 patients with PDA. We then analyzed five publicly available scRNA-seq datasets[3,14,15,31,32] and one single-nucleus RNA-seq (snRNA-seq) dataset[18]. The Elyada dataset was selected for initial cell state discovery purposes because it presented the highest cumulative percentage of variance explained (CPVE) and optimal entropy (Extended Data Fig. 1a–f).

Louvain cluster analysis[33,34] of metaVIPER-assessed TF and co-TF activity identified six distinct malignant PDA cell regulatory states (Fig. 1a,b), undetectable by gene expression-based cluster analysis (Extended Data Fig. 1g). Protein activity-based pseudotime trajectory inference, using Monocle[35], yielded nearly identical stratification (Fig. 1c). The six states represent three complementary PDA PDLs, which were designated GLS, MOS and PLS based on MRs representing distinct developmental markers. GLS was characterized by activation of gastrointestinal TFs (HNF4A, HNF1A, HNF4G, ONECUT3, GATA4, GATA6) (Fig. 1d); MOS was characterized by activation of epithelial–mesenchymal transition TFs (NOTCH1, ZEB2, SNAI1, SNAI2); and PLS was characterized by activation of pancreatic progenitor TFs (ONECUT1, NKX6.1, PDX1, GLIS3) (Fig. 1d).

Each PDL was further stratified into two substates, with differentially active MRs conserved across all three PDLs ($P \leq 1.5 \times 10^{-13}$, by two-sided gene set enrichment analysis (GSEA)) (Fig. 1e,f). Among the top MRs, the analysis identified active RAS signaling-associated regulators, such as YY1 and YBX1, and inactive acinar cell regulators, such as PTF1A and RBPJL, which are typically inactivated upon RAS activation[36–38]. As confirmed by the enrichment of MAPK signal across the two substates of each PDL (Extended Data Fig. 2a), these are probably associated with differential RAS signaling activity. To further test this hypothesis, we generated an experimental, PDA-specific MAPK

Fig. 1 | PDA transcriptional cell states. a, Principal component analysis (PCA) based on VIPER-inferred protein activity profiles showing the six cell states identified by cluster analysis. **b**, Silhouette plot showing the silhouette score of each individual cell in the assigned cluster. The red dashed line indicates a Silhouette Score of 0.25, which is used as a threshold for statistical significance. **c**, Pseudotrajectory plot generated by the Monocle algorithm on the VIPER-inferred protein activity profiles. **d**, Heatmap showing the VIPER-inferred most differentially activated 25 proteins of each cluster compared to all the other clusters, including the GLS and MOS markers (HNF1A, GATA6, NOTCH1, SNAI1, SNAI2, GLI3, GATA4, SOX9 and HNF4A). **e**, msVIPER plot of the MAPK signature identified across PDLs. Each row shows the distribution of the target genes (vertical lines) of a given regulatory protein (RP) in the gene expression signature distinguishing M^+ versus M^- cells (top bar). Only target genes with the highest likelihood across the four PDA networks used for metaVIPER are shown in the plot. Blue vertical lines represent negatively regulated target genes, and red vertical lines represent positively regulated target genes. The top bar represents the integrated gene expression signature across the three PDLs distinguishing M^+ versus M^- states, with differentially expressed genes ranked from most downregulated in M^- (blue) to the most upregulated in M^+ (orange). The three columns on the right indicate the activation status of each regulatory protein in the M^+ state compared to the M^- state within each PDL. Red indicates activation in the M^+ state, and blue indicates inactivation. **f**, GSEA plots showing the conservation of VIPER-inferred MRs between M^+ and M^- states across the three PDLs. Top: GSEA plot showing the two-sided enrichment analysis of the top 50 MRs (hits) (the 25 most activated and 25 most inactivated) distinguishing M^+ and M^- states of the GLS PDL in the protein activity signature (*x* axis) distinguishing M^+ and M^- states in the MOS PDL. The same analysis was done to perform a pairwise comparison across all the PDLs. Normalized enrichment scores (NES) and *P* values were estimated by two-sided enrichment analysis with 1,000 permutations. **g**, GSEA plot showing the conservation of the MAPK MRs (the 25 most activated and 25 most inactivated) inferred from patients, in the VIPER MAPK signature experimentally inferred by perturbing RAF-MEK-ERK in PDA cell lines. NES and *P* value were estimated by two-sided enrichment analysis with 1,000 permutations. Illustration created with Flaticon.com.

https://doi.org/10.1038/s41588-026-02714-8



differential protein activity signature from two cell lines (PANC1 and ASPC1) before and after treatment with 14 different RAF, MEK or ERK inhibitors. This signature was enriched in MRs differentially active between the two substates (Fig. 1g; $P = 1.8 \times 10^{-9}$, two-sided GSEA), which were thus denoted as M^+ and M^-. As expected, metabolic enzyme expression was significantly lower in M^- states (Extended Data Fig. 2b).

Consistent with RAS/MAPK pathway regulation by upstream growth factor receptor signaling, culturing six human PDA cell lines in serum-free media significantly depleted the M^+ substates ($P = 0.026$, paired Student's *t*-test) (Extended Data Fig. 2c). Moreover, treatment of 11 PDA cell lines with the MEK inhibitor trametinib for 24 h also significantly depleted the M^+ substate ($P = 0.0006$, paired Student's *t*-test) (Extended Data Fig. 2d). The short treatment duration suggests that PDA cells can rapidly transition between high and low MAPK activity states, independent of developmental lineages. Thus, PDA cell states are regulated by MR proteins controlling either developmental lineage or MAPK activity.

Robustness and histopathological features of PDA cell states

To assess robustness and generalizability of the six cell states from the Elyada dataset, we evaluated enrichment of their MR signatures in single cells from five other curated datasets[14,15,18,31,32]. Across all cohorts, ~87% of malignant cells were significantly enriched for only one of the six PDA cell states ($P \leq 0.05$, two-sided analytic rank-based enrichment analysis (aREA)[26]) (Extended Data Fig. 2e–i), indicating high reproducibility. Only 2% of cells could not be classified. The remaining 11% of cells were enriched for two or more state-specific MR signatures, consistent with plastic cross-state transdifferentiation. MOS cells were under-represented in the Lin, Peng, CSY and Steele datasets, consistent with the malignant cell selection strategies used by these studies, which relied on either experimental techniques (CSY) or expression of known epithelial markers (Lin, Peng and Steele), thus resulting in depletion of poorly differentiated malignant cells that had undergone epithelial–mesenchymal transition. By contrast, the Hwang snRNA-seq dataset from frozen tissue presented a much higher MOS cell fraction.

Analysis of seven human PDA cell lines and of a patient-derived xenograft (PDX) model (Extended Data Fig. 2j,k) revealed coexistence of cells from at least two PDL lineages, including both their M^+ and M^- substates, suggesting that PDA heterogeneity may emerge in an isogenic context. Although most cultured PDA cells were M^+, M^- cells dominated the primary human tumors, despite uniform oncogenic *KRAS* mutations. Although potentially counterintuitive, this likely reflects that culture media are replete with nutrients and oxygen, whereas nutrient and oxygen availability in the tumor milieu is limited[39]. Confirming these findings, most malignant cells in human PDA samples stained negatively for the MAPK pathway biomarker phosphorylated-ERK (p-ERK)[40] (Extended Data Fig. 3a–d).

The architecture of human PDA

We examined histopathological characteristics of PDA cells in different malignant states by 10x Xenium spatial transcriptomics analysis. For this purpose, we designed a custom 100-gene panel tailored to assess the activity of the MR governing the six identified states and profiled eight primary tumor samples from chemonaive PDA patients. We used a two-tier classification strategy: first, we separated malignant from non-malignant cells using an scRNA-seq-derived malignancy signature; then, we classified malignant cell states based on the enrichment of cell-state-specific MR targets in the custom panel. Alignment with H&E-stained tissue sections revealed striking spatial alignment between PDL annotation and histopathology (Fig. 2a–d and Extended Data Fig. 3e–j). GLS cells showed a well-differentiated appearance and were restricted to glandular structures (Fig. 2a). MOS cells were poorly differentiated and appeared either as isolated cells along the basolateral surface of glandular structures or invading the stromal microenvironment (Fig. 2c). PLS cells were found in glandular structures (Fig. 2a) and invading the adjacent stroma (Extended Data Fig. 3i), with morphologies matching their context. Different tumors showed high heterogeneity, including across regions within the same tumor, ranging from nearly pure representation of one developmental lineage (Extended Data Fig. 3e) to nearly equal intermixing of all three (Extended Data Fig. 3g). Unlike developmental lineages, MAPK states were not associated with histopathological features. Consistent with single-cell analyses and p-ERK immunohistochemistry results, most malignant cells were in an M^- state (Fig. 2b,d and Extended Data Fig. 3f,j), although occasional clusters of M^+ cells were also observed (Extended Data Fig. 3h). These results demonstrate that the overall differentiation state of human pancreatic tumors reflects the combined contributions of malignant cells from all three developmental lineages.

GLS and MOS cell states drive bulk-tissue clustering

To assess whether cell state heterogeneity explains variance in bulk-tissue expression cohorts, we performed *k*-medoid cluster analysis of metaVIPER-inferred protein activity profiles from TCGA and ICGC PDA cohorts. The analysis yielded a two-cluster optimal solution for each dataset, with nearly identical MR signatures ($P < 10 \times 10^{-40}$, two-sided aREA) (Extended Data Fig. 4a).

The two signatures were integrated using Stouffer's method[41] to generate a consensus bulk PDA regulatory state signature (Fig. 2e). MRs of these bulk-level states were highly enriched in GLS and MOS MRs from single cell analyses ($P \leq 2.8 \times 10^{-5}$ and $P \leq 2.4 \times 10^{-6}$, two-sided aREA, respectively) (Fig. 2f,g), defining them as 'bulk-GLS' and 'bulk-MOS' states or simply as GLS and MOS when the context is obvious (that is, bulk-GLS for sample and cell line-specific analyses). PLS-specific MR enrichment in bulk-profiles was mixed: PLS M^+ MRs showed modest bulk-GLS enrichment ($P = 1.4 \times 10^{-2}$) and PLS M^- MRs showed borderline bulk-MOS enrichment ($P = 5.5 \times 10^{-2}$). This observation is consistent with heterogeneous PLS morphology, as assessed from spatial transcriptome analysis, and explains the absence of a PLS-specific bulk cluster. Additional cohorts with bulk RNA-seq profiles replicated these findings (Supplementary Notes).

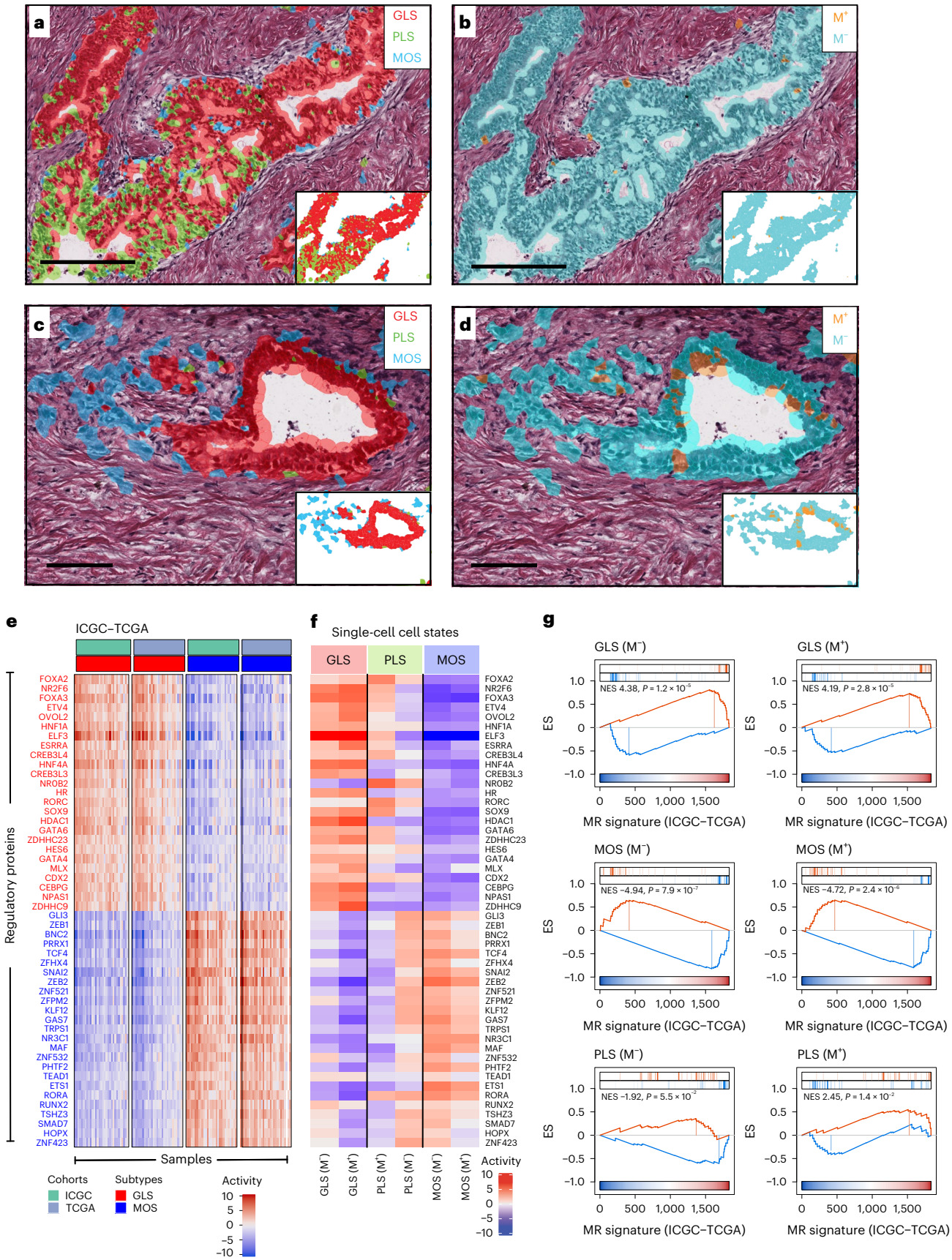
As previously shown, signatures representing less differentiated states—including the quasi-mesenchymal[4], basal-like[5], squamous[6] and basal-like A and B[14]—were significantly associated with aggressive tumors and poor survival outcomes, and correlated with the MOS ($P \leq 0.05$, one-sided Fisher's exact test) (Extended Data Fig. 4b). Indeed, MOS samples were associated with significantly worse survival than GLS samples in CUMC-E ($P = 5.2 \times 10^{-4}$, hazard ratio = 1.8) (Fig. 3a,b)

Fig. 2 | Spatial biology and bulk-based cluster analysis. a, Cell state map based on Xenium spatial transcriptomics showing the spatial distribution of PDL cell states, defined by the enrichment of cell-state-specific genes. Experiments were independently repeated in eight tumor samples, confirming the heterogeneous architecture of PDA. **b**, Cell state map based on Xenium spatial transcriptomics showing the spatial distribution of MAPK-dependent states in the same tissue section as in **a**. **c**, Cell state map based on Xenium spatial transcriptomics showing the spatial distribution of PDL cell states, based on the enrichment of cell-state-specific genes. **d**, Cell state map based on Xenium spatial transcriptomics showing the spatial distribution of MAPK-dependent cell states, based on the enrichment of cell-state-specific genes, in the same tissue section as in **c**. Insets (**a**–**d**): tumor cells isolated from the surrounding stroma. Scale bars, 100 µm. **e**, Heatmap showing clusters of patients from TCGA and ICGC based on the most differentially active MRs (tumor checkpoint) generated by the integration of ICGC and TCGA analyses. **f**, Heatmap showing the activity of the ICGC–TCGA top MRs in the cell states identified by single-cell analysis. **g**, GSEA plots showing the enrichment of the cell-state-specific MRs identified by single-cell analysis in the VIPER-inferred MR signature generated by integrating ICGC and TCGA analyses. NES and *P* values were estimated by two-sided GSEA with 1,000 permutations.

https://doi.org/10.1038/s41588-026-02714-8

and were more likely to be poorly differentiated ($P = 1.0 \times 10^{-3}$, χ^2 test) (Fig. 3c). Single-cell-based signatures confirmed the MOS association with poor prognosis or differentiation ($P < 0.05$, Fisher's exact test; Supplementary Notes).

Analysis of 190 PDA samples from an LCM cohort (CSY_{LCM}) identified *KRAS* imbalance associated with bulk state signatures ($P = 2.0 \times 10^{-4}$, two-sided χ^2 test) (Extended Data Fig. 4c), while TCGA DNA methylation data revealed distinctive patterns differentiating



https://doi.org/10.1038/s41588-026-02714-8

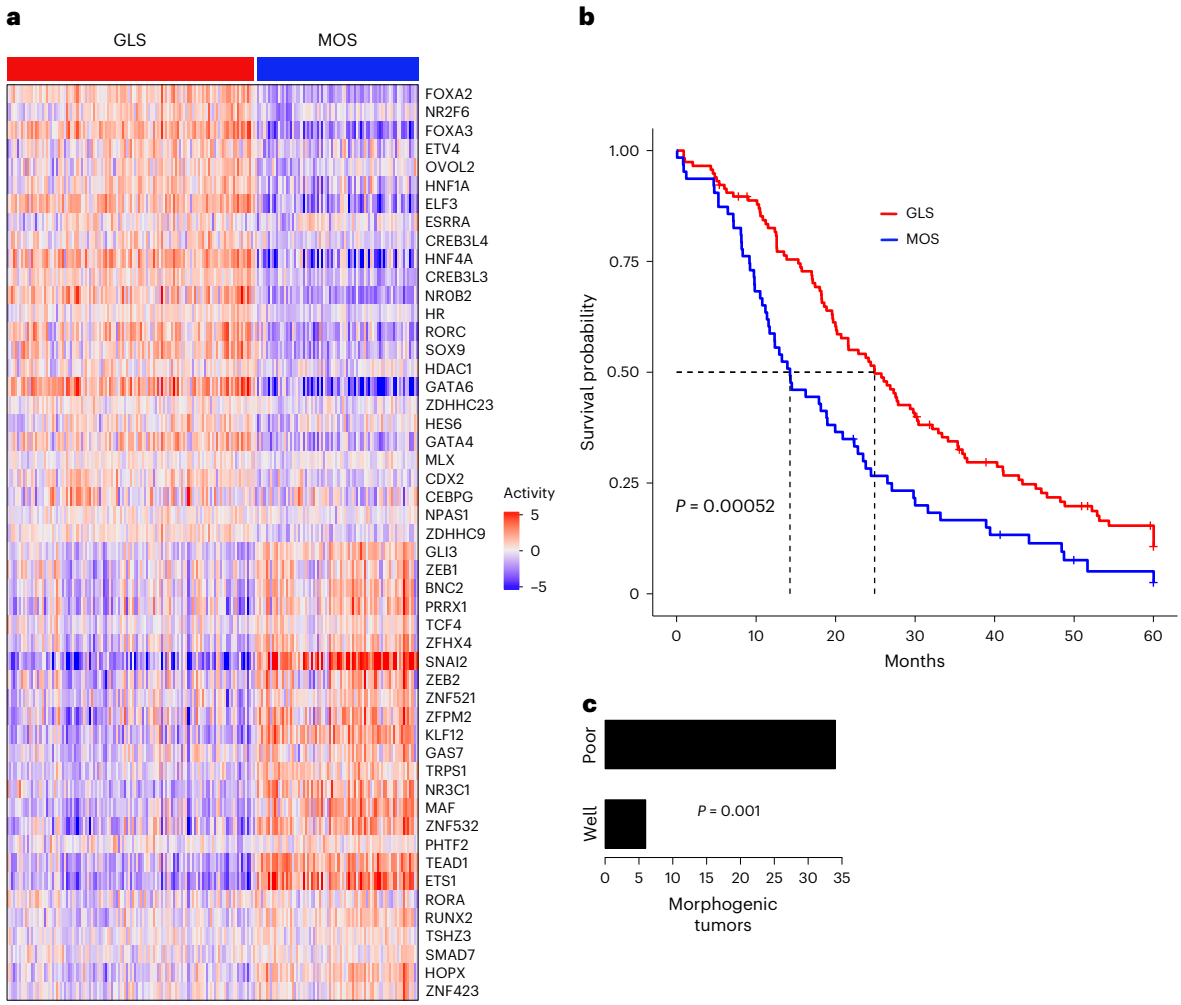


Fig. 3 | MR-based stratification of CUMC-E cohort. a, Heatmap showing patient (columns) stratification of the CUMC-E cohort based on the MR tumor checkpoint derived by the integration of ICGC and TCGA analyses. **b**, Kaplan–Meier curve showing that patient stratification of the CUMC-E cohort in MOS and GLS based on ICGC–TCGA MR tumor checkpoint correlates with patient survival (*P* value computed by log-rank test). **c**, Barplot showing that MOS tumors are enriched for poorly differentiated tumors (*P* value computed by one-sided chi-squared test).

GLS from MOS samples (Extended Data Fig. 4d). These data show that bulk-level PDA subtypes are determined by the ratio of malignant cells in epigenetically distinct regulatory states rather than by genetic drivers. Stratification of PDA cell lines into GLS or MOS based on bulk profiles nearly perfectly recapitulates their MOS versus GLS cell fraction (Extended Data Figs. 4e and 2j).

Barcode lineage tracing reveals mutation-independent plasticity

To assess PDA cell state plasticity, we performed barcode-based lineage tracing. Cell lines enriched in specific cell states were transduced with random 28-nucleotide barcode libraries, puromycin-selected, seeded in an 8,000 cells per well format, cultured for one to two population doublings to replicate the barcodes and then split into two pools. One pool (50% of cells) was collected at time point zero (T_0) (approx. one to two population doublings after seeding 8,000 transduced cells per well); the second pool was collected after ten additional population doublings (T_1). The scRNA-seq and GRN analysis were then used to identify and match barcodes in cells representing different states at the T_0 and T_1 time points.

First, we assessed plasticity between GLS, MOS and PLS states in the Hs766T cell line. This line shows adequate representation of all three states (Extended Data Fig. 2j), thus supporting pairwise state-transition analysis (Fig. 4a–c). Barcodes from each lineage at T_0 were found in cells representing all three lineages at T_1, with kinetics recapitulating equilibrium state admixture, thus confirming widespread spontaneous isogenic PDA lineage state plasticity.

We then assessed M^+/M^- plasticity. Given the limited fractional representation of each M^+ and M^- state in Hs766T cells, we repeated the analysis using PATU8988S and KP4 cells, highly enriched in GLS and MOS cells, respectively (no cell lines were enriched for PLS cells, thus preventing that additional analysis). Barcode analysis confirmed extensive transitions across MAPK states regardless of their GLS or MOS lineage (Fig. 4d,e), further confirming that transcriptional PDA cell states are highly plastic and undergo spontaneous transdifferentiation, including between the two MAPK substates. To further assess whether M^+/M^- transdifferentiation is causally induced or suppressed by RAS/MAPK signaling activity, we performed barcode-based lineage tracing in KP4 cells treated with 100 nM trametinib (a MEK inhibitor) or vehicle control (dimethyl sulfoxide (DMSO)) 24 h before the second time point (Fig. 5a). Consistent with spontaneous transitions, the M^+ state was depleted in DMSO-treated cells at T_1; however, trametinib-treated cells showed statistically significantly increased reprogramming compared to the DMSO baseline (Fig. 5b–d) ($P = 9.2 \times 10^{-152}$, using bootstrapping to control for the number of barcodes in each condition; Fig. 5c). This finding shows that PDA

https://doi.org/10.1038/s41588-026-02714-8

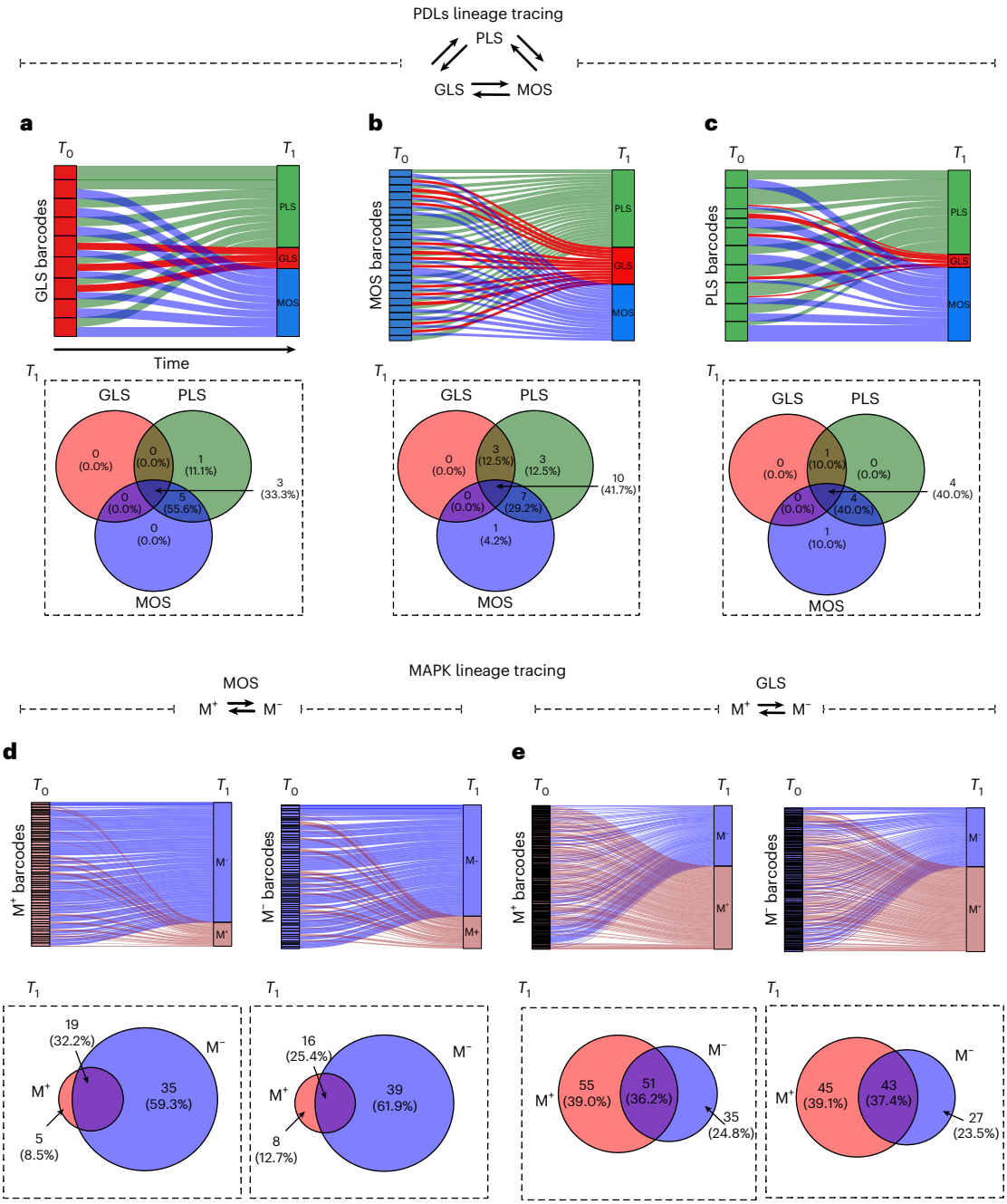


Fig. 4 | Single-cell lineage tracing across cell states. a, Top: alluvial plot representing the transition of GLS barcodes to the other cell states in the Hs766T model. Bottom: Venn diagram representing the redistribution of the GLS barcodes at T_1. **b**, Top: alluvial plot representing the transition of MOS barcodes to other cell states in the Hs766T model. Bottom: Venn diagram representing the redistribution of the MOS barcodes at T_1. **c**, Top: alluvial plot representing the transition of PLS barcodes to the other cell states in the Hs766T model. Bottom: Venn diagram showing the redistribution of the PLS barcodes at T_1. **d**, Top: alluvial plots representing the transitions of barcodes between M^+ and M^- states across the two time points in MOS-enriched (KP4) cells. Bottom: Venn diagram plots representing the redistribution of the barcodes at T_1 in KP4 cells. **e**, Top: alluvial plots representing the transitions of barcodes between M^+ and M^- states across the two time points in GLS-enriched (PATU8988S) cells. Bottom: Venn diagram plots representing the redistribution of the barcodes at T_1 in the lineage plasticity model.

MAPK states are directly controlled by RAS/MAPK signaling and, importantly, that malignant PDA cells can adaptively reprogram following MAPK inhibitor treatment, highlighting an important potential resistance mechanism.

MAPK substates are epigenetically distinct

Given the strong influence of MAPK signaling on cell proliferation and checkpoint progression, we questioned whether MAPK states represent stable, epigenetically distinct cell states or simply reflect cell cycle phases. For this purpose, we generated multiome profiles of KP4 cells (MOS-enriched), including single-cell ATAC-seq (scATAC-seq) and scRNA-seq. Uniform manifold approximation and projection visualization of protein activity-based clusters showed four clusters annotated for cell cycle phase using Seurat[33], revealing clear separation of G1 versus G2/M cells (Fig. 5e), with two sub-clusters per phase-specific cluster. Proteins differentially activated between the two sub-clusters of both G1 and G2/M cells were highly enriched in M^+ versus M^- state MRs (Fig. 5f), confirming that M^+ and M^- states are independent of cell cycle phase. To further assess whether these states are epigenetically distinct, we performed differential chromatin peak analysis using the

scATAC-seq profiles of M^+ versus M^- cells. The differential chromatin peak analysis was performed independently for G1 and G2/M cells and then integrated across both phases, yielding 2,027 differentially accessible chromatin peaks between the two MAPK states (Fig. 5g). Strikingly, the TFs whose motifs were enriched in these peaks were significantly enriched in M^+ versus M^- MRs (Fig. 5h), as well as in established RAS/MAPK-responsive TFs, such as MYC, HIF1A, ETS1 and ELK1/3, as assessed by EnrichR[42] analysis using the KEA database[43] (Fig. 5i,j). Taken together, these data establish M^+ and M^- as epigenetically distinct, cell cycle phase-independent cell states.

PDA essentiality is cell-state specific

To evaluate the functional relevance of identified PDA cell states, we assessed whether state-specific MRs were enriched in state-specific essential proteins. Given that PLS states are poorly represented in PDA cell lines, we focused on assessing MR viability effects based on GLS versus MOS MR signatures, as assessed from bulk or single cell profiles, respectively. For this purpose, we selected PATU8988S, HPAFII and CAPAN1 as GLS-enriched representatives and PANC1, KP4 and PK45H as MOS-enriched representatives. We performed pooled CRISPR–Cas9-mediated knockout (CRISPR KO) and CRISPR–dCas9-mediated inhibition (CRISPRi) screens in all six cell lines, focusing on transcriptional regulatory genes (that is, TFs and co-TFs) (Extended Data Fig. 5a). Cells were transduced with a pooled single guide RNA (sgRNA) library targeting a panel of over 3,000 genes (four sgRNAs per gene), including the subset of 1,835 regulatory proteins[26] whose activity was inferred by metaVIPER and over 600 essential, non-essential and control genes[44,45] (Extended Data Fig. 5b,c). Cells were collected and sequenced at $T_0 = 3$ d and $T_1 = 33$ d post transduction. The sgRNA counts were then integrated across both technical and biological replicates to assess their differential representation over time and then compared between MOS- and GLS-enriched cell lines to produce a differential, tumor state-specific, essentiality signature.

Top 100 GLS and top 100 MOS candidate MRs were significantly enriched in sgRNAs eliciting depletion of their matched cell state ($P = 9.1 \times 10^{-5}$, two-sided GSEA). Confirming single-cell and bulk-level MR overlap (Fig. 2g), the top 100 bulk-GLS and top 100 bulk-MOS MRs were similarly enriched ($P = 2.6 \times 10^{-4}$) (Extended Data Fig. 5d,e). CDX2, GATA6 and HNF1A emerged among the most essential GLS MRs, while MYBL1, ZEB1 and GLI2 emerged among the most essential MOS MRs (Fig. 6a,b). These functional dependencies provide paths as well as gene reporter assays for identifying drugs selectively targeting individual PDA cell types for combination therapy.

We then assessed whether more universal, cell state-independent MRs may emerge whose knockout or silencing may elicit essentiality in all cell lines. For this purpose, we identified candidate tumorigenesis MRs by metaVIPER analysis of genes differentially expressed in the malignant PDA epithelium (CUMC-E dataset) versus GTEx normal tissues[46] independent of PDL state. These were significantly enriched in essential genes across all six cell lines ($P = 5.5 \times 10^{-3}$, one-sided GSEA) (Fig. 6c,d).

To confirm state-specific MR essentiality in vivo, we transduced a smaller sgRNA library targeting 32 GLS and 61 MOS MRs, representing the most conserved state-specific MRs (based on activity, CRISPR KO and CRISPRi results integration) into Cas9-expressing KP4 (MOS-enriched) and HPAFII cells (GLS-enriched). After puromycin selection, cells were injected into seven immunocompromised mice, and tumor formation was monitored for 29 days, followed by genomic DNA extraction and next-generation sequencing (NGS) of the sgRNAs. After quality control (Extended Data Fig. 5f,g), CRISPR analysis confirmed the differential, state-specific essentiality of GLS and MOS MRs identified from both bulk and single-cell analyses. Specifically, the top 12 out of 32 (38%) GLS MRs and top 38 out of 61 (62%) MOS MRs elicited significant state-specific essentiality in vivo ($P = 6.7 \times 10^{-4}$, two-sided GSEA) (Fig. 6e). The top GLS essential MR was HNF1A, which was also among the top essential proteins identified by in vitro CRISPR KO and CRISPRi assays. Top MOS essential MRs included BPTF, a key $KRAS^{Mut}$ effector that remodels chromatin and regulates MYC activity[47].

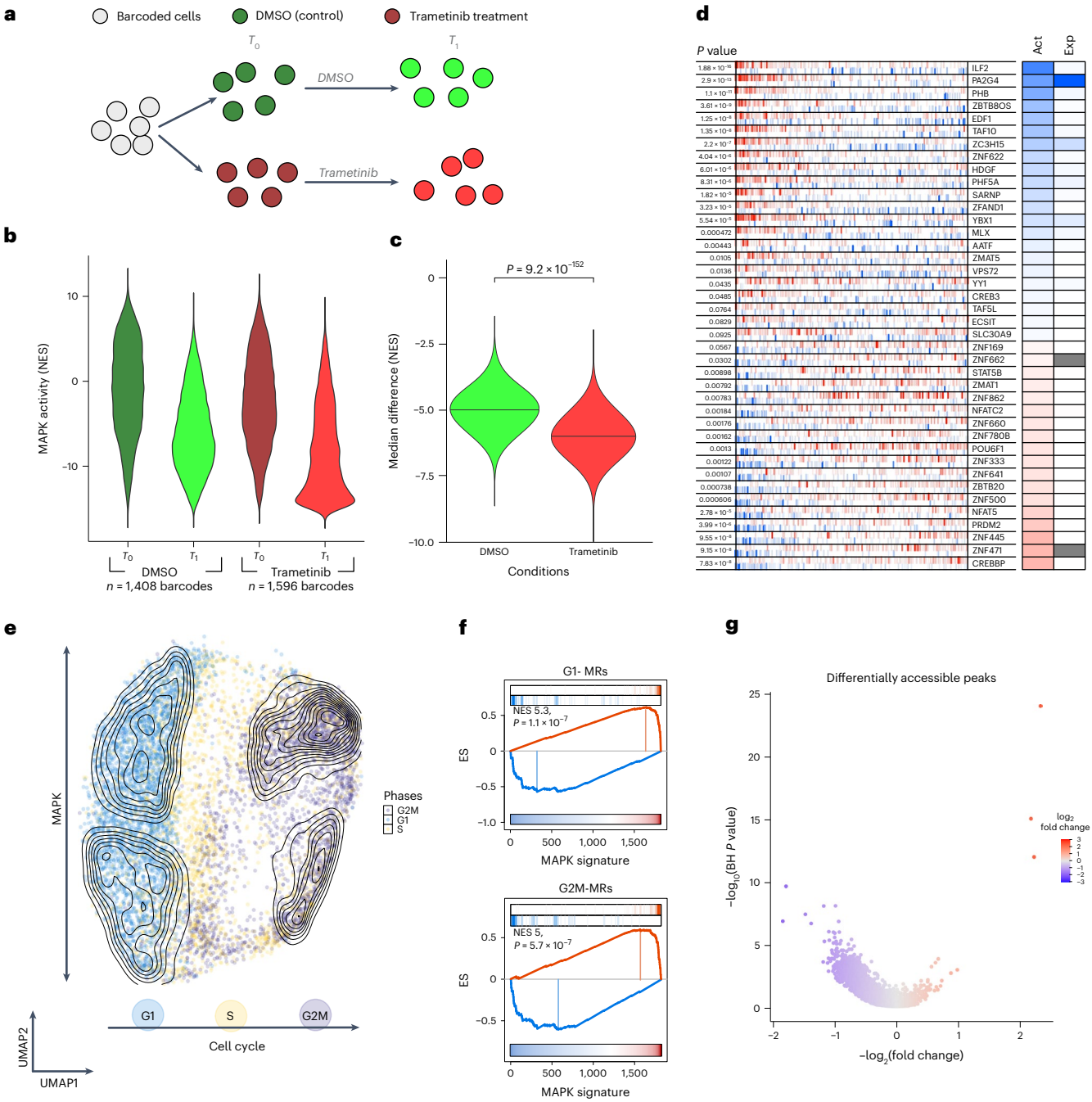
Ectopic MR expression reprograms PDA cell state

To assess whether candidate MRs mechanistically determine cell state, we asked whether ectopic expression of VIPER-inferred MRs could transdifferentiate PDA cells from the more aggressive, less differentiated MOS to the less aggressive, more differentiated GLS state. We performed lentiviral-mediated transduction of cDNAs encoding the eight most significant GLS MRs, from bulk-profile analysis, in MOS-enriched KP4 cells, using the tetracycline-inducible M2rtTA system[48]. However, selected MRs were also among the most significantly active in the single-cell-based GLS state (Extended Data Fig. 5h). Individually, each overexpressed MR effectively increased the activity of most other GLS-activated MRs (Fig. 7a). Yet, uniquely, ectopic OVOL2 expression also repressed most MOS-activated MRs, producing a virtually complete transdifferentiation of KP4 cells to a GLS-like state ($P = 4.3 \times 10^{-15}$, two-sided aREA) (Fig. 7a–c).

RNA-seq profiles following ectopic GLS MR expression in KP4 cells supported reconstructing transcriptional and post-translational MR → MR interactions by analyzing differential expression and VIPER-measured activity in cells overexpressing each MR versus controls. Consistent with the tumor checkpoint hypothesis[49], the

Fig. 5 | Trametinib-induced lineage tracing analysis and epigenetic changes associated with MAPK activity. a, Schematic representation of compound-induced reprogramming of PDA cells assessed by single-cell barcode lineage tracing. **b**, Violin plots showing the distribution of MAPK activity MR signature scores for matched barcodes across two time points in control (DMSO) and trametinib-treated conditions. **c**, Violin plots showing changes in MAPK activity between two time points in the DMSO (control) and trametinib-treated cells. Differences were assessed using a bootstrap procedure with 1,000 iterations. A two-sided *P* value was determined using the Wilcoxon test. **d**, MsVIPER plot showing the differential activity of the same RPs, inferred from patient data, between M^+ and M^- states as assessed by lineage tracing after trametinib treatment. **e**, Uniform manifold approximation and projection (UMAP) visualization based on protein activity of MOS cells. Cells (dots) are colored based on their cell cycle phase as predicted by Seurat analysis. The 2D density contours highlight cell clusters. **f**, GSEA plots showing the conservation of the MRs, inferred by comparing clusters within the same cell cycle phase, in the MAPK signature experimentally inferred by drug perturbation. The top GSEA plot shows the conservation of the MRs inferred between the two clusters in the G1 phase. The bottom GSEA plot shows the conservation of the MRs inferred between the two clusters in the G2M phase. Silhouette analysis, however, did not support the presence of well-separated clusters in the S phase. NES and *P* values were estimated by two-sided GSEA with 1,000 permutations. **g**, Volcano plot illustrating the differentially accessible peaks (DAP), as assessed by scATAC-seq, that showed differential MAPK activity across cell cycle phases. Differentially accessible peaks were identified using the logistic regression method as implemented in the Signac package. **h**, GSEA plot showing the enrichment of the TFs, whose motif was overrepresented in the differentially accessible peaks, in the protein activity MAPK signature experimentally inferred by a perturbation assay. The vertical gray line indicates the maximum value of the enrichment. NES and *P* value were estimated by two-sided GSEA with 1,000 permutations. **i**, Barplot showing the enrichment analysis by enrichR of the core set of TFs (leading edge in **h**) showing high activity in the MAPK signature and are overrepresented by motif enrichment analysis in DAP. *P* values were computed by enrichR with Fisher's exact test and adjusted for multiple testing using the Benjamini–Hochberg correction. **j**, Representative TF motifs of RPs showing differential activity between M^+ and M^- states and are overrepresented by motif enrichment analysis. Act, activity; Exp, expression.

https://doi.org/10.1038/s41588-026-02714-8



https://doi.org/10.1038/s41588-026-02714-8

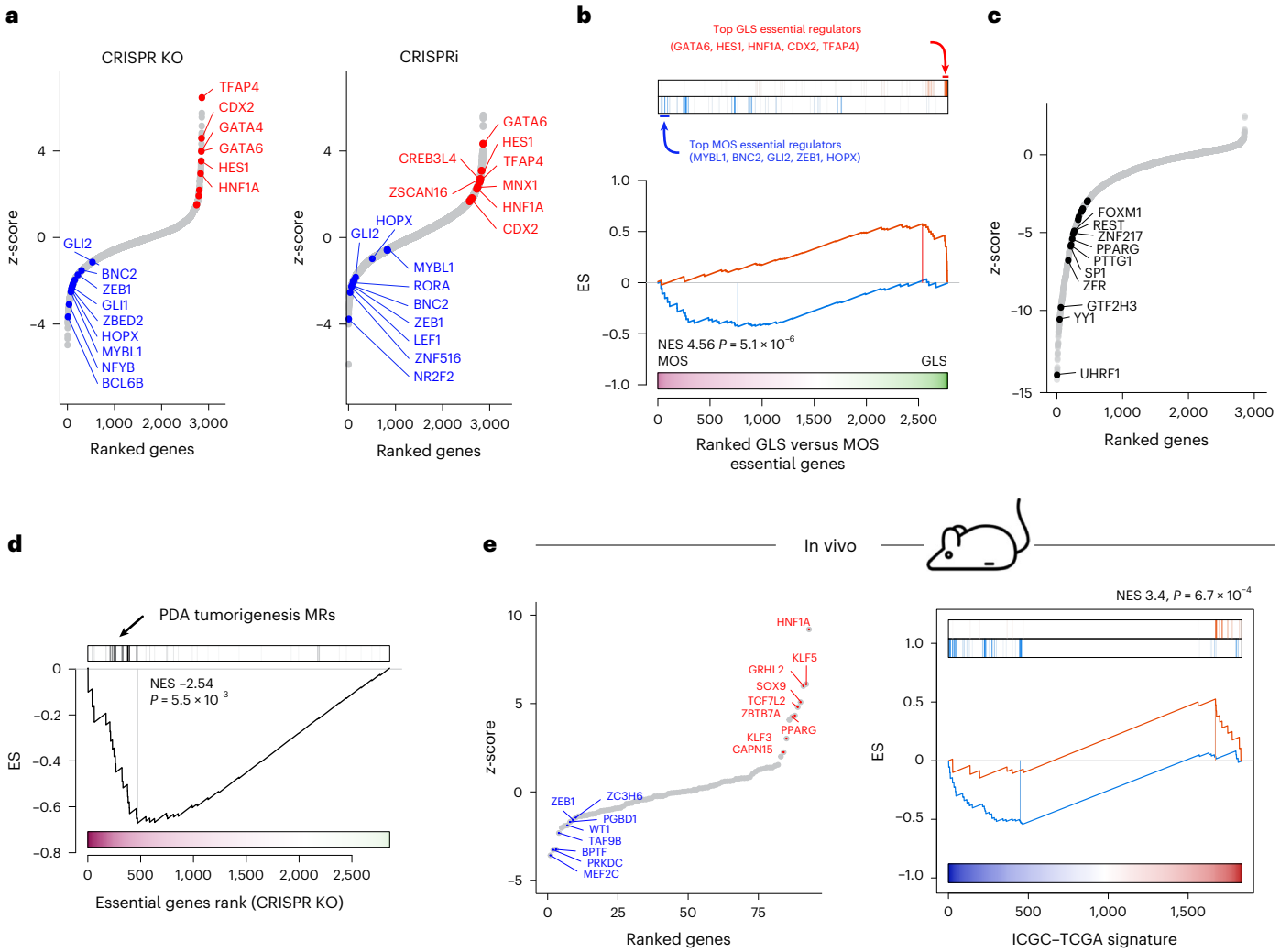


Fig. 6 | Functional CRISPR screening and essential MRs. a, Scatter plots showing differential gene essentiality signature between GLS- and MOS-enriched cell lines, as produced by CRISPR KO (left) and CRISPRi (right). Genes are ranked according to their differential essentiality score (*z*-score), from the most essential in MOS-enriched cells (left) to the most essential in GLS-enriched cells (right). VIPER-inferred GLS and MOS MRs are shown in red and blue, respectively. Bulk-profile MRs were used for this analysis since, similar to tumor samples, cell lines have a mixed composition. **b**, GSEA shows statistically significant enrichment of the 50 most differentially active GLS and MOS MRs in genes assessed as differentially essential between MOS and GLS cell lines, by integration of CRISPR KO and CRISPRi data (*P* value and NES were estimated by GSEA two-sided, with 1,000 permutations). **c**, Scatter plot showing genes ranked based on their PDA subtype-independent essentiality as determined by CRISPR KO. Some VIPER-inferred, subtype-independent MRs are highlighted. **d**, GSEA of subtype-independent PDA MRs, computed by comparing the 200 LCM CUMC-E samples against the gene expression centroid of all normal tissue samples in GTEx[46], in subtype-independent essential genes, as assessed by the pooled CRISPR KO screen (*P* value and NES were estimated by GSEA one-sided, with 1,000 permutations). **e**, Left: scatter plot showing differential gene essentiality signature between GLS- and MOS-enriched cells, as produced by CRISPR KO in vivo. GLS and MOS MRs are shown in red and blue, respectively. Right: GSEA plot shows statistically significant enrichment of the top 12 GLS and 38 MOS essential genes in proteins differentially active in GLS versus MOS patient-derived samples (ICGC–TCGA signature). Names in **a**–**c** and **e** are not italicized because they refer to the MR proteins although the analysis was performed by targeting the genes encoding them. NES and *P* value were estimated by two-sided GSEA with 1,000 permutations.

analysis revealed a complex modular structure with autoregulatory loops, in which the top eight GLS MRs positively regulate each other and repress the top eight MOS MRs (Extended Data Fig. 5i), producing a bistable (that is, on–off) switch responsible for state stability. This provides a mechanistic rationale for the role of these MRs in homeostatic PDL state control. Among the lineage MRs, OVOL2 emerged at the top of the regulatory hierarchy, directly activating or repressing the expression or activity of most other GLS and MOS MRs, respectively, explaining its singular ability to individually control transdifferentiation.

Western blotting confirmed OVOL2-mediated inactivation of its established transcriptional targets ZEB1 and vimentin[50,51]—known epithelial–mesenchymal transition effectors, a morphogenic subtype hallmark—and upregulation of epithelial markers, such as E-cadherin, undetectable in KP4 cells (Fig. 7d). OVOL2-mediated transdifferentiation was confirmed in two additional MOS-enriched cell lines (PANC1 and PK45H) (Extended Data Fig. 6a).

We then assessed whether failure to detect complete transdifferentiation by other MRs resulted from lower reprogramming efficiency, causing signature dilution in bulk profiles. We performed pooled overexpression of the same eight GLS MRs in individual KP4 cells at a multiplicity of infection (MOI) of 0.288 for each virus in a pool, followed by scRNA-seq profiling[52], such that most cells received one to three cDNAs.

We detected co-expression of 11 MR pairs and 14 higher-order MR combinations in ≥30 cells each (a threshold suitable for the analysis). Consistent with the initial experiment, 32% of KP4 cells were found in a GLS state following ectopic OVOL2 expression.

Moreover, other MRs and MR combinations induced statistically significant increase in GLS cells compared to controls, albeit in a smaller cellular fraction, except for the OVOL2–HNF1A pair, which significantly outperformed OVOL2 alone, resulting in 45% of the cells being classified as GLS (Fig. 7e and Extended Data Fig. 6b). This finding confirmed that low reprogramming efficiency prevented full effect detection in bulk cell line profiles. Other notable pairs included FOXA2–HNF1A, resulting in the third most efficient MOS → GLS reprogramming. Given that neither FOXA2 nor HNF1A produced significant effects individually, this observation shows synergistic interaction of the HNF1A–OVOL2 and HNF1A–FOXA2 MR pairs.

Finally, we tested whether ectopic expression of OVOL2 and HNF1A—the most synergistic pair—individually or combined could recapitulate MOS → GLS transdifferentiation in vivo. Plasmids containing open reading frame (ORF) constructs for OVOL2, HNF1A and the OVOL2–HNF1A combination, as well as mCherry (identical to in vitro assays), were lentivirally transduced into KP4 cells. Puromycin-selected cells were then pooled and injected into athymic mice ($n = 3$). Animals were killed at 14 days, and cells were scRNA-seq profiled, followed by metaVIPER-based cell state analysis to assess transdifferentiation towards a GLS state.

Enrichment of the most differentially active proteins, compared to negative controls, in proteins differentially active in the GLS → MOS signature was assessed by GSEA. Results confirmed OVOL2-mediated and HNF1A-mediated reprogramming of KP4 cells towards a GLS state ($P_{OVOL2} = 1.4 \times 10^{-6}$) and ($P_{HNF1A} = 5.3 \times 10^{-5}$), with the combination achieving virtually complete transdifferentiation ($P_{OVOL2/HNF1A} = 1.9 \times 10^{-9}$) (Fig. 7f,g), thus confirming their synergistic interaction. Intriguingly, although ectopic HNF1A expression was not effective in inducing statistically significant transdifferentiation in vitro, its effect in vivo was statistically significant, suggesting that MRs inferred from patient-derived samples may be better suited to in vivo than in vitro validation.

Discussion

Parsing molecular states from bulk PDA expression data is challenging because of tumor heterogeneity and cohort variability. However, consistent patterns have emerged, including well-differentiated versus poorly differentiated tumors, informing clinical practice. Single-cell analyses can improve assessment of intratumoral heterogeneity and reveal cellular contributions to progression and resistance. However, despite single-cell data availability, PDA epithelial characterization has mostly relied on bulk-derived signatures. Indeed, two challenges remain: gene dropout from low read depth per cell, and limited insight into causal drivers and non-oncogene dependencies.

We address both of these challenges using GRN analysis, identifying proteins maintaining six molecularly distinct tumor states, and confirming state-specific non-oncogene dependencies through CRISPR screens. Although effective protein markers were identified for two of the three main lineage states—GLS and MOS—a limitation of this study is the lack of validated PLS markers. This shortcoming is mitigated by spatial transcriptomics data, identifying PLS cells with distinct morphological characteristics.

The MR proteins identified by our work constitute a set of testable predictions for maintaining cell state identity and viability, and their ectopic expression validated them as PDA plasticity determinants. Functional screening confirmed that GLS and MOS MRs often contribute to cell state viability, providing a therapeutic roadmap targeting PDA heterogeneity.

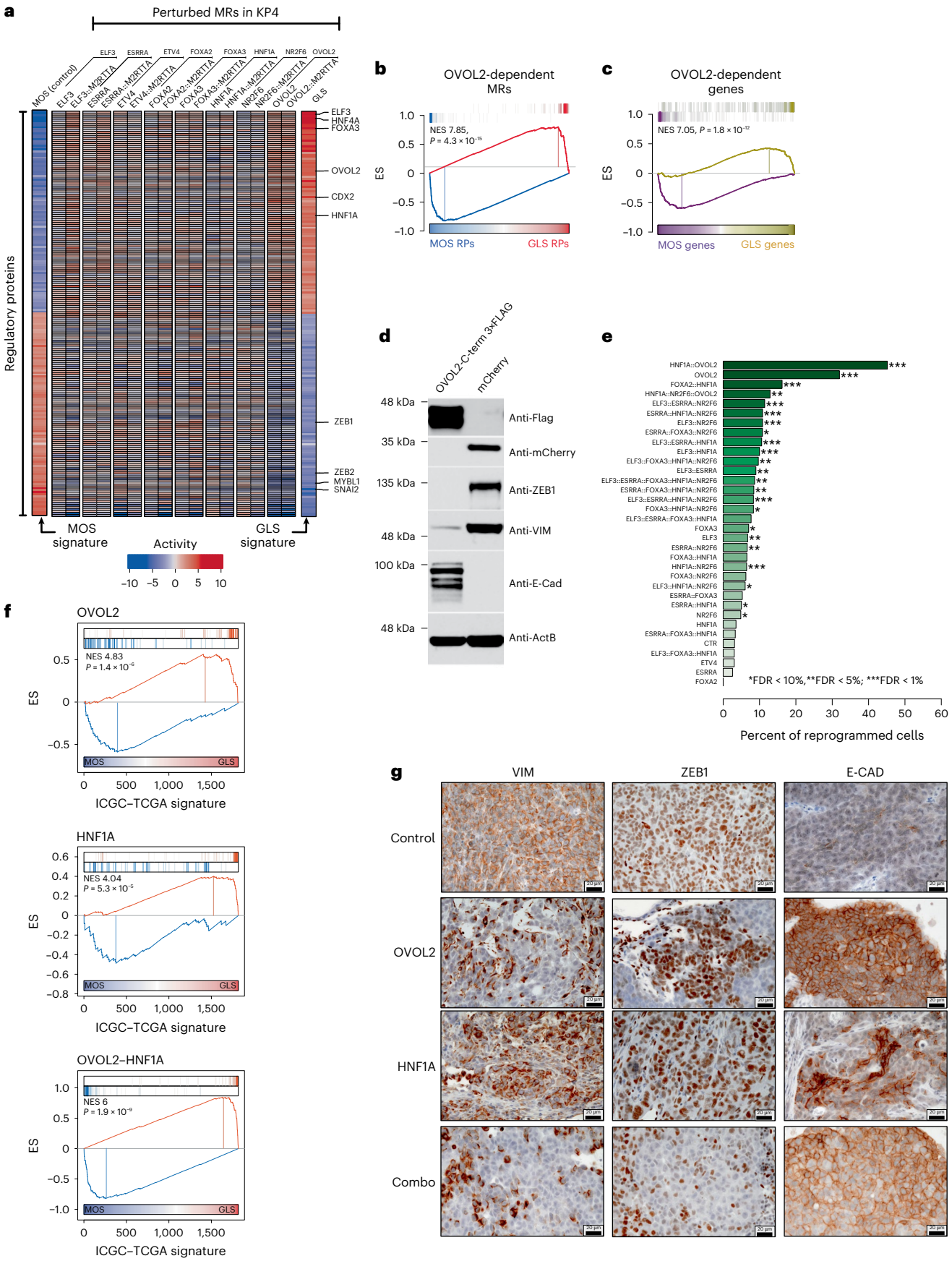
Within each of the three main lineages, we identified two substates that were distinguished by their relative RAS/MAPK activity levels; this observation was striking, given the near-universal presence of activating *KRAS* mutations in PDA. Yet, consistent with this finding, most malignant epithelial cells express low to undetectable p-ERK by immunohistochemistry, also consistent with the fact that mutant KRAS is ligand-dependent on signaling from upstream receptor tyrosine kinases such as EGFR[53]. Therefore, reduced exposure to growth factors, nutrients and oxygen probably limits in vivo proliferation, in contrast to in vitro conditions, during which PDA cells divide daily under nutrient-rich and oxygen-rich conditions. Our study suggests that the M^+ versus M^- dichotomy may reflect such nutrient-dependent proliferative heterogeneity. Indeed, consistent with this hypothesis, M^- cells dominate human PDA tumors in vivo, while M^+ cells dominate in culture.

Consistently, RAF/MEK/ERK inhibitors and growth factor withdrawal shift cells from M^+ to M^-, despite nutrient availability. Nevertheless, the M^- state is not long-term quiescent, as shown by barcode-based fate mapping and cell cycle analysis. Indeed, RAS/MAPK activity may toggle during proliferative cycles in a metabolic context-dependent fashion.

MAPK state divergence between patients and in vitro has profound implications for the development of effective therapies. Most drug screens use cultured cells, potentially biasing toward drugs that effectively target only a minority of malignant cells in vivo, suggesting that pushing cells into the M^- state may prove more informative. Additionally, our findings suggest that drugs depleting only one or two lineages may produce rapid non-genetic adaptation. Indeed, leveraging cell state MRs may facilitate identification of multi-drug regimens impacting all malignant cell states[54]. Identification of cell states and

Fig. 7 | Lineage reprogramming of MOS-enriched cells. a, Heatmap showing activity of top GLS and MOS MRs in the KP4 (MOS-enriched) cell line following ectopic expression (±M2rtTA) of the top eight GLS MRs ($n = 8$) from both bulk and single-cell analyses. These assays were done with three biological replicates. As a reference, to better visually assess transdifferentiation, the first and last columns represent the activity of the top-ranked bulk-MOS and bulk-GLS MRs, respectively, in KP4 (MOS-enriched) and HPAFII and PATU8988S (GLS-enriched) cells. **b**, GSEA plot showing the enrichment of the top 50 GLS (red) and top 50 MOS (blue) MRs in proteins differentially activated following ectopic OVOL2 expression in KP4 cells. *P* value and NES were assessed by two-sided GSEA with 1,000 permutations. **c**, GSEA plot showing the enrichment of the 200 most overexpressed (yellow) and underexpressed (purple) genes following ectopic OVOL2 expression in KP4 cells in genes differentially expressed in GLS- versus MOS-enriched cell lines (*P* value and NES assessed by two-sided GSEA with 1,000 permutations). **d**, Western blot showing inhibition of mesenchymal markers (ZEB1 and vimentin) and expression of epithelial markers (E-cadherin) following ectopic OVOL2-3×Flag expression in KP4 cells compared to ectopic expression of the negative control (mCherry). Western blots were conducted with three biological replicates. **e**, Bar plot showing the fraction of reprogrammed cells following ectopic expression of the top eight GLS MRs and their combinations in KP4 single cells. *P* values were assessed using Fisher's exact test, comparing the number of statistically significant GLS cells (false discovery rate (FDR) of <0.05) following ectopic expression of individual lineage MRs and their combination (MR set) in KP4 cells, against negative controls expressing mCherry (±M2rtTA). **f**, GSEA plots showing the in vivo reprogramming of a MOS-like tumor into a GLS-like tumor. The red and blue bars (gene set) represent the MRs differentially activated—increased (red) and decreased (blue), respectively—in a KP4 (MOS-enriched) xenograft following overexpression of GLS MRs (OVOL2, HNF1A and the OVOL2–HNF1A combination). The reference signature represents the lineage–morphogenic protein activity signature inferred from patients (ICGC–TCGA signature). *P* value and NES were estimated by GSEA two-sided, with 1,000 permutations. MRs of PDX tumor cells were computed using metaVIPER by comparing gene expression profiles of the tumor cells derived from mice injected with KP4 cells overexpressing GLS MRs versus tumor cells derived from mice injected with KP4 control cells (mCherry). **g**, Representative immunohistochemistry of vimentin (VIM), ZEB1 and E-cadherin (E-CAD) markers in xenograft tumors derived from KP4 cells transduced with control, OVOL2, HNF1A or OVOL2 + HNF1A (Combo) constructs and collected 14 days post injection. Scale bars, 20 µm.

https://doi.org/10.1038/s41588-026-02714-8



associated mechanistic determinants conserved across virtually all PDA tumors suggests that tumor states are 'quantized' into only a few stable states, rather than a continuum. Such a 'quantum cancer biology' paradigm suggests combination therapies targeting all six PDA states could effectively address plasticity across virtually all patients, negating the need to develop complex and expensive personalized therapies. Two alternative strategies may be possible: combining drugs targeting complementary states; or reprogramming treatment-insensitive states into sensitive ones, thus leveraging, for instance, the new class of mutant KRAS inhibitors. Both approaches could rely on sequential rather than co-occurring administration, thus limiting the potential toxicity of drug combinations.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-026-02714-8.

References

1. Rahib, L. et al. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* **74**, 2913–2921 (2014).
2. Neftel, C. et al. An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* **178**, 835–849.e21 (2019).
3. Elyada, E. et al. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discov.* **9**, 1102–1123 (2019).
4. Collisson, E. A. et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat. Med.* **17**, 500–503 (2011).
5. Moffitt, R. A. et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**, 1168–1178 (2015).
6. Bailey, P. et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52 (2016).
7. Cancer Genome Atlas Research Network. Integrated genomic characterization of pancreatic ductal adenocarcinoma. *Cancer Cell* 32, 185–203.e13 (2017).
8. Puleo, F. et al. Stratification of pancreatic ductal adenocarcinomas based on tumor and microenvironment features. *Gastroenterology* **155**, 1999–2013.e3 (2018).
9. Maurer, C. et al. Experimental microdissection enables functional harmonisation of pancreatic cancer subtypes. *Gut* **68**, 1034–1043 (2019).
10. Rashid, N. U. et al. Purity independent subtyping of tumors (PurIST), a clinically robust, single-sample classifier for tumor subtyping in pancreatic cancer. *Clin. Cancer Res.* **26**, 82–92 (2020).
11. Carstens, J. L. et al. Stabilized epithelial phenotype of cancer cells in primary tumors leads to increased colonization of liver metastasis in pancreatic cancer. *Cell Rep.* **35**, 108990 (2021).
12. Jiang, Z. et al. Tff2 defines transit-amplifying pancreatic acinar progenitors that lack regenerative potential and are protective against Kras-driven carcinogenesis. *Cell Stem Cell* **30**, 1091–1109.e7 (2023).
13. Raghavan, S. et al. Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. *Cell* **184**, 6119–6137.e26 (2021).
14. Chan-Seng-Yue, M. et al. Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. *Nat. Genet.* **52**, 231–240 (2020).
15. Peng, J. et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* **29**, 725–738 (2019).
16. Ligorio, M. et al. Stromal microenvironment shapes the intratumoral architecture of pancreatic cancer. *Cell* **178**, 160–175.e27 (2019).
17. Williams, H. L. et al. Spatially resolved single-cell assessment of pancreatic cancer expression subtypes reveals co-expressor phenotypes and extensive intratumoral heterogeneity. *Cancer Res.* **83**, 441–455 (2023).
18. Hwang, W. L. et al. Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment. *Nat. Genet.* **54**, 1178–1191 (2022).
19. Obradovic, A. et al. Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages. *Cell* **184**, 2988–3005.e16 (2021).
20. Paull, E. O. et al. A modular master regulator landscape controls cancer transcriptional identity. *Cell* **184**, 334–351 (2021).
21. Rajbhandari, P. et al. Cross-cohort analysis identifies a TEAD4–MYCN positive feedback loop as the core regulatory element of high-risk neuroblastoma. *Cancer Discov.* **8**, 582–599 (2018).
22. Aytes, A. et al. Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* **25**, 638–651 (2014).
23. Carro, M. S. et al. The transcriptional network for mesenchymal transformation of brain tumours. *Nature* **463**, 318–325 (2010).
24. Mundi, P. S. et al. A transcriptome-based precision oncology platform for patient-therapy alignment in a diverse set of treatment-resistant malignancies. *Cancer Discov.* **13**, 1386–1407 (2023).
25. Basso, K. et al. Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* **37**, 382–390 (2005).
26. Alvarez, M. J. et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).
27. Laise, P. et al. A model for network-based identification and pharmacological targeting of aberrant, replication-permissive transcriptional programs induced by viral infection. *Commun. Biol.* **5**, 714 (2022).
28. Alvarez, M. J. et al. A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. *Nat. Genet.* **50**, 979–989 (2018).
29. Ding, H. et al. Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nat. Commun.* **9**, 1471 (2018).
30. Curiel-Garcia, A. et al. Ras-dependent activation of BMAL2 regulates hypoxic metabolism in pancreatic cancer. Preprint at *bioRxiv* https://doi.org/10.1101/2023.03.19.533333 (2025).
31. Lin, W. et al. Single-cell transcriptome analysis of tumor and stromal compartments of pancreatic ductal adenocarcinoma primary tumors and metastatic lesions. *Genome Med.* **12**, 80 (2020).
32. Steele, N. G. et al. Multimodal mapping of the tumor and peripheral blood immune landscape in human pancreatic cancer. *Nat. Cancer* **1**, 1097–1112 (2020).
33. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
34. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
35. Qiu, X. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
36. Yuan, P. et al. KRAS/NF-κB/YY1/miR-489 signaling axis controls pancreatic cancer metastasis. *Cancer Res.* **77**, 100–111 (2017).
37. Lin, C. C. et al. Dysregulated Kras/YY1/ZNF322A/Shh transcriptional axis enhances neo-angiogenesis to promote lung cancer progression. *Theranostics* **10**, 10001–10015 (2020).

38. Yin, Q. et al. YB-1 as an oncoprotein: functions, regulation, post-translational modifications, and targeted therapy. *Cells* **11**, 1217 (2022).
39. Sullivan, M. R. et al. Quantification of microenvironmental metabolites in murine cancers reveals determinants of tumor nutrient availability. *eLife* **8**, e44235 (2019).
40. Kapoor, A. et al. Yap1 activation enables bypass of oncogenic kras addiction in pancreatic cancer. *Cell* **158**, 185–197 (2014).
41. Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A. & Williams, R. M. Jr. *The American soldier: Adjustment during Army life* (Princeton Univ. Press, 1949).
42. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
43. Lachmann, A. & Ma'ayan, A. KEA: kinase enrichment analysis. *Bioinformatics* **25**, 684–686 (2009).
44. Hart, T. et al. Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3 (Bethesda)* **7**, 2719–2727 (2017).
45. Palin, K. et al. Contribution of allelic imbalance to colorectal cancer. *Nat. Commun.* **9**, 3664 (2018).
46. Lonsdale, J. et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
47. Richart, L. et al. BPTF is required for c-MYC transcriptional activity and in vivo tumorigenesis. *Nat. Commun.* **7**, 10153 (2016).
48. Hockemeyer, D. et al. A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell* **3**, 346–353 (2008).
49. Califano, A. & Alvarez, M. J. The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nat. Rev. Cancer* **17**, 116–130 (2017).
50. Hong, T. et al. An Ovol2-Zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. *PLoS Comput. Biol.* **11**, e1004569 (2015).
51. Zeisberg, M. & Neilson, E. G. Biomarkers for epithelial–mesenchymal transitions. *J. Clin. Invest.* **119**, 1429–1437 (2009).
52. Son, J. et al. BACH2 inhibition reverses beta cell failure in type 2 diabetes models. *J. Clin. Invest.* **131**, e153876 (2021).
53. Boguski, M. S. & McCormick, F. Proteins regulating Ras and its relatives. *Nature* **366**, 643–654 (1993).
54. Fernández, E. C. et al. Systematic design of combination therapy by targeting master regulators of coexisting diffuse midline glioma cell states. *Nat. Genet.* **58**, 1112–1125 (2026).

Pasquale Laise[1,2,18], Mikko Turunen[1,18], Alvaro Curiel-Garcia[3,4,5,18], Lorenzo Tomassoni[1,2], H. Carlo Maurer[6], Ela Elyada[7,8], Bernhard Schmierer[9], Jeremy Worley[1], Jordan Kesner[1], Xiangtian Tan[1], Ester Calvo Fernandez[1], Yuanqing Xue[1], Yining Chen[1], Kelly Wong[1], Urszula N. Wasko[3], Somnath Tagore[1], Alexander L. E. Wang[1], Sabrina Ge[10,11], Alina C. Iuga[12], Aaron T. Griffin[1], Winston Wong[4,13], Gulam A. Manji[4,5,13], Mariano J. Alvarez[1,2], Faiyaz Notta[10,11,14], David A. Tuveson[7,8], Kenneth P. Olive[3,4,5] ✉ & Andrea Califano[1,4,5,15,16,17] ✉

[1]Department of Systems Biology, Columbia University, New York, NY, USA. [2]DarwinHealth Inc., New York, NY, USA. [3]Division of Digestive and Liver Diseases, Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. [4]Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. [5]Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY, USA. [6]Klinikum rechts der Isar, II. Medizinische Klinik, Technische Universität München, Munich, Germany. [7]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. [8]Lustgarten Foundation Pancreatic Cancer Research Laboratory, Cold Spring Harbor, NY, USA. [9]SciLifeLab and Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Biomedicum Solnavägen 9, Stockholm, Sweden. [10]Princess Margaret Cancer Centre, Toronto, Ontario, Canada. [11]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. [12]Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [13]Division of Hematology and Oncology, Columbia University Irving Medical Center and New York Presbyterian Hospital Herbert Irving Pavilion, New York, NY, USA. [14]PanCuRx Translational Research Initiative, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [15]Biohub, New York, NY, USA. [16]Department of Biomedical Informatics, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. [17]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA. [18]These authors contributed equally: Pasquale Laise, Mikko Turunen, Alvaro Curiel-Garcia. ✉e-mail: kenolive@columbia.edu; ac2248@cumc.columbia.edu

Methods

Experimental procedures and a general overview of the computational methods are described below, while comprehensive technical details, including computational analyses and experimental protocols, are provided in the Supplementary Methods. All human subject research in this study was approved by the Columbia University Institutional Review Board and performed under either protocol AAAS6152 or AAAR4172. Protocol AAAS6152 is further linked to two tissue collection and banking protocols: AAAB2667 and AAAL5871. For the LCM-RNA-seq cohort used to generate the CUMC-E PDA regulatory network in Fig. 3, all samples from the 232 human patients in this cohort (representing 200 PDA tumors and 32 precursor lesion samples) were covered by a Waiver of Consent granted by the Columbia University Institutional Review Board in association with protocol AAAB2667. Of these, samples collected from 228 patients were collected before the National Institutes of Health (NIH) Genomic Data Sharing (GDS) policy deadline of 25 January 2015. Two additional samples were collected after this deadline from patients who did not provide prospective research consent and are now deceased. Samples from the final two patients were collected after this deadline from patients who additionally provided prospective, written informed consent under protocol AAAL5871. For the spatial transcriptomics experiment, all eight subjects provided prospective, written informed consent for research under protocol AAAS6152 through a linkage to the Columbia University Biobank (protocol AAS7370). PDXs were generated under Institutional Review Board protocol AAAR4172. Animal studies were performed under Columbia University IACUC protocol AC-AABS3610. Tumors implanted orthotopically in the pancreas were dissected and subjected to RNA sequencing. The IACUC protocol sets ethical limitations based on a physical and behavioral scoring system. All animal experiments adhered to the ethical limitation scoring system.

Single-cell analysis

To select a dataset for initial discovery experiments, we examined five scRNA-seq datasets[3,14,15,31,32] and one snRNA-seq dataset[18]. Each dataset was analyzed independently to avoid technical batch effects. The snRNA-seq dataset was excluded because of its technical limitations, as it only assessed mRNA within the nucleus, limiting its ability to capture the full transcriptomic profile of the cells. Single-cell analysis was performed on the malignant cells only (Supplementary Methods). First, we transformed gene expression profiles into protein activity profiles using metaVIPER[29], integrating four PDA bulk networks and one PDA single-cell network. All networks were inferred from PDA samples by ARACNe[25,55,56] (Supplementary Methods). Then we performed principal component analysis and calculated the CPVE for each dataset, with a higher CPVE indicating a greater proportion of variance captured by the first principal components (Extended Data Fig. 1e). We also evaluated the entropy of each dataset (Extended Data Fig. 1f). Elevated entropy suggests random or less structured (that is, noisy) data, while a low entropy indicates lower information retained (Extended Data Fig. 1f). The most informative dataset is characterized by a high CPVE and moderate entropy, reflecting a clear and structured signal captured effectively by principal component analysis. From the Elyada dataset, we retained 1,856 malignant cells for the initial discovery phase. Cluster analysis was performed using the Louvain algorithm implemented in the Seurat package[33]; pseudotrajectory was inferred by the Monocle algorithm[57]. Both cluster analysis and pseudotrajectory inference were applied to the protein activity profiles of these cells, resulting in the identification of six cell states. Cross-cohort consistency of the signatures identified in the Elyada dataset was assessed in malignant cells from five additional independent PDA human cohorts, including the CSY-scSet[14], Peng dataset[15], Lin dataset[31], Steele dataset[32], Hwang dataset[18], seven PDA cell lines (CAPAN1, HPAFII, Hs766T, KP4, MIAPACA2, PANC0403, PATU8988S) and a PDA PDX model. Reproducibility was measured by evaluating the conservation of the signatures inferred in the Elyada dataset in each malignant cell, through enrichment analysis using the aREA algorithm[26,58].

Drug perturbation assays

A pooled library and transcriptome sequencing (PLATE-seq) experiment was performed in collaboration with Columbia University's Genome Center. Panc1 and Aspc1 pancreatic cancer cells were cultured in white 96-well tissue culture-treated plates at optimized density, in 100 µl of DMEM media supplemented with 10% FBS and 1% penicillin–streptomycin. After 24 h of incubation, the cells were treated with the following drugs: RAF inhibitors sorafenib, dabrafenib, RAF709, PLX8394 and GDC-0879; MEK inhibitors trametinib, cobimetinib, binimetinib, selumetinib and rafametinib; and ERK inhibitors SCH772984, ulixertinib, AZD0364 and ravoxertinib. Each drug was dosed at the concentration at which the cells were 80% viable after 48 h of treatment. After 24 h of treatment, the medium was replaced with 100 ml of FBS supplemented with 10% DMSO, and the plates were frozen at −80 °C before performing PLATE-seq. A detailed protocol for preparation of the automated PLATE-seq experiment has been previously described[59]. The PLATE-seq FASTQ files were pseudo-aligned to the GRCh38 human transcriptome (mRNA and ncRNA), and gene expression was quantified using Kallisto (v.0.44.0)[60], the tximport package[61] and the biomaRt package[62]. The ShortRead package[63] was used to assess the quality of sequencing data for each sample. The *z*-score method was used to generate differential gene expression signatures of each drug-treated sample with respect to the DMSO-treated samples. Protein activity profiles were computed using metaVIPER.

10x Genomics Xenium data

Eight fresh–frozen OCT-embedded tissue blocks of untreated samples from patients with pancreatic cancer were obtained from the Molecular Pathology Shared Resource at Columbia University. Tissue sections of 5 µm were sectioned and mounted onto Xenium slides (10x Genomics) and processed following the manufacturer's instructions. Imaging-based spatial transcriptomic profiling was performed following the manufacturer's standard workflow. To identify PDA malignant cell states, in addition to the Xenium Human Multi-Tissue and Cancer Gene Expression panel (n = 377 genes), we generated a custom panel of 100 genes predictive of the three PDLs and the MAPK-high cell state. Specifically, we selected 20 genes for each of the three PDLs and 20 genes predictive of MAPK pathway activation. An additional set of 20 stroma-related genes was included as a negative control. The Xenium analyzer performed iterative rounds of hybridization, washing and imaging to decode the spatial location and identity of each mRNA molecule. Raw image data generated by the Xenium analyzer was processed using the Xenium Onboard Analysis package (10x Genomics). The Xenium onboard analysis processes raw instrument images by first segmenting cell nuclei using DAPI stain, then decoding fluorescent probes to build a precise 3D map of individual RNA transcripts, and finally uses these transcript locations to refine the cell boundaries before generating the final data outputs. For cell type annotation, we used a previously published scRNA-seq dataset[15] as a reference. We generated cell type signatures from the authors' annotation by computing cell type centroids on normalized data. These signatures were then used for initial cell type annotation. Following initial cell type annotation, all cells identified as malignant were re-analyzed using enrichment analysis for the PDL-predictive and MAPK-predictive gene sets.

LCM dataset (CUMC-E)

Freshly frozen tissue samples were obtained from patients who underwent surgical resection at the Pancreas Center at Columbia University Medical Center as previously described[9]. Before surgery, all patients had given surgical informed consent, which was approved by the institutional review board. Immediately after surgical removal, the specimens were cryopreserved, sectioned and microscopically

evaluated by the Columbia University Tumor Bank (IRB AAAB2667). Suitable samples were transferred into OCT medium (Tissue Tek) and snap-frozen in a 2-methylbutane dry ice slurry. The tissue blocks were stored at −80 °C for later processing. H&E-stained sections of frozen PDA samples from the Tumor Bank were initially screened to confirm diagnosis, and overall sample RNA quality was assessed by the Pancreas Center-supported Next Generation Tumor Banking program using gel electrophoresis, with samples exhibiting high RNA quality used for subsequent analyses, including LCM, RNA-seq and gene expression quantification. LCM-RNA-seq was performed as previously described[9,64]. In brief, cryosections of OCT-embedded tissue blocks were transferred to PEN membrane glass slides and stained with cresyl violet acetate. Adjacent sections were H&E-stained for pathology review. LCM was performed on a PALM MicroBeam microscope (Zeiss), collecting at least 1,000 cells per compartment. RNA was extracted, and libraries were prepared using the Ovation RNA-seq System V2 kit (NuGEN). Libraries were sequenced to a depth of 30 million, 100 bp, single-end reads on an Illumina HiSeq 2000 platform. Computational analyses of the RNA-seq data for this dataset are described in the Supplementary Methods.

Single-cell lineage tracing

We modified the lentiviral vector also used in our TF overexpression assays (modified Tet-O-FUW EGFP-puro; Addgene, 30130) by using mCherry instead of EGFP, and cloned a unique molecular identifier (UMI-barcode; UMI-BC), which consists of a random 28-mer, 200 bp upstream of the lentiviral 3′ long terminal repeat region. This way, the UMI-BC site will be polyadenylated, and the barcode can be specifically amplified at the end of scRNA-seq library preps. In total, we cloned a plasmid library that contains approximately six million distinct UMI-BCs (Supplementary Table 24). The basic principle for lineage tracing experiments was similar to that described in previous work[65]. In brief, the UMI-BC-containing lentiviral constructs were transduced into KP4, PATU8988S and Hs766T cell lines (to approximately 15–30 million cells) with a MOI of <0.1, followed by puromycin selection. After all the non-transduced cells were dead, we seeded 8,000 cells per well, followed by incubation, which lasted approximately one to two population doublings (daughter cells were created for each UMI-BC-containing cell). At this point, we performed the first Chromium run (T_0) so that 50% of the cells were used for this initial time point run, and the remaining 50% of the cells were cultured further for the second time point (T_1). The T_1 Chromium run was performed approximately ten population doublings later. In addition, we conducted a lineage tracing assay for the KP4 cell line to compare the effects of treatment with 100 nM trametinib versus DMSO 24 h before the assay endpoint. During the 10x Chromium library preparations, we specifically amplified the UMI-BC-containing cDNA, similarly to that described in the section pertaining to the pooled TF overexpression assay (refer to Supplementary Methods for further details) using ORF_BC_amplif_oligo_F and ORF_BC_amplif_R-primers, and we spiked the product in the final NGS library at 10% total amounts. For both time points and all cell lines, the entire pool of UMI-BCs was processed using UMI-tools[66] to collapse clusters of UMI-BCs with fewer than four mismatches. Through this procedure, a reference list of all detected UMI-BCs was generated, and it allowed us to identify all the cells associated with one unique UMI-BC. Single-cell sequencing data of the PATU8988S, KP4 and Hs766T cell lines were processed using the Cell Ranger pipeline (v.3.0.2) from 10x Genomics (https://www.10xgenomics.com). FASTQ files were aligned using the human genome as a reference (v.GRCh38-2020-A). To avoid differences in sequencing depth across cells sequenced in different runs, which could significantly affect gene detection, the UMI counts of PATU8988S, KP4 and the other seven PDA cell lines were downsampled to make the probability distributions of the number of UMIs per cell comparable across different experiments. Single-cell gene expression profiles generated from PATU8988S and KP4 cell lines were normalized to log(counts per million) and subsequently transformed into gene expression signatures using the gene expression centroid of the seven PDA cell lines (as detailed in the single-cell analysis section and Supplementary Methods) as a reference (z-score transformation). Gene expression signatures were transformed into protein activity signatures using metaVIPER. For the trametinib-induced lineage tracing experiment in KP4 cells, we computed differential protein activity by comparing cells sharing the same barcode across the two time points (early and late) in both trametinib-treated and control (DMSO) conditions.

CRISPR KO and CRISPRi screening

sgRNAs containing lentiviruses were transduced into Cas9/dCas9-expressing PDA cell lines in duplicates (in the presence of 8 μg ml^{-1} polybrene), at an estimated MOI of 0.2 to 0.3. After 24 h, the lentivirus-containing media was removed, cells were washed with PBS and puromycin-containing media (2 μg ml^{-1}) was added to the cells for 48–72 h until all control cells (not virus-infected) were dead. Half of the cells were collected at this time point (day 0) and sequenced to assess sgRNA representation baseline. Cells were then maintained at >1,500 cells per guide throughout the screens (see Supplementary Methods for more details) and finally collected at 33 days (day 33) post-puromycin selection, followed by sgRNA amplicon-seq to assess gene essentiality. FASTQ files were analyzed with MAGeCK (v.0.5.6)[67], using RRA and total read count normalization, with default settings. Each replicate was analyzed independently by comparing the relative abundance of the gRNAs at day 33 against day 0. CRISPR KO and CRISPRi essentiality signatures for each replicate were computed by transforming the P value to a z-score (the sign was inferred from fold change). GLS and MOS essentiality signatures were computed by integrating the essentiality z-scores of the three GLS-enriched and three MOS-enriched cell lines, using Stouffer's method[41]. Finally, a differential GLS versus MOS essentiality signature was computed by comparing essentiality in MOS- versus GLS-enriched cells. To define a subtype-independent essentiality signature, we integrated the essentiality signatures as assessed by CRISPR KO across all cell lines, independent of cell state, using Stouffer's z-score method[41].

TF overexpression assay (PLATE-seq)

Full-length ORF clones for the top eight GLS MRs from bulk-profile analysis were ordered from CloneID (Harvard Medical School) and cloned into a modified Tet-O-FUW lentiviral expression vector (Addgene, 30130), which included the puromycin resistance gene. mCherry and EGFP ORFs were used as negative controls in the assay. All clones were sequence-verified. For each ORF, we introduced a unique 20 bp barcode sequence located 200 bp upstream of the lentiviral 3′ long terminal repeat region, as previously reported[68] (Supplementary Table 24). This produces a polyadenylated transcript, which contains the barcode proximal to its 3′ end. All viruses were produced, and viral titers were measured individually for each virus. ORFs containing lentivirus were transduced individually into KP4 morphogenic cell lines at an MOI of 2 in triplicate (six-well format in the presence of polybrene). In a second triplicate set, individual lentiviral ORFs were co-transduced with M2rtTA (FUW-M2rtTA; Addgene, 20342), a tetracycline-inducible transcriptional amplifier, allowing monitoring of MR overexpression at higher and lower levels[48]. At 24 h following viral transduction, media was changed, and puromycin (2.5 μg ml^{-1}) and doxycycline (0.6 μg ml^{-1}) were added to the cells, followed by a 5 day incubation period before total RNA was collected by the Direct-zol RNA MiniPrep Plus kit (Zymo Research). Then, 69 RNA-seq profiles, corresponding to 23 different conditions in triplicate, were generated by PLATE-seq[59] using 100 ng of total RNA as a template in each well. After all the overexpression assays were completed, we performed ORF construct verifications with Oxford Nanopore NGS, which revealed that the FOXA3 and ESRRA constructs had gene duplications, which were not seen with Sanger

sequencing. We do not expect that these gene duplications would have any effects on the final results. Single-end PLATE-seq reads were pseudo-aligned to the GRCh38 transcriptome (mRNA and ncRNA) and quantified using Kallisto (v.0.44.0)[60], with sequence-specific bias correction. Transcript-level counts were aggregated by Entrez IDs and compared between unperturbed GLS-enriched cells (PATU8988S and HPAFII) and unperturbed MOS-enriched cells (KP4). The corresponding gene expression signature was transformed into a protein activity signature and compared to the GLS–MOS protein activity signature (TCGA–ICGC signature) inferred from patient-derived profiles, to ensure conservation of protein activity signatures. In brief, this was performed by first computing a differential gene expression signature between unperturbed GLS and MOS-enriched cell lines using Student's *t*-test, as implemented in the VIPER package[26], on the normalized gene expression profiles. Then, the metaVIPER approach was used to transform this differential gene expression signature into a protein activity signature. Finally, the conservation between this protein activity signature (inferred by comparing unperturbed GLS versus MOS-enriched cell lines) and the GLS–MOS protein activity signature computed from patient-derived samples was assessed by a two-sided aREA test[26]. Specifically, this was done by assessing the conservation of the 50 most differentially activated and 50 most differentially inactivated proteins of the TCGA–ICGC signature in the protein activity signature inferred from the cell lines. This signature was then used to assess reprogramming by enrichment analysis (see Supplementary Methods).

Pooled TF overexpression assay (scRNA-seq)

The same barcoded constructs described in the PLATE-seq overexpression method section, representing the top eight GLS MR proteins from bulk-profile analysis (and mCherry as a negative control), were used in the single-cell overexpression assays. ORF viruses were pooled into two viral pools, such that on average, two to three ORFs were randomly transduced into every single cell (MOI = 0.288 for each virus). The M2rtTA construct was added to the second pool to increase the transcriptional output of the ORFs. KP4 cells were transduced with the two viral pools in a six-well format in the presence of polybrene. Media was changed 24 h post infection, and puromycin (2.5 μg ml^{-1}) and doxycycline (0.6 μg ml^{-1}) were added to the cultures. Cells were incubated for a total of 11 days before trypsinization, addition of MultiSeq barcodes[69] and 10x Chromium library preparation and sequencing. Non-transduced (control) KP4, HPAFII and PATU8988S cells were also included at this stage to represent the pre-treatment MOS-enriched cell state (KP4), while the average profile of non-transduced (control) GLS-enriched cell lines (HPAFII and PATU8988S) was used as a proxy for the desired GLS endpoint state. All cells (normal KP4, HPAFII, PATU8988S and KP4 transduced pools, ±M2rtTA) were MultiSeq-barcoded and mixed before the Chromium run. After the Chromium run, the cDNA was amplified with 1 μl of 2.5 μM MultiSeq additive primer added to the cDNA amplification mastermix. After this step, the material was divided into three portions: the whole transcriptome, the MultiSeq barcode and the ORF barcode. The MultiSeq barcode and ORF barcode portions were amplified with specific primers (MULTI-seq_TruSeq RPIX & MULTI-seq_Universal I5 for MultiSeq and ORF_BC_amplif_oligo_F and ORF_BC_amplif_R for ORF barcode) and spiked in the final NGS library at 1% and 10% total amounts, respectively. For the final analysis, the MultiSeq barcode data were not used (Supplementary Methods). For computational analyses, see Supplementary Methods.

Mouse strains and orthotopic injections. Immunocompromised mice (4–6 weeks old) were obtained from Jackson Laboratories (007850). Two and five million human lentivirally transduced PDA cells (KP4 and HPAFII) were orthotopically injected. For the CRISPR screen, tumor establishment, subsequent growth and overall health status were monitored by ultrasound (VEVO 3100, VisualSonics) weekly for 29 days; the mice were then killed, and genomic DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's protocol. For the OVOL2–HNF1A transdifferentiation experiments, mice carrying OVOL2, mCherry, HNF1A and both OVOL2 and HNF1A were injected, and single-cell isolation was performed after 14 days as described in Supplementary Methods.

CRISPR KO library design (in vivo). For in vivo CRISPR KO screening, we designed the target gene list to include 32 GLS and 61 MOS MRs identified by both bulk and single-cell analyses. All these genes were targeted with five sgRNAs per gene. For assay controls, we added six core-essential genes as strong positive controls, 100 negative intergenic controls and five non-targeting controls. For guide designs, we used CRISPick[70,71]. The candidates' essential MRs for the in vivo CRISPR were first selected by identifying the most conserved, subtype-specific, differentially activated proteins between patients and cell lines previously selected as models for GLS and MOS cell states. This was done by GSEA. Specifically, we selected the leading edges from the enrichment of the top 200 differentially active proteins of each cell line in the TCGA–ICGC signature derived from patients. Next, we performed another GSEA to identify, among the most conserved differentially active proteins between patients and cell lines, those that were enriched in the subtype-specific essential signature, defined by the in vitro CRISPR KO and CRISPRi screens, in each cell line. Then the most conserved differentially active proteins, as identified by metaVIPER and enriched in the essential genes, defined by the in vitro CRISPR screens, were selected for the in vivo CRISPR screen. For the MOS lineage, the highest enrichment was detected in the KP4 essential gene signature. For the GLS lineage, the highest MR enrichment was detected in HPAFII essential genes. After transducing the sgRNAs into KP4 and HPAFII cells and puromycin selection, cells were injected into seven immunocompromised mice, and tumor formation was monitored for 29 days, followed by genomic DNA extraction and NGS of the sgRNAs. Analysis of CRISPR data, including the identification of the essential signatures, was performed using Mageck software[67], following the same procedure described for the analysis in vitro screen.

In vivo MR overexpression-mediated transdifferentiation

In vivo overexpression assays were performed similarly to in vitro overexpression assays. However, only constructs for OVOL2, HNF1A, mCherry and the OVOL2–HNF1A combination were used in vivo. These viral constructs were transduced into KP4 cells in arrayed fashion (approximate MOI of 1–2), followed by puromycin selection and pooling the samples together in the following ratios based on their in vitro growth rate (OVOL2 50%, HNF1A 16.6%, mCherry 16.6% and OVOL2–HNF1A 16.6%). Pooled cells were then injected into athymic mouse models (three mice in total); 14 days later, the animals were killed, followed by scRNA-seq for the pooled samples. Similarly, as with the pooled in vitro assays, the ORFs for any given cell were identified using ORF barcodes, which were amplified from the amplified cDNA by using specific primers for ORF barcodes (same as with in vitro assays). M2RTTA was not used for in vivo overexpression assays. To assess MR-mediated transdifferentiation in vivo, we compared the tumor cells derived from mice injected with perturbed KP4 cell lines expressing HNF1A, OVOL2 and the combination of OVOL2–HNF1A against tumor cells derived from the mice injected with KP4 control cells (mCherry). This was done by comparing the gene expression centroid of perturbed cells against the gene expression centroid of control cells, and then applying metaVIPER. MetaVIPER generated three differential protein activity signatures: one signature for HNF1A overexpression, one signature for OVOL2 and one for the combination of OVOL2–HNF1A. These signatures were then interrogated with MOS and GLS MR signatures from bulk-profile analysis by two-sided GSEA to assess reprogramming.

Statistics and reproducibility

No statistical method was used to predetermine sample size. No data were excluded from the analyses. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

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

Data availability

Sequence-level data for LCM-RNA-seq samples used to generate the CUMC regulatory network in Fig. 3 are available in dbGAP (Study ID: phs004767). This includes samples from the 228 patients collected before the NIH GDS policy deadline of 25 January 2015 and the two patients collected after this deadline who additionally provided prospective research consent. To maintain compliance with NIH GDS policy, sequence-level data are not available for the two deceased patients covered by the Waiver of Consent whose samples were collected after 25 January 2015. Read count data for samples from all 232 patients are available in GEO (GSE143584). For the spatial transcriptomics experiment, all eight subjects provided prospective, written informed consent for research under protocol AAAS6152 through linkage to the Columbia University Biobank (protocol AAS7370).
CRISPR data and RNA-seq data related to cell lines, TF overexpression assays, single-cell xenograft lineage models and single-cell overexpression assays are available in GEO under accession GSE161369. scRNA-seq profiles from the Elyada dataset are accessible through NCBI dbGaP (phs001840.v1.p1), while profiles from the Chan-Seng-Yue dataset are available through the EGA database (EGAD00010001811; sample IDs: 100070, 91610, 91706, 95092 and 96460). Additional datasets include bulk RNA-seq of six PDA cell lines (GSE159918), pooled TF overexpression assay using scRNA-seq (GSE159921), single-cell data from a patient-derived xenograft model of PDA (GSE160977), CRISPR screening in PDA cell lines (GSE161365), TF overexpression assay in PDA cell lines using bulk RNA-seq (GSE161368), CRISPRi screening in PDA cell lines (GSE167157), in vitro drug screening of RAF/MEK/ERK inhibitors in two PDA cell lines (GSE252002), scRNA-seq of human PDA cell lines in normal versus serum-free media (GSE266034), single-cell barcoded lineage tracing in Hs766T (GSE268077), KP4 (GSE268229) and PATU8988S (GSE268441) PDA cell lines, in vivo ectopic expression of OVOL2 and HNF1A in PDA (GSE268388), scRNA-seq of seven PDA cell lines (GSE270117), in vivo pooled CRISPR KO screens in PDA cell lines (GSE270853), in vitro arrayed OVOL2 overexpression in PDA cell lines using bulk RNA-seq (GSE280160), single-cell barcoded lineage tracing in KP4 PDA cell line pre-trametinib and post-trametinib treatment (GSE292402), scRNA-seq of human PDA cell lines after trametinib treatment (GSE310584) and multiome profiles of KP4 cells (GSE312315). Source data are provided with this paper.

Code availability

No custom code was developed for this study; all software packages used are listed in the Methods and Supplementary Information.

References

55. Margolin, A. A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7**, S7 (2006).
56. Lachmann, A., Giorgi, F. M., Lopez, G. & Califano, A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* **32**, 2233–2235 (2016).
57. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
58. Alvarez, M. J. et al. Unbiased assessment of H-STS cells as high-fidelity models for gastro-enteropancreatic neuroendocrine tumor drug mechanism of action analysis. Preprint at *bioRxiv* https://doi.org/10.1101/677435 (2019).
59. Bush, E. C. et al. PLATE-Seq for genome-wide regulatory network analysis of high-throughput screens. *Nat. Commun.* **8**, 105 (2017).
60. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
61. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
62. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
63. Morgan, M. et al. ShortRead: a Bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* **25**, 2607–2608 (2009).
64. Maurer, H. C. & Olive, K. P. Laser capture microdissection on frozen sections for extraction of high-quality nucleic acids. *Methods Mol. Biol.* **1882**, 253–259 (2019).
65. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, eaaw3381 (2020).
66. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
67. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
68. Parekh, U. et al. Mapping cellular reprogramming via pooled overexpression screens with paired fitness and single-cell RNA-sequencing readout. *Cell Syst.* **7**, 548–555.e8 (2018).
69. McGinnis, C. S. et al. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).
70. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
71. Sanson, K. R. et al. Optimized libraries for CRISPR–Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, 5416 (2018).

Acknowledgements

This work was supported by a Clinical Translational Program Grant from the Lustgarten Foundation for Pancreatic Cancer Research (K.P.O.), a Precision Medicine Pilot Award from the Irving Institute for Clinical and Translational Research (K.P.O.), an NCI PSRC grant (U01CA274312), the Columbia University Digestive and Liver Disease Center (P30DK132710), NCI Research Centers for Cancer Systems Biology Consortium U54CA209997 and U54CA274506 Center for Cancer Systems Therapeutics (CaST, to A.C.), NIH Shared Instrumentation Grants (S10 OD012351, S10 OD021764 and S10OD032433 to A.C.) and NCI Cancer Center Support Grant (P30 CA013696) for the Herbert Irving Comprehensive Cancer Center (including the High-Throughput Sequencing, Single-Cell Analysis, OPTIC, Molecular Pathology and Database Shared Resources). Single-cell studies were supported by the NCI OIA Award (R35 CA197745 to A.C.). Funding was also provided by The Pancreas Center at Columbia/NY Presbyterian Hospital (K.P.O.), by the Sigrid Juselius Foundation (M.T.) and by a philanthropic gift from the family and friends of the late Muzzi Mirza. We acknowledge support from the Swedish National Infrastructure for Computing (SNIC; project SNIC 2017-7-265) and the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). A.C.G. received support from the Charles H. Revson Senior Fellowship in Biomedical Science

(grant no. 22-22). We thank G. Valenti for his comments on single-cell subtypes. We acknowledge the ORC MBMG core unit (Columbia University) for in vivo CRISPR KO screen genotype verification (grant 5P30DK026687).

Author contributions

A.C., K.P.O. and P.L. conceived the project. P.L. started the project, designed, performed and oversaw the computational analyses. M.T. designed, performed and oversaw the experimental analyses. A.C.-G., H.C.M., E.E., B.S., J.W., J.K., X.T., E.C.F., Y.C., K.W., U.N.W., S.G. and F.N. contributed to experimental execution, data acquisition and data generation. L.T., M.J.A., S.T., A.L.E.W., Y.X. and A.T.G. contributed to the computational analyses. A.C.I. reviewed histopathology for PDA samples in the CUMC cohort. G.A.M. and W.W. aided in the curation of human outcomes data. A.C., K.P.O. and P.L. wrote the paper with feedback from A.C.G., C.H.M., M.T. and M.J.A. K.P.O., D.A.T. and A.C. supervised the study.

Competing interests

A.C. is founder, equity holder and consultant of DarwinHealth, a company that has licensed some of the algorithms used in this paper from Columbia University (Patent Nos. 62/211,373 and 62/211,562; authors: A. Califano and M. Javier Alvarez). P.L. is Senior Director of Single-Cell Systems Pharmacology at DarwinHealth. P.L. worked on this project while he was an Associate Research Scientist at Columbia University and completed it after he joined DarwinHealth. L.T. is a full-time employee at DarwinHealth. M.J.A. is CSO and equity holder of DarwinHealth. Columbia University is also an equity holder in DarwinHealth. US patent numbers 10,777,299 and 10,790,040 have been awarded related to this work, assigned to Columbia University. All other authors declare no competing interests.

Additional information

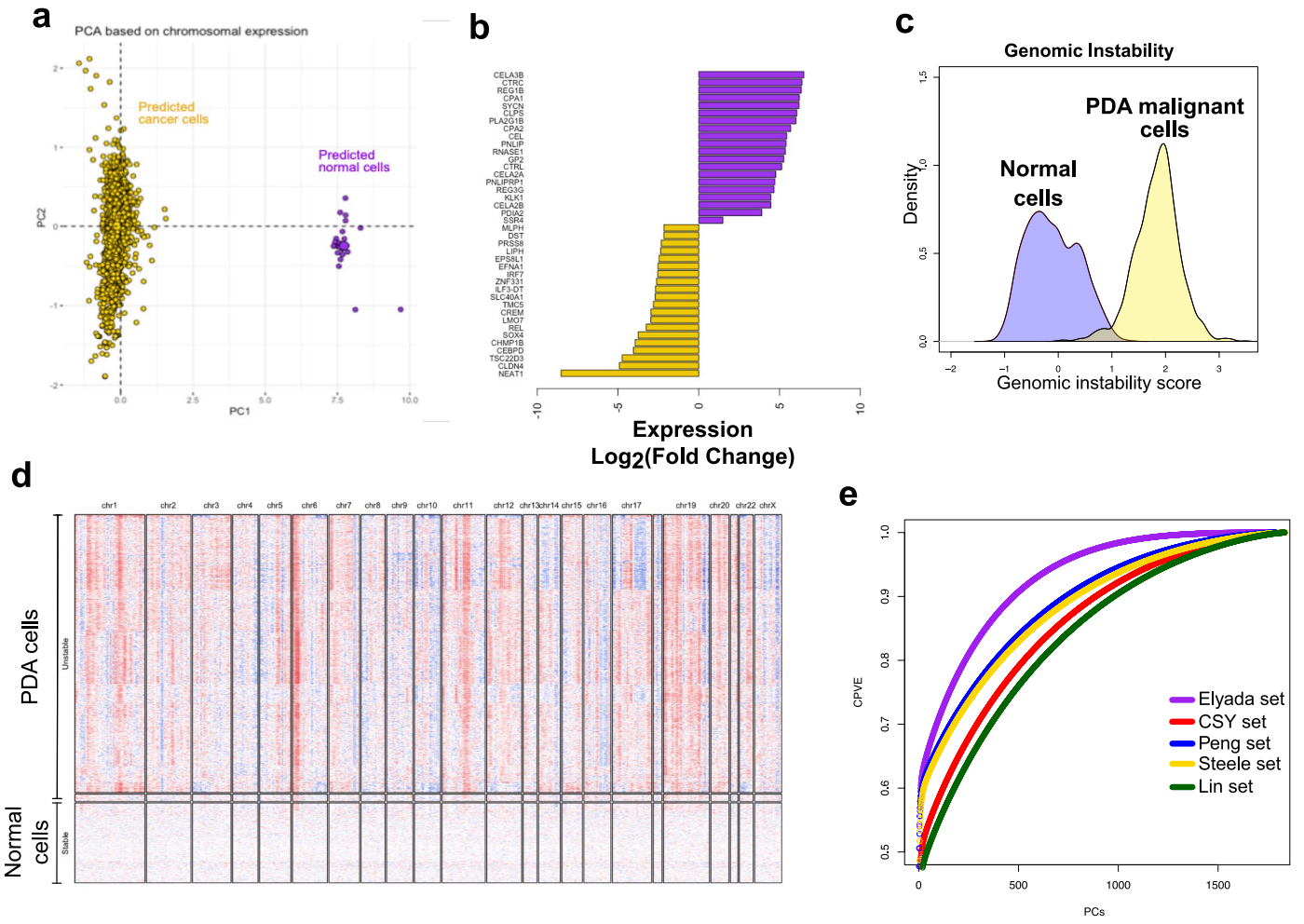
Extended data is available for this paper at https://doi.org/10.1038/s41588-026-02714-8.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-026-02714-8.

Correspondence and requests for materials should be addressed to Kenneth P. Olive or Andrea Califano.

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

https://doi.org/10.1038/s41588-026-02714-8

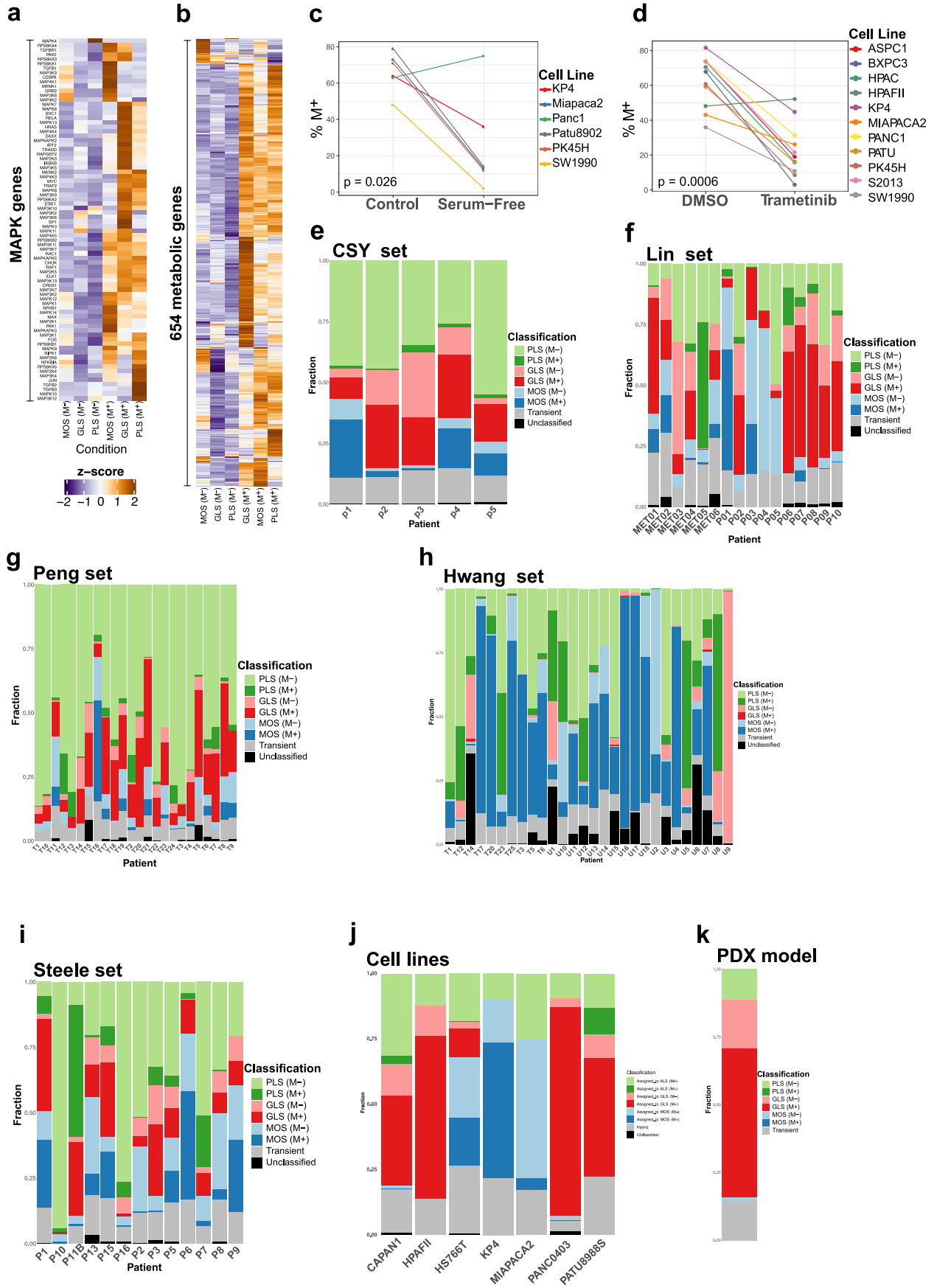


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s41588-026-02714-8

Extended Data Fig. 1 | Selection of malignant cells and single-cell datasets. **a)** PCA plot of PDA epithelial single cells from Elyada dataset based on chromosomal gene expression analysis effectively distinguishes aberrant ploidy tumor cells (yellow) vs. normal-ploidy cells (purple). **b)** Differentially expressed genes between non-transformed (normal) and transformed (cancer) single cells, as predicted by PCA of chromosomal expression. **c)** Density plot showing genomic instability scores between predicted PDA cancer cells by PCA on chromosomal expression compared to non-transformed pancreatic cells (Supplementary Methods). **d**) Heatmap showing CNVs and genomic instability, as inferred from the gene expression profiles of PDA cells compared to non-transformed pancreatic cells. Red indicates increased expression, blue indicates decreased expression**. e)** Cumulative Variance Explained (CPVE) plot. The CPVE plot shows the proportion of total variance captured as a function of the number of principal components (PCs). Each line represents one dataset. **f)** Barplot showing the entropy (nats) of each data set. Dashed line represents the median entropy across all the data sets. **g)** Box plots showing the distribution of silhouette scores of individual cells clustered by their metaVIPER-inferred protein activity profiles and by gene expression at different resolution values of Louvain algorithm.

https://doi.org/10.1038/s41588-026-02714-8

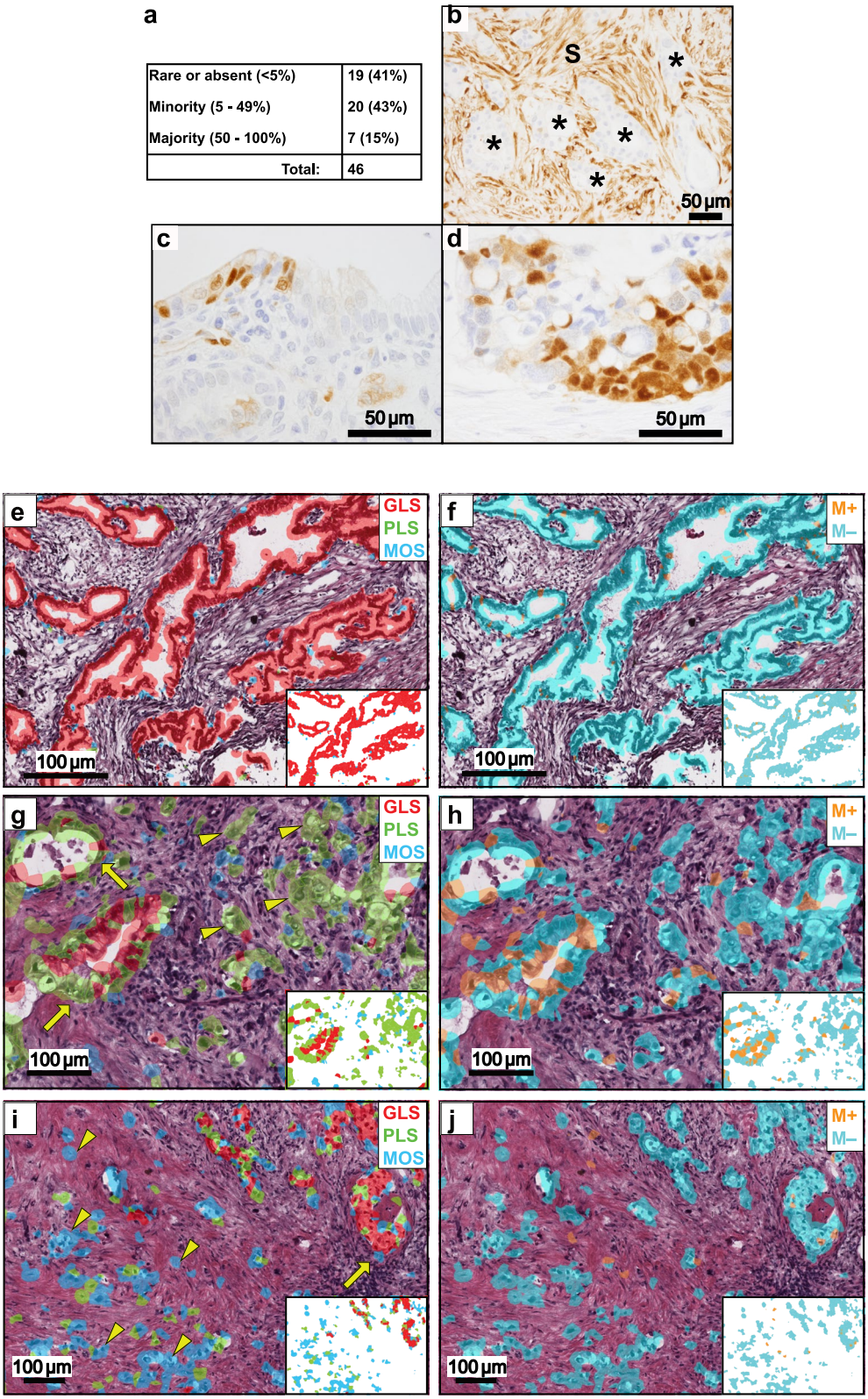


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41588-026-02714-8

Extended Data Fig. 2 | MAPK signaling and cell state conservation across single-cell datasets. a) Heatmap showing differential expression of MAPK pathway–associated genes (Biocarta database) across cell states. Expression is shown as z-scores of pseudobulk profiles, computed by averaging normalized gene expression (log_2(CPM + 1)) of individual cells within each cell state (cluster). **b)** Heatmap showing differential expression of genes associated with metabolic pathways (Reactome database) across cell states. **c)** Plot showing the percentage of M^+ and M^- cells across six PDA cell lines after switching from recommended (control) to serum-free media for 4 h. Cells were profiled by scRNA-seq, converted to protein activity profiles, and classified as M^+ or M^- based on enrichment of an experimentally inferred MAPK protein activity signature. P-values were calculated using a two-sided Student's *t*-test on log-transformed data. **d)** Plot showing the percentage of M^+ and M^- cells across ten PDA cell lines after Trametinib treatment relative to DMSO control. Cells were profiled by scRNA-seq, converted to protein activity profiles, and classified as M^+ or M^- based on enrichment of an experimentally inferred MAPK protein activity signature. P-values were calculated using a two-sided Student's *t*-test on log-transformed data. **e)** Stacked bar chart showing the fraction of PDA cell states in CSY single-cell dataset, inferred by enrichment analysis of cell state–specific protein activity signatures (25 most activated and 25 most inactivated) using the two-sided aREA test. **f)** Stacked barchart showing the fraction of PDA cell states in Lin dataset. **g)** Stacked barchart showing the fraction of PDA cell states in Peng dataset. **h)** Stacked barchart showing the fraction of PDA cell states in Hwang dataset. **i)** Stacked barchart showing the fraction of PDA cell states in Steele dataset. **j)** Stacked barchart showing the fraction of PDA cell states in PDA cell lines. **k)** Stacked barchart showing the fraction of PDA cell states in a PDX model.

https://doi.org/10.1038/s41588-026-02714-8

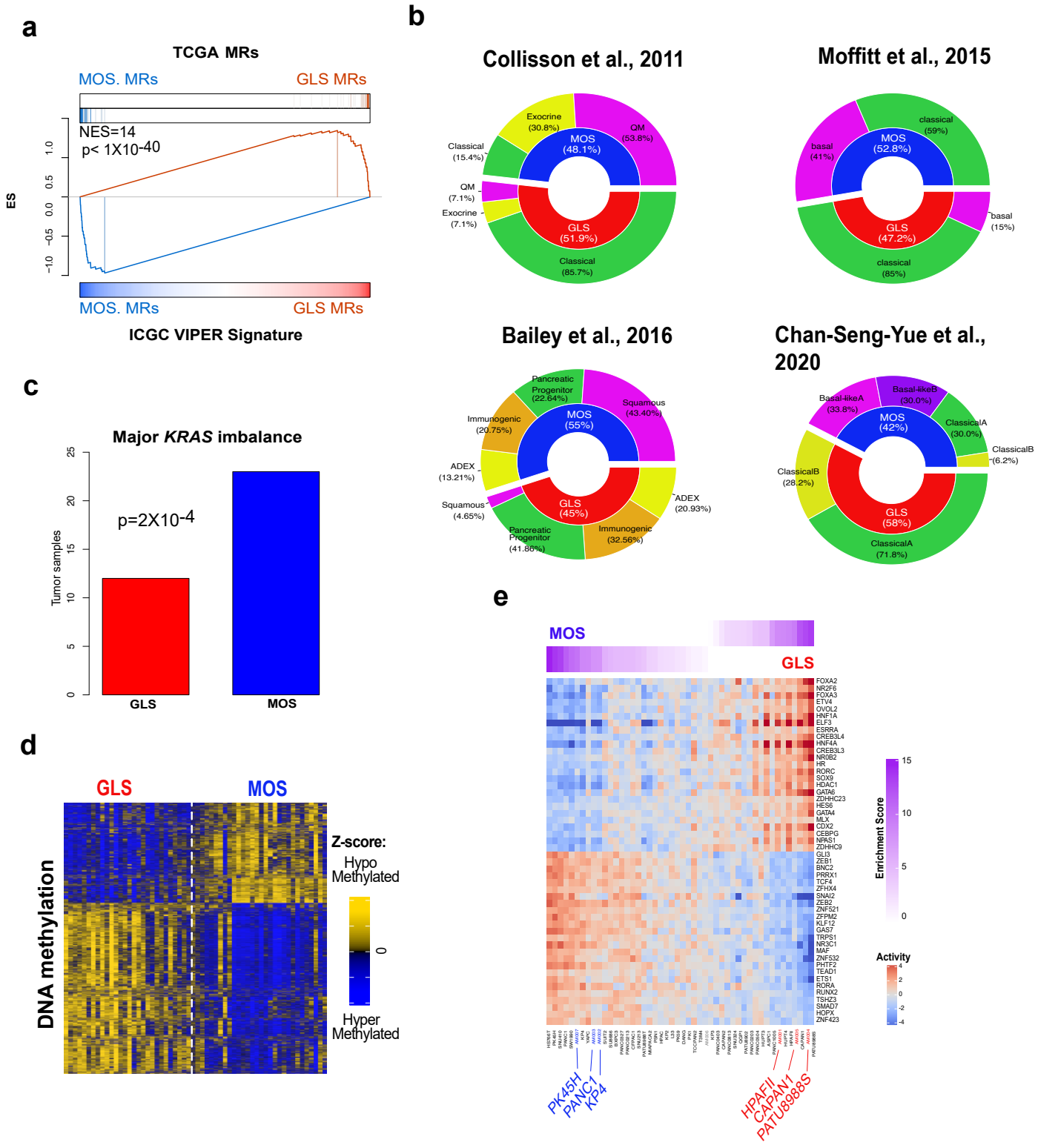


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41588-026-02714-8

Extended Data Fig. 3 | pERK staining and spatial architecture of PDA cell states. a) Table showing blinded scoring of percent pERK positive epithelial cells in a series of 46 human PDA samples from surgical resections. **b)** Example of absent pERK staining in malignant epithelia (*) despite positive stromal cells (S). **c)** Rare positive cells intermixed in a largely negative malignant epithelium. **d)** Malignant epithelium with a majority of pERK positive cells. Scale bars, 50 µm. **e)** Analysis of cell states by Xenium spatial transcriptomics. Eight human PDA samples were analyzed with a custom probe set designed to distinguish PDA cell states. Malignant cells were classified for their PDL states (**e,g,i**) and MAPK sub-states (**f,h,j**). Cell state maps are shown in inset. (**e,f**) Some tumors were dominated by one state, such as the GLS cells, in this tumor (**e**, red), which formed distinct, well-differentiated glands. Rare MOS cells (**e**, blue) were interspersed along the basolateral surface of the glandular structures, which were predominantly in the M^- state (**f**, teal). (**g,h**) In some tumors, PLS cells were found both in glandular structures (**g**, arrows) and invading the stroma (triangles) whereas MOS cells were found predominantly in stroma. This tumor exhibited clusters of M^+ cells in some structures (**h**, orange). (**i,j**) MOS cells (**i**, blue) were common in poorly differentiated tumors and were found throughout the stroma (**i**, triangles) as well as in association with the basolateral surface of residual glandular structures (**i**, arrow). Consistent with our single cell data, the PDL states of cells did not determine the MAPK state, with many of the EMT-like MOS cells found in the M^- state (**j**, teal). Scale bars, 100 µm.

https://doi.org/10.1038/s41588-026-02714-8

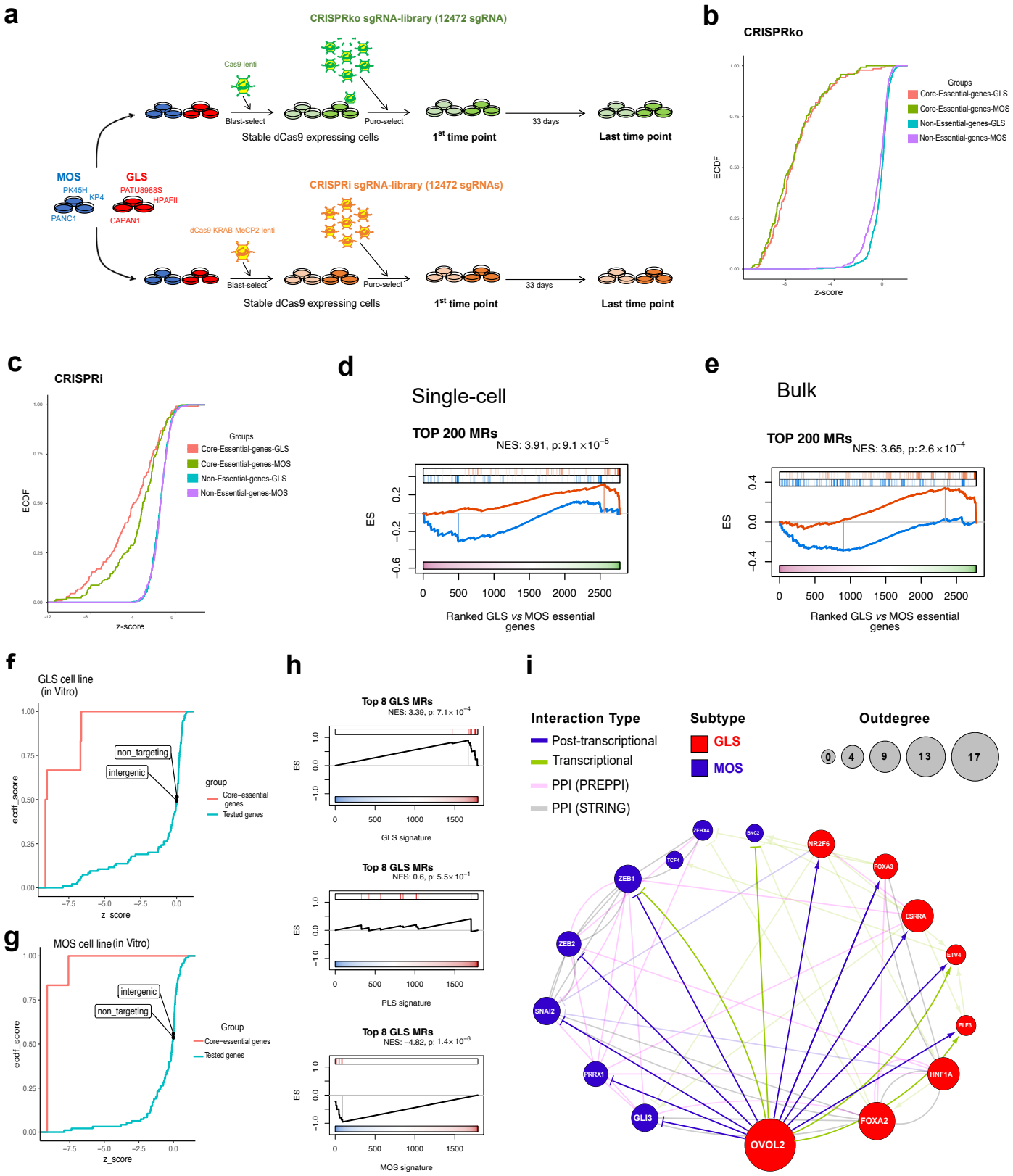


Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s41588-026-02714-8

Extended Data Fig. 4 | Bulk signatures conservation, *KRAS* imbalance and epigenetic differences. a) GSEA plot showing the enrichment of TCGA MRs (50 most overactivated and 50 most inactivated proteins) in the ICGC protein activity signature (p-value and NES were estimated by GSEA two-sided, with 1,000 permutations). **b)** PieDonut charts showing overlap between protein activity-based Lineage and Morphogenic classification and previously published classification schemes. **c)** Barplot showing the enrichment of *KRAS* imbalance in Morphogenic patients, in the CSY_{lcm} dataset. **d)** Heatmap showing differentially methylated sites in GLS vs. MOS TCGA samples (n = 58). **e)** Heatmap showing the enrichment of the 50 most differentially activated proteins (25 most activated and 25 most inactivated), as inferred by metaVIPER analysis of GLS vs. MOS subtype samples, integrated across TCGA and ICGC cohorts, in PDA cell lines from the Cancer Cell Line Encyclopedia (CCLE). The three cell lines labeled in red and blue were selected as GLS- and MOS-specific models, respectively, for a pooled CRISPR/Cas9 screen to validate predicted MR proteins. They were re-sequenced to ensure fidelity compared to their CCLE profiles. Specifically, re-sequenced Lineage cell lines HPAFII, CAPAN1, and PATU8988S were labeled as AM01, AM005, and AM004, respectively, while re-sequenced Morphogenic cell lines PK45H, PANC1, and KP4 were labeled as AM007, AM003, and AM002, respectively. The re-sequenced PANC0403 cell line (AM006) was identified as neither Lineage nor Morphogenic and was thus selected as control for the cell line selection.

https://doi.org/10.1038/s41588-026-02714-8

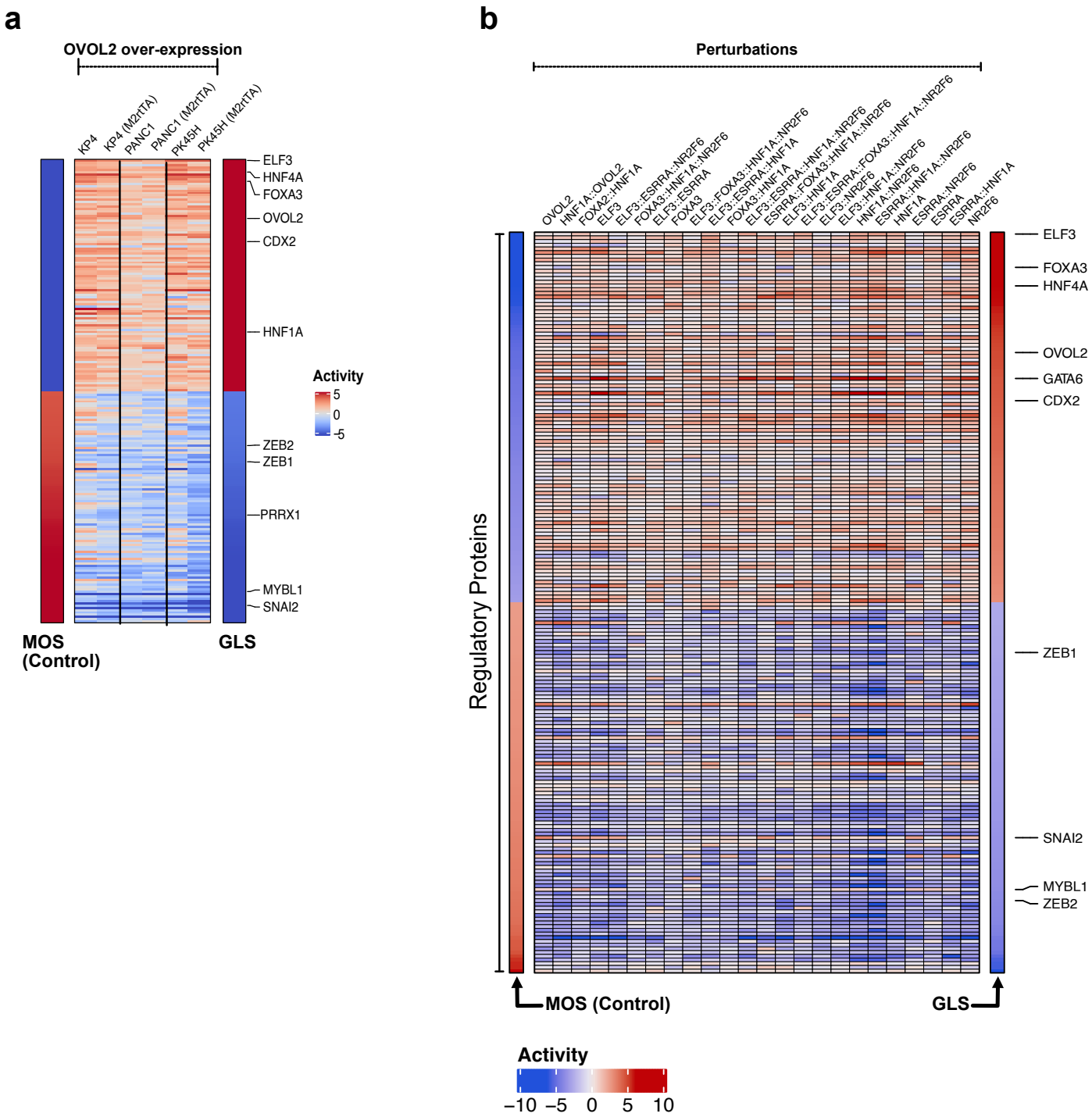


Extended Data Fig. 5 | See next page for caption.

https://doi.org/10.1038/s41588-026-02714-8

Extended Data Fig. 5 | CRISPR screening workflow, quality control and master regulators network. a) Schematic workflow for pooled CRISPR/dCas9 (CRISPRi and CRISPRko) screens. All in vitro screens were performed with two biological replicates per cell line for both CRISPRko and CRISPRi. **b)** ECDF plots showing z-score distributions of core-essential (positive control) versus non–core-essential (negative control) genes in pooled CRISPRko screens, demonstrating effective identification of essential genes and overall screen quality. **c)** ECDF plots for CRISPRi screens are shown as quality control metrics. **d)** GSEA plots showing enrichment of the top 200 differentially activated proteins (100 most activated and 100 most inactivated) inferred from single cells comparing GLS versus MOS states, tested against Lineage versus Morphogenic essential genes. P-values and NES were estimated by two-sided GSEA with 1,000 permutations. **e)** GSEA plots showing enrichment of the top 200 differentially activated proteins (100 most activated and 100 most inactivated) between Lineage and Morphogenic subtypes inferred from bulk analysis (ICGC–TCGA signature), tested against Lineage versus Morphogenic essential genes. P-values and NES were estimated by two-sided GSEA with 1,000 permutations. **f)** Quality control ECDF plot for Lineage MRs selected for the in vivo CRISPRko screen, showing z-score distributions of core-essential genes (positive controls) versus tested genes in vitro in the Lineage HPAFII cell line. Intergenic and non-targeting guide controls were included and highlighted. **g)** ECDF plot showing z-score distributions of core-essential genes (positive controls) versus tested genes assessed in vitro in the Morphogenic KP4 cell line. Intergenic and non-targeting guide controls were included and highlighted. **h)** GSEA plots showing enrichment of the top eight over-activated Lineage MRs in single-cell PDL signatures, computed by comparing each PDL to the other two. These results demonstrate conservation of the top eight Lineage MRs in the GLS signature. NES and p-values were estimated by two-sided GSEA with 1,000 permutations. **i)** Mechanistic regulatory network showing transcriptional and post-transcriptional regulation of the top eight Morphogenic and Lineage MRs, inferred from differential gene expression and protein activity following ectopic expression of each Lineage MR. Protein–protein interactions from STRING and PrePPI are included.

https://doi.org/10.1038/s41588-026-02714-8



Extended Data Fig. 6 | Master regulator-induced reprogramming at bulk and single-cell levels. a) Heatmap showing reproducibility of OVOL2-mediated Morphogenic → Lineage cell state reprogramming (+/- M2rtTA), across three distinct Morphogenic cell lines (KP4, PANC1 and PK45H). **b)** Heatmap showing the integrated protein activity profile of Lineage and Morphogenic RPs in the fraction of cells assessed as statistically significantly reprogrammed following ectopic expression of each MR set, integrated by the Stouffer's method. Rows represent Lineage and Morphogenic regulators with some notable MRs highlighted.

nature portfolio

Corresponding author(s): Califano Andrea, Kenneth P. Olive

Last updated by author(s): Mar 12, 2026

Reporting Summary

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

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	DNA methylation profiles were downloaded from TCGA using TCGAbiolinks package. No other software was used for data collection.
Data analysis	ARACNe-AP was used to reconstruct three PDA regulatory networks (CUMC, ICGC, and UNC),while the TCGA network was obtained from the aracne.networks R package. The VIPER R package was used to infer protein activity (VIPER algorithm). Seurat (R package) was used for single-cell preprocessing and clustering; Monocle was used for pseudo-trajectory analysis. For 10x Genomics data processing, Cell Ranger v5.0.1 was used for scRNA-seq processing of PDA cell lines, Cell Ranger v3 was used for PDX scRNA-seq processing , and Cell Ranger v3.0.2 was used for lineage tracing scRNA-seq processing;Xenium data were processed using the Xenium Onboard Analysis package , with downstream analysis in R v4.3.0. CITE-seq-Count was used (together with Seurat HTODemux) for hashing-based demultiplexing. For PLATE-Seq processing, kallisto v0.44.0 was used for pseudo-alignment and quantification, with tximport (R package) and biomaRt (R package) supporting aggregation and annotation, and ShortRead (R package)was used for sequencing QC. For bulk RNA-seq, DESeq2 (R package) was used for VST normalization and the cluster R package was used to perform PAM clustering. Survival analyses were performed using the survival R package and Kaplan–Meier plots were generated with survminer package. DNA methylation analyses used TCGAbiolinks to retrieve TCGA 450K data, minfi to convert beta values to M values, and limma for differential methylation testing. For lineage tracing barcode processing, UMI-tools was used to collapse UMI barcodes. For multiome analyses, Signac (R package) was used for scATAC-seq analysis including differential peak and motif analyses, mclust (R package;) was used for Gaussian mixture model-based clustering, p-values were aggregated using the HMP package , and KEA enrichment was performed using Enrichr (web tool). CRISPR screen analysis was performed using MAGeCK v0.5.6, and MR interaction networks were visualized using Cytoscape v3.8.2; additional steps used traineR (R package) for KNN training and base R stats for PCA.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

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

Data generated in this study were deposited on public repositories as described in the methods section.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Gender information are not reported in the manuscript
Reporting on race, ethnicity, or other socially relevant groupings	Information about race, ethnicity or other socially relevant groups are not reported in the manuscript
Population characteristics	Median age: 67; Sex: 52% M, 48% F; diagnosis: pancreatic adenocarcinoma or pancreatitis
Recruitment	Retrospective recruitment based on database search for “pancreatic adenocarcinoma” or “pancreatitis
Ethics oversight	Columbia IRB

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	The minimum sample size for LCM-RNAseq cohort was determined based on minimum number of samples required by the ARACNe algorithm. The sample size for CRISPRko, CRISPRi, Plate-seq over expression assay and western blot was determined based on the number of replicates reported in the scientific papers using the same approach
Data exclusions	No data were excluded.
Replication	All the experimental findings were reproduced as validated by at least two independent experiments
Randomization	Samples were not randomized
Blinding	There was no blinding

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

OVOL2-3xFlag: ANTI-FLAG M2 (Sigma-Aldrich)
mCherry: ab167453 (Abcam)
ZEB1: ab203829 (Abcam)
Vimentin: PIMA511880 (Thermo Fisher)
E-Cadherin: 3195S (Cell Signaling Technology)
Beta-Actin: sc-47778 (Santa Cruz Biotechnology)
anti-phosphoERK: 4370 (Cell Signaling Technology)
HNF4A : 3113 (Cell Signaling Technology)
OVOL2: PA5115700 (Thermo Fisher)
HNF1A: ab268116 (abcam)

Validation

OVOL2 and mCherry were validated by overexpressing the ORFs. For the other antibodies we relied on the validations reported on the manufacture's websites

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

CCLE, RIKEN, DSMZ and ATCC databases

Authentication

Cell lines were re-sequenced (RNASeq) to ensure fidelity compared to their CCLE profiles

Mycoplasma contamination

Cell lines were tested for mycoplasma contamination before and during the screening

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified lines were used

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals

8-12 week old NRG mice (mus musculus) were imported from Charles River Labs to serve as host for the PDX models, gender matched to the original patient

Wild animals

No wild animals were involved in this study

Reporting on sex

Sex was not considered in the study design

Field-collected samples

No samples were collected from the field in this study

Ethics oversight

Study protocols were approved by the Columbia IRB and IACUC committees

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration

Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.

Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*