Aldh1b1 expression defines progenitor cells in the adult pancreas and is required for Kras-induced pancreatic cancer

Ekaterina Mameishvili[1,2], Ioannis Serafimidis[3], Sara Iwaszkiewicz[1,2], Mathias Lesche[1,2,4], Susanne Reinhardt[4], Nora Bölicke[1,7], Maren Büttner[5], Dimitris Stellas[3], Adriana Papadimitropoulou[3], Matthias Szabolcs[6], Konstantinos Anastassiadis[4], Andreas Dahl[4], Fabian Theis[5], Argiris Efstratiadis[3], and Anthony Gavalas[1,2,7,*]

Running title: Adult pancreas progenitor cells and ductal adenocarcinoma

1 Paul Langerhans Institute Dresden (PLID) of Helmholtz Center Munich at the University Clinic Carl Gustav Carus of TU Dresden, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

2 German Centre for Diabetes Research (DZD), Germany

3 Biomedical Research Foundation of the Academy of Athens (BRFAA), Athens, Greece

4 Biotechnology Center (Biotec), TU Dresden, Dresden, Germany

5 Institute of Computational Biology (ICB), Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

6 Department of Pathology and Cell Biology, Columbia University Medical Center, New York, USA

7 Center for Regenerative Therapies Dresden (CRTD), Faculty of Medicine, TU Dresden, Dresden, Germany

* Correspondence: anthony.gavalas@tu-dresden.de

Keywords: adult stem and progenitor cells, pancreas, aldehyde dehydrogenase 1b1, organoids, single cell RNA sequencing, pancreatic ductal adenocarcinoma

ABSTRACT

The presence of progenitor or stem cells in the adult pancreas and their potential involvement in homeostasis and cancer development remain unresolved issues. Here, we show that mouse centroacinar cells can be identified and isolated by virtue of the mitochondrial enzyme Aldh1b1 that they uniquely express. These cells are necessary and sufficient for the formation of self-renewing adult pancreatic organoids in an Aldh1b1-dependent manner. Aldh1b1-expressing centroacinar cells, are largely quiescent, self-renew and, as shown by genetic lineage tracing, contribute to all three pancreatic lineages in the adult organ under homeostatic conditions. Single cell RNA sequencing analysis of these cells identified a novel progenitor cell population, established its molecular signature and determined distinct differentiation pathways to early progenitors. A distinct feature of these progenitor cells is the preferential expression of small GTPases, including Kras, suggesting that they might be susceptible to Kras driven oncogenic transformation. This finding and the overexpression of Aldh1b1 in human and mouse pancreatic cancers, driven by activated Kras, prompted us to examine the involvement of Aldh1b1 in oncogenesis. We demonstrated genetically that ablation of *Aldh1b1* completely abrogates tumor development in a mouse model of $Kras^{G12D}$-induced pancreatic cancer.

SIGNIFICANCE STATEMENT

Diabetes and pancreatic cancer are devastating diseases of the pancreas. The identification of organ-specific adult progenitor/stem cells is important for understanding their origin and designing therapeutic interventions. We identified progenitor cells in the adult mouse pancreas characterized by the expression of the mitochondrial enzyme Aldh1b1. These cells give rise to all three cell types in the organ and single cell gene expression analysis showed that they preferentially express Kras, oncogenic mutations of which account for more than 90% of the cases of human pancreatic cancer. Importantly, Aldh1b1 function is required for tumor development in a pancreatic cancer mouse model, suggesting that these cells play a key role in the disease and potentially constitute a novel therapeutic target for pancreatic cancer.

INTRODUCTION

The adult pancreas can respond to injury or metabolic stimuli to repair damage and regulate cell numbers. Although evidence for the presence of progenitor cells has been obtained under conditions of injury, there is an ongoing debate whether this is a result of the inherent plasticity of terminally differentiated cells or due to the presence of unidentified rare adult pancreatic stem/progenitor cells[1,2]. A related controversy concerns the cell-of-origin of pancreatic ductal adenocarcinoma (PDAC), which is the most lethal among malignancies (5-year survival rate ~6%). In more than 90% of the cases, PDAC is initiated by an oncogenic mutation of the GTPase Kras (Kras*) predominantly in codon 12 (e.g. $Kras^{G12D}$). It is thought that the precancerous lesion originates either in a stem/progenitor cell or in an acinar cell that undergoes acinar-to-ductal metaplasia (ADM), perhaps reinforced by cellular injury[3-8].

Rare acinar cells can be reprogrammed *in vivo* into endocrine cells or induced to differentiate into endocrine cells after streptozotocin-mediated ablation of β-cells or pancreatic duct ligation[9-12]. Other experiments suggested the presence of progenitor cells in the ductal tree. Pancreatic duct ligation induced the appearance of endocrine progenitors near the ducts, but lineage tracing experiments did not identify their origin conclusively[13-16]. Inactivation of the tumor suppressor gene *Fbw7* resulted in the conversion of some duct cells into β-cells[17], while ductal cells have been isolated and expanded clonally *in vitro* as organoid-like structures from human and mouse adult pancreata[18,19]. The outcome of these experiments may be attributed to the plasticity of acinar and ductal cells or, alternatively, to the presence of rare stem/progenitor cells.

Terminal duct/centroacinar cells, hereafter referred to simply as CACs, are rare, small cells with minimal cytoplasm, numerous mitochondria and long cytoplasmic extensions. They are contiguous with ductal cells at the end of the ductules forming a fenestrated lining on the luminal acinar surface[20,21]. In adult zebrafish, CACs generate new β cells following β cell ablation or partial pancreatectomy[22]. In the mouse, they have been prospectively isolated through their high aldhehyde dehydrogenase (Aldh) activity using the Aldefluor reagent that detects several Aldh isoforms[23,24]. These cells have been described as both progenitor-like and tumor initiating cells (TICs) [22,25-27] and it has been shown that they can generate *in vitro* pancreatospheres containing endocrine and exocrine cells[26]. However, the large number of the *Aldh* superfamily genes encoding enzymes with diverse specificities[28] had so far precluded further analysis. Interestingly, the mitochondrial enzyme Aldh1b1 is expressed in all mouse embryonic pancreatic

progenitors, but in the adult organ is confined to rare elongated cells[29]. The number of adult Aldh1b1$^+$ cells is dramatically upregulated upon pancreatic injury of either the acinar compartment, using cerulein-induced pancreatitis, or of the endocrine compartment, using streptozotocin-induced β cell ablation[29]. Importantly, the levels of ALDH1B1 in human pancreatic cancers, as assessed by immunohistochemistry, was found to be ~12-fold higher than normal[30].

RESULTS

Aldh1b1 is a specific molecular marker of centroacinar cells

First, we analyzed in the pancreas of adult mice (8-10 months old unless otherwise stated) the co-expression of Aldh1b1 with pancreas lineage and progenitor markers using a specific Aldh1b1 antibody (figure S1A, B). All these Aldh1b1$^+$ cells are epithelial (figure 1A, S1C) with invariably centroacinar location and morphology (figure 1A-F). The majority of Aldh1b1$^+$ cells were positive for the acinar surface marker PNA (71±12%) and the embryonic pancreas or duct progenitor marker Sox9 (64±11%) (figure 1B, C, S1C). Some Aldh1b1$^+$ cells were also positive for the ductal marker DBA (17±6%) and many (40±14%) were weakly stained for the ductal epithelial progenitor marker Prom1 (figure 1D, E, S1C). Aldh1b1$^+$ cells exhibiting weak amylase expression were also observed (figure 1F, S1D-F). For each marker examined, between 70 and 100 Aldh1b1$^+$ cells per pancreas (number of pancreata examined n=3) were scored. Adult Aldh1b1$^+$ cells were never found in islets (at least 500 islets examined, n=10) and were never positive for the progenitor and endocrine markers Pdx1 and Nkx6.1 (>1000 cells, n=6). There was no co-expression of Aldh1b1 with endothelial (PECAM), mesenchymal (Vimentin) or hematopoietic (CD45) markers (n=3) (figure S1G-I). The co-expression of Aldh1b1 with pancreas progenitor markers, but also with acinar and ductal markers, is consistent with the position of CACs at the interface of acinar and ductal cells and indicated that Aldh1b1$^+$ cells might include uncommitted stem/progenitor cells, as well as differentiating derivatives.

Aldh1b1-expressing cells are necessary and sufficient to generate adult mouse pancreatic organoids

Adult pancreatic organoids were derived from terminal duct fragments[15], but the identity of the cells involved remains elusive. Aldh$^+$ cells reside at the interface of terminal ducts

with the acinar compartment[26] and they also express Aldh1b1, while maintaining features of both ductal and acinar identity.

To examine whether Aldh$^+$ cells of the adult mouse pancreas can give rise to organoids, we FACS-isolated Aldh$^+$ cells using the Aldefluor reagent[23] (figure S2A, B) and placed them in a three-dimensional Matrigel-based culture[18,31]. This medium promotes expansion of progenitors and differs from the medium that allows only minimal expansion and formation of pancreatospheres containing endocrine and acinar cells[26]. Aldh$^+$ cells, which constituted 2.1±0.4% (*n*=12) of the cells in the adult pancreas, readily formed organoids at a seeding density of 500 cells/µl (figure 2A, B). These organoids could be passaged at least 11 times over 60 days. In contrast, the fraction of Aldh$^-$ cells did not form organoids (*n*=4). Similarly to duct-derived organoids[18], those derived from Aldh$^+$ cells contained epithelial E-cadherin$^+$ and Sox9$^+$ cells and were predominantly of ductal (Krt19$^+$) character. The organoids retained strong Aldh1b1 expression and contained Pdx1$^+$ cells (figure 2C-E, S2D). There was no expression of endocrine progenitor (Nkx6.1, Ngn3), terminal endocrine (Ins) or acinar (Ptf1a, Amy) markers. We also evaluated the expression levels of progenitor and terminal differentiation genes at successive organoid passages by qPCR and compared them to expression levels in 14.5dpc embryonic pancreata. The expression levels of *Aldh1b1*, *Sox9*, *Pdx1* and *Nkx6.1* were relatively stable but lower in comparison to the corresponding embryonic values, with the notable exception of *Sox9* (figure S2C). There was no appreciable expression of terminal endocrine or acinar genes (*Ins1, Ptf1a, Amy*).

These data indicated that adult mouse pancreatic organoids are derived exclusively from Aldh$^+$ cells. Considering that not all Aldh$^+$ cells are necessarily Aldh1b1$^+$, we assessed the contribution of Aldh1b1$^+$ cells in organoid formation by generating an *Aldh1b1* lineage-tracing mouse allele. To maintain *Aldh1b1* expression levels, while ensuring specific expression of a tamoxifen-activated Cre recombinase (Cre^{ERT2}), we inserted the Cre^{ERT2} transgene into the *Aldh1b1* locus immediately before the stop codon and preceded by a sequence encoding the self-cleaving peptide P2A. The selection cassette was removed *in vivo* by Flp mediated recombination. This strategy preserved all endogenous sequences, including the single intron of the gene, thus ensuring specific expression of Cre^{ERT2} (figure S2E). Mice carrying the *Aldh1b1^{CreERT2}* allele were then crossed with mice carrying the *ROSA26^{LSLtdTomato}* allele that encodes a Cre-inducible form of a stable red fluorescent GFP variant. Five consecutive doses of tamoxifen to 8-week old *Aldh1b1^{CreERT2}/ROSA26^{LSLtdTomato}* double

heterozygotes resulted in efficient labeling of 81.4±2.8 % of the Aldh1b1$^+$ cells as assessed by immunofluorescence one day later (*n*=3) (figure 2F, G). Tamoxifen treatment did not affect the number of Aldh1b1$^+$ cells. Immunofluorescence and image analysis showed that in 8-week old mice not treated with tamoxifen the percentage of Aldh1b1$^+$ cells was 1.4±0.5% (an average of 16K cells per pancreas scored; *n*=3) whereas the percentage of Aldh1b1$^+$ cells treated with five consecutive doses of tamoxifen was 1.8±0.6% (an average of 8.5K cells per pancreas scored; *n*=3). Similarly to Aldh1b1$^+$ cells, all tdTomato$^+$ cells were epithelial E-Cadherin$^+$ cells (478/479 cells scored; *n*=3) (figure 2H) and their majority (59.5±10%, *n*=3) expressed Hes1, another centroacinar cell marker[14] (figure 2I). Our lineage tracer allowed the occasional unequivocal assignment of centroacinar position (figure 2J, S2H). Similarly to Aldh1b1$^+$ cells, TdTomato$^+$ cells did not express CD45, a hematopoietic cell marker used for the isolation of resident macrophages[32] (448/450 cells scored, *n*=3)(figure S2F). Aldh1a1$^+$ cells, although also centroacinar by location and morphology (figure S2H), were clearly fewer and distinct from tdTomato$^+$ cells (figure S2G). A total of 608 tdTomato$^+$ and 137 Aldh1a1$^+$ cells were scored (n=3) in the same cryosections and no co-expression was detected. This indicated that Aldh1b1$^+$ cells are a subpopulation of centroacinar Aldh$^+$ cells and distinct from Aldh1a1$^+$ cells.

Aldh$^+$/tdTomato$^-$ and tdTomato$^+$ cells were FACS isolated (figure 2K, S2N, O) and placed in organoid culture. TdTomato$^+$ cells readily formed organoids at ≤ 400 cells/µl, which could be passaged at least 7 times, while retaining a similar doubling time (T_d=6.0±1.8 days) to that of Aldh$^+$ cells (T_d=5.3±1.0 days) (figure 2L-O). Time-lapse imaging suggested that single tdTomato$^+$ cells have the capacity to form organoids (Movies S1, S2). In contrast, Aldh$^+$/tdTomato$^-$ cells very rarely formed organoids that could not be passaged, indicating that adult pancreas organoids are derived from *Adlh1b1* expressing cells. Organoids derived from tdTomato$^+$ cells were identical to Aldh$^+$ organoids, as shown by immunofluorescence (*n*=4) (compare figure 2C-E with S2J-M). The expression levels of *Aldh1b1*, *Sox9*, *Pdx1* and *Nkx6.1* were also strikingly similar to Aldh$^+$ organoids and stable at successive passages, while there was no appreciable expression of terminal endocrine or acinar genes (*Ins1, Ptf1a, Amy*) (*n*=3) (compare figure S2C with S2P). *Aldh1b1* transcript levels were significantly lower in organoid cells than in embryonic pancreata. This could reflect an adaptation to culture conditions or a genuine difference in *Aldh1b1* expression between embryonic and adult progenitors. Embryonic progenitors need to quickly downregulate Aldh1b1 levels upon

differentiation[33,34]. This may necessitate lower protein stability and, therefore, correspondingly higher transcript levels.

To evaluate whether a functional *Aldh1b1* gene was necessary for adult pancreatic organoid formation, we generated a Cre conditionally inactivated *Aldh1b1* allele (*Aldh1b1*^{fl}) (figure S2Q). These mice were intercrossed with *ROSA26*^{CreERT2} mice to generate double homozygotes. To examine the consequences of *Aldh1b1* deletion in organoid formation and to ensure timely deletion while in culture we initiated *Aldh1b1* inactivation *in vivo* by tamoxifen administration to 8-week-old *Aldh1b1*$^{fl/l}$*/ROSA26*$^{CreERT2/CreERT2}$ mice. This regime resulted in partial *Aldh1b1* inactivation as seen by the presence of both intact and recombined alleles (figure S2R). Pancreatic Aldh^{+} cells, comparable in numbers to those recovered from untreated animals, were FACS-isolated and placed in organoid culture either in the presence of 1 μM 4-hydroxytamoxifen (4-OHT) to complete *Aldh1b1* inactivation, or in its absence (figure S2P). Exposure to 4-OHT in culture effected the completion of *Aldh1b1* deletion and this resulted in failure of these organoids to grow, in contrast to control organoids that have not been treated with 4-OHT (figure 2Q-R, S2R).

Altogether, these analyses revealed that *Aldh1b1* expressing cells are a subset of centroacinar cells of the adult mouse pancreas, which are necessary and sufficient for the formation of organoids. Importantly, *Aldh1b1* function was required for organoid self-renewal.

Aldh1b1-expressing cells self-renew and contribute to all three pancreatic lineages in homeostasis

Next, we examined whether *Aldh1b1* expressing cells contribute to all three lineages of the adult pancreas *in vivo* under homeostatic conditions. *Aldh1b1*^{CreERT2}/*ROSA26*^{LSLtdTomato} double heterozygous mice were treated with five consecutive doses of tamoxifen and the descendants of pancreatic *Aldh1b1* expressing cells, constitutively expressing tdTomato from the *ROSA26* locus (tdTomato$^+$ cells), were then analyzed at three time points; one day after the end of the labelling period, i.e. at the age of 8 weeks, as well as at the ages of 12 and 24 weeks (figure 3A). During this 16-week period, tdTomato$^+$ cells were found in all three pancreatic lineages at all three time points of analysis (figure 3B-G).

To analyze the contribution of the descendants of *Aldh1b1*-expressing cells in the adult pancreas under homeostatic conditions, we used flow cytometry to assess the numbers of descendants of *Aldh1b1*-expressing cells as tdTomato$^+$, acinar (tdTomato$^+$/PNA$^+$) and duct (tdTomato$^+$/DBA$^+$) cells. In these experiments, an average of 100K single viable cells were analyzed for each marker and time point (*n*=3-5 per marker and time point. This analysis showed a gradual increase of tdTomato$^+$ cells but the numbers of tdTomato$^+$/PNA$^+$ and tdTomato$^+$/DBA$^+$ cells were significantly increased only at the 24-week time point. This suggested that labeled cells first underwent expansion and then differentiation (figure 3H-J, S3A-D). Labeled duct cells were found primarily in terminal ducts, but some were also incorporated in larger ducts (figure 3G, S3E, F). Consistent with the flow cytometry analysis, acinar tdTomato$^+$ cells analyzed by immunofluorescence were most conspicuous at 24 weeks (figure 3E, F).

The contribution of the descendants of *Aldh1b1*-expressing cells to the endocrine lineage was assessed by immunofluorescence and image analysis at all three time points for β cells and at 24 weeks for α and δ cells. On average 50 islets per animal, consisting of a total of 2000-5000 C-peptide$^+$ cells, were imaged and analyzed (n=3-5 per time point). Similarly to the acinar and duct lineages, the increase in β cells was only apparent at 24 weeks (figure 3K, L). Mature tdTomato$^+$ β-cells were also detected within the islets as Ucn3$^+$ cells (figure S3G). To confirm that labelled β cells are derived from progenitors residing outside the islets we scored the percentage of islets containing tdTomato$^+$ β cells for all three time points of the analysis. We found that the number of islets containing tdTomato$^+$ cells increased over time, consistent with a model according to which tdTomato$^+$ β cells originate from progenitor cells outside the islets (figure 3M).

The distribution of tdTomato$^+$ β cells in islets over time was also consistent with such a model resulting in an increase of the number of islets containing single tdTomato$^+$ β cells as well as the increase islets containing more than one labelled β cell (figure S3H).

As expected, differentiation was accompanied by loss of Aldh1b1 expression. Therefore, most DBA$^+$/tdTomato$^+$ cells were Aldh1b1$^-$, whereas tdTomato$^+$ endocrine or acinar cells were always Aldh1b1$^-$ (figure S3E, F). Conversely, Aldh1b1$^+$ cells retained their centroacinar morphology throughout this period. We then asked whether *Aldh1b1*-expressing cells could switch on expression of Aldh1a1. Immunofluorescence analysis of 215 TdTomato$^+$ cells from three different 24 week old animals showed that 214/215 of these cells were Aldh1a1$^-$ firmly establishing that the Aldh1b1$^+$ and Aldh1a1$^+$ centroacinar cells are distinct subpopulations (figure S3M).

We then determined whether Aldh1b1-expressing cells self-renew *in vivo*. Despite the detection of differentiated progeny that were contributed by the Aldh1b1-expressing cells, the percentage of undifferentiated Aldh1b1$^+$ cells, as assessed by immunofluorescence, centroacinar morphology and localization outside of islets or ducts, remained comparable at 8 and 24 weeks of age (1.8±0.6% and 1.4±0.9%, *n*=3 and 4, respectively, an average of 170 Aldh1b1$^+$ cells scored per pancreas) (figure S3I-K). Importantly, the percentage of Aldh1b1$^+$ cells that were also tdTomato$^+$ remained the same between 8 and 24 weeks of age (81.4±2.8% and 84.8±5.2%, *n*=3 and 4, respectively,) (figure S3J, K, O). These results indicated that Aldh1b1$^+$ cells are not replenished from another source but self-renew *in vivo*. To confirm this, we compared the percentage of Aldh1b1$^+$ cells that remain proliferative with that of the general pancreatic cell population at 8 and 24 weeks of age by scoring at least 500K cells per pancreas (*n*=3 for each time point and condition). Immunofluorescence for Ki67 expression, a marker strictly associated with proliferative cells[35], showed that while the percentage of proliferative Aldh1b1$^+$ cells remained stable at 8 and 24 weeks (1.9±0.3% and 1.2±0.3% respectively), the corresponding percentage of the general population was substantially lower (0.2±0.02) at 8 weeks and declined further (0.04±0.01) at 24 weeks (figure 3N).

Considered altogether, these data show that *Aldh1b1*-expressing cells contribute to all three pancreatic lineages under conditions of homeostasis and are maintained as a long term self-renewing progenitor population.

Progenitor cell signature and early differentiation routes

Identification of a stem/progenitor cell population based solely on scRNA-seq results is extremely difficult, particularly in tissues with rare such cells[36,37]. Our FACS analysis indicated that Aldh1b1-expressing adult pancreatic cells represent just 0.5% of the cells in the adult organ. We used the *Aldh1b1*^{CreERT2} lineage tracer in combination with scRNA-seq to establish the molecular signature of these rare cells and infer their early differentiation routes.

TdTomato^{+} cells were FACS isolated one day after the completion of tamoxifen administration (early cells) and after one month (late cells) and their transcriptomes were determined by an adapted Smart-Seq2 protocol for scRNA-seq[38] (figure 4A). This experimental design introduced a key criterion for the identification of the adult progenitor cell population because early cells would contribute primarily, if not exclusively, to the original progenitor population. After initial mapping and computation of unique reads, only cells that passed all quality control criteria were retained. This resulted in a total of 578 tdTomato^{+} cells, of which 164 and 414 originated from the 8 week (early) and the 24 week (late) time points, respectively (figure 4A).

To identify distinct cell groups we used K-means unsupervised clustering (SC3)[39] for early cells and all cells. The results were visualized through the generation of diffusion maps, a nonlinear dimensionality reduction method that strongly reduces noise and is able to represent data from differentiation routes with branching points[40]. Early cells segregated in two groups, one large and one small (figure S4A). Cells in the small group expressed both embryonic pancreas progenitor markers and markers of the endocrine, acinar and ductal lineages indicating that they are differentiating progenitors (DP). In contrast, cells in the large group of progenitors (P) retained primarily expression of only acinar markers (figure S4B, C)(Table S1). To identify the progenitor population we proceeded with the analysis of all cells, isolated at both early and late time points. K-means unsupervised clustering identified five distinct groups (figure S4D), the largest of which contained the large majority (90%) of early cells and a large number (41%) of late cells (compare figure 4B, C with S4E). Considering that early time point cells contribute primarily to the original progenitor population, we conclude that this group constitutes the progenitor (P) cell group, the source of all other cells (figure 4B, C, S4E). Plotting the combined expression of either acinar, ductal, endocrine or embryonic progenitor marker genes (Table S1) on the diffusion map showed that all cells retained some expression of acinar terminal markers. Embryonic pancreas progenitor, endocrine

and ductal markers were all activated in a single cluster, the committed progenitor (CP) cluster (figure 4D-G, C). Apart from progenitor and committed progenitor populations, this analysis established the presence of acinar progenitors (AP1-2) that do not activate other lineage or progenitor markers and the presence of a transient progenitor population (TP). The latter was characterised by the expression of *Dclk1*, a recently identified marker of quiescent pancreatic progenitors[41] (figure 4C, S4F). Consistent with an ongoing, long, differentiation process between the two time points at which cells were collected, the number of expressed terminal differentiation markers for each of the three lineages increased substantially in the late cell population (Table S1).

To infer the early differentiation routes, we implemented pseudotemporal ordering of the cells using diffusion pseudotime (DPT)[42]. DPT orders cells on the diffusion map by comparing their differentiation probabilities toward different cell fates and suggests the most probable differentiation routes including branching points[42]. DPT analysis of all cells suggested two differentiation routes with a single branching point similar to the corresponding pattern in early cells (figure 4H, S4G).

Differential gene expression analysis identified the molecular signatures of the adult pancreas progenitor cells and their early progeny (table S2). Differentially expressed genes include cell surface markers, signal receptors and transcription factors that follow distinct expression patterns prefiguring the transitions between cell subpopulations (figure S4H, I). The information from these findings could be exploited for the formation of *bona fide* pancreatic organoids from adult progenitor cells *in vitro*, for directing adult pancreas regeneration *in vivo* and for studies that would increase our understanding of the pathogenesis of pancreatic disorders. In this regard, we inquired whether this analysis could offer insights into the origin of pancreatic cancer, as Gene Ontology analysis of differentially expressed genes suggested a key role of small GTPases, including Kras, in the progenitor cells (Table S3).

Aldh1b1 is required for the development of mouse PDAC

The scRNA-seq analysis indicated that *Kras* is preferentially expressed in the adult progenitor cells (figure 5A). This was confirmed by immunofluorescence with an antibody against wild-type Kras in sections of adult pancreata labelled and collected as in figure 2F. The results showed that the $Kras^+$ cells are relatively rare (~5%) and confined exclusively to the exocrine compartment exhibiting apparent centroacinar but not acinar morphology. $Kras^+$ cells were particularly enriched in the Aldh1b1 expressing

(tdTomato$^+$) population as 31.7±9.8% of the latter (*n*=3) were Kras$^+$ (figure 5B). This significant degree of Kras/Aldh1b1 co-expression was intriguing considering that ALDH1B1 protein expression was 12-fold higher than normal in human pancreatic cancers[30], while *ALDH1B1* gene expression was 5-fold higher in human PDAC specimens (n=96)[43]. Additionally, there is a nearly 7-fold upregulation of *Aldh1b1* expression (n=5), also reflected in immunofluorescence experiments (figure 5C, D) in our fully validated *Tg^{Pdx1Cre}/Actb^{LSLKras*}* mouse model which faithfully simulates human PDAC[44,45]. This model shows stage-wise, orderly, but rapid, progression to invasive PDAC[44] that was also confirmed during this work (*n*=10). The mice first develop low grade pancreatic intraepithelial neoplasias (PanIN) around postnatal day 10 (P10) and have a median survival time (T_{50}) of 42 days after birth[44] (figure 5E). Pairwise comparisons of microarray gene expression profiles with either the *Tg^{Pdx1Cre}/Kras^{LSLKras*}* model[6] or with human PDAC showed a very high degree of statistically significant overlap[44]. Despite nearly uniform expression of Kras* initiated during pancreas development of the *Tg^{Pdx1Cre}/Actb^{LSLKras*}* mice, Aldh1b1$^+$ cells at P1 were still exceedingly rare, as in control mice (figure S5A, B) indicating that *Aldh1b1* may not be downstream of Kras activation and that the number of Aldh1b1$^+$ cells increases only postnatally, during cancer progression.

We then asked whether Aldh1b1 is required for Kras-driven tumor initiation in the *Tg^{Pdx1Cre}/Actb^{LSLKras*}* mouse model[44,45]. Strikingly, transferring the *Tg^{Pdx1Cre}/Actb^{LSLKras*}* transgenes into the *Aldh1b1* null genetic background[33] completely abrogated cancer development and permitted normal lifespan. From a cohort of twelve animals, six were sacrificed at the different time points indicated on the graph and histological analysis revealed that they were tumor -and PanIN-free. The remaining six animals were healthy and tumor-free until the age of twelve months (figure 5E-H). This occurred despite sustained strong Kras* expression in the pancreata of the *Tg^{Pdx1Cre}/Actb^{LSLKras*}/Aldh1b1^{null}* rescued animals (figure 5I-K), indicating a strong requirement for Aldh1b1 in the development of mouse pancreatic cancer.

We then addressed whether Kras* induction, specifically in the *Aldh1b1* expressing cells, would result in development of PDAC. To this end, we activated the *Actb^{LSLKras*}* transgene using our *Aldh1b1^{CreERT2}* Cre driver. We observed that upon induction of Kras* expression by tamoxifen injection in adult *Aldh1b1^{CreERT2}/Actb^{LSLKras*}/ROSA26^{tdTomato}* triple heterozygotes, mice became cachectic and moribund necessitating their sacrifice after less than fifteen days. This early

euthanasia precluded the detection of PDAC that could have developed at a later time point. PanIN was not yet detectable histologically but this might not be surprising considering that in the *Tg*^{Pdx1Cre}*/Actb*^{LSLKras*} model cells have been exposed to Kras* activity for at least ten days during development and an additional ten days postnatally before the detection of low grade PanIN[40]. On the other hand, the autopsy revealed the presence of extensive intestinal tumors (to be analyzed in detail elsewhere), consistent with the strong Aldh1b1 expression in the intestinal crypts (figure S5C). Importantly, it indicated that Aldh1b1 expressing cells in other tissues may be particularly susceptible to Kras* driven oncogenic transformation. This information should be taken into consideration in future studies.

Our genetic analyses showed that the *Aldh1b1* expressing adult pancreatic progenitors might be a prime target of Kras*-mediated pancreatic oncogenesis and identified Aldh1b1 function as a requirement for mouse pancreatic cancer initiation.

DISCUSSION

Centroacinar cells have been prospectively isolated on the basis of their high Aldh activity using the live Aldh substrate Aldefluor which detects several Aldh isoforms[23,24]. When cultured under specific conditions in suspension these cells generated pancreatospheres containing endocrine and acinar cells raising the possibility that these cells could carry a similar function *in vivo*. Gene expression analysis for the *Aldh1a1, 1a2, 1a3, 1a7* and *8a1* genes suggested that *Aldh1a1* was the main isoform expressed[26]**.** However, this analysis covered only a small number of the 19 murine *Aldh* genes. It did not include *Aldh1b1,* which was at the time unknown[46]. Due to the lack of definitive molecular markers, the *in vivo* potential of these cells has not been addressed since then. In this report, we provide strong evidence that a subpopulation of Aldh^{+} cells, characterized by the expression of *Aldh1b1,* contribute to all three pancreatic lineages, self-renew *in vivo* and *in vitro* and constitute the cells of origin of pancreatic organoids. Characterization of this population and its early descendants by scRNA-seq provided its molecular signature and suggested the preferential expression of Kras in Aldh1b1^{+} centroacinar cells. This was confirmed by immunofluorescence, while subsequent genetic experiments showed an absolute requirement of *Aldh1b1* in the development of Kras^{G12D}-driven pancreatic cancer. We think that, collectively, these findings reconcile a large number of seemingly disparate reports postulating that either ducts or acinar cells contain stem or progenitor cells that may give rise to PanIN and PDAC.

The *Aldh1b1*-expressing progenitor cells appear to be centroacinar cells and this is consistent with many studies which suggested that progenitor cells are of ductal origin[13,47,48]. Aldh$^+$ centroacinar cells, as defined by location and morphology, are a heterogeneous population as indicated by the existence of centroacinar Aldh1a1$^+$ cells that are clearly distinct from Aldh1b1$^+$ cells. Previous studies suggested that Sox9$^+$ or Hes1$^+$ centroacinar cells could be a source of progenitor cells[20,21,26] but lineage tracing did not confirm this hypothesis[14,15]. This could be due to incomplete labelling and the heterogeneity of centroacinar cells described here could be an additional factor. Moreover, our scRNA-Seq analysis showed that *Hes1* and *Sox9* are highly, but variably, expressed in the committed progenitor population, but less so in the progenitor cells themselves (figure S4J, K).

Aldh1b1-expressing cells give rise to cells of all three lineages and self-renew *in vivo* for an extended period of time, while maintaining significantly higher proliferative capacity compared to the rest of the pancreas. On the other hand, and in line with the much lower regenerative capacity of the pancreas compared to that of the intestine, less than 2% of the Aldh1b1$^+$ cells are proliferative as opposed to 90% of the Lgr5$^+$ intestinal stem cells[49]. *Aldh1b1* expressing cells are distributed throughout the pancreas sometimes in proximity to islets or large ducts. Their mode of migration into islets and ducts would be of interest, particularly given that they may constitute cells of origin of pancreatic cancer and therefore their migration mechanism could play a role in metastasis.

Lineage tracing analysis over a period of four months showed a gradual increase in the number of tdTomato$^+$ cells, while the number of differentiated cells increased significantly only at the later time point. Additionally, there was an increase in the number of islets containing labelled β cells. These findings are consistent with a model according to which *Aldh1b1*-expressing cells first expand and then differentiate. The question arises whether they are tripotent progenitors or a heterogeneous population consisting of acinar, endocrine and duct progenitors. The scRNA-seq analysis favors the first hypothesis, considering that progenitors appear all in one cluster. Within this cluster there is some divergence based on whether cells are isolated early or late (figure 4B). At present, however, we cannot distinguish whether this reflects the genuine evolution of this population over time or is due to experimental variation. Collectively, these findings suggest the intriguing possibility that the population of *Aldh1b1*-expressing cells could include *bona fide* adult pancreatic stem cells, but additional molecular markers

and clonal analysis are required to conclusively address this issue.

Interestingly, it was recently shown that among ductal cells only a subset, expressing the cell surface marker CD24, were able to form organoids[50]. We have identified the same marker in the signature of committed progenitor cells (Table S2). Because Aldh$^+$ cells, in already formed human pancreatic organoids, have endocrine potential and have been identified as the self-renewing component of the same organoids[51], it will be important to identify the equivalent cells in the human pancreas.
The cell(s) of origin of pancreatic cancer remain(s) elusive. Since the demonstration that a *Pdx1-Cre* or *Ptf1a-Cre* driven expression of Kras^{G12D} in all embryonic pancreas progenitor cells is sufficient to induce pancreatic cancer[3,6], several Cre drivers have been used to identify adult cell populations that can undergo Kras*-driven oncogenic transformation. The use of acinar specific Cre drivers, such as *Ptf1a-CreER*, *Mist1-Cre* and *Ela-Cre* suggested that a small subset of acinar cells can give rise to pancreatic neoplasia and PDAC[5,7,52]. However, not all acinar cells can undergo oncogenic transformation. Acinar cells are generally resistant to Kras*-driven transformation, possibly because the mature acinar genes *Nr5a2*, *Ptf1a* and *Mist1* are effectively acting as Kras* tumor suppressors [53-55]. Thus, it is interesting that *Aldh1b1*-expressing progenitors express a number of acinar markers but not these tumor suppressor transcription factors (Table S1). Other reports suggested that centroacinar and ductal cells may also give rise to pancreatic neoplasias, particularly after PTEN ablation[25,27,56,57]. Whereas Sox9 is necessary for the development of pancreatic cancer, *Sox9-CreER* driven activation of Kras^{G12D} expression was not sufficient to induce oncogenic transformation of ductal or centroacinar cells[7]. This still leaves open the possibility that Aldh1b1$^+$/Sox9$^-$ centroacinar cells may serve as TICs in the pancreas. A common limitation of these approaches is that, while they reveal the potential of a cell population for neoplastic transformation, they do not show that the cells under consideration actually function as TICs. We propose that Kras$^+$ cells are more likely to undergo oncogenic transformation upon the acquisition of activating mutations in this gene. Our data suggested that Kras-expressing cells have centroacinar morphology and that *Aldh1b1*-expressing cells are enriched in Kras expression making them strong candidates for TICs. Pancreatic cancer risk factors, such as chronic pancreatitis, may induce Kras expression in dedifferentiating acinar cells. Therefore, it is not unlikely that more than one cell type could be susceptible to neoplastic transformation but in such a scenario, Aldh1b1 expression would need to be induced *de novo.*

Survival rates of PDAC patients have not been improved for the last 40 years. This is due to late diagnosis, but because the identified prevalent genetic mutations and genes essential for the early steps in disease progression are notoriously difficult to target[7,58-60]. Therefore, the identification of additional causal or necessary factors is an urgent endeavor. Thus, the association of high ALDH1B1 expression with increased PDAC invasiveness[30] and the upregulation of several ALDH genes that has been detected in all human PDAC samples examined[43] (figure S5D) are noteworthy. Additionally, high ALDH activity has been used as the most efficient means to identify rare tumor-initiating cells[61,62] but an essential role of this enzymatic activity in pancreatic progenitor cells and cancer had not been demonstrated. Small GTPase expression, and Kras expression in particular, is a hallmark of Aldh1b1^{+} cells and our study has revealed for the first time that the mitochondrial Aldh1b1 is necessary for Kras*-induced PDAC in mice. We think, therefore, that the Aldh1b1^{+} cell population, which already expresses Kras and maintains significantly higher proliferative capacity compared to the rest of the pancreas, would be particularly vulnerable to oncogenic transformation through the acquisition of Kras activating mutations. These findings render this population a strong candidate for the cells, or one type of the cells, of origin of pancreatic cancer.

The mechanism by which Aldh1b1 activity enables Kras* mediated oncogenic transformation requires investigation beyond the scope of this study. Aldh1b1 may enhance cell metabolism to promote cell proliferation but we also note that Kras* triggers a tumorigenic pathway that enhances ROS production through the alteration of mitochondrial metabolism [63]. To sustain tumorigenesis, ROS levels should be regulated to remain below toxic levels. Thus, as an example of another potential mechanism, this could be attained through the antioxidant action of ALDHs[64] and, more specifically, Aldh1b1. In the context of these hypotheses, we view the Aldh1b1 function, not as oncogenic *per se*, but as a precondition for Kras*-induced PDAC.

The findings reported here suggest novel possibilities for pancreas regeneration and a potential therapeutic target in pancreatic cancer.

MATERIALS AND METHODS

The *Aldh1b1*$^{tm2(CreERT2)Agav}$ (*Aldh1b1*^{CreERT2}) line was generated by inserting the bi-cistronic *Aldh1b1/P2A/CreERT2* transgene followed by the *FRT-PGK-neo-pA-FRT* selection cassette in the endogenous Aldh1b1 locus (Fig S4). Activation of the CreERT2 was obtained by intraperitoneal (IP) injection for five consecutive days of corn oil dissolved tamoxifen (20 mg/ml; Sigma) as described in the SI Appendix.

For FACS isolation of Aldh$^+$ cells and progeny of Aldh1b1 expressing cells, adult mouse pancreata were dissected, minced into small pieces and digested at 37°C by 1 mg/ml collagenase D (Roche, 11088858001). Aldh$^+$ cells were isolated based on their Aldh activity using the Aldefluor reagent (Aldefluor kit, Stem Cell Technologies 01700), whereas Aldh1b1 expressing cells and their progeny were isolated from *Aldh1b1*^{CreERT2}*/ROSA26*^{LSLtdTomato} double heterozygous mice by virtue of their tdTomato expression.

For the three-dimensional culture of adult pancreas progenitor organoids, FACS isolated Aldh$^+$ or tdTomato$^+$ cells were seeded in growth factor-reduced Matrigel (BO Biosciences, 356231) to a final density of 500 or 400 cells/µl, in a 96-well plate. After gelation of the matrigel, culture medium was added as described in the SI Appendix and replenished every 3 days. Organoids usually formed on the third day after seeding cells in culture and were passaged for the first time after 10-14 days.

Single cell library preparation, RNA sequencing, bioinformatics analysis as well as quantitative flow cytometry, time lapse fluorescent microscopy and other non-specialised methods are described in SI Appendix, along with a list of mouse strains, antibodies, genotyping primers and other biological reagents used.

ACKNOWLEDGMENTS

We thank Heiko Lickert, Mina Gouti, Lori Sussel and Gaetano Gargiulo for critically reading the manuscript and providing comments; Chris Wright, Michael Ray and Tetsuo Sudo for reagents; the CRTD FACS core facility and the MPI-CBG Dresden transgenic core for expert technical assistance; and the personnel of the animal houses at CRTD, Dresden and BRFAA, Athens for help with animal husbandry. Research in AE's laboratory was supported by funds provided to BRFAA by a Krauss Maffei Wegmann GmbH and Co Offset Program. Research in AG's laboratory was supported by grants from the German Center for Diabetes Research (DZD)(grant 82DZD00101), the German Research Foundation (DFG)(SFB 655, CRTD Seed Grant 2014, Project Grant GA

2004/3-1), the General Secretariat for Research and Technology (GSRT, Greece) (ESPA 2011, SYN-1-1496) and a TU Dresden Graduate Academy Bridging Grant to EM.

AUTHOR CONTRIBUTIONS

The study was conceived and designed by EM, IS and AG, experiments were designed, analyzed and interpreted by EM, IS, SI, MS, FT, AE and AG and performed by EM, IS, SI, SR and KA. Bioinformatics analysis was conducted by ML, MB, AD and FT, whereas DS and AP participated in the animal studies. AG wrote the manuscript with contributions from AE. AG supervised the project.

The authors declare no competing interests.

REFERENCES

1 Kopp, J. L., Grompe, M. & Sander, M. Stem cells versus plasticity in liver and pancreas regeneration. *Nature cell biology* **18**, 238-245, doi:10.1038/ncb3309 (2016).

2 Murtaugh, L. C. & Keefe, M. D. Regeneration and repair of the exocrine pancreas. *Annu Rev Physiol* **77**, 229-249, doi:10.1146/annurev-physiol-021014-071727 (2015).

3 Aguirre, A. J. *et al.* Activated Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal adenocarcinoma. *Genes & development* **17**, 3112-3126, doi:10.1101/gad.1158703 (2003).

4 De La, O. J. *et al.* Notch and Kras reprogram pancreatic acinar cells to ductal intraepithelial neoplasia. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 18907-18912, doi:10.1073/pnas.0810111105 (2008).

5 Habbe, N. *et al.* Spontaneous induction of murine pancreatic intraepithelial neoplasia (mPanIN) by acinar cell targeting of oncogenic Kras in adult mice. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 18913-18918, doi:10.1073/pnas.0810097105 (2008).

6 Hingorani, S. R. *et al.* Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer cell* **4**, 437-450 (2003).

7 Kopp, J. L. *et al.* Identification of Sox9-dependent acinar-to-ductal reprogramming as the principal mechanism for initiation of pancreatic ductal adenocarcinoma. *Cancer cell* **22**, 737-750, doi:10.1016/j.ccr.2012.10.025 (2012).

8 Storz, P. Acinar cell plasticity and development of pancreatic ductal adenocarcinoma. *Nat Rev Gastroenterol Hepatol* **14**, 296-304, doi:10.1038/nrgastro.2017.12 (2017).

9 Baeyens, L. *et al.* Transient cytokine treatment induces acinar cell reprogramming and regenerates functional beta cell mass in diabetic mice. *Nature biotechnology* **32**, 76-83, doi:10.1038/nbt.2747 (2014).

10 Cavelti-Weder, C. *et al.* Hyperglycaemia attenuates in vivo reprogramming of pancreatic exocrine cells to beta cells in mice. *Diabetologia* **59**, 522-532, doi:10.1007/s00125-015-3838-7 (2016).

11 Pan, F. C. *et al.* Spatiotemporal patterns of multipotentiality in Ptf1a-expressing cells during pancreas organogenesis and injury-induced facultative restoration. *Development* **140**, 751-764, doi:10.1242/dev.090159 (2013).

12 Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* **455**, 627-632 (2008).

13 Inada, A. *et al.* Carbonic anhydrase II-positive pancreatic cells are progenitors for both endocrine and exocrine pancreas after birth. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 19915-19919, doi:0805803105 [pii] 10.1073/pnas.0805803105 (2008).

14 Kopinke, D. *et al.* Lineage tracing reveals the dynamic contribution of Hes1+ cells to the developing and adult pancreas. *Development* **138**, 431-441, doi:138/3/431 [pii] 10.1242/dev.053843 (2011).

15 Kopp, J. L. *et al.* Sox9+ ductal cells are multipotent progenitors throughout development but do not produce new endocrine cells in the normal or injured adult pancreas. *Development* **138**, 653-665, doi:138/4/653 [pii] 10.1242/dev.056499 (2011).

16 Solar, M. *et al.* Pancreatic exocrine duct cells give rise to insulin-producing beta cells during embryogenesis but not after birth. *Developmental cell* **17**, 849-860, doi:S1534-5807(09)00477-8 [pii] 10.1016/j.devcel.2009.11.003 (2009).

17 Sancho, R., Gruber, R., Gu, G. & Behrens, A. Loss of Fbw7 reprograms adult pancreatic ductal cells into alpha, delta, and beta cells. *Cell stem cell* **15**, 139-153, doi:10.1016/j.stem.2014.06.019 (2014).

18 Huch, M. *et al.* Unlimited in vitro expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis. *The EMBO journal* **32**, 2708-2721, doi:10.1038/emboj.2013.204 (2013).

19 Lee, J. *et al.* Expansion and conversion of human pancreatic ductal cells into insulin-secreting endocrine cells. *Elife* **2**, e00940, doi:10.7554/eLife.00940 (2013).

20 Cleveland, M. H., Sawyer, J. M., Afelik, S., Jensen, J. & Leach, S. D. Exocrine ontogenies: on the development of pancreatic acinar, ductal and centroacinar cells. *Seminars in cell & developmental biology* **23**, 711-719, doi:10.1016/j.semcdb.2012.06.008 (2012).

21 Beer, R. L., Parsons, M. J. & Rovira, M. Centroacinar cells: At the center of pancreas regeneration. *Developmental biology* **413**, 8-15, doi:10.1016/j.ydbio.2016.02.027 (2016).

22 Delaspre, F. *et al.* Centroacinar Cells Are Progenitors That Contribute to Endocrine Pancreas Regeneration. *Diabetes* **64**, 3499-3509, doi:10.2337/db15-0153 (2015).

23 Balber, A. E. Concise review: aldehyde dehydrogenase bright stem and progenitor cell populations from normal tissues: characteristics, activities, and emerging uses in regenerative medicine. *Stem Cells* **29**, 570-575, doi:10.1002/stem.613 (2011).

24 Zhou, L. *et al.* Identification of cancer-type specific expression patterns for active aldehyde dehydrogenase (ALDH) isoforms in ALDEFLUOR assay. *Cell Biol Toxicol* **35**, 161-177, doi:10.1007/s10565-018-9444-y (2019).

25 Miyamoto, Y. *et al.* Notch mediates TGF alpha-induced changes in epithelial differentiation during pancreatic tumorigenesis. *Cancer cell* **3**, 565-576 (2003).

26 Rovira, M. *et al.* Isolation and characterization of centroacinar/terminal ductal progenitor cells in adult mouse pancreas. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 75-80 (2010).

27 Stanger, B. Z. *et al.* Pten constrains centroacinar cell expansion and malignant transformation in the pancreas. *Cancer cell* **8**, 185-195, doi:10.1016/j.ccr.2005.07.015 (2005).
28 Jackson, B. *et al.* Update on the aldehyde dehydrogenase gene (ALDH) superfamily. *Hum Genomics* **5**, 283-303, doi:7LM2078310429454 [pii] (2011).
29 Ioannou, M. *et al.* ALDH1B1 is a potential stem/progenitor marker for multiple pancreas progenitor pools. *Developmental biology* **374**, 153-163, doi:10.1016/j.ydbio.2012.10.030 (2013).
30 Singh, S. *et al.* Aldehyde Dehydrogenase 1B1 as a Modulator of Pancreatic Adenocarcinoma. *Pancreas* **45**, 117-122, doi:10.1097/MPA.0000000000000542 (2016).
31 Broutier, L. *et al.* Culture and establishment of self-renewing human and mouse adult liver and pancreas 3D organoids and their genetic manipulation. *Nature protocols* **11**, 1724-1743, doi:10.1038/nprot.2016.097 (2016).
32 Hashimoto, D. *et al.* Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. *Immunity* **38**, 792-804, doi:10.1016/j.immuni.2013.04.004 (2013).
33 Anastasiou, V. *et al.* Aldehyde dehydrogenase activity is necessary for beta cell development and functionality in mice. *Diabetologia* **59**, 139-150, doi:10.1007/s00125-015-3784-4 (2016).
34 Giannios, I. *et al.* Protein Methyltransferase Inhibition Decreases Endocrine Specification Through the Upregulation of Aldh1b1 Expression. *Stem Cells* **37**, 640-651, doi:10.1002/stem.2979 (2019).
35 Hutchins, J. R. *et al.* Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science* **328**, 593-599, doi:10.1126/science.1181348 (2010).
36 Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst* **3**, 346-360 e344, doi:10.1016/j.cels.2016.08.011 (2016).
37 Wollny, D. *et al.* Single-Cell Analysis Uncovers Clonal Acinar Cell Heterogeneity in the Adult Pancreas. *Developmental cell* **39**, 289-301, doi:10.1016/j.devcel.2016.10.002 (2016).
38 Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature methods* **10**, 1096-1098, doi:10.1038/nmeth.2639 (2013).
39 Kiselev, V. Y. *et al.* SC3: consensus clustering of single-cell RNA-seq data. *Nature methods* **14**, 483-486, doi:10.1038/nmeth.4236 (2017).
40 Angerer, P. *et al.* destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241-1243, doi:10.1093/bioinformatics/btv715 (2016).

41 Westphalen, C. B. *et al.* Dclk1 Defines Quiescent Pancreatic Progenitors that Promote Injury-Induced Regeneration and Tumorigenesis. *Cell stem cell* **18**, 441-455, doi:10.1016/j.stem.2016.03.016 (2016).

42 Haghverdi, L., Buttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nature methods* **13**, 845-848, doi:10.1038/nmeth.3971 (2016).

43 Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47-52, doi:10.1038/nature16965 (2016).

44 Stellas, D. *et al.* Therapeutic effects of an anti-Myc drug on mouse pancreatic cancer. *J Natl Cancer Inst* **106**, doi:10.1093/jnci/dju320 (2014).

45 Chio, II, Yordanov, G. & Tuveson, D. MAX-ing out MYC: a novel small molecule inhibitor against MYC-dependent tumors. *J Natl Cancer Inst* **106**, doi:10.1093/jnci/dju365 (2014).

46 Stagos, D., Chen, Y., Cantore, M., Jester, J. V. & Vasiliou, V. Corneal aldehyde dehydrogenases: multiple functions and novel nuclear localization. *Brain Res Bull* **81**, 211-218, doi:10.1016/j.brainresbull.2009.08.017 (2010).

47 Sancho, P. *et al.* MYC/PGC-1alpha Balance Determines the Metabolic Phenotype and Plasticity of Pancreatic Cancer Stem Cells. *Cell metabolism* **22**, 590-605, doi:10.1016/j.cmet.2015.08.015 (2015).

48 Xu, X. *et al.* Beta cells can be generated from endogenous progenitors in injured adult mouse pancreas. *Cell* **132**, 197-207 (2008).

49 Basak, O. *et al.* Mapping early fate determination in Lgr5+ crypt stem cells using a novel Ki67-RFP allele. *The EMBO journal* **33**, 2057-2068, doi:10.15252/embj.201488017 (2014).

50 Rezanejad, H. *et al.* Heterogeneity of SOX9 and HNF1beta in Pancreatic Ducts Is Dynamic. *Stem Cell Reports* **10**, 725-738, doi:10.1016/j.stemcr.2018.01.028 (2018).

51 Loomans, C. J. M. *et al.* Expansion of Adult Human Pancreatic Tissue Yields Organoids Harboring Progenitor Cells with Endocrine Differentiation Potential. *Stem Cell Reports* **10**, 712-724, doi:10.1016/j.stemcr.2018.02.005 (2018).

52 Kimura, Y. *et al.* ARID1A Maintains Differentiation of Pancreatic Ductal Cells and Inhibits Development of Pancreatic Ductal Adenocarcinoma in Mice. *Gastroenterology* **155**, 194-209 e192, doi:10.1053/j.gastro.2018.03.039 (2018).

53 von Figura, G., Morris, J. P. t., Wright, C. V. & Hebrok, M. Nr5a2 maintains acinar cell differentiation and constrains oncogenic Kras-mediated pancreatic neoplastic initiation. *Gut* **63**, 656-664, doi:10.1136/gutjnl-2012-304287 (2014).

54 Krah, N. M. *et al.* The acinar differentiation determinant PTF1A inhibits initiation of pancreatic ductal adenocarcinoma. *Elife* **4**, doi:10.7554/eLife.07125 (2015).

55 Shi, G. *et al.* Loss of the acinar-restricted transcription factor Mist1 accelerates Kras-induced pancreatic intraepithelial neoplasia. *Gastroenterology* **136**, 1368-1378, doi:10.1053/j.gastro.2008.12.066 (2009).
56 Guerra, C. *et al.* Chronic pancreatitis is essential for induction of pancreatic ductal adenocarcinoma by K-Ras oncogenes in adult mice. *Cancer cell* **11**, 291-302, doi:S1535-6108(07)00027-X [pii]
10.1016/j.ccr.2007.01.012 (2007).
57 Kopp, J. L. *et al.* Loss of Pten and Activation of Kras Synergistically Induce Formation of Intraductal Papillary Mucinous Neoplasia From Pancreatic Ductal Cells in Mice. *Gastroenterology* **154**, 1509-1523 e1505, doi:10.1053/j.gastro.2017.12.007 (2018).
58 Fox, R. G. *et al.* Image-based detection and targeting of therapy resistance in pancreatic adenocarcinoma. *Nature* **534**, 407-411, doi:10.1038/nature17988 (2016).
59 Kleeff, J. *et al.* Pancreatic cancer. *Nat Rev Dis Primers* **2**, 16022, doi:10.1038/nrdp.2016.22 (2016).
60 Saborowski, M. *et al.* A modular and flexible ESC-based mouse model of pancreatic cancer. *Genes & development* **28**, 85-97, doi:10.1101/gad.232082.113 (2014).
61 Ishizawa, K. *et al.* Tumor-initiating cells are rare in many human tumors. *Cell stem cell* **7**, 279-282, doi:10.1016/j.stem.2010.08.009 (2010).
62 Kim, M. P. *et al.* ALDH activity selectively defines an enhanced tumor-initiating cell population relative to CD133 expression in human pancreatic adenocarcinoma. *PloS one* **6**, e20636, doi:10.1371/journal.pone.0020636 (2011).
63 Liou, G. Y. *et al.* Mutant KRas-Induced Mitochondrial Oxidative Stress in Acinar Cells Upregulates EGFR Signaling to Drive Formation of Pancreatic Precancerous Lesions. *Cell reports* **14**, 2325-2336, doi:10.1016/j.celrep.2016.02.029 (2016).
64 Singh, S. *et al.* Aldehyde dehydrogenases in cellular responses to oxidative/electrophilic stress. *Free Radic Biol Med* **56**, 89-101, doi:10.1016/j.freeradbiomed.2012.11.010 (2013).

FIGURE LEGENDS

Figure 1: Co-expression of Aldh1b1 with differentiation and progenitor markers. (A-F) Immunofluorescence showed complete co-expression of Aldh1b1 with E-cadherin (A) and partial co-expression with Sox9 (B), PNA (C), DBA (D), Prom1 (E) and Amylase (F). Scale bars : 50 µm for main figures, 20 µm for insets.

Figure 2: *Aldh1b1* function is necessary for adult pancreas organoid formation. (A-E) Among FACS separated adult pancreatic cells (A) only Aldh$^+$ cells form organoids (B) which maintain expression of Aldh1b1 and E-cadherin (C) while expressing Sox9, Krt19 (D) and Pdx1 but not insulin (E). (F-J) Tamoxifen administration to *Aldh1b1*^{CreERT2}/*ROSA26*^{LSLtdTomato} double homozygotes (F) labels efficiently Aldh1b1$^+$ cells (arrows in G). All tdTomato$^+$ cells are E-cadherin$^+$ (H), many are Hes1$^+$ (I) and some are clearly centroacinar (J). (K-O) Among FACS fractionated adult pancreatic cells (K) only tdTomato$^+$ cells generate organoids (L, M). Growth rates of Aldh$^+$ (N) and tdTomato$^+$ organoids (O) are very similar (distinct symbols indicate independent experiments). (P-R) Deletion of *Aldh1b1* initiated *in vivo* by tamoxifen administration and completed in culture in the presence of 4-OHT after FACS isolation of Aldh$^+$ cells (P) showed that *Aldh1b1* null cells fail to generate self-renewing organoids (Q) while control cells grown *in vitro* in the absence of 4-OHT continue to grow (R) (distinct symbols indicate independent experiments). Scale bars B, L, M, Q 500 µm; C-E, G-I 50 µm; J 25 µm.

Figure 3: The progeny of *Aldh1b1* expressing cells contribute to all three pancreatic lineages. (A) *Aldh1b1*^{CreERT2}/*ROSA26*^{LSLtdTomato} double heterozygotes were treated with tamoxifen and analyzed one day later (8 weeks) as well as at 12 and 24 weeks of age. (B-G) tdTomato$^+$ progeny became C-pep$^+$ (C), Gcg$^+$ (C) or Som$^+$ (D) cells, acinar cells (E, F) or duct cells (G). (H-M) tdTomato$^+$ cells (H), tdTomato$^+$ acinar cells (I), tdTomato$^+$ duct cells (J), tdTomato$^+$ β-cells (I) and the number of islets containing tdTomato$^+$ β cells (M), increase over time. The contribution of tdTomato$^+$ cells to α and δ cells was quantified at 24 weeks (L). (N) Percentage of Aldh1b1$^+$ or all pancreatic cells that were Ki67$^+$ at 8 and 24 weeks of age. Individual data points for each time point represent animals analyzed. Scale bars: 50 µm. In (G-K) the mean (wide horizontal bar)

and the standard error of the mean (SEM) (narrow horizontal bars) are shown). * $p < 0.05$, *** $p < 0.001$.

Figure 4: Single cell RNA Seq of *Aldh1b1* expressing cells and their early progeny. (A) *Aldh1b1*^{CreERT2}/*ROSA26*^{LSLtdTomato} double heterozygotes were treated with tamoxifen and labeled cells were FACS isolated one day (early cells) or one month later (late cells) and processed for Smart-Seq2 sequencing and computational analysis. (B, C) Early or late cells are depicted on the diffusion map as red or grey dots respectively (B) and molecular marker analysis revealed the identity of the five identified clusters (C). (D-G) Mapping of the aggregate acinar (D), endocrine (E), duct (F) and progenitor (G) gene expression values. (H) Diffusion pseudotime suggested two differentiation pathways (grey line) with a single branching point. Numbers on the axes of all diffusion maps (B-H) represent diffusion coefficients (DC) and the grey line (H) represents differentiation pathways. Values on the gene expression heat maps (D-G) correspond to the aggregate $\log_{10}$(sequencing counts per million) values.

Figure 5: *Aldh1b1* is strongly upregulated in mouse PDAC and necessary for its initiation. (A) *Kras* is preferentially expressed in progenitor cells. Numbers on the axes represent diffusion coefficients and the values of the heat map correspond to the $\log_{10}$(normalized sequencing counts) values. (B) Immunofluorescence for Kras showed co-localization with tdTomato$^+$ cells. (C, D) *Aldh1b1* expression is dramatically upregulated in the *Pdx1*^{Cre}/*Actb*^{LSLKras*} PDAC model at both the mRNA (microarray gene expression analysis) (C) and protein levels as detected by immunofluorescence (D). E, Kaplan-Meier mouse survival curves showed that *Tg*$^{Pdx1\text{-}Cre}$/*Actb*^{LSLKras*} mice had a median survival rate (T_{50}) of 42 days (red line). In contrast, *Kras** was unable to exert its oncogenic action in *Tg*^{Pdx1Cre}/*Actb*^{LSLKras*}/*Aldh1b1*^{null} mice (blue line) (log-rank $p<0.0001$). From a cohort of twelve animals, six were sacrificed at the time points indicated on the graph and histological analysis revealed that they were tumor and PanIN free. The remaining six animals were healthy and tumor free until the age of twelve months. (F-H) Normal pancreas morphology in *Actb*^{LSLKras*} animals (F) and invasive PDAC in the *Pdx1*^{Cre}/*Actb*^{LSLKras*} PDAC model at P15 (G) which never appears in absence of functional *Aldh1b1* (H). (I-K) Immunofluorescence showed that induction of Kras* (E, F) is not affected in the absence of *Aldh1b1* (F). Scale bars: 20 um (B), 50

um (D), 50 um (F-H), 60 um (I-K). In (C) the mean (wide horizontal bar) and the SEM (narrow horizontal bar) are shown.

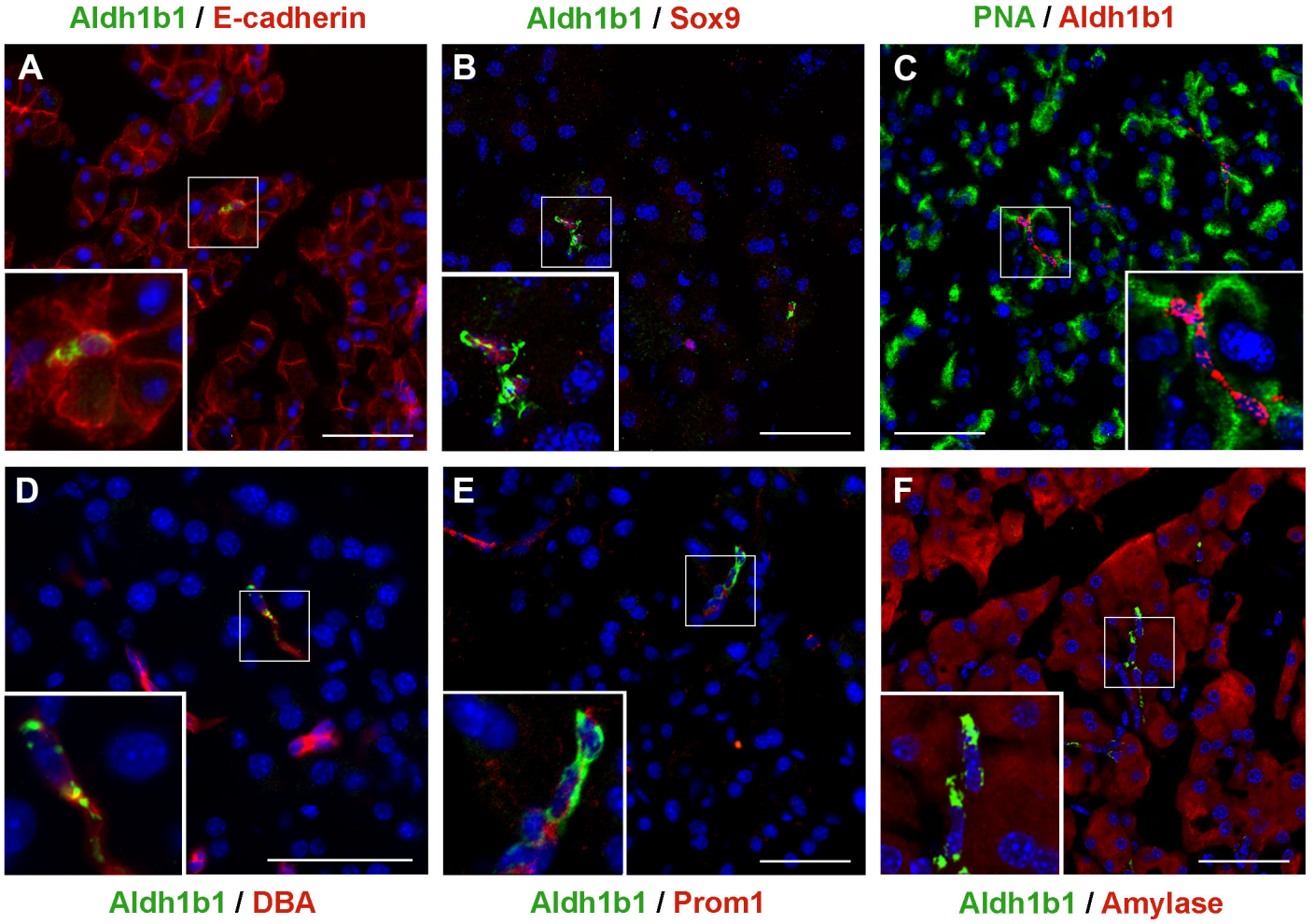
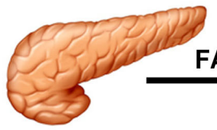

Aldh1b1 / E-cadherin
Aldh1b1 / Sox9
PNA / Aldh1b1
A
B
C
D
E
F
Aldh1b1 / DBA
Aldh1b1 / Prom1
Aldh1b1 / Amylase

Figure 1

A



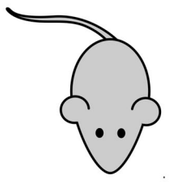
FACS

Aldh⁺ cells → organoids

Aldh⁻ cells → no organoids

wild type

F



tamoxifen

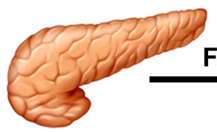
5 days

1 day

immunofluorescence or FACS

Aldh1b1 CreERT2 / ROSA26 LSLtdTomato

K



FACS

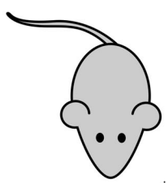
tdTomato⁺ → organoids

Aldh⁺ / tdTomato⁻ → very few organoids

Aldh⁻ / tdTomato⁻ → no organoids

Aldh1b1 CreERT2 / ROSA26 LSLtdTomato

P



+ tamoxifen

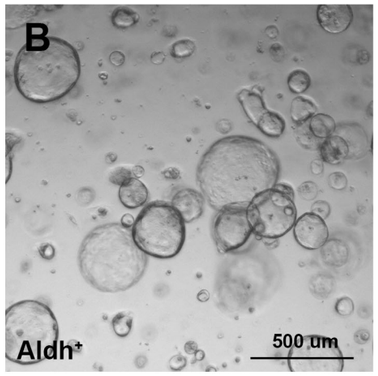
5 days

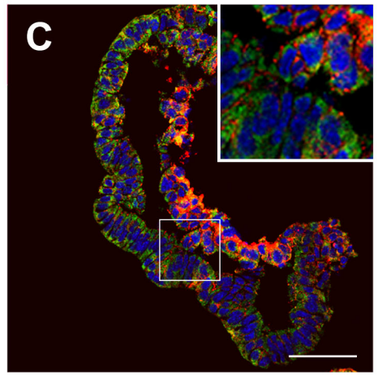
FACS

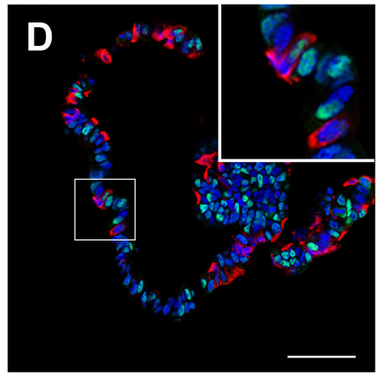
± 4 OH-tamoxifen

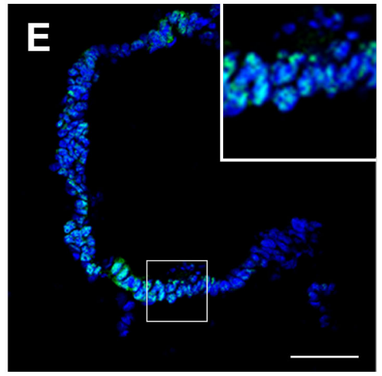
organoid culture

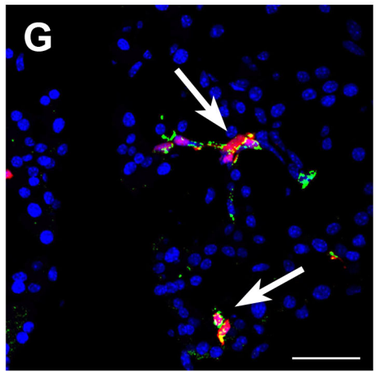
ROSA26 CreERT2 / Aldh1b1 fl/fl

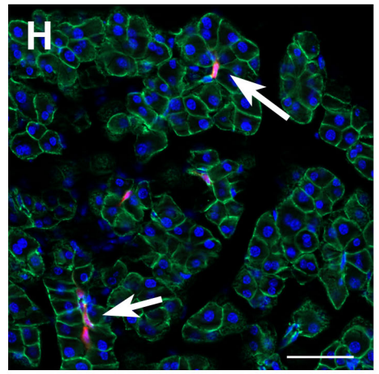


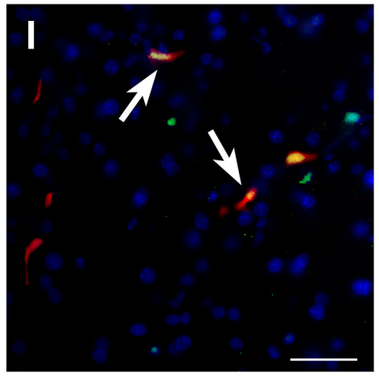


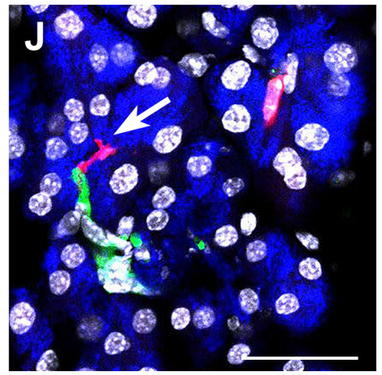


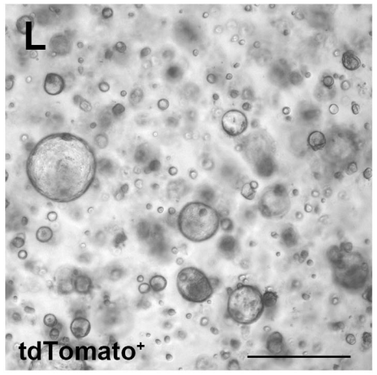


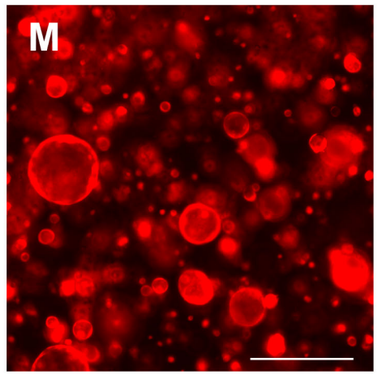


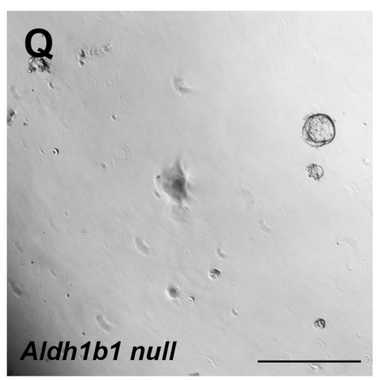


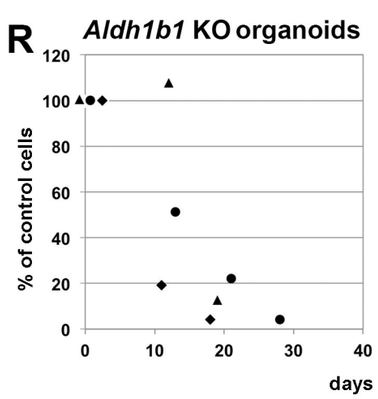


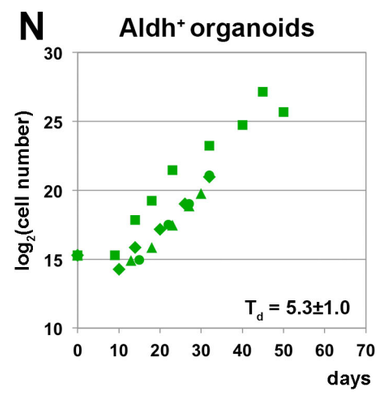












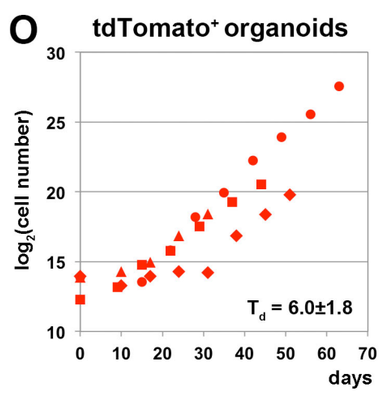


Figure 2

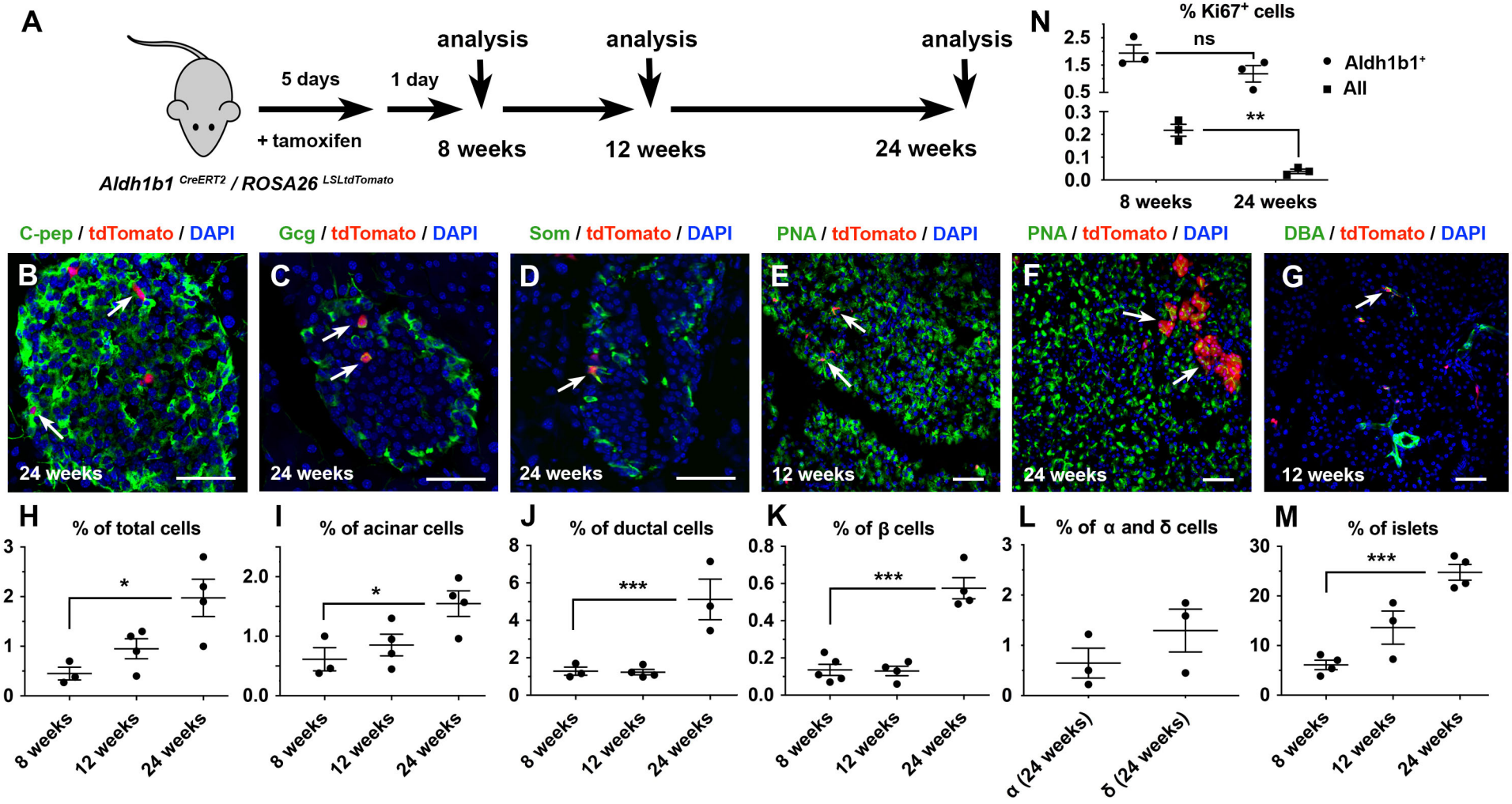

A
5 days
+ tamoxifen
1 day
analysis
8 weeks
analysis
12 weeks
analysis
24 weeks
Aldh1b1 CreERT2 / ROSA26 LSLtdTomato
N
% Ki67+ cells
ns
**
Aldh1b1+
All
8 weeks
24 weeks
C-pep / tdTomato / DAPI
Gcg / tdTomato / DAPI
Som / tdTomato / DAPI
PNA / tdTomato / DAPI
PNA / tdTomato / DAPI
DBA / tdTomato / DAPI
B
C
D
E
F
G
24 weeks
24 weeks
24 weeks
12 weeks
24 weeks
12 weeks
H
% of total cells
I
% of acinar cells
J
% of ductal cells
K
% of β cells
L
% of α and δ cells
M
% of islets
*
*
***
***
***
8 weeks
12 weeks
24 weeks
α (24 weeks)
δ (24 weeks)

Figure 3

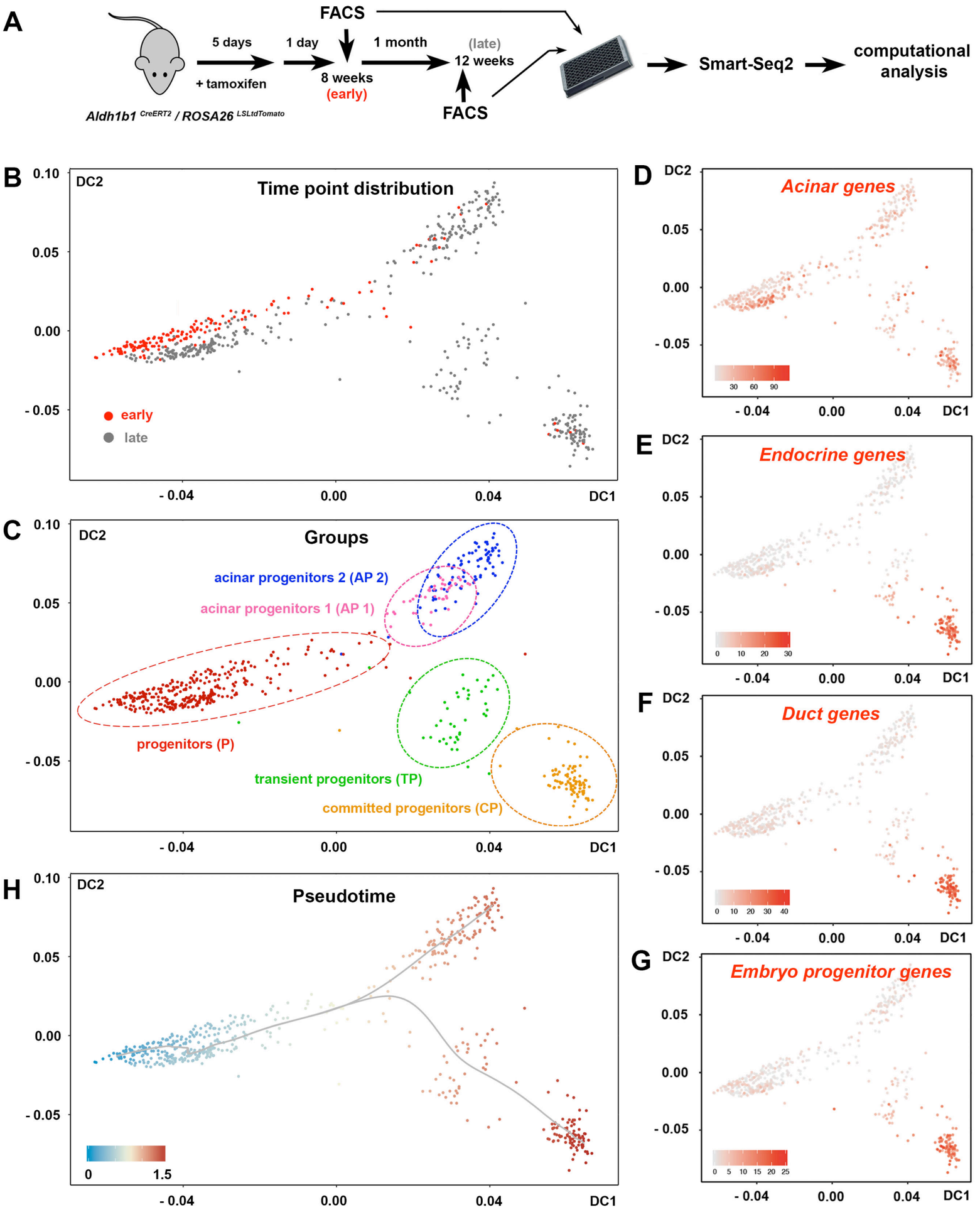

A
FACS
5 days
+ tamoxifen
1 day
1 month
(late)
12 weeks
8 weeks
(early)
FACS
Aldh1b1 CreERT2 / ROSA26 LSLtdTomato
Smart-Seq2
computational analysis
B
Time point distribution
DC2
DC1
early
late
C
Groups
acinar progenitors 2 (AP 2)
acinar progenitors 1 (AP 1)
progenitors (P)
transient progenitors (TP)
committed progenitors (CP)
H
Pseudotime
D
Acinar genes
E
Endocrine genes
F
Duct genes
G
Embryo progenitor genes

Figure 4

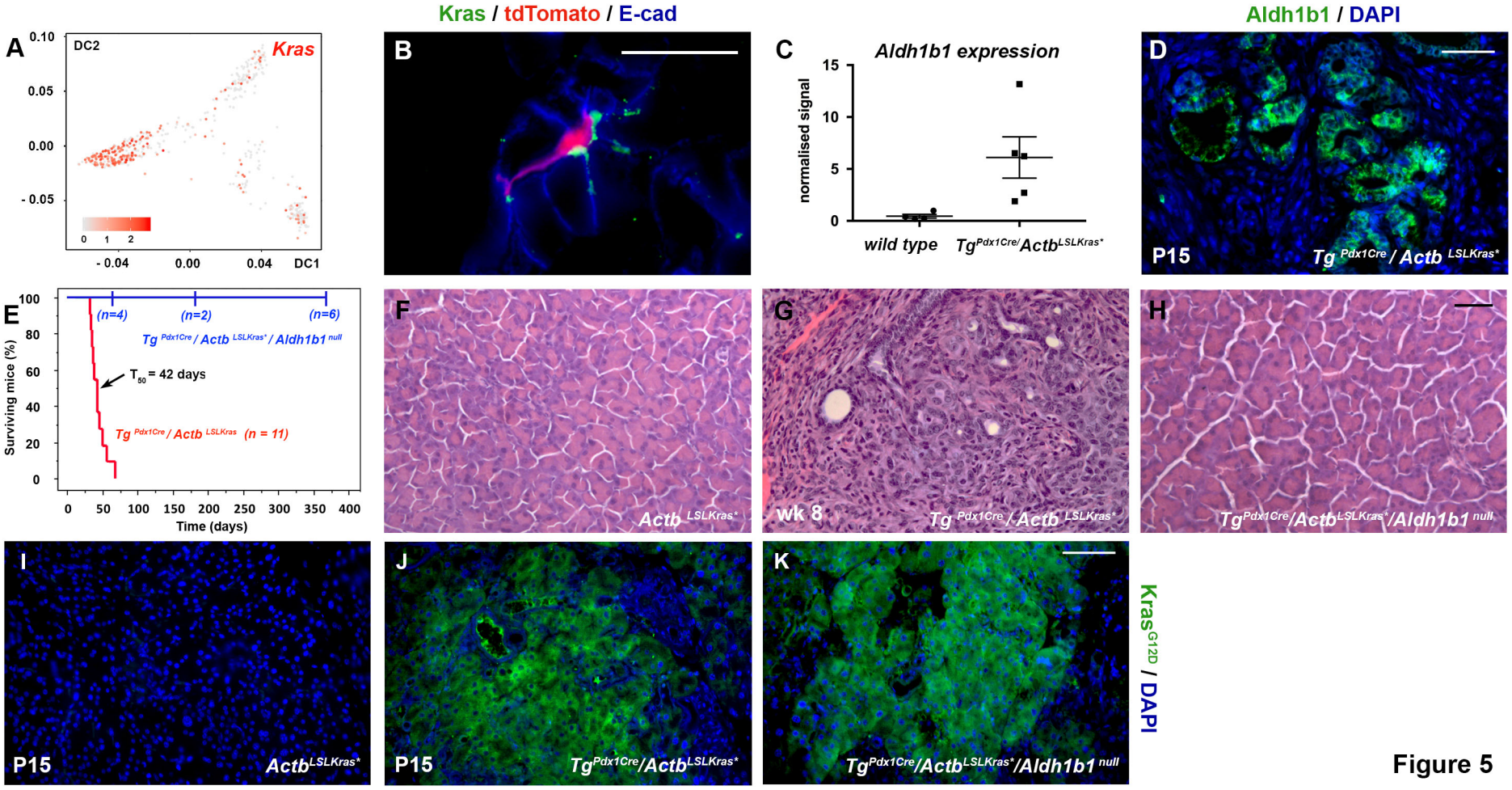

Kras / tdTomato / E-cad
Aldh1b1 / DAPI
A
DC2
Kras
0.10
0.05
0.00
- 0.05
0 1 2
- 0.04
0.00
0.04
DC1
B
C
Aldh1b1 expression
normalised signal
15
10
5
0
wild type
Tg Pdx1Cre/Actb LSLKras*
D
P15
Tg Pdx1Cre / Actb LSLKras*
E
Surviving mice (%)
100
80
60
40
20
0
(n=4)
(n=2)
(n=6)
Tg Pdx1Cre / Actb LSLKras* / Aldh1b1 null
T50 = 42 days
Tg Pdx1Cre / Actb LSLKras (n = 11)
0 50 100 150 200 250 300 350 400
Time (days)
F
Actb LSLKras*
G
wk 8
Tg Pdx1Cre / Actb LSLKras*
H
Tg Pdx1Cre/Actb LSLKras*/Aldh1b1 null
I
P15
Actb LSLKras*
J
P15
Tg Pdx1Cre/Actb LSLKras*
K
Tg Pdx1Cre/Actb LSLKras*/Aldh1b1 null
KrasG12D / DAPI

Figure 5