Long Noncoding RNAs Enforce Pancreatic Cancer Identity and Block Reprogramming

Dmytro Grygoryev[1,2*], Seung-Won Lee[1,2*], Connor Mitchell Frankston[1,3], Shauna Rakshe[4], Mark Berry[1], Alex Hirano[5], Taelor Ekstrom[1], Elise Manalo[1], Julien Tessier[1], Marilynn Chow-Castro[1], Jason Link[2,6], Dove Keith[6], Brett C. Sheppard[6, 7], Suzanne Fei[4], Terry Morgan[1, 7, 8], Helen E. Remotti[10], Wenli Yang[11], Emma Furth[12], Sudhir Thakurela[5], Rosalie Sears[2,6,8], and Jungsun Kim[1,2,8,13]

[1]Cancer Early Detection Advanced Research Center, Knight Cancer Institute, Oregon Health & Science (OHSU) School of Medicine, Portland, OR.

[2]Department of Molecular and Medical Genetics, OHSU School of Medicine, Portland, OR.

[3]Biomedical engineering graduate program, Department of Biomedical Engineering, OHSU School of Medicine, Portland, OR.

[4]Biostatistics Shared Resource, Knight Cancer Institute, OHSU, Portland, OR

[5]Lurie Center for Autism, Department of Pediatrics, Harvard Medical School, Boston, MA

[6]Brenden-Colson Center for Pancreatic Care, OHSU School of Medicine, Portland, OR.

[7]Department of Surgery, OHSU School of Medicine, Portland, OR.

[8]Knight Cancer Institute, OHSU School of Medicine, Portland, OR.

[9]Pathology & Laboratory Medicine, OHSU School of Medicine, Portland, OR.

[10]Department of Pathology, Columbia University, New York, NY.

[11]Institute for Regenerative Medicine, UPenn Perelman School of Medicine, Philadelphia, PA.

[12]Pathology and Laboratory Medicine, UPenn Perelman School of Medicine, Philadelphia, PA.

* Equal contribution

[13]corresponding author: Dr. Jungsun Kim (kimjungs@ohsu.edu)

Keywords (max 10): Cancer cell identity, cancer epigenetic memory, long noncoding RNA, master transcription factor-mediated cellular reprogramming, pancreatic Cancer, pluripotency, and plasticity.

Highlight

Summary (150 words or fewer)

Pancreatic ductal adenocarcinoma (PDAC), one of the deadliest malignancies, exemplifies a paradox of transcriptional plasticity and rigidity: it displays dynamic transcriptional states yet retains a stable classical epithelial identity, which resists therapeutic intervention. This rigidity reflects robust transcriptional memory, but the underlying mechanisms remain poorly understood. While the master transcription factors OCT4, SOX2, KLF4, and MYC (OSKM) can reprogram normal cells by erasing lineage identity, reprogramming in solid tumors such as PDAC is inefficient, with reprogrammed cells retaining features of their malignant origin—suggesting the presence of active epigenetic barriers. Long noncoding RNAs (lncRNAs) regulate chromatin state and cell identity, but whether they actively enforce cancer-specific transcriptional memory and restrict reprogramming remains largely unexplored. Herein, through CRISPR interference screens targeting cancer-associated lncRNAs, we identified a subset of cancer-associated lncRNAs whose depletion significantly enhanced OSKM-mediated reprogramming in PDAC cells. Knockdown of these lncRNAs enabled the acquisition of pluripotency-associated features, suppressed PDAC identity programs, and reduced tumorigenicity in vivo. Among the most potent repressors of reprogramming were ATXN7L3-AS1, INTS4P2, and AC079921.2, which were upregulated in PDAC but minimally expressed in normal pancreas. ATXN7L3-AS1 and AC079921.2 were associated with mitotic progression and extracellular matrix organization gene programs, respectively, and showed anticorrelation with translational and metabolic gene signatures across TCGA cancer cohorts. These findings uncover a previously unrecognized role for cancer-associated lncRNAs as key enforcers of transcriptional memory in PDAC. By stabilizing malignant identity and opposing reprogramming, these lncRNAs may represent a novel class of epigenetic regulators. Targeting them offers a conceptual and therapeutic framework for erasing cancer cell memory, resetting epigenetic state, and exposing latent vulnerabilities in PDAC.

Introduction (no subheading, max 7,000 including main figure legends).

Pancreatic ductal adenocarcinoma (PDAC) is one of the deadliest human cancers, emphasizing the urgent need for deeper biological understanding to develop new therapies[1]. Many cancer cells show remarkable transcriptional plasticity and heterogeneity. Yet, their epigenetic landscape plays a crucial role in limiting this plasticity, maintaining a stable, cancer-specific identity. PDAC exemplifies this paradox; it exhibits transcriptional plasticity with dynamic transitions between classical and basal subtypes, yet it keeps a stable core identity defined by oncogenic transcriptional programs and classical pancreatic epithelial signatures. This coexistence of plasticity and rigidity contributes to therapeutic resistance and aggressive clinical behavior. However, the molecular mechanisms that constrain their transcriptional memory and preserve malignant identity remain poorly understood.

Experimental reprogramming through forced expression of master transcription factors (TFs) has demonstrated the capacity to override somatic identity[2-7]. The best-known example is the induction of pluripotency via OCT4, SOX2, KLF4, and MYC (OSKM), which reprograms normal somatic cells into induced pluripotent stem cells (iPSCs)[6,7]. However, in solid tumors including PDAC, reprogramming is often impeded: reprogrammed cancer cells frequently retain epigenetic memory of their tissue of origin and resist full identity erasure[8-12]. This resistance is unexpected given the inherent plasticity of many cancers, suggesting the presence of robust, cancer-specific transcriptional memory mechanisms.

Cellular identity is maintained by gene regulatory networks (GRNs) composed of TFs, cofactors, and chromatin regulators[13-18]. Although GRNs can be reset during regeneration or experimental reprogramming, core networks encode a transcriptional memory preserved through mitotic bookmarking and chromatin state[16-20]. This memory stabilizes cell identity and acts as a barrier to fate conversion[21-32]. Among GRN components, long noncoding RNAs (lncRNAs) have emerged as critical regulators of epigenetic memory. By scaffolding chromatin remodeling complexes and guiding gene expression in cis and trans[33-43], lncRNAs orchestrate

cell-type-specific transcriptional programs. In cancer, lncRNAs contribute to progression and resistance[44-47], but whether they directly enforce malignant transcriptional memory and impede reprogramming remains largely unknown.

We hypothesized that a subset of cancer-associated lncRNAs enforces transcriptional memory in PDAC, thereby preserving malignant identity and preventing OSKM-induced reprogramming. To test this, we performed CRISPR interference (CRISPRi) screens targeting cancer-associated lncRNAs in normal pancreatic and PDAC cells, in the context of OSKM induction. We identified lncRNAs that constrain plasticity, maintain PDAC-specific transcriptional programs, and suppress the emergence of pluripotent states. Their suppression, in combination with OSKM, enables reprogramming of PDAC cells, erases malignant identity and reduces tumorigenic potential in vivo.

These findings uncover a previously unrecognized role for cancer-associated lncRNAs in enforcing transcriptional memory and stabilizing PDAC identity. By identifying lncRNAs that act as reprogramming barriers, our work reveals a new mechanism by which cancer cells resist fate change. Furthermore, these results demonstrate that depletion of such lncRNAs—together with OSKM—can overcome this resistance, erase malignant features, and induce functional reprogramming. These insights open new avenues for reprogramming-based therapeutic strategies aimed at resetting epigenetic state and uncovering latent therapeutic vulnerabilities in PDAC.

RESULTS

Restricted Plasticity of PDAC Cells Toward a Pluripotent State. Primary PDAC samples often include a large proportion of normal cells that are more amenable to reprogramming (Table S1). To accurately assess the reprogramming of human PDAC cells, we first enriched these cells using patient-derived xenografts (PDXs)[48]. To assess the reprogramming potential across PDAC subtypes, we employed both PDX-derived PDAC cells and established PDAC cell lines

representing classical (CFPAC1, ST-00014490 PDX-derived PDAC) and basal (Panc1, ST-00007599 PDX-derived PDAC) subtypes, using BJ human fibroblasts and H6c7 human normal pancreatic ductal epithelial cells (hnPDECs) as controls (Fig. 1A).

To enable transgene-free reprogramming, we utilized Sendai virus–mediated delivery of the polycistronic vector expressing OSKM. Because the Sendai virus is a non-integrating RNA virus, it provides a clean platform for reprogramming without genomic integration[49]. This system achieved efficient and consistent transgene expression across all primary PDAC cells, comparable to that in H6c7 hnPDECs and BJ fibroblasts[49].

We first validated the Sendai virus–mediated OSKM delivery system using human BJ fibroblasts and human acinar cells isolated from pancreatitis tissue of PDAC patients (Fig. S1A-S1B). The reprogramming efficiency of acinar cells (0.4%, 122/30,000 cells) was comparable to that of fibroblasts (Fig. S1C). The resulting iPSC lines expressed the canonical pluripotency marker TRA-1-60, exhibited normal karyotypes, and successfully differentiated into all three germ layers (Fig. S1D-F). These results confirmed that Sendai virus–mediated OSKM delivery effectively reprogrammed human pancreatic acinar cells and fibroblasts into a pluripotent state.

In contrast to pancreatic acinar cells, H6c7 hnPDECs failed to express TRA-1-60 by day 30 post-OSKM induction, although many expressed SSEA-1, a marker associated with the pluripotency of murine cells but with the early embryonic differentiation in human cells (Fig. S2A). TRA-1-60 expression eventually emerged by day 60 (data not shown), indicating that hnPDECs possessed a stronger lineage memory than acinar cells and fibroblasts. Classical-type PDAC cells expressed SSEA-1 but failed to express TRA-1-60, consistent with stalled progression toward pluripotency, after OSKC induction (Fig. S2A). OSKM induction did not change the SSEA-1 and TRA-1-60 expression pattern of basal-type PDAC cells (Fig.S2A), suggesting a failure to rewire their transcriptional identity.

The persistent expression of SSEA-1 in PDAC cells indicated partial lineage rewiring without successful acquisition of pluripotency. Even after 60 days of OSKM induction, PDAC

cells failed to achieve full reprogramming (data not shown). These findings suggested that PDAC cells retained a cancer-specific transcriptional memory that resisted reprogramming, creating a barrier to full cell fate conversion.

A Subset of PDAC Identity Genes Can Be Overridden by OSKM, but Many Others Cannot. Early molecular events following OSKM induction are critical in determining cell fate during reprogramming; OSKM typically initiates reprogramming of normal cells by silencing lineage-specific gene programs while transiently activating genes that facilitate cell fate transition[22,50-56]. To characterize early transcriptional responses in PDAC versus normal cells, we performed transcriptomic profiling of PDAC cells (ST-00007599, ST-00014490, Panc1, CFPAC1) and control cells (BJ fibroblasts and H6c7 hnPDECs) at day 7 post-OSKM induction (Fig. 1A). Notably, ectopic OSKM expression levels were comparable across all cell types (Fig. S2B), and PCA data show consistent patterns across technical replicates within each sample (Fig. S3A), enabling direct comparison of downstream transcriptional responses.

OSKM-mediated reprogramming represses original lineage identity in normal cells[53-56]. We therefore hypothesized that OSKM might downregulate a subset of the PDAC-specific transcriptome, particularly those involved in the malignant identity and potential vulnerabilities. To test this, we first defined PDAC-unique genes by identifying transcripts significantly upregulated across all PDAC cells relative to both fibroblasts and hnPDECs before OSKM treatment (Fig. 1B). Surprisingly, only 7 out of 375 of these PDAC-unique genes were significantly downregulated (>2-fold, $q < 0.05$) at day 7 following OSKM induction (Fig. 1B). These OSKM-responsive genes were enriched in pathways related to apoptosis and epithelial cell differentiation (Table S2A). *SPDEF*, encoding a transcription factor essential for classical PDAC tumorigenesis[57], was among OSKM-responsive genes and whose high expression in PDAC was associated with poor patient prognosis in TCGA datasets (Fig. 1C-D).

To systematically assess the extent of gene expression changes in PDAC cells following OSKM induction, we compared transcriptomic alterations across all cell types. BJ fibroblasts exhibited the highest number of differentially expressed genes (DEG), followed by H6c7 hnPDECs, whereas PDAC cells had markedly fewer DEG overall (Fig. S3B, Table S3). The limited number of DEGs observed in PDAC cells likely reflects inter-sample heterogeneity within the PDAC population. When analyzed at the individual cell-type level, each PDAC cell exhibited a greater number of DEGs compared to H6c7 hnPDECs (Fig. S3B, Table S3). This suggests that normal pancreatic ductal cells maintain a more rigid transcriptional state, whereas PDAC cells exhibit higher intrinsic plasticity. This plasticity, however, is constrained by a robust transcriptional memory that preserves core PDAC identity despite OSKM-mediated perturbation.

Collectively, these findings indicate that although OSKM can partially suppress tumor-associated gene programs, the core transcriptional identity of PDAC cells remains largely intact. This underscores a strong resistance to cell fate conversion and supports the concept that PDAC cells retain transcriptional memory, limiting the effectiveness of OSKM-mediated reprogramming.

Mesodermal Lineage Gene Programs are Significantly Enriched Among PDAC Genes That Are Unresponsive or Aberrantly Upregulated Upon OSKM Induction. We hypothesized that genes unresponsive to OSKM induction in PDAC compared to in normal cells (either BJ fibroblasts or H6c7 hnPDECs) function as safeguards to maintain PDAC cell identity. We also reasoned that genes upregulated in PDAC cells, relative to normal cells, following OSKM induction, represented aberrantly activated programs during reprogramming and might antagonize the reprogramming process instead of being required for successful reprogramming. Furthermore, upregulation of these genes could enhance PDAC cell plasticity in response to external stimuli, potentially leading to unexpected cell fate outcomes.

Establishing a uniform threshold for identifying DEGs is challenging due to the heterogeneity among PDAC samples. Thus, we developed a decision framework to classify genes based on cell-type-specific expression changes by incorporating variability across biological replicates (cancer n=4, normal n=2). Genes expressed differentially before and after OSKM treatment were identified in each sample using likelihood ratio testing (LRT), with selection criteria of $\log_2$(fold change: FC) > 0.2 and False Discovery Rate (FDR) < 0.05 (Fig.1E, Fig. S3C). Each DEG was then ranked based on a weighted Hamming distance[58] relative to predefined expression change patterns, including: (1) downregulation in normal with minimal change in cancer (OSKM unresponsive genes in PDAC), (2) no change or downregulation in normal cells and upregulation in cancer (OSKM aberrantly upregulated genes in PDAC) (Fig. 1F). The OSKM unresponsive genes in PDAC were significantly enriched in pathways involved in cell adhesion (Fig. 1G) whereas OSKM aberrantly upregulated genes in PDAC were significantly enriched in pathways related to mesodermal lineages including mesenchymal differentiation and morphogenesis (Fig. 1H). Strikingly, OSKM-unresponsive or aberrantly upregulated genes in PDAC were predicted to be regulated by PRC2 core components, including SUZ12 and EZH2 (Fig. 1G-H). Supporting this, the OSKM aberrantly upregulated PDAC genes significantly overlapped with known SUZ12 targets in both human and mouse embryonic stem (ES) cells (Fig. S3E). Notably, expression levels of PRC2 subunit genes remained unchanged upon OSKM induction (Fig. S3F), suggesting that the failure to suppress OSKM aberrantly upregulated genes in PDAC may stem from impaired recruitment rather than altered abundance of PRC2 complex.

In summary, OSKM induction downregulated some PDAC marker genes but did not affect, or even upregulated, the expression of the majority of PDAC-specific genes. Many of these were PRC2-regulated genes involved in mesodermal lineage, indicating a key epigenetic barrier to successful reprogramming in PDAC.

Suppression of Cancer-Associated lncRNAs in PDAC Cells Significantly Enhances the Reprogramming Efficiency. LncRNAs contribute to the specificity of PRC recruitment and function and also play a crucial role in maintaining cellular memory[34,43,59-61]. Thus, we hypothesized that a subset of cancer-associated lncRNAs in PDAC cells—specifically those unresponsive or aberrantly upregulated following OSKM induction—actively sustain PDAC identity, and that depleting these lncRNAs could facilitate reprogramming by weakening cancer-associated transcriptional memory. Accordingly, we performed CRISPRi–based screens targeting 691 cancer-associated lncRNAs using a pooled lentiviral library containing 7,087 gRNAs (10 gRNAs per lncRNA, plus non-targeting controls) (Table S4A)[62,63]. The screens were conducted in the context of OSKM-mediated reprogramming of normal ductal cells and PDAC cells, using the H6c7 and HPAFII cell lines engineered to stably express dCas9-KRAB (henceforth referred to as H6c7-dCas9-gRNA and HPAFII-dCas9-gRNA; Fig. 2A, Fig. S3G).

Strikingly, by day 30 post-OSKM induction, a significantly higher frequency of TRA-1-60^{+} pluripotent stem-like cells emerged in the HPAFII-dCas9-gRNA population compared to controls (Fig. 2B, Table S4B). Isolated TRA-1-60^{+} and TRA-1-60^{-} (Hereafter referred to as dCas9-gRNA-OSKM-TP and dCas9-gRNA-OSKM-TN, respectively) populations remained phenotypically stable over multiple passages (Fig. 2C). In contrast, the same gRNA pool targeting lncRNAs did not enhance reprogramming of H6c7 hnPDEC cells into pluripotency (Fig. S3H), suggesting that these lncRNAs function in a PDAC-specific manner to preserve cancer-associated transcriptional memory.

To investigate the molecular identity of PDAC cells reprogrammed with OSKM and depleted of lncRNAs, we performed single-cell RNA sequencing (scRNA-seq) on dCas9-gRNA-OSKM-TP and dCas9-gRNA-OSKM-TN populations, along with appropriate controls: dCas9, dCas9-gRNA, and dCas9-OSKM-TP (rare TRA-1-60^{+} cells generated from OSKM-reprogrammed HPAFII cells without lncRNA knockdown) (Fig. 3A).

After filtering out low-quality cells with >10% mitochondrial gene content and regressing out cell cycle–associated transcripts, we retained high-quality transcriptomes for downstream analysis (Fig. S4A-B). Using the Louvain algorithm, we merged all samples and identified 15 transcriptionally distinct clusters (Fig. 3A–C). HPAFII-dCas9 cells primarily occupied clusters corresponding to tubular epithelial (Group 0) and invasive epithelial (Group 1) cell types, along with several minor heterogeneous populations (Fig. 3B–C, Fig. S4C). As expected, dCas9-OSKM-TP cells (green) were transcriptionally indistinguishable from dCas9 cells (black), with no DEGs detected between the two groups (Fig. 3B), highlighting that OSKM alone was insufficient to reprogram PDAC cells.

Although dCas9-gRNA cells (light green, Fig. 3B) upregulated pyruvate metabolism genes and downregulated epithelial and oxidative phosphorylation genes compared to parental HPAFII-dCas9 cells (Fig. S4D), they remained restricted to Groups 0 and 1, consistent with persistent PDAC transcriptional identity (Fig. 3B–C).

In contrast, dCas9-gRNA-OSKM-TP (Red, Fig. 3B) cells were markedly depleted from Groups 0 and 1 and instead clustered predominantly in Group 5—a population unique to TRA-1-60^{+} cells (Fig.3B-C, Fig. S4E). Indeed, PCA analysis further confirmed that dCas9-gRNA-OSKM-TP cells were transcriptionally distinct from other subpopulations, reflecting reprogramming-induced identity shifts (Fig. S4F). Group 5 was defined by loss of pancreatic and germ layer markers and increased expression of pluripotency-associated genes (Fig. 3B–J), indicating a transition toward a pluripotent state.

To assess the developmental identity of HPAFII-dCas9-gRNA-OSKM-TP cells, we compared their transcriptomic profiles with reference datasets representing human PDAC[64], human and mouse normal pancreas[65], human early embryo before the three germ layers formed[66], and Parkinson's disease patients' fibroblast-derived iPSC[67]. PCA visualization showed that dCas9-gRNA-OSKM-TP cells clustered closely with early human embryos, and distinctly apart from PDAC or normal pancreatic cells (Fig. 3K).

Together, these findings demonstrated that lncRNA knockdown enabled OSKM-induced reprogramming to generate cells with transcriptional features of pluripotency. In contrast, OSKM expression alone failed to erase the transcriptional identity of PDAC cells.

Suppression of Cancer-Associated lncRNAs During OSKM Reprogramming Rewires PDAC Gene Programs. Next, we investigated whether the cell state transition induced by the combination of lncRNA depletion and OSKM induction was accompanied by functional erasure of PDAC identity and re-expression of tumor suppressor genes (TSG). Notably, cell clusters corresponding to mucinous (Group 0) and invasive epithelial (Group 1) PDAC phenotypes were markedly depleted in both dCas9-gRNA-OSKM-TP and -TN populations (Fig. 3B–C), suggesting that OSKM induction in combination with lncRNA knockdown broadly disrupts cancer-specific programs—even in TRA-1-60^{-} cells.

To systematically assess the extent to which PDAC identity is suppressed, we curated "PDAC-specific signature genes" from published scRNA-seq datasets by identifying transcripts enriched in PDAC cells but expressed in fewer than 20% of normal pancreatic cells[64,65]. PDAC-specific signature genes were significantly enriched in downregulated genes in dCas9-gRNA-OSKM-reprogrammed PDAC cells, especially TP cell populations, relative to dCas9-gRNA controls (Fig. 3L). By contrast, no significant downregulation was observed in cell populations treated with OSKM or gRNA only, indicating that both OSKM expression and lncRNA knockdown are required to repress PDAC transcriptional programs effectively.

We next asked whether this transcriptional reset extended to the re-expression of TSGs. Using the OncoKB database[68,69], we assembled a comprehensive list of TSGs that were expressed in normal pancreatic ductal cells but not in cancer and quantified their expression across conditions. As expected, parental HPAFII-dCas9 cells showed minimal or no expression of TSGs (Fig. 3M). In contrast, cells treated with either OSKM or dCas9-gRNA re-expressed

distinct subsets of TSGs. Notably, reprogrammed pluripotent cells (dCas9-gRNA-OSKM-TP) exhibited the most robust reactivation of TSGs (Fig. 3M).

We further analyzed TSG expression in dCas9-gRNA-OSKM-TP cells (Fig. 3B–C, group 5). Using unsupervised k-means clustering, TSGs were grouped into three distinct clusters based on their expression profiles (Fig. 3N). Comparing expression levels of 327 TSGs across cell states defined by NANOG and TRA-1-60 expressions: NANOG$^-$/TRA-1-60$^-$ (N$^-$/T$^-$), NANOG$^-$/TRA-1-60$^+$ (N$^-$/T$^+$), NANOG$^+$/TRA-1-60$^-$ (N$^+$/T$^-$), and NANOG$^+$/TRA-1-60$^+$ (N$^+$/T$^+$) showed that TSG expression was predominantly enriched in NANOG$^+$ states (Fig. 3O).

Together, these findings revealed that OSKM reprogramming in combination with lncRNA depletion not only induced a pluripotent-like state but also erased the PDAC identity and restored TSG expression, suggesting a true reversal of malignant cell fate.

LncRNA Depletion in Combination with OSKM Erases PDAC Identity and Reduces Malignant Tumor Formation In Vivo. We next assessed whether the pluripotency and suppression of the PDAC identity driven by lncRNA depletion in combination with OSKM induction altered the tumorigenic potential in vivo.

To this end, we xenografted HPAFII-dCas9-gRNA-OSKM-TP and -TN cells into immunodeficient NOD-SCID-IL2Rγ^null (NSG) mice, using HPAFII, HPAFII-dCas9-gRNA, and HPAFII-dCas9-OSKM-TP cells as controls. In addition, we included human ES (huES) cells as a pluripotency control (Fig. 4A, S5A-E). As expected, HPAFII cells rapidly formed tumors within one-month post-injection (Fig. 4B–C). All control groups similarly formed tumors, although dCas9-gRNAs produced slightly smaller tumors compared to HPAFII cells. Strikingly, mice injected with dCas9-gRNA-OSKM-TP cells either failed to develop tumors or formed significantly smaller tumors than dCas9-gRNA-OSKM-TN and all controls ($p < 0.005$; Fig. 4B–C).

Histopathological analysis confirmed that tumors from all control groups displayed classical grade 2 adenocarcinoma morphology, with strong human-specific marker TRA-1-85,

pancreatic ductal marker CK19, and proliferation marker Ki67 immunostainings and absence of apoptotic marker Cleaved Caspase-3 (CC3) immunostaining, indicating high proliferation and minimal apoptosis (Fig. 4D-G, 5A-N). Extensive central necrosis was observed, consistent with advanced PDAC pathology. In contrast, tumors derived from dCas9-gRNA-OSKM-TP cells lacked typical adenocarcinoma features (Fig. 4H, 5O-P) and showed robust CC3 immunostaining in addition to Ki67 immunostaining (Fig. 5Q-R), suggesting a transition toward a less aggressive, more benign-like phenotype.

Importantly, tumors from dCas9-OSKM-TP and dCas9-gRNAs cells retained classical adenocarcinoma histology (Fig. 4E-F and 5E, F, G, H), underscoring that neither OSKM expression nor lncRNA knockdown alone was sufficient to reprogram the malignant phenotype. Interestingly, dCas9-OSKM-TP tumors contained more differentiated, non-epithelial cell types in the tumor center, reflecting partial but incomplete identity shifts (Fig. 4F).

Consistent with being pluripotent, human ES cells formed fully mature teratomas within three months post-injection [Ref]. At the one-month point used for PDAC comparisons, ES cell-derived tumors were small and composed predominantly of neuronal epithelial cells, consistent with ectodermal differentiation upon early exit from pluripotency (Fig. S5A-C). These early teratomas possess both proliferation and apoptotic markers, further distinguishing them from HPAFII-derived tumors (Fig. S5D-E).

Together, these findings demonstrate that lncRNA depletion in combination with OSKM expression effectively reprogrammed PDAC cells, resulting in reduced tumorigenicity, loss of classical adenocarcinoma identity, and reactivation of apoptotic programs, in vivo. Importantly, this phenotypic reversal was not achieved by OSKM expression or lncRNA knockdown alone, highlighting the requirement of both to overcome the malignant PDAC state.

PDAC-Enriched lncRNAs Act as Barriers to Cellular Reprogramming and Promote the Tumor Identity. To characterize lncRNAs that maintained PDAC gene expression and were

selectively depleted during successful reprogramming, we performed next-generation sequencing of genomic DNA from FACS-sorted dCas9-gRNA-OSKM-TN, dCas9-gRNA-OSKM-TP, and dCas9-gRNA cells (Fig. 6A, S5A). We applied Z-score normalization and MAGeCK[70] negative control normalization to gRNA count data from two independently passaged cultures (separated by 3–6 passages) and selected gRNAs with Z-scores > 2 and positive beta scores ($p < 0.05$). To ensure robustness, raw gRNA counts from both experiments were merged, scaling the second dataset to correct for batch differences (Exp1/Exp2 ≈ 1.86), with counts rounded up to the nearest integer. gRNAs consistently enriched in both cultures, as defined by either Z-score or MAGeCK analysis, were retained (Table S6; Fig. S5A).

This stringent analysis identified several lncRNAs-targeting gRNAs that were enriched in dCas9-gRNA-OSKM-TP cells compared to dCas9-gRNA (Fig. 6A-B; Fig. S5F-I). Among these, ATXN7L3-AS1 (ENSG00000267394), AC079921.2 (ENSG00000249685), and INTS4P2 (ENSG00000273024, located within LINC03006:ENSG00000290926) were validated by qRT-PCR and found to be significantly downregulated in HPAFII-dCas9-gRNA-TP cells compared to control HPAFII-dCas9 (Fig. 6C; Table S6A). Importantly, these lncRNAs were minimally expressed or absent in normal pancreatic tissue (Fig. 6C-E) and their expression remained unchanged in response to OSKM alone in both PDAC and normal control cells—suggesting resistance to OSKM-mediated reprogramming (Fig. S6A).

We generated HPAFII PDAC cells expressing gRNAs targeting ATXN7L3-AS1, INTS4P2, or AC079921.2, as well as a non-targeting control. As a comparison, we introduced the same gRNAs into H6c7 non-malignant pancreatic ductal cells. In HPAFII cells, knockdown of ATXN7L3-AS1 or INTS4P2 led to reduced expression of basal PDAC markers KRT5 and TP63, as well as the classical marker GATA6 (Fig. 7A). ATXN7L3-AS1 knockdown modestly increased endodermal markers PDX1, FOXA2, and SOX9, whereas INTS4P2 knockdown had little effect on these genes. In contrast, the same perturbations in H6c7 cells produced the opposite pattern, underscoring that these lncRNAs function in a PDAC-specific manner.

To gain insight into the potential functions of these lncRNAs, we examined their genomic context and co-expression patterns in the TCGA Pan-Cancer RNA-seq dataset. Expression data were available for ATXN7L3-AS1 and AC079921. The expression of ATXN7L3-AS1 was positively correlated with the expression of genes involved in mitotic cell cycle progression (Fig. S6B). These genes were predicted to have UBTF as a potential upstream regulator based on motif analysis (Fig. S6C) and carried active enhancer marks such as H3K27ac and H3K4me2 associated with transcriptional activity and memory (Fig. S6D). The expression of AC079921.2 was positively correlated with the expression of genes involved in extracellular matrix (ECM) remodeling, cell adhesion, and cell migration (Fig. S6E), key hallmarks of invasive PDAC. These genes were predicted to be SUZ12 (PRC2) targets (Fig. S6F) and showed promoter enrichment for H3K27me3 (Fig. S6G), suggesting that AC079921.2 may antagonize PRC2-mediated silencing. Notably, OSKM-aberrantly upregulated or unresponsive PDAC genes (Fig. 1F–H) significantly overlapped with genes positively correlated with AC079921.2 (Fig. 7B-C). These overlapping genes are enriched for morphogenesis, cell migration, and ECM remodeling functions (Fig. 7D-E) and are predicted SUZ12 targets (Fig. 7F-G). Together, these findings suggest that AC079921.2 promotes expression of ECM and morphogenetic programs that resist OSKM-induced transcriptional resetting, thereby reinforcing PDAC cell identity. Consistently, AC079921.2-correlated genes are enriched for mesenchymal-like and invasive signatures, supporting a role for this lncRNA in maintaining a pro-invasive transcriptional state in PDAC cells.

Intriguingly, genes whose expressions negatively correlated with the expression of ATXN7L3-AS1 or AC079921.2 were enriched in translation and metabolic pathways (Fig. S6H-I), marked by histone modifications associated with active transcription H3K36me3 and H3K79me2 (Fig. S6J-K), and were predicted targets of MYC and MAX, TAF1, and GABPA (Fig. S6L-M). These findings linked these lncRNAs to the regulation of transcriptional programs that controlled biosynthesis and proliferation in PDAC.

Taken together, these data indicated ATXN7L3-AS1, INTS4P2, and AC079921.2 as key lncRNAs that helped maintain PDAC transcriptional identity.

Discussion

In this study, we identify a subset of cancer-associated lncRNAs as key regulators of transcriptional memory and malignant identity in PDAC. By performing CRISPRi screens during OSKM-induced reprogramming, we show that depletion of cancer-associated lncRNAs significantly enhances reprogramming efficiency and enables the transition of human PDAC cells into a pluripotent state. Although OSKM alone partially repressed PDAC gene programs, reprogramming remained incomplete, with cells stalling in early developmental states marked by activation of extracellular morphogenesis-related genes. In contrast, knockdown of lncRNAs in combination with OSKM led to robust induction of pluripotency-associated features, loss of adenocarcinoma identity, and reduced malignancy in vitro and in vivo. ATXN7L3-AS1, AC079921.2, and INTS4P2 were consistently downregulated in TRA-1-60^{+} reprogrammed cells. In particular, expression of ATXN7L3-AS1 and AC079921.2 correlated with mitotic and extracellular matrix gene programs, respectively, across TCGA cancers. Knockdown of ATXN7L3-AS1 also reduced expression of PDAC marker genes. Together, our findings reveal a previously unrecognized role for cancer-associated lncRNAs as molecular gatekeepers that preserve PDAC identity and constrain cell fate plasticity.

Are ATXN7L3-AS1 and AC079921.2 Gatekeepers of PDAC Identity and Plasticity?

Cellular memory is epigenetically inherited across cell divisions and is essential for maintaining cell identity. During mitosis, specific transcription factors act as mitotic bookmarks to preserve lineage fidelity[19,20,71]. Similarly, lncRNAs help maintain transcriptional memory and identity by regulating chromatin architecture and gene expression through interactions with chromatin-modifying enzymes.

ATXN7L3-AS1 is located on chromosome 17, upstream of *ATXN7L3* and *UBTF* in the antisense direction. Notably, ATXN7L3 is a component of the SAGA histone acetylation complex[72] and UBTF is an HMGB-box transcription factor essential for Pol I/II activity and known to be oncogenic[72]. Intriguingly, our motif analysis indicated that UBTF may be an upstream regulator of genes whose expression positively correlated with the ATXN7L3-AS1 level, suggesting that ATXN7L3-AS1 may reinforce a UBTF-driven transcriptional program. Although UBTF has not been directly implicated in PDAC, it has been reported as an oncogenic driver in other cancers[73]. Notably, UBTF::ATXN7L3 fusions define a high-risk subtype of B-ALL[74], and UBTF tandem duplications are recurrent in acute myeloid leukemia[75]. These findings support the hypothesis that ATXN7L3-AS1 acts as a molecular gatekeeper of cancer cell identity and may be a promising target for transcriptional memory and malignant persistence in PDAC.

In pan-cancer analyses, AC079921.2 expression is positively associated with expressions of genes involved in ECM remodeling, cell adhesion, and migration—processes characteristic of mesenchymal-like tumor states. In normal tissues, these genes are typically repressed by SUZ12 and marked by H3K27me3, suggesting that AC079921.2 may antagonize PRC2-mediated silencing. Consistent with this model, OSKM-aberrantly upregulated or unresponsive PDAC genes significantly overlap with AC079921.2-correlated genes in our dataset. These findings imply that AC079921.2 may help sustain a transcriptional program that resists OSKM-induced resetting, thereby preserving PDAC cell identity. Future studies should investigate whether AC079921.2 directly interacts with PRC2 components, such as SUZ12 or EZH2, to modulate their repressive function in PDAC cells. By maintaining expressions of ECM and morphogenetic genes, AC079921.2 may reinforce transcriptional plasticity and promote the adaptation to external cues. Targeting this lncRNA could be a strategy to constrain lineage plasticity and prevent undesired cell-state transitions during therapy or cellular reprogramming.

Intriguingly, levels of both ATXN7L3-AS1 and AC079921.2 are negatively correlated with the expression of genes involved in translation and metabolic pathways. This suggests that

these lncRNAs may help safeguard PDAC cell identity by repressing gene networks that are otherwise poised for high transcriptional activity and metabolic remodeling, thereby restricting the reprogramming potential.

Combining lncRNAs Depletion and OSKM Induction Is Required to Erase PDAC Identity

Although knockdown of ATXN7L3-AS1 and INTS4P2 reduced expression of a subset of PDAC marker genes, our results suggest that either lncRNA knockdown or OSKM induction alone is insufficient to fully erase PDAC identity or suppress malignancy in vivo. In xenograft experiments, lncRNA depletion without OSKM induction modestly reduced tumor size was, likely due to reduced proliferation without increased apoptosis, but did not reprogram PDAC cells or erase the adenocarcinoma identity. Similarly, OSKM induction alone resulted in tumors exhibiting more differentiated features with reduced proliferation but failed to fully erase the PDAC identity. In contrast, HPAFII-dCas9-gRNA-TP cells lost their original adenocarcinoma identity and reactivated apoptotic pathways. The observed tumor suppressor gene reactivation in NANOG^{+} pluripotent cells suggests that the loss of PDAC identity is coupled with a fate change toward a pluripotent state. We propose that lncRNA depletion sensitizes PDAC cells to OSKM reprogramming by weakening epigenetic barriers that maintain malignancy.

Beyond pluripotency, cancer reprogramming into other lineages, such as dendritic cells[76], has also been demonstrated. It will be important to test whether lncRNA depletion can facilitate PDAC reprogramming toward alternative or non-permissive lineages, e.g., acinar cells incompatible with oncogenic KRAS. We anticipate that targeting lncRNAs that regulate transcriptional memory and plasticity could enhance PDAC responsiveness to reprogramming cues and increase lineage fidelity.

Altogether, the selective expression of ATXN7L3-AS1, AC079921.2, and INTS4P2 in PDAC highlights their potential as regulators of transcriptional memory and plasticity in pancreatic cancer, and likely function as molecular barriers to reprogramming.

Limitations of the Study

Our study focuses specifically on lncRNAs that sustain the cancer identity in PDAC and uses OSKM to force reprogramming of PDAC toward pluripotency. However, lncRNAs that sustain the cancer identity may differ across cancer types and different lineage-specific transcription factors may be required for the reprogramming other cancer types.

While we observed a robust reduction in malignancy and activation of apoptosis in tumors derived from HPAFII-dCas9-gRNA-OSKM-TP cells, it remains unclear whether the erasure of malignancy is permanent. Long-term xenograft studies will be required to assess the durability of reprogramming-induced tumor suppression.

Conclusions

We have demonstrated that combining lncRNAs depletion and OSKM induction is necessary for erasing the identity and malignancy of PDAC cells. Our findings suggest that cancer-specific lncRNAs act as critical gatekeepers preserving PDAC cell identity and transcriptional memory, even in the presence of reprogramming cues. These lncRNAs likely underlie treatment resistance and disease recurrence and may represent high-value targets for therapeutic intervention.

AUTHOR CONTRIBUTIONS

Conceptualization, JK; Methodology, DG, SL, CMF, SR, MB, AH, TE, EM, JT, MCC, JL, and WY; Validation, DG, SL, CMF, SR, SF, TM, and WY; Formal Analysis, DG, SL, CMF, SR, AH, SF, TM, WY, EF, ST, and JK; Investigation, D.G., SL, CMF, SR, MB, SF, TM, WY, EF, ST, and JK.; Resources, JL, DK, BCS, EF, and RCS; Data Curation, DG, SL, CMF, SR, AH, SF, and ST; Writing-original draft, JK; Writing-Review & Editing, DG, SL, and JK; Visualization, DG, CMF, SR, AH, SF, and ST, Supervision, JK; Project Administration, JK; Funding Acquisition, JK.

ACKNOWLEDGMENTS

We thank Drs. Ken Zaret and Theresa Lusardi for comments on the manuscript. We thank the patients who donated their tissue for this study and P. Roberts and R. Armstrong for their technical assistance. We thank the Brenden-Colson Center for Pancreatic Care, the Oregon Pancreas Tissue Registry, the OHSU Knight Biostatistics Shared Resource and Biolibrary (NCI P30CA69533), and the Molecular Virology Core at the Oregon National Primate Research Center (NIH P51OD011092). The IRB protocol is the Oregon Pancreas Tissue Registry (IRB #3609). The work was funded by the CRUK-OHSU Project Award (2018-CRUK-OHSU-001 to JK), MRF New Investigator Grant (GCNCR1042A to JK), and Knight CEDAR (68182-906-000, 68182-933-000, 68182-939-000, 68182-954-000, Manuscript Prep 2024-1978) from the Cancer Early Detection Advanced Research Center at OHSU Knight Cancer Institute. This article was edited by Life Science Editors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Author contributions

DECLARATION OF INTERESTS

The authors declare no competing interests

STAR METHODS

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jungsun Kim (kimjungs@ohsu.edu).

Materials availability

- Plasmids, PDX tumors, and reprogrammed cancers generated in this study will be available upon request.
- This study did not generate new unique reagents.

Data and code availability

- Raw data and codes for sequencing data have been deposited at NCBI GEO and are publicly available as of publication. Accession numbers will be listed in the key resource table. The lead contact, upon request, will share any raw data reported in this paper.
- Codes for sequencing data have been deposited on GitHub and are publicly available as of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Derivation of Patient-Derived Xenografts (PDX) and Tissue Culture. Human PDAC specimens were obtained under the Oregon Pancreas Tissue Registry study (IRB00003609). Informed consent was obtained from all subjects. All experimental protocols were approved by the OHSU Institutional Review Board. All methods were carried out in accordance with relevant

guidelines and regulations. All animal works for PDX tumors were performed with the OHSU Institutional Animal Use and Care Committee (IACUC) approval. The derivation, expansion, culture, and characterization of PDAC PDXs, are described in our protocol paper[48]. Culture conditions for H6c7 and other PDAC cell lines are described[48]. BJ6 fibroblasts were cultured in DMEM supplemented with 10% fetal bovine serum (FBS)[55].

Reprogramming of Pancreatic Cancer Cells and Normal Controls. Reprogramming was performed using the CytoTune-iPS Sendai Reprogramming Kit (Invitrogen), which delivers the non-integrating Sendai virus encoding OCT4, SOX2, KLF4, and c-MYC (OSKM)[77]. Viral titers for Sendai virus and lentivirus were described in our earlier publication[49]. On day 7 post-transduction, cells were harvested for RNA sequencing and replated onto mitomycin C-treated mouse embryonic fibroblasts (MEFs) to support further reprogramming.

METHOD DETAILS

Tissue immunohistochemistry (IHC), cell immunostaining, antibodies, qRT-PCR, and Primers. Tissue IHC, immunofluorescence staining, and qRT-PCR were performed as previously described in our published work[8]. The following primary antibodies were used: Human TRA-1-85/CD147 Antibody (R&D Systems #MAB3195), human Cytokeratin 19 (Abcam, BA-17), Ki67 (Cell Signaling, #9449), and Cleaved Caspase-3 (Cell Signaling Technology, #12692). Primers used are described in Table Sx.

QUANTIFICATION AND STATISTICAL ANALYSIS

RNA-seq Data Processing and Normalization. Bulk RNA-seq was performed as previously described[49,78]. Raw FASTQ files were processed using fastp to generate clean reads by removing adapter sequences, reads containing poly-N, and low-quality reads. During this step,

quality metrics such as Q20, Q30, and GC content were also calculated. All downstream analyses were conducted using the high-quality clean reads. The reference genome and gene annotation files were downloaded directly from the genome database. The genome index was built using HISAT2 (v2.0.5), and paired-end clean reads were aligned to the reference genome using HISAT2 (v2.0.597)[79]. FeatureCounts (v1.5.0-p3) was used to quantify the number of reads mapped to each gene. FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values were then calculated for each gene based on its length and the number of mapped reads. Differential expression analysis was performed using the DESeq2 R package (v1.20.0), which employs statistical routines based on the negative binomial distribution to identify differentially expressed genes from count data[80]. P-values were adjusted for multiple testing using the Benjamini-Hochberg method to control FDR. Genes with an adjusted P-value ≤ 0.05 were considered significantly differentially expressed. RNA-seq data is available on NCBI GSExxxx.

To identify PDAC-specific gene categories, we first performed differential expression analysis using DESeq2 to compare each PDAC cell line to normal controls (BJ fibroblasts and H6c7 pancreatic epithelial cells). Genes that were significantly upregulated in PDAC cells (FDR < 0.05, $\log_2$ FC > 1) were selected. We then intersected the upregulated gene sets across all four PDAC cell lines to define a core set of PDAC-upregulated genes. To assess the impact of OSKM induction, we compared gene expression profiles before and after OSKM treatment and identified differentially expressed genes using the same thresholds (FDR < 0.05, $\log_2$FC> 1 or < –1).

Pattern Classification Following OSKM and Gene Ranking by Weighted Hamming Distance. DEGs before and after OSKM induction (day 7 vs. day 0) were identified for each cell type using the LRT implemented in DESeq2[80]. To determine robust thresholds for DEG calling, we evaluated mean absolute deviation (MAD) heatmaps across a range of FDR cutoffs and $\log_2$(FC) thresholds. Based on this analysis, genes with $\log_2$(FC) > 0.2 and FDR < 0.05 were

considered significantly differentially expressed. To classify gene expression patterns in response to OSKM induction, we queried DEGs across six cell types—two normal controls and four PDAC cell lines—using five predefined patterns: i) Downregulated in normal, upregulated in cancer: (–1, –1, 1, 1, 1, 1), ii) Downregulated in normal, unchanged in cancer: (–1, –1, 0, 0, 0, 0), iii) Unchanged in normal, upregulated in cancer: (0, 0, 1, 1, 1, 1), iv) Upregulated in normal, downregulated in cancer: (1, 1, –1, –1, –1, –1), v) Upregulated in normal, unchanged in cancer: (1, 1, 0, 0, 0, 0). These patterns were used to stratify gene responses to reprogramming and identify cancer-specific regulatory behaviors.

To rank genes within each predefined expression pattern category, we used a weighted Hamming distance metric relative to the query vector[58]. Let q denote the query vector representing a specific regulation profile (e.g., upregulated in normal with minimal change in cancer samples), and let cat(S) denote the categorical expression vector of a gene. The weighted Hamming distance d is calculated as: $d = \sum(i=1 \text{ to } N)\ w_i \times \mathrm{Ind}(q_i \neq \mathrm{cat}(S)_i)$ where N is the number of conditions or replicates, w_i is the weight assigned to the *i*th condition, and Ind is the indicator function that returns 1 if the query and gene vector differ at position *i*, and 0 otherwise. This metric penalizes deviations from the target pattern, with weights chosen to balance representation across biological conditions and to reflect the hierarchical structure of replicates within each condition. Genes with lower weighted Hamming distances are considered better matches to the specified query. For each query of interest, genes were ranked in ascending order based on their weighted Hamming distance.

Gene Ontology (GO) and Upstream Analysis. GO analysis was performed using Enrichment Analysis[81] or the ToppGene Suite[82]. To determine whether gene sets of interest are enriched for putative TF targets, we performed ChIP-X enrichment analysis using datasets from ENCODE or ChEA[83,84], accessed through Enrichr[85]or X2K[86]. ChIP-X refers to genome-wide TF binding data obtained from ChIP-ChIP, ChIP-Seq, ChIP-PET, or DamID experiments[83,84] . The

p-value is computed using Fisher's exact test or the hypergeometric test. This is a binomial proportion test that assumes a binomial distribution and independence for the probability of any gene belonging to any set. The q-value is an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypothesis testing. The combined score is calculated by taking the logarithm of the p-value from the Fisher exact test and multiplying it by the Z-score, which reflects the deviation from the expected rank. The odds ratio (used for rank-based enrichment) is derived by performing the Fisher exact test across numerous random gene sets to estimate the mean and standard deviation of the expected rank for each term in the gene-set library. A Z-score is then computed to quantify the deviation of the observed rank from expectation.

Single-cell RNAseq analysis. Single-cell RNA sequencing (scRNA-seq) was performed using the Chromium Next GEM Single Cell 5' Kit (v2.0, 10x Genomics). All scRNA-seq samples were processed using Seurat v5. Following quality control, cells were filtered, normalized, scaled, batch-corrected, and clustered in accordance with best practices outlined in the Seurat v5 vignette[87]. Cells expressing gRNA were filtered for BFP. Clustering was performed using the Seurat integration workflow with the CCA (Canonical Correlation Analysis) method, at a resolution of 0.5 using the Louvain algorithm. Harmony integration was also performed to assess potential differences compared to the Seurat CCA-based integration. scRNAseq data set was extracted from human PDAC (n=6, GSE212966[64]), human and mouse normal pancreas (n=4, GSE84133[65]), human early embryo before the three germ layers formed (D6-D14, n=555, GSE136447[66]), and iPSC derived from Parkinson's patients' fibroblast (n=12, GSE183248[67]). NT clusters were manually defined based on NANOG and TRA-1-60 expression profiles. Cells were grouped according to their expression of these markers, with clusters curated based on predefined thresholds rather than unsupervised clustering algorithms.

CRISPRi screen and next-generation sequencing. The lentivirus expressing a sgRNA library with 7087 different sgRNAs against cancer common lncRNAs and dCas9 protein was obtained through Millipore[62,63]. The HPAFII and H6c7 cells expressing dCas9 were generated by transducing HPAFII and H6c7 cells with lentiviruses carrying the dCas9-Bsr gene at an MOI of 1, followed by selecting cells for Blastidin. The CRISPRi efficiency was assessed in the clonal HPAFII and H6c7 cells expressing dCa9 with positive controls. The clonal HPAFII-dCas9 and H6c7-dCas9 were subsequently transduced with a lentiviral library of 7087 sgRNAs at an MOI of 0.1-0.3 with 300- 500x coverage to maintain an average 300-500-fold representation of each sgRNA. After puromycin selection, cells were collected for the reference controls (300x, 500x). HPAFII and H6c7 cells were reprogrammed with OSKM Sendai viruses[77]. The CRISPRi-targeted cells were always maintained at 300-500 coverage throughout the procedure. Technical replicate genomic DNA isolated from each group was subjected to nested PCR amplification to NGS libraries to identify the enriched or depleted sgRNAs in TRA-1-60^{+} reprogrammed cells over reference controls or TRA-1-60^{-}. The hit count corresponding to each sgRNA was computed from raw FastQ files using Millipore's custom pipelines. The raw hit counts were normalized by Z-score for exploratory analysis. The MAGeCK package with negative control sgRNA normalization was used to analyze the enriched lncRNAs[70]. The MAGeCK algorithm first performs normalization, then uses information borrowing across genes to estimate variance[70]. A negative binomial model is then fit and used to test for differential hits on genes across samples. Significance testing uses permutations. We used the MAGeCK mle module, since there were more than two conditions. The MAGeCKFlute R package was used with R version 4.4 for data analysis and visualization[88].

REFERENCE

1. Halbrook, C.J., Lyssiotis, C.A., Pasca di Magliano, M., and Maitra, A. (2023). Pancreatic cancer: Advances and challenges. Cell *186*, 1729-1754. 10.1016/j.cell.2023.02.014.
2. Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Sudhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. Nature *463*, 1035-1041. 10.1038/nature08797.
3. Laiosa, C.V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L., and Graf, T. (2006). Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. Immunity *25*, 731-744. 10.1016/j.immuni.2006.09.011.
4. Xie, H., Ye, M., Feng, R., and Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell *117*, 663-676. 10.1016/s0092-8674(04)00419-2.
5. Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987-1000. 10.1016/0092-8674(87)90585-x.
6. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872. 10.1016/j.cell.2007.11.019.
7. Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676. 10.1016/j.cell.2006.07.024.
8. Kim, J., Hoffman, J.P., Alpaugh, R.K., Rhimm, A.D., Reichert, M., Stanger, B.Z., Furth, E.E., Sepulveda, A., Yuan, C.X., Won, K.J., et al. (2013). An iPSC Line from Human Pancreatic Ductal Adenocarcinoma Undergoes Early to Invasive Stages of Pancreatic Cancer Progression. Cell Reports *3*, 2088-2099. 10.1016/j.celrep.2013.05.036.
9. Zhang, X., Cruz, F.D., Terry, M., Remotti, F., and Matushansky, I. (2013). Terminal differentiation and loss of tumorigenicity of human cancers via pluripotency-based reprogramming. Oncogene *32*, 2249-2260, 2260 e2241-2221. 10.1038/onc.2012.237.
10. Bernhardt, M., Novak, D., Assenov, Y., Orouji, E., Knappe, N., Weina, K., Reith, M., Larribere, L., Gebhardt, C., Plass, C., et al. (2017). Melanoma-Derived iPCCs Show Differential Tumorigenicity and Therapy Response. Stem Cell Reports *8*, 1379-1391. 10.1016/j.stemcr.2017.03.007.
11. Papapetrou, E.P. (2016). Patient-derived induced pluripotent stem cells in cancer research and precision oncology. Nat Med *22*, 1392-1401. 10.1038/nm.4238.
12. Kim, J., and Zaret, K.S. (2015). Reprogramming of human cancer cells to pluripotency for models of cancer progression. EMBO Journal *34*, 739-747. 10.15252/embj.201490736.
13. Davidson, E.H., and Erwin, D.H. (2006). Gene regulatory networks and the evolution of animal body plans. Science *311*, 796-800. 10.1126/science.1113832.
14. Shukla, V., Cetnarowska, A., Hyldahl, M., and Mandrup, S. (2022). Interplay between regulatory elements and chromatin topology in cellular lineage determination. Trends Genet *38*, 1048-1061. 10.1016/j.tig.2022.05.011.
15. Almeida, N., Chung, M.W.H., Drudi, E.M., Engquist, E.N., Hamrud, E., Isaacson, A., Tsang, V.S.K., Watt, F.M., and Spagnoli, F.M. (2021). Employing core regulatory circuits to define cell identity. EMBO J *40*, e106785. 10.15252/embj.2020106785.
16. Bulyk, M.L., Drouin, J., Harrison, M.M., Taipale, J., and Zaret, K.S. (2023). Pioneer factors - key regulators of chromatin and gene expression. Nat Rev Genet. 10.1038/s41576-023-00648-z.
17. Zaret, K.S., and Mango, S.E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr Opin Genet Dev *37*, 76-81. 10.1016/j.gde.2015.12.003.

18. Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956. 10.1016/j.cell.2005.08.020.
19. Caravaca, J.M., Donahue, G., Becker, J.S., He, X., Vinson, C., and Zaret, K.S. (2013). Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. Genes Dev *27*, 251-260. 10.1101/gad.206458.112.
20. Zhu, Z., Chen, X., Guo, A., Manzano, T., Walsh, P.J., Wills, K.M., Halliburton, R., Radko-Juettner, S., Carter, R.D., Partridge, J.F., et al. (2023). Mitotic bookmarking by SWI/SNF subunits. Nature *618*, 180-187. 10.1038/s41586-023-06085-6.
21. Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I., et al. (2010). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285-290. 10.1038/nature09342.
22. Polo, J.M., Liu, S., Figueroa, M.E., Kulalert, W., Eminli, S., Tan, K.Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T., et al. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat Biotechnol *28*, 848-855. 10.1038/nbt.1667.
23. Bar-Nur, O., Russ, H.A., Efrat, S., and Benvenisty, N. (2011). Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. Cell Stem Cell *9*, 17-23. 10.1016/j.stem.2011.06.007.
24. Ohi, Y., Qin, H., Hong, C., Blouin, L., Polo, J.M., Guo, T., Qi, Z., Downey, S.L., Manos, P.D., Rossi, D.J., et al. (2011). Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nat Cell Biol *13*, 541-549. 10.1038/ncb2239.
25. Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature *471*, 68-73. 10.1038/nature09798.
26. Buckberry, S., Liu, X., Poppe, D., Tan, J.P., Sun, G., Chen, J., Nguyen, T.V., de Mendoza, A., Pflueger, J., Frazer, T., et al. (2023). Transient naive reprogramming corrects hiPS cells functionally and epigenetically. Nature *620*, 863-872. 10.1038/s41586-023-06424-7.
27. Hu, Q., Friedrich, A.M., Johnson, L.V., and Clegg, D.O. (2010). Memory in induced pluripotent stem cells: reprogrammed human retinal-pigmented epithelial cells show tendency for spontaneous redifferentiation. Stem Cells *28*, 1981-1991. 10.1002/stem.531.
28. Thiagalingam, S. (2020). Epigenetic memory in development and disease: Unraveling the mechanism. Biochim Biophys Acta Rev Cancer *1873*, 188349. 10.1016/j.bbcan.2020.188349.
29. Cheloufi, S., Elling, U., Hopfgartner, B., Jung, Y.L., Murn, J., Ninova, M., Hubmann, M., Badeaux, A.I., Euong Ang, C., Tenen, D., et al. (2015). The histone chaperone CAF-1 safeguards somatic cell identity. Nature *528*, 218-224. 10.1038/nature15749.
30. Nefzger, C.M., Rossello, F.J., Chen, J., Liu, X., Knaupp, A.S., Firas, J., Paynter, J.M., Pflueger, J., Buckberry, S., Lim, S.M., et al. (2017). Cell Type of Origin Dictates the Route to Pluripotency. Cell Rep *21*, 2649-2660. 10.1016/j.celrep.2017.11.029.
31. Morris, S.A., Cahan, P., Li, H., Zhao, A.M., San Roman, A.K., Shivdasani, R.A., Collins, J.J., and Daley, G.Q. (2014). Dissecting engineered cell types and enhancing cell fate conversion via CellNet. Cell *158*, 889-902. 10.1016/j.cell.2014.07.021.
32. Ghosh, Z., Wilson, K.D., Wu, Y., Hu, S., Quertermous, T., and Wu, J.C. (2010). Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. PLoS One *5*, e8975. 10.1371/journal.pone.0008975.
33. Mattick, J.S., Amaral, P.P., Carninci, P., Carpenter, S., Chang, H.Y., Chen, L.L., Chen, R., Dean, C., Dinger, M.E., Fitzgerald, K.A., et al. (2023). Long non-coding RNAs: definitions, functions,

challenges and recommendations. Nat Rev Mol Cell Biol *24*, 430-447. 10.1038/s41580-022-00566-8.

34. Zaratiegui, M., Irvine, D.V., and Martienssen, R.A. (2007). Noncoding RNAs and gene silencing. Cell *128*, 763-776. 10.1016/j.cell.2007.02.016.
35. Flynn, R.A., and Chang, H.Y. (2014). Long noncoding RNAs in cell-fate programming and reprogramming. Cell Stem Cell *14*, 752-761. 10.1016/j.stem.2014.05.014.
36. Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., et al. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516-520. 10.1038/nature12210.
37. Kim, Y.J., Xie, P., Cao, L., Zhang, M.Q., and Kim, T.H. (2018). Global transcriptional activity dynamics reveal functional enhancer RNAs. Genome Res *28*, 1799-1811. 10.1101/gr.233486.117.
38. Kaikkonen, M.U., Lam, M.T., and Glass, C.K. (2011). Non-coding RNAs as regulators of gene expression and epigenetics. Cardiovasc Res *90*, 430-440. 10.1093/cvr/cvr097.
39. Jiao, W., Chen, Y., Song, H., Li, D., Mei, H., Yang, F., Fang, E., Wang, X., Huang, K., Zheng, L., and Tong, Q. (2018). HPSE enhancer RNA promotes cancer progression through driving chromatin looping and regulating hnRNPU/p300/EGR1/HPSE axis. Oncogene *37*, 2728-2745. 10.1038/s41388-018-0128-0.
40. Statello, L., Guo, C.J., Chen, L.L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol *22*, 96-118. 10.1038/s41580-020-00315-9.
41. Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. Nature *482*, 339-346. 10.1038/nature10887.
42. Ng, S.Y., Johnson, R., and Stanton, L.W. (2012). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. EMBO J *31*, 522-533. 10.1038/emboj.2011.459.
43. Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A *106*, 11667-11672. 10.1073/pnas.0904715106.
44. Wang, Z.W., Pan, J.J., Hu, J.F., Zhang, J.Q., Huang, L., Huang, Y., Liao, C.Y., Yang, C., Chen, Z.W., Wang, Y.D., et al. (2022). SRSF3-mediated regulation of N6-methyladenosine modification-related lncRNA ANRIL splicing promotes resistance of pancreatic cancer to gemcitabine. Cell Rep *39*, 110813. 10.1016/j.celrep.2022.110813.
45. Ye, M., Dong, S., Hou, H., Zhang, T., and Shen, M. (2021). Oncogenic Role of Long Noncoding RNAMALAT1 in Thyroid Cancer Progression through Regulation of the miR-204/IGF2BP2/m6A-MYC Signaling. Mol Ther Nucleic Acids *23*, 1-12. 10.1016/j.omtn.2020.09.023.
46. Jiang, M.C., Ni, J.J., Cui, W.Y., Wang, B.Y., and Zhuo, W. (2019). Emerging roles of lncRNA in cancer and therapeutic opportunities. Am J Cancer Res *9*, 1354-1366.
47. Arunkumar, G., Baek, S., Sturgill, D., Bui, M., and Dalal, Y. (2022). Oncogenic lncRNAs alter epigenetic memory at a fragile chromosomal site in human cancer cells. Sci Adv *8*, eabl5621. 10.1126/sciadv.abl5621.
48. Tessier, J., Grygoryev, D., Chow-Castro, M., Ekstrom, T., Manalo, E., Lee, S.W., Berry, M., Link, J.M., Keith, D., Allen-Petersen, B.L., et al. (2025). Protocol to purify and culture human pancreatic cancer cells from patient-derived xenografts. STAR Protoc *6*, 103672. 10.1016/j.xpro.2025.103672.
49. Grygoryev, D., Ekstrom, T., Manalo, E., Link, J.M., Alshaikh, A., Keith, D., Allen-Petersen, B.L., Sheppard, B., Morgan, T., Soufi, A., et al. (2024). Sendai virus is robust and consistent in

delivering genes into human pancreatic cancer cells. Heliyon *10*, e27221. 10.1016/j.heliyon.2024.e27221.

50. Cacchiarelli, D., Trapnell, C., Ziller, M.J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z.D., Ratanasirintrawoot, S., Zhang, X., et al. (2015). Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. Cell *162*, 412-424. 10.1016/j.cell.2015.06.016.
51. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Sasaki, A., Yamamoto, M., Nakamura, M., Sutou, K., Osafune, K., and Yamanaka, S. (2014). Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. Nat Commun *5*, 3678. 10.1038/ncomms4678.
52. Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209-1222. 10.1016/j.cell.2012.08.023.
53. Knaupp, A.S., Buckberry, S., Pflueger, J., Lim, S.M., Ford, E., Larcombe, M.R., Rossello, F.J., de Mendoza, A., Alaei, S., Firas, J., et al. (2017). Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. Cell Stem Cell *21*, 834-845 e836. 10.1016/j.stem.2017.11.007.
54. Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell *168*, 442-459 e420. 10.1016/j.cell.2016.12.016.
55. Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994-1004. 10.1016/j.cell.2012.09.045.
56. Li, D., Liu, J., Yang, X., Zhou, C., Guo, J., Wu, C., Qin, Y., Guo, L., He, J., Yu, S., et al. (2017). Chromatin Accessibility Dynamics during iPSC Reprogramming. Cell Stem Cell *21*, 819-833 e816. 10.1016/j.stem.2017.10.012.
57. Tonelli, C., Yordanov, G.N., Hao, Y., Deschenes, A., Hinds, J., Belleau, P., Klingbeil, O., Brosnan, E., Doshi, A., Park, Y., et al. (2024). A mucus production programme promotes classical pancreatic ductal adenocarcinoma. Gut *73*, 941-954. 10.1136/gutjnl-2023-329839.
58. Hamming, R.W. (1950). Error detecting and error correcting codes. The Bell System Technical Journal *29*, 147-160. 10.1002/j.1538-7305.1950.tb00463.x.
59. Petracovici, A., and Bonasio, R. (2021). Distinct PRC2 subunits regulate maintenance and establishment of Polycomb repression during differentiation. Mol Cell *81*, 2625-2639 e2625. 10.1016/j.molcel.2021.03.038.
60. Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell *40*, 939-953. 10.1016/j.molcel.2010.12.011.
61. Yang, F., Deng, X., Ma, W., Berletch, J.B., Rabaia, N., Wei, G., Moore, J.M., Filippova, G.N., Xu, J., Liu, Y., et al. (2015). The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. Genome Biol *16*, 52. 10.1186/s13059-015-0618-0.
62. Liu, S.J., Horlbeck, M.A., Cho, S.W., Birk, H.S., Malatesta, M., He, D., Attenello, F.J., Villalta, J.E., Cho, M.Y., Chen, Y., et al. (2017). CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science *355*. 10.1126/science.aah7111.
63. Horlbeck, M.A., Gilbert, L.A., Villalta, J.E., Adamson, B., Pak, R.A., Chen, Y., Fields, A.P., Park, C.Y., Corn, J.E., Kampmann, M., and Weissman, J.S. (2016). Compact and highly active next-

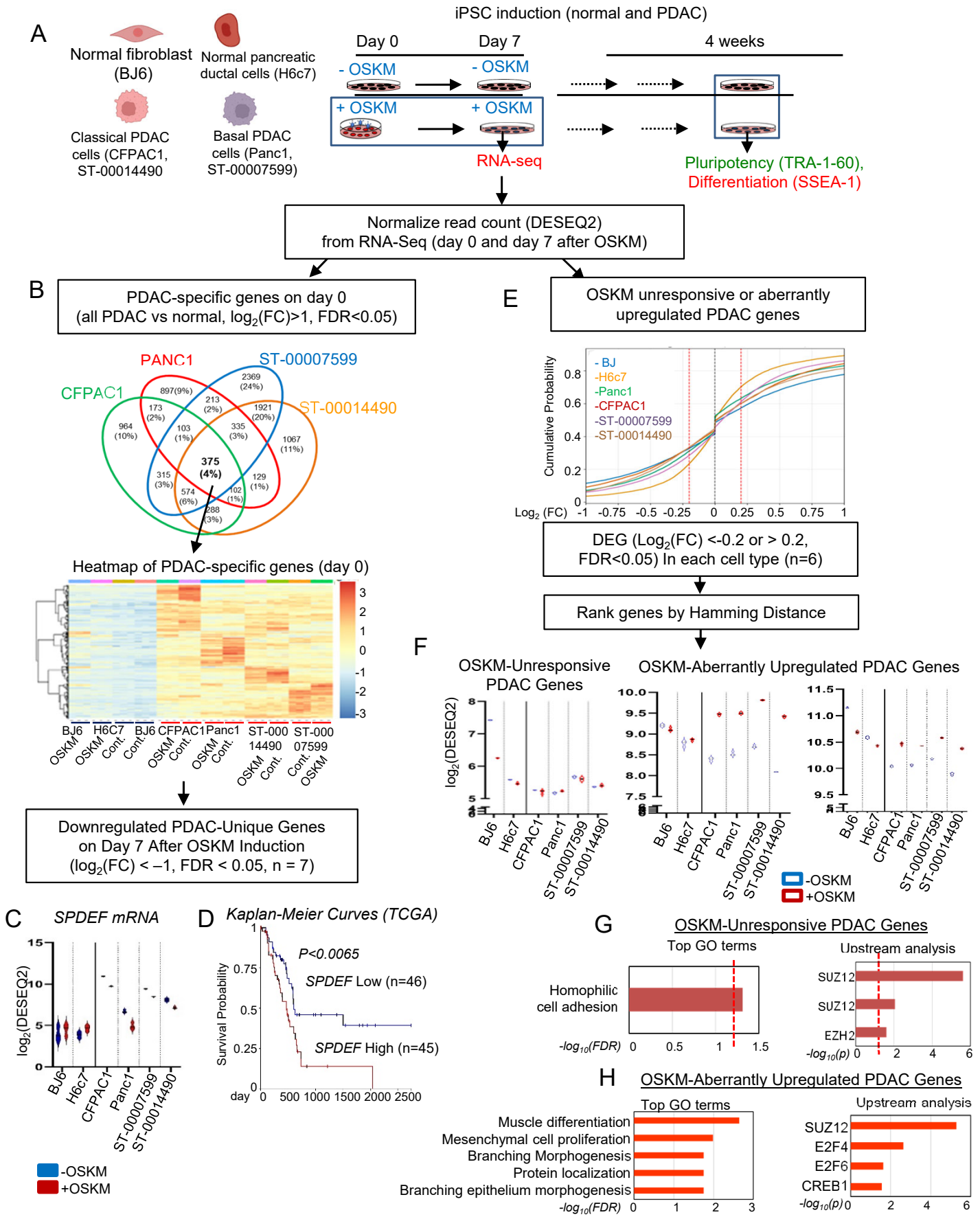
generation libraries for CRISPR-mediated gene repression and activation. Elife *5*. 10.7554/eLife.19760.

64. Chen, J., Liu, Z., Wu, Z., Li, W., and Tan, X. (2023). Identification of a chemoresistance-related prognostic gene signature by comprehensive analysis and experimental validation in pancreatic cancer. Front Oncol *13*, 1132424. 10.3389/fonc.2023.1132424.
65. Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., Ryu, J.H., Wagner, B.K., Shen-Orr, S.S., Klein, A.M., et al. (2016). A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. Cell Syst *3*, 346-360 e344. 10.1016/j.cels.2016.08.011.
66. Xiang, L., Yin, Y., Zheng, Y., Ma, Y., Li, Y., Zhao, Z., Guo, J., Ai, Z., Niu, Y., Duan, K., et al. (2020). A developmental landscape of 3D-cultured human pre-gastrulation embryos. Nature *577*, 537-542. 10.1038/s41586-019-1875-y.
67. Novak, G., Kyriakis, D., Grzyb, K., Bernini, M., Rodius, S., Dittmar, G., Finkbeiner, S., and Skupin, A. (2022). Single-cell transcriptomics of human iPSC differentiation dynamics reveal a core molecular network of Parkinson's disease. Commun Biol *5*, 49. 10.1038/s42003-021-02973-7.
68. Suehnholz, S.P., Nissan, M.H., Zhang, H., Kundra, R., Nandakumar, S., Lu, C., Carrero, S., Dhaneshwar, A., Fernandez, N., Xu, B.W., et al. (2024). Quantifying the Expanding Landscape of Clinical Actionability for Patients with Cancer. Cancer Discov *14*, 49-65. 10.1158/2159-8290.CD-23-0467.
69. Chakravarty, D., Gao, J., Phillips, S.M., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H., et al. (2017). OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol *2017*. 10.1200/PO.17.00011.
70. Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol *15*, 554. 10.1186/s13059-014-0554-4.
71. Palozola, K.C., Donahue, G., Liu, H., Grant, G.R., Becker, J.S., Cote, A., Yu, H., Raj, A., and Zaret, K.S. (2017). Mitotic transcription and waves of gene reactivation during mitotic exit. Science *358*, 119-122. 10.1126/science.aal4671.
72. Li, W., Atanassov, B.S., Lan, X., Mohan, R.D., Swanson, S.K., Farria, A.T., Florens, L., Washburn, M.P., Workman, J.L., and Dent, S.Y.R. (2016). Cytoplasmic ATXN7L3B Interferes with Nuclear Functions of the SAGA Deubiquitinase Module. Mol Cell Biol *36*, 2855-2866. 10.1128/MCB.00193-16.
73. Zhang, J., Zhang, J., Liu, W., Ge, R., Gao, T., Tian, Q., Mu, X., Zhao, L., and Li, X. (2021). UBTF facilitates melanoma progression via modulating MEK1/2-ERK1/2 signalling pathways by promoting GIT1 transcription. Cancer Cell International *21*, 543. 10.1186/s12935-021-02237-8.
74. Passet, M., Kim, R., Gachet, S., Sigaux, F., Chaumeil, J., Galland, A., Sexton, T., Quentin, S., Hernandez, L., Larcher, L., et al. (2022). Concurrent CDX2 cis-deregulation and UBTF::ATXN7L3 fusion define a novel high-risk subtype of B-cell ALL. Blood *139*, 3505-3518. 10.1182/blood.2021014723.
75. Umeda, M., Ma, J., Huang, B.J., Hagiwara, K., Westover, T., Abdelhamed, S., Barajas, J.M., Thomas, M.E., Walsh, M.P., Song, G., et al. (2022). Integrated Genomic Analysis Identifies UBTF Tandem Duplications as a Recurrent Lesion in Pediatric Acute Myeloid Leukemia. Blood Cancer Discov *3*, 194-207. 10.1158/2643-3230.BCD-21-0160.
76. Ascic, E., Akerstrom, F., Sreekumar Nair, M., Rosa, A., Kurochkin, I., Zimmermannova, O., Catena, X., Rotankova, N., Veser, C., Rudnik, M., et al. (2024). In vivo dendritic cell reprogramming for cancer immunotherapy. Science *386*, eadn9083. 10.1126/science.adn9083.
77. Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., and Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus

that does not integrate into the host genome. Proc Jpn Acad Ser B Phys Biol Sci *85*, 348-362. 10.2183/pjab.85.348.

78. Kim, J., Ekstrom, T., Yang, W., Donahue, G., Grygoryev, D., Ngo, T., Muschler, J.L., Morgan, T., and Zaret, K.S. (2021). Longitudinal analysis of human pancreatic adenocarcinoma development reveals transient gene expression signatures. Mol Cancer Res. 10.1158/1541-7786.MCR-21-0483.
79. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology *37*, 907-915. 10.1038/s41587-019-0201-4.
80. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. 10.1186/s13059-014-0550-8.
81. Thomas, P.D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L.P., and Mi, H. (2022). PANTHER: Making genome-scale phylogenetics accessible to all. Protein Sci *31*, 8-22. 10.1002/pro.4218.
82. Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res *37*, W305-311. 10.1093/nar/gkp427.
83. Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., and Ma'ayan, A. (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics *26*, 2438-2444. 10.1093/bioinformatics/btq466.
84. Keenan, A.B., Torre, D., Lachmann, A., Leong, A.K., Wojciechowicz, M.L., Utti, V., Jagodnik, K.M., Kropiwnicki, E., Wang, Z., and Ma'ayan, A. (2019). ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic Acids Res *47*, W212-W224. 10.1093/nar/gkz446.
85. Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M., et al. (2021). Gene Set Knowledge Discovery with Enrichr. Curr Protoc *1*, e90. 10.1002/cpz1.90.
86. Clarke, D.J.B., Kuleshov, M.V., Schilder, B.M., Torre, D., Duffy, M.E., Keenan, A.B., Lachmann, A., Feldmann, A.S., Gundersen, G.W., Silverstein, M.C., et al. (2018). eXpression2Kinases (X2K) Web: linking expression signatures to upstream cell signaling networks. Nucleic Acids Res *46*, W171-W179. 10.1093/nar/gky458.
87. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., and Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nature Biotechnology *42*, 293-304. 10.1038/s41587-023-01767-y.
88. Wang, B., Wang, M., Zhang, W., Xiao, T., Chen, C.H., Wu, A., Wu, F., Traugh, N., Wang, X., Li, Z., et al. (2019). Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. Nat Protoc *14*, 756-780. 10.1038/s41596-018-0113-7.

Grygoryev & Lee/Kim et al.,2025
Figure 1



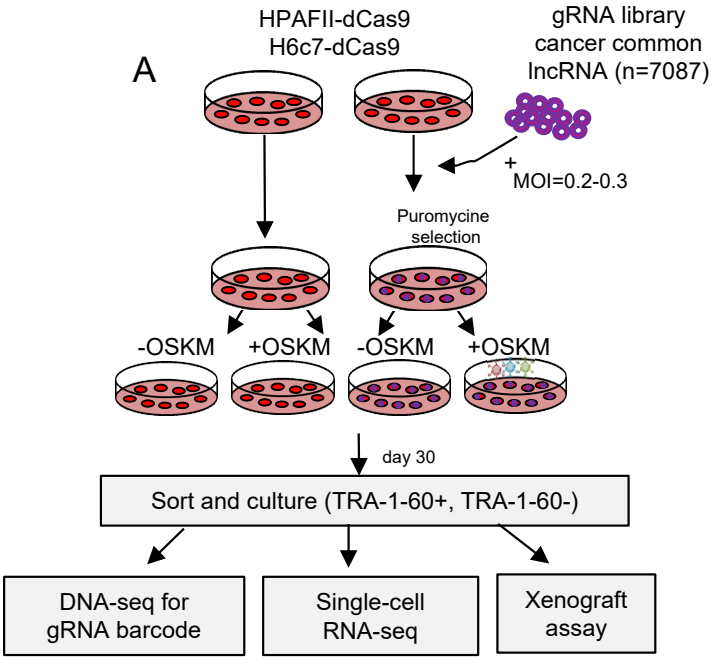
MAIN FIGURE TITLES and LEGENDS

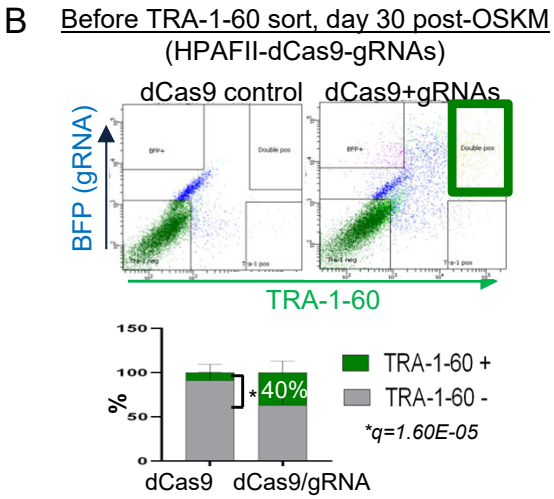
Figure 1. OSKM-mediated cellular reprogramming downregulates only a subset of PDAC unique genes.

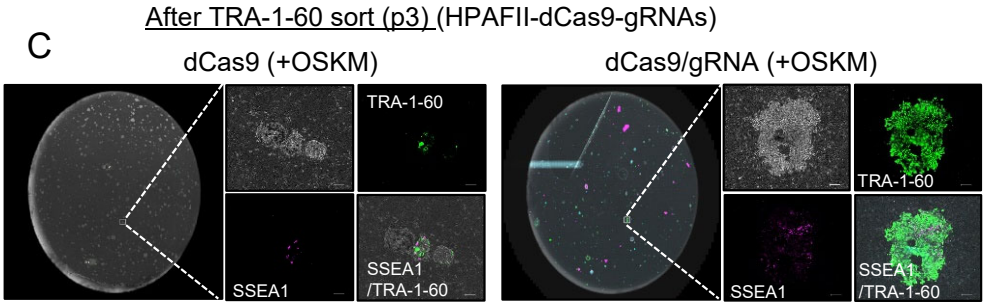
A. Schematic diagram of reprogramming of normal controls (BJ fibroblast and H6c7 hnPDEC) and human classical (CFPAC1, ST-00014490) and basal (Panc1, ST-00007599) subtypes of PDAC cells using Sendai virus-mediated expression of OSKM factors. Cells were transduced with OSKM-expressing Sendai virus on day 0. For transcriptomic analysis, cells were harvested on days 0 and 7. To assess pluripotency and early differentiation, cells were fixed on day 30 and immunostained for TRA-1-60 and SSEA1, respectively. B. Each Venn Diagram shows PDAC-specific genes that are upregulated relative to normal controls on day 0 (≥2-fold, FDR < 0.05), and the heatmap displays the expression of 375 PDAC-specific genes consistently upregulated across all PDAC samples, with Z-scores represented by the color key. Seven among the 375 PDAC unique genes were significantly downregulated by OSKM in all PDAC compared to normal cells on day 7 (<2-fold, FDR<0.05). C-D. Example of an OSKM-induced downregulated PDAC-specific gene. *SPDEF* mRNA levels in the indicated cells without and with OSKM induction (C). Kaplan–Meier survival analysis of *SPDEF* expression in the TCGA PDAC cohort (n = 150). Patients were stratified by SPDEF expression quartiles (Low: <7.611; High: >9.675), based on $\log_2$-transformed normalized RSEM values. P-value was calculated using the log-rank test (χ^2, rho = 0). The high *SPDEF* expression in PDAC is associated with poor overall survival of patients with PDAC (D). E. Identification of aberrantly increased or unchanged genes in PDAC following OSKM induction. The empirical cumulative distribution function (ECDF) displays the cumulative probability distribution of $\log_2$ fold-change values (day 7 vs. day 0) across samples, before and after OSKM induction. The X-axis indicates $\log_2$(FC), and the Y-axis indicates the cumulative fraction of genes. The plot is zoomed in to the range of -1 to 1 to highlight subtle shifts in gene expression. Genes with $\log_2$(FC) > 0.2 or < –

0.2 and FDR < 0.05 were defined as DEGs in each sample, based on inspection of the Mean Absolute Deviation (MAD) heatmap (Fig. S3C; see Methods). Each gene was then classified based on a pre-defined category, and the classification accuracy was determined using a weighted Hamming distance (see Methods). Genes with distances between 0 and 0.35 were selected for downstream analysis. F. Violin plots showing the distribution of gene expression changes across each transcriptional category. G-H. Gene Ontology (GO) and upstream regulator analyses for genes that were either unresponsive to OSKM in PDAC (G) or aberrantly upregulated following OSKM induction (H).

Grygoryev & Lee/Kim et al.,2025
Figure 2







	Total cell #	Total SSEA1+	Total TRA-1-60+	TRA-1-60+ SSEA1+	TRA-1-60+ only	SSEA1+ only
dCas9+OSKM	809	29 (3.6%)	35 (4.3%)	20 (2.5%)	15 (1.9.%)	9 (1%)
dCas9+gRNA	2665	23 (1%)	0	0	0	23 (1%)
dCas9+gRNA +OSKM	212	47 (22.2%)	96 (45.3%)	32 (15.1%)	64 (30.2%)	15 (7.1%)

Figure 2. CRISPRi against lncRNAs significantly improves the efficiency of reprogramming PDAC cells.

A. Schematic diagram of reprogramming HPAFII PDAC cells by CRISPRi screening. The reprogramming efficiency was measured by immunostaining for TRA-1-60 and SSEA1. The TRA-1-60+ pluripotent stem cell-like cells were quantified and sorted by flow cytometry at 30 days. B. Comparing percentages of TRA-1-60^{+} cells between HPAFII cells without and with gRNA against lncRNAs on day 30 after OSKM induction by flow cytometry. *indicates significant differences in the % of cells between TRA-1-60^{-} (91.2%) and TRA-1-60^{+} (8.8%) expressing cells in HPAFII-dCas9 (**q=1.60E-05, n=4,* Unpaired Welch t-test), whereas there were no significant differences between TRA-1-60^{-} and TRA-1-60^{+} cells in HPAFII-dCas9/gRNA cells. C. Representative images and summary of quantification of TRA-1-60^{+} (TRA-1-60^{+}/SSEA1^{+} and TRA-1-60^{+} only) and SSEA1^{+} (TRA-1-60^{+}/SSEA1^{+} and SSEA1^{+} only) expressing cells on the plates after several passages.

Grygoryev & Lee/Kim et al.,2025
Figure 3

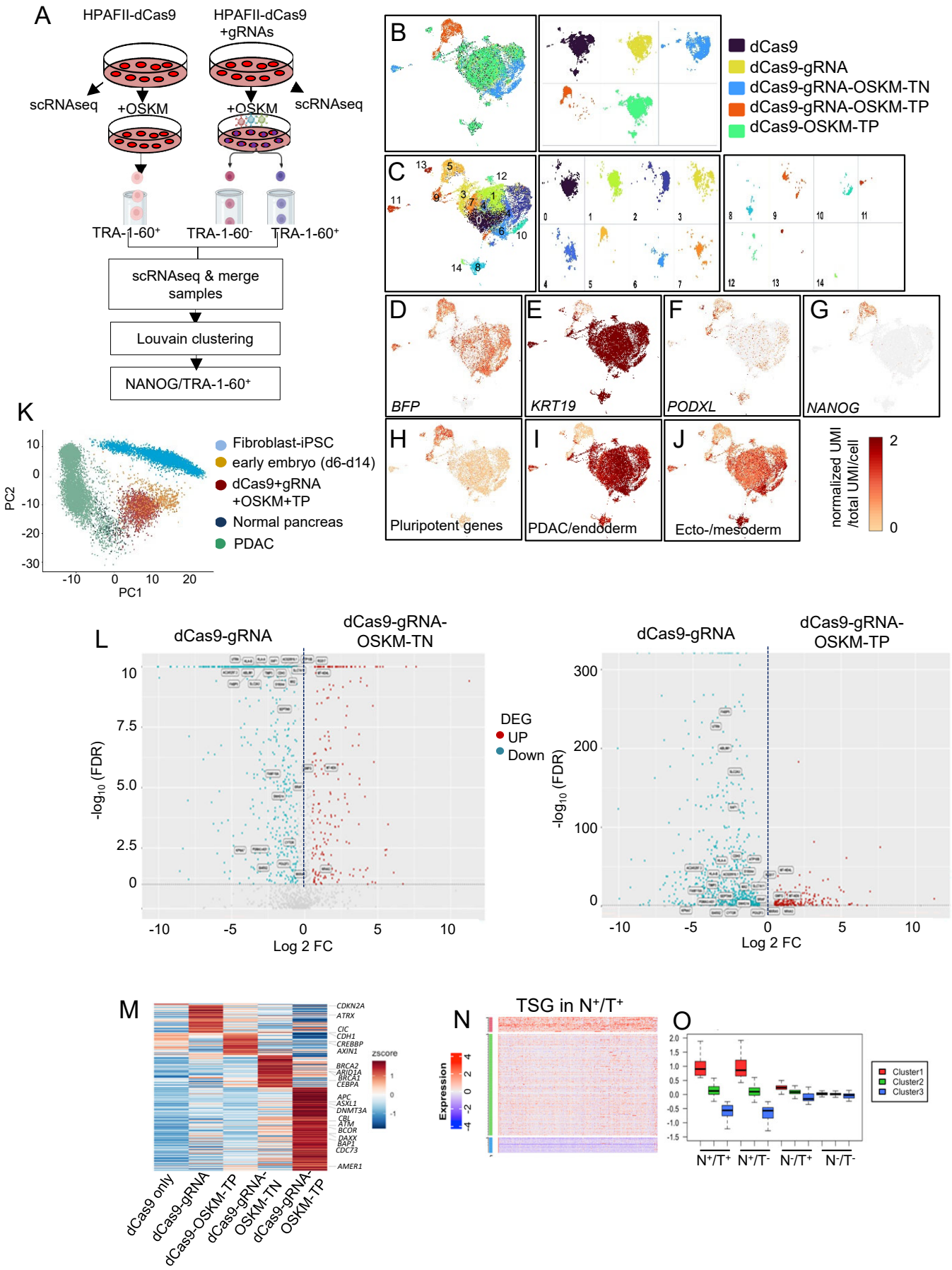


Figure 3. Suppression of PDAC gene programs and activation of pluripotency and tumor suppressor signatures in HPAFII-dCas9-gRNA-OSKM-TP cells.

A. Schematic diagram of the scRNA-seq workflow. B-J. UMAP visualizations showing distinct clusters by sample (B), 14 Louvain-defined clusters (C), as well as cells expressing BFP (D), the pancreatic ductal marker KRT19 (E), *PODXL* encoding TRA-1-60 (F), *NANOG* (G), pluripotency-associated genes (H), endodermal and PDAC-related transcription factors (I), and ectodermal/mesodermal lineage markers (J). Each dot represents a cell, and the color intensity indicates the normalized UMI counts (scaled by the total UMI per cell). K. Principal component analysis (PCA) of HPAFII-dCas9-gRNA-OSKM-TP cells alongside publicly available scRNA-seq datasets from human PDAC (n=6, GSE212966[64]), human and mouse normal pancreas (n=4, GSE84133[65]), human early embryo before the three germ layers formed (D6-D14, n=555, GSE136447[66]), and iPSC derived from Parkinson's disease patients' fibroblast (n=12, GSE183248[67]). L. Volcano plots of DEGs between dCas9-gRNA and either dCas9-gRNA-OSKM-TN (left) or HPAFII-dCas9-gRNA-OSKM-TP (right). PDAC-specific signature genes are annotated. M. Heatmap showing expression of 360 TSGs across the indicated cell populations. The plot shows z-scores of the mean TSG expression across cells in the sample. The distinct subsets show that each sample has a set of TSGs more highly expressed in that sample than in others. TSGs previously reported in PDAC are highlighted. N-O. Heatmap (N) and quantification (O) of TSG expression in the dCas9-gRNA-OSKM-TP cells. Each column represents an individual cell, each row corresponds to a TSG, and expression levels are shown as Z-scores normalized across all cells for each gene (N). Expression levels of TSGs in each cluster are shown across the indicated subpopulations; the Y-axis indicates Z-score.

Grygoryev & Lee/Kim et al.,2025

Figure 4

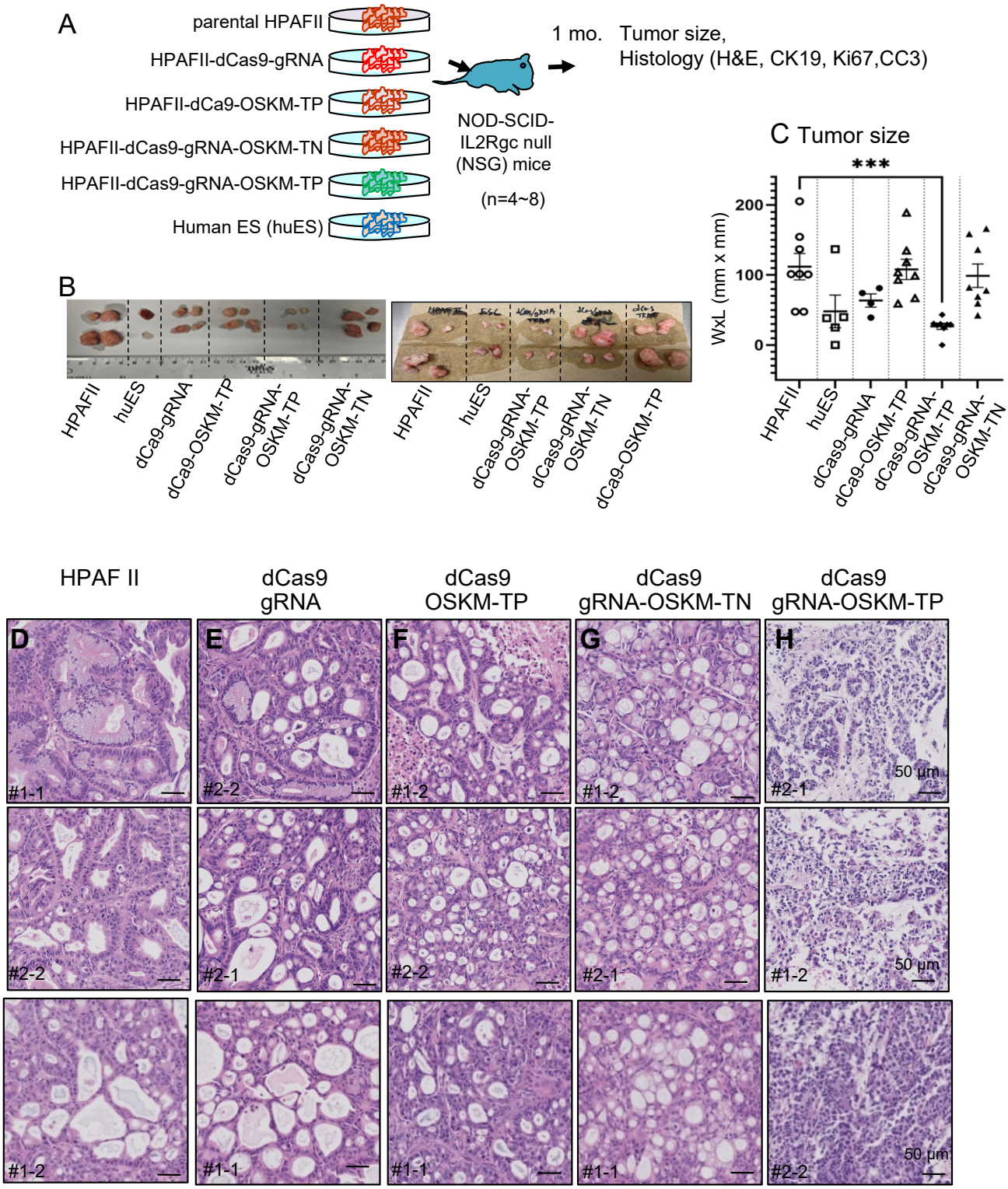


Figure 4. Reduced tumor size in xenografts derived from HPAFII-dCas9-gRNA-TP cells.

A. Schematic diagram of xenograft assay of TRA-1-60^{+} and TRA-1-60^{-}cells derived from OSKM-reprogrammed HPAFII-dCas9/gRNA, along with controls. Tumors were harvested from NSG mice one month after being subcutaneously injected with dCas9-gRNA-OSKM-TN and dCas9-gRNA-OSKM-TP, along with controls: Parental HPAFII, huES, dCas9-gRNA, dCas9-OSKM-TP (n=4-8 for each cell type). B. Representative images of tumors from two independent experiments; two to four mice for each cell type were used in each experiment. C. Comparison of tumor size among different cell types. The differences in tumor sizes (width x height) were tested pairwise using a non-parametric Mann-Whitney test. Only tumor sizes between parental HPAFII and dCas9-gRNA-OSKM-TP are significantly different (*** p=0.0002, n=8 each, Mann-Whitney test). D-H. Representative H&E images of the indicated xenografted tumor samples (D; HPAFII, E; dCas9-gRNA, F; dCas9-OSKM-TP, G; dCas9-gRNA-OSKM-TN, H; dCas9-gRNA-OSKM-TP) from three independent mice. Mouse identifiers (e.g., #1-1, #1-2, #2-1, #2-2) denote individual animals associated with each cell type. The scale bar indicates 50 μm.

Grygoryev & Lee/Kim et al.,2025
Figure 5

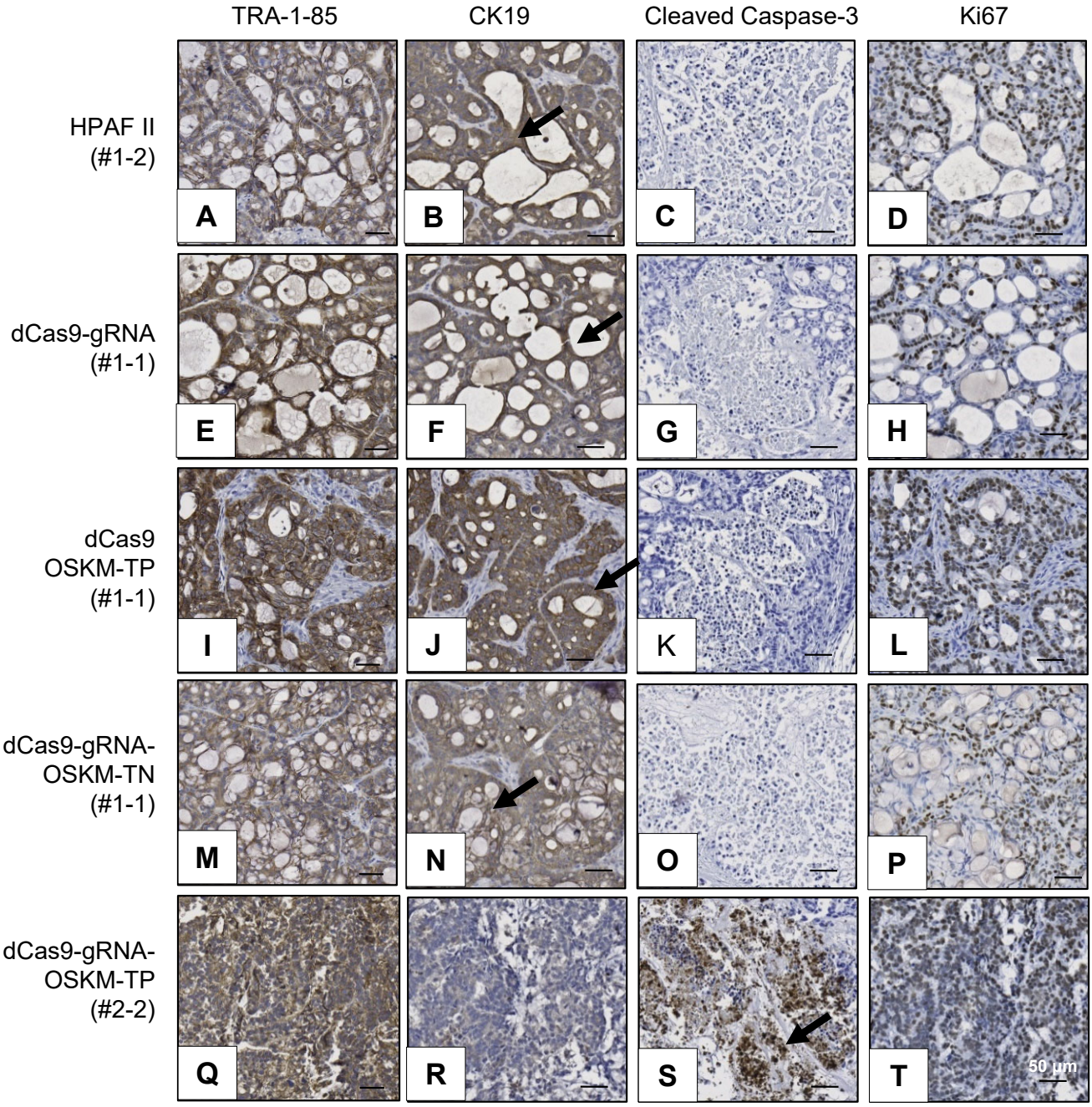


Figure 5. Loss of ductal identity and reactivation of apoptotic markers in xenograft tumors derived from PDAC cells reprogrammed with OSKM and lncRNA depletion.

Representative images of TRA-1-85 human specific marker, human CK19 pancreatic ductal marker, Cleaved Caspase-3 apoptotic marker, and Ki67 proliferation marker staining in xenograft tumors derived from parental HPAFII cells (A–D), dCas9-gRNA cells (E–H), dCas9-OSKM-TP cells (I–L), dCas9-gRNA-OSKM-TN cells (M–P), and dCas9-gRNA-OSKM-TP cells (Q–T). Mouse identifiers (e.g., #1-1, #1-2, #2-1, #2-2) denote individual animals associated with each cell type. Arrows indicate positive brown signals. Scale bar: 50 μm.

Grygoryev & Lee/Kim et al.,2025

Figure 6

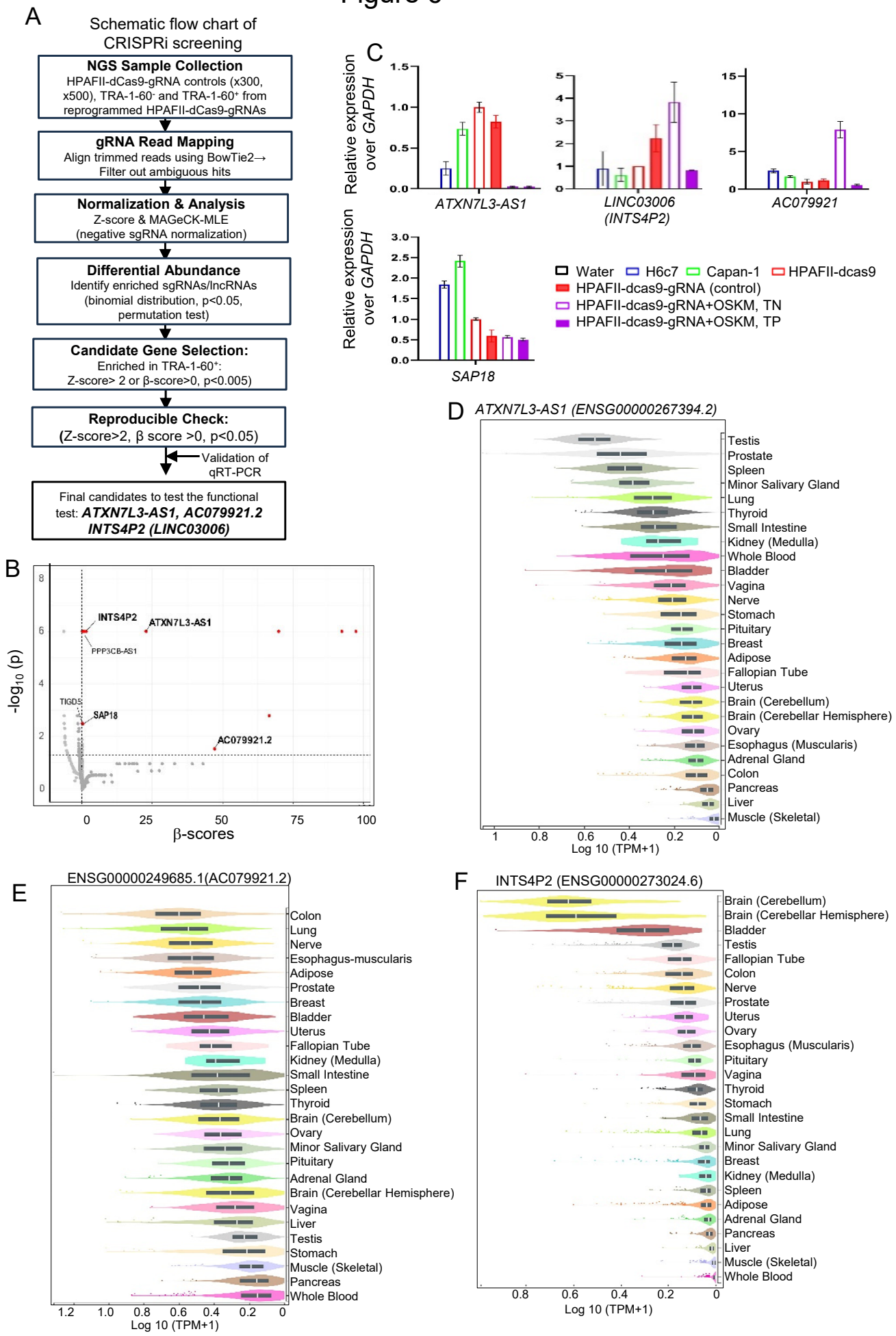


Figure 6. Emergence of ATXN7L3-AS1, INTS4P2, and AC079921 lncRNAs as PDAC cell identity factors.

A. Schematic flow chart of CRISPRi screening B. Significantly enriched lncRNAs in HPAFII-dCas9-gRNA-TP cells in comparison to the HPAFII-dCas9-gRNA reference sample are visualized in the volcano plot. The enrichment score of lncRNAs is shown as the "β scores." Positive and negative β scores indicate positive and negative selections in the TRA-1-60$^+$ compared to the reference condition, respectively. The X-axis shows the β-score of TRA-1-60$^+$ compared to the reference. The Y-axis displays –log10 (p-values). The p-value is calculated by ranking genes with a skewed distribution compared to a uniform null model and permuting to calculate the significance of the skew in distribution. β-score and p-values have been calculated separately from two independent experiments and combined to display the top-ranked lncRNAs in one volcano plot. The horizontal and vertical lines indicate the cutoff threshold. The red dots indicate lncRNAs with a β-score over zero and a permutation p-value <0.05. The candidate lncRNAs that were validated by qRT-PCR were highlighted with names. C. The expression of candidate lncRNAs in normal pancreatic cells and cancers, and HPAFII-dCas9-gRNA-OSKM-TN and TP cells (n=3, y-axis indicates relative expression over GAPDH, error bars=SD). D-F. Expression level of candidate lncRNAs in human normal tissues (GTEx). Expression values are in TPM (Transcripts per Million), calculated from a gene model with isoforms collapsed to a single gene. The Y-axis shows $\log_{10}$(TPM+1). Box plots are shown as median and 25th and 75th percentiles; points are displayed as outliers if they are above or below 1.5 times the interquartile range. Pancreatic tissues (n = 362) were included in the study.

Grygoryev & Lee/Kim et al.,2025
Figure 7

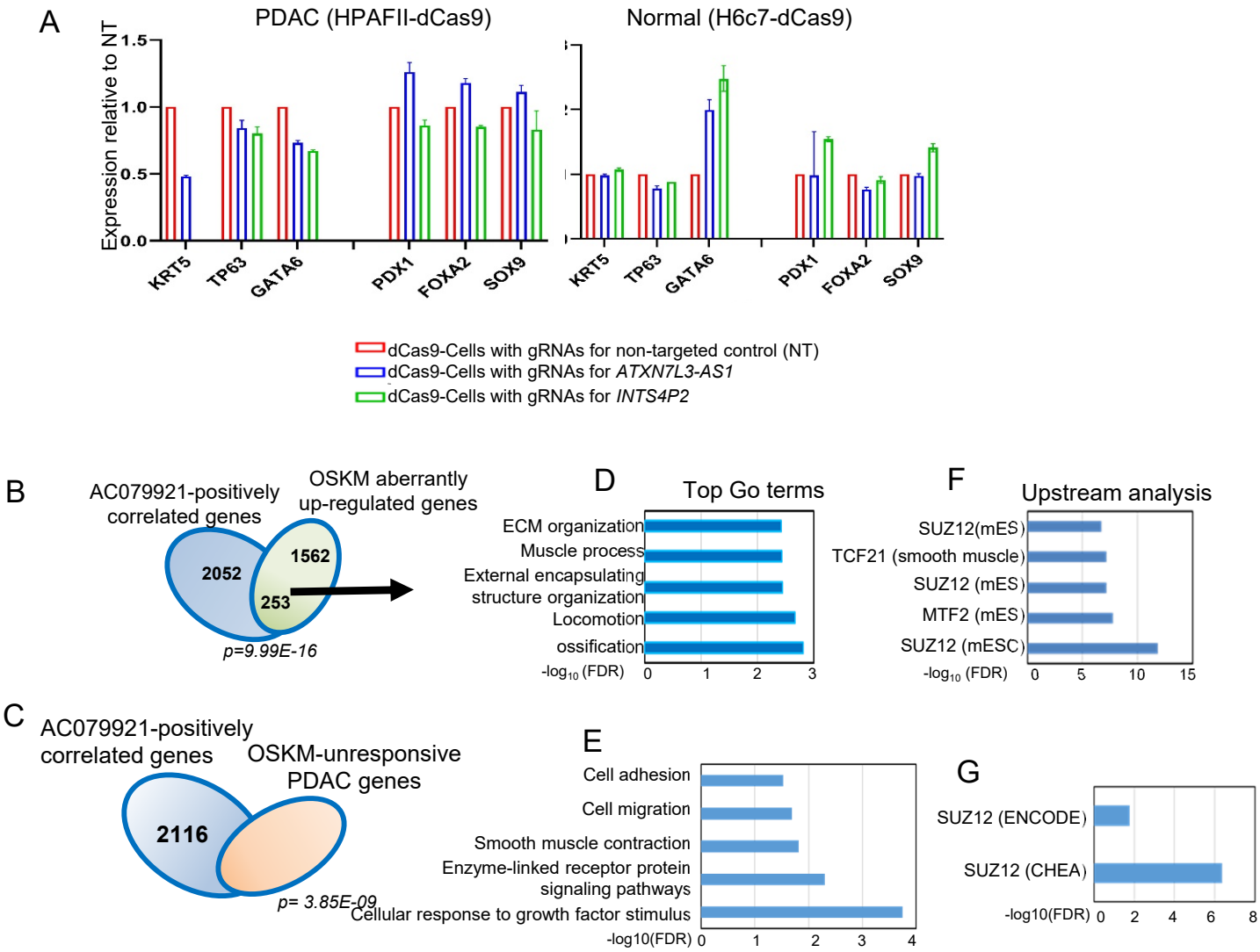
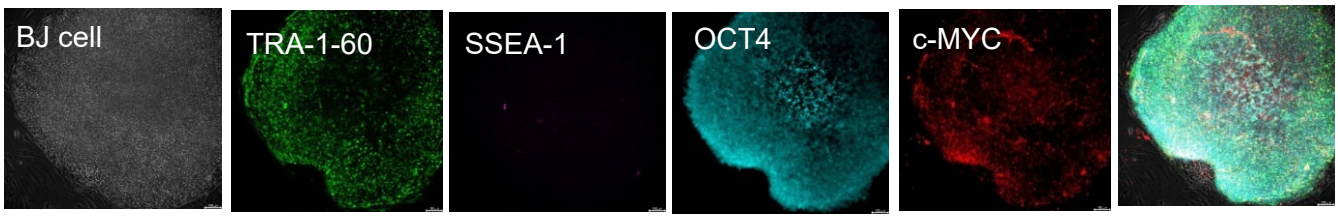


Figure 7. ATXN7L3-AS1 Knockdown Reduces PDAC Marker Expression; AC079921.2 Correlates with ECM Remodeling and Migration Programs

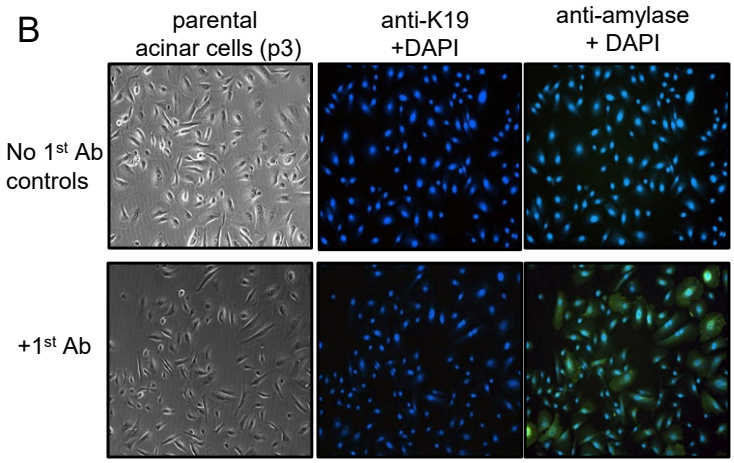
A. Relative expression of PDAC marker genes (basal: KRT5, TP63; classical: GATA6) and endodermal markers (PDX1, FOXA2, SOX9) in HPAFII PDAC and H6c7 hnPDEC cells expressing gRNAs targeting ATXN7L3-AS1 or INTS4P2 (n = 3; y-axis shows expression relative to non-targeting control; error bars indicate SD). B–G. Intersection of OSKM-aberrantly upregulated or unresponsive genes with AC079921.2–positively correlated genes in Pan TCGA tumors. Venn diagrams show overlaps between genes that were aberrantly upregulated (B) or unresponsive to OSKM (C) and genes that positively correlate with AC079921.2 expression across pan-TCGA tumors. Statistical significance was assessed using the hypergeometric test. Top 5 enriched GO terms among overlapping genes (D-E). The top 5 upstream regulators predicted from the overlapping gene sets (F-G).

Grygoryev & Lee/Kim et al.,2025
Figure S1

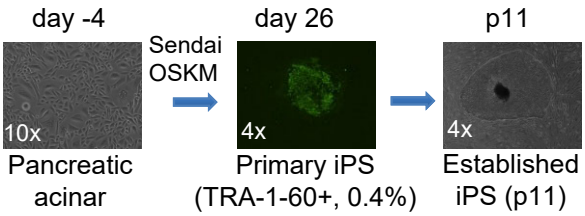
A. Immunostaining of pluripotency markers in BJ6 established iPSC



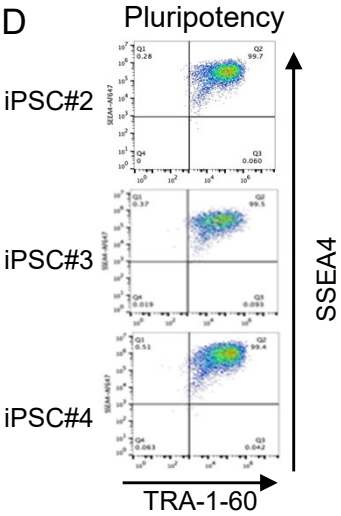
B



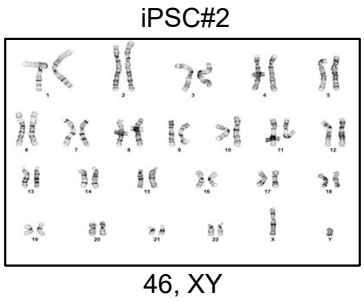
C



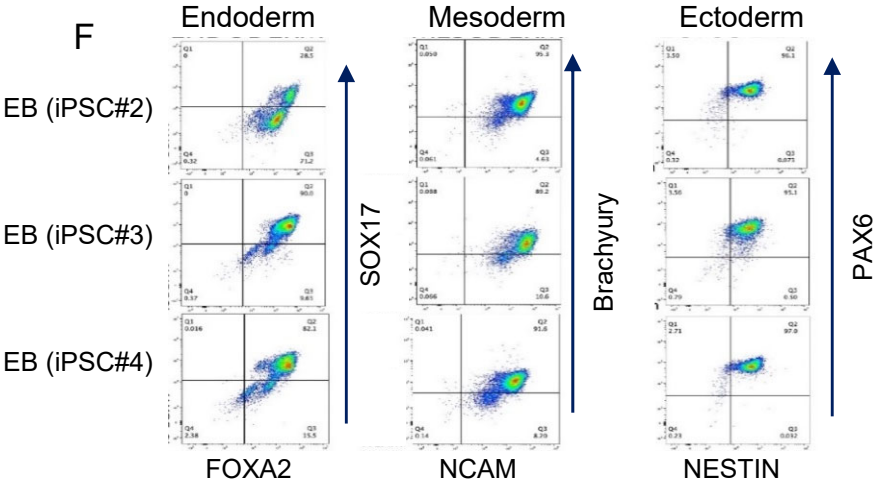
D



E



F



SUPPLEMENTAL FIGURE and TABLE LEGENDS

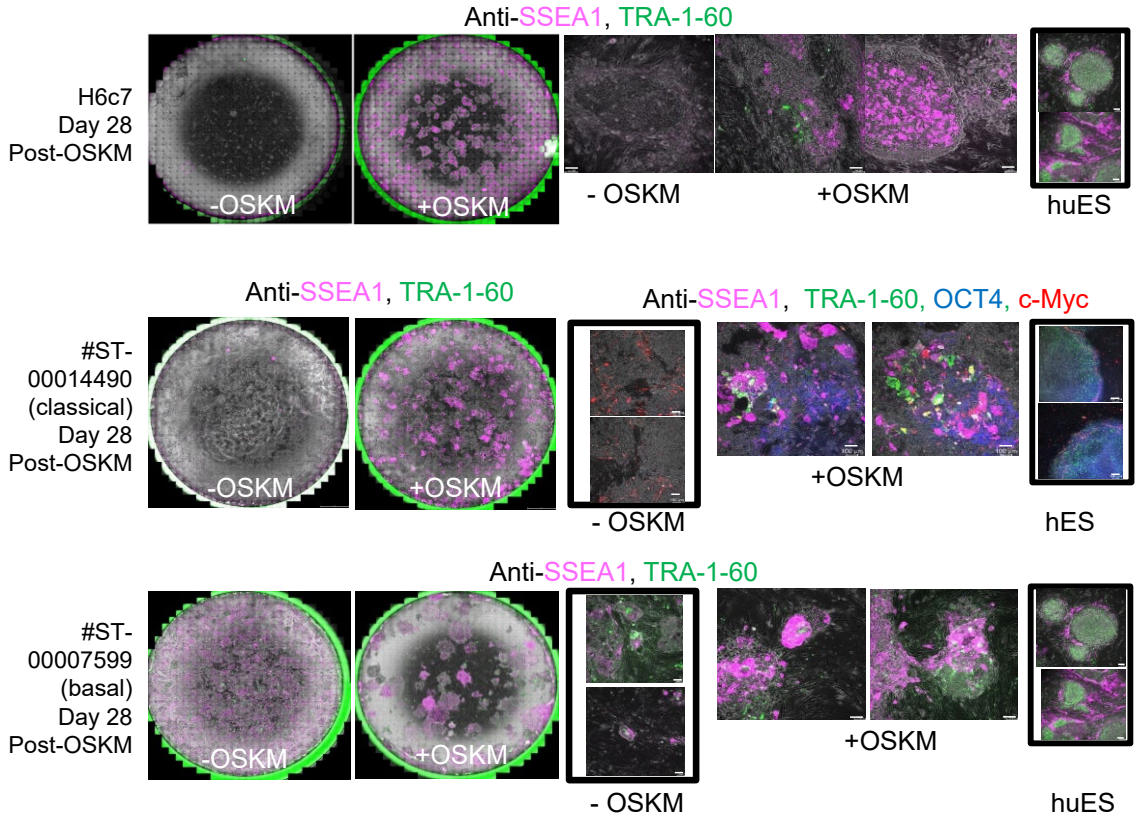
Supplementary Figures.

Figure S1 (Relevant to Figure 1). Transgene-free reprogramming of human normal fibroblast and pancreatic acinar cells into pluripotency using Sendai-virus mediated expression of OSKM master TFs.

A. Immunostaining of pluripotency markers in iPSCs established from BJ6 human normal fibroblast cells. B. Verification of expression of pancreatic acinar marker amylase and lack of pancreatic ductal marker K19 on pancreatic acinar cells using immunofluorescence (IF). C. Schematic diagram of reprogramming pancreatic acinar cells into iPSC using Sendai-OSKM. Initial numbers of cells and emerged TRA-1-60^{+} cells after 26 days were counted to determine the reprogramming efficiency. Three clones were established after five passages. D. Confirmation of expression of pluripotency surface markers TRA-1-60 and SSEA4 in established acinar iPS lines through flow cytometry. E. Representative image of cytogenetic analysis on 20 G-banded metaphase cells derived from pancreatic acinar-derived iPSC. F. Confirmation of functional pluripotency of acinar-derived iPSC clones. Their embryoid bodies (EB) differentiated into three germ layer lineages, which were verified by staining for three germ layer markers using flow cytometry.

Grygoryev & Lee/Kim et al.,2025
Figure S2

A. Immunostaining of pluripotency and early-differentiation markers



B. Expression of OSKM, pluripotency, and PDAC genes on day 7 after OSKM induction

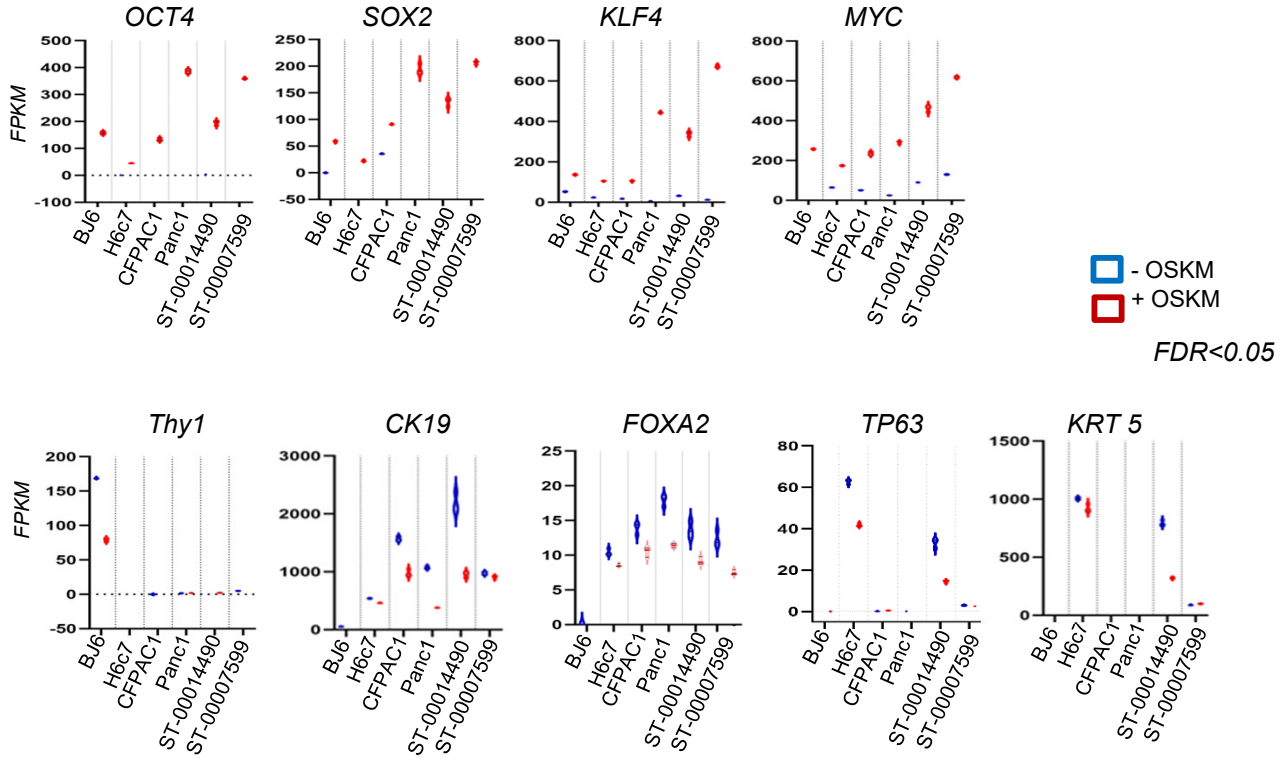


Figure S2 (Relevant to Figure 1). Reprogramming of human normal pancreatic ductal and cancer cells. A. Representative images of colonies stained for TRA-1-60 and SSEA-1 in indicated cells. Scale bars indicate 100 μm. B. Expression of OSKM, pluripotency, and PDAC genes in the indicated cells before and 7 days after OSKM induction.

Grygoryev & Lee/Kim et al.,2025
Figure S3

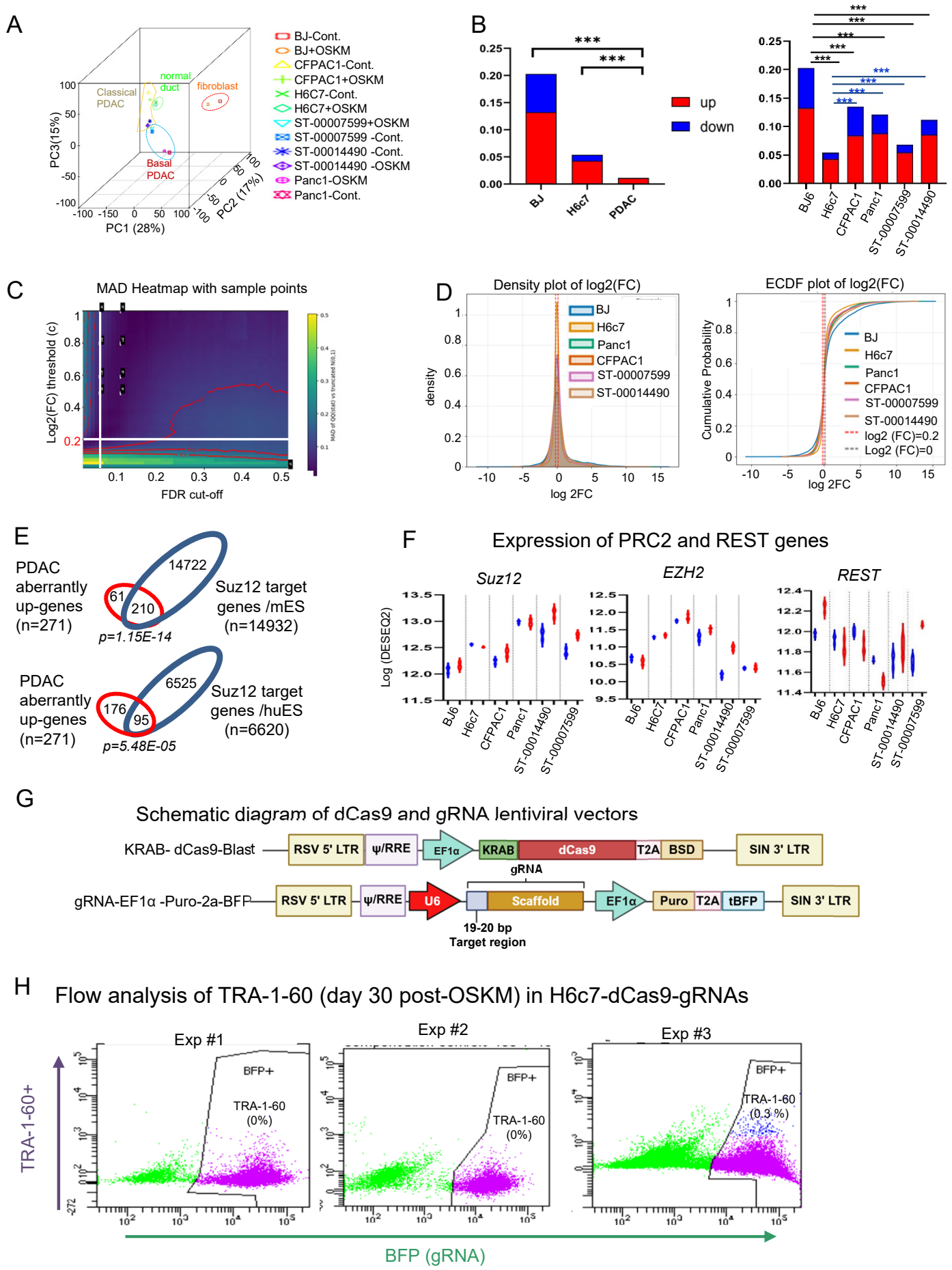


Figure S3 (Relevant to Figures 1-2). A-F. RNA-seq of classical (CFPAC1 and ST-00014490) and basal (Panc1 and ST-00007599) subtypes of PDAC and normal controls (human BJ6 fibroblast and H6c7 hnPDECs) before and 7 days after OSKM induction. (A) PCA plot. (B) Proportion of genes changed upon OSKM induction among a total of 302,383 genes in normal cells (BJ, H6c7) and PDAC (CFPAC1, ST-00014490, Panc1, and ST-00007599). A chi-squared test was used to compare the proportion of genes exhibiting expression changes following OSKM induction in PDAC cells relative to other normal cell types. The Y-axis indicates the proportion of DEGs relative to the total number of genes. *** indicates p-value < 2.2×10^{-13}. Black and blue bars indicate comparison between PDAC cells and BJ or H6c7 cells, respectively (See Table S3). (C) Mean Absolute Deviation (MAD) heatmap of Q–Q statistics across FDR cut-off points and various $\log_2$ (FC) truncated thresholds. The heatmap displays the MAD between observed Q–Q statistics and a truncated standard normal distribution (N(0,1)), computed across combinations of FDR cutoffs (x-axis) and $\log_2$ FC thresholds (y-axis). The color key indicates MAD of Q-Q (stat) vs truncated N (0,1). Lower MAD values (dark blue) indicate better agreement with the expected null distribution, suggesting more reliable gene selection. Red contours mark regions of increasing deviation. White horizontal and vertical lines indicate $\log_2$(FC) = 0.2 and FDR 0.05 thresholds, respectively. (D) Density and Empirical Cumulative Distribution Function (ECDF) plots throughout $\log_2$ FC values across all samples. (E) Intersection of aberrantly upregulated PDAC genes (downregulated in normal but upregulated in PDAC) and Suz12 target genes in human and mouse ES cells (hypergeometric p-value). F. Expressions of genes encoding PRC2 components (SUZ12 and EZH2) and REST transcriptional repressor in indicated cells before and after OSKM induction.

G-H. CRISPRi against cancer-associated lncRNAs. (G) Schematic diagram of KRAB-dCas9-Blastidin and gRNA-EF1a-Puro2a-BFP vectors. (H) Flow cytometry analysis of TRA-1-60$^+$ cells from H6c7 cells on day 30 after OSKM induction, along with lncRNA depletion.

Grygoryev & Lee/Kim et al.,2025
Figure S4

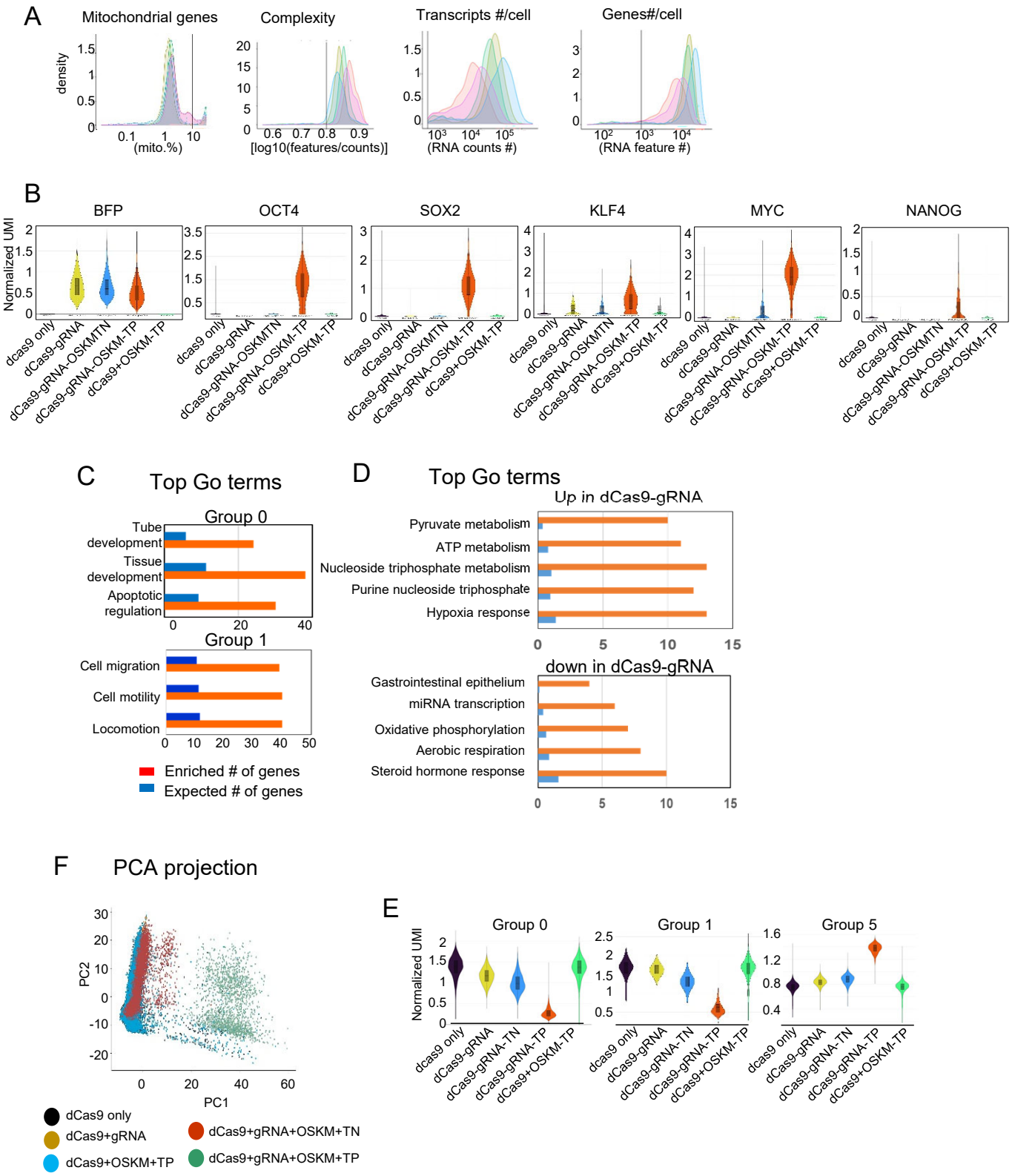
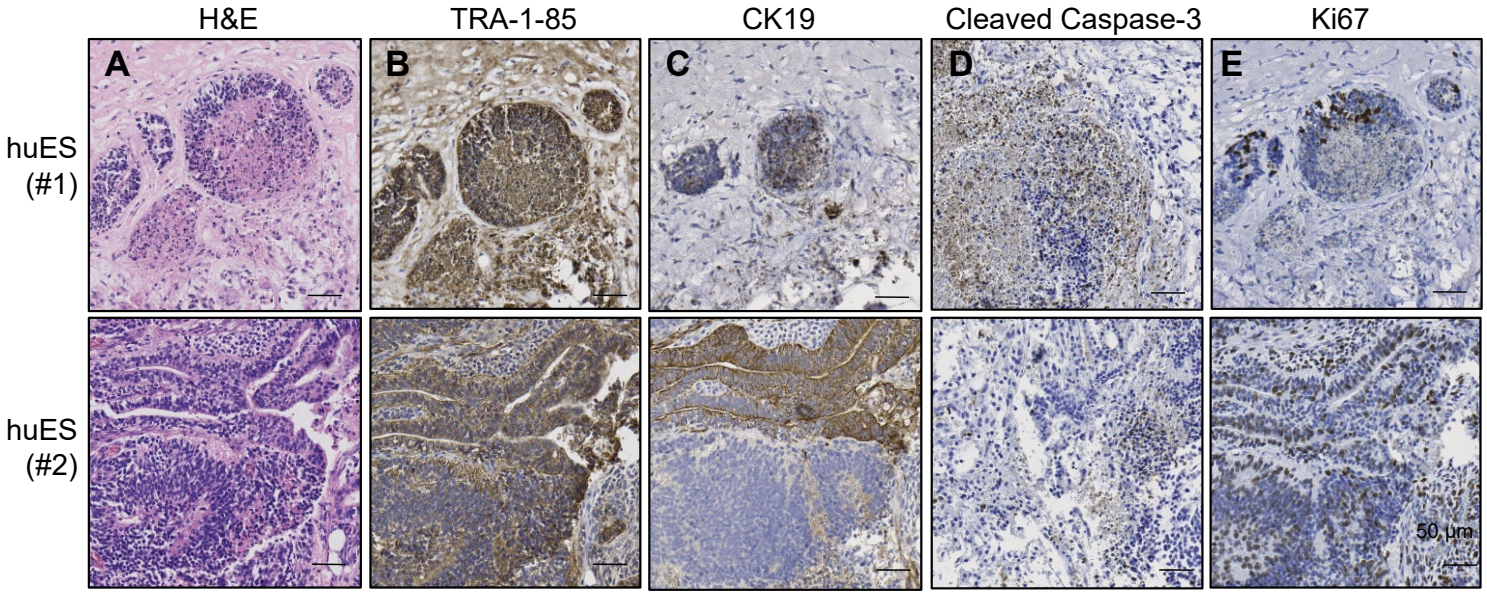
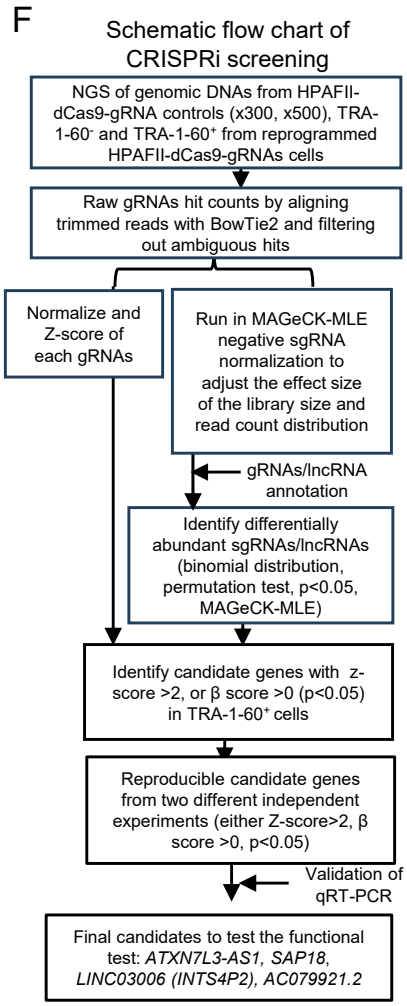
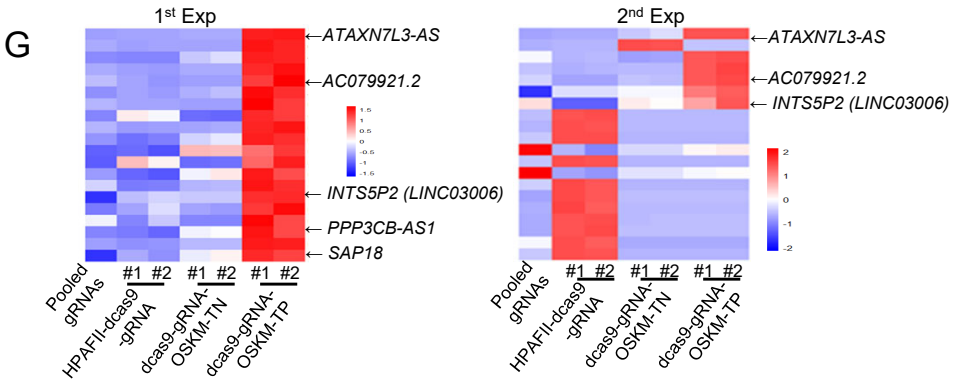


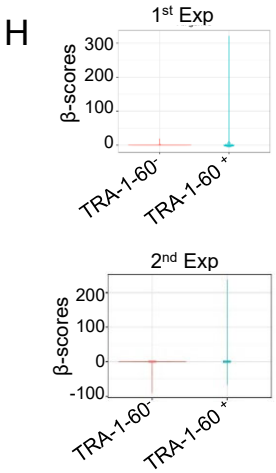
Figure S4 (Relevant to Figure 3) scRNAseq of HPAFII-dCas9-gRNA-OSKM-TP cells along with controls. A. Quality control density plots showing the distribution of cells by key metrics: percentage of mitochondrial gene expression, complexity (features per count), total transcripts per cell, and number of detected genes per cell. B. Expression levels of *BFP*, *OSKM*, and *NANOG* across samples (dCas9, dCas9-gRNA, dCas9-gRNA-OSKM-TN, dCas9-gRNA-OSKM-TP, dCas0-OSKM-TP). The X-axis represents individual samples, and the Y-axis shows normalized UMI counts. C. Top GO terms enriched in genes from Group 0 and Group 1 clusters. D. Top GO terms associated with DEGs in dCas9-gRNA cells compared to parental dCas9 cells. E. Violin plots showing normalized UMI counts for genes in cells from groups 0, 1, and 5 across samples. F. PCA projection of each sample.

Grygoryev & Lee/Kim et al.,2025
Figure S5









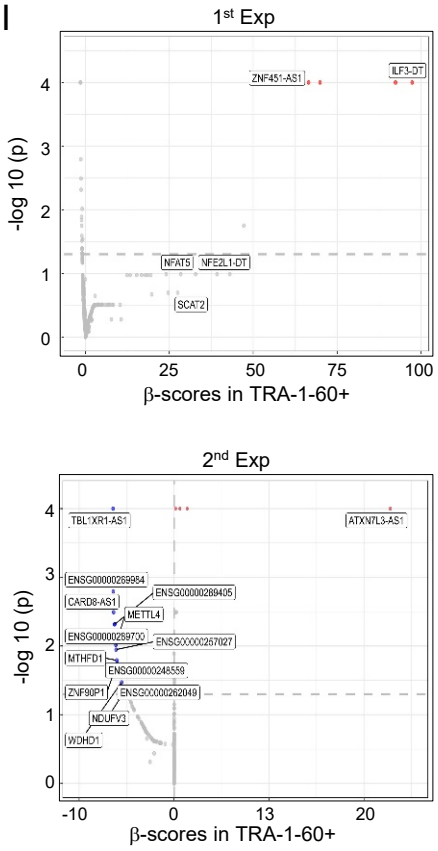


Figure S5 (Relevant to Figure 4-6) A-E. Xenograft of huES cells; Representative images of immunostaining of H&E (A), human-specific marker TRA-1-85 (B), human ductal marker CK19 (C), apoptotic marker Cleaved Caspase-3 (D), and proliferation marker Ki67 (E). F-I. Identification of candidate gRNAs reproducible from two independent experiments. Detailed diagram of selecting candidate lncRNAs (F). Z-score of top 20 gRNAs (G). The estimated β-score plots of TRA-1-60^{+} and TRA-1-60^{-} cells compared to references (HPAFII-dCas9-gRNA). The β scores of all gRNAs are normalized based on the negative control normalization (H). Volcano plot of top-ranked lncRNAs showing β scores enriched in TRA-1-60^{+} cells compared to the reference control. The Y-axis displays $-\log_{10}$ (p-values). The p-value is calculated by ranking genes with a skewed distribution compared to a uniform null model and permuting to calculate the significance of the skew in distribution. The X-axis indicates the β scores in TRA-1-60^{+}. Red indicates positive enrichment (beta score >0) and blue indicates negative enrichment (beta score <0) (I).

Grygoryev & Lee/Kim et al.,2025
Figure S6

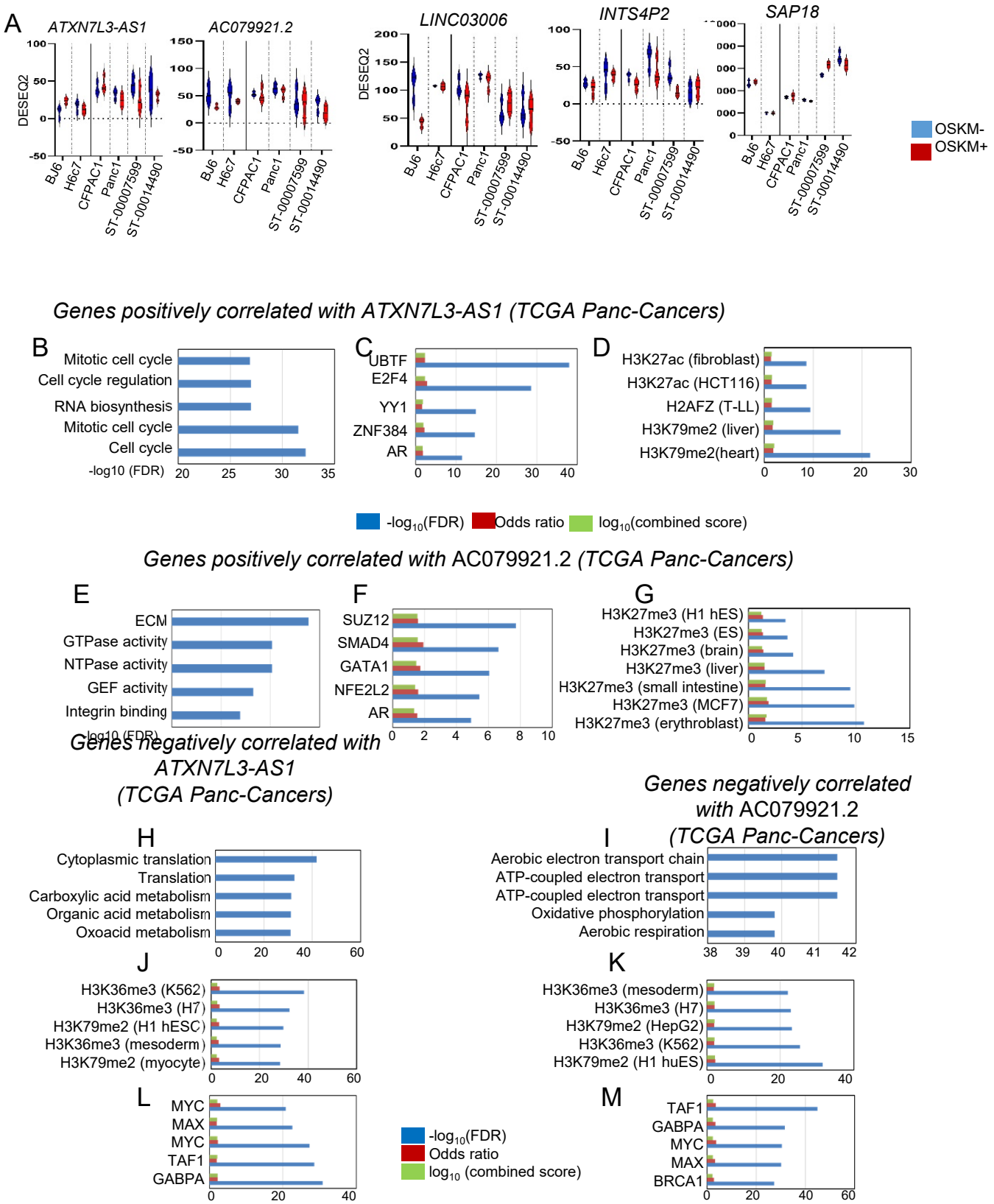


Figure S6 (Relevant to Figure 6-7) Association of ATXN7L3-AS1 with mitotic cell cycle progression and AC079921.2 with cell adhesion and migration in the TCGA dataset. A. Expression of candidate lncRNAs before and after OSKM induction. B-D. Top GO terms (B), potential upstream regulators (C), and histone modifications (D) enriched in the upstream of genes whose expressions are positively correlated with ATXN7L3-AS1 mRNA levels in TCGA Pan-cancer (Spearman correlation >0.4, n=2887). E-G. Top GO terms of (E), potential upstream regulators (F), and histone modifications of the upstream (G) of genes whose expressions are positively associated with AC079921.2 in TCGA Pan-cancer (Spearman correlation >0.4, 2305). H-M. Top GO terms of (H, I), histone modification of the upstream (J, K), potential upstream regulators (L, M) in of genes whose expressions are negatively associated with ATXN7L3-AS1 and AC079921.2 in TCGA Pan-cancer, respectively (Spearman correlation <-0.4, ATXN7L3-AS1, n=1928; AC079921.2, n=667).