Long Noncoding RNAs Preserve Pancreatic Cancer Identity and Resist Cell Fate Conversion

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 E. Furth[12], Sudhir Thakurela[5,13,14], Rosalie Sears[2,6,8], and Jungsun Kim[1,2,8,15]

[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, Massachusetts General Hospital, 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.
[13]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA
[14]Broad Institute of MIT and Harvard, Cambridge, MA, USA,

* Equal contribution
[15]corresponding author: Dr. Jungsun Kim (kimjungs@ohsu.edu)

Keywords: cancer cell identity; transcriptional memory; transcriptional rigidity; long noncoding RNA; master transcription factor-mediated cellular reprogramming; pancreatic cancer; pluripotency; plasticity; *ATXN7L3-AS1*; *AC079921.2*

Summary

The Yamanaka factors (OCT4, SOX2, KLF4, and MYC; OSKM) can rejuvenate aging phenotypes in somatic cell types by resetting the epigenetic landscape. Curiously, most solid tumor cells remain largely resistant to reprogramming despite their well-documented plasticity, and the underlying mechanisms are unclear. Here, we combined genomic profiling and *in vivo* assays to investigate OSKM-mediated reprogramming of pancreatic ductal adenocarcinoma (PDAC). In the initial stages, we found that cancer-specific genes were refractory while mesodermal/ECM programs, normally silenced by PRC2, were aberrantly upregulated. A CRISPR interference screen for OSKM reprogramming coupled with functional analyses revealed that suppression of cancer-associated long noncoding RNAs (lncRNAs) erased malignant epithelial programs, restored tumor suppressor activity, and impaired tumorigenicity *in vivo*. We further identified that *ATXN7L3-AS1* lncRNA sustains the PDAC malignant identity through its association with active epithelial oncogenic programs and poised PRC2-targeted developmental genes, thereby supporting both plasticity and memory. Thus, by exploring why cancer cells are resistant to reprogramming, we identify lncRNAs as gatekeepers of malignant identity, suggesting that targeting lncRNAs could be a generalizable therapeutic strategy in treating solid tumors.

Introduction

Cellular reprogramming through forced expression of lineage-defining transcription factors (TFs), including pioneer factors, can erase established cell identity and impose new cell fates[1-7]. The first demonstration of induced pluripotency was achieved by combined expression of the TFs OCT4, SOX2, KLF4, and MYC (OSKM, or the Yamanaka factors), which reprogrammed normal somatic cells into induced pluripotent stem cells (iPSCs)[6, 7]. Beyond generating iPSCs, OSKM expression has been reported to rejuvenate aging phenotypes by reversing age-associated epigenetic alterations[8, 9]. Because many cancers segregate with aging and are partly

driven by epigenetic dysregulation, reprogramming has been proposed as a potential strategy for resetting the malignant cell fate.

However, with the exception of neuronal lineage and hematologic malignancies[10-12], solid tumors exhibit broad resistance to OSKM reprogramming[13-17]. While combined expression of the four factors partially reprograms solid cancer cells, the resulting states are often unstable: cells either stall in intermediate states[13] or revert to malignant phenotypes[14, 15]. Persistent ectopic factor expression can attenuate malignant features and maintain a “tamed” state; yet these cells remain refractory and do not undergo complete reprogramming[16]. In some contexts, ectopic expressions of OSKM or partial reprogramming even reinforce tumor-like programs[17]. These findings suggest that reprogrammed cancer cells retain a transcriptional and epigenetic memory of their tissue of origin and therefore resist complete erasure of cell identity[18, 19]. Given that the plasticity of cancer cells is well-documented [20-25], this striking resistance suggests the existence of robust and cancer-specific transcriptional memory mechanisms. Yet, the molecular basis of such a memory mechanism remains poorly understood.

PDAC, one of the deadliest human cancers in aging patients[26], exemplifies the paradox of a cell state that is both plastic and fixed. In PDAC cancer cells, transcriptional transitions between classic and basal subtypes are dynamic[27-31]. Yet, cells maintain a stable malignant identity defined by oncogenic programs and ductal epithelial signatures that underpin therapeutic resistance[32, 33]. Identifying mechanisms that can maintain this malignant identity, despite its dynamic transcriptional states, is crucial for developing effective therapeutic strategies against this lethal cancer.

Gene regulatory networks comprise TFs, cofactors, and chromatin regulators that tightly coordinate the maintenance of cell identity[34-37]. Recently, long noncoding RNAs (lncRNAs) have emerged as key epigenetic regulators, exhibiting greater tissue specificity than most protein-coding genes[38-41]. By scaffolding chromatin-remodeling complexes and guiding gene expression in trans and cis, lncRNAs are well poised to orchestrate cell-type-specific

transcriptional programs[42-49]. As such, they have essential roles in embryonic development, pluripotency, and differentiation[49-53], and influence reprogramming by modulating microRNAs that control core pluripotency factors[53, 54]. Conversely, dysregulation of lncRNAs is increasingly recognized as a disease-driving mechanism in cancer, contributing to both disease progression and therapeutic resistance[55-58]. Yet, specific lncRNAs in solid tumors such as PDAC that act to preserve malignant identity and block cell fate conversion have not been identified.

Here, we leveraged PDAC as an aging-associated cancer model to investigate how OSKM reprogramming is constrained in the context of cancer-specific transcriptional memory. Specifically, we coupled a systematic CRISPR interference (CRISPRi) screen with OSKM reprogramming, integrative genomic analyses, and *in vivo* xenograft assays. Through this approach, we identified a subset of cancer-associated lncRNAs that enforce the coexistence of plasticity and epithelial memory, thereby preserving malignant ductal identity and impeding reprogramming.

RESULTS

PDAC Cells Resist OSKM-Reprogramming, Revealing Restricted Plasticity. PDAC comprises two major transcriptional subtypes: classic and basal-like (also referred to as squamous)[27-29]. In prior work, we implanted human primary PDAC tissue into immunodeficient mice to establish and fully characterize patient-derived xenograft (PDXs) tumor models, thereby avoiding selection against PDAC during reprogramming (Table S1)[59]. We thus performed OSKM reprogramming in PDAC cells representing both classical (CFPAC1, ST-00014490 PDX) and basal subtypes (Panc1, ST-00007599 PDX) [59, 60] using BJ fibroblasts and H6c7 human normal pancreatic duct epithelial cells (hnPDECs) as normal controls (Fig. 1A).

Instead of transgene-mediated reprogramming, we utilized Sendai virus to deliver the OSKM-expressing polycistronic vector[60] as it provides a clean platform for reprogramming somatic cells without causing genomic integration[61]. Importantly, this system achieved efficient

and consistent transgene expression across all primary PDAC cells, comparable to that in H6c7 hnPDECs and BJ fibroblasts[60].

We first validated the Sendai virus–mediated OSKM delivery system using BJ fibroblasts and acinar cells isolated from pancreatitis tissue of PDAC patients (Fig. S1A-S1B). The reprogramming efficiency of acinar cells (0.4%) was comparable to that of fibroblasts (Fig. S1C). The resulting iPSC lines expressed the canonical pluripotency marker TRA-1-60[62, 63], exhibited normal karyotypes, and successfully differentiated into all three germ layers (Fig. S1D-F), validating 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. However, many cells expressed SSEA-1, a marker associated with pluripotency in mouse cells and with early embryonic differentiation in human cells[64, 65] (Fig. S2A). Because TRA-1-60 expression emerged by day 45-60[66], we concluded that hnPDECs retain a stronger lineage memory than acinar cells or fibroblasts. When we conducted OSKM induction in classical-type PDAC cells, we detected expression of SSEA-1 but not TRA-1-60, consistent with stalled progression toward pluripotency (Fig. S2A). Similarly, in OSKM-induced basal subtype PDAC cells, TRA-1-60 expression was sporadic while SSEA-1 expression was unaffected, indicating resistance to transcriptional rewiring (Fig.S2A). Even after 60 days of OSKM induction, PDAC cells failed to achieve full reprogramming. Altogether, these findings indicate that PDAC cells maintain stable, cancer-associated programs that prevent complete cell fate conversion.

OSKM Overrides Only a Subset of PDAC Identity Genes. Early molecular events following OSKM induction are critical in determining cell fate outcomes[3, 67-73]. To characterize such early transcriptional responses, we examined transcriptomic profiles 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), with Principal Component Analysis (PCA) showing consistent patterns across technical replicates within each sample (Fig. S3A).

OSKM initiates reprogramming of somatic cells by silencing their original lineage-specific gene programs[3, 67-73]. We therefore hypothesized that OSKM might downregulate a subset of the PDAC-specific transcriptome, particularly those involved in the malignant identity. To test this hypothesis, we defined PDAC-unique genes by identifying transcripts significantly upregulated across all PDAC cells relative to both fibroblasts and hnPDECs prior to OSKM reprogramming (Fig. 1B, Table S2A). Out of 375 PDAC-unique genes, we only found 7 that were significantly downregulated on day 7 following OSKM induction (Fig. 1B, Table S2B), several of which were associated with apoptosis and epithelial differentiation pathways (Table S2C). These findings suggest that early reprogramming in PDAC is constrained by restrictive regulatory mechanisms. Notably, one of the OSKM-responsive genes was SPDEF, which encodes the ETS transcription factor with a SAM-pointed domain[74], a newly reported driver of classical PDAC tumorigenesis[75]. As expected, we found that high expression of SPDEF correlated with poor prognosis (Fig. 1C-D).

Next, we assessed transcriptional responses to OSKM 7 days after induction across different cell types. We detected the greatest number of differentially expressed genes (DEGs) in BJ fibroblasts. In contrast, we detected fewer in H6c7 hnPDECs, and markedly fewer DEGs in PDAC cells overall (Fig. S3B, Table S2D). These variations likely reflect inter-sample heterogeneity, as we detected more DEGs in each PDAC sample than in H6c7 hnPDECs (Fig. S3B, Table S2E) but also indicate higher intrinsic transcriptional plasticity in PDAC. Importantly, this plasticity was not accompanied by loss of malignant identity. Instead, it was constrained by robust transcriptional memory that preserved PDAC-specific programs despite OSKM induction.

Thus, while OSKM partially suppressed tumor-associated gene programs, the core transcriptional identity of PDAC cells remained largely intact.

PRC2-Linked Lineage Programs Persist During Reprogramming. Based on our data thus far, we hypothesized that OSKM-unresponsive genes in PDAC act as gatekeepers that preserve malignant identity, whereas OSKM-aberrantly upregulated genes (silenced or unchanged in normal cells) represent maladaptive, cancer-specific barriers that antagonize cell fate conversion. Such aberrant responses could potentially drive maladaptive PDAC plasticity as well as impeding reprogramming.

To identify such mechanisms, and account for inter-sample heterogeneity, we developed a classification framework incorporating variability across PDAC (n=4) and normal (n=2) samples. DEGs before and after OSKM induction were identified using likelihood ratio testing (LRT)[76] and ranked by a weighted Hamming distance[77] relative to predefined patterns. These included: (1) downregulated in normal cells but minimally changed in cancer ("OSKM unresponsive in PDAC"), and (2) unchanged or downregulated in normal cells but upregulated in cancer ("OSKM aberrantly upregulated in PDAC") (Fig. 1E-F and S3C–D, Table S2F-H, method). While OSKM-unresponsive genes were enriched in cell adhesion pathways, OSKM-aberrantly upregulated genes were associated with mesodermal and morphogenetic lineage programs (Fig. 1G–H). Both categories were predicted targets of the PRC2 core components (SUZ12, EZH2) (Fig. 1G-H) and significantly overlapped with known SUZ12 targets in human and mouse embryonic stem cells (mES and huES) (Fig. S3E). Notably, expression of the PRC2 subunit remained unchanged post-OSKM (Fig. S3F), suggesting that impaired PRC2 recruitment—rather than reduced subunit abundance—underlies the failure to suppress these programs.

In summary, OSKM downregulated a small subset of PDAC marker genes and left most PDAC-specific genes unchanged or upregulated. Many of these belonged to PRC2-regulated lineage programs.

Suppressing Cancer-Associated lncRNAs Unlocks Reprogramming Efficiency. Given the broad role of lncRNAs in regulating epigenetic machinery, including recruiting PRC2[42, 78-82], we next performed CRISPRi screens targeting 691 cancer-associated lncRNAs (10 gRNAs per lncRNA plus controls; 7,087 gRNAs total; Table S3A) to identify those sustaining malignant identity[83, 84]. The screens were conducted in the context of OSKM-mediated reprogramming of normal ductal cells and PDAC cells, using the H6c7 and HPAFII cells expressing catalytically inactive Cas9 (dCas9)-KRAB (henceforth referred to as H6c7-dCas9-gRNA and HPAFII-dCas9-gRNA; Fig. 2A).

As expected, TRA-1-60^{+} pluripotent-like colonies were rare in HPAFII dCas9 control cells by day 30 post-OSKM induction. In contrast, cells transduced with the lncRNA-targeting gRNA pool showed an increased frequency of TRA-1-60^{+} cells (Fig. 2B). The isolated TRA-1-60^{+} and TRA-1-60^{-} populations (hereafter referred to as dCas9+gRNA+OSKM-T^{+} and dCas9+gRNA+OSKM-T^{-}, respectively) remained phenotypically stable over multiple passages (Fig. 2C). In contrast, the same gRNA pool targeting lncRNAs did not enhance reprogramming of H6c7 hnPDEC cells (Fig. S3G), suggesting that these lncRNAs act in a PDAC-specific manner to preserve malignant transcriptional memory.

Pooled lncRNA Depletion Enables OSKM to Erase PDAC Identity and Restore Tumor Suppression. Next, we profiled single-cell transcriptomes from OSKM-induced, pooled-CRISPRi PDAC cells (dCas9+gRNA+OSKM-TRA-1-60^{+} [T+] and TRA-1-60^{-} [T-]) (Fig. 3A, S4A–B) to assess whether depletion of cancer-associated lncRNAs was sufficient to override PDAC cell identity and establish a new gene program (pluripotency). While Louvain clustering resolved 15 distinct clusters, control dCas9 cells localized mainly to tubular ductal (Group 0) and invasive epithelial (Group 1) clusters, consistent with PDAC identity (Fig. 3B–C, S4C). As expected, OSKM expression alone was insufficient to erase PDAC programs (Fig. 3B-C, S4C).

Similarly, dCas9–gRNA cells displayed metabolic shifts (pyruvate metabolism up, oxidative phosphorylation down; Fig. S4D) yet remained confined to PDAC-like clusters (Fig. 3B–C).

In contrast, dCas9+gRNA+OSKM-T$^+$ cells redistributed away from PDAC clusters (Groups 0 and 1) and localized predominantly to Group 5, characterized by pluripotency-associated gene expression and loss of pancreatic/germ layer markers (Fig. 3B–J, S4E-F, Table S4A–C). PCA confirmed that T$^+$ cells occupied a transcriptional space distinct from PDAC or normal pancreas and aligned most closely with early human embryo profiles (Fig. 3K).

Both T$^+$ and T$^-$ populations were depleted of mucus-producing and invasive epithelial clusters (Fig. 3B–C), indicating a broad disruption of malignant programs. A curated PDAC-specific gene signature (Table S4D)[85, 86] was significantly downregulated only in the combined lncRNA-depleted OSKM condition (Fig. 3L). While controls expressed few tumor suppressor genes (TSGs), dCas9+gRNA+OSKM-T$^+$ cells showed robust and broad TSG re-expression, enriched within NANOG$^+$ states (Fig. 3M–O), based on comparative analysis of the OncoKB TSG (Table S4E)[87, 88].

Together, these results demonstrate that pooled suppression of cancer-associated lncRNAs during reprogramming erases PDAC transcriptional identity and produce pluripotent-like states where tumor suppressor programs are reactivated, confirming the reversal of malignant fate.

Pooled lncRNA Depletion with OSKM Erases PDAC Identity and Limits Tumorigenesis In Vivo. We next investigated xenograft outcomes in immunodeficient NOD-SCID-IL2Rγ^{null} (NSG) mice[89] to assess whether depleting PDAC identity-associated transcriptional programs would tame cancer progression *in vivo* (Fig. 4A). As expected, parental HPAFII cells formed tumors within one-month after injection, with dCas9-gRNAs producing slightly smaller tumors compared to their parental (Fig. 4B–C). Strikingly, mice injected with dCas9+gRNA+OSKM-T$^+$ cells either

failed to develop tumors or formed significantly smaller tumors than dCas9+gRNA+OSKM-T$^-$ and all controls ($p < 0.005$; Fig. 4B–C).

In histopathology analysis, we confirmed classical grade 2 adenocarcinoma morphology in tumors from all control groups, 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 (Fig. 4D-G, 5A-P). We observed extensive central necrosis, consistent with advanced PDAC pathology. In contrast, tumors derived from dCas9+gRNA+OSKM-T$^+$ cells lacked typical adenocarcinoma features, lacking CK19 and showing robust CC3 immunostaining (Fig. 4H, 5Q-T), suggesting a transition toward a less aggressive phenotype.

In all, our analysis underscores that neither OSKM expression nor lncRNA knockdown alone was sufficient to reprogram the malignant phenotype. At the one-month time point, huES-derived tumors were small, composed predominantly of neuronal epithelial cells[90, 91], which are distinct from HPAFII-derived tumors (Fig. S5A-E). Together, these findings demonstrate that lncRNA depletion in combination with OSKM expression effectively reprogrammed PDAC cells *in vivo*, resulting in reduced tumorigenicity, loss of classical adenocarcinoma identity, and reactivation of apoptotic programs. 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.

Identification of *ATXN7L3-AS1*, *AC079921.2*, and *INTS4P2* as Cancer-Associated lncRNAs That Restrict OSKM Reprogramming. To identify lncRNAs that maintain PDAC gene expression and are selectively depleted during successful reprogramming, we sequenced genomic DNA from FACS-sorted dCas9–gRNA controls, and gRNA+OSKM-T$^+$ and gRNA+OSKM-T$^-$ cells (Fig. 6A, Table S5A–E). After normalizing gRNA counts from two independent cultures using Z-score and MAGeCK negative control normalization[92], we retained

lncRNAs with Z-scores > 2 and positive β-scores ($p < 0.05$) in both cultures (Fig. 6A and S5F-H, Table S5F–K, methods).

This stringent analysis identified several lncRNAs-targeting gRNAs enriched in gRNA+OSKM-T$^+$ cells (Fig. 6B). Among these, *ATXN7L3-AS1* (ENSG00000267394), *AC079921.2* (ENSG00000249685), and *INTS4P2* (ENSG00000273024, within LINC03006:ENSG00000290926) were validated by qRT-PCR and found to be significantly downregulated in HPAFII-dCas9+gRNA+OSKM-T$^+$ cells compared to control HPAFII-dCas9 cells (Fig. S6A). Importantly, these lncRNAs were minimally expressed or absent in normal pancreatic tissue[93] (Fig. 6C-E) and remained unchanged after OSKM induction alone (Fig. S6B).

***ATXN7L3-AS1* Occupies Both PRC2-Poised Developmental and Active Oncogenic Gene Programs in PDAC.** To assess functional properties of candidate lncRNAs, we targeted *ATXN7L3-AS1*, *AC079921.2*, and *INTS4P2* with CRISPR-dCas9 in PDAC cells (HPAFII and CAPAN1) and normal ductal cells (H6c7). Knockdown of *ATXN7L3-AS1*, *AC079921.2*, and *INTS4P2* reduced expression of PDAC-enriched genes, including basal markers (*KRT5, TP63*) and the lineage factor GATA6, while modestly increasing progenitor markers (*PDX1, FOXA2, SOX9*) upon depletion of *ATXN7L3-AS1* and *AC079921.2* (Fig. S6C). Importantly, depletion of *ATXN7L3-AS1* and *AC079921.2* had the opposite effect on normal ductal H6c7 cells, supporting a cancer-specific role for these lncRNAs (Fig. S6D). Notably, the *ATXN7L3-AS1* locus harbors recurrent pathogenic mutations across multiple cancers, underscoring its potential role in establishing and maintaining malignant states (OSMIC[94]; Fig. 6F, Table S6A-B).

Chromatin Isolation of RNA purification (ChIRP-seq)[95] analysis identified that *ATXN7L3-AS1* primarily localizes to CpG island promoters, enriched for ZNF transcription factor motifs, of genes involved in RNA Pol II regulation, actin cytoskeletal remodeling, and cell migration, including neuronal migration (Fig. 7A–D, S7A–B, Table S6C–D). Notably, *ATXN7L3-AS1*–

occupied genes overlapped with binding sites of UBTF, ZBTB7A, and PRC2 components (SUZ12, EZH2), as well as regions marked by H3K27me3 (Fig. 7D, S7C–E).

Next, we examined the expression of *ATXN7L3-AS1*–occupied genes across normal and PDAC cells[96] (Fig. 7E, Table S6E-H). Ward's clustering analysis[97] partitioned these genes into four transcriptional modules with distinct biological signatures: a fibroblast-specific cluster (Group 3, 20%) enriched in developmental programs such as neuronal differentiation and tube formation; a normal ductal cell–associated cluster (Group 2, 13%) dominated by RNA metabolism and transcriptional regulation; a basal subtype–associated cluster (Group 1, 16%) enriched in synaptic signaling and neuronal differentiation, consistent with neuronal mimicry of basal subtypes; and a broad classical subtype–associated cluster (Group 4, 52%) enriched in RNA metabolism, tyrosine kinase signaling, small GTPase pathways, and epithelial morphogenesis (Fig. 7E, Table S6E-H). Upstream regulator analysis revealed cluster-specific differences: SUZ12/PRC2 targets were preferentially enriched in the fibroblast- and basal subtype–associated groups (Groups 3 and 1), consistent with Polycomb-maintained developmental programs. By contrast, UBTF emerged as the top regulator in the broad classical cluster (Group 4), where SUZ12 ranked lower (Fig. 7E).

Consistent with this model, *ATXN7L3-AS1*–bound tubular ductal epithelial development genes are normally silenced by PRC2/H3K27me3 but remain active in PDAC (Fig. 8A–B). Similarly, OSKM-aberrantly upregulated genes, which significantly overlap with *ATXN7L3-AS1*–bound genes, were particularly enriched in migration and morphogenesis pathways (Fig. 8C). Together, these findings suggest that *ATXN7L3-AS1* lncRNA sustains malignant PDAC identity by maintaining epithelial cancer–associated programs in an active state while keeping PRC2-target developmental genes poised, thereby supporting both plasticity and malignant transcriptional memory.

Similarly, *AC079921.2* was associated with the expression of invasive hallmark PDAC genes involved in extracellular matrix (ECM) remodeling (Fig. S8A–C, Table S7) across TCGA Pan-Cancers. OSKM-unresponsive or aberrantly upregulated genes overlapping with these sets were likewise enriched in morphogenesis, migration, and predicted SUZ12 targets (Fig. S8D–E). Thus, *AC079921.2* reinforces the malignant identity of PDAC by sustaining pro-invasive and morphogenesis programs resistant to OSKM.

Finally, functional experiments further validated that *ATXN7L3-AS1* depletion markedly enhanced OSKM-mediated reprogramming in PDAC cells, but not in normal ductal cells (Fig. 8D–F), underscoring its cancer-specific role in maintaining malignant transcriptional memory. Collectively, these findings reveal that *ATXN7L3-AS1* and *AC079921.2* sustain malignant PDAC programs by engaging both PRC2-regulated developmental pathways and active promoters of cancer-associated genes, thereby reinforcing plasticity and malignant memory (Fig. 8G).

Discussion

OSKM-mediated reprogramming can reset cell identity in normal and aged cells by remodeling the epigenome, yet most solid tumors remain largely refractory[13-19]. This resistance has long remained an enigma, given the well-known plasticity of cancer cells and their expression of pluripotency factors. Here, we optimized a high-efficiency, non-integrating gene delivery system[60], amplified primary PDAC cells[59], and developed an analytical framework to account for inter-sample heterogeneity. While normal pancreatic ductal epithelial cells are intrinsically more resistant to reprogramming than fibroblasts or acinar cells, they can ultimately be reprogrammed with sustained expression of OSKM. In contrast, under the optimized conditions, we found that PDAC cells failed to undergo full reprogramming, in part due to the aberrant activation of PRC2-target developmental genes, including those involved in ECM remodeling and mesodermal lineage programs. A subsequent CRISPRi screen revealed that lncRNAs act

as major gatekeepers of malignant identity, driving resistance to reprogramming. Because lncRNAs often act as scaffolds for PRC2 and are expressed at low levels, conventional RNA-seq pipelines may fail to detect those functionally involved in preserving malignant memory. Whether oncogenic KRAS directly contributes to reinforcing such lncRNA-mediated epigenetic barriers is an important question for future studies.

***ATXN7L3-AS1* and *AC079921.2* as Gatekeepers of PDAC Identity.** LncRNAs can contribute to transcriptional memory by shaping chromatin architecture and regulating gene expression through chromatin-modifying enzymes[38-49]. Our data suggest that *ATXN7L3-AS1* acts as a dual regulator of transcriptional memory in PDAC. By binding CpG-rich promoters enriched for ZNF TF binding motifs, it preserves PRC2-target developmental programs such as ECM remodeling and cell migration in a poised, while simultaneously maintaining activation of epithelial cancer-associated genes. This dual activity enables PDAC cells to sustain both latent plasticity and active malignant transcriptional programs, providing a mechanistic basis for their resistance to reprogramming. This is consistent with prior work reporting that lncRNAs scaffold chromatin modifiers preserve lineage fidelity[91-94]. Importantly, *ATXN7L3-AS1* loci harbor pathogenic PDAC mutations (Fig. 6F) and *UBTF*, a known oncogenic TF in other cancers[96-98], suggesting that its role in PDAC spans both structural and regulatory mechanisms.

By contrast, *AC079921.2* expression was correlated with ECM remodeling, cell adhesion, and migration genes—hallmarks of mesenchymal-like and invasive phenotypes (Fig. S8A-C). Like *ATXN7L3-AS1*, these gene sets were enriched for PRC2 targets, suggesting that *AC079921.2* antagonizes PRC2-mediated repression. Supporting this view, genes aberrantly upregulated or unresponsive to OSKM reprogramming significantly overlapped with *AC079921.2*–correlated genes in PDAC (Fig. S8D-I). Thus, *AC079921.2* sustains a mesenchymal-like, invasion-prone transcriptional program that resists reprogramming while promoting adaptability to microenvironmental cues. Whether *ATXN7L3-AS1* or *AC079921.2*

physically interacts with PRC2 to modulate its activity is unknown. Together, these findings establish *ATXN7L3-AS1* and *AC079921.2* as gatekeepers of PDAC identity, acting through distinct yet complementary mechanisms to preserve transcriptional memory and plasticity, thereby stabilizing malignant states.

Combining lncRNA Depletion with OSKM Induction is Required to Erase PDAC Identity. While knockdown of *ATXN7L3-AS1*, *AC079921.2*, or *INTS4P2* moderately reduced PDAC marker expression, neither lncRNA depletion nor OSKM induction alone was sufficient to fully erase PDAC identity or suppress malignancy *in vitro* or *in vivo* (Figs. 3-5). In xenografts, lncRNA depletion modestly reduced tumor size, likely due to decreased proliferation without increased apoptosis, but did not induce reprogramming or erase adenocarcinoma features. Similarly, OSKM expression yielded tumors with more differentiated histology and reduced proliferation, yet these tumors retained core transcriptional programs characteristic of PDAC.

In contrast, combining lncRNA knockdown with OSKM induction (dCas9+gRNA+OSKM-T$^+$ cells) erased adenocarcinoma identity, reactivated apoptotic pathways, and markedly impaired tumorigenesis *in vivo*. Under these conditions, OSKM converted cells into NANOG$^+$ pluripotent-like states in which TSGs were robustly reactivated. Notably, this requirement for pluripotent conversion to enable TSG re-expression is distinct from other reprogramming contexts, where NANOG$^+$ states are often linked to malignant potential. Together, cancer-associated lncRNAs emerge as key determinants of malignant persistence, whose suppression is necessary to unlock the reprogramming potential of OSKM and redirect it away from malignancy.

Toward a new reprogramming cocktail: lncRNA depletion sensitizes PDAC to OSKM-induced cell fate transitions. Our prior work demonstrated that full reprogramming of PDAC cells into a pluripotent state was impeded, with partially reprogrammed states requiring

sustained ectopic OSKM expression[16]. Herein, we extend these findings by proposing that lncRNA depletion sensitizes PDAC cells to OSKM-mediated reprogramming by weakening epigenetic barriers that preserve malignant identity. To enhance clinical relevance, we optimized a non-integrating, Sendai virus–mediated system for transient OSKM expression across multiple patient-derived PDAC cells[60].

Beyond pluripotency, lineage reprogramming of cancer cells into alternative fates—such as dendritic cells—has been demonstrated[98-100]. Whether lncRNA depletion similarly sensitizes PDAC cells to acquire non-permissive cell states[101-103] remains an open question. Targeting lncRNAs enforcing transcriptional memory could enhance responsiveness to reprogramming cues and improve lineage fidelity.

Conclusions. Here, we integrate prior work on OSKM reprogramming with PRC2 biology and lncRNA-mediated transcriptional memory to explain how PDAC cells can simultaneously be plastic and rigid. Specifically, we show that cancer-associated lncRNAs act as gatekeepers of malignant identity, preserving ductal epithelial and invasion-prone programs that resist reprogramming. Targeting these lncRNAs removes this barrier, unlocking the capacity of OSKM to erase malignant identity, reactivate tumor suppressor pathways, and impair tumorigenesis. This conceptual framework establishes malignant memory as an actively maintained epigenetic state and suggests that lncRNA-directed therapies may sensitize tumors to differentiation- and reprogramming-based therapeutic interventions in pancreatic cancer.

Limitations of the Study. Our study focuses on PDAC-specific lncRNAs that sustain the cancer identity and utilizes OSKM to induce the reprogramming of PDAC toward pluripotency. However, lncRNAs that sustain the cancer identity may differ across cancer types, and the effective reprogramming into other lineages will likely require tailored combinations of lineage-specific transcription factors.

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 on 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.

ACKNOWLEDGMENTS We thank Drs. Ken Zaret and Theresa Lusardi for comments on the manuscript. We thank Dr. Jeffrey Drebin for providing de-identified resected tumor specimens for the initial phase of this study, which was exempt from IRB review. 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. Life Science Editors edited this article.

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; scRNAseq analysis supervision, 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, CMF, SR, and JK; Visualization, DG, CMF, SR, AH, SF, and ST, Supervision, JK; Project Administration, JK; Funding Acquisition, JK.

DECLARATION OF INTERESTS

The authors declare no competing interests.

METHODS

Derivation of Patient-Derived Xenografts (PDX) and Tissue Culture. De-identified human PDAC specimens were obtained under the University of Pennsylvania study (IRB-exempted) or the Oregon Pancreas Tissue Registry protocol (IRB00003609). Written informed consent was obtained from all participants. The OHSU Institutional Review Board approved all experimental protocols. All methods were carried out in accordance with relevant guidelines and regulations. All animal work for PDX tumors was 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[59]. Culture conditions for H6c7 and other PDAC cell lines are described[59]. BJ6 fibroblasts were cultured in DMEM supplemented with 10% fetal bovine serum (FBS)[3].

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)[104]. Viral titers for Sendai virus and lentivirus were described in our earlier publication[60]. On day 7 post-transduction, cells were harvested for RNA sequencing and re-seeded onto irradiated mouse embryonic fibroblasts (MEFs) to support further reprogramming until 30 days.

CRISPR interference (CRISPRi) Screen. 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, Millipore, Table S3A)[83, 84]. The HPAFII and H6c7 cells expressing dCas9-KRAB 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 dCas9

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 reference controls (300x, 500x). HPAFII and H6c7 cells were reprogrammed with OSKM Sendai viruses using the OSKM Sendai viruses[104]. The CRISPRi-targeted cells were always maintained at 300-500 coverage throughout the procedure.

For downstream analysis, cells were harvested after 30 min incubation with AccuMAX and processed using a mouse cell depletion kit as previously described[59]. For flow cytometry analysis or sorting, DNase I (0.1 mg/ml) was added to prevent cell aggregation, and TRA-1-85^{+} expression was used to identify human cells, while BFP marked gRNA-expressing cells.

RNA-seq Data Processing and Normalization. Bulk RNA-seq was performed as previously described[60, 105]. 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)[106]. FeatureCounts (v1.5.0-p3) were 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 as needed. 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 DEGs from count data[76]. 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 are available on NCBI GSE247243.

Identification of PDAC Gene Signatures Repressed by OSKM. To identify PDAC-specific gene categories, we first performed differential expression analysis using DESeq2 to compare each PDAC cell 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 samples to define a core set of PDAC-upregulated genes (Table S2A). 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, Table S2B).

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[76]. To establish robust thresholds for DEG calling, we evaluated how closely observed DEG statistics matched the expected null hypothesis distribution. Specifically, we constructed empirical quantile-quantile (Q-Q) plots comparing DEG statistics from genes presumed to be non-differentially expressed (those with simultaneously high adjusted p-values and low log2(FC)) against their expected null distribution. We applied Huber regression to these Q-Q plots and measured the mean absolute deviation (MAD) between the fitted regression line and the identity line over the range of observed statistics. Higher MAD values indicate greater divergence from the null distribution, reflecting genuine biological signals. MAD values were computed across a range of adjusted p-values and log2(FC) threshold combinations and visualized as a heatmap. Based on this analysis, genes with log2(FC) > 0.2 and FDR < 0.05 were considered significantly differentially expressed, as these thresholds corresponded to a clear elevation in MAD.

To classify gene expression responses to OSKM induction, we analyzed DEGs across six cell types—two normal controls and four PDAC cells ([normal#1][normal#2][cancer#1][cancer#2][cancer#3][cancer#4])—using six predefined expression patterns: i) Downregulated in normal, unchanged in cancer: [-1][-1][0][0][0][0], ii) Downregulated in normal, upregulated in cancer: [-1][-1][1][1][1][1], iii) Downregulated in normal, upregulated or unchanged in cancer: [-1][-1][0/+1][0/+1][0/+1][0/+1], iv) Unchanged in normal, upregulated in cancer: [0][0][1][1][1][1], v) Upregulated in normal, downregulated in cancer: [1][1][-1][-1][-1][-1], vi) Upregulated in normal, unchanged in cancer: [1][1][0][0][0][0]. These patterns were used to stratify gene responses to OSKM and identify cancer-specific regulatory behaviors. We classified genes in Class I ([-1][-1][0][0][0][0]) as OSKM-unresponsive genes, and genes in Class II ([-1][-1][1][1][1][1]), III ([-1][-1][0/+1][0/+1][0/+1][0/+1]), and IV [0][0][1][1][1][1] as OSKM-aberrantly upregulated genes (Table S2F-H).

To rank genes within each predefined expression pattern category, we used a weighted Hamming distance metric[77] relative to the query vector. 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 values of the 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 (Table S2F-H).

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 normalized, scaled, batch-corrected, and filtered in accordance with best practices outlined in the Seurat v5 vignette[107]. Cells expressing gRNA were filtered for BFP. 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). 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[85]), human and mouse normal pancreas (n=4, GSE84133[86]), human early embryo before the three germ layers formed (D6-D14, n=555, GSE136447[108]), and iPSC derived from Parkinson's patients' fibroblasts (n=12, GSE183248[109]). "PDAC-specific signature genes" were curated from published scRNA-seq datasets by identifying transcripts enriched in PDAC cells but expressed in fewer than 20% of normal pancreatic cells (Table S4D)[85, 86]. Tumor suppressor genes (TSG) were curated from the OncoKB database[88] (Table S4E). 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.

Next-Generation Sequencing for Identification of Depleted gRNAs in Reprogrammed Cells (Table S5). Genomic DNA from each group was subjected to nested PCR amplification to generate NGS libraries, enabling identification of enriched or depleted sgRNAs in TRA-1-60^{+} reprogrammed cells compared to either reference controls or TRA-1-60^{-} cells. These analyses were performed on two independently passaged biological cultures, separated by 3–6 passages.

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[92]. The MAGeCK algorithm first performs normalization and then uses information borrowing across genes to estimate variance[92]. A negative binomial model is then fit and used to test for differential hits on lncRNAs across samples. Significance testing used 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[110]. 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 (Table S5J). gRNAs consistently enriched from two independent samples, as defined by either Z-scores > 2 and positive beta scores ($p < 0.05$), were retained (Table S5K).

Chromatin Isolation by RNA Purification followed by sequencing (ChIRP-seq). CAPANI cells and HPAFII were transfected with an *ATXN7L3-AS1* overexpression plasmid to generate cells with elevated nuclear expression of *ATXN7L3-AS1*. Cells (2×10^7) were washed three times with ice-cold PBS and crosslinked by adding 37% formaldehyde solution (1:36 dilution) for 10 min at room temperature (25 °C). Crosslinking was quenched with 2.5 M glycine (1:20 dilution) for 5 min. Crosslinked cells were scraped, collected by centrifugation at 1,000 × g for 3 min, washed twice with pre-cooled PBS, and pelleted by centrifugation. Cell pellets were resuspended in 1 mL cold lysis buffer per 2×10^7 cells and sonicated (30 s on/30 s off, five cycles) to obtain chromatin fragments. Lysates were clarified by centrifugation at 12,000 rpm for 10 min at 4 °C, and supernatants were stored at −80 °C until ChIRP.

Cell lysates were mixed with RNA–protein hybridization buffer at a 1:2 ratio and incubated overnight at room temperature with 5 μL of biotin-labeled probes tiling the target RNA. Streptavidin magnetic beads (50 μL) were washed three times with hybridization buffer and then

incubated with the sample–probe mixture for 4 h at room temperature. Beads were washed five times with 1 mL wash buffer and bound complexes were eluted with 50 μL elution buffer at 37 °C, 1,000 rpm for 5 min. Eluates were treated with proteinase K (2.5 μL) overnight at 65 °C, and DNA was purified using a DNA purification kit. Purified DNA was dissolved in 50 μL DEPC-treated water for downstream library preparation. ChIRP-enriched DNA was used to construct sequencing libraries with the VAHTS mRNA-seq V3 Library Prep Kit. Libraries were prepared by end repair, adaptor ligation, magnetic bead purification, and PCR amplification, followed by stringent quality control. The sequencing was performed on the Illumina platform.

ChIRP-seq data processing and analysis. Raw sequencing reads were quality-filtered to remove low-quality reads, adaptor sequences, and PCR duplicates. Clean reads were aligned to the human genome (hg38) using STAR2[111], allowing up to two mismatches. Only uniquely mapped reads were retained for downstream analyses. Peak calling was performed with MACS[112] using a p-value cutoff of 1e-5, and binding peaks were defined by clustering overlapping reads. Peak characteristics, including genomic distribution, length, spacing, and density, were analyzed. Annotation of *ATXN7L3-AS1* binding sites was performed with HOMER (annotatePeaks.pl) to assign peaks to genomic features (TSS, −1 kb to +100 bp; TTS, −100 bp to +1 kb; CDS exons; 5′ UTR exons; 3′ UTR exons; CpG islands; repeats; introns; and intergenic regions)[113]. Genes associated with significant peaks were subjected to functional enrichment analyses to identify biological pathways potentially regulated by *ATXN7L3-AS1*. Motif enrichment analysis was performed using HOMER.

Gene Ontology (GO) and Upstream Analysis. GO analysis was performed using Enrichment Analysis[114] or the ToppGene Suite[115]. To determine whether gene sets of interest are enriched for putative TF targets, we performed ChIP-X enrichment analysis using datasets from ENCODE[116] or ChEA[117], accessed through Enrichr[118] or X2K[119]. ChIP-X refers to genome-wide

TF binding data obtained from ChIP-ChIP, ChIP-Seq, ChIP-PET, or DamID experiments[117]. The p-value is computed using Fisher's exact test or the hypergeometric test. This is a binomial proportion test that assumes 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 Fisher's 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.

Identification of Disease-Associated Somatic Variants. Disease-associated somatic variant data for *ATXN7L3-AS1* were collected from the COSMIC[94] and ClinVar[120] databases via LncBook 2.0[121]. We retained variants labeled as “confirmed somatic mutations” in COSMIC. Variants were classified as pathogenic if they were annotated as having a significant impact on disease in ClinVar or if they had a FATHMM-MKL score greater than 0.7, a threshold previously shown to predict pathogenic potential of noncoding variants[122] (Table S6A-B).

Expression of lncRNAs in normal Tissues and TCGA pan-cancer cohorts. The expression profiles of lncRNAs in normal tissues were obtained from the GTEx database, and those in the TCGA Pan-Cancer dataset were obtained through The Atlas of Non-coding RNAs in Cancer (TANRIC)[123] (Table S7A-D).

ADDITIONAL RESOURCES. In vivo amplification and characterization of patient-derived xenograft (PDX) tumors have been described in STAR Protocols[59]. Detailed protocols for

transduction of PDAC cells and subtype characterization of PDX tumors were published in Heliyon[60].

Tissue immunohistochemistry (IHC), cell immunofluorescence (IF) staining, and qRT-PCR. Tissue IHC, IF, and qRT-PCR were performed as previously described in our published work[16].

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Human TRA-1-85/CD147 Antibody	R&D Systems	Cat # MAB3195
PE Mouse anti-Human TRA-1-85 Antigen	BD Biosciences	Cat # 563021
human Cytokeratin 19 (Abcam, BA-17)	Abcam	Cat # ab7755
Ki67	Cell Signaling Technology,	Cat # 9449
Cleaved Caspase-3	Cell Signaling Technology	Cat # 12692
FITC conjugate, Anti-TRA-1-60 Antibody, clone TRA-1-60, FITC conjugate	Millipore-Sigma	FCMAB115F
TRA-1-60 Alexa Fluor™ 488 Conjugate Kit for Live Cell Imaging	Thermo Scientific	Cat # A25618
Human/Mouse SSEA-1 Alexa Fluor« 594-conjugated Antibody	R&D systems	IC2155T-050UG
GloLIVE Human/Mouse SSEA-1 NL557 Live Cell Antibody	R&D Systems	Cat # NLLC2155R
Sendai virus HN Monoclonal Antibody (1A6), Alexa Fluor™ 647, eBioscience™	Invitrogen	Cat # 51-6494-82
OCT4 Antibody (OTI9B7) [Alexa Fluor« 405]	Novus Biologicals	NBP2-71225AF405
Anti-Oct4 antibody - ChIP Grade	Abcam	Cat# ab19857
Biological samples		
CF-1 MEFS 2 MILLION	Thermo Fisher Scientific	A34180
Dr4 Mouse Embryonic Fibroblasts, irradiated	Thermo Fisher Scientific	A34966
Chemicals, peptides, and recombinant proteins		
Iscript™ Advanced cDNA Synthesis Kit	Bio-Rad	1725037
RNase-Free DNase Set (50)	Qiagen	79254
Applied Biosystems TaqMan Fast Advanced Master Mix	Thermo Fisher Scientific	44-445-57
SPRIselect, Liquid	Beckman Coulter	B23318
Keratinocyte-SFM (1X)	Invitrogen	17005042
Primocin™	Thermo Fisher Scientific	ant-pm-05
Agilent RNA 6000 Pico Reagents	Agilent Technologies	5067-1514
ACCUMAX™ Cell Detachment Solution, Sterile, Liquid	Thermo Fisher Scientific	07921

InSolution™ Y-27632	Calbiochem	688001-500UG
Epidermal Growth Factor, human, animal component free,	Millipore-Sigma	E5036
Gibco FGF Basic (aa 10 155) Recombinant Human Protein	Invitrogen	PHG0021
Knockout™ Serum Replacement	Invitrogen	10828-028
L Glutamine (200mM)	Invitrogen	25030081
Eagle's Minimum Essential Medium (EMEM)	ATCC	30-2003
0.1% Gelatin in Water	STEMCELL Technologies	07903
Embryonic Stem Cell FBS, Qualified, One Shot™ Format, US Origin	Thermo Fisher Scientific	16141002
Human TruStain FcX™ (Fc Receptor Blocking Solution) 50 tests	Bio Legend	422301
Cholera Toxin from Vibrio cholerae	Thermo Fisher Scientific	C8052-.5MG
Blasticidin, Sterile, ≥95%, 10 mg/mL, Liquid	InvivoGen	ANTBL05
Puromycin, 100 mg (10 x 1 ml tubes)	InvivoGen	ANTPR1
DEFINED KERATINOCYTE-SFM, COMBO COMBINATION	Invitrogen	10744019
Protein Free, Thermo Scientific, Pierce™ Protein-free T20 (PBS) blocking buffer	Pierce	PI37573
Bovine Pituitary Extract (BPE)	Invitrogen	13028014
PowerTrack™ SYBR Green Master Mix	Applied Biosystems	A46109
Collagen I Rat Protein, Tail	Gibco	A1048301
FIX VIABILITY STAIN 660 100UG	Bd Biosciences	BDB564405
Cytotune™ EmGFP Sendai Fluorescence Reporter	Life Science	A16519
Critical commercial assays		
CytoTune™-iPS 2.0 Sendai Reprogramming Kit	*Invitrogen*	A16518
Human CRISPRi Common Cancer Cell Lines Kit	Millipore-Sigma	CRISPRIS10
Human CRISPR Control Kit	Millipore	CRISPRICON-1KT
QIAamp UCP DNA Micro Kit	Qiagen	56204
Lenti-X™ Provirus Quantitation Kit	Takara	631239
Deposited data		
Raw and analyzed data	This paper	GEO: GSE247243
Table S3A. List of gRNAs targeting cancer-associated lncRNAs (CRISPRIL17_hg38).	This paper, Supplementary Table	Table S3A.
Experimental models: Cell lines		
Human: BJ fibroblast cells	ATCC	CRL-2522
Human: H6c7 normal pancreatic ductal cells	KeraFast	ECA001-FP
Human: hTERT-HPNE,	ATCC	CRL-4023
Human: Passage 40 H1 and H9 ES cells	Wicell	WA01
Human: HPAFII PDAC cells	ATCC	CRL-1997
Human: Panc1 PDAC cells	ATCC	CRL-1469
Human: CFPAC1 cells	ATCC	CRL-1918
Human: PDAC cells from PDX (ST-00014490)	Star Protocols[59]	n.a.

Human: PDAC cells from PDX (ST-00007599)	Star Protocols[59]	n.a.
Experimental models: Organisms/strains		
Mouse: NOD-SCID-IL2Rgc null (NSG) (4-6 weeks old, female/male)	The Jackson Laboratory	JAX: #005557
Oligonucleotides		
Primer: ATXN7L3-AS1	IDT	Forward primer: TCTCCAAGAGCCCCAAAGTC Reverse primer: GCGTCCTGTCCTGATAACCT
Primer: AC079921.2 #2	IDT	Forward: AGAGCATGGGGTGAAAGACC Reverse: AGAGACTTTCTCAGGGTGGC
Primer: INTS4P2 #1:	IDT	Forward: AGCAGCAAAACAGACTTTACCC Reverse: TGTCCACAAAGCTGTCCTGC
PrimerINTS4P2 #2	IDT	Forward: AGATACGGACCCTGGAGCAT Reverse: GAACCACCCAAGAGCCATCA
RAB1A	Life Technologies	Hs00366313_m1
GAPDH	Life Technologies	Hs02758991_g1
GATA6	Life Technologies	Hs00232018_m1
SOX9	Life Technologies	Hs00165814_m1
KRT19	Life Technologies	Hs00761767_s1
Foxa2	Life Technologies	Hs05036278_s1
Hnf1b	Life Technologies	Hs01001602_m1
Onecut1	Life Technologies	Hs00413554_m1
FUT4	Life Technologies	Hs01106466_s1
MNX1	Life Technologies	Hs00907365_m1
MNX1	Life Technologies	Hs03454395_m1
PDX1	Life Technologies	Hs00426216_m1
Sendai-CMYC	Life Technologies	Mr04269876_mr
Sendai-OCT3-4	Life Technologies	Mr04269878_mr
Sendai-KLF4	Life Technologies	Mr04421256_mr
SEV-SOX2	Life Technologies	Mr04269881_mr
Sendai	Life Technologies	Mr04269880_mr
Recombinant DNA		

Software and algorithms		
DESEQ2	Love et al.,[76]	https://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html
MAGeCK	Li et al., 2014[92] and Wang e al., 2019[110]	https://github.com/liulab-dfci/MAGeCK
Other		

REFERENCE

1. Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035-1041 (2010).
2. Laiosa, C.V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L. & Graf, T. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. *Immunity* **25**, 731-744 (2006).
3. Soufi, A., Donahue, G. & Zaret, K.S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994-1004 (2012).
4. Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663-676 (2004).
5. Davis, R.L., Weintraub, H. & Lassar, A.B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987-1000 (1987).
6. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872 (2007).
7. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).
8. Lu, Y. *et al.* Reprogramming to recover youthful epigenetic information and restore vision. *Nature* **588**, 124-129 (2020).
9. Ocampo, A. *et al.* In Vivo Amelioration of Age-Associated Hallmarks by Partial Reprogramming. *Cell* **167**, 1719-1733 e1712 (2016).
10. Stricker, S.H. *et al.* Widespread resetting of DNA methylation in glioblastoma-initiating cells suppresses malignant cellular behavior in a lineage-dependent manner. *Genes Dev* **27**, 654-669 (2013).
11. Kotini, A.G. *et al.* Stage-Specific Human Induced Pluripotent Stem Cells Map the Progression of Myeloid Transformation to Transplantable Leukemia. *Cell Stem Cell* **20**, 315-328 e317 (2017).
12. Chao, M.P. *et al.* Human AML-iPSCs Reacquire Leukemic Properties after Differentiation and Model Clonal Variation of Disease. *Cell Stem Cell* **20**, 329-344 e327 (2017).
13. Zhang, X., Cruz, F.D., Terry, M., Remotti, F. & Matushansky, I. Terminal differentiation and loss of tumorigenicity of human cancers via pluripotency-based reprogramming. *Oncogene* **32**, 2249-2260, 2260 e2241-2221 (2013).
14. Miyoshi, N. *et al.* Defined factors induce reprogramming of gastrointestinal cancer cells. *Proc Natl Acad Sci U S A* **107**, 40-45 (2010).
15. Nagai, K. *et al.* Long-term culture following ES-like gene-induced reprogramming elicits an aggressive phenotype in mutated cholangiocellular carcinoma cells. *Biochem Biophys Res Commun* **395**, 258-263 (2010).
16. Kim, J. *et al.* An iPSC Line from Human Pancreatic Ductal Adenocarcinoma Undergoes Early to Invasive Stages of Pancreatic Cancer Progression. *Cell Reports* **3**, 2088-2099 (2013).
17. Knappe, N. *et al.* Directed Dedifferentiation Using Partial Reprogramming Induces Invasive Phenotype in Melanoma Cells. *Stem Cells* **34**, 832-846 (2016).
18. Papapetrou, E.P. Patient-derived induced pluripotent stem cells in cancer research and precision oncology. *Nat Med* **22**, 1392-1401 (2016).
19. Kim, J. & Zaret, K.S. Reprogramming of human cancer cells to pluripotency for models of cancer progression. *EMBO Journal* **34**, 739-747 (2015).
20. Shendy, N.A.M., Zimmerman, M.W., Abraham, B.J. & Durbin, A.D. Intrinsic transcriptional heterogeneity in neuroblastoma guides mechanistic and therapeutic insights. *Cell Rep Med* **3**, 100632 (2022).

21. Sanghvi, N. *et al.* Charting the transcriptomic landscape of primary and metastatic cancers in relation to their origin and target normal tissues. *Sci Adv* **10**, eadn0220 (2024).
22. Quintana, E. *et al.* Phenotypic heterogeneity among tumorigenic melanoma cells from patients that is reversible and not hierarchically organized. *Cancer Cell* **18**, 510-523 (2010).
23. Flavahan, W.A., Gaskell, E. & Bernstein, B.E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357** (2017).
24. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849 e821 (2019).
25. Elia, I. *et al.* Breast cancer cells rely on environmental pyruvate to shape the metastatic niche. *Nature* **568**, 117-121 (2019).
26. Halbrook, C.J., Lyssiotis, C.A., Pasca di Magliano, M. & Maitra, A. Pancreatic cancer: Advances and challenges. *Cell* **186**, 1729-1754 (2023).
27. Moffitt, R.A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet* **47**, 1168-1178 (2015).
28. Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47-52 (2016).
29. Chan-Seng-Yue, M. *et al.* Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. *Nat Genet* **52**, 231-240 (2020).
30. Raghavan, S. *et al.* Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. *Cell* **184**, 6119-6137 e6126 (2021).
31. Burdziak, C. *et al.* Epigenetic plasticity cooperates with cell-cell interactions to direct pancreatic tumorigenesis. *Science* **380**, eadd5327 (2023).
32. Baba, T. *et al.* A Lesson in Transcriptional Plasticity: Classical Identity Is Silenced, but Not Lost, in Pancreatic Ductal Adenocarcinoma Cell Lines. *Gastroenterology* **163**, 1450-1453 e1453 (2022).
33. Singhal, A. *et al.* A Classical Epithelial State Drives Acute Resistance to KRAS Inhibition in Pancreatic Cancer. *Cancer Discov* **14**, 2122-2134 (2024).
34. Davidson, E.H. & Erwin, D.H. Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796-800 (2006).
35. Shukla, V., Cetnarowska, A., Hyldahl, M. & Mandrup, S. Interplay between regulatory elements and chromatin topology in cellular lineage determination. *Trends Genet* **38**, 1048-1061 (2022).
36. Bulyk, M.L., Drouin, J., Harrison, M.M., Taipale, J. & Zaret, K.S. Pioneer factors - key regulators of chromatin and gene expression. *Nat Rev Genet* (2023).
37. Boyer, L.A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-956 (2005).
38. Cabili, M.N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**, 1915-1927 (2011).
39. Lee, J.T. Disruption of imprinted X inactivation by parent-of-origin effects at Tsix. *Cell* **103**, 17-27 (2000).
40. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22**, 1775-1789 (2012).
41. Mattick, J.S. *et al.* Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol* **24**, 430-447 (2023).
42. Khalil, A.M. *et al.* 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 (2009).
43. Guttman, M. & Rinn, J.L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339-346 (2012).

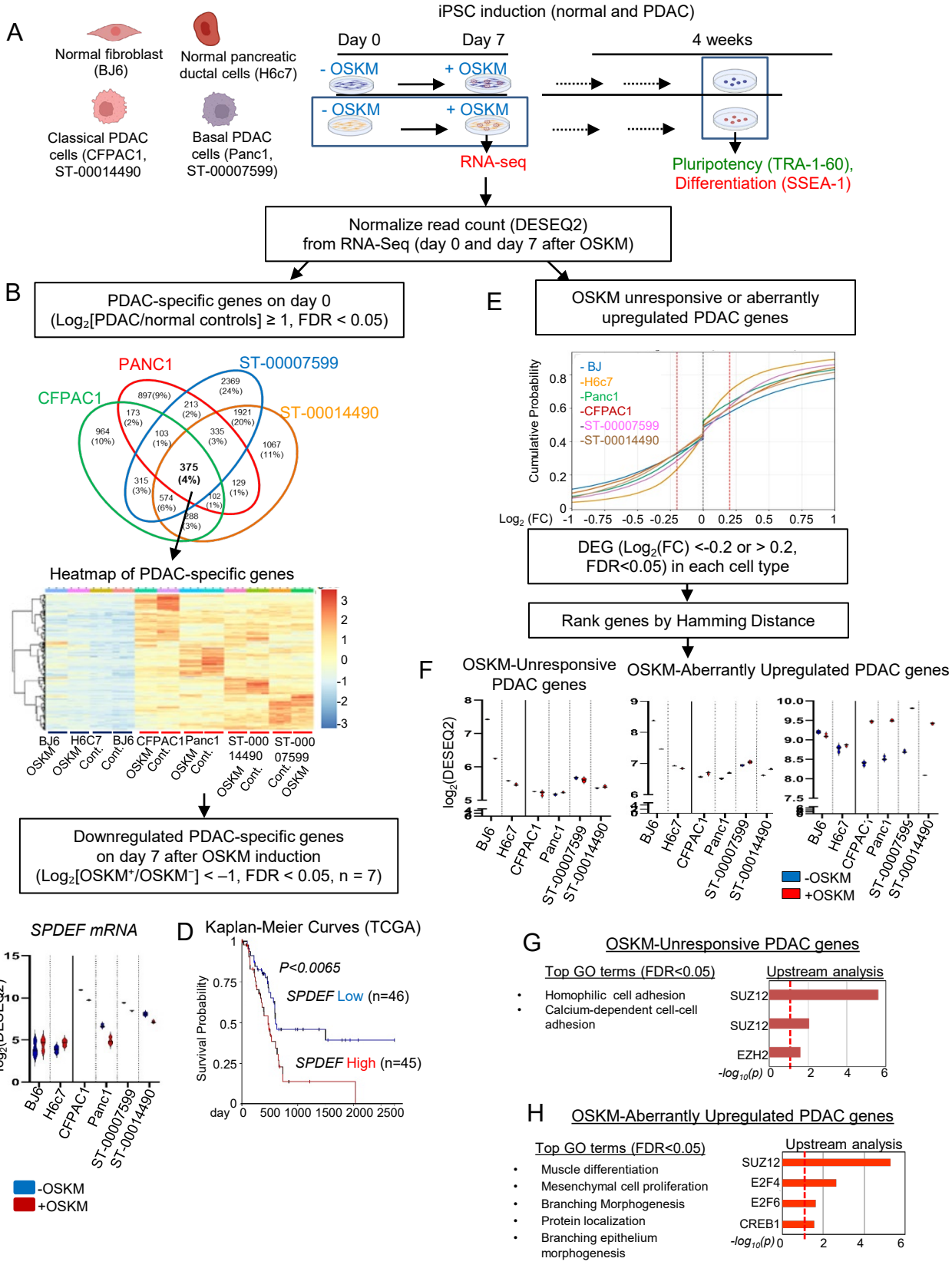
44. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409-419 (2010).
45. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516-520 (2013).
46. Kim, Y.J., Xie, P., Cao, L., Zhang, M.Q. & Kim, T.H. Global transcriptional activity dynamics reveal functional enhancer RNAs. *Genome Res* **28**, 1799-1811 (2018).
47. Jiao, W. *et al.* HPSE enhancer RNA promotes cancer progression through driving chromatin looping and regulating hnRNPU/p300/EGR1/HPSE axis. *Oncogene* **37**, 2728-2745 (2018).
48. Statello, L., Guo, C.J., Chen, L.L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* **22**, 96-118 (2021).
49. Ng, S.Y., Johnson, R. & Stanton, L.W. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J* **31**, 522-533 (2012).
50. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295-300 (2011).
51. Wu, D. *et al.* Dual genome-wide coding and lncRNA screens in neural induction of induced pluripotent stem cells. *Cell Genom* **2** (2022).
52. Lin, N. *et al.* An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol Cell* **53**, 1005-1019 (2014).
53. Wang, Y. *et al.* Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Dev Cell* **25**, 69-80 (2013).
54. Loewer, S. *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet* **42**, 1113-1117 (2010).
55. Wang, Z.W. *et al.* SRSF3-mediated regulation of N6-methyladenosine modification-related lncRNA ANRIL splicing promotes resistance of pancreatic cancer to gemcitabine. *Cell Rep* **39**, 110813 (2022).
56. Ye, M., Dong, S., Hou, H., Zhang, T. & Shen, M. 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 (2021).
57. Jiang, M.C., Ni, J.J., Cui, W.Y., Wang, B.Y. & Zhuo, W. Emerging roles of lncRNA in cancer and therapeutic opportunities. *Am J Cancer Res* **9**, 1354-1366 (2019).
58. Arunkumar, G., Baek, S., Sturgill, D., Bui, M. & Dalal, Y. Oncogenic lncRNAs alter epigenetic memory at a fragile chromosomal site in human cancer cells. *Sci Adv* **8**, eabl5621 (2022).
59. Tessier, J. *et al.* Protocol to purify and culture human pancreatic cancer cells from patient-derived xenografts. *STAR Protoc* **6**, 103672 (2025).
60. Grygoryev, D. *et al.* Sendai virus is robust and consistent in delivering genes into human pancreatic cancer cells. *Heliyon* **10**, e27221 (2024).
61. Ban, H. *et al.* Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors. *Proc Natl Acad Sci U S A* **108**, 14234-14239 (2011).
62. Draper, J.S., Pigott, C., Thomson, J.A. & Andrews, P.W. Surface antigens of human embryonic stem cells: changes upon differentiation in culture. *J Anat* **200**, 249-258 (2002).
63. Schopperle, W.M. & DeWolf, W.C. The TRA-1-60 and TRA-1-81 human pluripotent stem cell markers are expressed on podocalyxin in embryonal carcinoma. *Stem Cells* **25**, 723-730 (2007).
64. Noisa, P. *et al.* Identification and characterisation of the early differentiating cells in neural differentiation of human embryonic stem cells. *PLoS One* **7**, e37129 (2012).
65. Fenderson, B.A., Andrews, P.W., Nudelman, E., Clausen, H. & Hakomori, S. Glycolipid core structure switching from globo- to lacto- and ganglio-series during retinoic acid-induced differentiation of TERA-2-derived human embryonal carcinoma cells. *Dev Biol* **122**, 21-34 (1987).

66. Alshaikh, A. *et al.* Reprogramming Pancreatic Ductal Adenocarcinoma to Pluripotency. *J Vis Exp* (2024).
67. Polo, J.M. *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol* **28**, 848-855 (2010).
68. Cacchiarelli, D. *et al.* Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. *Cell* **162**, 412-424 (2015).
69. Takahashi, K. *et al.* Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. *Nat Commun* **5**, 3678 (2014).
70. Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209-1222 (2012).
71. Knaupp, A.S. *et al.* Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. *Cell Stem Cell* **21**, 834-845 e836 (2017).
72. Chronis, C. *et al.* Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* **168**, 442-459 e420 (2017).
73. Li, D. *et al.* Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell Stem Cell* **21**, 819-833 e816 (2017).
74. Oettgen, P. *et al.* PDEF, a novel prostate epithelium-specific ets transcription factor, interacts with the androgen receptor and activates prostate-specific antigen gene expression. *J Biol Chem* **275**, 1216-1225 (2000).
75. Tonelli, C. *et al.* A mucus production programme promotes classical pancreatic ductal adenocarcinoma. *Gut* **73**, 941-954 (2024).
76. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
77. Hamming, R.W. Error detecting and error correcting codes. *The Bell System Technical Journal* **29**, 147-160 (1950).
78. Zaratiegui, M., Irvine, D.V. & Martienssen, R.A. Noncoding RNAs and gene silencing. *Cell* **128**, 763-776 (2007).
79. Petracovici, A. & Bonasio, R. Distinct PRC2 subunits regulate maintenance and establishment of Polycomb repression during differentiation. *Mol Cell* **81**, 2625-2639 e2625 (2021).
80. Zhao, J. *et al.* Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* **40**, 939-953 (2010).
81. Yang, F. *et al.* The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol* **16**, 52 (2015).
82. Aguilar, R., Rosenberg, M., Levy, V. & Lee, J.T. An evolving landscape of PRC2-RNA interactions in chromatin regulation. *Nat Rev Mol Cell Biol* **26**, 631-642 (2025).
83. Liu, S.J. *et al.* CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* **355** (2017).
84. Horlbeck, M.A. *et al.* Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* **5** (2016).
85. Chen, J., Liu, Z., Wu, Z., Li, W. & Tan, X. Identification of a chemoresistance-related prognostic gene signature by comprehensive analysis and experimental validation in pancreatic cancer. *Front Oncol* **13**, 1132424 (2023).
86. Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst* **3**, 346-360 e344 (2016).
87. Suehnholz, S.P. *et al.* Quantifying the Expanding Landscape of Clinical Actionability for Patients with Cancer. *Cancer Discov* **14**, 49-65 (2024).
88. Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* **2017** (2017).

89. Shultz, L.D. *et al.* Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. *J Immunol* **174**, 6477-6489 (2005).
90. Gossler, A., Joyner, A.L., Rossant, J. & Skarnes, W.C. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463-465 (1989).
91. Turner, D.A., Trott, J., Hayward, P., Rue, P. & Martinez Arias, A. An interplay between extracellular signalling and the dynamics of the exit from pluripotency drives cell fate decisions in mouse ES cells. *Biol Open* **3**, 614-626 (2014).
92. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15**, 554 (2014).
93. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585 (2013).
94. Sondka, Z. *et al.* COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res* **52**, D1210-D1217 (2024).
95. Chu, C., Qu, K., Zhong, Franklin L., Artandi, Steven E. & Chang, Howard Y. Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Molecular Cell* **44**, 667-678 (2011).
96. Diaferia, G.R. *et al.* Dissection of transcriptional and cis-regulatory control of differentiation in human pancreatic cancer. *EMBO J* **35**, 595-617 (2016).
97. Murtagh, F. & Legendre, P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification* **31**, 274-295 (2014).
98. Ascic, E. *et al.* In vivo dendritic cell reprogramming for cancer immunotherapy. *Science* **386**, eadn9083 (2024).
99. Linde, M.H. *et al.* Reprogramming Cancer into Antigen Presenting Cells as a Novel Immunotherapy. *Cancer Discov* (2023).
100. Zimmermannova, O. *et al.* Restoring tumor immunogenicity with dendritic cell reprogramming. *Sci Immunol* **8**, eadd4817 (2023).
101. Habbe, N. *et al.* Spontaneous induction of murine pancreatic intraepithelial neoplasia (mPanIN) by acinar cell targeting of oncogenic Kras in adult mice. *Proc Natl Acad Sci U S A* **105**, 18913-18918 (2008).
102. Krah, N.M. *et al.* Prevention and Reversion of Pancreatic Tumorigenesis through a Differentiation-Based Mechanism. *Dev Cell* **50**, 744-754 e744 (2019).
103. Krah, N.M. *et al.* The acinar differentiation determinant PTF1A inhibits initiation of pancreatic ductal adenocarcinoma. *Elife* **4** (2015).
104. Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. & Hasegawa, M. 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 (2009).
105. Kim, J. *et al.* Longitudinal analysis of human pancreatic adenocarcinoma development reveals transient gene expression signatures. *Mol Cancer Res* (2021).
106. Kim, D., Paggi, J.M., Park, C., Bennett, C. & Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**, 907-915 (2019).
107. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology* **42**, 293-304 (2024).
108. Xiang, L. *et al.* A developmental landscape of 3D-cultured human pre-gastrulation embryos. *Nature* **577**, 537-542 (2020).
109. Novak, G. *et al.* Single-cell transcriptomics of human iPSC differentiation dynamics reveal a core molecular network of Parkinson's disease. *Commun Biol* **5**, 49 (2022).
110. Wang, B. *et al.* Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat Protoc* **14**, 756-780 (2019).

111. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
112. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
113. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).
114. Thomas, P.D. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci* **31**, 8-22 (2022).
115. Chen, J., Bardes, E.E., Aronow, B.J. & Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* **37**, W305-311 (2009).
116. Consortium, E.P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636-640 (2004).
117. Lachmann, A. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438-2444 (2010).
118. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* **1**, e90 (2021).
119. Clarke, D.J.B. *et al.* eXpression2Kinases (X2K) Web: linking expression signatures to upstream cell signaling networks. *Nucleic Acids Res* **46**, W171-W179 (2018).
120. Landrum, M.J. *et al.* ClinVar: updates to support classifications of both germline and somatic variants. *Nucleic Acids Res* **53**, D1313-D1321 (2025).
121. Ma, L. *et al.* LncBook: a curated knowledgebase of human long non-coding RNAs. *Nucleic Acids Res* **47**, D128-D134 (2019).
122. Shihab, H.A. *et al.* An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* **31**, 1536-1543 (2015).
123. Li, J. *et al.* TANRIC: An Interactive Open Platform to Explore the Function of lncRNAs in Cancer. *Cancer research* **75**, 3728-3737 (2015).

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

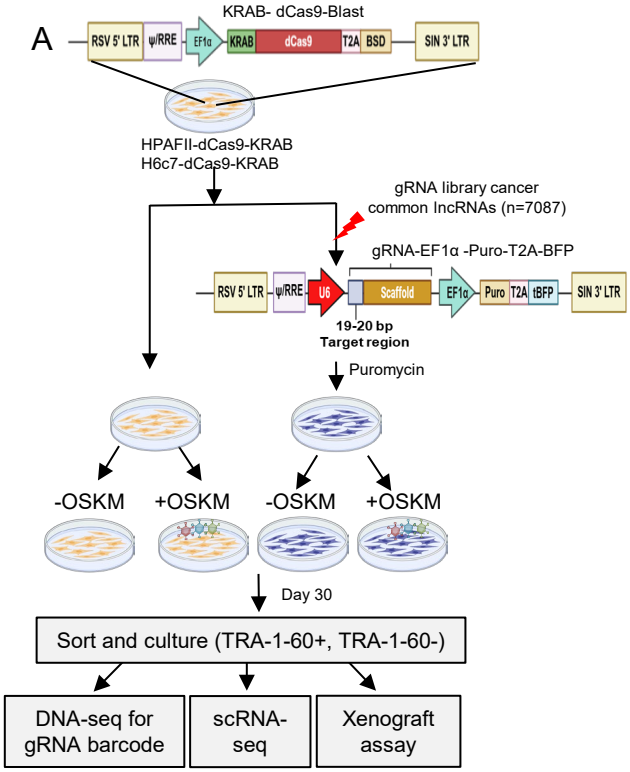


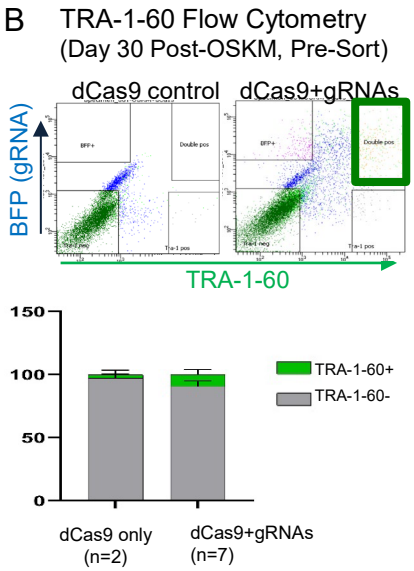
MAIN FIGURE TITLES and LEGENDS

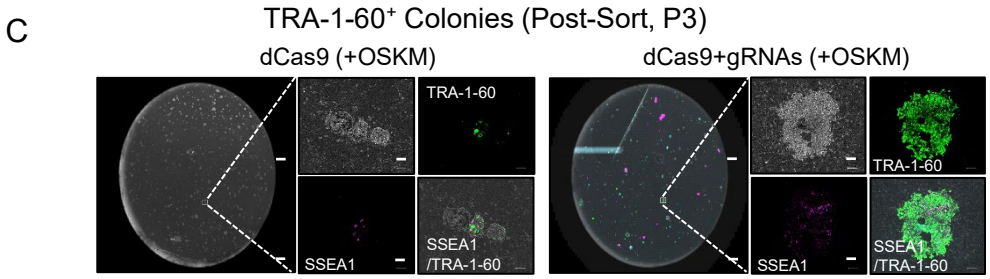
Figure 1. OSKM Reprogramming Partially Suppresses PDAC-Unique Genes, Leaving Others Unresponsive or Aberrantly Induced. A. Schematic diagram of reprogramming of normal controls (BJ fibroblast and H6c7 hnPDEC) and human classical (CFPAC1, ST-00014490) and basal (Panc1, ST-00007599) subtypes of PDAC cells 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 ($\log_2$ [PDAC/normal controls] ≥ 1, FDR < 0.05). The heatmap displays the expression profiles of 375 PDAC-specific genes consistently upregulated across all PDAC samples, with Z-scores indicated by the color key. Seven of the 375 PDAC-unique genes were significantly downregulated by OSKM across all PDAC cells ($\log_2$[OSKM$^+$/OSKM$^-$] < –1, FDR < 0.05). C-D. Example of an OSKM-induced downregulated PDAC-specific gene. (C) *SPDEF* mRNA levels in the indicated cells without and with OSKM induction. (D) 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. 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 (FC) values across samples, before and after OSKM induction (day 7 vs. day 0). 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 as an empirical measure of deviation from the null hypothesis (Fig. S3C; see Methods). Genes were then queried against pre-defined patterns of expression changes across the samples. A weighted Hamming distance was used to measure how well a gene's pattern of expression changes across samples matched with each query (see Methods). For each query, genes with weighted Hamming distances no greater than 0.35 were selected as hits for downstream analysis. F. Violin plots showing expression levels for OSKM-unresponsive PDAC genes (left: distance < 0.375, n = 1,466; Table S2F) and OSKM-aberrantly upregulated genes (middle: distance < 0.25, n = 1,815; right: distance = 0, n = 82; Tables S2G–S2H; see Methods). G-H. Gene Ontology (GO)[102] and X2K upstream regulator[103, 104] analyses for genes that were either unresponsive to OSKM in PDAC (G, Table S2F) or aberrantly upregulated following OSKM induction (H, Table S2G). The red dashed line indicates the significance threshold (FDR or p-value < 0.05).

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







	Total colonies #	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+gRNAs	2665	23 (1%)	0	0	0	23 (1%)
dCas9+gRNAs+OSKM	212	47 (22.2%)	96 (45.3%)	32 (15.1%)	64 (30.2%)	15 (7.1%)

Figure 2. CRISPRi of Cancer-Associated lncRNAs Enhances Reprogramming Efficiency in PDAC Cells. A. Schematic diagram of reprogramming HPAFII PDAC and H6c7 hnPDEC 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. Flow cytometry–based quantification of TRA-1-60^{+} and TRA-1-60^{-} fractions in HPAFII-dCas9 control cells and HPAFII-dCas9 cells expressing lncRNA-targeting gRNAs (BFP^{+}) at day 30 after OSKM induction. C. Representative images and quantification summary of TRA-1-60^{+} and SSEA1^{+} cells on plates after multiple passages (Scale bar: 284 μm).

Grygoryev & Lee/Kim et al.,2025

Figure 3

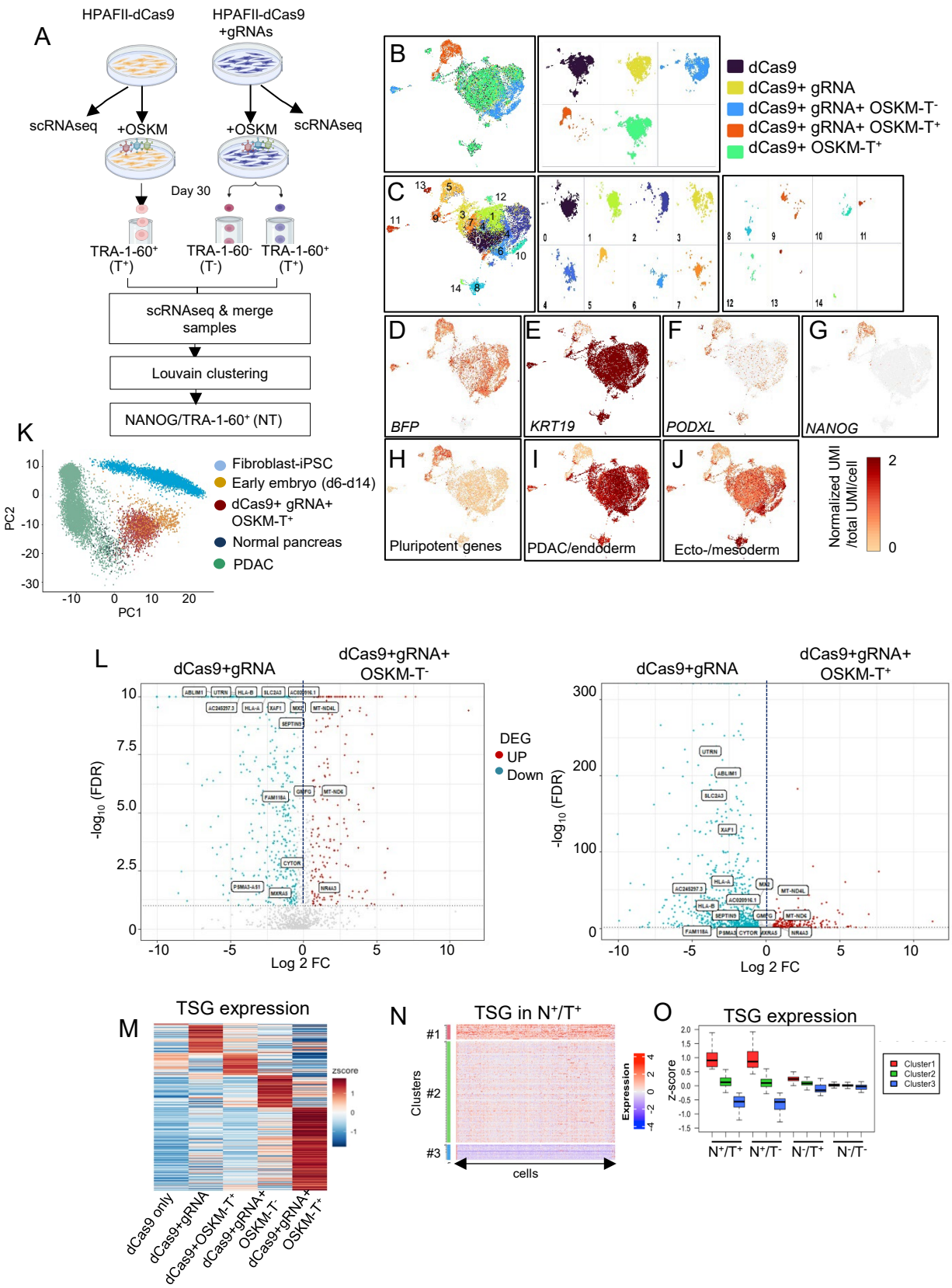


Figure 3. Repression of PDAC Programs and De-repression of Pluripotency and Tumor Suppressor Signatures in PDAC Cells Reprogrammed with OSKM and lncRNA Depletion. A. Schematic diagram of the scRNA-seq workflow. B-J. UMAP visualizations showing distinct clusters by sample (B), 15 Louvain-defined clusters (C), cells expressing BFP (D), pancreatic ductal marker KRT19 (E), *PODXL* encoding TRA-1-60 (F), *NANOG* (G), pluripotency-associated genes (H, Table S4A), endodermal and PDAC-related genes (I, Table S4B), and ectodermal/mesodermal lineage markers (J, Table S4C). Each dot represents a cell, and the color intensity indicates the normalized UMI counts (scaled by the total UMI per cell). K. PCA of HPAFII-dCas9+gRNA+OSKM-T^{+} cells alongside publicly available scRNA-seq datasets from human PDAC (n=6, GSE212966[83]), human and mouse normal pancreas (n=4, GSE84133[84]), human early embryo before the three germ layers formed (D6-D14, n=555, GSE136447[105]), and iPSC derived from Parkinson's disease patients' fibroblast (n=12, GSE183248[106]). L. Volcano plots of DEGs between dCas9+gRNA and either dCas9+gRNA+OSKM-T^{-} (left) or dCas9+gRNA+OSKM-T^{+} (right). PDAC-specific signature genes are annotated (n=332, Table S4D). M. Heatmap showing expression of 318 TSGs across the indicated cell populations (Table S4E). 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. N-O. Heatmap (N) and quantification (O) of TSG expression in the dCas9+gRNA+OSKM-T^{+} cells. (N) 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. TSGs were grouped into three distinct clusters based on their expression profiles using unsupervised k-means clustering. (O) Expression levels of TSGs in each cluster are shown across the indicated subpopulations based on 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^{+}). The Y-axis indicates Z-score.

Grygoryev & Lee/Kim et al.,2025

Figure 4

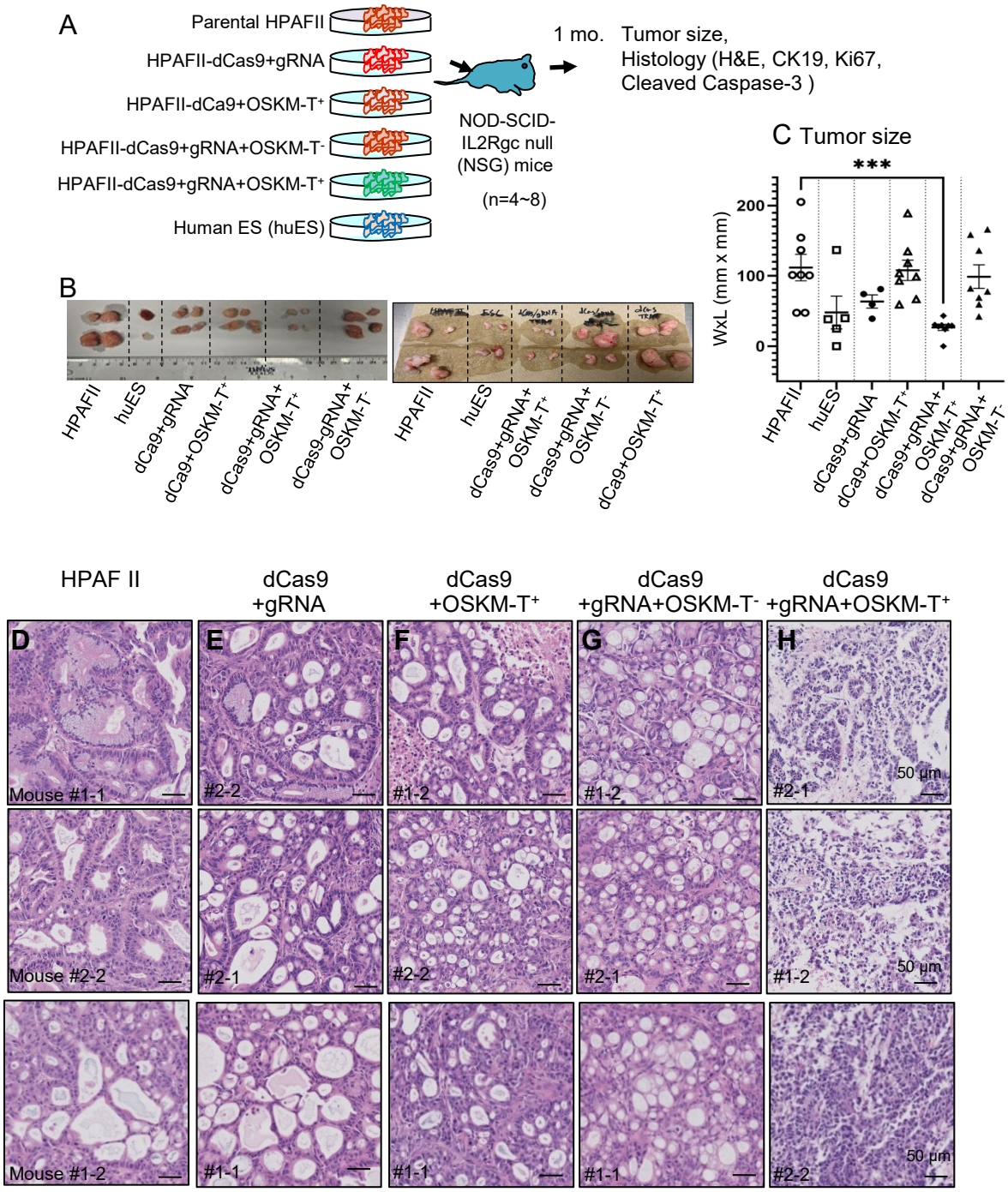


Figure 4. Xenograft Tumor Growth Is Reduced in PDAC Cells Reprogrammed with OSKM and lncRNA Depletion. A. Schematic diagram of xenograft assay of TRA-1-60^{+} and TRA-1-60^{-}cells derived from OSKM-reprogrammed HPAFII-dCas9/gRNA, along with controls. Tumors were harvested from NSG mice one month after being subcutaneously injected with dCas9+gRNA+OSKM-T^{-} and dCas9+gRNA+OSKM-T^{+}, along with controls: Parental HPAFII, huES, dCas9+gRNA, dCas9+OSKM-T^{+} (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-T^{+} 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-T^{+}, G; dCas9+gRNA+OSKM-T^{-}, H; dCas9+gRNA+OSKM-T^{+}) 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

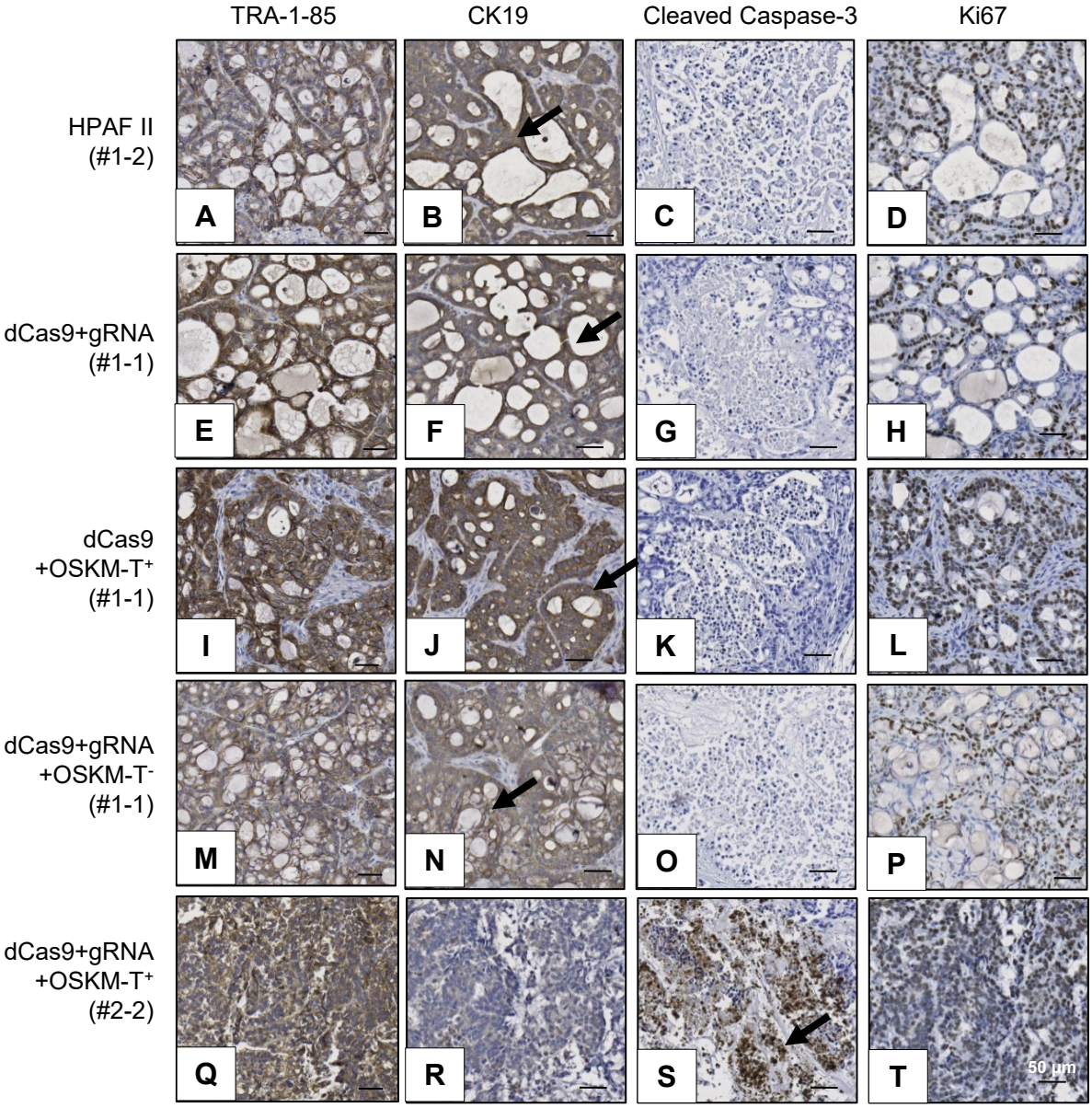


Figure 5. Xenograft Tumors from OSKM/lncRNA-Depleted PDAC Cells Lose Ductal Identity and Express Apoptotic Markers. Representative images of TRA-1-85 human specific marker, human CK19 pancreatic ductal marker, Cleaved Caspase-3 apoptotic marker, and Ki67 proliferation marker staining in xenograft tumors derived from parental HPAFII cells (A–D), dCas9+gRNA cells (E–H), dCas9+OSKM-T^{+} cells (I–L), dCas9+gRNA+OSKM-T^{-} cells (M–P), and dCas9+gRNA+OSKM-T^{+} 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

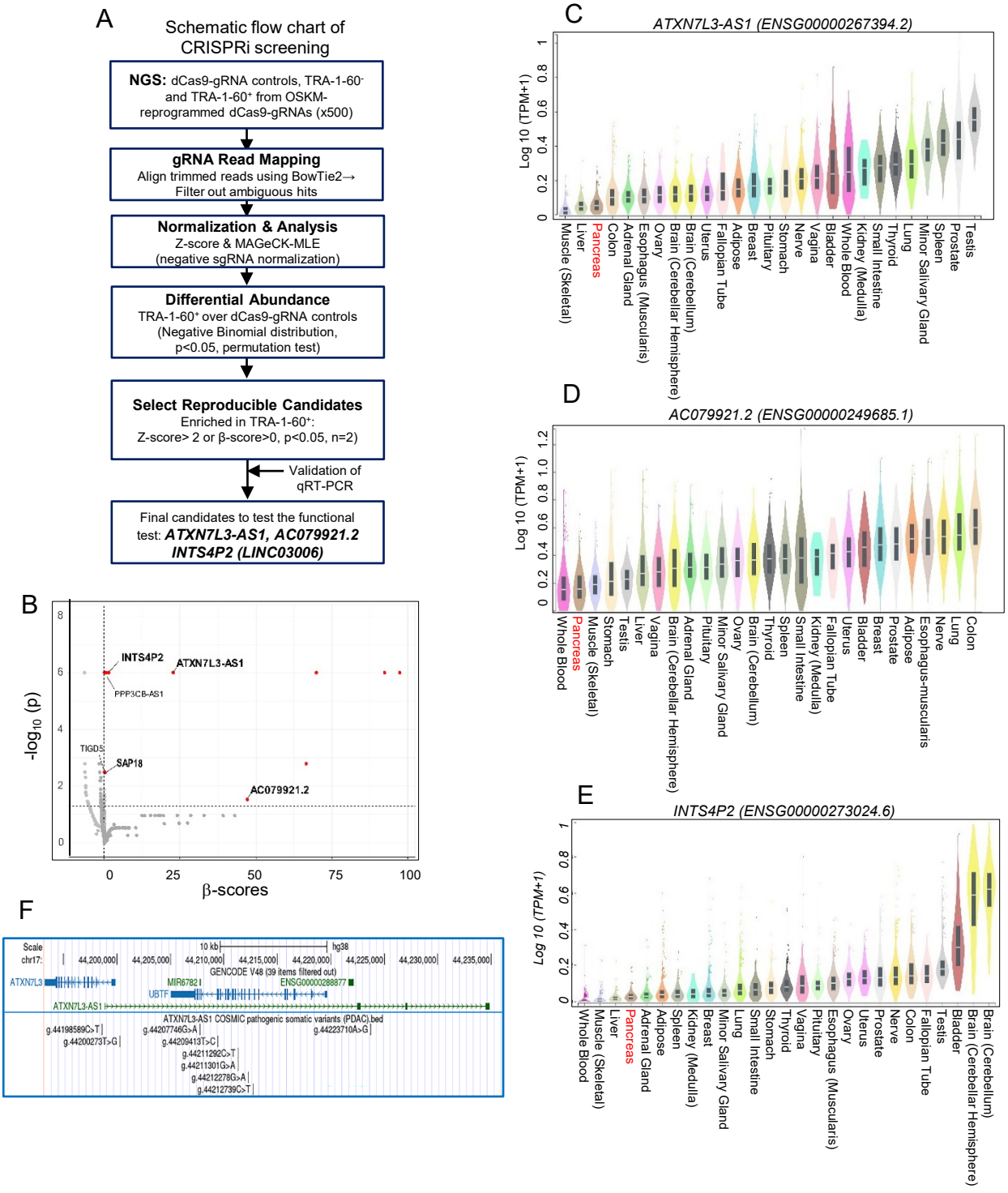
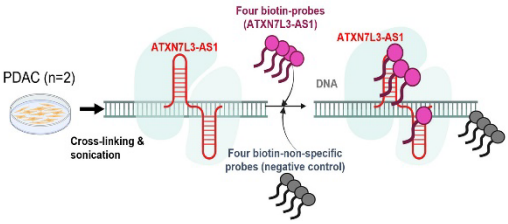


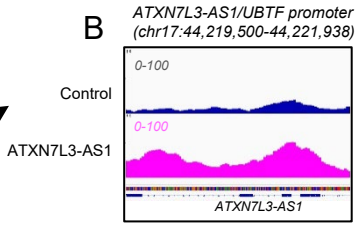
Figure 6. Identification of *ATXN7L3-AS1, AC079921.2, and INTS4P2* as Candidate Barriers to Reprogramming in OSKM/lncRNA-Depleted PDAC Cells. A. Schematic flow chart of CRISPRi screening. B. Significantly depleted lncRNAs in HPAFII-dCas9+gRNA+OSKM-T$^+$ cells in comparison to the HPAFII-dCas9-gRNA reference sample are visualized in the volcano plot. The enrichment score of gRNAs is shown as the "β-scores." Positive and negative β-scores indicate positive and negative selections of gRNAs 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 was calculated by ranking lncRNAs 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 were 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-E. Expression level of candidate lncRNAs in human normal tissues (GTEx)[91]. Expression values are in TPM (Transcripts per Million), calculated from a gene model with isoforms collapsed to a single gene. The X-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) are highlighted in red. F. UCSC genome browser view of the *ATXN7L3-AS1* locus (chr17:44,198,841–44,234,868; hg38), located between the *ATXN7L3* and *UBTF* genes. Pathogenic somatic variants identified in PDAC from the COSMIC database are displayed along the locus.

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

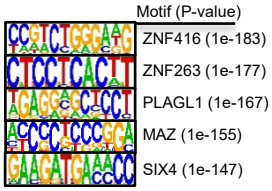
A



B

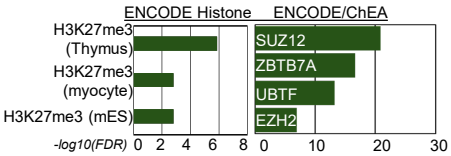


C All ATXN7L3-AS1 peaks in PDAC (p<1e-5, n=5622)

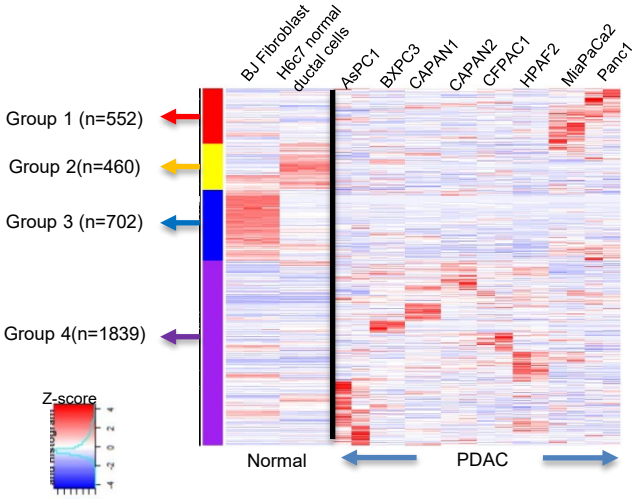


D Genes with ATXN7L3-AS1 binding sites in PDAC (p<1e-5, n=4626)

Top GO biological process (FDR)
- Actin cytoskeleton organization (2.2E-06)
- Neuron migration (1.9E-06)
- Axon guidance (2.10E-11)
- RNA pol II transcriptional regulation (1.2E-12)



E Ward's minimum variance hierarchical clustering of genes with ATXN7L3-AS1 (n=3553)



Go Terms and ENCODE/ChEA Upstream (FDR)

Group 1 (Basal subtype-associated, n=552, 16%)
- Synaptic signaling (1.10E-04)
- Neuronal development (5.01E-04)
- Cell-Cell signaling (6.97E-03)
- Upstream: EZH2 (4.5E-03), SUZ12 (4.5E-03), REST (4.3E-03)

Group 2 (Normal ductal cell-associated, n=460, 13%)
- RNA biosynthesis (1.62E-04)
- RNA pol II transcription (2.08E-03)
- Upstream: TP63 (2.23E-04), ZBTB7A (2.23E-04), E2F6 (2.23E-04), UBTF (2.36E-04), POU5F1 (2.56E-03)

Group 3 (Fibroblast-associated, n=702, 20%)
- Tube development (1.09E-16)
- Generation of neurons (1.95E-16)
- Neuron differentiation (5.73E-16)
- Circulatory system development (5.73E-16),
- Upstream: SUZ12 (6.95E-15), UBTF (2.47E-06), AR (3.40E-06), NFE2L2 (3.65E-06), ZBTB7A (7.48E-06)

Group 4 (Classical subtype-associated, n=1839, 52%)
- Tyrosine kinase signaling pathway (2.44E-04)
- Cellular response to endogenous stimulus (4.56E-04)
- Small GTPase-mediated signal transduction (4.64E-04)
- Epithelium development (1.64E-03)
- Morphogenesis (2.06E-03)
- RNA metabolic process (2.44E-04)
- RNA pol II transcription (4.56E-04)
- Upstream: ZBTB7A (8.11E-14), UBTF (5.86E-11), RCOR1 (3.85E-06), RUNX1 (5.63E-06), HNF4A (4.42E-05), SMC3 (4.94E-05), E2F6 (2.35E-04), GATA2 (3.22E-04), TCF3 (3.40E-04), CTCF (3.40E-04), SMAD4 (3.59E-04), BCL3 (6.82E-04), SUZ12 (7.25E-04)

Figure 7. *ATXN7L3-AS1* Binds Two Categories of Genes in PDAC: Developmental Programs Such as Cell Migration, Which It Preserves in a Poised State, and Cancer-Associated Genes, Which It Maintains in an Active State. A-C. Mapping *ATXN7L3-AS1* binding sites using ChIRP-seq. (A) Schematic of ChIRP-seq strategy for mapping *ATXN7L3-AS1* genomic binding sites in PDAC cells (HPAFII and CAPAN1, n=2 each cell). *ATXN7L3-AS1*–RNA/DNA complexes were cross-linked, captured with biotinylated antisense probes, and sequenced; non-specific probes served as controls. (B) Representative genome browser tracks of *ATXN7L3-AS1* ChIRP-seq peaks. *ATXN7L3-AS1* binding is enriched at the *UBTF* promoter, which lies within the *ATXN7L3-AS1* genomic locus. (C) HOMER motif analysis of the sequence of *ATXN7L3-AS1* peaks. De novo motif discovery was performed with HOMER on all 5,622 *ATXN7L3-AS1* ChIRP-seq peaks ($p<1e-5$). Significantly enriched motifs are shown, ranked by $-\log_{10}(p)$. D-E. Genes with *ATXN7L3-AS1* Peaks. (D) Functional enrichment of genes with *ATXN7L3-AS1* binding peaks ($p < 1e-5$), including representative GO terms, ENCODE histone mark annotations, and ENCODE/ChEA-predicted regulators. (E) Gene signatures associated with *ATXN7L3-AS1*–bound genes. Ward minimum variance hierarchical clustering of genes with *ATXN7L3-AS1* across normal and PDAC cell lines. The heatmap indicates Z-score expression. Representative Go categories and ENCODE/ChEA Consensus upstream analysis of genes in each group (FDR<0.05).

Grygoryev & Lee/Kim et al.,2025
Figure 8

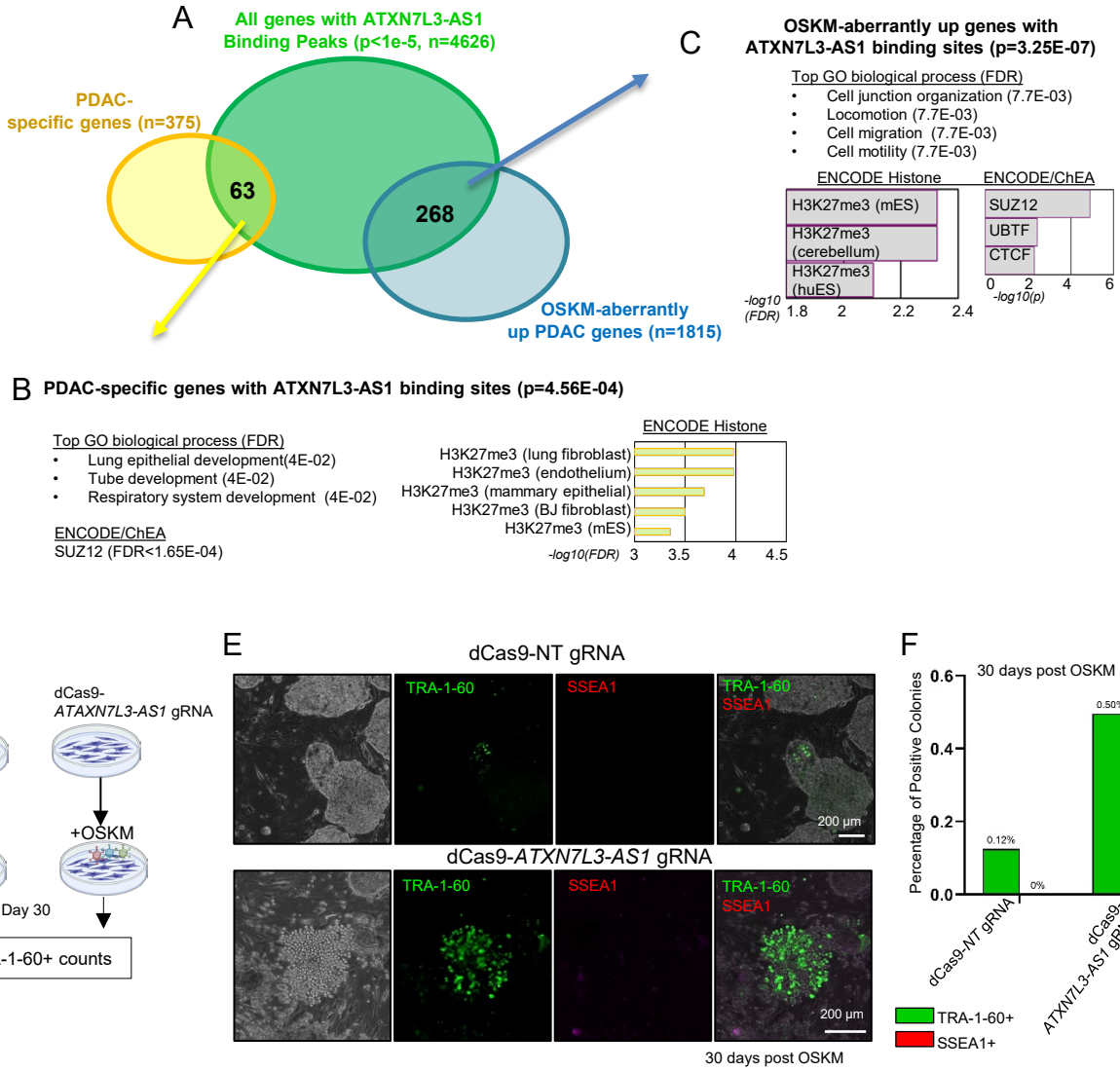
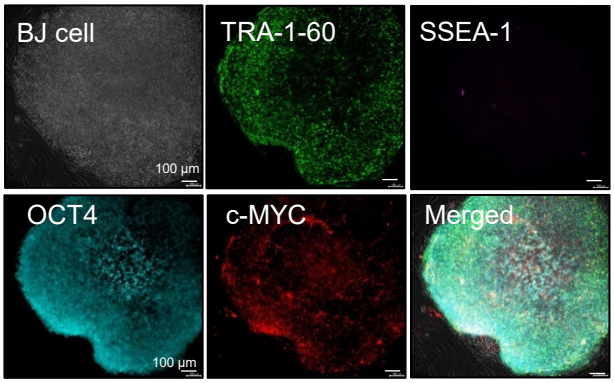


Figure 8. *ATXN7L3-AS1* Binds Malignant Genes and Impedes OSKM Reprogramming in PDAC. A. Intersection of genes with *ATXN7L3-AS1* binding sites with PDAC-specific genes and OSKM-aberrantly upregulated genes (p=4.56E-04, 3.25E-07, respectively, hypergeometric test). B-C. Enriched GO categories, histone modifications, and upstream regulators associated with *ATXN7L3-AS1* peaks overlapping (B) PDAC-specific genes or (C) OSKM-aberrantly upregulated genes. D-F. Reprogramming of PDAC cells depleting *ATXN7L3-AS1* lncRNAs. (D) Schematic of CRISPRi-based depletion of *ATXN7L3-AS1* during OSKM induction in PDAC cells. (E) Representative images of TRA-1-60$^+$ colonies derived from HPAFII-dCas9–*ATXN7L3-AS1* gRNA cells at day 30 of OSKM reprogramming; HPAFII-dCas9–NT gRNA cells serve as controls. (F) Quantification of colony phenotypes showing the percentage of TRA-1-60$^+$ (green) or SSEA1$^+$ (red) colonies among total colonies counted. G. Model of *ATXN7L3-AS1* function in sustaining malignant PDAC transcriptional programs. In normal ductal epithelial cells, migration and invasion programs are silenced. In PDAC, *ATXN7L3-AS1* associates with CpG island promoters to maintain epithelial development, small GTPase signaling, and transcriptional programs in an active state, while keeping migration and invasion pathways in a poised state. Upon *ATXN7L3-AS1* knockdown during OSKM reprogramming, this protection of the cancer genome is lost, allowing OSKM to silence PDAC identity gene programs and thereby enable reprogramming.

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

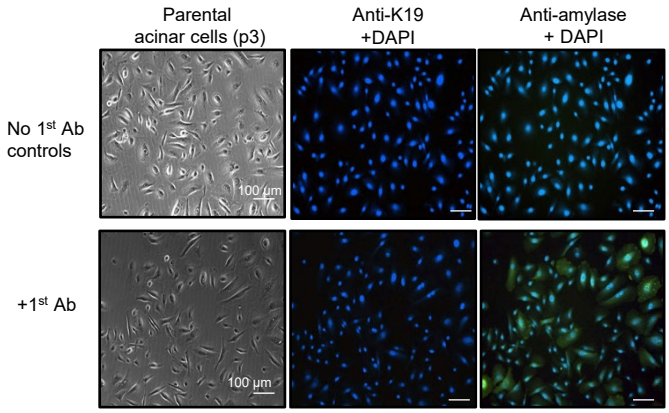
A

Immunostaining of pluripotency markers in BJ6 established iPSC



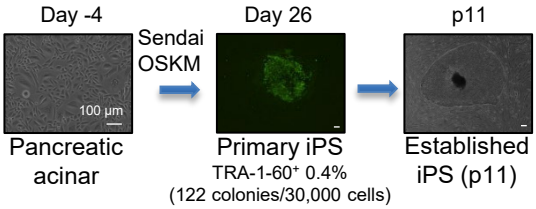
B

Immunostaining of acinar and ductal markers in patient-derived acinar cells



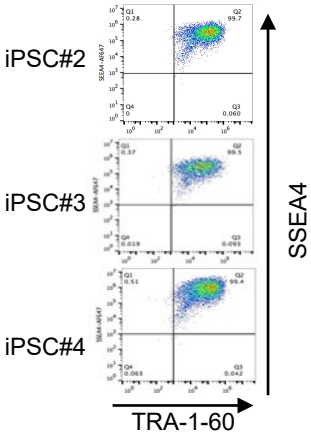
C

Reprogramming of pancreatic acinar cells



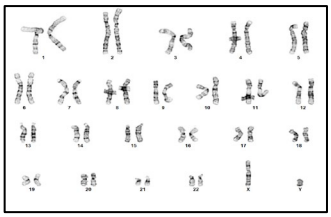
D

Pluripotency



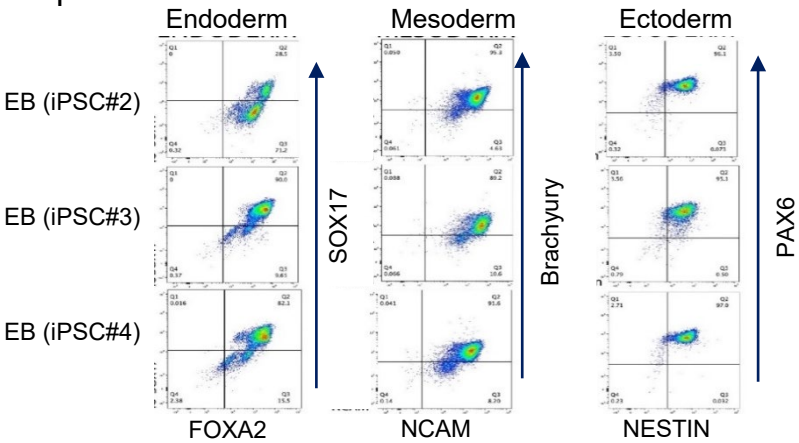
E

iPSC#2



46, XY

F

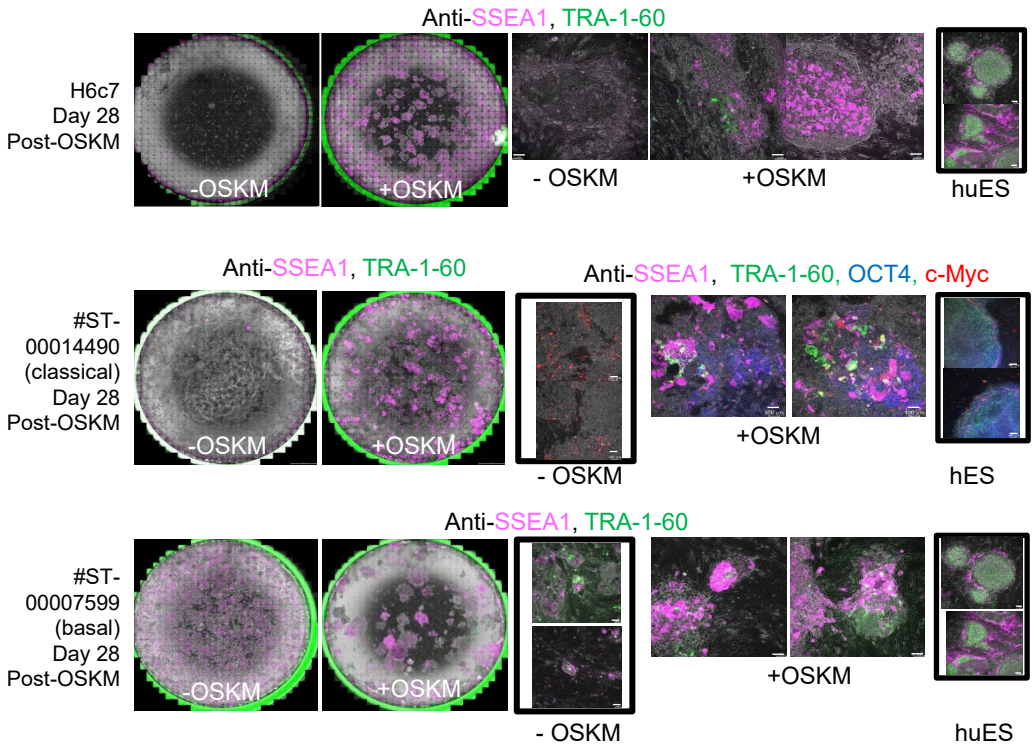


SUPPLEMENTAL FIGURE TITLES and LEGENDS

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

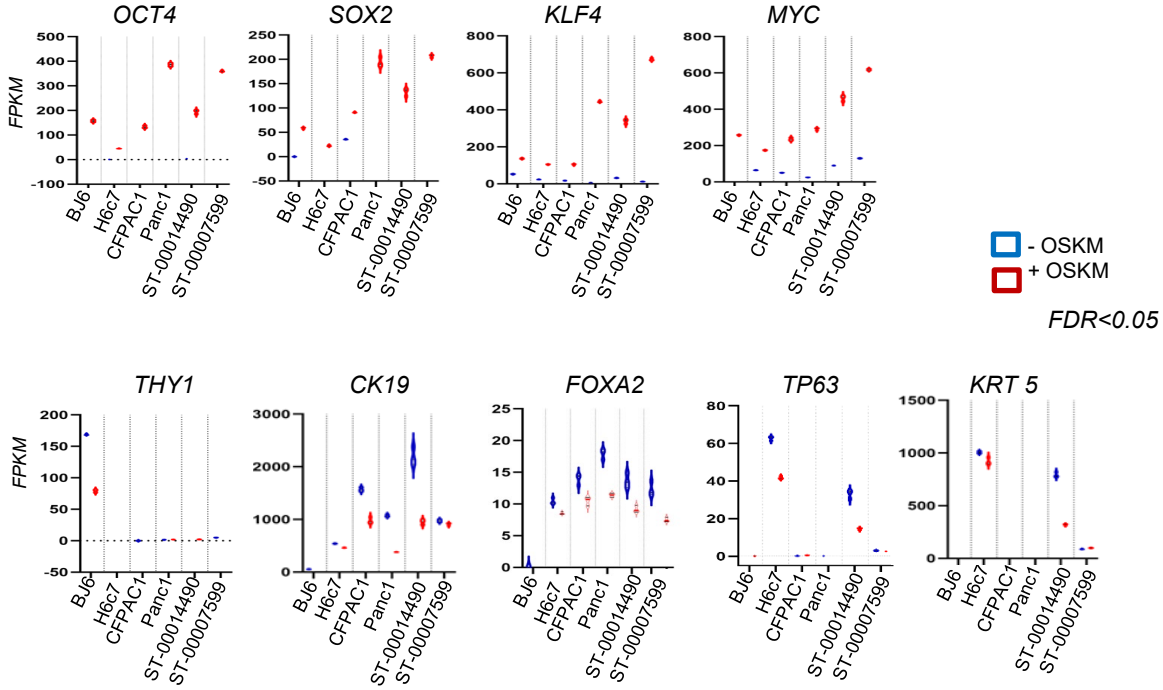


Figure S2 (Relevant to Figure 1). Reprogramming of human normal pancreatic ductal and cancer cells. A. Representative images of colonies stained for TRA-1-60 and SSEA-1 in indicated cells. Scale bars indicate 100 μm. B. Expression of OSKM, pluripotency, and PDAC genes in the indicated cells before and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

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

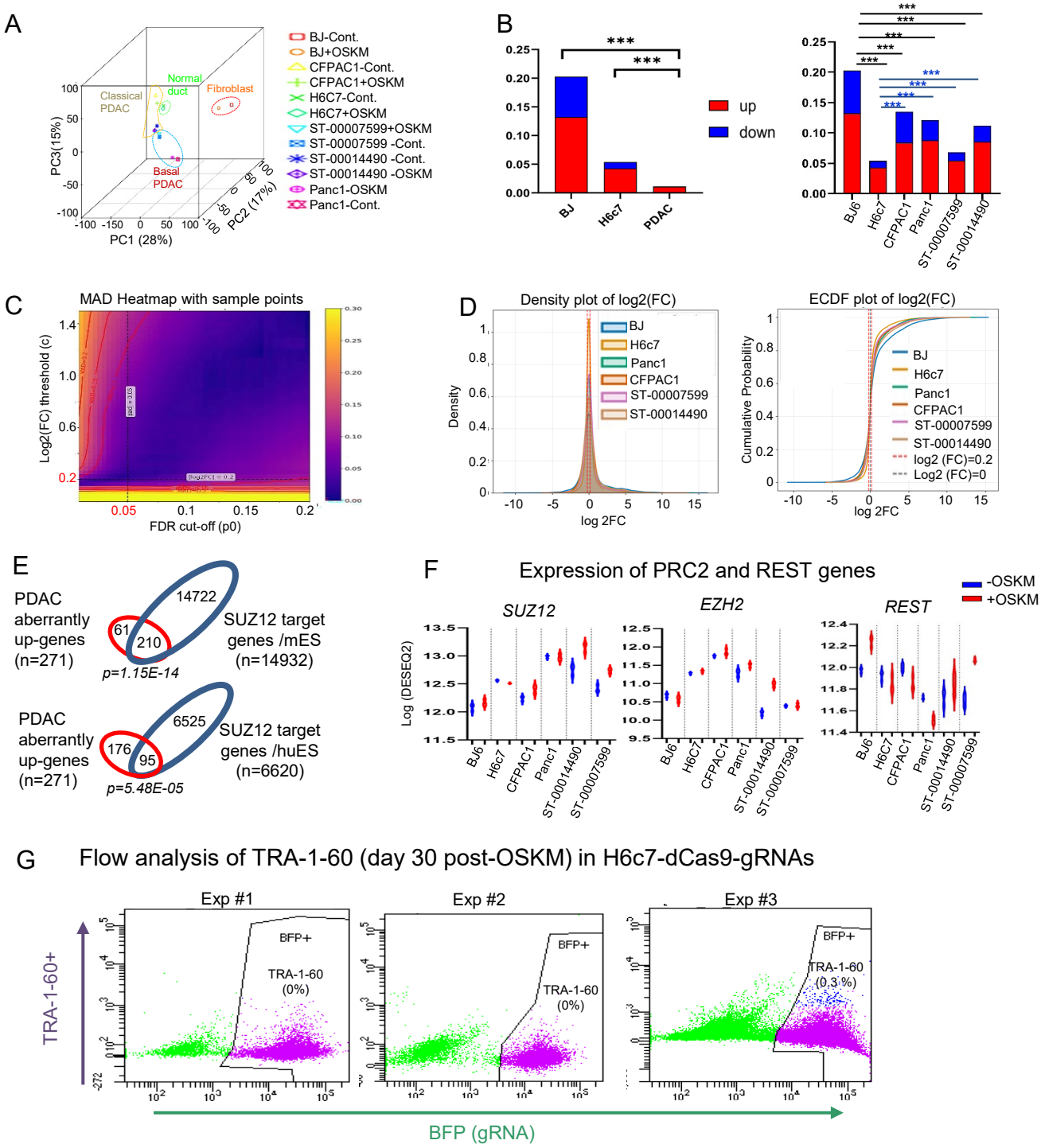


Figure S3 (Relevant to Figures 1-2). A-F. RNA-seq of classical (CFPAC1 and ST-00014490) and basal (Panc1 and ST-00007599) subtypes of PDAC and normal controls (human BJ6 fibroblast and H6c7 hnPDECs) before and 7 days after OSKM induction. (A) PCA plot. (B) Proportion of genes changed upon OSKM induction 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 $< 2.2 \times 10^{-13}$. Black and blue bars indicate comparison between PDAC cells and BJ or H6c7 cells, respectively (See Table S2D-E). (C) Mean Absolute Deviation (MAD) heatmap of Q–Q statistics across FDR cut-off points and $\log_2$ (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. Black 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 of log2 FC values across all samples. Red dashed lines indicate |log2 FC|=0.2. (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. Flow cytometry analysis of TRA-1-60$^+$ cells on day 30 after OSKM induction in H6c7-dCas9 cells expressing gRNAs targeting lncRNAs.

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

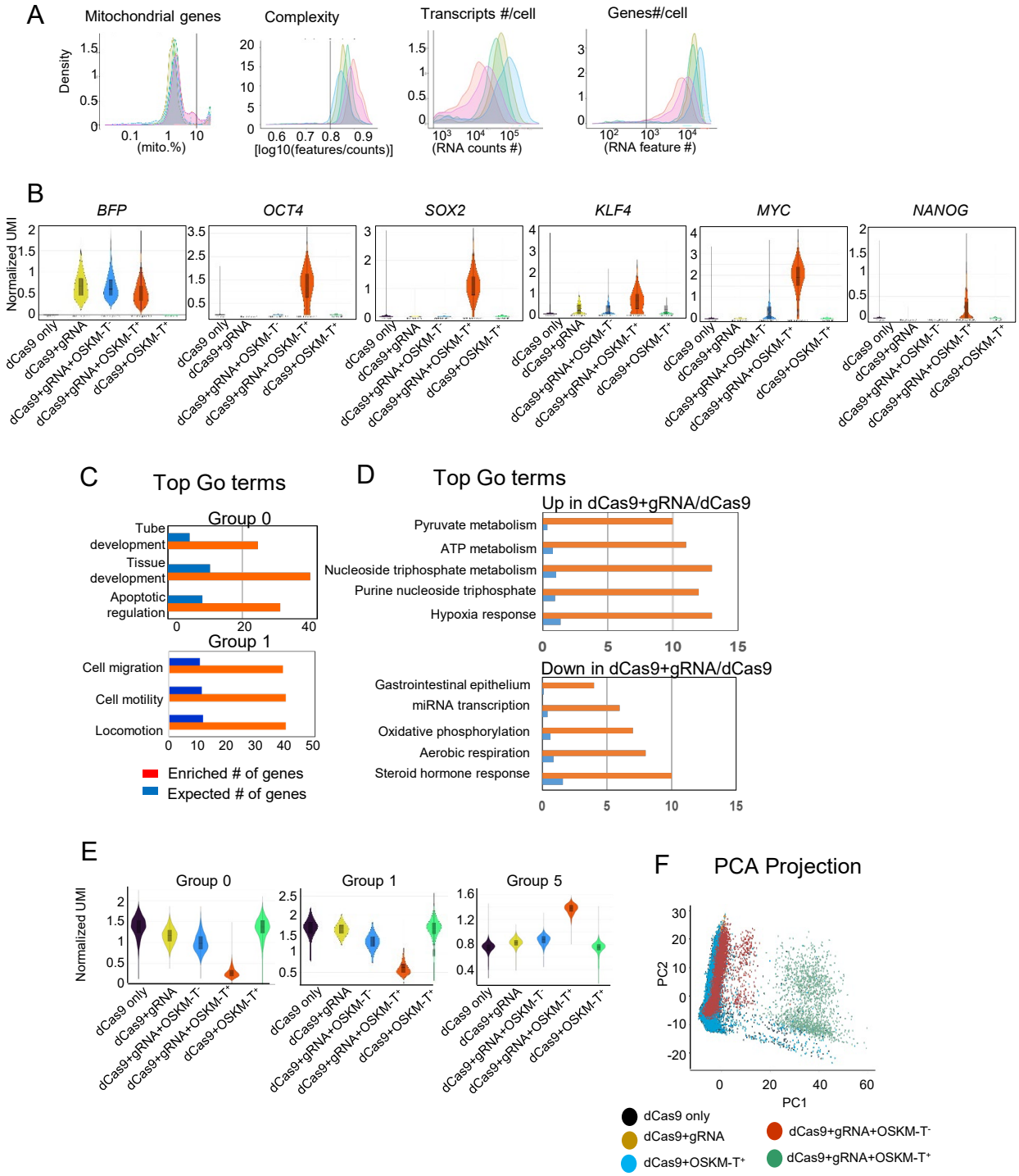
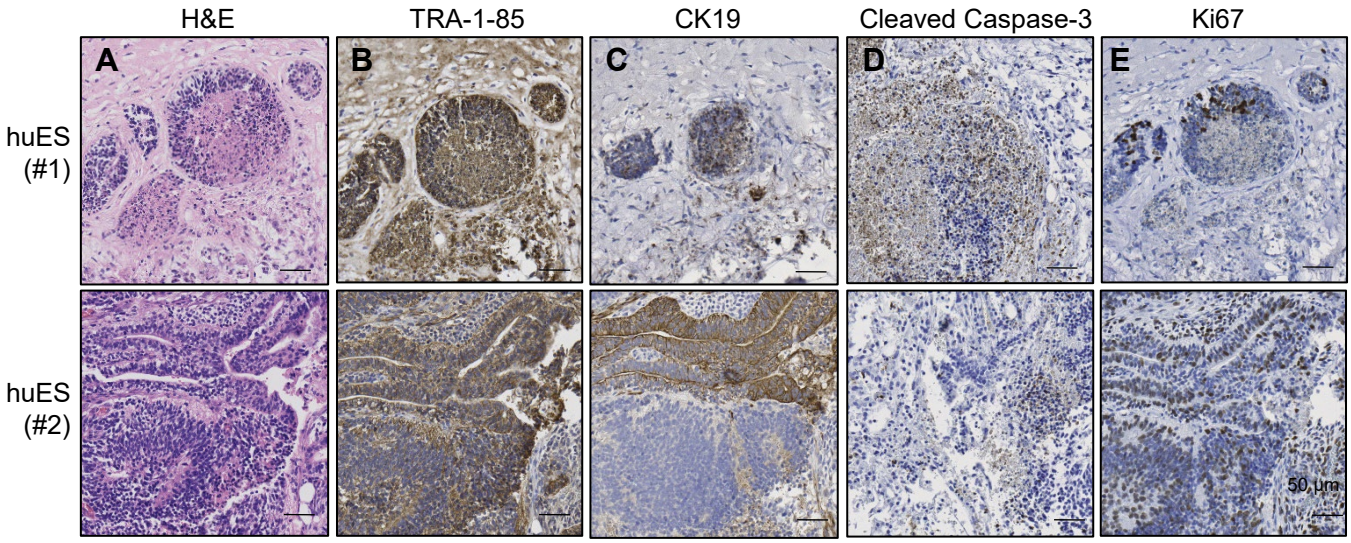
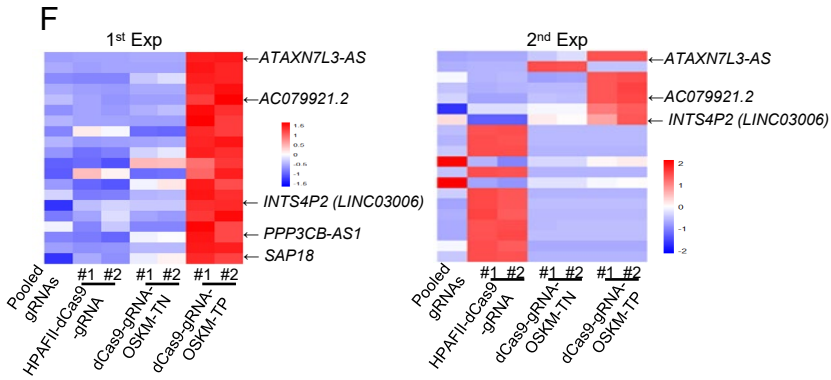
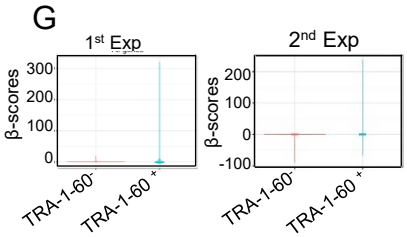


Figure S4 (Relevant to Figure 3) scRNAseq of HPAFII-dCas9+gRNA+OSKM-T^{+} cells along with controls. A. Quality control density plots showing the distribution of cells by key metrics: percentage of mitochondrial gene expression, complexity (features per count), total transcripts per cell, and number of detected genes per cell. B. Expression levels of *BFP*, *OSKM*, and *NANOG* across samples (dCas9 only, dCas9+gRNA, dCas9+gRNA+OSKM-T^{-}, dCas9+gRNA+OSKM-T^{+}, dCas9+OSKM-T^{+}). 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 show 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







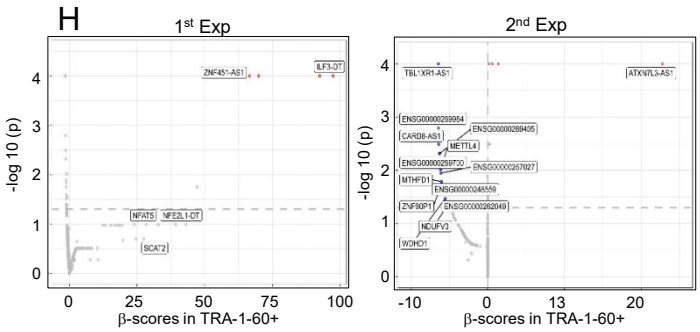
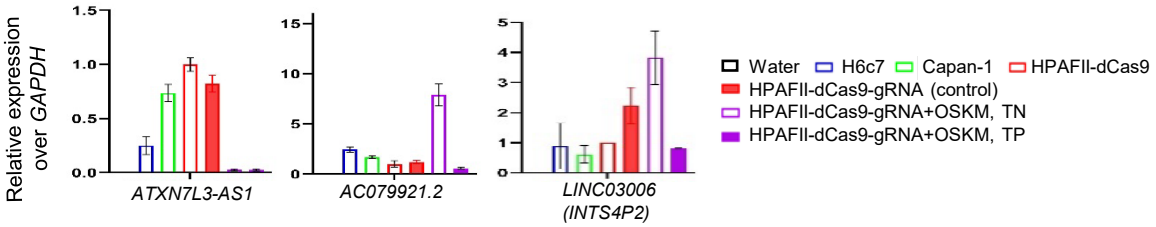


Figure S5 (Relevant to Figure 4-6) A-E. Xenograft of huES cells; (A) Representative images of immunostaining of H&E, (B) human-specific marker TRA-1-85, (C) human ductal marker CK19, (D) apoptotic marker Cleaved Caspase-3, and (E) proliferation marker Ki67. F-H. Identification of PDAC-enriched lncRNAs depleted during reprogramming; (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).

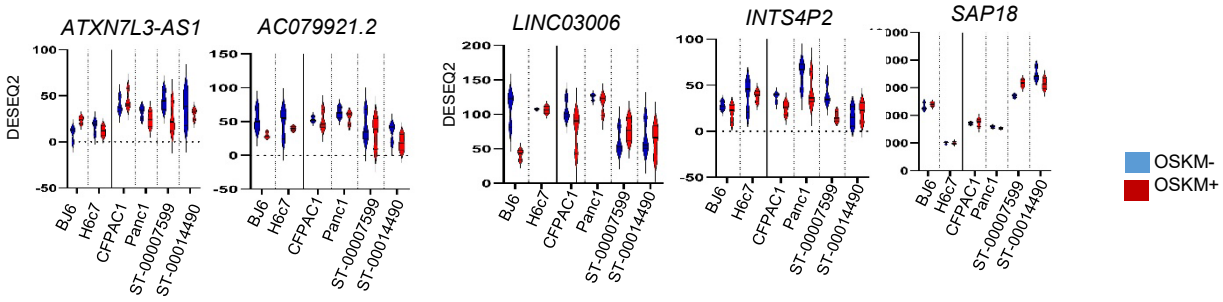
Grygoryev & Lee/Kim et al.,2025

Figure S6

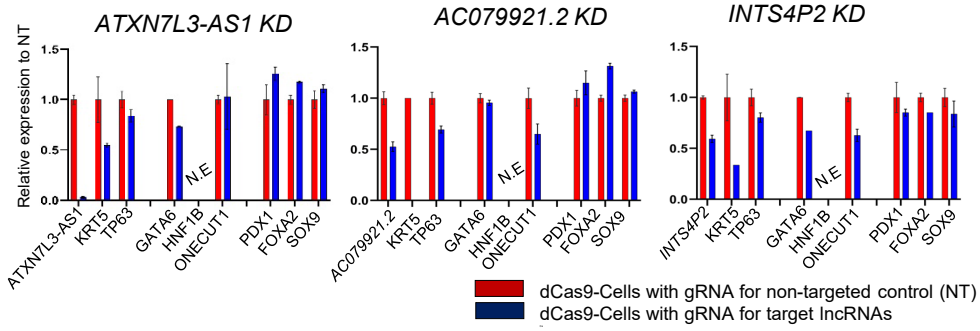
A



B



C PDAC



D Normal ductal cells

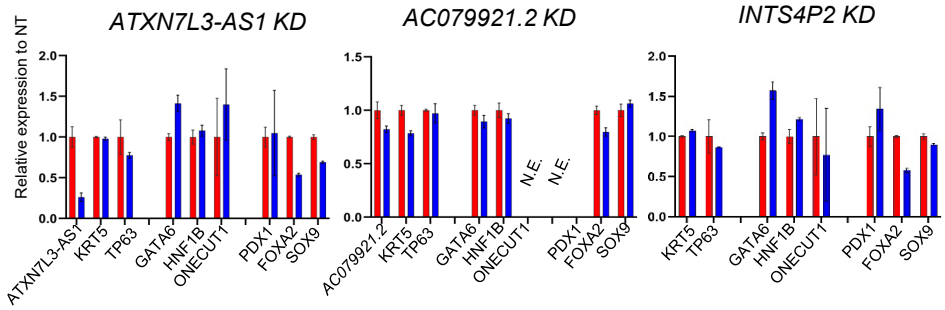


Figure S6 (Relevant to Figure 6) A. The expression of candidate lncRNAs in normal pancreatic cells and cancers, and HPAFII-dCas9+gRNA+OSKM-T^- and T^+ cells (n=3, y-axis indicates relative expression over *GAPDH*, error bars=SD). B. Expression of candidate lncRNAs before and after OSKM induction (n=3, each cell). C-D. Expression of basal subtype markers (*KRT5, TP63*), classical ductal TFs (*GATA6, HNF1B,* and *ONECUT1*), and pancreatic progenitor markers (*PDX1, FOXA2,* and *SOX9*) in PDAC (HPAFII, C) and normal ductal (H6c7, D) cells following CRISPRi-mediated knockdown of *ATXN7L3-AS1*, *AC079921.2*, or *INTS4P2*. Y-axis indicates relative expression normalized to *GAPDH* and scaled to non-targeting controls (NT); error bars represent SD. N=2~3. N.E. indicates no expression.

Grygoryev & Lee/Kim et al.,2025
Figure S7

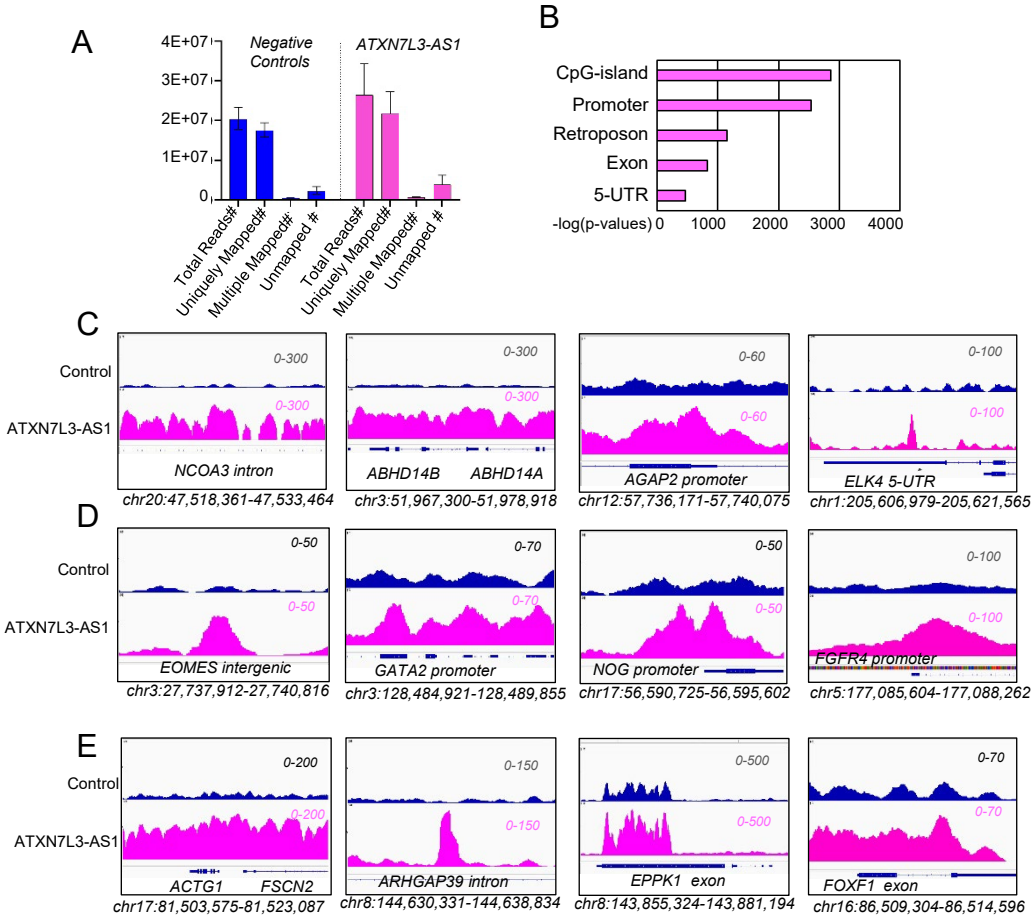


Figure S7 (Relevant to Figure 7) A–E. ChIRP-seq analysis of *ATXN7L3-AS1* genomic occupancy. (A) Read counts from negative control versus pooled *ATXN7L3-AS1* ChIRP samples. (B) Genome-wide distribution of *ATXN7L3-AS1* binding sites. (C–E) Representative genome browser views of *ATXN7L3-AS1* binding over (C) transcriptional activators, (D) mesodermal/neuronal/ ductal developmental genes, and (E) actin cytoskeletal genes.

Grygoryev & Lee/Kim et al.,2025
Figure S8

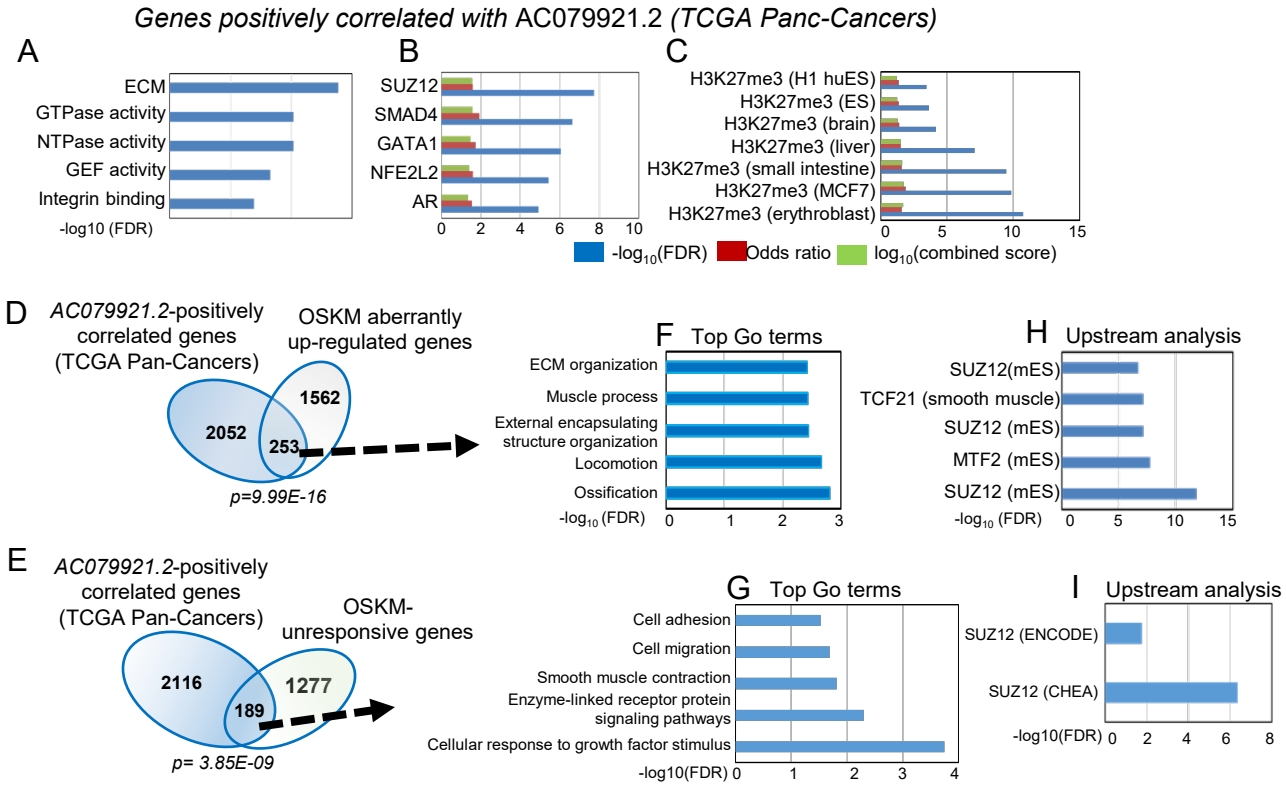


Figure S8. (Relevant to Figure 8) *AC079921.2* Associate with Pro-Invasive Gene Programs. Association of *AC079921.2* with cell adhesion and migration in the TCGA Pan-cancer dataset (TANRIC[107]). A–C Top GO terms (A), potential regulators (B), and histone modifications (C) enriched upstream of genes positively correlated with AC079921.2 in TCGA pan-cancer (Spearman correlation > 0.4; n = 2,305, Table S7B). D-E. Overlap of *AC079921.2*–positively correlated genes in pan-TCGA tumors with OSKM-aberrantly upregulated (D) or unresponsive (E) genes in PDAC reprogramming. The hypergeometric test was used for significance. F-G. GO analysis of overlapping genes from (D) and (E), enriched for ECM organization, muscle contraction, cell adhesion, and migration pathways. H-I. Upstream regulator analysis identifies SUZ12 and chromatin-associated complexes as potential regulators of AC079921.2–associated transcriptional programs.

SUPPLEMENTAL TABLE TITLES and LEGENDS

Table S1. (*Related to Figure 1*) Table S1. KRAS genotyping of iPSCs derived directly from primary PDAC samples.

Table S2. (*Related to Figure 1*) OSKM-mediated reprogramming is not sufficient to override PDAC identity.

- **Table S2A.** Regularized $\log_2$-transformed DESeq2 expression values for 375 PDAC-unique genes.
- **Table S2B.** Regularized $\log_2$-transformed DESeq2 expression values for 7 PDAC-unique genes regulated by OSKM induction.
- **Table S2C.** Gene ontology (GO) analysis of 7 PDAC-unique genes downregulated by OSKM induction.
- **Table S2D.** Chi-square test comparing the proportion of genes with expression changes in pooled PDAC vs. normal cells following OSKM induction.
- **Table S2E.** Chi-square test comparing the proportion of genes with expression changes in each PDAC and normal sample following OSKM induction.
- **Table S2F.** List of OSKM-unresponsive PDAC genes (signature: [-1][-1][0][0][0][0], outliers removed, n=1,466, distance < 0.375; related to Fig. 1F).
- **Table S2G.** List of OSKM-aberrantly upregulated PDAC genes (signature: [-1][-1][0/+1][0/+1][0/+1][0/+1], outliers removed, n=1,815, distance < 0.25; related to Fig. 1F).
- **Table S2H.** List of strongly OSKM-aberrantly upregulated PDAC genes (signature: [-1/0][-1/0][1][1][1][1], n=82, distance = 0; related to Fig. 1F).

Table S3. (*Related to Figure 2*) Depletion of cancer-associated lncRNAs enhances reprogramming of PDAC.

- **Table S3A.** List of gRNAs targeting cancer-associated lncRNAs (CRISPRIL17_hg38).

Table S4. (*Related to Figure 3)* Single-cell RNA-seq reveals transcriptional reprogramming toward pluripotency, suppression of PDAC identity, and reactivation of tumor suppressor genes (TSGs)

- **Table S4A.** List of pluripotency marker genes.
- **Table S4B.** List of endodermal and PDAC marker genes.
- **Table S4C.** List of ectodermal and mesodermal marker genes.
- **Table S4D.** List of "PDAC-specific signature genes" (n=332)
- **Table S4E**. List of OncoKB-annotated tumor suppressor genes (total: n=379; excluding oncogenes: n=318).

Table S5. (*Related to Figure 6)* Identification of depleted gRNAs in HPAFII-dCas9+gRNA+OSKM-T^{+} cells.

- **Table S5A.** Sample barcodes and Illumina adaptor sequences.
- **Table S5B.** Read counts aligned to gRNAs (Experiment 1).
- **Table S5C.** Read counts aligned to gRNAs (Experiment 2).
- **Table S5D.** Raw hit counts (Experiment 1).
- **Table S5E.** Raw hit counts (Experiment 2).
- **Table S5F.** Z-scores of each gRNA (Experiment 1).
- **Table S5G.** Z-scores of each gRNA (Experiment 2).
- **Table S5H.** MAGeCK-MLE output (aggregated reference x500, Experiment 1).
- **Table S5I.** MAGeCK-MLE output (Experiment 2).
- **Table S5J.** MAGeCK-MLE output (combined Experiment 1 and 2).
- **Table S5K.** Final list of selected candidate gRNAs.

Table S6. *(Related to Figure 6-7)* COSMIC Pathogenic somatic variants in *ATXN7L3-AS1* (GRCh38) and ChIRP-Seq for *ATXN7L3-AS1*.

- **Table S6A**. Pathogenic somatic variants in *ATXN7L3-AS1* identified in pan-cancers.

- **Table S6B**. Pathogenic somatic variants in *ATXN7L3-AS1* identified in pancreatic cancer.
- **Table S6C**. *ATXN7L3-AS1* CHIRP probes
- **Table S6D**. *ATXN7L3-AS1* ChIRP-Seq DNA Mapping
- **Table S6E**. Normalized DESeq2 counts of genes in Group 1 (red; basal subtype–associated, n = 552).
- **Table S6F**. Normalized DESeq2 counts of genes in Group 2 (yellow; normal pancreatic ductal cell–associated, n = 460).
- **Table S6G**. Normalized DESeq2 counts of genes in Group 3 (blue; fibroblast-associated, n = 702).
- **Table S6H**. Normalized DESeq2 counts of genes in Group 4 (purple; cancer-associated, n = 1,839).

Table S7. (Related to Figures 7 and S8) TCGA Pan-Cancer Genes Correlated with *ATXN7L3-AS1* and *AC079921.2*.

- **Table S7A.** TCGA Pan-Cancer Genes Positively Correlated with *ATXN7L3-AS1* (Spearman correlation > 0.4)
- **Table S7B.** TCGA Pan-Cancer Genes Positively Correlated with *AC079921.2* (Spearman correlation > 0.4)
- **Table S7C.** TCGA Pan-Cancer Genes Negatively Correlated with *ATXN7L3-AS1* (Spearman correlation <- 0.4)
- **Table S7D.** TCGA Pan-Cancer Genes Negatively Correlated with *AC079921.2* (Spearman correlation <- 0.4)