Single-cell mapping identifies MSI+ cells as a common origin for diverse subtypes of pancreatic cancer

Nirakar Rajbhandari[1,2,†], Michael Hamilton[1,2,†], Cynthia M. Quintero[3,4], L. Paige Ferguson[1,2], Raymond Fox[1,2], Christian M. Schürch[5], Jun Wang[6], Mari Nakamura[3,4], Nikki K. Lytle[1,2], Matthew McDermott[1,2], Emily Diaz[1,2], Hannah Pettit[3,4], Marcie Kritzik[1,2,3,4], Haiyong Han[7], Derek Cridebring[7], Kwun Wah Wen[8], Susan Tsai[9], Michael G. Goggins[10], Andrew M. Lowy[11], Robert J. Wechsler-Reya[4,6,12,13], Daniel D. Von Hoff[7], Aaron M. Newman[14], Tannishtha Reya[1,2,3,4,15,*]

[1]Departments of Pharmacology and Medicine, University of California San Diego School of Medicine La Jolla, CA

[2]Moores Cancer Center, University of California San Diego School of Medicine, La Jolla, CA

[3]Department of Physiology and Cellular Biophysics, Columbia University Medical Center, New York City, NY

[4]Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York City, NY

[5]Department of Pathology and Neuropathology, University Hospital and Comprehensive Cancer Center Tübingen, Tübingen, Germany

[6]Tumor Initiation and Maintenance Program, NCI-Designated Cancer Center, SBP Medical Discovery Institute, La Jolla, CA

[7]Molecular Medicine Division, The Translational Genomics Research Institute, Phoenix, Arizona

[8]Department of Pathology, Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, CA

[9]Department of Surgery, The Medical College of Wisconsin, Milwaukee, WI

[10]Departments of Pathology, Medicine and Oncology, Johns Hopkins School of Medicine, Baltimore, MD

[*]Correspondence: t.reya@columbia.edu.
[†]These authors contributed equally

[11]Department of Surgery, Division of Surgical Oncology, University of California San Diego School of Medicine, La Jolla, CA

[12]Rady Children's Institute for Genomic Medicine, San Diego CA

[13]Department of Neurology, Columbia University Medical Center, New York City, NY

[14]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA

[15]Lead Contact

Abstract

Identifying the cells from which cancers arise is critical for understanding the molecular underpinnings of tumor evolution. To determine whether stem/progenitor cells can serve as cells of origin, we created a *Msi2-Cre^{ERT2}* knock-in mouse. When crossed to *CAG-LSL-Myc^{T58A}* mice, *Msi2-Cre^{ERT2}* mice developed multiple pancreatic cancer subtypes: ductal, acinar, adenosquamous and rare anaplastic tumors. Combining single-cell genomics with computational analysis of developmental states and lineage trajectories, we demonstrate that MYC preferentially triggers transformation of the most immature MSI2$^+$ pancreas cells into multi-lineage pre-cancer cells. These pre-cancer cells subsequently diverge to establish pancreatic cancer subtypes by activating distinct transcriptional programs and large-scale genomic changes, and enforced expression of specific signals like Ras can redirect subtype specification. This study shows that multiple pancreatic cancer subtypes can arise from a common pool of MSI2$^+$ cells and provides a powerful model to understand and control the programs that shape divergent fates in pancreatic cancer.

Blurb (In brief):

Rajbhandari *et al.* report the development of a *Msi2-Cre^{ERT2}* mouse model which shows that multiple pancreatic cancer subtypes can arise from a common pool of MSI2$^+$ cells and provides a powerful framework to understand the programs that shape divergent fates in pancreatic cancer.

Graphical Abstract
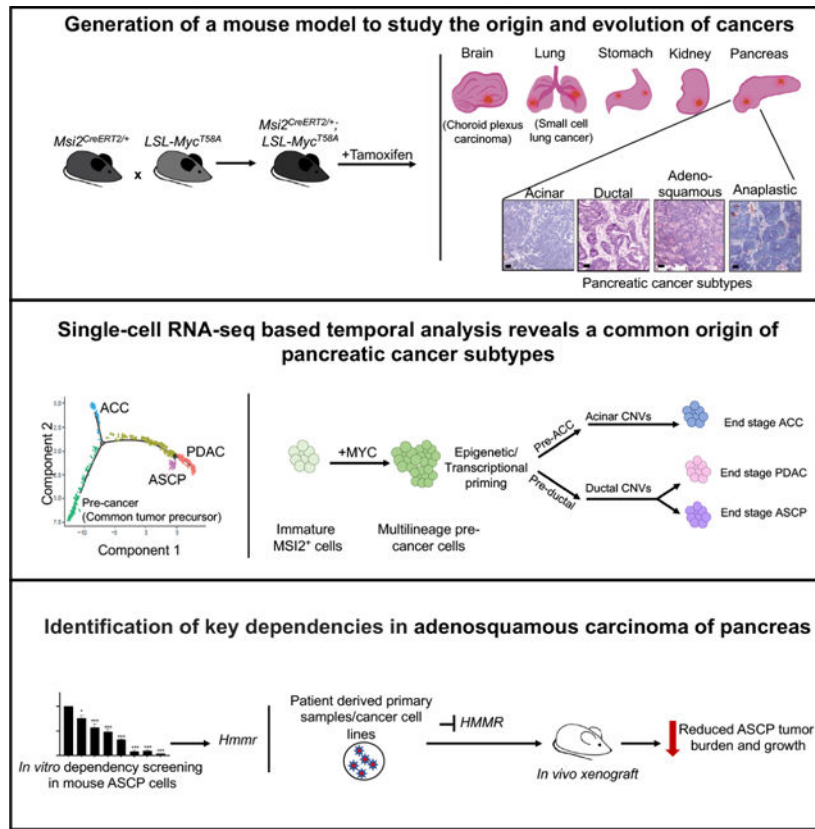
Pancreatic cancer has the highest mortality and lowest 5-year survival rate among all major cancers[1]. Pancreatic ductal adenocarcinoma (PDAC) is an aggressive and common form of exocrine cancer[2,3], and has received the most attention in terms of understanding its genesis and its vulnerabilities. However, within the exocrine pancreas, multiple other forms of malignancies can occur. This includes adenosquamous carcinoma of the pancreas (ASCP), an aggressive subtype of PDAC that has the worst clinical prognosis[4–6]; acinar cell carcinoma (ACC), which accounts for 15% of all childhood pancreatic cancers[7] and is only moderately responsive to chemotherapies[8]; and the highly aggressive anaplastic tumors, characterized by their undifferentiated state[9]. While originally considered rare, recent studies have indicated that a substantially higher fraction of PDAC (18%) harbors admixed adenosquamous features than was previously recognized[10]. Further, what is considered the 'basal subtype' (40% of PDAC) in fact aligns well with the molecular signature of the adenosquamous subtype. This raises the possibility that a substantive fraction of pancreatic cancers may not have properly matched mouse models and highlights the need for new models that better reflect disease variance and subtype [10,11]. While not as rapidly progressing as PDAC or ASCP, acinar cell carcinomas are nonetheless highly aggressive with an overall 5-year survival rate of less than 10%; 50% of patients present with metastatic disease and have a median survival of 14 months[12]. Thus, better models can enable understanding the biological basis of ACC, and the development of therapeutic approaches to ensure more durable remissions for this disease as well.

Although there are multiple subtypes of pancreatic cancers, they are currently treated as one disease. Defining where these tumors arise and how they differ during initiation and progression can enable an understanding of their molecular underpinnings and the development of subtype-specific therapies. While multiple studies have focused on acinar versus ductal cells as an origin for PDAC[13–18], they have almost exclusively used mutant *Kras*. Gain-of-function mutations in *KRAS* are present in nearly 90% of human PDAC cases[19,20], and are considered the primary driver of pancreatic cancer, However, mutant *Kras* generally drives classical PDAC in animal models[21]. Thus, how other pancreas cancer subtypes arise, and the trajectories cells take to establish these malignancies have remained relatively unexplored. Here we develop a mouse model in which overexpressing oncogenic c-MYC in adult pancreatic precursor cells triggered the development of PDAC, ASCP, ACC, and anaplastic tumors. Our current model is particularly exciting and highly relevant to human disease, as MYC has been reported to be overexpressed in more than 40% of primary patient tumors and amplified in over 30% of them[22]. Deregulated expression or amplification of MYC is particularly associated with a higher incidence of ASCP (47%) and ACCs (54% in pure and up to 100% in mixed ACCs) in human pancreatic cancer patients[5,23]. *MYC* can be an initiating event even in the absence of *KRAS* mutation[5]. Further, a whole exome sequencing study of pancreatic cancers found that MYC amplified cases were not significantly associated with established genetic lesions, including *KRAS*, *TP5*3, *SMAD4* and *CDKN2A*[24]. In addition, genetic alterations frequently associated with pancreatic cancer including KRAS are uncommon in ACC[23] but a significant percentage of ACC contains MYC alterations. This study thus provides a previously unavailable model for multiple subtypes of pancreatic cancer, and a unique opportunity to trace, at single-cell resolution, the origins of the pre-malignant state and its evolution to distinct lethal malignancies.

Results

MSI2 expression marks the cell of origin for cancers across diverse tissues

To understand the tumor-initiating potential of Musashi-2 (MSI2[+]) stem/progenitor cells, we developed a *Msi2-Cre*[ERT2] knock-in mouse (*Msi2*[CreERT2/+]) model in which a *CreERT2*-cassette was integrated into the "ATG" codon of the *Msi2* locus (Figure 1A–B). To assess the fidelity of this model, we analyzed endogenous MSI2 expression in tissues. MSI2 was found to be expressed in stem/progenitor enriched populations in SRY box transcription factor 2 (SOX2[+]) cells of the subventricular zone (Figure 1C), club cells at the bronchoalveolar junction in the lung (Figure 1D), calbindin+ cells in the kidneys (Figure 1E). Msi2 was also expressed in ductal epithelial cells in the pancreas (Figure 1F). We crossed *Msi2-Cre*[ERT2] mice to *R26R-LSL-tdTomato* reporters[25] to compare *Cre* activity to *Msi2* expression. TdTomato expression following tamoxifen-driven *Cre* activation was strongly aligned with Msi2 expression (Figure 1G–J and Figures S1A–D), confirming the fidelity and specificity of this knock-in model.

To determine if normal MSI2[+] cells can serve as an origin for cancers, we crossed *Msi2-Cre*[ERT2] mice with the *CAG-LSL-Myc*[T58A] model[26] (Figure 1K) to target MYC, a critical oncogene to Msi+ cells. Although MYC is usually amplified rather than mutated, expression

of wildtype MYC is often insufficient to drive cancers in mouse models [27,28]; therefore the *Myc^T58A* allele, which stabilizes the MYC protein thereby increasing its levels, has been used extensively to model MYC amplification[26,29,30]. Tamoxifen treated *Msi2-Cre^ERT2*; *CAG-LSL-Myc^T58A* mice (referred to here as *Msi2-Myc*) developed a range of cancers, including choroid plexus carcinoma (Figure 1L and Figure S1E), small-cell lung cancer (Figure 1M and Figure S1F), gastric carcinoma (Figure 1N and Figure S1G), and pancreatic cancer (Figure 1O). In the kidney, MYC drove dysplasia (Figures S1H–I) and cooperated with VHL deletion to drive kidney cancer (data not shown). Tumor penetrance varied between tissues; of the mice analyzed, 80% had lung tumors, 68% gastric tumors, 52% brain tumors, and 45% pancreatic tumors (Figure 1P). These data identify MSI2$^+$ cells as a cell of origin for diverse cancers and define a new autochthonous model for understanding the progression of MYC-driven tumors.

A model for multiple pancreatic cancer subtypes

The tumors that developed in the pancreas of the *Msi2-Myc* model were of particular interest, as they represented a diversity of pancreatic cancer subtypes: pancreatic ductal adenocarcinoma, adenosquamous carcinoma of the pancreas, acinar cell carcinoma, and anaplastic tumors. Tumor subtypes were identified using histology and canonical markers: cytokeratin19 and absence of amylase (AMY$^−$) for PDAC (Figure 2A), cytokeratin 5/14 for tumors with squamous features (SF) (Figure 2B), and/or p63 for squamous differentiation (SD) (Figure 2C) within squamous tumors; amylase for ACC (Figure 2D); and the absence of differentiation markers for anaplastic tumors (Figure 2E). While some *Msi2-Myc* mice harbored only one subtype, others presented with two or more (Figure 2F). Overall, ductal tumors with squamous features (CK5, CK14) or frank ASCP (p63+) were found in 64% (16/25), ACC in 72% (18/25), and anaplastic tumors in 8% (2/25) of mice analyzed (Figure 2G).

RNA sequencing on flow-sorted epithelial tumor cells revealed that ACC and ASCP subtypes harbored transcriptomic landscapes distinct from each other and the cell of origin (Figure 2H), where ASCP tumors were more closely related to *Kras/p53* driven (KPC) tumors[31] consistent with their ductal nature. In addition, ASCP tumors showed higher expression of squamous markers such as *Krt5, Krt6a,* and *Trp63* (Figure 2I) and enriched (N) *Trp63* transcripts (Figures 2J).

To determine the concordance between pancreatic tumors from the *Msi2-Myc* model and human disease, we compared transcriptomic and genomic signatures. Comparison of the transcriptomic landscape of human and mouse ACC tumors revealed the conservation of both enriched and depleted pathways (Figure 3A). Within depleted pathways, genes that are commonly deleted or have loss-of-function mutations in human ACC were also depleted in mouse ACC (n=4) (Figure 3B). While a number of deregulated pathways are shared between mouse and human ACC tumors, enriched and depleted genes were less conserved at the gene level in mouse and human ACC (21 enriched, FDR<5%, Figure 3C). Genome-level analysis revealed the conservation of amplified genes (Figure 3D) in mouse and human ACC specifically in programs related to autophagy, amino acid metabolism, oxidative phosphorylation and apoptosis (Figure 3E).

The transcriptomic landscape of mouse and human ASCP tumors showed conservation of enriched pathways (Figures 3F–G), along with gene sets previously implicated in pancreatic cancer progression (Figure 3H). At a genomic level, mouse and human ASCP tumors showed conservation of amplified genes (Figure 3I) implicated in the regulation of critical programs such as Ras/Notch signaling, MAP kinase signaling, and apoptosis (Figure 3J).

To further validate the relevance of MYC in human pancreatic malignancy, we immunostained human pancreatic pre-malignant lesions and end-stage cancers for MYC. MYC was highly expressed in a subset of human pancreatic preneoplastic lesions (n=6 out of 15 tissue spots analyzed) (Figure S2A) and in all end-stage ASCP tumors analyzed (n=3) (Figure S2B). MYC was also expressed in human ACC tumors (n=3) albeit at lower levels (Figure S2C). As MYC expression is typically not observed in the normal mouse and human pancreas, these findings suggest MYC could be an early oncogenic event that contributes to the development of human ACC and ASCP. Thus, the *Msi2-Myc* model reflects the diversity of human pancreatic cancers with high molecular and histologic fidelity and provides a foundation for understanding the molecular basis of subtype establishment.

Single-cell map of the cell of origin and pre-cancer states in pancreas cancer subtypes

The fact that multiple subtypes of pancreatic cancers arose from a common pool of MSI2$^+$ cells provided a rare opportunity to define how normal cells transform and diverge into distinctly fated tumors. Thus, we used single-cell sequencing to trace the molecular trajectory of Msi2+ cells (*t=0*) as they transitioned in real-time through a pre-cancerous state (5 weeks post-tamoxifen; *t=5*) to PDAC, ASCP, or ACC tumors (10–12 weeks post-tamoxifen, *t=12*) (Figure 4A).

Analysis of Msi2-GFP knock-in reporter mice (*Msi2$^{eGFP/+}$*) showed that at initiation, MSI2 was expressed predominantly in ducts and islet cells and a few acinar cells (Figure S3A–B). MSI2$^+$ duct cells fell into three clusters (D1-D3, Figure 4B). To define the hierarchy of these clusters, we used CytoTRACE, a recently developed computational pipeline that can predict the most immature cells within single-cell populations by tracking transcriptional diversity[32] predicted D1/D2 as the most undifferentiated populations (dark red) and D3 as more differentiated (green) (Figure 4C). Consistent with this, D1 and D2 clusters were enriched for stem cell and re-programming genes such as *Prom1, Sox4, and Gata6* (Figure 4D) as well as *Oct4, Tead2,* and *Notch* (not shown). However, D3 was more differentiated, with lower expression of most stem cell genes and elevated expression of Forkhead box J1 (*Foxj1*), a transcription factor required for ciliated cell differentiation (Figure 4E).

Since tumors of various lineages were found to arise from MSI2$^+$ cells, we used lineage tracing to define if MSI2$^+$ cells normally give rise to multiple lineages. *Msi2-Cre^{ERT2}*; *R26R-LSL-tdTomato$^{+/−}$* reporters were tamoxifen pulsed postnatally and showed labeling predominantly in duct cells and occasional acinar cells (range = 0.0 to 0.15%, mean = 0.05%) at t=0. Following a chase (*t* = 10 weeks*)*, only labeled duct cells expanded (from 12.6 ± 0.5% at pulse to 34.1 ± 6.4% at chase) (Figures S4A–C). Further, lineage tracing of adult mice (8 weeks) showed no significant expansion of either duct or acinar cells (Figures S4D–E). These data indicate that while postnatal, MSI2$^+$ ductal cells do not appear to give

rise to cells of other lineages, they harbor a latent potential to form both ductal and acinar lineages that can be triggered by oncogenic MYC.

To define how MYC initiates transformation, we mapped the transcriptomic diversity of cells following MYC induction. *Msi2-Myc* mice crossed to *R26R-LSL-YFP* mice[33], and YFP+ (i.e MYC-expressing) epithelial cells were sorted and analyzed at 5 weeks (*t=5*). Interestingly these pre-cancer cells were dominated by large pool of undifferentiated cells (Figure 4F) expressing both ductal and acinar signatures (not shown) and marked by the expression of *Hmmr* (Hyaluronan mediated motility receptor) (Figure 4F). Histologically, these pre-cancer cells generally presented as patches of undifferentiated cell clusters or 'tumorlets' (Figure S4F–I) with some lesions harboring aberrant ductal morphology (Figure 4G and Figure S4J). Cell cycle and chromatin-modifying genes (*Hmgb2* and *Top2a*) and a multitude of oncogenic programs (*Aldh1a1, Ran, Ube2c,* and *Klf2*) (Figure 4H) were enriched in these pre-malignant lesions relative to the normal pancreas (t=0). Embryonic stem cell and cell-cycle signatures (Figure 4I) were also enriched in these uncommitted cells at t=5. Metagene analysis showed that gene signatures of D1 and D2 were the most enriched within the pre-cancer cells (Figure 4J), suggesting that the more undifferentiated D1/D2 duct cells may be the most sensitive to MYC-driven transformation.

Developmental trajectories of pancreas cancer subtypes

To understand the basis of the divergence of distinct tumor subtypes from this common pool of pre-cancerous cells, we analyzed end-stage (t=12) acinar and adenosquamous tumor cells from *Msi2-Myc* mice. Single-cell RNA-sequencing of acinar tumors were defined by lipase (*Cel*) and β-catenin (*Ctnnb1*) (Figure 5A) and harbored a distinct population predicted by CytoTRACE as less differentiated. CytoTRACE and differential expression analysis identified *Hmgb2* and *Top2a* among the genes most correlated and enriched within the less differentiated population, respectively. Further, the less differentiated cell population harbored the fewest copy number variations (CNVs) genomic changes which usually accumulate during tumor evolution and can drive progression (Figure 5B). Monocle, an algorithm that can project single-cell relationships based on the kinetics of gene expression[34] built a single trajectory originating from this predicted precursor population (Figure 5C, cells in green), to established ACC cells (Figure 5C).

To understand whether ductal/squamous lineage tumors evolve from distinct origins and progress along distinct paths relative to ACC, we analyzed three independent tumors, one PDAC with squamous features i.e. PDAC/ASCP(SF) (Figure 5D–F), one with p63+ squamous differentiation i.e. ASCP(SD), (Figures 5G–I) and one with more classical PDAC signatures (not shown). The most immature clusters in the bi-fated ductal/squamous tumors and the p63+ ASCP tumor were similar to the most immature cluster in ACC: all expressed *Hmgb2/Top2a* (Figures 5E, 5H left and middle), and contained the fewest CNVs. CNVs accumulated progressively with increasing differentiation, highlighting a remarkable alignment of how the genomic evolution of the tumors paralleled its epigenetically controlled progression towards a differentiated lineage (Figures 5E and 5H right). In the bi-fated tumor, both PDAC and ASCP arose from the *Top2a/Hmgb2+* population (Figures 5E–F) but diverged at a branch point either continuing to a PDAC fate (*Sox9, Agr2,* and *Tff2*

at endpoint) or turning towards a squamous fate (*Krt5, Krt6a, Krt15 and Krt7* at endpoint) (Figure 5F).

Mixed pancreatic cancer subtypes in *Msi2-Myc* mice arise from a common pool of precancer cells

Based on the computational data above, we experimentally tested whether the common pool of transcriptionally uncommitted cells (t=5) can give rise to multiple pancreatic cancer subtypes. To this end, we orthotopically transplanted this pool of recombined cells into the pancreas of recipient (NSG) mice (Figure 6A). Pre-cancer cells were sorted from *Msi2-Myc; R26-LSL-tdTomato* mice to identify cells with MYC activation and an enriched expression of *Hmmr* (Figure 4F, bottom) (*n* = 3 biological replicates). Following transplantation, all recipient mice developed tumors. Single-cell RNA-sequencing on sorted cancer cells from the end-stage tumors (n=3) revealed significant heterogeneity with both acinar and ductal subtypes classified based on lineage-specific genes (Figures 6B and 6C), suggesting these cells can recreate the mixed nature of the primary tumors (Figures 5A–I). In addition, uncommitted progenitor cells persisted in these mixed tumors (Figure 6C) reflecting the undifferentiated cells we identified in primary autochthonous tumors. Together these findings provide experimental evidence that multiple subtypes of pancreatic cancer can arise from a common pool of pre-cancerous cells which serves as common tumor precursors in *Msi2-Myc* mice (Figure 6D).

Consistent with this observation, genomic analysis of ACC, PDAC and ASCP isolated from the same mice shared common CNVs, indicating both lineages arose from a common tumor precursor that later diverged (Figure 6E, top, and bottom). This finding was further supported by the integration of all the scRNA-seq data from midpoint t=5 samples and all tumor subtypes (n = 5) showing a transcriptional overlap of cell populations (Figure 6F). Cell populations found at t=5 were also found in both ACC and ASCP tumors. The pre-cancer or common tumor precursor population was found in all t=5 and t=12 samples with a common transcriptomic gene signature, including enrichment of *Hmgb2* and *Top2a* expression (Figure 6G). An equivalent ancestral precursor pool may also be conserved in human ASCP (Figure S5A),[35] with cell populations with a less differentiated state (Figure S5B, left, red), low CNV frequency (Figure S5B, right), and precancer gene signature (Figure S5C). Further, *TOP2A* and *HMGB2*, two key gene markers of precancer cells, were preferentially expressed in the less differentiated cells (Figure S5C, right), where their expression correlated with poor survival in PDAC patients (Figure S5D–G; [www.proteinatlas.org[36]].

Differential contribution of genomic and epigenetic changes in PDAC/ASCP specification

The fact that ACC and ASCP tumor cells displayed distinct patterns of conserved CNVs (Figures 6H–I; many confirmed with exome-seq analysis-green arrows) raised the possibility that CNVs may drive subtype specification during tumor growth. Although lineage-specific CNVs had already begun to emerge in pre-cancerous lesions (t=5) (Figure 6J) aligned with lineage-specific transcriptional programs in distinct pools of cells (Figures 6K–L), we found that lineage-specific CNVs occurred after fate-specific gene expression had already emerged, suggesting that the transcriptional changes that ultimately dictate subtypes were

likely initiated by epigenetic mechanisms (Figure 6J). In support of this, single-cell ATAC-seq revealed that loci for both duct and acinar lineages were open in pre-cancer cells (t=5, Figure S6A), but that acinar loci were epigenetically suppressed as tumors committed to the ductal/squamous lineage (t=12, Figure S6B). These suggest that open ductal and acinar epigenetic landscapes lead to a hybrid multilineage state in pre-cancer cells. This is then followed by epigenetically driven transcriptomic changes that initiate distinct fates, with subsequent selection of CNVs that differentially support the growth of distinct tumor subtypes.

To test if transcriptional programs enriched in ductal lineages can functionally confer ductal fate, we focused on RAS signaling which was found to be amplified and transcriptionally upregulated in ductal-fated cells (Figures 7A–B). Additionally, we also found MYC expression in the current *Msi2-Myc* model was sufficient to activate mutant *Kras* signaling in the later stages of tumor development in ASCP tumors (Figure 7C). We thus crossed *Msi2-Cre^ERT2*; *CAG-LSL-Myc^T58A* mice with *Kras^LSL-G12D/+* mice to generate *Msi2-Myc*; *Kras* mice (Figure 7D). In contrast to the ACC and PDAC/ASCP tumors that occurred when MYC was induced in MSI2+ cells, the inclusion of oncogenic *Kras* in these cells drove a very aggressive form of PDAC with squamous features (Figures 7E–G) with no detectable ACC. This indicates that activation of RAS signaling is sufficient to shift fates towards ductal/squamous lineage and away from ACC and highlights the fidelity and predictive power of the computational modeling.

Identification of dependencies of pancreatic adenosquamous cancers

ASCP remains essentially unresponsive to the current standard of care. Thus, the adenosquamous model we developed provided a unique opportunity to understand not only how ASCPs develop but also identify their vulnerabilities. To test potential dependencies of squamous tumors at later stages, we selected genes that were enriched in end-stage ASCP tumors (Figure 8A), and/or in early precancer lesions (*Hmmr* and *Atf3*), and yet, whose role in pancreatic cancers remains largely unknown. Of these, shRNA-mediated knockdown of *Ifne, Atf3, Hmmr,* and *Sprr3* most significantly and consistently decreased the organoid-forming potential of two independent cancer cell lines derived from *Msi2-Myc* squamous tumors (Figure 8B and Figure S7A and S7B), highlighting the importance of these genes in the continued maintenance of squamous tumor cells. Among these genes, *Hmmr*, was particularly interesting as higher expression of this unconventional extracellular matrix receptor correlates with poor prognosis in many cancer types[37]. Given that our model allowed the identification of molecular programs that could be of impact in early-stage disease we first assessed whether HMMR may be of relevance as an approach to interception. HMMR is undetectable or low in both mouse and human normal pancreas but highly upregulated and localized in the membranes in mouse and patient-derived precancerous lesions (Figures 8C–D). shRNA-mediated knockdown of *Hmmr* in mouse pre-cancer cells (t=5 lesions) significantly reduced organoid formation *in vitro* (Figure 8E), identifying *Hmmr* as a potential dependency during the early stages of tumorigenesis.

In late-stage disease, *Hmmr* inhibition in primary mouse ASCP cells reduced organoid growth by over 50% (Figure 8F). Immunostaining of a clinically annotated pancreatic cancer

tissue microarray (Figure 8G) and real-time quantitative PCR of classical and squamous pancreatic cancer cell lines (Figures S7C–E) showed that HMMR is highly expressed in primary patient samples and patient-derived cell lines, of both ductal and squamous lineages. Given its expression, we knocked down *HMMR* in human pancreatic cancer cell lines (Figure S7F) and found it markedly inhibited the 3D growth of most human pancreatic cancer cells *in vitro* (Figure 8H) with the deepest impact on FG and KLM-1, both adenosquamous lines (Figure S7G). Knockdown of *HMMR* also reduced the growth of orthotopically transplanted cell lines *in vivo* by greater than 2-fold (Figure 8I) with a remarkable reduction in the invasiveness of tumor cells indicated by the presence of uninvolved normal tissues adjacent to the tumor cells (Figure 8I). Consistent with these findings, HMMR knockdown significantly reduced the flank tumor burden of human cell lines in transplanted mice (Figure S7H), thus strongly suggesting that HMMR is important for the growth of advanced cancer, specifically, adenosquamous cancer *in vivo*.

Finally, we knocked down *HMMR* in patient-derived xenograft (PDX) tumors, where PDX tumors were dissociated and infected with GFP-tagged lentiviral shRNA *in vitro*, and then re-transplanted subcutaneously in NSG mice (Figure 8J). While each PDX sample was transduced equivalently at $t = 0$ (Figure 8K), the relative frequency of GFP$^+$ EpCAM$^+$ tumor cells were reduced by 5–20-fold in sh*HMMR* tumors at the endpoint (Figure 8K) identifying HMMR as an exciting key functional dependency of mouse as well as human adenosquamous carcinoma *in vivo*. Collectively, these data demonstrate that the *Msi2-Myc* model can be useful to identify functional dependencies of ASCP and provide a foundation to develop approaches to the interception and targeting of the most lethal subtype of pancreatic malignancies.

Discussion

The work we describe here provides an exciting model of pancreatic cancer and demonstrates that MSI2$^+$ cells can serve as a common origin for multiple pancreatic tumor subtypes (Figure 8L). Specifically, *Msi2-Cre$^{ERT2/+}$*-driven overexpression of MYC preferentially targeted the most immature subfractions of Msi2+ pancreas cells, leading to the expansion of hybrid multilineage pre-cancer cells with both acinar and ductal loci open. Subsequently, subsets of these pre-cancer cells initiated distinct transcriptional programs that in turn triggered specific fates and the development of ductal and acinar tumor subtypes. Unique genomic changes that favor oncogenic growth along distinct lineages were selected as fully differentiated acinar or ductal/ASCP tumors were established. As an example of how distinct programs can shift fates, we found that while MYC alone allowed the development of both ACC and PDAC/ASCP, enforced genetic activation of RAS signaling which was enriched in ductal tumors, drove tumors towards a ductal/squamous fate and away from acinar cell carcinoma. This highlights a critical discovery that defining the underpinnings of each subtype has the potential to enable redirecting subtype specification potentially to more therapy-responsive states.

One of the most important findings of our current work is the demonstration that MSI2$^+$ cells in the pancreas cells can serve as an origin for multiple subtypes of pancreas cancer. Most prior models have focused on driving oncogenic *Kras* and *Trp53* that resulted in

classical PDAC [13,14,18,38–40]. Previously, models overexpressing MYC in the pancreas have been developed (either together with *Kras* or embryonically)[27,28,41,42]. However, none have resulted in the broad range of tumor subtypes we report, and therefore, single-cell tracking of how subtypes arise had remained unexplored.

Although our primary interest has been in defining whether MSI2[+] cells can serve as a cell of origin, the data have provided some clues about whether these tumors arise from MSI[+] ductal or acinar cells. A vast majority of MSI2[+] cells are duct cells and *Kras* activation in the *Msi2-Cre^{ERT2}* model failed to induce pancreatic tumors, which strongly suggests that we predominantly targeted duct cells that are known to be highly resistant to transformation by mutant *Kras*[17,43]. Additionally, all PDACs in our model demonstrated significantly enriched AGR2 signature, which has been associated with a ductal origin[13]. However, because MSI2 is expressed in rare acinar cells, and because expression of a *Myc* transgene in acinar cells under the control of the elastase promoter induced acinar cell carcinoma along with some ductal neoplasia (though not squamous or anaplastic to our knowledge)[42], we cannot rule out that rare MSI2[+] acinar cells may also serve as a cell-of-origin in this model.

The *Msi2-Myc* mice may also serve as a model for pancreatic carcinomas with mixed phenotypes including acinar-ductal adenocarcinoma and PDAC-ASCP. Such mixed carcinomas are considered aggressive and are being reported at an increasing prevalence[44–46], especially with the advent of single-cell sequencing strategies[47]. In addition, observations that some tumors shift fates after therapy may be consistent with our finding that common tumor precursors persist as an ancestral pool in PDAC patient samples and raise the possibility they may adopt a distinct fate in response to epigenetic and genetic events.

The *Msi2-Myc* model is particularly important because pancreatic cancer is currently treated equivalently, and subtype-specific treatments have not been defined. Given recent data showing that a higher fraction of PDACs are admixed with adenosquamous cells[10,11] than previously realized, effectively modeling this heterogeneity has become more urgent. In this context, the *Msi2-Myc* mice present a unique, timely, and important model to better understand this disease as mapping the transcriptomic and microenvironmental landscape of ASCP may allow defining key dependencies and synthetic lethal screens. A combination of single-cell transcriptomics with functional genomics allowed us to identify HMMR as a dependency in *Msi2-Myc* induced tumor cells. Using both mouse and human pancreatic adenosquamous cancer cell lines, we directly demonstrated that *HMMR* inhibition impaired the growth of these tumors *in vitro* as well as *in vivo*. Consistent with this, HMMR was required for the propagation of patient-derived xenografts *in vivo*, providing strong evidence that *HMMR* is required for the growth of aggressive adenosquamous carcinoma. This finding aligns with the prior work suggesting a pro-tumorigenic role for HMMR during *Ras*-mediated oncogenesis of fibroblasts in a cell transplant model[48], in neuroendocrine tumors[49] and aggressive fibromatosis[50].

The model we have described here provides an interesting and important complement to another model targeting stem/progenitor cells using the Prominin-1 (CD133) promoter[43,51]; in this model, several mutations (except *Myc*) were activated and while this resulted in

tumors in multiple organs no pancreas tumors arose. The distinct tumor phenotypes in the two models highlight how the nature of oncogene (MYC), as well as the cell in which it is targeted (MSI2[+]), are key in directing diverse tumor types. While MSI2[+] cells are stem and progenitor cells across tissues[52–54], whether they harbored the latent capacity to launch tumorigenic programs was unknown. Our data provide evidence that normal MSI2[+] cells are primed for, and highly sensitive to, MYC-driven transformation across organs. To our knowledge, no other reported MYC model has resulted in such a diverse range of tumor types. Thus, in a broader context, this may serve as an important resource for understanding how alterations in MYC drive tumor initiation and progression from progenitor cells and allow the development of strategies for early detection and interception based on the unique vulnerabilities of diverse cancers.

Limitations of the study

The *Msi2-Myc* model presented a unique opportunity to trace the evolution and dependencies of multiple pancreatic cancer subtypes. While we were able to more fully explore dependencies in ASCP, ACC cells were more fragile, and their dependencies were not analyzed in this work but would be of interest in the future. Additionally, our study also included comparisons between the *Msi2-Myc* model and human pancreatic cancer subtypes in order to establish the clinical relevance of our work. Although we were able to make direct comparisons of endpoint ACC and ASCP tumors in the mouse model and patient samples and show the fidelity of the model, there were fewer samples of human precancer lesions and expanding these cohorts would be of interest to strengthen these comparisons.

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to Tannishtha Reya (t.reya@cumc.columbia.edu).

Materials availability—Request regarding reagents should be addressed to Tannishtha Reya (t.reya@cumc.columbia.edu).

Data and code availability—The bulk RNA-seq, scRNA-seq and scATAC-seq datasets have been deposited in the Gene Expression Omnibus (GEO) under the accession numbers GSE181166, GSE241226 (bulk RNA-seq), GSE182396, GSE241231 (scRNA-seq) and GSE181408 (scATAC-seq). Exome-seq datasets have been deposited in the Read Archive (SRA) with accession number SRP455227. This study analyzes existing, publicly available data and the accession numbers for the dataset are listed in the key resources table. All deposited data are publicly available as of the date of publication.

This paper does not report the original code. Any additional information required to reanalyze the data reported in this paper is available upon request to the lead contact.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cells—Mouse pancreatic cancer cell lines were generated from established end-stage tumors (10–12 weeks after tamoxifen induction) in *Msi2-Myc* mice. Tumors were aseptically isolated and dissociated into the single-cell suspension as mentioned above.

Dissociated single cells were then either plated in three-dimension (3-D) in organoid culture conditions in PancreaCult organoid media (Stem Cell Technologies, Catalog #06040) or plated in 2D in 1x DMEM containing 10% FBS, 1x Pen/Strep, 1x Glutamine, and 1x non-essential amino acids (GIBCO, Life Technologies). Experiments were performed using cells from low passages (less than 10). All human pancreatic cancer cell lines were purchased from ATCC, except KLM-1, T3M-4 and FG human pancreatic cancer cells (COLO-357). KLM-1 and T3M-4 cells were gifted by RIKEN Cell Bank, Japan. FG cells were provided by Dr. Andrew Lowy; these cells were derived from a PDAC metastasis and have been validated. These cells were maintained in 2D culture in 1x DMEM containing 10% FBS, 1x pen/strep, and 1x nonessential amino acids (NEAA). Cells were periodically tested for the presence of mycoplasma and verified to be negative. Patient-derived xenograft (PDX) tumors and organoids were derived from originally consented PDAC patients in accordance with the Declaration of Helsinki and use was approved by UCSD's IRB; samples were de-identified and no further information on patient status, treatment or otherwise, is available. Dissociated PDX tumors were maintained as organoids in matrigel domes covered with organoid growth media *in vitro*. Briefly, organoid growth media contained advanced DMEM/F12, 10mM HEPES (pH 7.2–7.5), 1X GlutaMAX, 100 ug/mL primocin, 50% Wnt3a conditioned media, 10% R-Spondin1-conditioned media, 1X-B27 supplement, 10mM nicotinamide, 1.25 mM N-acetyl cysteine, 100 ng/mL murine noggin, 50 ng/mL human-EGF, 100 ng/mL human-FGF, 10 nM human gastrin, 500 nM A-83–01.

Mice—*Msi2-Cre^{ERT2}* mice were generated by conventional gene targeting (genOway Inc, France). These mice were only bred into heterozygosity (*Msi2^{CreERT2/+}*). Heterozygous mutant mice were obtained at the expected Mendelian ratio and developed normally through adulthood. Mice of both genders are fertile and lack any phenotypic abnormalities. Generation of *Msi2e^{GFP/+}* reporter mice was previously described[54]. All the GFP-reporter mice used in experiments were heterozygous for the Msi2 allele. *Rosa26-LSL-YFP* mice[33] were previously described. The *Lox-Stop-Lox (LSL)-Myc^{T58A}-IRES-Luciferase* mice were generated as described[26]. The *LSL-Kras G12D* (*Kras^{LSL-G12D/+}*) mouse (B6.129S4-Kras^{tm4Tyj/J}; Stock No. 008179), *Rosa26-LSL-tdTomato* mice (B6.Cg-Gt(ROSA)26Sor^{tm14(CAG-tdTomato)Hze/J}; stock number: 007914) were purchased from Jackson Laboratory[25]. Immuno-compromised NSG (NOD.Cg-PrkdcscidIL2rgtm1Wji/SzJ, Stock No: 005557) mice were purchased from The Jackson Laboratory. All mice were bred and maintained in the specific-pathogen-free animal care facilities at the University of California San Diego. Animals had access to food and water ad libitum and were housed in ventilated cages under controlled temperature and humidity with a 12-hour light-dark cycle. All animal experiments were performed according to protocols approved by the University of California San Diego Institutional Animal Care and Use Committee. Therefore, males and females of each strain were equally used for experimental purposes, and both sexes are represented in all datasets. All mice enrolled in experimental studies were treatment-naive and not previously enrolled in any other experimental study. Only mice heterozygous for *CAG-LSL-Myc^{T58A}* were used for *Msi2-Myc* tumor studies. All the genotyping primers are listed in Table S1.

Patient samples—Snap-frozen tumor tissues were received from Department of Pathology, Stanford University and the Biobank of the European Pancreas Centre at the Department of General and Visceral Surgery of the University Hospital (EPZ-Pancobank), working in accordance with the regulations of the tissue banks and upon an approval of the Ethics Committee of Heidelberg University (Ethic votes S-708/2019 and S-083/2021). Collection & processing of the samples and clinical data by the EPZ- Pancobank was supported by the Heidelberger Stiftung Chirurgie and in part by the German Ministry of Science and Education (BMBF) grants 01ZX1305C, 01ZX1605C, 01KT1506.

METHOD DETAILS

***In vivo* tamoxifen delivery**—Fresh tamoxifen (Sigma T-5648) stock was prepared by dissolving tamoxifen in corn oil (Sigma C8267) at 20 mg/ml concentration. For Cre-reporter labeling and lineage tracing experiments, adult mice were injected intraperitoneally with a single dose of 2 mg per 20 g body weight. MYC induction in *Msi2-Cre^{ERT2}* mice was initiated at 6–8 weeks of age with 3 doses of tamoxifen (2 mg per 20 g of body weight) delivered over 5 days. The first day of tamoxifen injection was considered as day-0 (d0).

Tissue dissociation, cell isolation, and FACS analysis—All mouse pancreatic tumors were aseptically taken out from tumor-bearing mice, washed with cold 1x PBS (GIBCO, Life Technologies), and minced into 1–2 mm pieces using scalpel blades immediately following resection. Tumor pieces were collected into a 50 ml Falcon tube containing 10 mL Gey's balanced salt solution (GBSS, Sigma, #G9779), 5 mg Collagenase P (Roche, #11249002001), and 0.2 μg DNase I (Roche, #4716728001). Samples were incubated for 20 to 30 minutes at 37°C, then pipetted and returned to 37°C for 15 minutes. Samples were re-pipetted then passed through a 70 μm nylon mesh (Corning). Red blood cells were lysed using RBC lysis buffer (eBioscience, #00–4333-57), and remaining tumor cells were washed, resuspended in GBSS (GIBCO, Life Technologies) containing 2.5% FBS and 2 mM EDTA for staining, FACS analysis, and cell sorting.

For dissociation of the normal and pre-cancerous pancreas, pancreata were dissociated using a modification of a protocol previously developed by *Assi* and colleagues[77]. In brief, pancreata were minced with scissors and collected into 50 ml Falcon tubes containing 10 ml of EGTA buffer with 0.6 mg/ml Collagenase P. Samples were incubated for 15 minutes in a rotor at 37°C. Samples were washed once with cold PBS without mixing the samples and centrifuged at 500g for 3 minutes. Pellets were then further dissociated using Calcium Buffer with 0.6 mg/ml of *Collagenase P* and placed in a rotor at 37 C for 20 minutes. At the end of incubation, samples were pipetted, and then neutralized with 10 ml of cold PBS and passed through 70 um nylon mesh. Cells were centrifuged at 500G for 3 minutes, and pellets were processed for FACS staining without the RBC lysis step. For analysis of cell surface markers by flow cytometry, 0.5×10^6 to 1×10^6 cells/ml were resuspended in EGTA buffer containing 0.5% BSA (for the normal or pre-cancerous pancreas) or HBSS buffer containing 2.5% FBS (for tumors) with DNase I (20ug/ml), then stained with respective antibodies in the presence of Fc block. The following fluorochrome-conjugated rat antibodies were used: anti-mouse EpCAM (eBioscience), anti-mouse CD45-PE/Cy7 (eBioscience), anti-mouse

CD31-PE (BD Bioscience). Propidium-iodide (Life Technologies) or Sytox Blue (Life Technologies) was used to stain dead cells.

Human patient-derived xenograft (PDX) tumors were dissociated in a way similar to mouse tumors, except with the addition of 2 mg Pronase (Roche) along with 5 mg Collagenase P and 0.2 μg DNase in 10 ml GBSS. For FACS analysis, human epithelial tumor cells were stained with anti-human EpCAM-PE (eBioscience, catalog # 12–9326-42) mouse antibody for FACS analysis.

Immunofluorescence staining and imaging—All fresh tissues, including normal and cancerous pancreas, were fixed in 4% PFA for at least 24 hours (Fisher Scientific) and paraffin-embedded in the Immunohistochemistry Core at Moores Cancer Center, University of California San Diego using standard protocols. 5 um sections were obtained and deparaffinized in xylene. The paraffin-embedded formalin-fixed (FFPE) human preneoplastic pancreas tissue arrays (TMA 546 and TMA 1146) were generously gifted by Dr. Michael G. Goggins at the Johns Hopkins School of Medicine. The FFPE human pancreatic cancer tissue array (TMA T144A) was purchased from US Biomax, Inc. The FFPE human pancreatic adenosquamous and acinar cell carcinoma sections were generously provided by Dr. Kwun Wah Wen at the University of California San Francisco. For paraffin-embedded mouse and human pancreas tissues, antigen retrieval was performed at 95–100 °C for 40 minutes using 1x Citrate Buffer, pH 6.0 (GeneTex, catalog # GTX30936). Sections from paraffin-embedded tissues were blocked in PBS containing 0.25% Tween-20 (Sigma-Aldrich) and 5% Goat Serum (Fisher Scientific). All incubations with primary antibodies [(GFP (Abcam #ab13970), Amylase (Abcam #ab21156), Amylase (Santa Cruz, #sc-46657), CK19 (Millipore, #MABT293), DBA-biotinylated (Vector Labs #B-1035–5), Musashi-2 (Abcam #ab76148), RFP/tdTomato (Rockland #600–401-379 and ThermoFisher #MA5–15257), E-Cadherin (BD #610182), Calbindin (Abcam #ab82812), MYC (Abcam #ab32072/Y69), p63 (Biolegend #619002), Chromogranin A (Invitrogen #PA5–16685), Synaptophysin (Cell Marque #336R-24), Sox2 (ThermoFisher #14–9811-82), CK5 (Biolegend #905501), CK14 (Invitrogen #MA5–11599 and Biolegend #906004), CCSP (Millipore #07623), and HMMR (LSBio #LS-B7037)] were carried out overnight at 4°C. Incubation with Alexa fluor-conjugated secondary antibodies (Molecular Probes) was performed for 1 hour at room temperature. DAPI (Molecular Probes) was used to detect DNA and images were obtained with a Leica TCS SP5 II Confocal Laser Scanning Microscope (Leica Microsystems).

Cloning of shRNA constructs and lentivirus production—Short hairpin RNA (shRNA) constructs were designed and cloned into the pLV-hU6-EF1a-green vector by Biosettia. All the shRNA target oligo sequences are listed in Table S2. The virus was produced in 293T cells transfected with shRNA constructs along with pRSV/REV, pMDLg/pRRE, and pHCMVG constructs mixed at a 2:1:1:1 proportion in Opti-MEM media with a dropwise addition of X-tremeGENE HP DNA transfection reagent (Roche, Catalog #6366236001). Viral supernatants were collected for two days, filtered through 0.45 μm filter, and then concentrated by ultracentrifugation at 20,000 rpm for 2 hours at 4°C. For infection of human pancreatic cancer cells, viral supernatants were also concentrated using Lenti-X concentrator kit (Takara Bio, catalog #631232) using the manufacturer's protocol.

In vitro matrigel colony formation assay for mouse pancreatic cancer cells—

For *in vitro* colony-forming assay, 300 to 500 of flow-sorted (GFP+) mouse tumor cells transduced with GFP-tagged lentiviral shRNA were mixed in 10 uL of cold liquefied growth factor reduced Matrigel (BD Biosciences, 354230). Cells suspended in Matrigel were plated in the center of the 48-well plate as a spherical dome in triplicate for organoid cultures. After incubation at 37°C for 5 min cells plated in Matrigel were layered with 300 uL of PancreaCult organoid media (Stem Cell Technologies, Catalog #06040). Colonies were counted after 7 to 10 days of plating.

In vitro matrigel colony formation assay for human pancreatic cancer cells—

Human pancreatic cancer cell lines were infected with GFP-tagged shRNAs and sorted after 48 to 72 hours of infection. 300 to 500 sorted tumor cells were resuspended in 50 μL of growth factor reduced Matrigel mixed with 50 μl of freshly prepared PancreaCult™ human organoid media (Stem Cell Technologies, catalog #100–0781) with only 50% growth factor (1:1 ratio) and plated in 96-well ultra-low attachment plates (Costar) in triplicate. After incubation at 37 °C for 5 min, 100 μl media was placed over the Matrigel layer. Colonies were counted after 12 to 14 days of plating.

Orthotopic transplantation of flow-sorted mouse precancer cells for their multilineage tumor-forming potential—

Precancer cells were taken from *Msi2-Myc; R26-LSL-tdTomato* mice after 6–7 weeks of tamoxifen administration, and ascetically dissociated using the above-mentioned protocol. Recombined (tomato+) precancer cells from individually dissociated pancreas were sorted separately at a low pressure (10 psi) using 130 um nozzle. Flow-sorted pure precancer cells were mixed in resuspension media containing sterile DMEM and Matrigel (BD Biosciences) at 1:1 ratio. 50,000 to 100,000 live cells suspended in 50 μL of resuspension media were orthotopically injected into the pancreas of 6- to 8-week-old NOD/SCID recipient mice. All the recipient mice were monitored weekly for their health and tumor growth.

Flank and orthotopic transplantation of tumor cells for dependency studies—

For *in vivo* tumor growth assay, low passaged pancreatic cancer cells were infected with GFP-tagged shRNA encoding lentiviral particles. Positively infected (GFP) cells were sorted between 48 to 72 hours of transduction. 10,000 to 20,000 freshly isolated shRNA-infected tumor cells were resuspended in 50 μL DMEM and mixed with 50 μL Matrigel (BD Biosciences). Cells were injected orthotopically (FG) or subcutaneously into the right flank of 6- to 8-week-old NOD/SCID recipient mice. All the recipient mice were monitored weekly for their health and tumor growth. Subcutaneous tumor dimensions were monitored weekly for 8–10 weeks. Two to three independent transplant experiments were conducted using one shRNA at n = 3 independent tumors per group.

Patient-derived xenograft (PDX) transplantation and dependency analysis *in vivo*—

PDX tumor cells were dissociated as described above. Dissociated 250,000 tumor cells were plated in 96-well ultra-low attachment plates in 200 μL freshly prepared human organoid media with GFP-tagged lentiviral shRNA (MOI = 25) with 8 μg/mL polybrene. Cells were spin-infected at 600G for 1 hour at room temperature followed by overnight

incubation at 37 C. Infected cells were resuspended in 50 μL organoid media after 24 hours of infection. 15 to 20 μL of cells were set aside and replated in a 96-well ultra-low attachment plate; these cells were cultured to 48 h post-transduction and stained with EpCAM-PE for FACS analysis to assess transduction efficiency (%GFP+/EpCAM-PE+) at t = 0. The remaining cells were mixed 1:1 with matrigel and transplanted into the right flank of NSG recipient mice. 8–12 weeks after the transplantation, tumors were dissociated for endpoint analysis. PDX shRNA studies *in vivo* were conducted using one PDX sample run in duplicate using one shRNA. The lentiviral vector encoding EGFP and shRNA targeting human *HMMR* used to infect PDX in our study, pLV[shRNA]-EGFP-U6>[huHMMR], was constructed and packaged by VectorBuilder. The vector ID is VB221013–1385ygq, which can be used to retrieve detailed information about the vector on vectorbuilder.com.

Real-time quantitative PCR analysis—RNA was isolated using RNeasy Plus Mini kits (QIAGEN) and converted to cDNA using Superscript Mix (Quantabio). Quantitative real-time PCR was performed using an iCycler (BioRad) by mixing cDNAs, iQ SYBR Green Supermix (BioRad), and gene-specific primers. All the primers are listed separately in Table S3. All real-time data was normalized to b2-microglobulin (mouse) or GAPDH (human).

RNA extraction and library preparation for bulk RNA sequencing—Tumors from 10–12 weeks old *Msi2-Myc* mice were harvested and dissociated into a single cell suspension as described above. Tumor cells were stained with anti-CD45 and anti-EpCAM antibodies (eBioscience) in the presence of Fc block. 50,000–100,000 CD45-/EpCAM+ cells were sorted, washed with 1x PBS, and the pellet was lysed in RLT-Plus Buffer (Qiagen # 74034) in the presence of β-mercaptoethanol. For RNA extraction from human patient samples, snap-frozen tumor tissues were sliced into 1–2 mm sized pieces and immediately crushed in an Eppendorf tube with a sterile microcentrifuge pestle in the presence of RLT-Plus buffer and β-mercaptoethanol. Total RNA was isolated using the RNeasy Micro Plus kit (Qiagen # 74034). Total RNA was assessed for quality using an Agilent Tapestation, and all samples had RIN >7.0. RNA libraries were generated from >50 to 100 ng of RNA using Illumina's TruSeq Stranded mRNA Sample Prep Kit following the manufacturer's instructions. RNA libraries were multiplexed and sequenced with 75 base pairs (bp) single-end reads or 100 bp paired-end reads to a depth of approximately 30 million reads per sample on an Illumina HiSeq4000 or NovaSeq 6000.

DNA extraction and library preparation for the whole exome-sequencing—Genomic DNAs from each sorted mouse tumor population and control normal CD45+ immune cells sample was isolated using "AllPrep DNA/RNA FFPE Kit" (Qiagen). A total of 300 to 500 ng of extracted genomic DNA was used for library preparation. Library preparation and sequencing for whole-exome analysis was performed in Genewiz Inc's sequencing facility, using established protocols. The prepared libraries were sequenced for 150×2 base paired-end reads on an Illumina HiSeq platform.

Single-cell capture and single-cell RNA sequencing (scRNA-Seq)—Pancreas dissociations were performed using the protocol as described above. For the determination

of heterogeneity in Msi2 expressing cells in the normal adult pancreas at baseline (time, t=0), pancreata from *Msi2eGFP/+* reporter mice at 8 weeks of age were used. All viable GFP+ cells were collected after negative selection for CD45/CD31+ cells. For the analysis of pancreata at mid-point of tumor progression after the onset of MYC expression following tamoxifen delivery, CD45-/EpCAM+/Cre-reporter+ (YFP+) cells were collected at 5 weeks after tamoxifen treatment. Cre-reporter+, epithelial cells were used to identify MYC-expressing cells in *Msi2-CreERT2; LSL-MycT58A; R26-LSL-YFP* mice. In most cases, viable epithelial (EpCAM+, CD45-) cells were sorted from the tumor, while in cases with high tumor cell content, all viable cells were sorted in preparation for single-cell sequencing.

Sorted cells were washed once and resuspended in 1x PBS with 0.05% BSA. Sorted cells were manually counted and viability was assessed using Trypan Blue staining. Up to 16,000 cells were loaded into one well of the 10x chip to produce Gel Bead-in-Emulsions (GEMs) and capture individual cells. We utilized the Chromium Single Cell 3' Reagent Version 1 Kit (10X Genomics). GEMs underwent reverse transcription to barcode RNA before cleanup and cDNA amplification. Libraries were prepared with the Chromium Single Cell 3' Reagent Version 3 Kit. Each sample was sequenced on HiSeq4000 (Illumina) in Rapid Run Mode with paired-end sequencing parameters for an average depth of 50,000 reads per cell.

Single-cell capture for assay for transposase accessible chromatin sequencing (scATAC-seq)—Mouse pancreas or tumors were dissociated and processed for single-cell capture using the protocol described above. In the analysis of pancreata at mid-point of tumor progression (t = 5), CD45-/EpCAM+/Cre-reporter+ (tdTomato+) cells were collected at 5 weeks after tamoxifen treatment. Cre-reporter+, epithelial cells were used to identify MYC-expressing cells in *Msi2-Myc; R26-LSL-tdTomato* mice. Epithelial cells from tumors were sorted based on CD45-/EpCAM+ expression. Captured single cells were then further processed for single-cell ATAC sequencing using 10x Genomics protocol.

Computational analyses of sequencing data

Bulk RNA-seq analysis: The mouse GRCm38 assembly was acquired from the Ensembl FTP site (http://uswest.ensembl.org/info/data/ftp/index.html) and indexed using Kallisto using default settings[78]. Transcript quantifications were performed using 50 bootstrap samples with the remaining options on default settings. The sleuth R package was used for pairwise differential expression using the Wald test, with an FDR <0.05 and a beta value >1[79]. Gene and transcript quantifications were also generated using sleuth. Gene enrichment analysis was performed with differentially expressed genes using the Enrichr software[61] (https://maayanlab.cloud/Enrichr/).

Single-cell RNA sequencing (scRNA-Seq) analysis: Mouse and human single-cell RNA sequencing were aligned to the 10x genomics pre-built mm10 and hg38 references, respectively, and feature-barcode matrices were generated using Cellranger v3 (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger). Secondary analysis was performed using the Seurat v3.1/4.0 R packages [80]. Cells were filtered for a minimum of 500 features, a maximum of 2,500 features, and a mitochondrial percentage less than 10% per cell. Read counts were

normalized using log normalization and 2,000 variable features were identified using a vst selection method. PCA dimensionality reduction was performed, and elbow plots were used to determine dimensionality. Cluster resolutions were adjusted between 0.3–0.6 accordingly to obtain discrete gene signatures among the clusters. Uniform Manifold Approximation (UMAP) was used to render final single-cell composition plots.

The FindAllMarkers function within Seurat was used to identify the top 20 differentially expressed genes, ranked accordingly to average log fold change and an adjusted p-value <0.05. Gene set enrichment of scRNA-seq data was performed using PRESTO v1.0[81] and fgsea v1.1 R packages[82]. The Wilcoxauc function within the presto package was used to develop pre-ranked gene lists, and fgsea was to perform gene set enrichment using 1,000 permutations and the C2 curated gene sets from MSigDB (https://www.gsea-msigdb.org/gsea/msigdb). Significantly enriched gene sets were defined by an FDR <0.05 and an NES >2.5. Metagene analysis was performed with AUCell R package[83]. In mouse cell pre-cancer cells, AUCell was used to determine the percentage of genes, of the top 20 differentially expressed genes identified from t=0, expressed at t=5. The more conservative AUC cutoff selected by AUCell was used, to minimize false positives, to determine the percentage of pre-cancer cells expressing each of the normal cells t=0 signatures. For the analysis in human tumor cells, the conserved 59 genes enriched in mouse pre-cancer cells (across the 3 biological replicates) were projected onto the human tumor cells. Using a global k1 threshold, 218 of the 920 human tumor cells expressed >9 genes of the mouse pre-cancer signature. Single-cell differentiation status was determined using non-normalized read count matrices, generated from the filtered cells described above, with the CytoTRACE R package[32] or the web-based software (https://cytotrace.stanford.edu). Genes correlated with pre-cancer cells (common tumor precursors) were identified using CytoTRACE software.

For trajectory analysis, Monocle v2.90 was used[34,84]. Single-cell matrices files were initially processed, as described above, and then read in as a cell dataset with a lower detection limit of 0.7 and filtered using a minimum expression of 0.5. Dimensionality was determined using elbow plots and reduced using a UMAP reduction method. Unsupervised clustering of the cells was performed, and the rho and delta values were adjusted to the number of clusters previously identified by Seurat[80]. Differential expression analysis according to the clusters was performed and then ranked according to q-value. Dimensionality reduction was performed again using a DDRTree method and cells were ordered along a trajectory. The number of end-point states was not specified in the ordering of the cells.

Genomic alterations from scRNA-seq data were determined using InferCNV[68]. Filtered non-normalized cell count matrices, described above, were used as an input for InferCNV. As no normal pancreas cells were observed in pre-cancer lesions and end-stage tumors, pre-cancer cells were used as reference cells in the InferCNV analysis. Furthermore, the 'common tumor precursors' served as a natural "spike-in" control as pre-cancer cells were found in all scRNA-seq allowing for direct comparison between samples. Similarly, analysis of the human PDAC/ASAP scRNA-seq sample (Figure S5A–C), all immature cell clusters, as determined using the Seurat and CytoTRACE pipelines described above, were used as reference cells in the InferCNV analysis. As a secondary analysis, each of the samples was run with no references cells, using the average across the cells, which yielded similar

results (data not shown). A cutoff of 0.1 and an HMM-based CNV prediction method were used in the InferCNV analysis. The total number of CNVs was determined using InferCNV for each of the individual biological replicates, and total CNVs for each cells were projected on UMAP plots. The statistical significance of conservation between mouse and human amplified genes was determined using a hypergeometric distribution (http://nemates.org/MA/progs/overlap_stats.html) with a genome size of 20,000 genes.

Exome-seq analysis: Mouse exome-sequencing was aligned using BWA v0.7.5 (BWA-mem) to the mouse mm10 assembly[85]. Samtools was used to convert sam files to sorted bam files [73]. Duplicate reads were removed from sorted bam files using Picard tools ("Picard Toolkit." 2019. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/; Broad Institute). Sequenza [70] was used for copy number estimation from processed bam files (Favero et al., 2015, Annals of Oncology). The mm10 assembly fasta file was used to produce the GC wiggle track file. Mouse CD45 sorted cells from an adult *CAG-LSL-Myc^{T58A}* mouse were used as the normal reference when running sequenza-utils bam2seqz to each of the mouse tumors. Post-processing binning and secondary analysis, including normalization, segmentation and cellularity, and ploidy calculations, were performed using default settings. The chromosome positions of exome-seq identified CNVs were used to compare with the chromosome positions of CNVs identified from InferCNV.

Single-cell assay for transposase accessible chromatin sequencing (scATAC-seq) analysis: Mouse single-cell ATAC sequencing was aligned and accessibility counts were generated using Cell Ranger ATAC v2.0.0 with the Cell Ranger provided mm10–1.2.0 mouse assembly. Signac v1.1.1 was used for secondary analysis of mouse scATAC-seq results[86]. Cells were filtered for outliers with the following QC metrics: peak region fragment length of >3,000 and <100,000, reads in peaks >40, blacklist ration <0.025, nucleosome signal >4 and TSS enrichment > 2. The first LSI component was not used, as it captured sequencing depth variation. UMAP non-linear reduction was used and clustering was performed using the SLM algorithm at a resolution of 0.2–1.2. Differentially accessible regions were determined using a logistic regression framework, and normalized coverage maps showing Tn5 frequency insertion events were plotted using *CoveragePlot*.

STRING interactome network analysis: To understand the broader pathways that might be driven by amplified genes identified in mouse ACC and ASCP samples, we used network analysis in *Cytoscape* [87]. Briefly, we used a core set of amplified genes that were conserved across three replicate ACC and ASCP mouse tumors to seed and propagate independent ACC and ASCP networks within the mouse STRING interactome, including only high confidence interactions (>0.8)[88]. The STRING interactome contains known and predicted functional protein-protein interactions. The interactions are assembled from a variety of sources, including genomic context predictions, high throughput lab experiments, and co-expression databases; interaction confidence is based on a weighted combination of all lines of evidence, with higher quality experiments contributing to a higher confidence score (score is out of 1.0 as the highest). To explore neighboring genes, we integrated 66 conserved ACC-amplified genes or 87 conserved ASCP-amplified genes into the STRING interactome and used them to run the diffusion algorithm within Cytoscape. The diffusion algorithm

ranks all nodes within the interactome using a heat value, indicating how closely each node interacts with our set of amplified seed genes. Extracting the top 500 rankings "hot" genes, we assembled a propagated "hot" network for both ACC and ASCP. To identify the larger programs enriched within these networks, we next wanted to cluster closely related nodes and determine their functional annotations. We used the community clustering algorithm in clusterMaker to cluster each network, extracting the largest connected, clustered component for further functional enrichment analysis and visualization. Functional enrichment was run within *Cytoscape* using STRING.

<u>Sequencing data availability:</u> All the datasets generated during and/or analyzed during the current study have been deposited at NCBI. Bulk RNA-Seq, single-cell RNA-Seq, and single-cell ATAC-Seq data deposited in Gene Expression Omnibus (GEO) are available under the accession numbers GSE181166, GSE241226 (Bulk RNA-Seq), GSE182396, GSE241231 (scRNA-Seq) and GSE181408 (scATAC-Seq). Exome Seq datasets have been deposited in the Read Archive (SRA) with accession number SRP455227.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were carried out using GraphPad Prism software version 9.5.1 (GraphPad Software Inc.). For *in vivo* flank transplantation and tumor growth studies, recipient animals within each group were randomly assigned to control or experimental groups. The number of replicates for each *in vitro* and *in vivo* study is indicated in the figure legends. All data are shown as the mean ± SEM. Two-tailed unpaired Student's *t*-tests or one-way analysis of variance (ANOVA) for multiple comparisons were used when appropriate to determine statistical significance (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

DECLARATION OF INTERESTS

T.R. is a founder, and member of the Board of Directors, and holds executive roles at Tiger Hill Therapeutics.

References:

1. Siegel RL, Miller KD, and Jemal A. (2020). Cancer statistics, 2020. CA Cancer J Clin 70, 7–30.10.3322/caac.21590 [PubMed: 31912902]

2. Collisson EA, Bailey P, Chang DK, and Biankin AV (2019). Molecular subtypes of pancreatic cancer. Nat Rev Gastroenterol Hepatol 16, 207–220.10.1038/s41575-019-0109-y [PubMed: 30718832]

3. Kleeff J, Korc M, Apte M, La Vecchia C, Johnson CD, Biankin AV, Neale RE, Tempero M, Tuveson DA, Hruban RH, and Neoptolemos JP (2016). Pancreatic cancer. Nat Rev Dis Primers 2, 16022.10.1038/nrdp.2016.22 [PubMed: 27158978]

4. Boyd CA, Benarroch-Gampel J, Sheffield KM, Cooksley CD, and Riall TS (2012). 415 patients with adenosquamous carcinoma of the pancreas: a population-based analysis of prognosis and survival. J Surg Res 174, 12–19.10.1016/j.jss.2011.06.015 [PubMed: 21816433]

5. Lenkiewicz E, Malasi S, Hogenson TL, Flores LF, Barham W, Phillips WJ, Roesler AS, Chambers KR, Rajbhandari N, Hayashi A, et al. (2020). Genomic and Epigenomic Landscaping Defines New Therapeutic Targets for Adenosquamous Carcinoma of the Pancreas. Cancer Research.10.1158/0008-5472.CAN-20-0078

6. Borazanci E, Millis SZ, Korn R, Han H, Whatcott CJ, Gatalica Z, Barrett MT, Cridebring D, and Von Hoff DD (2015). Adenosquamous carcinoma of the pancreas: Molecular characterization of 23 patients along with a literature review. World J Gastrointest Oncol 7, 132–140.10.4251/wjgo.v7.i9.132 [PubMed: 26380056]

7. La Rosa S, Sessa F, and Capella C. (2015). Acinar Cell Carcinoma of the Pancreas: Overview of Clinicopathologic Features and Insights into the Molecular Pathology. Front Med (Lausanne) 2, 41-41.10.3389/fmed.2015.00041 [PubMed: 26137463]

8. Al-Hader A, Al-Rohil RN, Han H, and Von Hoff D. (2017). Pancreatic acinar cell carcinoma: A review on molecular profiling of patient tumors. World J Gastroenterol 23, 7945–7951.10.3748/wjg.v23.i45.7945 [PubMed: 29259370]

9. Clark CJ, Graham RP, Arun JS, Harmsen WS, and Reid-Lombardo KM (2012). Clinical outcomes for anaplastic pancreatic cancer: a population-based study. J Am Coll Surg 215, 627–634.10.1016/j.jamcollsurg.2012.06.418 [PubMed: 23084492]

10. Hayashi A, Fan J, Chen R, Ho Y. j., Makohon-Moore AP, Lecomte N, Zhong Y, Hong J, Huang J, Sakamoto H, et al. (2020). A unifying paradigm for transcriptional heterogeneity and squamous features in pancreatic ductal adenocarcinoma. Nature Cancer 1, 59–74.10.1038/s43018-019-0010-1 [PubMed: 35118421]

11. Chan-Seng-Yue M, Kim JC, Wilson GW, Ng K, Figueroa EF, O'Kane GM, Connor AA, Denroche RE, Grant RC, McLeod J, et al. (2020). Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. Nat Genet 52, 231–240.10.1038/s41588-019-0566-9 [PubMed: 31932696]

12. Holen KD, Klimstra DS, Hummer A, Gonen M, Conlon K, Brennan M, and Saltz LB (2002). Clinical characteristics and outcomes from an institutional series of acinar cell carcinoma of the pancreas and related tumors. J Clin Oncol 20, 4673–4678.10.1200/jco.2002.02.005 [PubMed: 12488412]

13. Ferreira RMM, Sancho R, Messal HA, Nye E, Spencer-Dene B, Stone RK, Stamp G, Rosewell I, Quaglia A, and Behrens A. (2017). Duct- and Acinar-Derived Pancreatic Ductal Adenocarcinomas Show Distinct Tumor Progression and Marker Expression. Cell Rep 21, 966–978.10.1016/j.celrep.2017.09.093 [PubMed: 29069604]

14. Lee AYL, Dubois CL, Sarai K, Zarei S, Schaeffer DF, Sander M, and Kopp JL (2019). Cell of origin affects tumour development and phenotype in pancreatic ductal adenocarcinoma. Gut 68, 487–498.10.1136/gutjnl-2017-314426 [PubMed: 29363536]

15. Guerra C, Collado M, Navas C, Schuhmacher AJ, Hernández-Porras I, Cañamero M, Rodriguez-Justo M, Serrano M, and Barbacid M. (2011). Pancreatitis-induced inflammation contributes to pancreatic cancer by inhibiting oncogene-induced senescence. Cancer cell 19, 728–739.10.1016/j.ccr.2011.05.011 [PubMed: 21665147]

16. Westphalen CB, Takemoto Y, Tanaka T, Macchini M, Jiang Z, Renz BW, Chen X, Ormanns S, Nagar K, Tailor Y, et al. (2016). Dclk1 Defines Quiescent Pancreatic Progenitors that Promote Injury-Induced Regeneration and Tumorigenesis. Cell Stem Cell 18, 441–455.10.1016/j.stem.2016.03.016 [PubMed: 27058937]

17. Kopp JL, von Figura G, Mayes E, Liu F-F, Dubois CL, Morris J.P.t., Pan FC, Akiyama H, Wright CVE, Jensen K, et al. (2012). Identification of Sox9-dependent acinar-to-ductal reprogramming

as the principal mechanism for initiation of pancreatic ductal adenocarcinoma. Cancer cell 22, 737–750.10.1016/j.ccr.2012.10.025 [PubMed: 23201164]

18. Guerra C, Schuhmacher AJ, Cañamero M, Grippo PJ, Verdaguer L, Pérez-Gallego L, Dubus P, Sandgren EP, and Barbacid M. (2007). Chronic pancreatitis is essential for induction of pancreatic ductal adenocarcinoma by K-Ras oncogenes in adult mice. Cancer Cell 11, 291–302.10.1016/j.ccr.2007.01.012 [PubMed: 17349585]

19. Hayashi A, Hong J, and Iacobuzio-Donahue CA (2021). The pancreatic cancer genome revisited. Nature Reviews Gastroenterology & Hepatology 18, 469–481.10.1038/s41575-021-00463-z

20. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, et al. (2008). Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321, 1801–1806.10.1126/science.1164368

21. Dennaoui R, Shrestha H, and Wagner KU (2021). Models of pancreatic ductal adenocarcinoma. Cancer Metastasis Rev 40, 803–818.10.1007/s10555-021-09989-9 [PubMed: 34491463]

22. Schleger C, Verbeke C, Hildenbrand R, Zentgraf H, and Bleyl U. (2002). c-MYC Activation in Primary and Metastatic Ductal Adenocarcinoma of the Pancreas: Incidence, Mechanisms, and Clinical Significance. Modern Pathology 15, 462–469.10.1038/modpathol.3880547 [PubMed: 11950922]

23. La Rosa S, Bernasconi B, Vanoli A, Sciarra A, Notohara K, Albarello L, Casnedi S, Billo P, Zhang L, Tibiletti MG, and Sessa F. (2018). c-MYC amplification and c-myc protein expression in pancreatic acinar cell carcinomas. New insights into the molecular signature of these rare cancers. Virchows Archiv 473, 435–441.10.1007/s00428-018-2366-5 [PubMed: 29721608]

24. Witkiewicz AK, McMillan EA, Balaji U, Baek G, Lin WC, Mansour J, Mollaee M, Wagner KU, Koduru P, Yopp A, et al. (2015). Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. Nat Commun 6, 6744.10.1038/ncomms7744 [PubMed: 25855536]

25. Madisen L, Zwingman TA, Sunkin SM, Oh SW, Zariwala HA, Gu H, Ng LL, Palmiter RD, Hawrylycz MJ, Jones AR, et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat Neurosci 13, 133–140.10.1038/nn.2467 [PubMed: 20023653]

26. Mollaoglu G, Guthrie MR, Bohm S, Bragelmann J, Can I, Ballieu PM, Marx A, George J, Heinen C, Chalishazar MD, et al. (2017). MYC Drives Progression of Small Cell Lung Cancer to a Variant Neuroendocrine Subtype with Vulnerability to Aurora Kinase Inhibition. Cancer Cell 31, 270–285.10.1016/j.ccell.2016.12.005 [PubMed: 28089889]

27. Sodir NM, Kortlever RM, Barthet VJA, Campos T, Pellegrinet L, Kupczak S, Anastasiou P, Swigart LB, Soucek L, Arends MJ, et al. (2020). MYC Instructs and Maintains Pancreatic Adenocarcinoma Phenotype. Cancer Discovery 10, 588.10.1158/2159-8290.CD-19-0435 [PubMed: 31941709]

28. Farrell AS, Joly MM, Allen-Petersen BL, Worth PJ, Lanciault C, Sauer D, Link J, Pelz C, Heiser LM, Morton JP, et al. (2017). MYC regulates ductal-neuroendocrine lineage plasticity in pancreatic ductal adenocarcinoma associated with poor outcome and chemoresistance. Nature Communications 8, 1728.10.1038/s41467-017-01967-6

29. Wang X, Cunningham M, Zhang X, Tokarz S, Laraway B, Troxell M, and Sears RC (2011). Phosphorylation Regulates c-Myc's Oncogenic Activity in the Mammary Gland. Cancer Research 71, 925.10.1158/0008-5472.CAN-10-1032 [PubMed: 21266350]

30. Link JM, Ota S, Zhou Z-Q, Daniel CJ, Sears RC, and Hurlin PJ (2012). A critical role for Mnt in Myc-driven T-cell proliferation and oncogenesis. Proceedings of the National Academy of Sciences 109, 19685.10.1073/pnas.1206406109

31. Johnson BL, d'Alincourt Salazar M, Mackenzie-Dyck S, D'Apuzzo M, Shih HP, Manuel ER, and Diamond DJ (2019). Desmoplasia and oncogene driven acinar-to-ductal metaplasia are concurrent events during acinar cell-derived pancreatic cancer initiation in young adult mice. PLoS One 14, e0221810.10.1371/journal.pone.0221810

32. Gulati GS, Sikandar SS, Wesche DJ, Manjunath A, Bharadwaj A, Berger MJ, Ilagan F, Kuo AH, Hsieh RW, Cai S, et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. Science 367, 405–411.10.1126/science.aax0249 [PubMed: 31974247]

33. Srinivas S, Watanabe T, Lin CS, William CM, Tanabe Y, Jessell TM, and Costantini F. (2001). Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Dev Biol 1, 4.10.1186/1471-213x-1-4 [PubMed: 11299042]

34. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, and Rinn JL (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32, 381–386.10.1038/nbt.2859 [PubMed: 24658644]

35. Peng J, Sun B-F, Chen C-Y, Zhou J-Y, Chen Y-S, Chen H, Liu L, Huang D, Jiang J, Cui G-S, et al. (2019). Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. Cell Research 29, 725–738.10.1038/s41422-019-0195-y [PubMed: 31273297]

36. Pontén F, Jirström K, and Uhlen M. (2008). The Human Protein Atlas--a tool for pathology. J Pathol 216, 387–393.10.1002/path.2440 [PubMed: 18853439]

37. Shang J, Zhang X, Hou G, and Qi Y. (2022). HMMR potential as a diagnostic and prognostic biomarker of cancer-speculation based on a pan-cancer analysis. Front Surg 9, 998598.10.3389/fsurg.2022.998598 [PubMed: 36704516]

38. Hingorani SR, Wang L, Multani AS, Combs C, Deramaudt TB, Hruban RH, Rustgi AK, Chang S, and Tuveson DA (2005). Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. Cancer Cell 7, 469–483.10.1016/j.ccr.2005.04.023 [PubMed: 15894267]

39. Hingorani SR, Petricoin EF, Maitra A, Rajapakse V, King C, Jacobetz MA, Ross S, Conrads TP, Veenstra TD, Hitt BA, et al. (2003). Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. Cancer Cell 4, 437–450.10.1016/s1535-6108(03)00309-x [PubMed: 14706336]

40. Aguirre AJ, Bardeesy N, Sinha M, Lopez L, Tuveson DA, Horner J, Redston MS, and DePinho RA (2003). Activated Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal adenocarcinoma. Genes Dev 17, 3112–3126.10.1101/gad.1158703 [PubMed: 14681207]

41. Lin WC, Rajbhandari N, Liu C, Sakamoto K, Zhang Q, Triplett AA, Batra SK, Opavsky R, Felsher DW, DiMaio DJ, et al. (2013). Dormant cancer cells contribute to residual disease in a model of reversible pancreatic cancer. Cancer Res 73, 1821–1830.10.1158/0008-5472.CAN-12-2067 [PubMed: 23467612]

42. Sandgren EP, Quaife CJ, Paulovich AG, Palmiter RD, and Brinster RL (1991). Pancreatic tumor pathogenesis reflects the causative genetic lesion. Proc Natl Acad Sci U S A 88, 93–97.10.1073/pnas.88.1.93 [PubMed: 1986386]

43. Zhu L, Finkelstein D, Gao C, Shi L, Wang Y, Lopez-Terrada D, Wang K, Utley S, Pounds S, Neale G, et al. (2016). Multi-organ Mapping of Cancer Risk. Cell 166, 1132–1146 e1137.10.1016/j.cell.2016.07.045 [PubMed: 27565343]

44. Anderson MJ, Kwong CA, Atieh M, and Pappas SG (2016). Mixed acinar-neuroendocrine-ductal carcinoma of the pancreas: a tale of three lineages. BMJ Case Rep 2016.10.1136/bcr-2015-213661

45. Bosman FT, World Health O, and International Agency for Research on, C. (2010). WHO classification of tumours of the digestive system (International Agency for Research on Cancer).

46. Cancer, I.A.f.R.o. (2019). WHO classification of tumours of the digestive system (International Agency for Research on Cancer).

47. Juiz N, Elkaoutari A, Bigonnet M, Gayet O, Roques J, Nicolle R, Iovanna J, and Dusetti N. (2020). Basal-like and Classical cells coexistence in pancreatic cancer revealed by single cell analysis. bioRxiv, 2020.2001.2007.897454.10.1101/2020.01.07.897454

48. Hall CL, Yang B, Yang X, Zhang S, Turley M, Samuel S, Lange LA, Wang C, Curpen GD, Savani RC, et al. (1995). Overexpression of the hyaluronan receptor RHAMM is transforming and is also required for H-ras transformation. Cell 82, 19–26.10.1016/0092-8674(95)90048-9 [PubMed: 7541721]

49. Chen X, Lee SK, Song M, Zhang T, Han MS, Chen YT, Chen Z, Ma X, Tung CH, and Du YN (2021). RHAMM(B)-mediated bifunctional nanotherapy targeting Bcl-xL and mitochondria for pancreatic neuroendocrine tumor treatment. Mol Ther Oncolytics 23, 277–287.10.1016/j.omto.2021.10.002 [PubMed: 34761107]

50. Tolg C, Poon R, Fodde R, Turley EA, and Alman BA (2003). Genetic deletion of receptor for hyaluronan-mediated motility (Rhamm) attenuates the formation of aggressive fibromatosis (desmoid tumor). Oncogene 22, 6873–6882.10.1038/sj.onc.1206811 [PubMed: 14534534]

51. Zhu L, Gibson P, Currle DS, Tong Y, Richardson RJ, Bayazitov IT, Poppleton H, Zakharenko S, Ellison DW, and Gilbertson RJ (2009). Prominin 1 marks intestinal stem cells that are susceptible to neoplastic transformation. Nature 457, 603–607.10.1038/nature07589

52. Fox RG, Park FD, Koechlein CS, Kritzik M, and Reya T. (2015). Musashi signaling in stem cells and cancer. Annu Rev Cell Dev Biol 31, 249–267.10.1146/annurev-cellbio-100814-125446 [PubMed: 26566113]

53. Yousefi M, Li N, Nakauka-Ddamba A, Wang S, Davidow K, Schoenberger J, Yu Z, Jensen ST, Kharas MG, and Lengner CJ (2016). Msi RNA-binding proteins control reserve intestinal stem cell quiescence. J Cell Biol 215, 401–413.10.1083/jcb.201604119 [PubMed: 27799368]

54. Fox RG, Lytle NK, Jaquish DV, Park FD, Ito T, Bajaj J, Koechlein CS, Zimdahl B, Yano M, Kopp J, et al. (2016). Image-based detection and targeting of therapy resistance in pancreatic adenocarcinoma. Nature 534, 407–411.10.1038/nature17988 [PubMed: 27281208]

55. Jakel C, Bergmann F, Toth R, Assenov Y, van der Duin D, Strobel O, Hank T, Kloppel G, Dorrell C, Grompe M, et al. (2017). Genome-wide genetic and epigenetic analyses of pancreatic acinar cell carcinomas reveal aberrations in genome stability. Nat Commun 8, 1323.10.1038/s41467-017-01118-x [PubMed: 29109526]

56. Suntsova M, Gaifullin N, Allina D, Reshetun A, Li X, Mendeleeva L, Surin V, Sergeeva A, Spirin P, Prassolov V, et al. (2019). Atlas of RNA sequencing profiles for normal human tissues. Sci Data 6, 36.10.1038/s41597-019-0043-4 [PubMed: 31015567]

57. Morgan RT, Woods LK, Moore GE, Quinn LA, McGavran L, and Gordon SG (1980). Human cell line (COLO 357) of metastatic pancreatic adenocarcinoma. Int J Cancer 25, 591–598.10.1002/ijc.2910250507 [PubMed: 6989766]

58. Fox RG, Lytle NK, Jaquish DV, Park FD, Ito T, Bajaj J, Koechlein CS, Zimdahl B, Yano M, Kopp JL, et al. (2016). Image-based detection and targeting of therapy resistance in pancreatic adenocarcinoma. Nature 534, 407–411.10.1038/nature17988 [PubMed: 27281208]

59. Bray NL, Pimentel H, Melsted P, and Pachter L. (2016). Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology 34, 525–527.10.1038/nbt.3519

60. Anders S, and Huber W. (2010). Differential expression analysis for sequence count data. Genome Biology 11, R106.10.1186/gb-2010-11-10-r106 [PubMed: 20979621]

61. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 44, W90–97.10.1093/nar/gkw377 [PubMed: 27141961]

62. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nature Communications 8, 14049.10.1038/ncomms14049

63. Satija R, Farrell JA, Gennert D, Schier AF, and Regev A. (2015). Spatial reconstruction of single-cell gene expression data. Nature Biotechnology 33, 495–502.10.1038/nbt.3192

64. Browning BL (2008). PRESTO: rapid calculation of order statistic distributions and multiple-testing adjusted P-values via permutation for one and two-stage genetic association studies. BMC Bioinformatics 9, 309.10.1186/1471-2105-9-309 [PubMed: 18620604]

65. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine J-C, Geurts P, Aerts J, et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nature Methods 14, 1083–1086.10.1038/nmeth.4463 [PubMed: 28991892]

66. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, and Rinn JL (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature Biotechnology 32, 381–386.10.1038/nbt.2859

67. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, and Trapnell C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nature Methods 14, 979–982.10.1038/nmeth.4402 [PubMed: 28825705]

68. Tickle Timothy TI, Georgescu Christophe, Brown Maxwell, Haas Brian. (2019). inferCNV of the Trinity CTAT Project. Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

69. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and Subgroup GPDP (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.10.1093/bioinformatics/btp352 [PubMed: 19505943]

70. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, Szallasi Z, and Eklund AC (2015). Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. Annals of Oncology 26, 64–70.10.1093/annonc/mdu479 [PubMed: 25319062]

71. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13, 2498–2504.10.1101/gr.1239303 [PubMed: 14597658]

72. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. (2014). STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Research 43, D447–D452.10.1093/nar/gku1003 [PubMed: 25352553]

73. Li H, and Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.10.1093/bioinformatics/btp324 [PubMed: 19451168]

74. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, and Mesirov JP (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545–15550.10.1073/pnas.0506580102 [PubMed: 16199517]

75. Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, Olsen BN, Mumbach MR, Pierce SE, Corces MR, et al. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat Biotechnol 37, 925–936.10.1038/s41587-019-0206-z [PubMed: 31375813]

76. Stuart T, Srivastava A, Madad S, Lareau CA, and Satija R. (2021). Single-cell chromatin state analysis with Signac. Nature Methods 18, 1333–1341.10.1038/s41592-021-01282-5 [PubMed: 34725479]

77. Assi M, Dauguet N, and Jacquemin P. (2018). DIE-RNA: A Reproducible Strategy for the Digestion of Normal and Injured Pancreas, Isolation of Pancreatic Cells from Genetically Engineered Mouse Models and Extraction of High Quality RNA. Front Physiol 9, 129.10.3389/fphys.2018.00129 [PubMed: 29535635]

78. Bray NL, Pimentel H, Melsted P, and Pachter L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34, 525–527.10.1038/nbt.3519 [PubMed: 27043002]

79. Pimentel H, Bray NL, Puente S, Melsted P, and Pachter L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods 14, 687–690.10.1038/nmeth.4324 [PubMed: 28581496]

80. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, and Satija R. (2019). Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902.e1821.10.1016/j.cell.2019.05.031 [PubMed: 31178118]

81. Browning BL (2008). PRESTO: Rapid calculation of order statistic distributions and multiple-testing adjusted P-values via permutation for one and two-stage genetic association studies. BMC Bioinformatics 9, 309.10.1186/1471-2105-9-309 [PubMed: 18620604]

82. Korotkevich G, Sukhov V, and Sergushichev A. (2019). Fast gene set enrichment analysis. bioRxiv, 060012.10.1101/060012

83. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine JC, Geurts P, Aerts J, et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat Methods 14, 1083–1086.10.1038/nmeth.4463

84. Qiu X, Hill A, Packer J, Lin D, Ma YA, and Trapnell C. (2017). Single-cell mRNA quantification and differential analysis with Census. Nat Methods 14, 309–315.10.1038/nmeth.4150 [PubMed: 28114287]

85. Li H, and Durbin R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595.10.1093/bioinformatics/btp698 [PubMed: 20080505]

86. Stuart T, Srivastava A, Lareau C, and Satija R. (2020). Multimodal single-cell chromatin analysis with Signac. bioRxiv, 2020.2011.2009.373613.10.1101/2020.11.09.373613

87. Morris JH, Apeltsin L, Newman AM, Baumbach J, Wittkop T, Su G, Bader GD, and Ferrin TE (2011). clusterMaker: a multi-algorithm clustering plugin for Cytoscape. BMC Bioinformatics 12, 436.10.1186/1471-2105-12-436 [PubMed: 22070249]

88. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 45, D362–d368.10.1093/nar/gkw937 [PubMed: 27924014]

Highlights:

- Development of *Msi2-Cre^ERT2* knock-in mouse to express MYC in stem/progenitor cells

- MSI2$^+$ cells are cells of origin for multiple cancers such as lung, brain, pancreas

- *Msi2-Myc* mice form diverse pancreatic cancer subtypes from a common tumor precursor

- Mapping dependencies of adenosquamous tumors identifies HMMR as a potential target
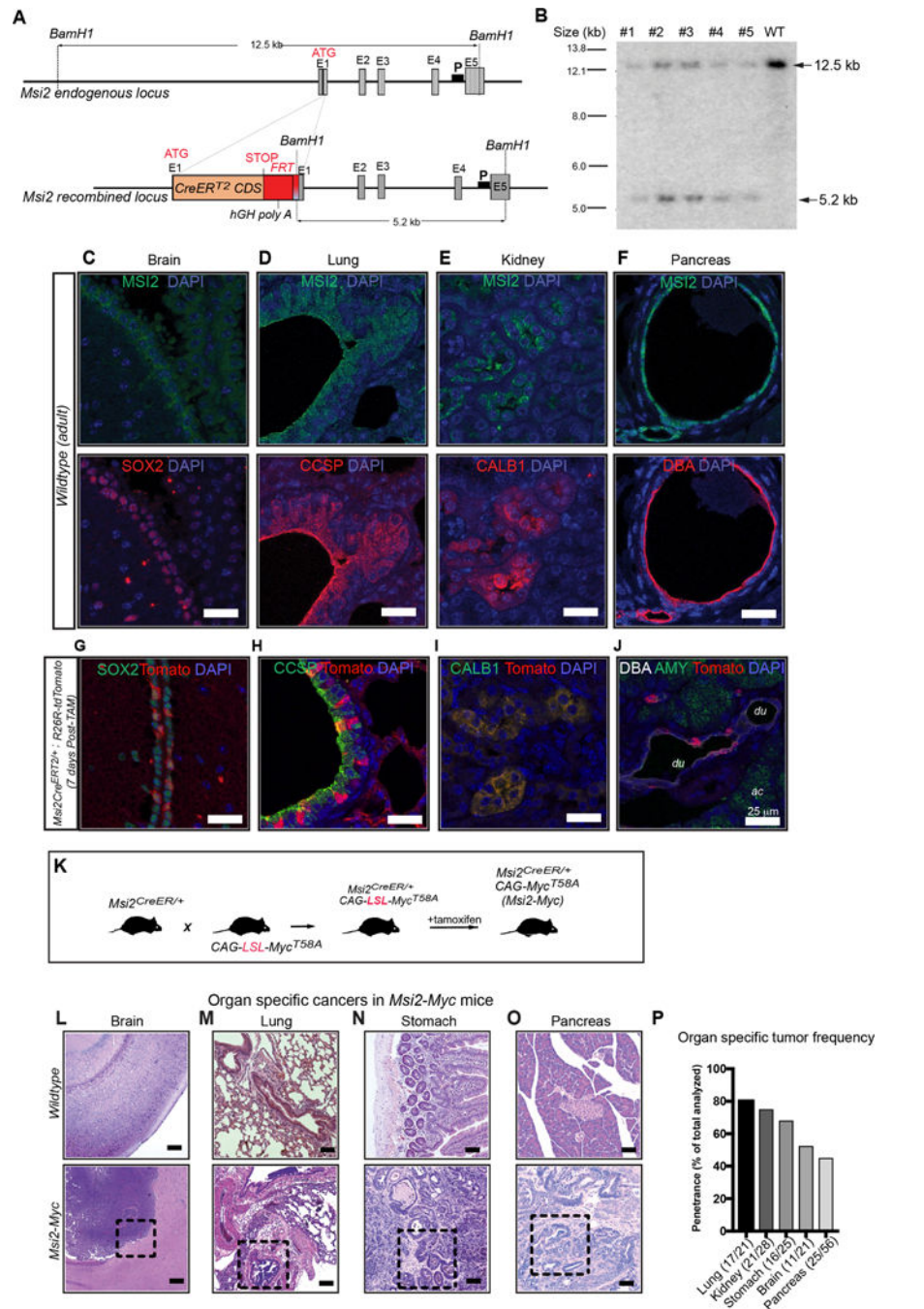
Figure 1: MSI2 expression marks the cell-of-origin for cancers across multiple tissues.

(A) Schematic of wildtype and recombined *Msi2* alleles with CreER^{T2} coding sequence (CDS) cloned into the ATG site of *Msi2*. P= Southern blot probe.

(B) Southern blot of tissues from mice heterozygous for targeted allele (n=5), compared to wildtype (WT).

(C-F) Expression of MSI2 (green, upper panel) and tissue-specific markers (red, lower panel) in brain (SOX2, C), lung (CCSP, D), kidney (CALB1, E), pancreas (DBA, F). n=3, Scale= 25 µm. (G-J) Expression of tdTomato (red, *Msi2*-driven *Cre* activity) in brain (SOX2,

green, G), lung (CCSP, green, H), kidney (CALB1, green, I), pancreas (DBA, white, J; Amylase, green, J; *du* = duct and *ac*= acinar cells) following tamoxifen (n=3). Scale= 25 μm. (K) Scheme of the generation of *Msi2-Myc* mice.

(L-O) Histology of normal tissue from wild-type (upper panels) and tumor tissues from *Msi2-Myc* mice (lower panels). Brain (L), lung (M), stomach (N), pancreas (O). Boxed areas show tumor regions.

(P) Tumor penetrance in *Msi2-Myc* model. Co-occurrence of tumors is common; Scale= 50 μm.
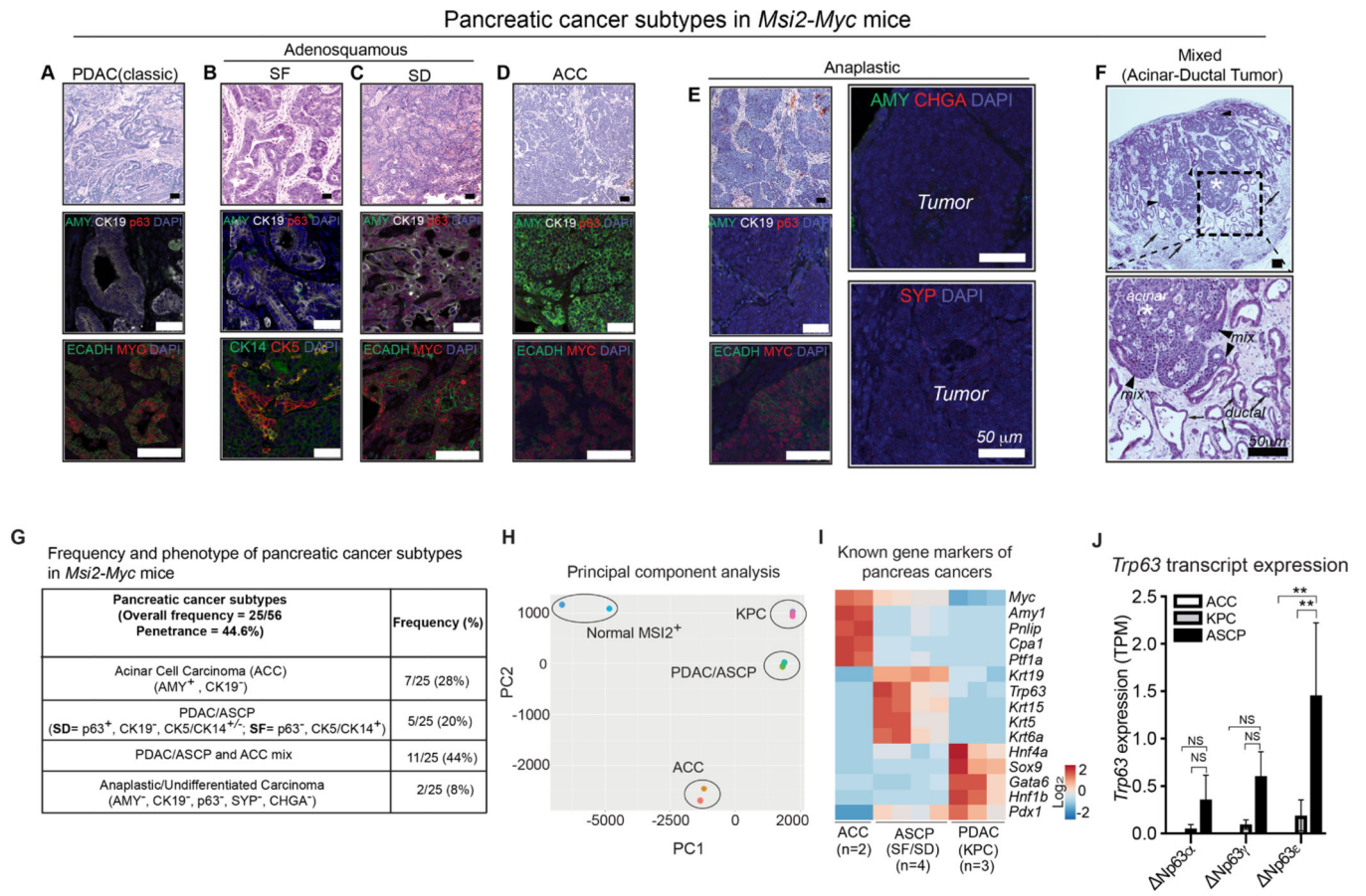
See also Figure S1.

Figure 2: Targeted expression of MYC in MSI2[+] cells drive multiple pancreatic cancer subtypes in Msi2-Myc mice.

(A-E) Characterization of tumor subtypes using lineage-specific markers in PDAC (A), ASCP-SF (B), ASCP-SD (C), ACC (D), and anaplastic (E) tumors in *Msi2-Myc* mice. (E, right) shows the lack of neuroendocrine markers (SYP and CHGA) expression. ECADH = E-Cadherin, CHGA = Chromogranin A, AMY = Amylase, SYP = Synaptophysin. Scale= 50 μm.

(F) Histology of acinar-ductal mixed tumor (top: low magnification; bottom: higher magnification); arrow= Ductal tumor, asterisk (*) = ACC, arrowhead = ACC/PDAC admix. Scale= 50 μm.

(G) Table showing the overall frequency and phenotype of tumor subtypes in *Msi2-Myc* mice.

(H) Principal components analysis of MSI2[+] pancreatic cells (*Msi2eGFP/+*, *t=0*), ACC and PDAC/ASCP tumor cells from *Msi2-Myc* mice and PDAC (KPC)[31] tumor cells.

(I) Heatmap of classification markers in *Msi2-Myc* ACC (n=2 technical) and ASCP (n=2 biological replicates with 2 technical replicates each) and PDAC (KPC)[31] tumors (n=3).

(J) Bar graphs show Np63 (*Trp63*) variant expression in ACC, ASCP, and PDAC (KPC) pancreatic tumors. Data shown as the mean ± SEM; NS=not significant (*p value > 0.05*), **p value < 0.01* using One-way ANOVA with multiple comparisons.
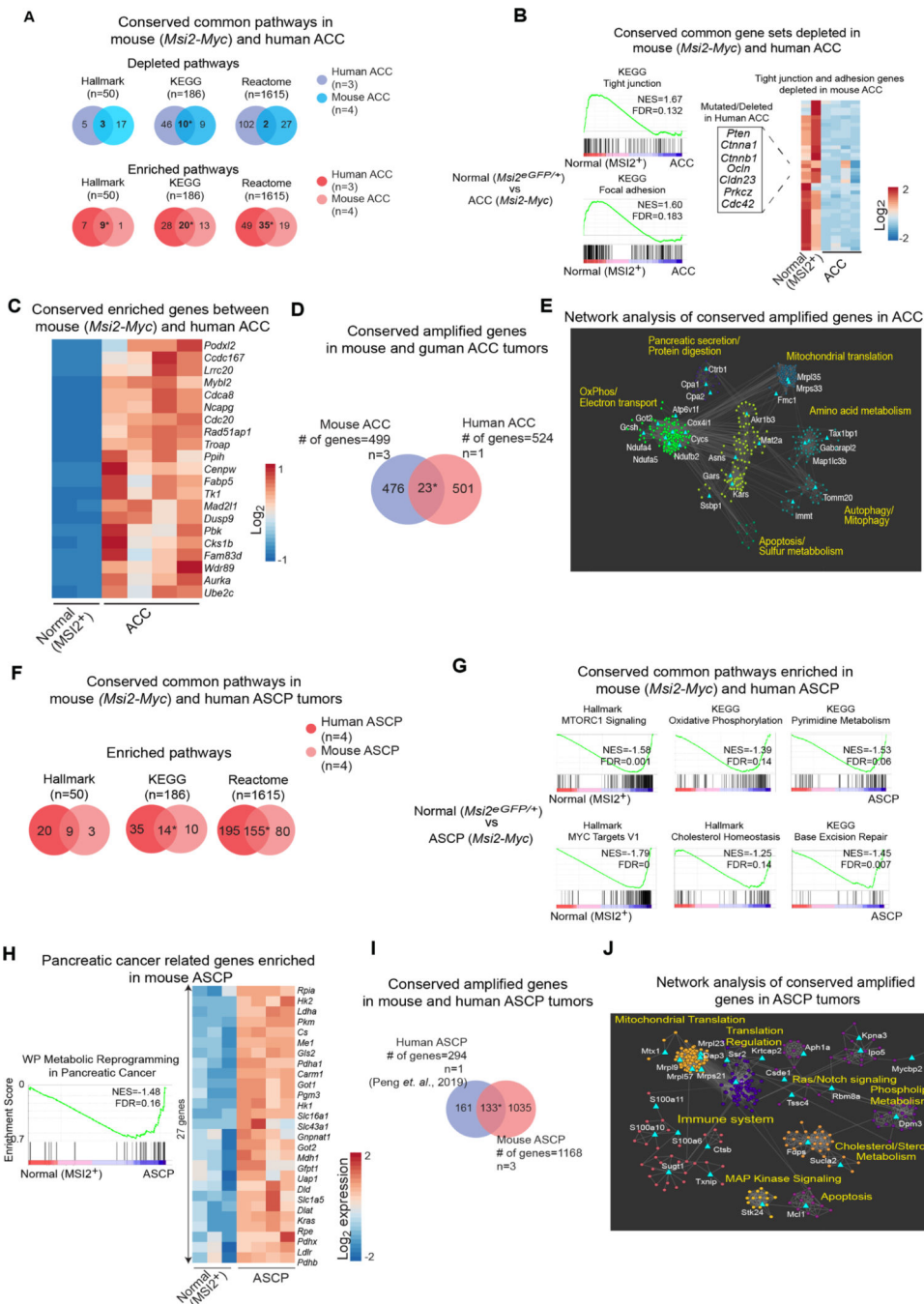
Figure 3: Transcriptomic and genomic alterations are conserved in Msi2-Myc tumors and corresponding human cancers.

(A) Conserved gene sets between ACC tumors in patients and in *Msi2-Myc* model.

(B) GSEA of normal pancreas relative to ACC tumors (FDR<0.25); KEGG pathways shown are depleted in human and mouse ACC tumors. Heatmap shows pathway genes that are mutated/deleted in human ACC and suppressed in mouse ACC.

(C) Heatmap of genes enriched in human and mouse ACC tumors (FDR<5%).

(D) Conserved amplified genes within the predicted amplified genomic regions in human[55] and mouse ACC tumors.

(E) STRING network analysis of shared amplified genes showing top pathways altered in human and mouse ACC tumors.

(F) Conserved gene sets between ASCP in human patients and *Msi2-Myc* mice.

(G) GSEA comparing normal pancreas to ASCP tumors (FDR<0.25); KEGG pathways enriched in human and mouse ASCP.

(H) GSEA shows enriched metabolic programming in pancreatic cancer. Heatmap highlights upregulated *Kras* and *Kras*-related genes in *Msi2-Myc* ASCP.

(I) Conserved amplified genes within the predicted amplified genomic regions in human[35] and mouse ASCP tumors.

(J) STRING network analysis of shared amplified genes showing top pathways altered in human and mouse ASCP tumors.
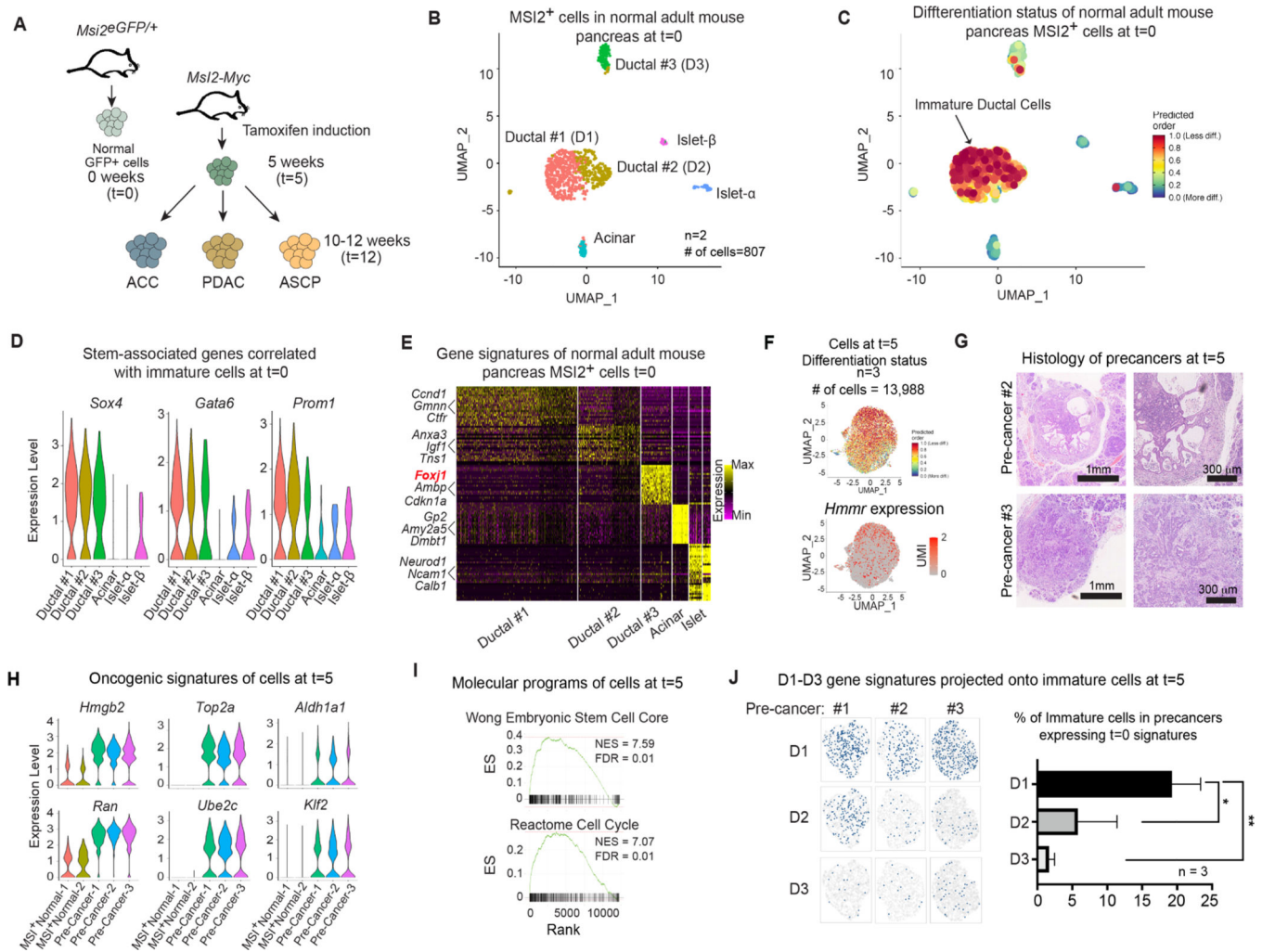
See also Figure S2.

Figure 4: Immature Msi2+ cells are the target of transformation by MYC.

(A) Schematic for scRNA-seq of pancreas cells expressing oncogenic *Myc* (n=2 at *t=0*, n=3 at *t=5*, n=5 at *t=10–12*, biological samples).

(B) UMAP plot showing *Msi2*-expressing clusters from normal pancreas (t=0) (integration plot of n=2 biological samples).

(C) UMAP plot generated by CytoTRACE analysis predicts differentiation status of ductal cell clusters in normal pancreas. D1: most undifferentiated (red); D2: differentiated (yellow/orange); D3: most differentiated (green).

(D) Expression of prominent stem cell-associated genes in Msi2-expressing pancreatic cell clusters.

(E) Heatmap showing top 10 genes enriched in each MSI2-expressing cluster in pancreas; 3 genes within each cluster are highlighted.

(F) UMAP plot showing most undifferentiated population predicted by CytoTRACE (top), identified by *Hmmr* mRNA expression (bottom) in pre-cancer cells from *Msi2-Myc* mice 5-weeks (t=5) post tamoxifen (integration plot of n=3 biological samples shown).

(G) Representative images showing pre-cancer lesions in *Msi2-Myc* mice 5 weeks (t=5) post-tamoxifen (left: lower magnification; right: higher magnification).

(H) Enriched expression of oncogenic and stemness-associated signature genes at t=5.

(I) GSEA map showing enrichment of stem cell and cell cycle genes in undifferentiated pre-cancer cells.

(J) Metagene analysis showing the projection of ductal gene signatures (blue) from normal *Msi2*-expressing pancreatic cells at *t=0* on the undifferentiated clusters from three pre-cancer lesions (left, precancers 1–3), with the D1 signature being most aligned with pre-cancer lesions. Cells above the AUCell-determined thresholds are blue, and cells below the AUCell thresholds are grey (left). Graphs showing percentage of cells expressing indicated gene signatures (right). Data shown as mean ± SEM; *p value < 0.05*, **p value < 0.01* using One-way ANOVA with multiple comparisons.
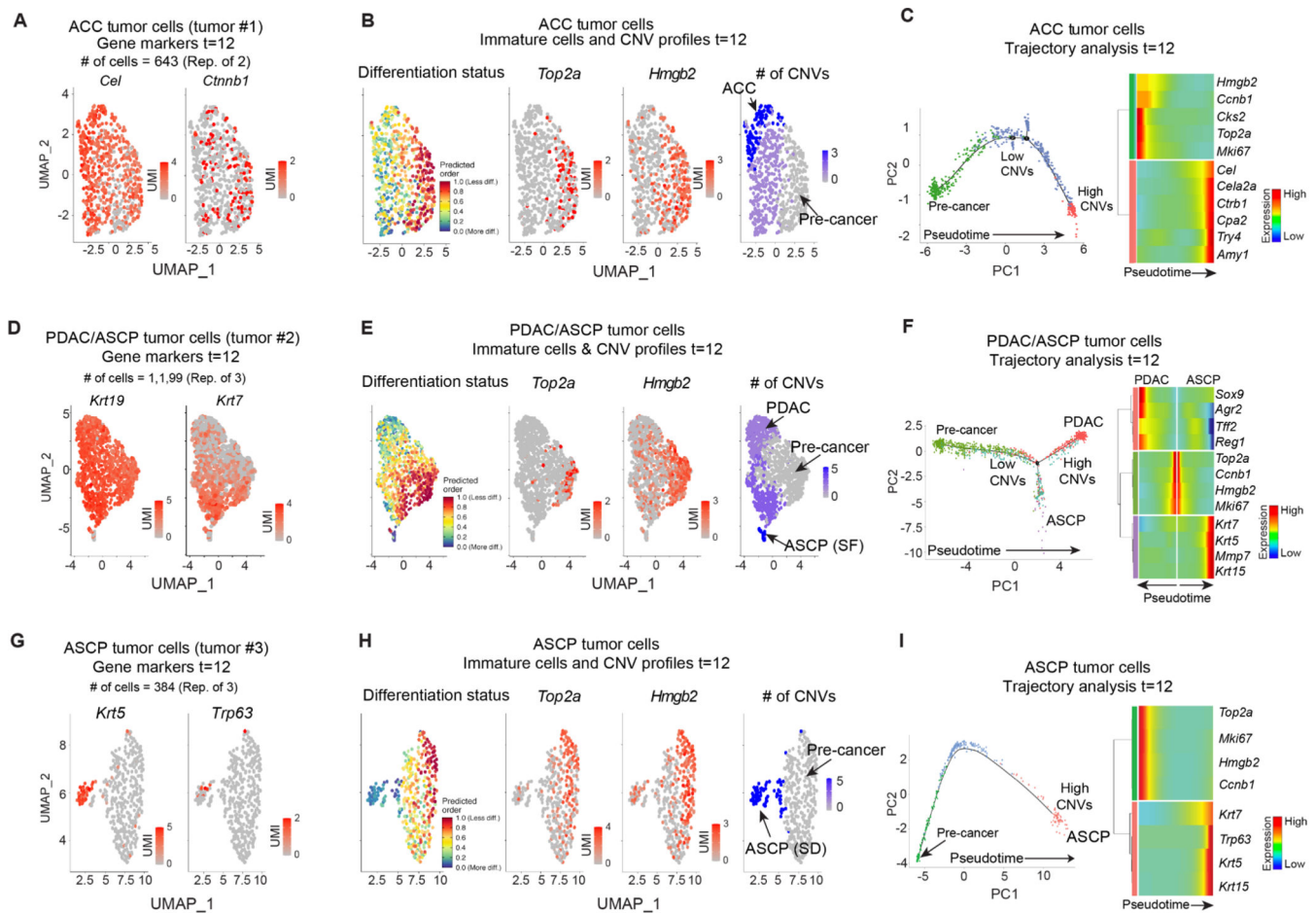
See also Figures S3 and S4.

Figure 5: Tracing the evolution of divergent pancreatic cancer subtypes.

Five independent end-stage tumors were analyzed as biological replicates: n=2, pure ACCs; n=3, mixed ACC-ASCP tumors.

(A) UMAP plot showing expression of common classification markers in a representative ACC tumor at t=12 (n=2).

(B) UMAP plot showing differentiation status and high concordance of the undifferentiated cells (left, red) with *Hmg2b* and *Top2a* expression (middle). Rise in CNVs (purple, right) parallels increasing differentiation in ACC.

(C) Trajectory map showing pseudo-time progression of ACC. *Top2a/Hmg2b* cells at origin (green), most advanced ACC tumor cells (red, left); heatmap showing gene expression profile changes with ACC progression (right).

(D) UMAP plot showing expression of common classification markers in representative PDAC/ASCP(SF) tumor at t=12 (n=3).

(E) UMAP plot showing differentiation status and high concordance of undifferentiated cells (left, red) with *Hmg2b* and *Top2a* expression (middle). Rise in CNVs (right) parallels increasing differentiation in end-stage PDAC/ASCP(SF) tumor.

(F) Trajectory map of a PDAC/ASCP(SF) tumor (left); heatmap showing gene expression profile changes during ASCP(SF) progression (right).

(G) Representative UMAP plot showing expression of common classification markers in a representative ASCP(SD) at t =12 (n=3).

(H) UMAP plot showing differentiation status and high concordance of undifferentiated cells (left) with *Hmg2b* and *Top2a* expression (middle). Rise in CNVs (right) parallels increasing differentiation in ASCP(SD).

(I) Trajectory map of ASCP(SD) tumor (left); heatmap showing gene expression profile changes during ASCP(SD) progression (right).
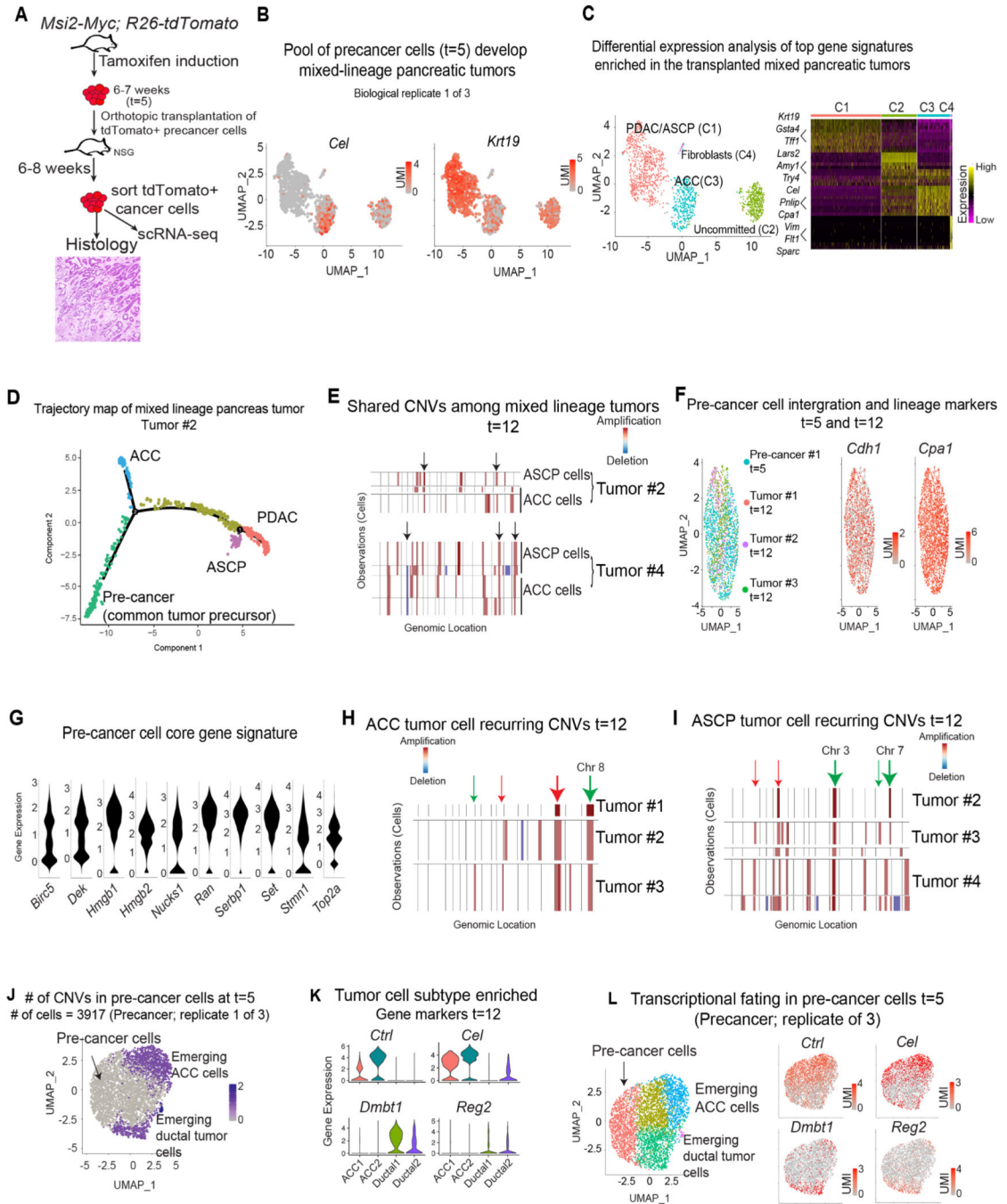
Figure 6: Mixed pancreatic cancer subtypes in Msi2-Myc mice arise from a common pool of uncommitted precancer cells (common tumor precursors).

(A) Schematic of scRNA-seq of pancreatic cells expressing oncogenic MYC (n=3 biological replicates at *t=5*).

(B) Expression of *Cel* (ACC marker) and *Krt19* (ductal marker) demonstrating mixed-lineage pancreatic cancer characteristics (n=3).

(C) Differential expression analysis (Wilcox method using default setting) highlighting gene signatures of identified cell populations with categorical classifications. Three canonical genes commonly enriched in the identified populations are indicated (n=3).

(D) Trajectory map of pseudo-time progression of tri-lineage tumor showing ACC, PDAC, and ASCP tumor cells arising from *Top2a$^+$/Hmg2b$^+$* common pool of pre-cancer cells (representative, n=3).

(E) InferCNV analysis of multi-lineage pancreatic tumors (t=12) showing shared CNV (arrow) signature among different fates (ACC/ASCP) in 2 independent samples, indicating a shared origin.

(F) Integration of scRNA-seq data from *t=5* and *t=12* shows shared immature cell population (referred to as pre-cancer cells). This population shared CNVs and a common gene signature. All three ASCP tumors harbored some ACC tumor cells, while both ACC had no detectable ASCP tumor cells.

(G) Core gene signature of pre-cancer cells (common tumor precursors) across time (*t=5* and *t=12*).

(H) Mouse ACC tumors with highly conserved (n=3, bold arrows) and less frequently conserved (thin arrow) CNVs. Green arrows indicate exome-seq-confirmed CNVs.

(I) Independent mouse ASCP/Squamous tumors with highly conserved (n=3 of 3, bold arrows) and less frequently conserved (n=2 of 3, thin arrows) CNVs. Green arrows indicate exome-seq-confirmed CNVs.

(J) UMAP plot showing emergence of distinct CNV footprints in representative early-stage pre-cancer lesion at t=5 (n=3).

(K) Enriched gene markers observed in pancreatic tumor subtypes, ACC and ASCP (*t=12*; n=2 independent end-stage tumors for each).

(L) UMAP plot of pre-cancer (t=5) showing early transcriptional divergence towards ACC and PDAC/ASCP fates preceding emergence of CNVs.

See also Figures S5 and S6.
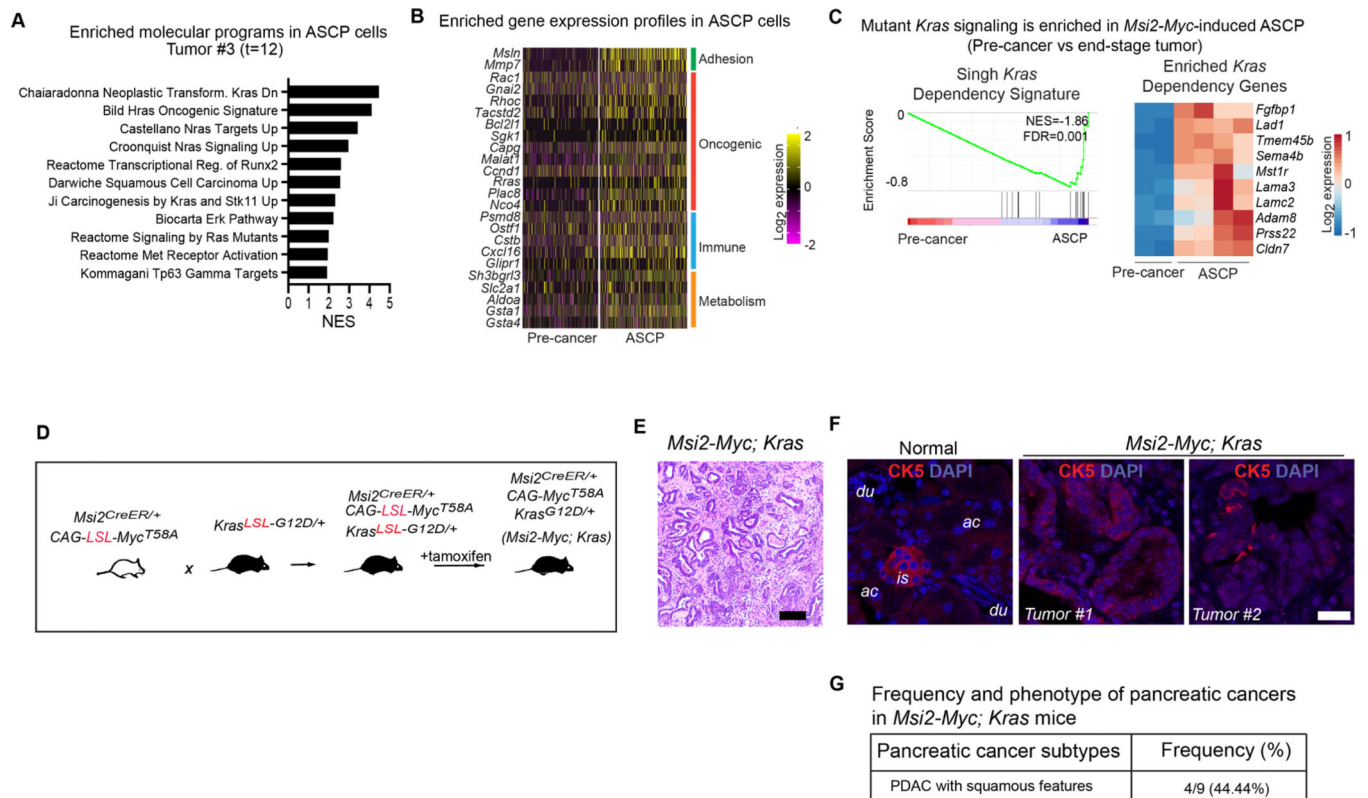
Figure 7: RAS signaling is sufficient to shift the fate of pancreatic cancers towards ductal and adenosquamous lineage in Msi2-Myc model.

(A) GSEA showing top enriched pathways in differentiated mouse ASCP tumor cells (ASCP tumor #3).

(B) Heatmap of gene expression changes as pre-cancer cells (common tumor precursors) progress to end-stage ASCP (ASCP tumor #3).

(C) GSEA showing upregulation of *Kras* and *Kras*-related genes in the *Msi2-Myc* ASCP tumors. Analysis was performed using all t=5 (n=2) and ASCP (n=4) bulk RNA-seq data sets.

(D) Generation of *Msi2-Myc; Kras* mice.

(E) Representative images showing extensive ductal tumors in pancreas of *Msi2-Myc; Kras* mice. Scale= 25 μm

(F) Images showing squamous tumor cells (CK5=red, DAPI=Blue) in 2 independent pancreatic ductal tumors in *Msi2-Myc; Kras* mice. Scale= 25 μm.

(G) Table showing frequency and phenotype of tumor subtype in *Msi2-Myc; Kras* mice.
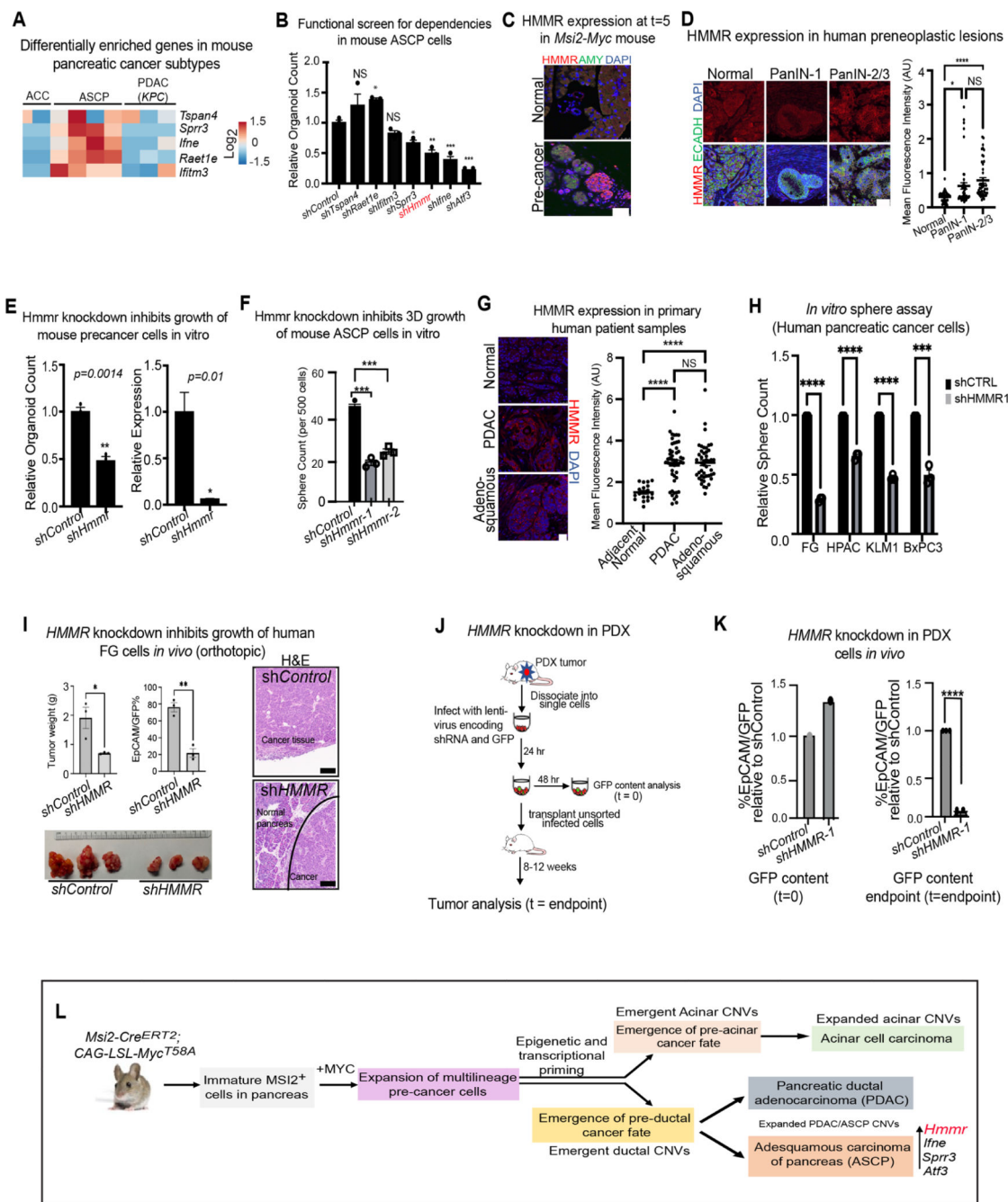
Figure 8: Identification of dependencies of pancreatic adenosquamous cancers.

(A) Heatmap of enriched genes in bulk *Msi2-Myc* ASCP compared to ACC and KPC PDAC.

(B) Impact of shRNA-mediated inhibition of indicated genes on organoid growth of mouse ASCP cells (*n*=2 biological replicates, *n*=3 technical replicates/shRNA. Data represented as mean ± SEM. **p* < 0.05, ***p* < 0.01, and ****p value* < 0.001 by ordinary one-way ANOVA with multiple comparisons).

(C) Increased expression of HMMR in a representative pre-cancer lesion in *Msi2-Myc* mice (n=2 biological replicates). Scale bar = 50 μm.

(D) Representative images (left) and MFI (right) showing increased expression of HMMR in human pancreatic preneoplastic (PanIN) tissue array (n=8 independent tissue spots for normal pancreas, n=9 for PanINs; n=5 frames per case; One-way ANOVA with multiple comparisons, mean ± SEM; scale bar = 50 μm). NS = not significant (*p value > 0.05*), *p value < 0.05*, ****p value < 0.0001*.

(E) Impact of sh*Hmmr* on *Msi2-Myc* preneoplastic organoids t=5 (left, n=3 technical replicates, 1 biological sample). Right: Relative expression of *Hmmr* in shControl and sh*Hmmr* cells. Data represented as mean ± SEM. NS = not significant (*p value > 0.05*), *p value < 0.05*, **p value < 0.01* by unpaired two-sample t-test.

(F) Impact of sh*Hmmr* on mouse ASCP organoid growth (representative, n=2 biological replicates at n=3; One-way ANOVA with multiple comparisons, mean ± SEM; ***p value < 0.001*).

(G) Representative images (left) and dotplot showing MFI for HMMR expression in human pancreatic cancer tissue array (n=3 for adjacent normal pancreas, n=10 for PDAC, n= 9 for ASCP, n=5 or more frames per case; data represented as mean ± SEM, One-way ANOVA with multiple comparisons). Scale bar = 50 μm.

(H) Impact of sh*Hmmr* in human ASCP sphere formation (n=2 biological, n=3 technical replicates per biological; ANOVA with multiple comparisons, mean ± SEM. ***p value < 0.001*, ****p value < 0.0001*).

(I) Left: Impact of sh*HMMR* on FG cells *in vivo* [n=3 biological replicates (1 orthotopic/2 flank transplants), n=3 technical per biological, mean ± SEM, unpaired two-sample t-test]. NS=not significant (*p value > 0.05*), *p value < 0.05*, ***p value < 0.001*.

Right: Representative staining showing end-stage tumors from orthotopically transplanted sh*Control* (top) and sh*HMMR* (bottom) transduced FG cells. Solid-line separates tumor and adjacent normal pancreas. Scale bar = 100 μm.

(J) Schematic of *in vivo* HMMR knockdown of patient-derived xenograft (PDX) tumor cells.

(K) Impact of sh*HMMR* on patient-derived xenograft growth. Tumors analyzed EpCAM+GFP+ shRNA tumor cells at t=0 (left) and end-stage (right) (representative, n=1 patient-derived xenograft, n=2 independent experiments, n=12 mice; unpaired two-sample t-test, mean ± SEM). NS= Not significant (*p value > 0.05*), *p value < 0.05*, **p value < 0.01*, ***p value < 0.001*, ****p value < 0.0001*.

(L) Model of pancreatic cancer subtype initiation and evolution in *Msi2-Myc* tumor model. See also Figure S7.

KEY RESOURCE TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
GFP	Abcam	Cat# ab13970, RRID: AB_300798
Amylase	Abcam	Cat# ab21156, RRID: AB_300798
Amylase	Santa Cruz (G-10)	Cat# sc-46657, RRID: AB_626668
CK19	Millipore	Cat# TROMA-III / MABT293, RRID: AB_2892523
Msi2	Abcam	Cat# ab76148, RRID: AB_1523981
RFP/mcherry/tdTomato	Rockland	Cat# 600-401-379, RRID: AB_2209751
RFP/mcherry/tdTomato	ThermoFisher	Cat# MA5-15257, RRID: AB_10999796
Ecadherin	BD	Cat# 610182, RRID: AB_397581
Calbindin	Abcam	Cat# ab82812, RRID: AB_1658451
c-MYC	Abcam	Cat# ab32072 (Y69), RRID: AB_731658
Sox9	Millipore	Cat# ab5535, RRID: AB_2239761
EpCAM	Abcam	Cat# ab71916, RRID: AB_1603782
DBA-Biotinylated	Vector	Cat# B-1035, RRID: AB_2314288
p63(deltaN)	Biolegend	Cat# 619002, RRID: AB_2207170
Chromogranin A	Invitrogen	Cat# PA5-16685, RRID: AB_11002183
CK5	Biolegend	Cat# 905501, RRID: AB_2565050
CK14	Invitrogen	Cat# MA5-11599, RRID: AB_10982092
CCSP (CC10)	Millipore Sigma	Cat# 07-623, RRID: AB_310759
Synaptophysin	Cell Marque	Cat# 336R-94, RRID: AB_2857955
Sox2	ThermoFisher	Cat# 14-9811-82, RRID: AB_11219471
Nestin	Millipore Sigma	Cat# MAB353, RRID: AB_94911
HMMR	LS Bio	Cat# LS-B7037, RRID: AB_11189262
anti-mouse EpCAM-PE/Cy7	Thermo Fisher/Invitrogen	Cat# 50-112-9753, RRID: AB_1724047
anti-mouse EpCAM-APC	Thermo Fisher/Invitrogen	Cat# 17-5791-82, RRID: AB_2716944
anti-mouse EpCAM-FITC	Thermo Fisher/Invitrogen	Cat# 11-5791-82, RRID: AB_11151709
anti-mouse CD45-PE/Cy7	Thermo Fisher/Invitrogen	Cat# 25-0451-82, RRID: AB_2734986
anti-mouse CD31-PE	Thermo Fisher/Invitrogen	Cat# 12-0311-82, RRID: AB_465632
anti-mouse PDGFRa-BV421	Biolegend	Cat# 135923, RRID: AB_2814036
anti-human EpCAM-PE	eBioscience (ThermoFisher)	Cat# 12-9326-42, RRID: AB_11044497
Bacterial and virus strains		
Chemically competent OneShot Stbl3 cells	Invitrogen	Cat# C737303
pLV[shRNA]-EGFP-U6>[huHMMR]	VectorBuilder	VB221013-1385ygq
Biological samples		
Patient-derived xenografts (PDX)	Moores Cancer Center, University of California San Diego	NA

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Dnase I	Millipore-Sigma	Cat# 10104159001
Collagenase P	Millipore-Sigma	Cat# 11213873001
Pronase	Millipore-Sigma	Cat# 11459643001
UltraPure™ 0.5M EDTA, pH 8.0	ThermoFisher	Cat# 15575020
Tamoxifen	Millipore-Sigma	Cat# T5648
DPBS (1X)	ThermoFisher	Cat# 14190250
DPBS (10X)	ThermoFisher	Cat# 14200166
Opti-MEM™ I Reduced Serum Medium	ThermoFisher	Cat# 11058021
DMEM, high glucose	ThermoFisher	Cat# 11965118
Trypsin-EDTA (0.05%)	ThermoFisher	Cat# 25300062
Gey's Balanced Salt Solution	Millipore-Sigma	Cat# G9779
eBioscience™ 1X RBC Lysis Buffer	ThermoFisher	Cat# 00-4333-57
Foundation B™ Fetal Bovine Serum	Gemini Bio	Cat# 900-208-500
EGTA Buffer 0.5 M pH 8.0	ThermoFisher	Cat# 50-255-956
Paraformaldehyde Solution, 4% in PBS	ThermoFisher	Cat# J19943.K2
Citrate buffer, 10X pH 6.0	GeneTex	Cat# GTX30936
Tween 20	ThermoFisher	Cat# J20605.AP
Normal Goat Serum Blocking Solution	Vector Laboratories	Cat# S-1000-20
Lenti-X Concentrator	Takara Bio	Cat# 631231
XtremeGENE HP DNA Transfection Reagent	Millipore-Sigma	Cat# 6366546001
Mouse PancreaCult Organoid Media	Stem Cell Technologies	Cat# 06040
Human PancreaCult Organoid Media	Stem Cell Technologies	Cat# 100-0781
Matrigel® Growth Factor Reduced (GFR) Basement Membrane Matrix	Corning	Cat# 354230
Penicillin-Streptomycin	ThermoFisher	Cat# 15140122
MEM Non-Essential Amino Acids Solution (100X)	ThermoFisher	Cat# 11140050
GlutaMAX™ Supplement	ThermoFisher	Cat# 35050061
N-Acetyl-L-cysteine	Millipore-Sigma	Cat# A9165
Nicotinamide	Millipore-Sigma	Cat# N0636
Recombinant Human EGF	Peprotech	Cat# AF-100-15
Human FGF-basic (FGF-2/bFGF) Recombinant Protein	ThermoFisher	Cat# 13256-029
Gastrin I (human)	Tocris	Cat# 3006
A 83-01	Tocris	Cat# 2939
B-27™ Supplement	ThermoFisher	Cat# 17504044
Y-27632 dihydrochloride	Millipore-Sigma	Cat# Y0503
Advanced DMEM/F-12	ThermoFisher	Cat# 12634010
HEPES (1M)	ThermoFisher	Cat# 15630080
TrypLE™ Express Enzyme	ThermoFisher	Cat# 12604021

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recovery™ Cell Culture Freezing Medium	ThermoFisher	Cat# 12648010
Primocin	Invivogen	Cat# ant-pm-1
Recombinant Murine Noggin	Peprotech	Cat# 250-38
Polybrene Infection / Transfection Reagent	Millipore-Sigma	Cat# TR-1003-G
SYTOX™ Blue Dead Cell Stain	ThermoFisher	Cat# S34857
Critical commercial assays		
RNeasy Micro Plus kit	Qiagen	Cat# 74034
Illumina's TruSeq Stranded mRNA Sample Kit	Illumina	Cat# 20020594
Chromium Single Cell 3' GEM Library and Gel Bead Kit v3	10x Genomics	Cat# PN-1000269
Chromium Single Cell ATAC Library and Gel Bead Kit	10x Genomics	Cat# 1000176
RNeasy Plus Mini kits	Qiagen	Cat# 74134
iQ SYBR Green Supermix	BioRad	Cat# 1708882
qScript cDNA Supermix	Quantabio	Cat# 101414-102
AllPrep DNA/RNA FFPE Kit	Qiagen	Cat# 80234
Zippy Plasmid Miniprep kit	Zymo Research	Cat# 11-309
Deposited data		
Data set for bulk RNA-Seq (mouse)	This paper	GEO: GSE181166
Data set for bulk RNA-Seq (human)	This paper	GEO: GSE241226
Data set for single-cell RNA-Seq (mouse)	This paper	GEO: GSE182396
Data set for single-cell RNA-Seq (mouse)	This paper	GEO: GSE241231
Data set for single-cell ATAC-Seq (mouse)	This paper	GEO: GSE181408
Data set for Exome-Seq (mouse)	This paper	SRA: SRP455227
Data set for bulk RNA sequencing of Kras/p53 driven mouse pancreatic cancers (KPC)	Johnson et al., 2019[31]	GEO: GSE111540
Data set for single cell RNA sequencing of human adenosquamous pancreatic cancer	Peng et al., 2019[35]	GSA: CRA001160
Data set for bulk RNA sequencing of normal human pancreas	Suntsova et al., 2019[56]	GEO: GSM3415808
Data set for exome sequencing of human acinar cell carcinoma	Jäkel et al., 2017[55]	EGAS0000100253 3
Experimental models: Cell lines		
Human: HEK293T	ATCC	Cat# ATCC CRT-3216
Human: AsPC1	ATCC	Cat# ATCC CRL-1682
Human: BxPC3	ATCC	Cat# ATCC CRL-1687
Human: FG	Gifted by Dr. Andrew Lowy (Morgan et al., 1980[57])	RRID: CVCL_8196
Human: HPAC	ATCC	Cat# ATCC CRL-2119
Human: KLM1	RIKEN Cell Bank, Japan	RRID: CVCL_5146
Human: MiaPaca2	ATCC	Cat# ATCC CRM-CRL-1420
Human: T3M4	RIKEN Cell Bank, Japan	RRID: CVCL_4056

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Human: Panc1	ATCC	Cat# ATCC CRL-1469
Human: Panc0813	ATCC	Cat# ATCC CRL-2551
Human: HeLa	ATCC	Cat# ATCC CRM-CCL-2)
Mouse: *Msi2-Myc*	This paper	N/A
Experimental models: Organisms/strains		
Msi2CreERT2	This paper	
LSL-Myc(T58A)	JAX	Strain# 035557
Rosa26-LSL-tdTomato	JAX	Strain# 007914
Rosa26-LSL-YFP	JAX	Strain# 006148
REM2 (Msi2eGFP/+) reporter	Fox *et al.* 2016[58]	N/A
LSL-KrasG12D: B6.129S4-Krastm4Tyj/J	JAX	Strain# 008179
NSG; NOD.Cg-Prkdc scid IL2rg tm1Wjl/SzJ	JAX	Strain# 005557
Oligonucleotides		
All Genotyping primer sequences	JAX	Table S1
Genotyping primers for *Msi2CreER*: Msi2CreER-WT: GGA GGA GGG ACG GAG ATC TG	This paper	N/A
Genotyping primers for *Msi2CreER*: Msi2CreER-R: GAG TCG TTG GCG CTG CC	This paper	N/A
Genotyping primers for *Msi2CreER*: Msi2CreER-F: CAT TTG AGT TGC TTG CTT GGC	This paper	N/A
shRNA target sequences	This paper	Table S2
All real-time PCR primer sequences	This paper	Table S3
Real-time PCR primers for mouse *Hmmr*. Forward: ACT CAG GAC AAA CGG ATC CAG	This paper	N/A
Real-time PCR primers for mouse *Hmmr*. Reverse: ACT GCT GCA TTG AGC TTT GC	This paper	N/A
Real-time PCR primers for human *HMMR*: Forward: GAA CGT GGT GCC CAG GAC AG	This paper	N/A
Real-time PCR primers for human *HMMR*: Reverse: TGC AGC ATT TAG CCT TGC TTC CA	This paper	N/A
Recombinant DNA		
pRSV/REV	Addgene	Cat# 12253
pMDLg/pRRE	Addgene	Cat# 12251
pCMV-VSV-G	Addgene	Cat# 8454
pLV-hU6-EF1a-green	Biosettia	SORT-B05
Software and algorithms		
ImageJ Web	National Institute of Health	https://ij.imjoy.io/
Leica LAS AF 1.8.2 software	Leica	https://leica-las-af-lite.software.informer.com/
GraphPad Prism software version 7.0d	Graphpad Software Inc.	https://www.graphpad.com/scientifics oftware/prism/
FlowJo Software, version 10.8.1	Tree Star	https://www.flowjo.com/solutions/flowjo/downloads/v10

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Kallisto version 0.46.0	Bray et al., 2016[59]	https://pachterlab.github.io/kallisto/
DESeq2_1.26.0	Anders and Huber, 2010[60]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Enrichr	Kuleshov et al., 2016[61]	https://maayanlab.cloud/Enrichr/
Cell Ranger version 1.1.0	Zheng et al., 2017[62]	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation
Seurat v3.1 and 4.0	Satija et al., 2015[63]	https://github.com/satijalab/seurat
PRESTO version 1.0	Browning et al., 2008[64]	http://www.stat.auckland.ac.nz/~browning/presto/presto.html
AUCell R package, version 1.9.1	Aibar et a., 2017[65]	https://bioconductor.org/packages/AUCell/
CytoTRACE R, version 0.1.0	Gulati et al., 2020[32]	https://cytotrace.stanford.edu/
Monocle v2.90	Trapnell et al., 2014[66] and Qiu et al., 2017[67]	https://bioconductor.org/packages/monocle/
InferCNV, version 1.2.2	Tickle et al., 2019[68]	https://www.bioconductor.org/packages/infercnv/
Samtools, version 1.10.2	Li et al., 2009[69]	https://sourceforge.net/projects/samtools/files/samtools/1.10.2/
Picard Toolkit, version 1.98	Broad Institute	https://broadinstitute.github.io/picard/
Sequenza, version 3.0.0	Favero et al., 2015[70]	https://sequenzatools.bitbucket.io/
Cytoscape, version 3.7.2	Shannon et al., 2003[71]	https://cytoscape.org/
STRING Mouse Interactome 12.0	Szklarczyk et al., 2014[72]	https://stringdb.org/
Burrow-Wheeler Aligner (BWA) tool, v 0.7.5	Li et al., 2009[73]	https://sourceforge.net/projects/biobwa/
Gene Set Enrichment Analysis 2.2.3	Subramanian et al., 2005[74]	http://software.broadinstitute.org/gsea/index.jsp
Cell Ranger ATAC v2.0.0	Satpathy et al., 2019[75]	https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/installation
Signac v1.1.1	Stuart et al., 2021[76]	https://stuartlab.org/signac
Other		